<table>
<thead>
<tr>
<th>Title</th>
<th>Assessment of environmental compliance of waterbodies through integration of monitoring and modelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>McGovern, Joseph Vincent</td>
</tr>
<tr>
<td>Publication Date</td>
<td>2016-09-30</td>
</tr>
<tr>
<td>Item record</td>
<td><a href="http://hdl.handle.net/10379/6270">http://hdl.handle.net/10379/6270</a></td>
</tr>
</tbody>
</table>
ASSESSMENT OF ENVIRONMENTAL COMPLIANCE OF WATERBODIES THROUGH INTEGRATION OF MONITORING AND MODELLING

BY JOSEPH V. McGOVERN

DISSERTATION SUBMITTED TO NUI GALWAY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE Degree OF PhD

Department of Civil Engineering
College of Engineering and Informatics
September 2016
Declaration

I declare that this dissertation, in whole or in part, has not been submitted to any University as an exercise for a degree. I further declare that, except where reference is given, the work is entirely my own.

Signed:

__________________
Joseph Vincent McGovern           September 2016
Abstract

Design of water quality monitoring networks in estuarine and coastal waters is hampered by a lack of a clear methodology. Monitoring networks are designed on an ad-hoc basis supported by experts in the area. Guidance on the design of monitoring networks focuses on inland surface and ground water, with little thought given to the highly dynamic estuarine environment. Monitoring data from the marine environment may be used for assessment purposes or in the detection of an emerging trend. Numerical models may be used to assist in the understanding of underlying processes and predict the outcome of management options. However, high resolution numerical models are not directly used as a proxy for monitoring data.

Entropy analysis theory has been applied to monitoring data and numerical model water quality simulation data at Cork Harbour in order to prioritise the existing monitoring network and establish the optimum monitoring network with respect to spatial and temporal coverage for Chlorophyll-a, molybdate reactive phosphorus and dissolved inorganic nitrogen. The same method has been applied to rank all 30x30m grid cells used in the numerical model. The optimised monitoring networks returned the same trophic status for each of the seven water bodies in Cork Harbour as were returned by the current monitoring programme adopted by the Irish EPA.

To address any perceived shortcoming in redesigning monitoring networks and in order to utilise a calibrated and validated numerical model to supplement the monitoring network, the Bayesian Maximum Entropy method of geostatistical estimation was tested on the proposed optimised monitoring networks. Estimates of the timeseries at removed monitoring points were generated using monitoring data from the retained monitoring network in addition to numerical model outputs. Using the water quality data estimates and monitoring data from the retained network yielded the same trophic compliance status that would be returned were all monitoring points retained. Leave-one-out cross validation of daily averaged monitoring data proved that BME or kriging could be adopted to derive probability density functions of values across Cork Harbour in each constituent waterbody with a high level of confidence.
A framework proposal was suggested whereby a reciprocal arrangement could be adopted of improvement of numerical model accuracy would assist in the ongoing optimisation of the monitoring network, the results of which would be used to re-calibrate and validate the numerical model on an ongoing basis as and when new information would come to light.
Acknowledgements

Thank you to my friends and family for the support and encouragement throughout my endeavours.

I would like to sincerely thank Professor Padraic O’Donoghue for granting me the opportunity to undertake this research in NUI Galway.

A special thanks to Professor Michael Hartnett for his support, advice and input throughout our many discussions relating to my studies. He has granted me a great opportunity to engage in a unique area of research which I hope to explore further and expand upon in the years to come. Thank you also to Dr. Stephen Nash for his support and assistance throughout my studies.

Thank you to Marc L. Serre in the University of North Carolina at Chapel Hill for allowing me to build upon the previous work carried out by the BMElab research group by allowing me to use the BMElib of MATLAB functions.

I would like to thank the EPA for the funds that were provided by them through the STRIVE programme.

Most importantly, thank you to my girlfriend Aoife for her relentless support since she came into my life.
Table of contents

Declaration........................................................................................................................................ii

Abstract......................................................................................................................................... iii

Acknowledgements...................................................................................................................... v

Table of contents ........................................................................................................................ vi

List of figures.................................................................................................................................. xii

List of tables................................................................................................................................... xxi

Mathematical Notations ................................................................................................................ xxv

CHAPTER 1. INTRODUCTION ................................................................................................. 1

1.1. Estuarine and coastal water quality in Ireland .................................................................... 1

1.2. Design of monitoring networks ......................................................................................... 2

1.3. Combined use of water quality monitoring and modelling ........................................... 3

1.4. Aims and objectives ............................................................................................................ 4

1.5. Thesis overview .................................................................................................................. 5

CHAPTER 2. LITERATURE REVIEW .................................................................................... 7

2.1. Introduction............................................................................................................................... 7

2.2. Water quality legislation ...................................................................................................... 8

2.2.1. European Union................................................................................................................ 8

2.2.2. Irish Coastal and Transitional Waters WFD Monitoring............................................. 10
<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.2.1. Surveillance Monitoring</td>
<td>11</td>
</tr>
<tr>
<td>2.2.2.2. Operational Monitoring</td>
<td>12</td>
</tr>
<tr>
<td>2.2.2.3. Investigative Monitoring</td>
<td>12</td>
</tr>
<tr>
<td>2.2.2.4. Sampling frequency</td>
<td>13</td>
</tr>
<tr>
<td>2.2.3. Irish estuarine and coastal water quality</td>
<td>13</td>
</tr>
<tr>
<td>2.2.3.1. TSAS</td>
<td>15</td>
</tr>
<tr>
<td>2.2.4. OSPAR</td>
<td>17</td>
</tr>
<tr>
<td>2.2.5. UNEP MEDPOL</td>
<td>19</td>
</tr>
<tr>
<td>2.3. Monitoring practice &amp; guidelines</td>
<td>20</td>
</tr>
<tr>
<td>2.3.1. Statistical considerations</td>
<td>20</td>
</tr>
<tr>
<td>2.3.2. US monitoring network guidance</td>
<td>21</td>
</tr>
<tr>
<td>2.3.3. Functional data analysis</td>
<td>30</td>
</tr>
<tr>
<td>2.3.4. Discriminant analysis</td>
<td>31</td>
</tr>
<tr>
<td>2.3.5. Matter element analysis</td>
<td>32</td>
</tr>
<tr>
<td>2.3.6. Principal component analysis</td>
<td>34</td>
</tr>
<tr>
<td>2.3.7. Other Research Methods for monitoring optimisation</td>
<td>36</td>
</tr>
<tr>
<td>2.3.8. WFD monitoring network guidance</td>
<td>38</td>
</tr>
<tr>
<td>2.3.9. JEEP EC Tidal Monitoring Manual</td>
<td>40</td>
</tr>
<tr>
<td>2.4. Combined use of modelling and monitoring for environmental compliance</td>
<td>41</td>
</tr>
<tr>
<td>2.5. Information theory and Entropy methods</td>
<td>45</td>
</tr>
<tr>
<td>2.5.1. Application of Entropy methods to monitoring network design</td>
<td>50</td>
</tr>
<tr>
<td>2.6. Geostatistics and BME</td>
<td>52</td>
</tr>
</tbody>
</table>
CHAPTER 4. ENTROPY METHODS

4.2. Entropy analysis ................................................................................................................. 102

4.2.1. Spatial Entropy Analysis .............................................................................................. 103

4.2.2. Spatial – Temporal Entropy Analysis ........................................................................... 106

4.3. BME analysis ...................................................................................................................... 108

4.3.1. Solution of BME mean estimate and estimation error variance - Example ............... 115

4.3.2. BME algorithm ............................................................................................................. 118

4.4. Conclusions ......................................................................................................................... 123

CHAPTER 5. OPTIMISATION OF WATER QUALITY MONITORING PROGRAMMES VIA ENTROPY METHODS .................................................. 124

5.1. Introduction ........................................................................................................................ 124

5.2. Case Study Description ...................................................................................................... 125

5.3. Comparison of model output and EPA data via marginal entropy ............................... 128

5.4. Existing monitoring network ............................................................................................ 134

5.4.1. Spatial analysis ............................................................................................................. 134

5.4.2. Spatial/Temporal analysis ............................................................................................ 145

5.5. Ab initio spatial monitoring design ................................................................................... 154

5.5.1. Monitoring design by salinity class .............................................................................. 154

5.5.1.1. Freshwater ............................................................................................................ 155

5.5.1.2. Intermediate salinity ............................................................................................. 158

5.5.1.3. Seawater ............................................................................................................... 161

5.5.2. Monitoring design by TSAS compliance level ................................................................ 163

5.5.2.1. Intermediate waters ............................................................................................... 164
5.5.2.1.1. Eutrophic................................................................. 164
5.5.2.1.2. Potentially eutrophic ............................................. 167
5.5.2.1.3. Intermediate trophic status................................. 170
5.5.2.2. Full salinity waters.................................................. 172
5.5.2.2.1. Eutrophic................................................................. 172
5.5.2.2.2. Potentially eutrophic ............................................. 175
5.5.2.2.3. Intermediate trophic status................................. 177
5.6. Conclusions........................................................................ 180

CHAPTER 6. BME AND GEOSPATIAL ESTIMATION.............. 185
6.1. Introduction...................................................................... 185
6.2. EPA monitoring data – ‘Hard Data’................................. 186
6.2.1. Spatial and temporal data trends............................... 191
6.3. Derivation of soft data..................................................... 202
6.4. BME and kriging estimation.......................................... 207
6.4.1. Standard estimation.................................................... 208
6.4.2. Log estimation........................................................... 217
6.4.3. Comparison of overall accuracy of lognormal and standard interpolation.... 223
6.4.4. Comparison of water quality assessment outcomes........ 228
6.5. Quasi-stationary BME and kriging................................. 230
6.6. Summary and conclusions............................................. 242

CHAPTER 7. SUMMARY AND CONCLUSIONS..................... 247
7.1. Summary......................................................................... 247
7.2. Critique of research output ................................................................. 250

7.2.1. Entropy analysis .............................................................................. 251

7.2.2. BME and kriging estimation ............................................................ 252

7.3. Conclusions ...................................................................................... 255

7.3.1. Entropy methods ............................................................................ 255

7.3.2. BME ................................................................................................ 258

7.4. Future Work ...................................................................................... 260

APPENDIX 1. DIVAST SOLUTION ............................................................... 263

APPENDIX 2. KRIGING AND BME DERIVATION .................................... 269

Simple Kriging .......................................................................................... 269

BME ........................................................................................................... 273

BME Mode estimate .................................................................................. 284

BME Mean estimate .................................................................................. 286

BME posterior PDF - variance and standard deviation ............................... 288

REFERENCES .......................................................................................... 289
List of figures

Figure 2.1 Identification of homogeneous river reaches by matter-element analysis, taken from Chen et al. (2012). Deviation of monitoring points from the straight line joining the top left and bottom right corner indicates the requirement for extra monitoring [Image removed for copyright reasons] ................................................................................................................................................ 33

Figure 3.1 A zero-dimensional mass transport model representing the net exchange of mass, influenced by riverine inflow, and exchange at the tidal boundary; modified from Lung (1993) [Image removed for copyright reasons] ............................................................................................................ 67

Figure 3.2 1D hydrodynamic model diagram; discharge Q and water elevation $\zeta$ are spatially referenced in one plane. ............................................................................................................................................... 68

Figure 3.3 Velocity profile with depth in (a) a 3-D stratified model and (b) in a shallow, depth integrated model ........................................................................................................................................................................... 72

Figure 3.4 A shallow, well mixed water body (M.W.L. = mean water level) $\zeta =$ water elevation above or below M.W.L., $h =$ water depth below M.W.L. ............................................................................................................... 72

Figure 3.5 The space-staggered grid scheme and ($I,J$) coordinate system, adapted from Nash (2010) [Image removed for copyright reasons] ............................................................................................................................................. 75

Figure 3.6 (a) Sample model grid showing x-direction integration sections and (b) individual section (land cells are grey, wet cells white and open boundary cells blue) (Nash 2010) [Image removed for copyright reasons] ............................................................................................................................................... 78

Figure 3.7 Schematic of the water quality interactions as represented within DIVAST, taken from Nash et al. (2011) [Image removed for copyright reasons] ............................................................................................................... 81

Figure 3.8 Locations of wastewater discharges, rivers, catchments and monitoring points around the greater Cork harbour area ............................................................................................................................................. 88

Figure 3.9: Cork Harbour model domain and calibration/validation locations (Nash et al. 2011) [Image removed for copyright reasons] ............................................................................................................................................... 89

Figure 3.10: (a) Water surface elevation calibration for spring tide at Curraubinny and (b) linear regression analysis of measured and modelled data (Nash et al. 2011) ............................................................................................................................................. 90
Figure 3.11: (a) Water surface elevation validation for neap tide at Passage West and (b) linear regression analysis of measured and modelled data (Nash et al. 2011) [Image removed for copyright reasons] ................................................................................................................................................ 90

Figure 3.12: Dye plume predicted by the calibrated model (concentrations in µg/l) (Nash et al. 2011) [Image removed for copyright reasons] .............................................................................................................. 92

Figure 3.13: Dye plume recorded in the field (Nash et al. 2011) [Image removed for copyright reasons] ................................................................................................................................................ 92

Figure 3.14: Salinity concentrations predicted using 35ppt (Run 1) and 0ppt (Run 2) initial conditions (Nash et al. 2011) [Image removed for copyright reasons] ........................................................................................................ 93

Figure 3.15: Monthly-averaged waterbody temperature (Nash et al. 2011) [Image removed for copyright reasons] ........................................................................................................................................... 94

Figure 3.16: Monthly-averaged light intensity and photoperiod (Nash et al. 2011) [Image removed for copyright reasons] ........................................................................................................................................... 94

Figure 3.17: Model predictions of Chlorophyll_a in Cork Harbour using the model default and site specific light attenuation functions. Results shown for Point D1(Nash et al. 2011) [Image removed for copyright reasons] ........................................................................................................................................... 95

Figure 3.18 Flow chart of DIVAST FORTRAN subroutines .................................................................................................................. 97

Figure 3.19: Predicted Chlorophyll_a concentrations for spring high water, September and validation locations (1-25) (Nash et al. 2011) [Image removed for copyright reasons] ................. 98

Figure 3.20: Remotely sensed Chlorophyll_a for spring high water, September (Nash et al. 2011) [Image removed for copyright reasons] ........................................................................................................................................... 98

Figure 3.21: Linear regression analysis of the modelled and measured Chlorophyll_a data presented in Figures 3.20 and 3.22 respectively at the validation locations shown in Figure 3.20 (Nash et al. 2011) [Image removed for copyright reasons] ........................................................................................................................................... 99

Figure 4.1 Algorithm describing the application of spatial entropy analysis to n_m potential monitoring locations ........................................................................................................................................................................... 105

Figure 4.2 Algorithm describing the procedure to follow to determine the optimum sampling frequency using temporal entropy analysis, based on the optimum monitoring network derived from Figure 4.1 ........................................................................................................................................................................... 107
Figure 4.3 Flow chart of computational implementation of BME via MATLAB.................108

Figure 4.4 Example of BME estimation using adjacent hard data and soft probabilistic data....113

Figure 4.5 Outline of procedure of BME estimation with the relevant MATLAB functions referenced................................................................................................................114

Figure 4.6 BME algorithm as applied to EPA water quality data and DIVAST simulated water quality datasets via MATLAB ...........................................................................................................................................120

Figure 4.7 Data processing algorithm of EPA_EPA_DATA, DIVAST_EPA_DATA, EPAsel ect and DIVASTselect for BME...........................................................................................................................................121

Figure 4.8 Simple kriging algorithm...........................................................................................121

Figure 4.9 Framework for interaction between BME, monitoring network optimisation by entropy analysis and numerical model calibration ...........................................................................................................................................122

Figure 5.1 Map showing the monitoring sites in the Cork Harbour area that lie within the numerical model domain...........................................................................................................................................127

Figure 5.2 Map of 39 EPA monitoring locations in the Cork Harbour area................................131

Figure 5.3 Residence times in the greater Cork Harbour area (Dabrowski 2005) [Image removed for copyright reasons] ...........................................................................................................................................132

Figure 5.4 Comparison of marginal entropy of PDFs of MRP at each monitoring point, derived from monitoring data and simulated water quality data...........................................................................................................................................133

Figure 5.5 Comparison of marginal entropy of PDFs of Chlorophyll_a at each monitoring point, derived from monitoring data and simulated water quality data ...........................................................................................................................................133

Figure 5.6 Comparison of marginal entropy of PDFs of DIN at each monitoring point, derived from monitoring data and simulated water quality data...........................................................................................................................................134

Figure 5.7 Monitoring network size versus the percentage of total monitoring network transinformation for each of the TSAS parameters for a summer and winter water quality model simulation in Cork Harbour...........................................................................................................................................138

Figure 5.8 Monitoring station priority from 1 to 24 for Chlorophyll_a sampling during summer ...........................................................................................................................................139
Figure 5.9 Monitoring station priority from 1 to 24 for Chlorophyll_a sampling during winter 139
Figure 5.10 Monitoring station priority from 1 to 24 for MRP sampling during summer .......... 140
Figure 5.11 Monitoring station priority from 1 to 24 for MRP sampling during winter .......... 141
Figure 5.12 Monitoring station priority from 1 to 24 for DIN sampling during summer .......... 142
Figure 5.13 Monitoring station priority from 1 to 24 for DIN sampling during winter .......... 142
Figure 5.14 Monitoring station priority from 1 to 24 for DO sampling during summer .......... 143
Figure 5.15 Monitoring station priority from 1 to 24 for DO sampling during winter .......... 144
Figure 5.16 Salinity categories in the greater Cork Harbour area .......................................... 155
Figure 5.17 Priority of potential monitoring locations for Chlorophyll_a in freshwater .......... 156
Figure 5.18 Priority of potential monitoring locations for DIN in freshwater ......................... 156
Figure 5.19 Priority of potential monitoring locations for MRP in freshwater ....................... 157
Figure 5.20 Priority of potential monitoring locations for DO sat% in freshwater .................. 157
Figure 5.21 Priority of potential monitoring locations for Chlorophyll_a in brackish waters .... 158
Figure 5.22 Priority of potential monitoring locations for DIN in brackish waters ................. 159
Figure 5.23 Priority of potential monitoring locations for MRP in brackish waters ............... 160
Figure 5.24 Priority of potential monitoring locations for DO sat% in brackish waters ............ 160
Figure 5.25 Priority of potential monitoring locations for Chlorophyll_a in seawater .......... 161
Figure 5.26 Priority of potential monitoring locations for DIN in seawater ......................... 162
Figure 5.27 Priority of potential monitoring locations for MRP in seawater ....................... 162
Figure 5.28 Priority of potential monitoring locations for DO sat% in seawater .................... 163
Figure 5.29 Priority of potential monitoring locations for Chlorophyll_a in brackish eutrophic waters ................................................................................................................................................. 165
Figure 5.30 Ranking of potential monitoring locations for DIN in brackish eutrophic waters...166

Figure 5.31 Priority of potential monitoring locations for MRP in brackish eutrophic waters...166

Figure 5.32 Ranking of potential monitoring locations for DO sat% in brackish eutrophic waters ............................................................................................................................................................167

Figure 5.33 Ranking of potential monitoring locations for Chlorophyll\_a in brackish potentially eutrophic waters .................................................................................................................................168

Figure 5.34 Ranking of potential monitoring locations for DIN in brackish potentially eutrophic waters .................................................................................................................................................168

Figure 5.35 Priority of potential monitoring locations for MRP in brackish potentially eutrophic waters .................................................................................................................................................169

Figure 5.36 Ranking of potential monitoring locations for DO sat% in brackish potentially eutrophic waters ............................................................................................................................................................169

Figure 5.37 Priority of potential monitoring locations for Chlorophyll\_a in brackish intermediate risk waters ............................................................................................................................................................170

Figure 5.38 Priority of potential monitoring locations for DIN in brackish intermediate risk waters ............................................................................................................................................................171

Figure 5.39 Priority of potential monitoring locations for MRP in brackish intermediate risk waters ............................................................................................................................................................171

Figure 5.40 Priority of potential monitoring locations for DO sat% in brackish intermediate risk waters ............................................................................................................................................................172

Figure 5.41 Priority of potential monitoring locations for Chlorophyll\_a in eutrophic seawaters ............................................................................................................................................................173

Figure 5.42 Priority of potential monitoring locations for DIN in eutrophic seawaters..............173

Figure 5.43 Priority of potential monitoring locations for MRP in eutrophic seawaters ............174

Figure 5.44 Priority of potential monitoring locations for DO sat% in eutrophic seawaters ......174

Figure 5.45 Priority of potential monitoring locations for Chlorophyll\_a in potentially eutrophic seawaters ............................................................................................................................................................175
Figure 5.46 Priority of potential monitoring locations for DIN in potentially eutrophic seawaters ........................................................................................................................................... 176

Figure 5.47 Priority of potential monitoring locations for MRP in potentially eutrophic seawaters ........................................................................................................................................... 176

Figure 5.48 Priority of potential monitoring locations for DO sat% in potentially eutrophic seawaters ........................................................................................................................................... 177

Figure 5.49 Priority of potential monitoring locations for Chlorophyll_a in intermediate risk eutrophic seawaters ........................................................................................................................................... 178

Figure 5.50 Priority of potential monitoring locations for DIN in intermediate risk eutrophic seawaters ........................................................................................................................................... 178

Figure 5.51 Priority of potential monitoring locations for MRP in intermediate risk eutrophic seawaters ........................................................................................................................................... 179

Figure 5.52 Priority of potential monitoring locations for DO sat% in intermediate risk eutrophic seawaters ........................................................................................................................................... 179

Figure 6.1 Histograms of Chlorophyll_a data and log transformed Chlorophyll_a data .......... 189

Figure 6.2 Histograms of MRP data and log transformed MRP data ........................................ 190

Figure 6.3 Histograms of DIN data and log transformed DIN data .......................................... 190

Figure 6.4 Temporal trend of standard and log transformed Chlorophyll_a within each waterbody and across Cork Harbour ........................................................................................................................................... 192

Figure 6.5 Temporal trend of standard and log transformed MRP within each waterbody and across Cork Harbour ........................................................................................................................................... 193

Figure 6.6 Temporal trend of standard and log transformed DIN within each waterbody and across Cork Harbour ........................................................................................................................................... 193

Figure 6.7 Spatial trend of Chlorophyll_a throughout Cork Harbour ........................................ 194

Figure 6.8 Spatial trend of MRP throughout Cork Harbour ....................................................... 194

Figure 6.9 Spatial trend of DIN throughout Cork Harbour ....................................................... 195
Figure 6.10 Histograms of residual data after removal of the spatial and temporal mean trend data for standard and log transformed Chlorophyll_a ................................................................. 197

Figure 6.11 Histograms of residual data after removal of the spatial and temporal mean trend data for standard and log transformed MRP ............................................................................. 197

Figure 6.12 Histograms of residual data after removal of the spatial and temporal mean trend data for standard and log transformed DIN ........................................................................... 198

Figure 6.13 Three commonly used covariance models and covariance model nomenclature .... 199

Figure 6.14 Covariance models derived from the residual data from optimised MRP monitoring network with the lowest TST of 25% ................................................................................... 200

Figure 6.15 Covariance models derived from the residual data from the log-transformed Chlorophyll_a monitoring data the optimised monitoring network with the TST of 75% .................. 200

Figure 6.16 Chlorophyll_a monitoring data versus model Chlorophyll_a data .............................. 204

Figure 6.17 MRP monitoring data versus model MRP data .......................................................... 204

Figure 6.18 DIN monitoring data versus model DIN data .......................................................... 205

Figure 6.19 Log transformed Chlorophyll_a monitoring data versus log transformed model Chlorophyll_a data ............................................................................................................. 205

Figure 6.20 Log transformed MRP monitoring data versus log transformed model MRP data .. 206

Figure 6.21 Log transformed DIN monitoring data versus log transformed model DIN data .... 206

Figure 6.22 Estimated timeseries of Chlorophyll_a at monitoring point LE430 using standard BME and monitoring data from the 75% TST monitoring network for Chlorophyll_a using a covariance model derived for the North Channel ......................................................... 211

Figure 6.23 Estimated timeseries of Chlorophyll_a at monitoring point LE430 using standard BME and monitoring data from the 75% TST monitoring network for Chlorophyll_a using a covariance model derived for the whole of Cork Harbour using all available data ...................... 211

Figure 6.24 standard BME estimation of DIN in the lower Lee Estuary using 25% TST monitoring network ....................................................................................................................... 212
Figure 6.25 standard BME estimation of DIN in the lower Lee Estuary using 50% TST monitoring network ............................................................................................................................ 212

Figure 6.26 Estimated timeseries of Chlorophyll_a at monitoring point LE380 using log-kriging and monitoring data from the 25% TST monitoring network for Chlorophyll_a using a single covariance model derived for Cork inner harbour ............................................................................. 218

Figure 6.27 Estimated timeseries of Chlorophyll_a at monitoring point LE380 using standard kriging and monitoring data from the 25% TST monitoring network for Chlorophyll_a using a single covariance model derived for Cork inner harbour ............................................................................. 218

Figure 6.28 Comparison of Chlorophyll_a estimation datasets for all monitoring optimisation schemes and geostatistical estimation combinations ............................................................................. 225

Figure 6.29 Comparison of MRP estimation datasets for all monitoring optimisation schemes and geostatistical estimation combinations ............................................................................. 225

Figure 6.30 Comparison of DIN estimation datasets for all monitoring optimisation schemes and geostatistical estimation combinations ............................................................................. 226

Figure 6.31 Covariance models derived from log-transformed MRP residual data in the lower Lee Estuary ............................................................................................................................................... 232

Figure 6.32 Covariance models derived from log-transformed MRP residual data throughout Cork Harbour ..................................................................................................................................... 233

Figure 6.33 Leave-one-out BME estimation of Chlorophyll_a in Cork Inner harbour using quasi-stationary covariance models for estimation centred around each monitoring point ............................................................................................................................ 233

Figure 6.34 Leave-one-out BME estimation of Chlorophyll_a in the lower Lee Estuary using quasi-stationary covariance models for estimation centred around each monitoring point ............................................................................................................................ 234

Figure 6.35 Leave-one-out log-BME estimation of DIN in the lower Lee Estuary using quasi-stationary covariance models derived from data for all of Cork Harbour ..................................................................................................................................... 234

Figure 6.36 Leave-one-out log-kriging estimates of DIN in Cork inner harbour using quasi-stationary covariance models for estimation centred on each monitoring point ............................................................................................................................ 235

Figure 6.37 Comparison of standard kriging estimates of DIN in Cork inner harbour using quasi-stationary covariance models for estimation centred on each monitoring point ............................................................................................................................ 235
Figure 6.38 Comparison of histograms of datasets generated by carrying out leave-one-out cross validation following 4 estimation procedures using quasi-stationary covariance models derived from data for all of Cork Harbour................................................................. 241

Figure 6.39 Comparison of histograms of datasets generated by carrying out leave-one-out cross validation following estimation procedures using separate quasi-stationary covariance models for each waterbody in Cork Harbour........................................................................................................... 241
List of tables

Table 2.1 TSAS water quality compliance criteria ................................................................. 16

Table 3.1 Pre-Carrigrennan numerical model industrial and WWTP volumetric flow rates and their licence discharge standards .................................................................................. 86

Table 3.2 Pre-Carrigrennan numerical model freshwater discharge flow rates and salinity and Chlorophyll_a concentrations for the Lee, Glashaboy, Owenbuidhe and Owenacurra rivers and the three catchments indicated in Figure 3.8 ........................................................................... 87

Table 3.3 Pre-Carrigrennan numerical model inflow characteristics for rivers and catchments: nutrients, DO and BOD ........................................................................................................ 88

Table 3.4 Hydrodynamic constants and coefficients from Falconer et al., 2001 used for calibration (Nash et al. 2011). ................................................................................................................. 91

Table 3.5 Parameters used in nutrient model formulations and associated values (Nash et al. 2011) [Figure removed for copyright reasons] [Sources: 1-Brown and Barnwell (1985), 2-Bowie et al. (1985), 3-Chapra (1997), 4-Falconer et al. (2001)] .......................................................... 96

Table 4.1 Hard and soft data for BME and simple kriging estimation .................................. 116

Table 4.2 Cross validation results in BME estimation example ........................................ 117

Table 5.1 TSAS trophic compliance reporting by the EPA for Cork Harbour from 1995 – 2009 ......................................................................................................................... 129

Table 5.2 Rank of each of the 39 EPA monitoring stations in Cork Harbour by marginal entropy values .......................................................................................................................... 130

Table 5.3 Spatial optimisation of the DIN monitoring network from a 1 year numerical model simulation .......................................................................................................................... 136

Table 5.4 Summary of monitoring station spatial priority in Cork Harbour ....................... 137

Table 5.5 Transinformation between sampling Chlorophyll_a at LE330 on a monthly basis for a one year period, and each additional monitoring point from LE610 to LE530 at multiples of the monthly frequency .................................................................................................................. 147
Table 5.6 Transinformation between sampling MRP at LE150 on a monthly basis during summer, and each additional monitoring point from LE220 to LE520 at multiples of the monthly frequency 148

Table 5.7 The optimum sampling frequency for all management options ................................. 149

Table 5.8 The optimum number of sampling events required per monitoring location for all management options .................................................................................................................. 150

Table 5.9 The optimum number of monitoring locations for all management options .......... 150

Table 5.10 The monitoring effort required to implement each optimum monitoring network size ........................................................................................................................................ 151

Table 5.11 Comparison of ultimate monitoring network transinformation at weekly and monthly resolution ......................................................................................................................... 151

Table 5.12 Comparison of the level of compliance of the proposed monitoring regime versus the existing monitoring regime; sampling is carried out on data generated from a 3 year model simulation. The corresponding EPA reported compliance status is also presented for comparison .............................. 153

Table 6.1 Comparison of skewness and kurtosis for water quality parameters before and after log-transformation ................................................................. 189

Table 6.2 Skewness and kurtosis data for the histograms contained in Figures 6.10 – 6.12 196

Table 6.3 Covariance model parameters for log transformed MRP for each of the 3 monitoring networks ............................................................................................................. 201

Table 6.4 ME of kriging cross validation of monitoring data for 3 monitoring scenarios ......... 213

Table 6.5 ME of BME cross validation of monitoring data for 3 monitoring scenarios ........... 213

Table 6.6 MAE of kriging cross validation of monitoring data for 3 monitoring scenarios ...... 214

Table 6.7 MAE of BME cross validation of monitoring data for 3 monitoring scenarios ....... 214

Table 6.8 RMSE of kriging cross validation of monitoring data for 3 monitoring scenarios ...... 215

Table 6.9 RMSE of BME cross validation of monitoring data for 3 monitoring scenarios ....... 215

Table 6.10 R² values of kriging cross validation of monitoring data for 3 monitoring scenarios 216
Table 6.11 $R^2$ values of BME cross validation of monitoring data for 3 monitoring scenarios ................................................................................................................................. 216

Table 6.12 Mean Error (M.E.) of log-kriging cross validation of EPA monitoring data for 3 different monitoring scenarios ........................................................................................................................................ 219

Table 6.13 M.E. of log-BME cross validation of EPA monitoring data for 3 different monitoring scenarios ........................................................................................................................................ 219

Table 6.14 Mean Absolute Error (M.A.E.) of log-kriging cross validation of EPA monitoring data for 3 different monitoring scenarios ........................................................................................................................................ 220

Table 6.15 M.A.E. of log-BME cross validation of EPA monitoring data for 3 different monitoring scenarios ........................................................................................................................................ 220

Table 6.16 Root Mean Squared Error (R.M.S.E.) of log-kriging cross validation of EPA monitoring data for 3 different monitoring scenarios ........................................................................................................................................ 221

Table 6.17 R.M.S.E. of log-BME cross validation of EPA monitoring data for 3 different monitoring scenarios ........................................................................................................................................ 221

Table 6.18 $R^2$ values of log-kriging cross validation of EPA monitoring data for 3 different monitoring scenarios ........................................................................................................................................ 222

Table 6.19 $R^2$ values of log-BME cross validation of EPA monitoring data for 3 different monitoring scenarios ........................................................................................................................................ 222

Table 6.20 Skewness and kurtosis parameters for combinations of kriging and BME of standard and log transformed monitoring data ........................................................................................................................................ 227

Table 6.21 Comparison of the optimum geostatistical interpolation technique for each water quality parameter, waterbody and monitoring network size ........................................................................................................................................ 227

Table 6.22 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout the lower Lee Estuary ........................................................................................................................................ 228

Table 6.23 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout the Glashaboy Estuary ........................................................................................................................................ 228
Table 6.24 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout the North Channel ................................................................. 229

Table 6.25 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout Lough Mahon .............................................................................................................. 229

Table 6.26 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout the Owenacurra Estuary .................................................................................................................. 229

Table 6.27 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout Cork inner harbour .................................................................................................................. 229

Table 6.28 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout Cork outer harbour .................................................................................................................. 230

Table 6.29 \( R^2 \) values of leave-one-out BME cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from the entire monitoring dataset ........................................................................................................ 236

Table 6.30 \( R^2 \) values of leave-one-out BME cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from monitoring data for each waterbody ........................................................................................................ 236

Table 6.31 \( R^2 \) values of leave-one-out log-BME cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from the entire monitoring dataset ........................................................................................................ 237

Table 6.32 \( R^2 \) values of leave-one-out log-BME cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from monitoring data for each waterbody ........................................................................................................ 237

Table 6.33 Summary of the optimum estimation procedures of leave-one-out cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from the entire monitoring dataset ........................................................................................................ 237
Table 6.34 Summary of the optimum estimation procedures of leave-one-out cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from monitoring data for each waterbody .................................................................238

Table 6.35 A comparison of the percentage of monitoring data in each waterbody falling within the 68%, 95% and 99% when estimated using standard kriging and quasi-stationary covariance models derived from the entire dataset or from the respective waterbody ...........................................239

Table 6.36 A comparison of the percentage of monitoring data in each waterbody falling within the 68%, 95% and 99% when estimated using standard BME and quasi-stationary covariance models derived from the entire dataset or from the respective waterbody. ..................................................239

Table 6.37 A comparison of the percentage of monitoring data in each waterbody falling within the 68%, 95% and 99% when estimated using log-kriging and quasi-stationary covariance models derived from the entire dataset or from the respective waterbody. ..................................................240

Table 6.38 A comparison of the percentage of monitoring data in each waterbody falling within the 68%, 95% and 99% when estimated using log-BME and quasi-stationary covariance models derived from the entire dataset or from the respective waterbody. ..................................................240

Table 6.39 Skewness and kurtosis values for the datasets estimated by kriging and BME using standard and log transformed data using quasi-stationary covariance models centred around each monitoring point derived using monitoring data from all over Cork Harbour ........................................242

Table 6.40 Skewness and kurtosis values for the datasets estimated by kriging and BME using standard and log transformed data using quasi-stationary covariance models centred around each monitoring point derived using monitoring data restricted to the waterbody in which the monitoring point is located .................................................................................................................................242
### Mathematical Notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>Cross-sectional area</td>
</tr>
<tr>
<td>$A_{nc}$</td>
<td>Nitrogen to Chlorophyll_a ratio (mgN mgchla$^{-1}$)</td>
</tr>
<tr>
<td>$A_{oc}$</td>
<td>Oxygen to Chlorophyll_a ratio (mgO$_2$ mgchla$^{-1}$)</td>
</tr>
<tr>
<td>$A_{pc}$</td>
<td>Phosphorus to Chlorophyll_a ratio (mgP mgchla$^{-1}$)</td>
</tr>
<tr>
<td>$B$</td>
<td>Bayesian update matrix</td>
</tr>
<tr>
<td>$\beta_{stat}$</td>
<td>Type II error rate</td>
</tr>
<tr>
<td>$1-\beta_{stat}$</td>
<td>Statistical power</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Momentum correction factor for non-uniform vertical velocity profile</td>
</tr>
<tr>
<td>$C_{tot}$</td>
<td>Total cost of sampling</td>
</tr>
<tr>
<td>$C_0$</td>
<td>Initial fixed cost of sampling</td>
</tr>
<tr>
<td>$C_s$</td>
<td>Cost of establishing monitoring site</td>
</tr>
<tr>
<td>$C_v$</td>
<td>Cost per site visit</td>
</tr>
<tr>
<td>$C$</td>
<td>Chezy coefficient</td>
</tr>
<tr>
<td>$C_n$</td>
<td>Courant number</td>
</tr>
<tr>
<td>$C^*$</td>
<td>Air-water interface resistance coefficient</td>
</tr>
<tr>
<td>$C_{BOD}$</td>
<td>BOD concentration (mg l$^{-1}$)</td>
</tr>
<tr>
<td>$C_{DO}$</td>
<td>Dissolved oxygen concentration (mg l$^{-1}$)</td>
</tr>
<tr>
<td>$C_s$</td>
<td>Dissolved oxygen saturation concentration (mg l$^{-1}$)</td>
</tr>
<tr>
<td>$C_{NH3}$</td>
<td>Ammoniacal nitrogen concentration (mg l$^{-1}$)</td>
</tr>
<tr>
<td>$C_p$</td>
<td>Phytoplankton population (mg m$^{-3}$)</td>
</tr>
<tr>
<td>$C_T$</td>
<td>Water quality standard concentration</td>
</tr>
<tr>
<td>$CV$</td>
<td>Coefficient of variation</td>
</tr>
<tr>
<td>$(C_{m_i}^*)_i$</td>
<td>Proportion of $C_m$ which is caused by the upstream segment i</td>
</tr>
<tr>
<td>$C_x(p_i, p_j)$</td>
<td>The covariance between points i and j</td>
</tr>
<tr>
<td>$d$</td>
<td>Depth of flow</td>
</tr>
<tr>
<td>$d_a$</td>
<td>Allowable difference in mean</td>
</tr>
<tr>
<td>$D$</td>
<td>Total water depth (m)</td>
</tr>
<tr>
<td>$D_{PI}$</td>
<td>Phytoplankton death plus respiration rate constant (d$^{-1}$)</td>
</tr>
<tr>
<td>$D_w$</td>
<td>Wind induced dispersion coefficient</td>
</tr>
<tr>
<td>$D_\sigma$</td>
<td>Longitudinal dispersion coefficient</td>
</tr>
<tr>
<td>$D_{xx}, D_{xy}, D_{yx}, D_{yy}$</td>
<td>Dispersion coefficients in x and y directions</td>
</tr>
<tr>
<td>$e$</td>
<td>Estimation error</td>
</tr>
<tr>
<td>$E(I_i)$</td>
<td>The expected value of Moran’s I</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>Depth mean eddy viscosity</td>
</tr>
<tr>
<td>$f$</td>
<td>Fraction of day exposed to sunlight</td>
</tr>
<tr>
<td>$f_{D3}$</td>
<td>Fraction of dissolved inorganic P in water column</td>
</tr>
<tr>
<td>$f_{D8}$</td>
<td>Fraction of dissolved organic P</td>
</tr>
<tr>
<td>$f_s$</td>
<td>Sampling frequency of spatial stations</td>
</tr>
</tbody>
</table>
$f_i$ Sampling frequency at temporal stations
$f_x(\chi)$ Probability density function of parameter $\chi$
$F_x(\chi)$ Cumulative density function of parameter $\chi$
g Acceleration due to gravity
$G_L$ Growth limiting factor
$G_N$ Nutrient limitation factor
$G_{PI}$ Phytoplankton growth rate (d$^{-1}$)
$G_T$ Temperature adjustment factor
$(G_{PI})_{max}$ Maximum growth rate under optimum lighting and nutrient conditions (d$^{-1}$)
H Total water depth
H(X) Marginal entropy of parameter X
$I_H$ Light level which causes growth rate of half of $(G_{PI})_{max}$ (ly d$^{-1}$)
$I_i$ Moran’s information measure
$I_O$ Surface light intensity (ly d$^{-1}$)
k Cross section index
$k_{12}$ Nitrification rate
$k_{83}$ Dissolved organic P mineralisation rate (d$^{-1}$)
$k_a$ Aeration rate
$k_{BOD}$ Half saturation constant for BOD
$k_{NIT}$ Half saturation constant for Nitrogen
$k_d$ Deoxygenation rate
$k_{ir}$ Rate of dissolved oxygen uptake through phytoplankton respiration and excretion
$k_l$ Depth averaged turbulent diffusion constant
$k_p$ Longitudinal depth averaged dispersion constant
K Number of temporal stations
$K_a$, $K_b$ Normalised deviation of parameter from lower and upper extremae
$K_A$ Light attenuation coefficient (m$^{-1}$)
$K_{MPC}$ Half saturation constant for phytoplankton limitation of phosphorus recycle (mgp l$^{-1}$)
l Water quality parameter index
L Number of spatial stations
$L_y$ Distance for complete lateral mixing in-stream
$m_{max}$ Maximum number of compliance categories
$m$ The number of principal components
$n_{tot}$ Total number of multiples of the base lag unit considered
$n_{sample}$ Sample size
$n_s$ Number of sites
$n_{wq}$ Number of water quality parameters
n Timestep
n*  Effective number of observations
N_0  Total number of samples to be taken
N_T  Dimensionless number which accounts for the trend level and
      length of series
p   Number of periods
p_{init}  Initial segment priority
p_{freq}  Interval between sampling events
p_i  The probability at each of the discretisations of a PDF
p*  Ultimate segment priority
p_{ij}  The discrete probability values of a bivariate probability
        distribution
P_j  Weight factor
P  Wetted perimeter
P_{NH3}  Ammoniacal nitrogen preference factor
\rho  Fluid density
\rho_a  Density of air (1.292 kg/m^3)
\rho_i  Autocorrelation coefficient for lag i
q   The order of principal component
q_x  Depth integrated volumetric flux in the x direction
q_y  Depth integrated volumetric flux in the y direction
Q_i  Volume of seawater entering the bay for the first time
Q_L  Lateral discharge per unit length
Q_o  Volume of well-mixed water leaving the bay during ebb tide
Q_R  River inflow
r   Station number
r_p  Autocorrelation coefficient for interval p
R   Hydraulic radius
R_{iq}  Cross correlation between segments i and q
s   The coordinate along the reach
S   Estimated population standard deviation
S_e  Slope of energy grade line
S_f  Friction slope
S_i  The ambient seawater salinity
S_i^2  Parameter sample variance
S_{jk}  Threshold value of compliance for parameter j for class of
        compliance k
S_x  Standard deviation of same
SOD  Sediment oxygen demand (mg l^{-1})
t   Student’s t
T(X_1,X_2)  Transinformation entropy between the parameter X at point 1
            and 2
T_0  Anticipated average duration of compliance between violations
$T_1$ Anticipated average duration of violation
$\text{Tr}$ Increase in mean level to be detected from monitoring
$u$ $X$ direction velocity
$U$ $X$ direction depth integrated velocity
$\mu$ Mean stream velocity
$\mu^*$ Shear velocity
$\mu_j$ The fuzzy relative membership degree
$v_{\text{tot}}$ Number of visits per period
$v_i$ Depth-average mean eddy viscosity
$v$ $Y$ direction velocity
$V$ $Y$ direction depth integrated velocity
$V(I_i)$ The variance of Moran’s $I$
$V_{S3}$ Organic matter settling velocity ($\text{m s}^{-1}$)
$V_{S4}$ Inorganic sediment settling velocity ($\text{m s}^{-1}$)
$w_i$ Allocated weighting of stratum
$W$ Wind velocity
$W_j$ Weight
$\omega$ Angular velocity of earth’s rotation
$\omega_j$ The principal component weight
$\omega_{npj}$ The spatial weight between point $p$ and neighbouring n-1 points
$x_i$ Parameter value
$x_{i\text{ max}}$ Maxima of the decision matrix
$x_{i\text{ min}}$ Minima of the decision matrix
$X$ Water quality data
$X_i^{(l)}$ $l^{th}$ parameter value
$X(p)$ Spatiotemporal random field where $p(s, t)$
$x[x_1, x_2, \ldots, x_m]^T$ Random variables at the $m$ points in the domain
$X[x_1, x_2, \ldots, x_m]^T$ Vector of realizations at the $m$ points in the domain
$Z_{1-\alpha/2}$ Quantile of the normal distribution for a probability of non-exceedance of $1-\frac{\alpha}{2}$
$Z_{IIi}$ The $z$ score at a given confidence level
$Z_{\alpha/2}$ Standard normal deviate
$\zeta$ Water elevation above/below mean water level
$\Delta$ Sampling interval
$\Delta t$ Timestep
$\Delta x$ Grid spacing
$\lambda$ Linear kriging weights
$\sigma_c$ Expected standard deviation of water quality variable
$\sigma_e$ Standard deviation of a detrended time-series
$\sigma_y$ Distance from farthest bank of stream to point of discharge
$\sigma^2_e$ Variance of estimation error
\( \sigma^2 \) \hspace{0.5cm} \text{Variance of } x_t \\
\phi \hspace{0.5cm} \text{Solute concentration} \\
\theta \hspace{0.5cm} \text{Geographic latitude} \\
\theta_{83} \hspace{0.5cm} \text{Temperature coefficient}
Chapter 1. Introduction

1.1. Estuarine and coastal water quality in Ireland

In the Republic of Ireland, 67 estuarine and coastal waterbodies are monitored for water quality by various agencies including the Environmental Protection Agency and Marine Institute. In advance of the first cycle of the Water Framework Directive (WFD), 104 transitional and 30 coastal water bodies were categorised as either at risk or probably at risk, of which 74 and 13 respectively were so categorised as a consequences of point and diffuse pollution (EPA 2005). Recent trends in water quality in Irish transitional and coastal waters point towards improvements in water quality as a consequence of improved wastewater treatment and implementation of nitrates action plans (Bradley et al. 2015). So far in anticipation of the next cycle of the WFD, 53 transitional water bodies and 17 coastal water bodies have been identified as at risk (Deakin 2015).

In the most recent evaluation of Irish wastewater infrastructure, deficits in wastewater treatment capacity were focused on the east, south and south-east of the country; the same receiving waters have also displayed strong N concentrations in recent reporting on the Nitrates directive due to the highly productive agricultural lands within their upstream catchments (EPA 2013).

Reporting under Article 29(1) of the 2010 Nitrates Regulations (EPA 2013) highlighted some interesting trends, with average nitrate concentrations the highest in the south and south–east in groundwater bodies, river water bodies and estuarine and coastal water bodies. Phosphate concentrations in groundwater were broadly dispersed with no noteworthy spatial trend as are the observed average concentrations in river water bodies. MRP concentrations in estuarine and coastal waters are high along the east, south and south west coastline.

Eutrophication as a direct consequence of nutrient enrichment from agriculture and wastewater discharges remains the most prominent issue currently impacting water quality (DECLG 2015) with both sources contributing 93 per cent of nitrogen load
and 88 per cent of phosphorus load. Diffuse discharge from agriculture generates the majority of both nutrients whilst municipal discharges release almost a third of phosphorus. In the most recent assessment report on Irish water bodies, 55 per cent of transitional water bodies and 7 per cent of coastal water bodies were of less than good status (Bradley et al. 2015). The moderate reduction in nutrient emissions from the two primary sources has been attributed in the same report to an improvement in farm management and better management of existing municipal wastewater treatment facilities.

1.2. Design of monitoring networks

There is no rigorous methodology for developing water quality monitoring networks for transitional and coastal waters. A number of guides on monitoring network design have been published since the 1970s. These guides however focus on inland surface and ground waters with no consideration given to coastal waters, with the exception of a few guides which give a qualitative guidance.

Amongst the suggestions made in relation to transitional and coastal water quality monitoring are increasing the density of sampling in space and time to counteract high heterogeneity, using targeted sampling to account for seasonal variation, making due consideration for tidal effects, residence time, river discharge, density and frequency of sampling and sampling locations and frequency.

Statistical assistance in monitoring network design is limited to approximating the number of monitoring points necessary for determining the mean value, a linear trend or a step trend as covered above but no guidance exists on prioritising monitoring locations in a methodological manner. Simple random sampling is not suitable for marine environment due to its heterogeneous nature. Systematic sampling at a regular spacing or frequency may be used to avoid clustering of data points but the approach may not capture adequate information on water quality trends or average conditions.

A number of methods have been proposed in literature as a means of optimal monitoring but none have thus far been adopted or proposed in official guidelines for monitoring network design. Those methods include cluster analysis, discriminant
Introduction

analysis, principal component analysis and variography. Whilst a number of texts contain equations for the design of monitoring networks on the basis of cost, the question of where to sample remains unanswered.

1.3. Combined use of water quality monitoring and modelling

Monitoring and numerical modelling of water quality in transitional and coastal waters are often carried out with a shared purpose but without due consideration of the potential interaction of both activities in the long term. One of the main impediments to the use of numerical models in water quality management is the lack of confidence in the accuracy of numerical models. Poor availability of comprehensive datasets for calibration and validation also inhibits long term use of numerical models. This lack of confidence in the integration of modelling in the monitoring process leads to the collection of monitoring data for analysis and water quality assessment alone. Data is thus inadequate for model calibration and validation. There are many potential uses for a calibrated and validated water quality model, including the identification of pressures and assessment of susceptibility to those pressures and the definition of reference conditions.

The possibility of designing water quality monitoring networks with water quality models remains unexplored to the author’s knowledge. Any efforts to date in the area have been restricted due to efforts focusing on a single priority with regard to monitoring. The potential for ongoing interaction between monitoring and modelling is clear; if water quality monitoring was carried out with a view to ongoing calibration and reassessment of numerical model accuracy, the numerical model could underpin the ongoing reassessment of the validity of each monitoring point due to emerging spatial and temporal patterns in water quality parameters. In addition, divergence between the results of water quality monitoring and modelling over time may indicate that water quality model inputs require modification.
1.4. Aims and objectives

The overarching objective of this research was to derive a methodology for the design and optimisation of the water quality monitoring network within Cork Harbour through the combined use of pre-existing water quality data provided by the Irish EPA and a calibrated and validated water quality model. For the purposes of this research, four water quality parameters related to trophic status assessment were included, namely: Chlorophyll\textsubscript{a}, Dissolved Oxygen saturation percentage (D.O. Sat\%), and Inorganic Nitrogen and Phosphorus, which are referred to throughout as Molybdate Reactive Phosphorus (MRP) and Dissolved Inorganic Nitrogen (DIN) respectively. DIN refers to the sum total of Ammonia and Nitrate.

Two areas of research were pursued. An approach utilising entropy analysis theory was applied to prioritise the existing EPA surface water quality monitoring network and whilst also determining three stages of monitoring network contraction for each water quality parameter such that each monitoring network for each water quality parameter would deliver the least redundant information. The same methods were used to prioritise zones within Cork Harbour for monitoring of each parameter, without direct consideration of the existing monitoring network.

To support the prospect of reduced monitoring coverage or possible relocation of monitoring points, the Bayesian Maximum Entropy (BME) method of modern geostatistics was adapted to the Cork Harbour area. The BME method is proposed as an approach to estimate water quality data at monitoring points which have been removed, due to a desire to downscale the monitoring network, or which have been identified as redundant based on entropy analysis results. BME has the capability to process numerical model simulated data and measured site data adjacent to the estimation site producing a probability density function describing the likely range of values at that estimation point. The results of BME and simple kriging estimates were both compared to collocated monitoring data to verify the veracity of the estimation procedures in the marine environment.

A framework was proposed to coordinate the interaction between water quality monitoring, monitoring network optimisation, and water quality modelling and water quality assessment.
1.5. Thesis overview

The content and layout of this thesis is as follows:

Chapter 2 contains the literature review covering current water quality legislative requirements and the monitoring practice and guidelines which may be applied to monitoring network design. Previous examples of combined use of modelling and monitoring for environmental compliance are also discussed. An overview is given of entropy theory and its potential applications to monitoring network design. The theory behind the BME area of modern geostatistics, which is a development of entropy theory and Bayes’ rule, is also given.

Chapter 3 describes the theory behind 1-D, 2-D and 3-D numerical modelling of water quality in transitional and coastal waters. The solution scheme which underpins the DIVAST numerical model used in this thesis is outlined.

Chapter 4 describes the methodologies that were followed in the application of entropy theory and geostatistical theory in Chapters 5 and 6 respectively. The chapter also includes a framework methodology to follow for integrating the results of entropy analysis and BME estimation in a feedback loop to regularly improve the accuracy of water quality modelling and thus the accuracy of monitoring network optimisation analysis using entropy theory and geospatial probabilistic estimates using BME which are both underpinned by the numerical model water quality simulations.

Chapter 5 implements the methodology for application of entropy theory outlined in Chapter 4. The chapter contains recommendations on the ranking of the existing monitoring network with specific recommendations for spatial and temporal coverage given for each of the three water quality parameters of concern. A high resolution analysis of potential monitoring network prioritisation is carried out for each water quality parameter for each salinity class and each compliance class as defined by the trophic status assessment system currently endorsed by the Irish EPA. Finally, the trophic status assessment results yielded by using each optimised and/or downsized monitoring network are compared to the trophic status assessment result yielded by using the existing monitoring regime used by the Irish EPA.
Chapter 6 contains the results of BME and kriging estimation in Cork Harbour. The reliability of these estimation procedures are considered from the perspectives of their prospective use in estimation of water quality parameters in redesigned monitoring networks where monitoring points have been removed or as a means of estimating water quality parameters in areas where there is no monitoring carried out, on the basis of a pre-existing comprehensive monitoring network. The monitoring data and water quality parameter BME estimates are subsequently used in trophic status assessment and the deviation, from the result that would be otherwise returned is quantified.

Chapter 7 concludes the thesis with a summary of the outcomes from Chapters 5 and 6, a critical commentary on the results contained in this thesis and the associated recommendations for future work.
Chapter 2. Literature Review

2.1. Introduction

As a legacy of the age before commercial aviation, when shipping was the principal means of transporting food, fuel and raw material for manufacturing, towns located near the mouth of estuaries thrived on the convenient access to basic commodities, industry for employment and transportation links. Due to the large populations of these towns and cities, the associated burden of wastewater treatment and disposal, and the burden from diffuse pollution from nutrient dense river inflows, estuarine water quality has often suffered from undesirable disturbances such as eutrophication and dissolved oxygen fluctuations.

Many factors influence estuarine water quality including rainfall runoff, agricultural practices in catchments, municipal and industrial wastewater treatment and discharge, the salinity gradient and mixing, flushing times, and seasonal factors such as sunlight, photoperiod and light intensity.

The following chapter outlines the approach to maintaining estuarine water quality applied in Ireland and the wider European Union. The qualitative guidance provided to direct monitoring program design is reviewed and discussed, while monitoring programs design considerations are also detailed such as the number of samples and sampling frequency required for different monitoring programs. Novel methods for monitoring network design are discussed. Previous combined uses of water quality monitoring and modelling are outlined, before the area of entropy theory, and its relevance to monitoring network design is discussed. An extension of entropy theory known as the BME method of geostatistical estimation is described in detail, as a means of combining water quality monitoring and modelling, before a final justification for the author’s work is given.
2.2. Water quality legislation

Surface waters spanning from inland freshwater lakes and rivers, to transitional estuarine waters through to fully saline coastal waters require regular monitoring for a variety of purposes, including long-term trend observation, investigation into pollution events, assessing improvements in response to measures, intercalibration for determining standards for similar water bodies and baseline monitoring to determine a departure from water quality standards.

In the following section, the approach used for water quality governance in the EU is discussed in detail.

2.2.1. European Union

The Water Framework Directive (WFD) (European Commission 2000a) sets the primary objective for all surface water bodies and groundwater bodies that those currently classified as being of “good” status will be maintained at their present status or improved. Water bodies which are not currently in attainment of “good” water quality status are directed to achieve at least “good” status by the end of the river basin management planning cycle. The directive sets out that member states shall divide the country into river basin districts for planning and management purposes. River basin management plans are assembled for 6 year implementation cycles; the most recent cycle was the first, which covered the period from 2009 to 2015.

Article 4 of the WFD establishes a number of generic environmental objectives for all surface waters including:

- Implementing the necessary measures to maintain the status
- Restore all water bodies to good surface water status
- Protect and enhance all heavily modified water bodies ecological potential and surface water chemical status
- Implement measures to scale down and eventually suspend the emission of all priority hazardous substances
An evaluation of the applicability of each objective is carried out at each water body. Following evaluation, exemptions are allowed on the basis of prioritisation of environmental, social and economic considerations outside of water management or in circumstances where the environmental objectives cannot be met within the RBMP cycles (European Commission 2009b). Exemptions are approved in the form of extended deadlines, relaxation of objectives, temporary decline of status, alterations to water bodies during the RBMP cycle or designation as a heavily modified water body. Exemptions are allowed under the conditions established in article 4, such as a time extension for implementation longer than the management cycle, or the timescale for effecting change in the aspect of the biogeochemical cycle which pertains to the management issue exceeds the management cycle timeframe.

Article 5 stipulates the requirement for characterisation of water bodies and impact assessment of anthropogenic activities on water bodies. Annex II sets out the technical specification to be followed in article 5 evaluations. Characterisation of the river basin districts involves the identification of constituent water bodies and their boundaries, the classification as individual water body types, the typology of water bodies into sub-types based on hydromorphology and physicochemical parameters and the establishment of type specific reference conditions. Reference standards are inter-calibrated across member states using knowledge of parameter values for water body sub-types. The intercalibration exercise involves the harmonisation of ecological assessment systems, ecological quality criteria and specifically the ecological quality criteria for good status (European Commission 2003a). The ecological quality ratio is also determined to indicate the boundaries between different status categories.

The assessment of anthropogenic impacts on water bodies entails the identifying and quantifying the significant anthropogenic pressures in the form of point and diffuse pollution, water abstractions, flow regulations and any other anthropogenic impacts. An assessment is then completed to determine the vulnerability of each water body to the associated pressures. The initial risk assessment of 2004 prior to WFD monitoring was completed using existing data in combination with estimates of environmental quality standards and limited knowledge of identified pressures and impacts to determine initial risk status (UKTAG WP 12a).
The initial risk assessment was subject to alteration following on from the intercalibration exercise between participating countries. Intercalibration resulted in harmonised ecological quality ratios for parameters, yielding a definition of good status via EQRs to provide a basis for the final risk assessment.

Article 11 states the requirement for a programme of measures to correct or maintain and improve water quality status. Basic measures include those required by legislation, relating to sustainable water abstraction, improvement in water quality to lower the level of water purification for potable water treatment, regulation of point and diffuse pollution or any other measures required to counteract pressures highlighted by article 5 assessments (Kavanagh and Bree 2009). Both emission limit values for end of pipe emissions and environmental quality standards for whole water bodies are applied within the WFD where previous EU water quality directives have utilised either of these controls but never both in combination.

2.2.2. Irish Coastal and Transitional Waters WFD Monitoring

Ireland is obliged to assemble and implement a national WFD monitoring programme under the auspices of nominated local authorities and state agencies. Article 8 and Annex V of the WFD state the requirement for surveillance, operational and investigative monitoring programmes. The monitoring programme came into effect on 22nd December 2006. The WFD monitoring programme substitutes the existing national water body monitoring programmes for individual water body types (EPA 2006). The individual monitoring programmes are directly linked with the programmes of measures which specify actions intended to correct water quality. The WFD monitoring programme sets out the nature, frequency and location of monitoring.

The WFD monitoring programme allows ongoing assessment of the efficacy of programmes of measures to influence water quality whilst also emphasizes shortcoming in the programmes of measures as new threats to water quality can be identified. The individual monitoring programmes presented by the EPA are interlinked with the programmes of measures which are contained in all Irish river basin management plans. Each monitoring type is directed at achieving specific objectives. Each monitoring type within the monitoring programmes is associated
with sub networks which are selected to assess the efficacy of the monitoring. Each sub network pertains to one or more of the objectives of that monitoring type.

Water bodies included under the surveillance and operational monitoring programmes were selected to be indicative of overall transitional and coastal water status. The transitional and coastal waters monitoring programme comprises four surveillance monitoring subnetworks and six operational monitoring subnetworks. Physicochemical sampling is carried using minisondes to measure salinity, temperature, DO, turbidity and depth. Discrete water sampling is carried out and analysed for pH, ammonia, total oxidised nitrogen, phosphate, silica, Chlorophyll_a and BOD. Automated sampling is used in circumstances where greater temporal accuracy or greater knowledge of nutrient patterns is required (EPA 2006).

2.2.2.1. Surveillance Monitoring

Surveillance monitoring is completed to assist with the impact assessment procedure, the optimisation of future monitoring programmes and the identification of long term changes in the natural conditions and the impacts of anthropogenic activities. At selected monitoring sites, surveillance monitoring is carried out for a single year of the term of a river basin management plan, covering all biological quality parameters, hydromorphological quality parameters, general physico-chemical parameters, priority pollutants, and all pollutants released in significant quantities to the river basin.

The four subnets which are common to all surface water surveillance monitoring are:

- Monitoring at representative waterbodies to determine surface water status,
- Assessment of significant changes in natural conditions of water body,
- Impact of anthropogenic activities, to assist in the risk assessment
- Water bodies which are highlighted in the WFD, including bodies containing a significant volume of water relative to the RBD, bodies with high flow rates within the RBD, bodies located on member state boundaries and locations where pollutant loads are transferred across state boundaries.
2.2.2.2. Operational Monitoring

Operational monitoring is aimed at assessing the success of programmes of measures at maintaining or improving water quality status in water bodies. The five subnets common to all surface water operational monitoring as stated in the monitoring programme are for the assessment of:

- The success of pollution control measures in reducing impact of point and combined point sources of pollution, diffuse pollution control measures,
- The measures to limit hydromorphological impacts,
- The success of programmes of measures in maintaining good or high water quality status in water bodies not considered as being of high risk in the characterisation report,
- Water bodies classified as Species Protected Areas or Habitat Protected Areas that have been deemed at high risk of not attaining good water quality status.

2.2.2.3. Investigative Monitoring

Investigative monitoring is prompted by events or observations which are not included within the remit of operational or surveillance monitoring. Investigative monitoring would be carried out in situations where the cause of a parameter exceedance is unknown, the degree and extent of pollution must be determined and hence a programme of measures can be assembled to address the pollution, or where surveillance monitoring has indicated that a water body will not achieve “good” ecological or chemical water quality status, in order to identify the cause of not meeting the environmental objectives. Snapshot monitoring is also used in order to assess the contribution of water bodies not included in operational and surveillance modelling programmes to water quality of adjacent water bodies. Electronic alert networks are used to identify potential pollution sources by analysis of the parameter values over time. Remote sensing is also used in order to locate or refine the approximate location of pollution.
2.2.2.4. Sampling frequency

Minimum sampling frequencies for biological, hydromorphological and physicochemical parameters and other pollutants and priority substances are noted in the main WFD text. These frequencies are as follows:

- Monthly – priority substances
- Quarterly – physicochemical quality elements. Thermal conditions, oxygenation, salinity, nutrient status and acidification status
- Biannually - phytoplankton
- Every 3 years for other biological elements
- Every 6 years for hydromorphological elements

Increased sampling frequencies are promoted within the WFD if the departure is based on expert knowledge, experience or technical reasoning. Frequencies are chosen to take into account seasonal variation and anthropogenic activities.

2.2.3. Irish estuarine and coastal water quality

The earliest known appraisal of Irish coastal and estuarine water quality (Letts et al. 1908) acknowledged the considerable issue of discharging raw untreated wastewater to the mouth of the river Liffey which resulted in issues relating to de-oxygenation and algal blooms (Hartnett et al. 2011b), while the same report also suggested that the prevalence of macroalgae blooms were due to excess nitrogen.

In an effort to deal with the water quality management issues, the 1977 Local Government Pollution Act ushered in the requirement for water quality management plans for each of the major waterbodies in Ireland. Of the 100 waterbodies to be considered, a review by the EPA indicated that only 20 WQMPs had been assembled by 1999 (EPA 1999).

In more recent times, no intensive coastal monitoring campaign was undertaken until the 1970s and 1980s (O'Sullivan et al. 1991) where 14 estuaries, harbours, bays and marine dumping sites around the Irish coast were sampled for physicochemical parameters, heavy metals and chlorinated hydrocarbons in shellfish, water and
sediment. Monitoring data was collated from both the independent efforts of the Irish Department of the Marine and monitoring under the Oslo and Paris Conventions for the purposes of the Joint Monitoring Programme at five of Ireland’s industrialised estuaries. Some of the notable observations from that report included elevated mercury levels in shellfish in Cork Harbour, occasional exceedances of shellfish cadmium guideline values but no consistent trend of non-compliance, and exceedances of guideline values for lead in Dublin, Cork and Waterford. No issues were detected regarding heavy metal toxicity in fish, while the limited nutrient observations indicated high nitrogen levels in the Boyne Estuary, Dublin Bay and Wexford Harbour due to sewage input.

Although monitoring campaigns were ongoing in Irish waters over many years linked to international commitments such as the Convention for the Protection of the Marine Environment of the North-East Atlantic (OSPAR), the first comprehensive national monitoring programme for coastal, transitional and marine waters (NEMP) was only published as recent as 2003, as a precursor and a preparatory effort to the EU WFD monitoring programme (EPA et al. 2003). The document covered all aspects of concern in tidal waters including trophic assessment, marine ecological assessment, radiological protection, and impact of all natural and anthropogenic loadings.

In a recent evaluation of tidal water quality, monitoring data collected under the NEMP prior to initiation of the WFD monitoring programme were used to evaluate trophic status amongst other parameters (Clabby et al. 2008). Two bodies were downgraded from potentially eutrophic to eutrophic and three were downgraded from unpolluted to intermediate trophic status. Lough Mahon in Cork Harbour was the only previously eutrophic body to improve in status due to the Cork main drainage project which was completed in 2004. A number of improvements were found in other regards in Cork Harbour and environs, including an improvement in the level of oxygen depletion, ammonium concentrations.

The most recent coastal and transitional water quality report (McGarrigle et al. 2010) indicates that a marked improvement has been effected by improved wastewater treatment, lower chlorophyll levels and better oxygen conditions. The percentage of eutrophic coastal or transitional bodies decreased from 18.9 to 10.1%. 20 additional
bodies were included due to WFD monitoring as well. However, non-compliances were observed in three areas for the DIN environmental quality standard (S.I. No. 272 of 2009) including Cork Harbour and the outer Cork Harbour. MRP non-compliances were observed in four bodies including Lough Mahon at Harper’s Island. Although oxygen conditions were largely healthy, a number of waters were notable for deoxygenation including the lower Lee Estuary.

As part of Ireland’s commitments to OSPAR, an effort was made recently to quantify the reductions required to achieve non-problem trophic status in the application of the comprehensive procedure (O’Boyle et al. 2011). The authors applied the Dynamic Combined Phytoplankton and Macroalgae (DCPM) box model to five problematic Irish estuaries. Data requirements included knowledge of area and depth, water exchange rates, microplankton loss rates, light attenuation coefficients, knowledge of loading during summer and winter, riverine inflows, tidal range and tidal phase. The study revealed that a 15% reduction in nutrient loading to the studied estuarine waters would result in non-problem status with regard to eutrophication. This same reduction would prove to be a conservative aim for other less problematic estuaries within Irish waters.

2.2.3.1. TSAS

Trophic status is determined by the Irish EPA using the trophic status assessment system or TSAS (Toner et al. 2005). Table 2.1 presents the criterion and the levels of compliance for three different salinity classes.

There are 4 levels of compliance: eutrophic, where more than 2 parameters are non-compliant, potentially eutrophic where 2 parameters are non-compliant and a 3rd parameter is within 15% of the compliance limit, intermediate where there are 1 or 2 non compliances or unpolluted.

Otherwise waters are classified as unpolluted. Summer is defined under TSAS as April to September inclusive summing to 30 weeks, and winter is defined as October to March inclusive, summing to 22 weeks.
### Table 2.1 TSAS water quality compliance criteria

<table>
<thead>
<tr>
<th>Parameter/ Waterbody type</th>
<th>Criterion</th>
<th>Statistic</th>
<th>Period to which criterion applies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category A:</strong> Nutrient enrichment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dissolved Organic Nitrogen</strong></td>
<td>mg/l N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tidal fresh waters</td>
<td>&gt;2.6</td>
<td>Median</td>
<td>Winter or summer</td>
</tr>
<tr>
<td>Intermediate waters</td>
<td>&gt;1.4</td>
<td>Median</td>
<td>Winter or summer</td>
</tr>
<tr>
<td>Full salinity water</td>
<td>&gt;0.25</td>
<td>Median</td>
<td>Winter or summer</td>
</tr>
<tr>
<td><strong>Orthophosphate</strong></td>
<td>μg/l P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tidal fresh waters</td>
<td>&gt;60</td>
<td>Median</td>
<td>Winter or summer</td>
</tr>
<tr>
<td>Intermediate waters</td>
<td>&gt;60</td>
<td>Median</td>
<td>Winter or summer</td>
</tr>
<tr>
<td>Full salinity water</td>
<td>&gt;40</td>
<td>Median</td>
<td>Winter or summer</td>
</tr>
<tr>
<td><strong>Category B:</strong> Accelerated growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorophyll_a</td>
<td>mg/m³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tidal fresh waters</td>
<td>&lt;15</td>
<td>Median</td>
<td>Summer</td>
</tr>
<tr>
<td>or</td>
<td>&gt;30</td>
<td>90th Percentile</td>
<td>Summer</td>
</tr>
<tr>
<td>Intermediate waters</td>
<td>&lt;15</td>
<td>Median</td>
<td>Summer</td>
</tr>
<tr>
<td>or</td>
<td>&gt;30</td>
<td>90th Percentile</td>
<td>Summer</td>
</tr>
<tr>
<td>Full salinity water</td>
<td>&lt;10</td>
<td>Median</td>
<td>Summer</td>
</tr>
<tr>
<td>or</td>
<td>&gt;20</td>
<td>90th Percentile</td>
<td>Summer</td>
</tr>
<tr>
<td><strong>Category C:</strong> Undesirable disturbance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DO Sat %</td>
<td>% saturation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tidal fresh waters</td>
<td>&lt;70</td>
<td>5th Percentile</td>
<td>Summer</td>
</tr>
<tr>
<td>or</td>
<td>&gt;130</td>
<td>90th Percentile</td>
<td>Summer</td>
</tr>
<tr>
<td>Intermediate waters</td>
<td>&lt;70</td>
<td>5th Percentile</td>
<td>Summer</td>
</tr>
<tr>
<td>or</td>
<td>&gt;130</td>
<td>90th Percentile</td>
<td>Summer</td>
</tr>
<tr>
<td>Full salinity water</td>
<td>&lt;80</td>
<td>5th Percentile</td>
<td>Summer</td>
</tr>
<tr>
<td>or</td>
<td>&gt;120</td>
<td>90th Percentile</td>
<td>Summer</td>
</tr>
</tbody>
</table>
2.2.4. OSPAR

The OSPAR convention for the protection of the marine environment of the North East Atlantic was signed in Paris on September 22nd 1992 by 15 European countries as a formal commitment to the prevention and elimination of pollution of the marine environment from offshore activities, actions in river basins adjacent to the north east Atlantic Ocean and environmental assessment of the same marine waters. The OSPAR convention resulted from the conglomeration of the subscribing parties and the combining of the objectives of the Oslo and Paris conventions.

Comprehensive physicochemical and biological monitoring of the marine environment is carried out by all contracting parties with the monitoring results feeding directly into the common procedure for the identification of the eutrophic status of the OSPAR maritime area (OSPAR Commission 2005b). A consistent approach to monitoring is carried out in all member countries as a result of the qualitative recommendations on monitoring of oxygen, nutrients and Chlorophyll_a for the joint assessment and monitoring programme. The OSPAR Eutrophication monitoring programme (OSPAR Commission 2005a) assesses waters under three priority levels: Non problem areas, problem areas and potential problem areas which are dealt with in the following manner:

- Non-problem areas are sampled in order to confirm ongoing adherence to non-problem status but with a limited number of parameters and sampled infrequently,
- Problem areas are monitored at a high frequency to observe trends in nutrient concentrations and eutrophication effect parameters (Chlorophyll_a and O₂),
- Potential problem areas are treated as problem areas for up to 5 years and reclassified as a problem area or non-problem afterwards,

Monitoring is carried out with a spatial coverage which encloses the expected eutrophic area. However, contracting parties must decide on optimum sampling frequencies and locations, although minimum sampling frequencies are suggested in the programme.
The guidelines provided for the joint assessment and monitoring programme are provided for the monitoring of nutrients (OSPAR Commission 2013a), dissolved oxygen (OSPAR Commission 2013b) and Chlorophyll_a (OSPAR Commission 2012).

Nutrient monitoring must capture the spatial extent and the temporal trends within the spatial extent for the nutrients in the area. Monitoring of nutrients is recommended during the winter when algal activity is lowest, which results in a conservative view of nutrients due to low algal uptake. In riverine areas, monitoring is to be carried out along salinity gradients. Temporally, nutrient monitoring must capture highest winter nutrient concentrations throughout the range of expected salinity values in order to fully facilitate the use of mixing diagrams which are integral to the assessment of temporal trend as part of the common procedure for assessment of trophic status (OSPAR Commission 2005b). Dissolved oxygen monitoring must be completed during and after peak primary production (OSPAR Commission 2013b). Monitoring of Chlorophyll_a must capture the spatial and temporal extent of phytoplankton bloom, as well as the long term trend in order to capture the frequency and duration of bloom events. A control site is stipulated outside the eutrophic region in order to monitor the influence of other factors such as seasonal and interannual variability.

The requirement for monitoring of the causative and direct effects along a salinity gradient in transient waters for the JAMP is due to the need for a mixing diagram to determine the salinity standardised concentration, which is compared for compliance annually with the salinity related background concentration for each of the parameters of concern. In the first application of the comprehensive procedure, exceedance was defined as 50% above background concentrations for DIN and phosphorus, the N/P ratio and maximum and mean Chlorophyll_a values. Area specific limit values on mean concentrations are defined for coastal waters. All assessment parameters at their respective assessment levels are considered as a whole for the classification. A qualitative assessment of the results is then completed and justification of the score is given with reasons.

Of note in the common procedure document is the following remark: “Particular attention is drawn to the need of updated guidance on the frequency and spatial
coverage of monitoring in the eutrophication monitoring programme” (OSPAR Commission 2005b).

2.2.5. UNEP MEDPOL

The Mediterranean Pollution Program was established under the Mediterranean Action Plan under the auspices of the United Nations Environment Program to assess marine pollution in the Mediterranean Sea. Under MEDPOL, 21 countries bordering the sea participate in the water quality monitoring and assessment to identify and remedy pollution caused by wastewater treatment plant discharges, nonpoint discharge of nutrients to watercourses or any other activities which negatively affect Mediterranean water quality. MEDPOL monitoring guidance is quite direct, with specific reference to the type of site to monitor, the selection of monitoring points at that site and qualitative advice on sampling frequency (UNEP 2003).

The monitoring focus is placed on marine sites, off-shore fish farms and coastal lagoons. Marine and fish farm sites which are eutrophic according to the TRIX framework (Vollenweider et al. 1998) are selected as impact sites with one comparable undisturbed site selected for comparison. Sampling of all TRIX parameters are required including DO, CHLA, total nitrogen, DIN, total phosphorus and total phosphorus. Seasonal sampling is recommended with an emphasis on sampling during productive periods displaying high variability, and the frequency used must have a statistical power to detect a change in concentration over a stated period. Particular attention is given to the spatial coverage of monitoring at selected sites, with a minimum of one transect perpendicular to the coast required for each site and three monitoring points distributed along each transect, the spacing of which is determined based on the bathymetry.
2.3. Monitoring practice & guidelines

2.3.1. Statistical considerations

The issue of serial correlation in establishing monitoring frequencies has proven contentious (Loftis et al. 1991) in the context of estimating average conditions. Loftis notes that in the derivation of a long term annual mean and confidence interval from serially correlated data from Equations (2.1) and (2.2), as correlation between samples increases, so does the variance and hence the confidence interval.

\[
\bar{x} \pm Z_{\alpha/2} \left[ \text{var}(\bar{x}) \right]^{1/2} \\
\text{var}(\bar{x}) = \frac{\sigma^2}{2} \left[ 1 + \frac{2}{n_{\text{tot}}} \sum_{i=1}^{n_{\text{tot}-1}} (n_{\text{tot}} - 1) \rho_i \right] \\
\rho_k = \frac{E[(x_t - \mu_t)(x_{t-k} - \mu_{t-k})]}{[\text{var}(x_t)\text{var}(x_{t-k})]^{1/2}}
\]

where:
- \(\bar{x}\) = sample mean
- \(Z_{\alpha/2}\) = standard normal deviate
- \(\sigma^2\) = variance of \(x_t\)
- \(\rho_i\) = autocorrelation coefficient for lag \(i\)
- \(n_{\text{tot}}\) = total number of multiples of the base lag unit considered
- \(E((x_t-\mu_t)(x_{t-k}-\mu_{t-k}))\) = covariance between \(x\) at time \(t\) and \(x\) after a time lag \(k\)

It can be seen that the correlation coefficient is a function of the covariance between two monitoring samples. Loftis stated however that autocorrelation in data utilised for deriving a mean over a specific time reduces the variance in the final estimate. Hence autocorrelation has different implications depending on the objective being pursued by monitoring.
2.3.2. US monitoring network guidance

Monitoring of surface waters in the USA is primarily linked with activities sanctioned under the Clean Water Act (C.W.A.), including water quality implementation plan assessment, total maximum daily load derivations and discharge permit compliance. However, guidance on water quality monitoring network design is biased towards sampling within the river basin in rivers, streams and lakes, with a scarcity of guidance on the monitoring of water quality in the estuarine and coastal environment. In the following section, various methods which have been proposed for designing monitoring networks in surface waters are discussed and their relevance to the estuarine environment is considered.

The national water quality handbook (USDA and NRCS 2003) contains a number of references to key monitoring guidance publications. The authors suggest the use of the mixing length $L_y$ after Sanders et al. (1983) in Equation (2.4) to determine a suitable distance downstream from an outfall to monitor stream water quality.

$$L_y = 2.17 \frac{\sigma_y^2}{d} \times \frac{\mu}{\mu^*}$$  \hspace{1cm} (2.4)

where:

$L_y$ = distance for complete lateral mixing in-stream

$\sigma_y$ = distance from farthest bank of stream to point of discharge

$d$ = depth of flow

$\mu$ = mean stream velocity

$\mu^*$ = shear velocity = $(gRS_e)^{1/2}$

$g$ = acceleration due to gravity

$R$ = hydraulic radius = $A/P$

$A$ = cross-sectional area

$P$ = wetted perimeter

$S_e$ = slope of energy grade line
Equation (2.4) may be applied to determine the distance $L_y$ downstream of a junction with an adjoining tributary or outfall, and instances where concentrations vary with stream depth or across the stream width would require sampling across a transect to detect average conditions.

In order to determine the spatial representativeness of a monitoring site in a riverine lake or a long and narrow lake with a longitudinal water quality parameter gradient, the same document nominates the method of Potash and Henson (1978), see Equation (2.5):

$$\pm Distance = \frac{[\bar{X} \pm S_x t] - a}{b}$$

Equation (2.5)

where

$\bar{X} = \text{mean of water quality parameter at monitoring station}$

$S_x = \text{standard deviation of same}$

$t = \text{student’s t at p = 0.05}$

$a$ and $b$ are the regression intercept and slope of a linear regression plot of the water quality variable with distance along the lake. The author suggests repeated linear regression at different monitoring points to determine the representative distance around each monitoring point and hence ensure that there is no overlap between the representative zones around adjacent monitoring points. This exercise may be curtailed by limited monitoring data, which would suggest the use of a validated numerical model may be beneficial. The aforementioned method assumes that knowledge of water quality in individual strata is not required.

Optimizing the location of sampling was suggested on a pragmatic basis in the same publication, based on the cost analysis method proposed by Reckhow and Chapra (1983). With knowledge of all of the sources of variance, the cost of individual sampling sites and a cap on costs, the number of sites $S$ that are affordable is determinable using Equations (2.6) to (2.11):
\[ C_{\text{tot}} = C_0 + n_s C_s + n_s p v_{\text{tot}} C_v \] (2.6)

\[ v_{\text{tot}} = \left( \frac{C_{\text{tot}} K_v + C_s}{p C_v (p K_s + K_{sv})} \right)^{\frac{1}{2}} \] (2.7)

\[ K_s = \frac{\sigma_s^2}{\sigma_e^2} \] (2.8)

\[ K_v = \frac{\sigma_v^2}{\sigma_e^2} \] (2.9)

\[ K_{sv} = \frac{\sigma_{sv}^2}{\sigma_e^2} \] (2.10)

\[ n_s = \frac{C_{\text{tot}}}{C_s + p v_{\text{tot}} C_v} \] (2.11)

where:

\( C_{\text{tot}} \) = total cost of sampling

\( C_0 \) = initial fixed cost

\( C_s \) = cost of establishing monitoring site

\( C_v \) = cost per site visit

\( n_s \) = number of sites

\( p \) = number of periods

\( v_{\text{tot}} \) = number of visits per period

Variance term \( \sigma \) refers to site variation (s), variation due to visits (v), co-variation between site and visits (sv) and error variance (e). The method presented above would only apply to the case where a monitoring network is scaled back on the basis of data collected on an initial monitoring excursion. No attention is given to the process of selecting the subset of monitoring points from the initial network.

The number of samples \( n \) required in a monitoring expedition is a basic requirement of monitoring. The number \( n \) is dependent on the chosen sampling scheme and the specific objective of the monitoring effort.
The range of expressions for n are as follows in Equations (2.12) to (2.17), related to 4 specific objectives:

- Estimation of sample mean value with simple random sampling:

  \[ n_{\text{sample}} = \frac{t^2S^2}{d_a^2} \]  
  \[ \text{Using standard deviation (Reckhow and Chapra 1983)} \]  
  \[ n_{\text{sample}} = \frac{t^2CV^2}{\%X^2} \]  
  \[ \text{Using coefficient of variation (Moser and Huibregtse 1976)} \]

- Linear trend detection with simple random sampling:

  \[ n_{\text{sample}} = \frac{12t^2S^2}{d_a^2} \]  
  \[ \text{(Ward et al. 1990)} \]  
  \[ S^2 = \frac{\sum X_i^2 - \left(\frac{\sum X_i}{m}\right)^2}{m - 1} \]  

- Step trend detection with simple random sampling:

  \[ n_{\text{sample}} = \frac{2t^2S^2}{d_a^2} \]  
  \[ \text{(Sanders et al. 1983)} \]

- Estimation of sample mean value with stratified random sampling (Reckhow and Chapra 1983):

  \[ n_{\text{sample}} = \frac{t^2(\sum w_iS_i)^2}{d_a^2} \]  
  \[ \text{Overall sample size} \]  
  \[ n_i = \frac{n_{\text{sample}}w_iS_i}{\sum(w_iS_i)} \]  
  \[ \text{Strata sample size} \]

where

- \( n_{\text{sample}} \) = sample size
- \( t \) = student’s t value at n-1 degrees of freedom and confidence level of p
- \( S \) = estimated population standard deviation of m data points
- \( d_a \) = allowable difference in mean
- \( CV \) = coefficient of variation
- \( \%X \) = allowable deviation from the mean
- \( X_i \) = \( i^{th} \) parameter value
- \( w_i \) = allocated weighting of stratum
Simple random sampling has been criticised for its inability to capture a representative sample in particular in the marine environment (USEPA 2001). Stratified sampling may be more suited to application in coastal waters based on salinity, flushing rates or variance in the variable being monitored.

Systematic sampling of the parameter of interest at a regular spacing or frequency will avoid clustering of data points but the method may still avoid capturing a representative sample (USEPA 1997), especially where periodic fluctuations in values would be expected (Gilbert 1987), as per tidal waters. Monitoring required in assembling mixing diagrams for the OSPAR comprehensive procedure would require monitoring points to be located along a salinity gradient to capture the entire range of salinity values which are likely to be observed. Neither simple, stratified nor systematic samples would adequately capture this salinity gradient.

Spatial and temporal sampling guidance published in support of water quality monitoring may point towards the number of monitoring points necessary for determining the mean value, a linear trend or a step trend as covered above, but an approach to prioritising monitoring locations and frequencies has not been pursued by many.

The first notable attempt (Beckers et al. 1972) involved the 1-D numerical modelling of the coupled BOD and dissolved oxygen of the Wabash river basin with a number of point discharges. The river basin was simplified into contiguous segments which the authors suggested was a simplification for computational simplicity, but also allowed the network to be viewed on a macroscopic level. An initial priority was allocated to each segment, using Equation (2.18), based on ambient water quality standards and the simulated water quality data for that segment, while ultimate priority given by Equation (2.19), ranked each segment based on its overall importance relative to the remaining segments.

\[
P_{\text{init}} = \begin{cases} 
\frac{C_T - C_m}{\sigma_c} & \text{for “not less than” WQS} \\
\frac{C_m - C_T}{\sigma_c} & \text{for “not to exceed” WQS}
\end{cases}
\]

(2.18)
\( p_i^* = \begin{cases} \frac{C_T - C_{m^*}}{\sigma_c} & \text{for “not less than” WQS} \\ \frac{C_{m^*} - C_T}{\sigma_c} & \text{for “not to exceed” WQS} \end{cases} \) 

(2.19)

\( \bar{C}_{m^*} = \bar{C}_m(1 - R_{iq}) \) 

(2.20)

\( R_{ij} = \frac{(C_{m^*})_i}{C_m} \) 

\( i = j - 1 \) 

(2.21)

\( R_{ik} = R_{ij}R_{jk} \) 

\( i = j - 1 = k - 2 \)

The prioritising would continue until all segments have been included, or otherwise optimising could be terminated early upon reaching the desired network size.

where:

- \( p_{\text{init}} \) = initial segment priority
- \( p^* \) = ultimate segment priority
- \( C_m \) = highest (not to exceed) or lowest (not less than) observed concentration
- \( C_{m^*} \) = highest (not to exceed) or lowest (not less than) observed concentration after monitoring initiates at point \( q \) which has highest initial segment priority
- \( (C_{m^*})_i \) = proportion of \( C_m \) which is caused by the upstream segment \( i \)
- \( C_T \) = water quality standard concentration
- \( \sigma_c \) = expected standard deviation of water quality variable
- \( R_{sq} \) = cross correlation between segments \( i \) and \( q \)
Beckers also presented a method for prioritising monitoring frequencies, using Equations (2.23) and (2.30), and the associated expressions in Equation (2.24) to (2.29), although the method was recommended to be used in conjunction with cost analysis as high frequency sampling would be ideal given no constraint on cost.

\[ M(\Delta, T_0, T_1) = \frac{T_0}{\Delta} \left[ C_0 \exp \left( -\frac{\Delta}{T_0} \right) + C_1 \exp \left( -\frac{\Delta}{T_1} \right) \right] \]  

(2.23)

\[ C_0 = \sum_{n=1}^{\infty} \frac{T_0^{n-1}T_1^{n}}{(T_0-T_1)^2} f_{0n} \]  

(2.24)

\[ C_1 = \sum_{n=1}^{\infty} \frac{T_0^{n-1}T_1^{n}}{(T_0-T_1)^2} f_{1n} \]  

(2.25)

\[ f_{0n} = \sum_{k=1}^{\infty} a_k^n \Delta^{k-1} \left[ \frac{T_0-T_1}{T_0T_1} \right]^{k-1} \]  

(2.26)

\[ f_{1n} = \sum_{k=1}^{\infty} b_k^n \Delta^{k-1} \left[ \frac{T_0-T_1}{T_0T_1} \right]^{k-1} \]  

(2.27)

\[ a_k^n = \frac{(-1)^{n-k}(2n-k-1)!}{(n-k)! (n-1)! (k-1)!} \]  

(2.28)

\[ b_k^n = (-1)^k a^n \]  

(2.29)

\[ M(\Delta, T_0, T_1) = \frac{T_0}{\Delta} \left[ \frac{\exp \left( -\frac{\Delta}{T_0} \right) - \exp \left( -\frac{\Delta}{T_1} \right)}{T_0/T_1 - 1} \right] \]  

(2.30)

where:

\( T_0 = \) anticipated average duration of compliance between violations

\( T_1 = \) anticipated average duration of violation

\( \Delta = \) sampling interval

and \( \Delta < T_0 \) is assumed
The observed value of M, described as a “priority measure”, would range between 0 and 1 indicating complete ineffective and effective sampling respectively. The second simplified term for M given by Equation (2.30) applies in the case where $\Delta < T_0 + T_1$, which means that the sampling interval is shorter than the average time between violations. The method was presented as a means of determining the optimum sampling frequency for streams and rivers and would require comprehensive data in order to obtain reliable estimates of $T_0$ and $T_1$.

The method presented by Beckers fails to evaluate a trade-off between the size of the monitoring network, the priority of monitoring points and a suitable sampling frequency which was an objective pursued by Ward et al. (1990) in the monitoring of the acid lakes of Quebec. Ward applied correspondence analysis and ascending hierarchical classification to identify patterns and similarities between lakes with respect to patterns in biochemical variables. Application of correspondence analysis to monitoring data consisting of pH, alkalinity, and sulphate and combined calcium and magnesium measurements from 158 lakes revealed that pH, alkalinity, calcium and magnesium concentrations varied similarly and sulphate was distinct in its pattern of variation. Five classes of lake were revealed from ascending hierarchical classification. Spatial optimization of monitoring was completed by stratified random sampling of 100 of the available lakes based on knowledge of individual standard deviations for each of the parameters in each stratum. Ward also applied the methods of Lettenmaier (Lettenmaier 1976) to determine an optimum sampling scheme.

Ward established following the works of Lettenmaier that assuming a statistical power $(1-\beta_{\text{stat}})$ of 0.9 in Equation (2.31), and expressing $N_T$ as a function of $Z_{1-\alpha/2}$, equation in Equation (2.32) would lead to Equation (2.33) which yields the desired number of samples required to deliver a statistical power of 0.9 for a trend $T_r$ given the error variance of the detrended time-series. Equation (2.33) assumes that each observation is independent. Ward presented Equation (2.34) as given elsewhere (Lettenmaier 1976) as a means of deriving an effective number of observations $n^*$ in the case where dependence exists between the $n$ observations at a monitoring location. Ward concluded that monitoring on a weekly basis for a year would yield between 5 and 10 effective observations $n^*$. 
(Lettenmaier 1976) \[ 1 - \beta_{\text{stat}} = F \left( N_T - Z_{1-\alpha/2} \right) \] (2.31)

(Lettenmaier 1976) \[ N_T = \frac{Tr}{\sigma_e} \sqrt{\frac{n^*}{12}} \] (2.32)

(Ward et al. 1990) \[ n^* = \frac{12 \left( 1.282 + Z_{1-\alpha/2} \right)}{(Tr/\sigma_e)^2} \] (2.33)

(Ward et al. 1990) \[ \frac{1}{n^*} = \frac{1}{n_{\text{sample}}} + \frac{2}{n_{\text{sample}}^2} \sum_{p=1}^{n-1} \left( n_{\text{sample}} - p_{\text{freq}} \right) r_p \] (2.34)

where

\( \beta_{\text{stat}} \) = Type II error rate

1-\( \beta_{\text{stat}} \) = Statistical power

\( F \) = standardized normal cumulative distribution function

\( N_T \) = dimensionless number which accounts for the trend level and length of series

\( Z_{1-\alpha/2} \) = quantile of the normal distribution for a probability of non-exceedance of 1-\( \alpha/2 \)

\( Tr \) = increase in mean level to be detected from monitoring

\( \sigma_e \) = standard deviation of a detrended time-series

\( n_{\text{sample}} \) = sample size

\( n^* \) = effective number of observations

\( p_{\text{freq}} \) = interval between sampling events

\( r_p \) = autocorrelation coefficient for interval p

Ward also considered the trade-off between emphasising spatial and temporal sampling with Equations (2.35) and (2.37), with the latter equation describing the situation where spatial and temporal sampling efforts are synchronized and temporal data is used as spatial observations. \( N_0 \) would be set to a fixed value due to resources. The measure \( \pi \) contained in Equation (2.37) represents the priority given to temporal monitoring over spatial monitoring. Hence for a number of fixed values for \( \pi \) the trade-off between spatial and temporal monitoring can be considered.
Therefore, with a fixed available resource \( N_0 \), based on cost, the term \( \pi \) could be plotted based on a number of values of \( f_t \) and \( f_s \) to visualise the trade-off between focusing on spatial or temporal sampling.

\[
N_0 = K f_t + L f_s \quad (2.35)
\]

\[
N_0 = K(f_t - f_s) + M f_s \quad (2.36)
\]

\[
\pi = \frac{K f_t}{L f_s} \quad (2.37)
\]

where

\( N_0 \) = total number of samples to be taken  
\( K \) = number of temporal stations  
\( f_t \) = sampling frequency at temporal stations  
\( L \) = number of spatial stations  
\( f_s \) = sampling frequency of spatial stations  
\( M = K + L \)

### 2.3.3. Functional data analysis

Pastres et al. (2011) applied the functional data analysis (FDA) method as described by James and Sugar (James and Sugar 2003) to determine spatial similarity in Chlorophyll\_a, MRP, ammonium, nitrate and salinity monitoring data, which revealed two or three groups of monitoring points for each water quality variable. Pastres noted that in future applications, FDA should continue to be used on individual parameters, as multivariate applications would lead to grouping with no rational in a spatial context. The method yielded 5 groups, of which 3 groups containing a total of 22 points could clearly be related back to proximity to pressures such as wastewater outfalls, sea water and riverine inputs. Of the remaining 8 points, 4 were proposed to represent areas for investigative monitoring as termed under the WFD due to the existence of confounding conditions with low Chlorophyll\_a but high nutrient concentrations. Pastres suggested that some redundancy existed between these 22 points and hence the network could be reduced in size but did not propose a methodology for prioritising the monitoring points to be retained. Pastre’s
work lacked a measure of covariance and the procedure assumed that all data was independent with no interaction between sites.

2.3.4. Discriminant analysis

The Sado Estuary in Portugal has been a test bed for a number of monitoring optimization techniques. In a number of publications, methods including cluster analysis and discriminant analysis (Caeiro et al. 2003a), simulated annealing and indicator kriging (Caeiro et al. 2003b) and variography (Caeiro et al. 2003c). In the application of cluster analysis and discriminant analysis to delineate estuarine management areas, those areas described by cluster analysis better reflected the underlying water quality issues in the estuary although the locations of management areas were largely the same. Caeiro scaled down the heavy metal monitoring network in the Sado Estuary initially comprising 77 locations by applying simulated annealing conditioned on maintaining the proportions of both the initial organic matter content classifications and the initial ecological status classifications. Simulated annealing proceeded on the basis of minimising the kriging error variance upon removal of a redundant monitoring location. The size of the scaled back monitoring network was determined after Cochran (Cochran 1977) using Equation (2.14). This equation assumes that monitoring locations are independent and identically distributed which may be untrue in an estuarine environment. However, the application of kriging assumes an inherent dependence which is represented by a semivariogram or covariograms, therefore the iid assumption must be brought into question. The only monitoring locations to be considered were those already in use due to limited data availability which also limits the scope of the optimized monitoring network.

Design of an optimum water quality monitoring network usually targets a single objective such as representativeness of the underlying waterbody. Multiple criteria can be considered at once in optimizing monitoring networks using genetic algorithms (Park et al. 2006). Park used GA to seek the maximum fitness score comprising the sum of four fitness functions to represent compliance with water quality standards, supervision of water use, surveillance of pollution sources and examination of water quality changes. The optimized monitoring network was
selected from a discretized watershed basin of 15,000m x 15,000m. Of the 110 original monitoring points, only 35 were carried over to the optimized river basin monitoring which also contained 110 locations.

2.3.5. Matter element analysis

A lack of resources in terms of both finance and manpower is a common driver behind the objective optimising a monitoring network and in so doing, making optimum use of those resources. However, data sparsity limits the scope of monitoring program optimisation. Spatial optimisation is limited to removing monitoring points if only discrete measurements are available. Chen overcame this limitation by utilising monitoring data and additional data from a supplementary monitoring campaign to calibrate and validate a commercial 1-D finite difference hydrodynamic and water quality model which facilitated the application of matter element analysis to simulated data for a 1,890km long reach of the Heilongjiang River (Chen et al. 2012). Equations (2.38) – (2.43) underpin matter element analysis as applied by Chen:

\[
K_a(X^k) = \frac{X^k_i - c_i}{c_i - a_i} \quad (2.38)
\]

\[
K_a(X^k) = \sum_{i=1}^{n_{wq}} W_i K_a(X^k_i) \quad (2.39)
\]

\[
K_b(X^k) = \frac{X^k_i - c_i}{c_i - b_i} \quad (2.40)
\]

\[
K_b(X^k) = \sum_{i=1}^{n_{wq}} W_i K_b(X^k_i) \quad (2.41)
\]

\[
W_i = \frac{P_i}{\sum_{i=1}^{n} P_i} \quad (2.42)
\]

\[
P_i = \frac{1}{2} (a_i + b_i) \quad \frac{1}{m} \sum_{k=1}^{m} s_{ik} \quad (2.43)
\]

where

\[n_{wq} = \text{number of water quality parameters}\]

\[X = \text{water quality data}\]
a, b, c = lowest, highest and average value of parameter over entire river

k = cross section index

l = water quality parameter index

K_\text{a}, K_\text{b} = Normalised deviation of parameter from lower and upper extremae

W_j = weight

P_j = weight factor

S_{jk} = threshold value of compliance for parameter j for class of compliance k

m_{\text{max}} = maximum number of compliance categories

Matter element analysis proceeds by plotting K_a against K_b for each parameter. A straight line from top left to bottom right should be observed with clustering of monitoring points along the line as in Figure 2.1 representing redundancy. In Chen’s application of MEA to the existing monitoring network at high, average and low flow conditions, homogeneity was detected in groups of 2-3 points in the existing network whilst MEA of the simulated data at 94 locations in the river basin yielded clusters of monitoring points ranging in size from 4 to 12. 3 unmonitored homogenous river reaches were identified for future monitoring and 3 existing cluster of monitoring points each consisting of two monitoring points were reduced from a total of 6 points to 4.

Figure 2.1 Identification of homogeneous river reaches by matter-element analysis, taken from Chen et al. (2012). Deviation of monitoring points from the straight line joining the top left and bottom right corner indicates the requirement for extra monitoring [Image removed for copyright reasons]
2.3.6. Principal component analysis

Principal Component Analysis was also identified as a means of clustering water quality variables into distinct groups of variables which are correlated with one another (Karpuzcu et al. 1987, Ouyang 2005). Solidoro utilised principal component analysis to determine the spatial similarity of monitoring points in the lagoon of Venice (Solidoro et al. 2004). One of the two principal components was found to reflect the trophic status. The component also correlated well with residence time and salinity.

More recently, Ou considered the application of principal component analysis, kriging error variance analysis, Moran’s information measure and fuzzy optimal modelling to determine an optimum lake water quality monitoring network (Ou et al. 2012). Ou applied principal component analysis and found that approximately 90% of data variance could be described by three principal components. Principal component analysis proceeds by finding the vector of weights $W$ that explain the maximum proportion of the variance in the original data vector $X$. Upon removing the product of the weights of the first principal component, analysis continues on the new dataset and continues until all of the data variance has been removed. Each principal component is given by:

$$PC = \sum_{i=1}^{n} W_i x_i$$

where $x_i$ refers to the variable under scrutiny and $W_i$ is the PCA weighting which maximises the proportion of variance of $X$ described. Subsequently, kriging analysis was applied to determine the change in error variance of each of the principal components upon removing each monitoring station when compared to the variance of the data at the monitoring point before removal. A low change in variance would indicate redundancy while high variance would represent a high priority. Local Moran’s information analysis investigates the spatial correlation between each monitoring point $p$ and the remaining monitoring locations. Positive values of Moran’s I and the associated $Z$ scores indicate that neighbouring monitoring points are similar and redundancy may exist and a negative score indicates that a monitoring location is a priority monitoring point.
where $I_p$ is Moran’s information measure, $x_p$ refers to the parameter value, $X$ is the mean parameter value, $S_p^2$ is the parameter sample variance, $\omega_{mpj}$ is the spatial weight between point $p$ and neighbouring $n-1$ points, $Z_{Ip}$ is the z score at a given confidence level, $E(I_p)$ is the expected value of Moran’s I and $V(I_p)$ is the variance of Moran’s I.

Kriging error variance analysis and Moran’s I are two distinct ways to prioritise monitoring and eliminate redundancy. In order to find consensus between these disparate methods, Ou utilised fuzzy optimal modelling. The process involves transforming the elements $x_{qj}$ of the decision matrix $X$ into the fuzzy matrix $F$ which is completed element wise using Equation (2.50) or Equation (2.51), depending on whether monitoring priority is indicated by high or low values in the decision matrix respectively:

$$f_{qr} = \frac{x_{q\max} - x_{qr}}{x_{q\max} - x_{q\min}}$$  \hspace{1cm} (2.50)

$$f_{qr} = \frac{x_{qr} - x_{q\min}}{x_{q\max} - x_{q\min}}$$  \hspace{1cm} (2.51)

$$\mu_r = \left(1 + \frac{\sum_{q=1}^{m}\omega_q(1 - f_{qr})^2}{\sum_{q=1}^{m}(\omega_q f_{qr})^2}\right)^{-1}$$  \hspace{1cm} (2.52)
where \( q \) refers to the order of principal component, \( r \) is the station number, \( m \) is the number of principal components, \( x_{q \text{ max}} \) and \( x_{q \text{ min}} \) refer to the extreme values of the decision matrix, \( \omega_q \) is the principal component weight and \( \mu_r \) is the fuzzy relative membership degree. In the final ranking of monitoring points the lowest values of \( \mu_r \) correspond to the highest priority monitoring points which should be retained.

In Ou’s application of the aforementioned methods to the 30 monitoring points in Lake Winnipeg, when kriging error variance of principal components is compared with Moran’s I analysis separately, 7 monitoring points are common to the 10 lowest priority monitoring points by each method. The highest priority monitoring points share no common ground however. Application of fuzzy optimal modelling to the initial results, thus combining the priority measures of both the preceding methods, results in a blending of the high and low priority monitoring points depending upon the weights appointed to kriging and Moran’s I results. Ou notes that the kriging may result in discontinuities around estimation points and hence would be better suited to densely populated datasets. This is due to kriging yielding a best linear unbiased estimate such that the estimation error variance reaches a maximum value between two data points. Ou also cautions the application of Moran’s I to cases where less than 30 monitoring points are available as the Z score test at a given confidence level is more suited to a number of data points exceeding 30.

2.3.7. Other Research Methods for monitoring optimisation

Regression and harmonic analysis can also be used to determine or lower the frequency of sampling respectively (Khalil and Ouarda 2009). Multivariate linear regression has been used in the lowering of the sampling frequency of specific conductance, by maintaining the coefficient of determination for multivariate linear regression (Tirsch and Male 1984). None of the aforementioned methods dictate the location of monitoring.

Trend analysis as a means of identifying the optimum sampling frequency has been developed by a number of authors (Bayley and Hammersley 1946, Lettenmaier 1976, Sanders and Adrian 1978). The process entails the calculation of the maximum number of samples which can be taken in any given year whilst avoiding autocorrelation, and determining the duration of sampling required to establish a
trend with different confidence levels and powers. The number of samples required to result in an estimated mean falling within prescribed confidence limits has been proposed as another means of designing a monitoring network (Sanders and Adrian 1978, Sanders et al. 1983).

Both the number of monitoring points and the frequency of sampling of water quality variables has been prioritised over the exact orientation of monitoring points in previous literature discussing temporal trend detection in river and stream monitoring networks (Dixon and Chiswell 1996, Lettenmaier 1978, Lettenmaier and Burges 1977).

Naddeo investigated the influence of varying sampling frequencies from 1 up to 6 months on the WFD water quality characterisation of sections of river in the Sele basin in Italy (Naddeo et al. 2007). Lowering the sampling frequency from monthly to every 2 months at one monitoring point or every 4 months at another resulted in a drop in the water quality status of the stretches in question from good to moderate.

Dowdall utilised the semivariance, which provides a basis for many geostatistical methods, and kriging, as a means of optimising the sampling frequency for technetium-99 in the Norwegian Arctic marine environment (Dowdall et al. 2005). The technique involved applying kriging to generate cross validation estimates of technetium-99 using semivariance models of measurements from marine and seaweed data. The author observed that a time series of monthly sampling at two monitoring points could be estimated whilst preserving a minimum average kriging standard deviation, by using measurements at less than monthly intervals. Dowdall concluded that sampling more frequently than every 40 days for seaweed data or every 50 days for seawater concentrations would yield minimum gains. One shortcoming of reducing the sampling frequency is that it limits the accuracy of future updated semivariograms and subsequent reevaluation of kriging estimates. The author suggested using simulated time series data from a calibrated NAOSIM (North Atlantic – Arctic Ocean Sea Ice Model) model in that eventuality, or if no historical data existed (Karcher et al. 2004).
In the context of lake water quality monitoring, residence times may provide constructive guidance on the optimum location and frequency of monitoring. Dabrowski and Berry (2009) proposed a sampling methodology for Lough Corrib guided by the spatial distribution of lake residence time. The ECOMSED finite difference hydrodynamic and sediment transport model was used to determine the flushing time in each of the 14,490 computational cells. Waters were divided into three categories: poorly, moderately and quickly flushed. It was suggested that priority high frequency sampling to be carried out at quickly flushed waters, such as locations near river inlets/outlets. The quarterly minimum sampling frequency suggested in the WFD (European Commission 2000b) was considered in the analyses. Representative monitoring points were suggested as those cells with a monthly percentage change of conservative tracer of within 10% of the lake’s overall value. The location of the proposed locations changed each month, with a restricted area maintaining the criterion for the entire year. The proposed methodology was not restricted to specific pollutant types.

In the context of estuarine monitoring, Radford utilised a numerical model, along with cumulative sum and control charts to reduce monitoring efforts, with a practical application to the Severn Estuary with salinity and dissolved cadmium data (Radford and West 1986). Radford derived a relationship between the minimum sampling frequencies required to determine a single pollution incident with the retention time of the Severn Estuary.

2.3.8. WFD monitoring network guidance

The monitoring guidance provided under the Common Implementation Strategy for the WFD lacks a rigorous methodology for the assembly of monitoring programmes. Advice relating to WFD monitoring programmes is provided by the European Commission in the areas of monitoring point location and quantity and monitoring frequency (European Commission 2003b). Suggestions include the grouping of water bodies with similar natural conditions and anthropogenic pressures, increasing the density of sampling in space and time to counteract high heterogeneity, using targeted sampling to account for seasonal variation and locating monitoring points at the most sensitive locations (European Commission 2009a).
For operational monitoring, it is suggested that a number of monitoring points should be used within a water body subjected to a number of pollution sources, and the location of the monitoring points should reflect the location of the sources, while water bodies polluted by a single source should be monitored at the location most sensitive to the pollutant (European Commission 2003b). Statistical assistance is limited to approximating the number of monitoring points for a desired confidence level and precision.

The Irish EPA utilise the coefficient of variation approach to determine the number of samples required to determine the mean values as expressed in Equation (2.12b) for the purposes of the national WFD monitoring programme (EPA 2006). Irish waters are routinely sampled 4, 6 or 12 times per year. The same publication does not however explicitly refer to utilising the coefficient of variation approach in estuarine or coastal waters, with references only made to tables of nationally averaged CVs for a range of water quality parameters for lakes and rivers.

It is also proposed that the WHO mixing zone principle should be applied where part of a body in isolation is impacted by a point source. The mixing zone principle entails sampling just outside the zone of impact. UK task advisory group 12a on monitoring for the WFD also suggest sampling within the zone of impact, where a body is subjected to widespread pollution, in order to assess the status of the body and the severity of the pressures on the body (UKTAG 2005). UKTAG guidance specific to operational monitoring in transitional and coastal waters suggests a unique approach to each water type in two case studies (UKTAG 2006). Coastal waters are divided by a 1km² square grid within which simple random sampling is applied. In the same publication, stratified sampling of nutrients in the transitional waters of Poole harbour is suggested with the area subdivided based on salinity.

The guidance paper drafted by UKTAG also refers monitoring agencies to the Improved Environmental Monitoring software which was developed as part of a collaborative project between the UK Environment Agency, the Scotland and Northern Ireland Forum on Environmental Research, the Po River Authority and the Italian National Research Council (Caggiati et al. 2000). The project resulted in the publication of manuals of best practice for the development of water quality monitoring programmes for the UK and Italy. The UK manual of best practice
(Wyatt et al. 1998) is an accompaniment to the improved environmental monitoring software. The UK manual comprises the monitoring concepts and a detailed development of each monitoring strategies. Six monitoring strategies are discussed which are quality characterisation, spatial comparison, spatial trend detection, temporal comparison, temporal trend detection and Before-After-Control-Impact. The software and the manual do not give detail of the statistical procedures involved in the six monitoring strategies but the Equations (2.12) to (2.17) above would be the basis for any calculations.

In the Italian application of the improved environmental monitoring manual, the iterative process of monitoring programme optimisation is completed without the assistance of planning or data analysis tools. A case study of the Sesia River included as part of the Italian manual of best practice illustrated this elementary method of programme optimisation whereby an initial programme was implemented and following inspection of the results, uninformative monitoring points were removed from the plan and sampling frequency was reduced from weekly to monthly (Caggiati et al. 2000).

2.3.9. JEEP EC Tidal Monitoring Manual

The tidal monitoring manual published under the joint European estuaries research project advises on the consideration of tidal effects, residence time, river discharge, density and frequency of sampling and sampling locations and frequency (Kramer et al. 1994). Mid period sampling is advised in daily monitoring and observations on the spring-neap cycle should be taken at the midpoint in the spring neap cycle at mid-tide. The author also suggests that the residence time should guide the duration of the sampling effort, as sample data collected over two days in a well flushed waterbody will contain more variance than data collected over the same duration in poorly flushed waters. The sampling frequency recommended for physico-chemical parameters during the peak season for primary production is monthly and off peak sampling is bimonthly. Kramer also suggested that salinity would be a key consideration in the sampling of estuarine waters, due to its relationship with key water quality parameters. Regular monitoring at 1km intervals is recommended for
the upper freshwater part of the estuary and sampling along the salinity gradient should be carried out for the remainder of the estuary.

2.4. Combined use of modelling and monitoring for environmental compliance

The use of modelling as part of the planning process for WFD implementation is advocated by the European Commission (European Commission 2003c). Suggested uses include the identification of pressures, assessment of susceptibility and identification of interactions. Modelling is also promoted as a tool for the definition of reference conditions where no satisfactory waterbody exists, and for investigative purposes where a greater understanding is required (European Commission 2003c). Rekolainen has proposed an alternative DPCER framework (Rekolainen et al. 2003) to the well-known DPSIR framework (European Environment Agency 1999). This framework (Driving force – Pressures – Chemical state – Biological state – Response) for WFD implementation acknowledges the role of numerical models in the derivation of programmes of measures and estimation of maximum tolerable loads.

Within the context of WFD monitoring, modelling has a number of beneficial uses (Hojberg et al. 2007): data quality assurance, interpolation in space and time, conceptual model validation, assessment of anthropogenic activities, and design of monitoring programmes. Numerical modelling has the potential to assist in the assessment of anthropogenic activities and programmes of measures, as models may be able to account for natural variation such as rainfall-runoff which cannot be evaluated through monitoring data alone.

To date, the utilisation of models in the implementation of monitoring programmes for the EU WFD has been limited by time, skill, data availability (Harmoni-CA 2005) and the lack of confidence of policy makers in modelling (Brugnach et al. 2006). This lack of confidence in the integration of modelling in the monitoring process has resulted in the collection of monitoring data for analysis alone. Data is thus often unsuitable for model calibration and validation (Hojberg et al. 2007). Data scarcity has been cited as one of the main justifications for reliance on simplistic
models as opposed to mechanistic numerical models (Jorgensen et al. 2007). WFD article 5 reports generally indicated limited usage of models in the characterisation of water bodies due to poor data availability whilst classification and characterisation was completed relying on expert knowledge (Jorgensen et al. 2007).

Participant countries in the HarmoniRib project were found to have historically directed monitoring at the most persistent issues, thus resulting in the availability of comprehensive datasets for complex models to be assembled for further analysis (Jorgensen et al. 2007). Other potential restrictions to the usage of models include cost, lack of digitisation of paper data and incompatible datasets, as well as rivalry between competing organisations. Jorgensen concluded that, although usage of numerical models may not be inherently limited by data availability, water managers ought to be more mindful of the potential uses modelling throughout monitoring programme design, in order to preclude the possibility of data scarcity impeding future modelling activity.

A number of proposals have been made as to how models can be integrated into the monitoring process, including the establishment of an open central database for monitoring data and model results, standardisation of models/reduction in the number of models due to the variation in detail and usage amongst models, and regular communication to policy makers/water managers as to the abilities and limitations of models (Harmoni-CA 2005).

Surveillance monitoring data has been suggested as a good starting point for model assembly and validation, as surveillance data covers a wide range of chemical, biochemical and ecological parameters. Models produced from surveillance data could facilitate the identification of further data needs for operational and investigative monitoring (Harmoni-CA 2005). Quality assurance of monitoring data is currently incorporated in lab analysis of samples (European Commission 2003b). A suitably calibrated and validated model is potentially useful for confirmation that a parameter concentration is close to a model estimate. Agreement of monitoring and model data may also serve to bolster the confidence of policy makers in the uses of models. (Hojberg et al. 2007).
Mechanistic numerical models account for the governing processes which relate the water quality at one monitoring point to another non-measured location. Integration of monitoring data into a numerical model on an ongoing basis may strengthen the predictive and interpolative features. (Højberg et al. 2007). As most deterministic numerical models include the processes which form the basis for conceptual models, the use of monitoring data within numerical models also serves as a means of keeping the conceptual model updated.

Nielsen et al. (2003) showcased the ability of a numerical model to identify reference conditions in river basins in the absence of high status waterbodies in Randers Fjord, Denmark. Pristine conditions were simulated by only applying nutrient loads generated by leaching of non-agricultural areas to the fjord. Three models were applied to the fjord – a simple dilution model, the grey box model and a coupled hydrodynamic and eutrophication MIKE 12 model. Reference conditions, present conditions and water action plan conditions were modelled. MIKE 12 was found to produce the most accurate simulation of present conditions, whilst the simulation of the water action plan measures lead to the judgement that the programme of measures was inadequate and would not result in good ecological status.

Model output has been used for Lithuanian reporting to the European Commission. MIKE BASIN modelling of N, P and BOD was used to identify and rank pollution sources and risk of waterbody compliance, and test scenarios for potential programmes of measures (Harmoni-CA 2005).

Lessin and Raudsepp (2006) concluded that the spatial averages of water parameters yielded from a calibrated MIKE3 model corresponded closely with the results of one monitoring station, determining that monitoring point to be an optimum location.

Ferreira promotes the use of ecological model simulations as a means of predicting the response of the Ria Formosa, Portugal to a change in pressure. The results of model simulations were compared with the ASSETS screening model to determine the ecological status brought about by a change in nutrient loading (Ferreira et al. 2007).

Cozzi utilised the Princeton Ocean Model to identify the hydrological processes in the Gulf of Trieste, claiming a combination of monitoring and modelling was
achieved (Cozzi et al. 2004). However, monitoring data was discarded after model validation. The resultant model enabled tracer analysis to be undertaken to determine the fate of treated wastewater under different constant winds.

Hartnett et al. (2011a) evaluated a novel approach for determining TSAS trophic status in Cork Harbour with promising results. The DIVAST finite difference numerical model was utilised in order to overcome data scarcity and the uncertainty surrounding trophic assessment with limited monitoring data which purportedly represent spatial and temporal variability. The numerical model, which contains hydrodynamic, solute transport and water quality modules was fully calibrated and validated for the Cork Harbour area. A conservative tracer study was then carried out using the numerical model in order to determine the residence time for each cell in a 30m x 30m discretisation of the model domain. Residence times were used to delineate homogenous trophic assessment areas. TSAS trophic status assessment was carried out by running the validated model for the full suite of water quality parameters relevant to TSAS compliance for a 2 month representative summer period and 1 month winter period, pre and post installation of the Carrigrennan wastewater treatment plant. Water quality parameter concentrations were averaged spatially and temporally to determine representative values for each of the homogenous zones for DIN, MRP, dissolved oxygen and Chlorophyll_a. The results indicated the merits of using well calibrated and validated numerical model for determining representative homogenous water quality management zones and determining TSAS trophic status compliance. The resultant TSAS status for each management area matched the Irish EPA derived TSAS values for the same areas based on monitoring data. The procedure also shows promise in guiding decision makers in water quality management towards the correct actions which may result in improvement in water quality status.
### 2.5. Information theory and Entropy methods

Information entropy was established in the 1940s by Claude Shannon, initially to investigate noise in broadcasting of telegraphs (Shannon 1948, Shannon and Weaver 1949). Shannon’s measure of information content has had many applications since then in the broad fields of science and mathematics. Shannon’s entropy measures the information within a discretised PDF (probability density function)\(^1\), with broader PDFs generally yielding larger entropy than a narrow PDF, whilst a single value with absolute certainty has entropy of zero.

The measure can be applied to PDFs of any shape or size. Generally entropy theory is applied to PDFs developed from comprehensive datasets. Considering a PDF, the location of the peak indicates the mean value for a symmetrical, normal distribution, whilst the horizontal spread of the PDF can be described by the standard deviation or error variance. As the standard deviation or variance increases, so does the dispersion in the range of likely values for the parameter in question.

A PDF can be assembled from parameter data collected over a fixed space and time period at a continuous resolution, or from sparse data collected at irregular intervals and spatial locations. The confidence of the end user in the PDF is dependent upon prior knowledge of the data sampling routine in space and time; with a comprehensive dataset, one may have full confidence in the PDF to describe the data throughout the spatiotemporal domain of coverage. Shannon’s information entropy provides a suitable measure of the information contained in a PDF (Shannon 1948).

Shannon information entropy \((H(X))\), as applied to a PDF provides a measure of its shape; the measure incorporates the summation of the product of the probability densities and logarithms of the probabilities as shown in Equation (2.53):

\[
H(X) = -K \sum_{i} p_i \ln(p_i) = -\int_{-\infty}^{+\infty} f_x \ln(f_x) \, dx \tag{2.53}
\]

where \(p_i\) refers to the probability at each of the discretisations of the PDF. The latter term incorporating \(f_x\) describes the procedure where the PDF is a continuous distribution. \(K\) is a term introduced by the use of logarithms in the derivation of the entropy.

---

\(^1\) A PDF describes the likelihood of the occurrence of each value in a dataset; the area under the curve of a PDF must sum to 1, indicating all eventualities are described therein.
entropy of a PDF. The value for K must be positive (Shannon 1948). In all applications in the author’s work, a value of unity is used for K, which is a requirement in using natural logarithms in the determining entropy. Standard units for entropy determined using natural log are *nats*, while entropy determined using logs to the base 2 are *bits*.

Consider the situation where we have complete certainty about the value of a datum, we can say that the probability of the datum taking the value we ascribe to that point is 1. Determining the marginal entropy of that datum as follows, using Equation (2.53):

\[ H(X) = -1 \times (1 \times \log_e 1) = 0 \]  

(2.54)

We can see that marginal entropy is at a minimum of zero when we have absolute certainty about a point. There is little information contained in a single datum at a single coordinate in space and time regarding likely values at another time or spatial coordinate. If a PDF is discretized over the range of values it covers, the entropy at each probability in each of the value discretisations can be determined. For example, if in a given PDF, we take the values which are likely to occur with probabilities 0.001, 0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 0.99, 0.999, the entropy value for these probabilities is, using Equation (2.53):

\[ H(X) = -1 \times [(-0.0069) + (-0.046) + (-0.23) + (-0.347) \\
+ (-0.347) + (-0.215) + (-0.094) + (-0.0099) \\
+ (-0.000999)] = 1.296 \]  

(2.55)

It can be seen that one of the characteristics of the marginal Shannon information entropy measure results in values closer to the probability 0.5 contribute the most to the total entropy summation for the PDF, whilst values with extremely high or low probabilities contribute the least to the total entropy values.

It follows that with greater knowledge of the individual values that a variable may take and its probabilities, the entropy of the probability distribution increases.
Entropy theory henceforth shall refer to the application of the Shannon Information measure (Shannon and Weaver 1949). Entropy theory is underpinned by four measures:

- marginal entropy,
- joint entropy,
- conditional entropy,
- transinformation

The marginal entropy of a probability distribution is defined in Equation (2.53). The marginal entropy of a normal distribution is proportional to the standard deviation of the data, while marginal entropy of a uniform distribution is proportional to the interval length (Fuentes et al. 2007). Marginal entropy is a measure of the spread of a probability distribution. Marginal Entropy takes a value of 0 when a data series consists of a single value with a probability of 1 i.e. a completely certain outcome. A broad normal distribution will have higher entropy.

Joint entropy of a bivariate probability distribution of X and Y is defined by developing Equation (2.53) further, integrating the bivariate PDF over the two variables as presented in Equation (2.56) (Ozkul et al. 2000):

$$H(X,Y) = -K \sum_i \sum_j p_{ij} \ln(p_{ij}) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{xy} \ln(f_{xy}) \, dx \, dy$$

(2.56)

where $p_{ij}$ refers to the discrete values of the bivariate probability distribution of variables X and Y while $f_{xy}$ is the continuous PDF approximation of the discrete probability values of X and Y.

The total entropy of two independent variables is the sum of the respective entropies, while the total entropy of two stochastically dependent variables is less than the sum of the separate marginal entropies.

$$H(X,Y) \leq H(X) + H(Y)$$

(2.57)
Joint entropy of a multivariate probability distribution of $X_1,\ldots,X_n$ is defined as (Ozkul et al. 2000):

$$H(X_1,\ldots,X_n) = -K \sum_{i_1} \sum_{i_n} p_{i_1\ldots i_n} \ln(p_{i_1\ldots i_n}) \quad (2.58)$$

where $X_1,\ldots,X_n$ denotes the collection of random variables $X$ at $n$ locations, $n$ is the number of locations and $p_{i_n}$ refers to the discrete values of the multivariate probability distribution of variables $X_n$.

Considering a range of PDFs for a variable, with each PDF sourced from data at different locations, and each PDF is discretized at fixed intervals, it follows that one of the PDFs may have a greater spread of values and associated probabilities when compared to all others. This PDF would have maximum entropy and would describe the greatest variability. In the event that multiple PDFs are available at different locations and instances from comprehensive datasets, one may wish to identify the similarity or dissimilarity between different PDFs. Conditional entropy and transinformation are two complimentary entropy measures which determine the similarity between PDFs.

Conditional entropy is defined as:

$$H(X_2 | X_1) = H(X_1, X_2) - H(X_1) \quad (2.59)$$

where $H(X_2 | X_1)$ is the entropy of $X_2$ given $X_1$, $H(X_1, X_2)$ is the joint entropy between $X_1$ and $X_2$ and $H(X_1)$ is the marginal entropy of $X_1$. This is a measure of the information content in the data series of $X_2$ that cannot be accounted for in the data series of $X_1$.

Transinformation, which is a measure of the redundancy or commonality of information between two variables $X_1$ and $X_2$, is given by Equation (2.60):

$$T(X_1, X_2) = H(X_2) - H(X_2 | X_1) = H(X_2) - [H(X_1, X_2) - H(X_1)] \quad (2.60)$$

$$= H(X_1) + H(X_2) - H(X_1, X_2)$$
From Equation (2.60) above, it can be seen that the transinformation $T(X_1,X_2)$ between $X_1$ and $X_2$ is the difference between the marginal entropy of $X_2$ and the conditional entropy of $X_2$ given $X_1$. Thus, transinformation in the above example quantifies the difference between the uncertainty in relation to $X_2$ known beforehand and the uncertainty in relation to $X_2$ which remains given $X_1$.

Equation (2.61) expands upon Equation (2.60) by describing in the general sense the transinformation between $(m-1)$ monitoring points which have been added to an optimised monitoring network, and each of the candidate monitoring points to be added with $m^{th}$ priority.

$$T((X_1,\ldots,X_{m-1}),X_m) = H(X_1,\ldots,X_{m-1}) - H(X_1,\ldots,X_{m-1}|X_m)$$

(2.61)

where

$$H(X_1,\ldots,X_{m-1}|X_m) = H(X_1,\ldots,X_m) - H(X_m)$$

(2.62)

where $T((X_1,\ldots,X_{m-1}),X_m)$ refers to the transinformation between the $m-1$ existing monitoring points and the $m^{th}$ monitoring point to be assessed as an additional monitoring point, $H(X_1,\ldots,X_{m-1})$ refers to the multivariate entropy and $H(X_1,\ldots,X_{m-1}|X_m)$ is the conditional entropy multivariate entropy conditioned on the PDF at point $m$. $m$ refers to the size of the optimised monitoring network.

Spatial-temporal optimisation is possible after spatial optimisation, using Equation (2.63) after Ozkul et al. (2000). Starting with the smallest monitoring network of the two priority stations, the transinformation between the priority monitoring station $s_1$ at zero lag and the second priority station at multiples of the lag are considered. For each combination of $s_{\text{max}}$ stations, transinformation reaches a local minimum, which is considered the optimum monitoring frequency for that combination of stations.

$$T(X_{s_1,0},X_{s_2,l},X_{s_3,l},\ldots,X_{s_{\text{max}},l}) = H(X_{s_1,0}) - H(X_{s_1,0}|X_{s_2,l},X_{s_3,l},\ldots,X_{s_{\text{max}},l})$$

(2.63)
where $s_{\text{max}}$ is the maximum number of monitoring locations considered, $l$ is the sampling frequency, $T(X_{s1,0}, X_{s2,l}, X_{s2,l}, \ldots, X_{s_{\text{max}},l})$ is the transinformation between the base monitoring point which has the highest marginal entropy and datasets sampled at the remaining $s_{\text{max}}$ monitoring points; $H(X_{s1,0})$ refers to the marginal entropy of the PDF of all data from monitoring point $s_1$; $H(X_{s1,0} \mid X_{s2,l}, X_{s2,l}, \ldots, X_{s_{\text{max}},l})$ is the entropy of monitoring point $s_1$ conditioned on monitoring data from all of the remaining monitoring points sampled at sampling frequency $l$.

### 2.5.1. Application of Entropy methods to monitoring network design

Entropy analysis methods have been applied to a variety of monitoring network optimisation scenarios. Chen derived datasets at potential monitoring sites using kriging interpolation between existing monitoring points (Chen et al. 2008). An exponential model was fitted to the transinformation entropy between all candidate monitoring points to determine the number of monitoring points which would adequately describe rainfall. The outcome of the analysis determined that the majority of monitoring points should be relocated.

Mogheir determined the optimum spatial grid of groundwater monitoring sites by determining the relationship between station transinformation and distance (Mogheir et al. 2006). The limit value at which point transinformation reached a constant value was deemed the optimum spacing for monitoring points. A square grid constructed using the optimum spacing as side length was superimposed over the existing monitoring stations in order to augment or downsize the monitoring network. The spatial density of monitoring points was reduced from 1.9 wells per km$^2$ to 0.9 wells per km$^2$.

Karamouz applied entropy methods to the combined sampling of dissolved oxygen and electrical conductivity in a river monitoring network (Karamouz et al. 2009). Existing water quality data was used to calibrate the river water quality model QUAL2K (Chapra and Pellletier 2003). Time series of EC and DO were derived for each existing and potential monitoring location, while the two time series were aggregated using 3 different weightings. In each case, the first 10 out of the 28 proposed monitoring points were elected as priority locations, with the order varying slightly in each case. The optimum sampling frequency for the elected monitoring
network is determined using a transinformation measure between the first priority station at the existing sampling interval and the remaining monitoring points at a lag of multiples of the existing sampling interval. In each case, monthly sampling of EC and DO was recommended over the existing weekly sampling, while biweekly sampling was recommended for nutrients. The optimum sampling interval was deemed to be the interval for which transinformation reached a minimum.

Harmancioglu applied entropy to discrete river water quality data (Harmancioglu and Alpaslan 1992). It was concluded that monitoring could be extended to a bimonthly sampling interval with minimal loss of information, although consideration of daily, weekly or fortnightly sampling intervals was precluded due to the monitoring data being available at monthly resolution.

Through entropy methods, space-time trade-off can be considered explicitly (Harmancioglu and Alpaslan 1992, Karamouz et al. 2009, Ozkul et al. 2000). For a constant level of transinformation, a number of combinations of monitoring stations and sampling frequencies can be considered on cost grounds.

Contraction of monitoring programmes has been the objective in previous monitoring optimisation studies (Caeiro et al. 2004, Caeiro et al. 2003b, Dowdall et al. 2005, Karpuzcu et al. 1987, Naddeo et al. 2007, Tirsch and Male 1984). Entropy analysis alternatively facilitates an expansion of the existing monitoring programme instead of pursuing a contraction of the existing monitoring programme. While contraction of monitoring programmes may be targeted, there is a resultant loss of information from sampling at a lower spatial density or lower frequency; this may necessitate linear regression of discontinued monitoring points or parameters. Expansion or redesign of monitoring networks using entropy analysis would ensure that sampling in space and time reflected the underlying hydrodynamic and biogeochemical processes.

While previous work by other authors has been limited by poor resolution data (Harmancioglu and Alpaslan 1992, Ozkul et al. 2000), numerical models can generate results at varying resolution with the potential to investigate lags from daily intervals to quarterly. Numerical models have been utilised in previous entropy
studies to derive high resolution data for entropy analysis in river reaches (Karamouz et al. 2009).

2.6. Geostatistics and BME

Various measures of entropy analysis methods facilitate the identification of informative datasets which may describe a specific location or time; consequently the measures of marginal entropy and transinformation may assist in the prioritisation of monitoring, whether spatially or temporally. However, the process of monitoring network optimization, or redesign, may lead to discontinuity of monitoring, whilst monitoring cannot be practically carried out at every location or on a near continuous basis. In this context, geostatistical interpolation in its many forms is extremely relevant to the area of water quality monitoring as it enables the extrapolation of data from its coordinate to an unmonitored location.

The most relevant forms of classical geostatistical interpolation are simple and ordinary kriging. Simple and ordinary kriging differ in that ordinary kriging assumes that the mean trend is unknown, whereas simple kriging assumes that the mean trend is known and constant. Both kriging estimates are derived from the product of pre-existing data and weights determined in light of knowledge of spatial or temporal covariance. A full derivation of the simple kriging estimator is provided in Appendix 1. The simple kriging estimate given in Equation (2.64), as established by Krige and developed by Matheron (Krige 1953, Matheron 1963), is based on a number of assumptions:

- The generated estimate is a linearly weighted combination of the pre-existing observed values,
- Unbiasedness of the estimate is assumed. The expected value of the estimate is assumed to be equal to the expected value of the expression for the estimator,
- Mean squared error is minimised,
- The mean trend and the estimated covariance model are both known,
- Only hard, certain data is available
The simple kriging estimate $\hat{\chi}_k$ is given as follows:

$$\hat{\chi}_k = \bar{x}_k + \sum_{i=1}^{m} \frac{c_{ik}}{c_{kk}} (\chi_i - \bar{x}_i)$$

where $\bar{x}_k$ is the mean trend at the estimation point, $c_{ik}$ refers to the covariance vector between each of the $m$ data points and the estimation point $k$ and $c_{kk}$ refers to the variance at the estimation point. $\chi_i$ and $\bar{x}_i$ refer to the data value and the mean trend respectively at each of the $m$ data points.

BME was established by George Christakos (Christakos 1990, 1991) and further developed by Marc L. Serre (Serre 1999) as a more appealing alternative to kriging. Maximisation of Shannon’s information entropy is used as one of the constraints placed on the derivation of a PDF which describes all of the available monitoring data throughout the spatial extent of the monitoring network and the full duration of monitoring. The PDF is then updated using Bayes’ rule to describe the probable value of the monitoring parameter at a location where monitoring data is not available. The Shannon information criterion underpins BME where data availability is limited and entropy analysis theory where data scarcity is not an issue. The Shannon information criterion is used comparatively to identify after the fact the PDF with the greatest entropy or information. The other constraints placed upon the PDF take into account the spatial and temporal mean trend and covariance of the monitoring data throughout the data domain.

The PDF fully reflects the underlying data, resulting in an estimator which reflects all the available information. The Bayesian updating of the PDF results in a PDF describing the probable range of the water quality parameter at a location where no monitoring data is available, by making an allowance for the spatiotemporal covariance between the estimation point and the neighbouring data points, which can be described as hard or soft data points. Hard data points are locations where monitoring data can be considered definite, whereas soft data points are locations where there is uncertainty surrounding the exact value of the monitoring parameter and the range of probable values can be described by a PDF.

The spatial and temporal mean trend and covariance data derived from all available data for the parameter under investigation provide constraints to the solution of a
PDF which has maximum entropy. This PDF is deemed representative of the entire domain of the dataset in space and time. The PDF is then updated to take into account data in proximity to the estimation point and the final estimate is generated by Bayesian conditionalisation. The final estimate is the PDF of the data contained in the entire mapping space including the estimation point, given the data contained in the mapping space without the estimation point. Hence the PDF produced by Bayesian conditionalisation is a PDF that describes solely the estimation point.

Where kriging methods return estimates with a mean and standard deviation of a normal posterior PDF, there are no assumptions regarding the shape of the PDF. Hence the PDF accurately reflects the underlying data which could take the form of either high quality data with zero uncertainty, deemed hard data, or soft data which contained inherent uncertainty, such as interval data or PDFs. The method yields non-linear estimators. The BME estimator reduces to the simple kriging estimator in the limiting case where no soft probabilistic data exists.

The BME method of geostatistics can reduce uncertainty and inform the decision making process by processing soft uncertain data to support interpolation whilst avoiding the limiting assumption of a normal Gaussian distribution to the predictions for mapping purposes. The BME method has been proven to date in numerous journal publications in many diverse fields. BME has been applied to the Catawba River reservoir system in North Carolina, whereby hard monitoring data consisting of seasonal Chlorophyll_a levels were combined with model predictions of Chlorophyll_a from the BATHTUB empirical eutrophication model to produce posterior PDFs of Chlorophyll_a at all stations within the system (LoBuglio et al. 2007).

2.6.1. Spatiotemporal Random Field Theory

Spatiotemporal random field theory underpins BME (Christakos and Bogaert 1996). Christakos defines a spatiotemporal random field (STRF) X(p) as “a collection of complementary field realisations χ associated with the values of a natural variable at points p = (s,t) of a spatiotemporal continuum S x T” (Christakos 2000).
A STRF X(p) of the random variable x comprises a collection of random variables throughout the mapping space; all of the random variables are referred to in their entirety as x_{map}. x_{map} comprises x_{data}, which is the collection of individual random variables describing each monitoring point (x_{data} = (x_1, x_2, ..., x_m)), and x_k which refers to the as yet unknown random field at the estimation points. The subscript “map” denotes the entire mapping space, which includes all soft and hard data points, including the estimation points which are denoted with the subscript “k”. Each variable x_i is positioned at location p_i=(s_i, t_i) where p_i refers to the coordinate label, while s_i refers to the spatial location and t_i refers to the time. A realisation χ of an STRF takes a value at each of the k data points i.e. χ_{map} = (χ_1, χ_2, ..., χ_m, χ_k). The probability P(x<χ) that a random variable x takes a value less than one of its realisations χ can be summarised for any random variable by its CDF \( F_x(χ) = P[x < χ] \). A cumulative distribution function (CDF) indicates the probability that the value of a variable x will be less than a realisation χ.

The PDF is given by differentiating the CDF w.r.t. χ:

\[
f_x(χ) = \frac{dF_x(χ)}{dχ}
\]

Likewise the CDF is obtained by integrating the PDF w.r.t. χ:

\[
F_x(χ) = P[x < χ] = \int f(χ)dχ
\]

A PDF of a single random variable is described as univariate. A bivariate PDF describes the likely values of occurrence of two random variables whilst a multivariate PDF describes the likely value of occurrence of a collection of variables. The multivariate CDF \( F_x(χ_{map}) \) and the PDF \( f_x(χ_{map}) \) characterising all random variables (x_{map} = (x_1, x_2, ..., x_m, x_k)) in the random field X is defined as (Yu et al. 2008):

\[
F_x(χ_{map}) = P[χ_1 < x_1 < χ_1 + dχ_1, ..., χ_k < x_k < χ_k + dχ_k] = \int f(χ_{map}; p_{map})dχ_1 ... dχ_k \]

\( χ_{map} = χ_1, ..., χ_k \)
This can be simplified as:

\[ P[x_{map} < x_{map} < x_{map} + x_{map}] = \int f(x_{map}; p_{map}) dx_{map} \]  

(2.67)

The integral of the PDF \( f_x(x_{map}) \) by all realisations gives a value of one. The greater the number of random variables included in the multivariate PDF, the greater the knowledge retained in the PDF. However, the shape of the multivariate PDF \( f_x(x_{map}) \) is unknown. A satisfactory estimate of the multivariate PDF \( f_x(x_{map}) \) can be derived by using the general knowledge given by mean trend and covariance. An unsolved general solution to the multivariate PDF is contained in the general knowledge or moment equations used in the prior stage below. All known information pertaining to the mean trend and covariance is utilised resulting in a number of similar equations. The only unknowns in these equations lie within the unsolved general solution to the multivariate PDF.

BME analysis involves three distinct stages (Christakos 2000):

- Prior: assembles a prior general knowledge-based PDF which accounts for the relevant physical laws and the distribution of soft and hard field data using the mean and the covariance of data field
- Meta-prior: involves the processing of data and division of data into hard and soft data
- Posterior: integrates the general knowledge-based PDF with the specific knowledge base to give a PDF of the estimation point(s) in question.

The term prior PDF is conceptual; in effect, the term refers to the general expression for the posterior PDF before the expression is adapted to neighbouring data points that are adjacent to the estimation point; see appendix 2 for the derivation of the general solution of the posterior PDF.
2.6.2. The prior stage and entropy maximisation

At the prior stage of BME, the amount of information or entropy yielded from general knowledge within the prior PDF is maximised so that the optimally informative PDF is updated using Bayes’ rule at the posterior stage. The more informative a piece of knowledge is, the greater the number of eventualities which are deemed unlikely to occur. A more informative PDF represents a lesser likelihood, and hence there is an inverse relation between information and probability. As entropy and information are synonymous in this discussion, a PDF with maximum entropy contains the most information. As there is an inverse relation between information and probability, the PDF with maximum entropy or information has the narrowest probability range. Therefore, the PDF is as narrow as is necessary to describe all the available data which is available in terms of spatiotemporal mean trend and covariance.

The Shannon information metric (Shannon 1948) is as follows:

$$\text{Info}_G[\chi_{map}] = \log\{\text{Prob}_G[\chi_{map}]\}^{-1} = -\log\{\text{Prob}_G[\chi_{map}]\}$$ (2.68)

where \(\text{Info}_G\) refers to the general information contained in the realisations in \(\chi_{map}\) and \(\text{Prob}_G\) refers to the all of the probabilities of the realisations of \(\chi_{map}\). Extending Shannon’s information rule to all of the probabilities contained in the PDF \(f_G\) requires integration of the PDF \(f_G\) in a similar manner to Equation (2.65b), resulting in an updated Shannon information measure in Equation (2.69):

$$\text{Info}_G[\chi_{map}] = -\int d\chi \ f_G(\chi) \log f_G(\chi)$$ (2.69)

Equation (2.69) is expanded and maximised by solving the Euler Lagrange equation (Christakos 2000):

$$\frac{\partial (-f_G(\chi) \log f_G(\chi))}{\partial f_G} + \sum_{\alpha=0}^{N_c} \mu_\alpha \frac{\partial (g_\alpha(\chi)f_G(\chi))}{\partial f_G} = 0$$ (2.70)
Solving the above equation yields the general solution of the multivariate PDF $f_G$ parameterized by the Lagrange coefficients $\mu_\alpha$:

$$f_G(x_{map}) = \exp \left\{ \mu_0 + \sum_{i=1}^{n} \mu_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \mu_{ij} x_i x_j \right\} \quad (2.71)$$

where $\mu_0$ is the Lagrange weight corresponding to the normality constraint, the Lagrange weights $\mu_i$ correspond to the mean trend constraint and the Lagrange weights $\mu_{ij}$ correspond to the covariance constraint. The prior stage assembles a general knowledge-based PDF $f_G$, which accounts for the distribution of all available data by maximising the Shannon information entropy based on the mean trend and the covariance derived from the entire dataset. The general knowledge base $G$ can be summarised by a set of functions $G_\alpha (\alpha=0, 1, \ldots, N_c)$. $G$ is given by the following equations:

$$\bar{h}_\alpha(p_{map}) = G_\alpha[x_{map}, p_{map}; f_G] \quad (2.72)$$

$$g_\alpha(x_{map}) = \int d{x_{map}} g_\alpha(x_{map}) f_G(x_{map}, p_{map}) \quad (2.73)$$

$$\bar{h}_\alpha(p_{map}) = g_\alpha(x_{map}), \quad \alpha = 0, 1, \ldots, N_c \quad (2.74)$$

The number of constraints $N_C$ consists of one constraint accounting for normalisation, $n$ for the number of data points in the dataset and $n \times (n-1)/2$ for the covariance between all data points. There are $n \times (n-1)/2$ constraints derived from knowledge of the covariance between $n$ points because the covariance matrix sized $n \times n$ contains duplicates either side of the diagonal and hence half of the covariance matrix is discarded.

Equations are organised to follow the format $\bar{h}_\alpha = g_\alpha$ where $h_\alpha$ contains terms representing the mean or covariance of data, whilst $g_\alpha$ comprises terms such as the general formula for the prior PDF $f_G$ and a function of field values and co-ordinates.

$g_\alpha(x_{map})$ are known functions in $x_{map}$ such as means $\bar{x}_i$, variances $(x_i - \bar{x}_i)^2$ and centred covariances $(x_i - \bar{x}_i)(x_k - \bar{x}_k)$. A normalisation constraint is required to provide the initial normal shape of the PDF in advance of the implementation of further constraints derived from general knowledge. A zero order constraint accounts for normalisation such that $g_0 = 1$ and $g_0(x_{map}) = 1$. 

58
\( h_\alpha(p_{map}) \) and \( g_\alpha(x_{map}) \) do not always have similar forms. \( h_\alpha(p_{map}) \) relates to statistical expectation whereas \( g_\alpha(x_{map}) \) is a set of equations which are functions of \( x_{map}, p_{map} \) and the PDF \( f_G \).

All available general knowledge is utilised at the prior stage to generate these equations, each of which adds an additional constraint to the shape of the prior PDF \( f_G \). The PDF \( f_G \) is assumed to take the basic form \( f_G = e^{\mu_0 + \mu^T g} \). “\( g \)” refers to the vector of aforementioned general knowledge equations. “\( \mu_0 \)” and “\( \mu \)” refer to coefficients associated with the normalisation and general constraints respectively. The prior PDF is determined by substituting the unsolved general PDF into the general knowledge equations and solving for the vector of \( \mu \) coefficients (Christakos et al. 2002).

The spatiotemporal mean trend at each data point is required for the calculation of the covariance or variance which assists in the characterisation of the spatiotemporal field. Upon removal of the mean trend from all data points, the centred covariance of the residual is calculated as a measure of the autocorrelation between any two locations. Centred covariance is the covariance determined after removal of the mean trend, whereas non-centred covariance describes autocorrelation where the mean trend is unknown or not determined. Both centred and non-centred covariances can be utilised in BME, although non-centred covariance values derived without a mean trend model must describe all data variability which results in very short spatial and temporal autocorrelation ranges which are not meaningful or useful for geostatistical estimation. Centred covariance is defined by Equation (2.75) based on the assumption that the mean trend is determinable at all locations.

\[
C_x(p_i, p_j) = (x_i - \bar{x}_i)(x_j - \bar{x}_j)
\]

(2.75)

\( C_x(p_i, p_j) \) refers to the covariance between points \( i \) and \( j \) where \( p_i = (s_i, t_i) \) and \( (x_i - \bar{x}_i) \) refers to the residual \( res(s_i, t_i) \) and \( \bar{x}_i \) is the mean trend at point \( i \).
Once the covariances for all paired locations in the mapping area are determined, separable or non-separable models are fitted to the data which take into account the spatial distance or temporal time lapse between points. The selected model should be more accurate for closer points separated by smaller distances and time lapses, as adjacent data should have a greater influence on estimates than remote data. An example of a space-time covariance model is given by (Christakos et al. 2004):

\[
c_{s}(r_{ij}, \tau_{ij}) = c_{1}e^{\frac{3r_{ij}^{2}}{a_{11}^{2}}} + c_{2}e^{\frac{3r_{ij}^{2}}{a_{22}^{2}}}
\]

In the example in Equation (2.76), the primary model covariance model is Gaussian in space and exponential in time, although any range of combinations are possible, consisting of gaussian, exponential, cosine, sine, spherical or nugget models.

It is assumed in fitting a covariance model to raw covariance data that covariance is stationary at the scale and the spatial and temporal extent over which raw covariance is determined. The scale of spatial and temporal covariance stationarity has been assumed on a number of scales in previous applications of BME. LoBuglio et al. (2007) utilised the same covariance model in the derivation of BME estimates of Chlorophyll_a in a number of lakes remotely connected within the Catawba River basin in North Carolina, although no effort was made to cross validate estimates. Li et al. (2013) divided the adjoining Indian and Western Pacific Ocean into 8 adjacent subregions and derived separate sea surface temperature covariance models for each subregion thus assuming stationarity within each subregion. Christakos et al. (2004) used separate covariance models for BME estimation of ozone in 4 months. Akita et al. (2012) determined a separate covariance model centred on each estimation point using a moving window thus each concluding that fine particulate matter over the United States was quasi-stationary or stationary within a region. This was evidenced by a spatial gradient in the sill and spatial range of covariance across the contiguous United States.
2.6.3. The Meta-Prior Stage

The meta-prior stage involves the examination and division of data into (a) definite “hard” data and (b) probabilistic/interval “soft” data. In the context of water quality analysis, hard data can be readily identified as any site sample which has undergone lab analysis. Soft data can be split up as follows:

- Interval data: the value of a parameter will fall between two limits e.g. the 90% confidence interval.
- Probabilistic data: the probable value of a parameter at a location can be summarised by a PDF.

In the context of model data usage for BME, the model results at each node must be presented in the form of a mean and standard deviation of a Gaussian normal distribution.

2.7. The Posterior Stage

The posterior stage updates the general knowledge-based PDF by means of Bayesian conditionalization, whereby the prior PDF is adapted at each grid location to account for adjacent data in space and time. Bayes’ rule is incorporated into the formulation of the posterior PDF. The equation for the posterior PDF is a mathematical expression of the probable value at the location \( p_k \) given the knowledge of the spatiotemporal mean trend and covariance between the estimation point \( p_k \) and all other data coordinates.

Expression (2.77) below states after Bayes’ rule that the knowledge of the probable range of values for \( \chi_k \) at point \( p_k \) given the knowledge of all of the available data is obtained by dividing the general descriptor for the probable range of values throughout the dataset including the estimation point \( p_k \), by the specific descriptor for the probable range of values for the available dataset.

\[
Prob_K\{\chi_k|\chi_{data}(S)\} = \frac{Prob_K\{\chi_k \text{ and } \chi_{data}(S)\}}{Prob_K\{\chi_{data}(S)\}} \tag{2.77}
\]
If a bivariate PDF $f_{xx'}(\chi,\chi')$ describing the possible values of the variables $x$ and $x'$ is known, and the univariate PDF $f_{x'}(\chi')$ describing the possible values of $x'$ is known, then the univariate or marginal PDF $f_x(\chi)$ can be derived by integrating the bivariate PDF $f_{xx'}(\chi,\chi')$ with respect to $\chi'$ (Papoulis 2002).

Given a multivariate PDF $f_G(\chi_1, \chi_2, \chi_3, \chi_4)$ which describes the likely values of four random variables $x_1, x_2, x_3, x_4$, a marginal multivariate PDF with respect to $x_1, x_2$ and $x_3$ can be derived by integrating the multivariate by $d\chi_4$ hence removing the influence of the random variable $x_4$ on the PDF.

In the same manner, in the Bayesian conditionalisation process below the PDF $f_G(\chi_{data}) (= f_G(\chi_{hard\ data}, \chi_{soft\ data}))$ is derived by integration of the multivariate PDF $f_G(\chi_{map}) (= f_G(\chi_{hard\ data}, \chi_{soft\ data}, \chi_k))$ by $d\chi_k$ hence removing the information stored relating to the estimation point $p_k$. Through the Bayesian conditionalisation process, an equation for the univariate PDF $f_K(\chi_k) = f_k(\chi_k | \chi_{data})$ is derived.

The more specific descriptor for the posterior PDF, $f_{Kbc}(\chi_K)$ is given below.

$$f_{Kbc}(\chi_K) = A^{-1} \int_D d\Xi_S(\chi_{soft}) f_G(\chi_{map})$$

where

$$A = \int_D d\Xi_S(\chi_{soft}) f_G(\chi_{data})$$

Note with the equation for the Bayesian conditionnalised solution for the PDF at location $p_k$ that to the right of the integral sign, the integral $\int_D d\Xi_S(\chi_{soft})$ results in a multivariate PDF $f_S(\chi_{soft})$ of the likely values of the soft values adjacent to the point $p_k$. Note also that the multiplication of the resulting multivariate PDF $f_S(\chi_{soft})$ by $f_G(\chi_{map})$ yields an updated multivariate PDF of the likely values updated taking into account the adjacent soft data.

The integrand $\Xi_S$ takes on different forms depending upon the soft data type which is being integrated e.g. for interval data, the integrand becomes, for probabilistic data, the integrand is $F_S(\chi_{soft})$, which is the CDF of the data.
2.8. Conclusions

Ambient water quality legislation internationally shares a common objective of achieving satisfactory status, specifically described in a European context as “good surface water status” or “good ecological potential” under the WFD.

Numerical modelling has been suggested as a valuable adjunct to water quality management under the WFD (European Commission 2000a), with uses such as lowering uncertainty in water quality monitoring data, establishing reference conditions and assessing anthropogenic activities (Hojberg et al. 2007). Although numerical models are useful, their use is hampered by time requirements, skill shortage and data scarcity (Harmoni-CA 2005). In some cases, monitoring campaigns may be focused on data collection for model calibration and validation purposes, with data subsequently discarded.

Water quality monitoring and numerical modelling of water quality are oftentimes complimentary in terms of their uses and their end goals. While in some instances monitoring may be carried out with the specific purpose of aiding in model calibration and validation, monitoring is largely carried out with no consideration of the use of data in an eventual numerical model.

Locations for monitoring points have previously been chosen based on expert opinion to find a representative location (EPA 2006). These efforts are arbitrary and there is no basis outside of expert knowledge for sampling at the selected locations. The mandatory minimum sampling frequency set out in the WFD may be overruled based on expert knowledge but no instructive procedure exists to determine what frequency should be selected.

A number of conflicting priorities can also influence monitoring network design. In determining average conditions, serial correlation may be desirable, whereas many expressions which yield the number of samples required in a monitoring campaign assume all data points are independent and identically distributed. A large proportion of monitoring guidance is aimed towards monitoring within the river basin network in lakes, streams and rivers, with little reference to the estuarine environment which is the ultimate sink for watershed nutrients.
The highly dynamic estuarine environment is influenced by many factors such as riverine inflow, point and nonpoint discharge locations and quantities and tidal conditions. Qualitative guidelines on monitoring in estuaries suggest an approach based on capturing the salinity gradient in the waterbody. The location of monitoring points is however still open to interpretation. The spatial and temporal coverage provided by one monitoring point or one sampling event is unknown and within the highly heterogeneous estuarine environment is likely to vary throughout the waterbody.

Entropy theory is a flexible method by which monitoring points can be prioritised. It facilitates prioritisation of existing monitoring points based on minimising data redundancy, also known as transinformation, between monitoring points, based on their respective datasets. Entropy theory and analysis has considerable potential in transitional waters as a means of designing monitoring networks, given the highly dynamic nature of estuaries and the resultant spatial/temporal heterogeneity.

The BME modern method of geostatistics is suited to addressing the issues of monitoring data discontinuity, and uncertainty in trophic status assessment, that may be brought about by monitoring network redesign or contraction by methods such as entropy analysis. BME generates probabilistic estimates based on data with varying probabilistic certainty, such as water quality monitoring data and high resolution simulated water quality datasets generated by numerical models, which may have an inherent uncertainty.

Entropy theory yields dividends in two situations: one where data is scarce and one wishes to maximise the amount of information gleaned from the available data through derivation of BME estimates, and the other being the use of entropy theory to analyse the information contained in PDFs where data scarcity is not an issue due to comprehensive datasets which can provide PDFs. In the event that data scarcity is an issue and only a number of points with definite data exist and some uncertain soft data points exist, a parametric PDF can be derived to describe the likely values within the domain of the data, which can be manipulated to describe a PDF at any point within the data range. To the author’s knowledge, neither method has been applied to estuarine water bodies.
There is no clear methodology for monitoring in the estuarine environment. Additionally, there is no framework for synergistic iterative water quality monitoring and modelling in the estuarine environment.

The primary aim of this thesis is to address the shortage of suitable guidance on the optimisation of water quality monitoring in the estuarine environment, by establishing a methodology to optimise an existing monitoring regime or design a monitoring network outright with no previous monitoring network in place. As the optimal locations for monitoring may change due to environmental pressures over time, there may be a requirement to relocate monitoring points whilst retaining an understanding of water quality at redundant monitoring points.

The second aim of this thesis is to establish a means of deriving a reliable estimate of water quality at locations where monitoring has ceased or has yet to begin, as an adjunct to any monitoring regime.

The final aim of this work is to establish a framework for monitoring and modelling water quality in the transitional and coastal zone.
Chapter 3. Numerical Modelling

3.1. Introduction

The advent of the computer and the advancement of computing capabilities throughout the 20th century, coupled with a greater understanding of the solution to the partial differential equations which describe the conservation of mass and momentum and the advection and diffusion of solute, have facilitated great progress in the field of numerical modelling of physical scenarios in the engineering field. Numerical modelling of hydrodynamics, water quality and solute transport in the estuarine environment is a central pillar in water quality monitoring and environmental management and planning.

Amongst the factors influencing the choice of numerical model in a water quality study are the scale of the model domain, complexity, computational cost, data availability and the physical scenario under examination. Numerical modelling of the estuarine environment should focus on solving processes in the dimensions of interest and the underlying equations should be expressed in terms of the dimensions along which variation occurs and the remaining dimensions should be integrated to simplify the solution. Hence depth averaging would be suitable in a vertically well-mixed body, whilst areal averaging would be applicable to vertically stratified reservoirs and cross sectional averaging would be used in longitudinal modelling of rivers and estuarine reaches with uniform cross section. Simplification in modelling may lead to the possibility of inaccuracies. In addition, fewer parameters are available to calibrate the model against measurements, whilst calibration with remaining parameters may never result in the optimal simulation (Martin 1999).

The following chapter outlines the theory behind the numerical modelling methods which are available when considering the solution of the hydrodynamic and water quality and solute transport characteristics in surface waters. The approaches covered are rudimentary 0D modelling of estuaries, 1D longitudinal modelling of a riverine/estuarine reach and 3D modelling in complex estuarine and coastal
environments. Finally, the chosen 2D depth averaged hydrodynamic model is
detailed and the justification for its application in the author’s work is presented.

3.2. 0-D mass transport modelling

The most basic approach to modelling an embayment is to consider it as a box type
model as depicted in Figure 3.1. The Bay is assumed homogenous and the results
generated from Equations (3.1) to (3.4) provide an insight into the flushing in the
domain.

Figure 3.1 A zero-dimensional mass transport model representing the net exchange of
mass, influenced by riverine inflow, and exchange at the tidal boundary; modified from
Lung (1993) [Image removed for copyright reasons]

Continuity of mass in the system is used to obtain values for $Q_i$, the volume of
seawater entering the Bay for the first time, $Q_o$ is the volume of well-mixed water
leaving the Bay during ebb tide. From these values, the flushing rate can be
determined (Lung 1993). Knowledge of the river inflow $Q_R$, the embayment salinity
$S$ and the ambient seawater salinity $S_i$ is a requirement for the calculation. Assuming
the Bay is at steady state and tidally averaged:
\[ Q_l S_l = Q_o S \]  
(3.1)

\[ Q_l + Q_R = Q_o \]  
(3.2)

\[ Q_l = \frac{Q_R S}{S_l - S} \]  
(3.3)

\[ Q_o = \frac{Q_R S_l}{S_l - S} \]  
(3.4)

The aforementioned approach is suitable where an outline view of hydrodynamics is wanted for inter-comparison between a number of similar waterbodies. Some shortcomings exist in the application of this approach; for example, the method does not account for localised variation in the flushing rate.

3.3. 1D hydrodynamic modelling

One dimensional hydrodynamic modelling may be applicable where flow is homogenous throughout a cross section as presented in Figure 3.2. The solution of the following St Venant equations of one dimensional conservation of mass and momentum yields predictions of water velocity and elevations (Liang et al. 2007):
Continuity equation
\[ \frac{\partial A}{\partial t} + \frac{\partial Q}{\partial s} = Q_L \]  \hspace{1cm} (3.5)

Momentum equation
\[ \frac{\partial Q}{\partial t} + \frac{\partial \left( \frac{\beta Q^2}{A} \right)}{\partial s} = -gA\frac{\partial \zeta}{\partial s} - gAS_f \]  \hspace{1cm} (3.6)

Advection diffusion equation
\[ A \frac{\partial \phi}{\partial t} + \frac{\partial (Q\phi)}{\partial s} = AD_x \frac{\partial^2 \phi}{\partial x^2} \]  \hspace{1cm} (3.7)

where \( A \) is the cross sectional area, \( t \) is time, \( Q \) is the discharge, \( s \) is the coordinate along the reach, \( Q_L \) is the lateral discharge per unit length, \( \beta \) is the momentum correction factor, \( \zeta \) is the water surface elevation, \( S_f \) is the friction slope, defined in terms of the Chezy coefficient \( C \) as \( S_f = \frac{\frac{|Q|}{C^2A^2R}}{C^2A^2R} \) and the Chezy coefficient is defined as \( C = 1.49 \frac{R^{\frac{1}{n}}}{n} \). \( \phi \) refers to the solute concentration while \( D_x \) is the longitudinal dispersion coefficient.

Examples of 1D modelling include HEC-RAS (Brunner 2010), Mike11 (DHI 2009), and ISIS (Halcrow and HR.Wallingford 1999). Suitable finite difference solution schemes for calculation of water elevation and velocity include the Preissman implicit 4 point box scheme (Preissmann 1961), the leapfrog implicit scheme or 6 point Abbott-Ionescu scheme (Abbott and Ionescu 1967). The application of 1D hydrodynamic modelling in estuaries is limited to shallow waters with little variation in velocity with depth and width and is most suitable for the interface between the river and tidal waters. Although the results generated by 1D models are limited, there is scope for further development with off channel storage cells for flood modelling (Liang et al. 2007), referred to as a pseudo-2D method. Coupled 1D-2D models are available commercially (DHI 2007) whilst the standalone ISIS 1D and DIVAST 2D models have also been assembled successfully (Lin et al. 2006).
3.4. 3D Navier-Stokes equations

The 3D Navier-Stokes equations form the basis of detailed modelling of estuarine and coastal waters. The equations describe continuity of mass and momentum, incorporating Newton’s equations of motion. The 3D momentum equation accounts for local (1) and advective acceleration (2), the Coriolis force (3), pressure gradient (4), wind shear stress (5), bottom shear stress (6) and turbulent shear forces (7). A number of finite difference models solve the 3D Navier-Stokes equations by dividing the domain into layers which are depth averaged (Liang et al. 2007, Lin and Falconer 1997, USEPA 2002). The layer integrated 3D Navier-Stokes and advection-diffusion equations are given as follows in Equations (3.8) – (3.11) (Falconer 1993):

Continuity equation

\[
\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0
\]

(3.8)

X direction momentum equation

\[
\frac{\partial u}{\partial t} + \frac{\partial u^2}{\partial x} + \frac{\partial uv}{\partial y} + \frac{\partial uw}{\partial z} = f v - \frac{gu}{4} + \frac{\rho a C W_x (W_x^2 + W_y^2 + W_z^2)^{\frac{1}{2}}}{\rho}
\]

\[- \frac{gu(u^2 + v^2 + w^2)^{\frac{1}{2}}}{C^2} + \frac{2}{\partial x} \left[ \frac{\partial u}{\partial x} \right] + \frac{\partial}{\partial y} \left[ \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right] + \frac{\partial}{\partial z} \left[ \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right]
\]

(3.9)

Y direction momentum equation

\[
\frac{\partial v}{\partial t} + \frac{\partial uv}{\partial x} + \frac{\partial v^2}{\partial y} + \frac{\partial vw}{\partial z} = -fu - \frac{g\xi}{\partial y} + \frac{\rho a C W_y (W_x^2 + W_y^2 + W_z^2)^{\frac{1}{2}}}{\rho}
\]

\[- \frac{gv(u^2 + v^2 + w^2)^{\frac{1}{2}}}{C^2} + \frac{\partial}{\partial x} \left[ \frac{\partial v}{\partial x} + \frac{\partial u}{\partial x} \right] + \frac{2}{\partial y} \left[ \frac{\partial v}{\partial y} + \frac{\partial u}{\partial y} \right]
\]

(3.10)

Advection diffusion equation
where:

\[ \frac{\partial \phi}{\partial t} + \frac{\partial u\phi}{\partial x} + \frac{\partial v\phi}{\partial y} + \frac{\partial w\phi}{\partial z} = D_{xx} \frac{\partial^2 \phi}{\partial x^2} + D_{yy} \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial}{\partial z} \left[ D_{zz} \frac{\partial \phi}{\partial z} - w_s \phi \right] + \Phi \]

\[ (3.11) \]

\text{and:}

\[ u, v = x \& y \text{ direction velocity} \]
\[ U, V = x \& y \text{ direction depth integrated velocity} \]
\[ \zeta = \text{water elevation above/below mean water level} \]
\[ q_x, q_y = \text{depth integrated volumetric flux in } x \text{ and } y \text{ direction } \left( q_x = UH, q_y = VH \right) \]
\[ \beta = \text{momentum correction factor for non-uniform vertical velocity profile} \]
\[ f = 2\omega \sin \theta = \text{Coriolis parameter} \]
\[ \omega = \text{angular velocity of earth’s rotation} \]
\[ \theta = \text{geographic latitude} \]
\[ C = \text{Chezy bed roughness coefficient} \]
\[ \nu_t = \text{depth-average mean eddy viscosity} \]
\[ \rho_a = \text{density of air (1.292 kg/m}^3) \]
\[ \rho = \text{fluid density} \]
\[ \varepsilon = \text{depth mean eddy viscosity} \]
\[ g = \text{gravitational acceleration} \]
\[ W_x, W_y, W_z = \text{wind velocity in } x \text{ and } y \text{ direction} \]
\[ C^* = \text{air-water interface resistance coefficient} \]
\[ H = \text{total water depth} \]
\[ \phi = \text{depth averaged solute concentration} \]
\[ D_{xx}, D_{xy}, D_{yx}, D_{yy} = \text{Dispersion coefficients in } x \text{ and } y \text{ directions} \]
\[ \Phi_\phi = \text{all other sources and sinks: outfalls, rivers, chemical and biological transformations} \]

\textbf{3.5. 2D depth averaged hydrodynamic modelling}

In shallow, well-mixed estuaries with no stratification, the Navier-Stokes equations can be reduced to describe continuity of mass and momentum in two horizontal orthogonal directions, with the third dimension of depth removed by integration. DIVAST is a finite difference model which solves the depth integrated continuity and Navier-Stokes equations as seen in Equations (3.14) to (3.16), yielding current velocities and water surface elevations, whilst the advection diffusion equation forms the basis for solute transport and water quality modules. The model suited the chosen site of Cork Harbour as the waterbody is vertically well-mixed.

Depth integration of Equations (3.8) to (3.11) reduces the governing equations from three dimensions to two, as shown in Figure 3.3, which simplifies the computational complexity of the solution scheme. Depth integration is allowable in shallow, vertically well-mixed waters due to vertical velocities being negligible compared to
the horizontal orthogonal components and vertical accelerations being negligible compared to gravity. In a stratified water body, the vertical velocity profile is parabolic (Schlichting 1979), whereas depth integration in a well vertically mixed water body assumes a constant velocity with depth.

Figure 3.3 Velocity profile with depth in (a) a 3-D stratified model and (b) in a shallow, depth integrated model

The orthogonal horizontal velocity components U and V which are integrated over the depth H in the vertical z direction are given by Equations (3.12) and (3.13) where h is the water depth below mean water level and ζ is the water elevation above or below mean water level, as depicted in Figure 3.4.

\[
U = \frac{1}{H} \int_{-h}^{\zeta} u \, dz \\
V = \frac{1}{H} \int_{-h}^{\zeta} v \, dz
\] (3.12) (3.13)

Figure 3.4 A shallow, well mixed water body (M.W.L. = mean water level) \( \zeta = \) water elevation above or below M.W.L., \( h = \) water depth below M.W.L.
The continuity Equation (3.14) considers that the mass entering a finite space must be equal to the mass leaving that space. The X or Y direction momentum equations as presented in Equations (3.15) and (3.16) contain the momentum terms on the left hand side from local and advective accelerations, whilst the terms on the right hand side represent the causative forces which consist of the Coriolis affect, pressure gradients, wind shear stress, bed resistance and turbulence. The depth integrated hydrodynamic equations are given by (Falconer 1994):

**Continuity equation**

\[
\frac{\partial \zeta}{\partial t} + \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} = 0
\]  

(3.14)

**X direction momentum equation**

\[
\frac{\partial q_x}{\partial t} + \beta \left[ \frac{\partial U q_x}{\partial x} + \frac{\partial U q_y}{\partial y} \right] = f q_y - g H \frac{\partial \zeta}{\partial x} + \frac{\rho a CW_x (W_x^2 + W_y^2)^{\frac{1}{2}}}{\rho} - \frac{g U (U^2 + V^2)^{\frac{1}{2}}}{C^2} + 2 \frac{\partial}{\partial x} \left[ \varepsilon H \frac{\partial U}{\partial x} \right] + \frac{\partial}{\partial y} \left[ \varepsilon H \left( \frac{\partial U}{\partial y} + \frac{\partial V}{\partial x} \right) \right]
\]  

(3.15)

**Y direction momentum equation**

\[
\frac{\partial q_y}{\partial t} + \beta \left[ \frac{\partial V q_x}{\partial x} + \frac{\partial V q_y}{\partial y} \right] = f q_x - g H \frac{\partial \eta}{\partial y} + \frac{\rho a CW_y (W_x^2 + W_y^2)^{\frac{1}{2}}}{\rho} - \frac{g V (U^2 + V^2)^{\frac{1}{2}}}{C^2} + 2 \frac{\partial}{\partial y} \left[ \varepsilon H \frac{\partial V}{\partial y} \right] + \frac{\partial}{\partial x} \left[ \varepsilon H \left( \frac{\partial V}{\partial x} + \frac{\partial U}{\partial y} \right) \right]
\]  

(3.16)

**Advection diffusion equation**

\[
\frac{\partial H \phi}{\partial t} + \frac{\partial H U \phi}{\partial x} + \frac{\partial H V \phi}{\partial y} = \frac{\partial}{\partial x} \left[ D_{xx} \frac{\partial \phi}{\partial x} + D_{xy} H \frac{\partial \phi}{\partial y} \right] + \frac{\partial}{\partial y} \left[ D_{yy} H \frac{\partial \phi}{\partial y} + D_{yx} \frac{\partial \phi}{\partial x} \right] + \Phi \phi
\]  

(3.17a)

\[
D_{xx} = \frac{(k_p U^2 + k_U V^2) \sqrt{H \rho}}{c \sqrt{U^2 + V^2}} + D_w
\]  

(3.17b)

\[
D_{yy} = \frac{(k_p V^2 + k_U U^2) \sqrt{H \rho}}{c \sqrt{U^2 + V^2}} + D_w
\]  

(3.17c)

\[
D_{xy} = D_{yx} = \frac{(k_p - k_U) H U V \sqrt{H \rho}}{c \sqrt{U^2 + V^2}}
\]  

(3.17d)

where:

- \( u, v = x & y \) direction velocity
- \( U, V = x & y \) direction depth integrated velocity
- \( \zeta \) = water elevation above/below mean water level
- \( q_x, q_y \) = depth integrated volumetric flux in x and y direction (\( q_x = U H, q_y = V H \))
- \( \beta \) = momentum correction factor for non-uniform vertical velocity profile
Numerical Modelling

\[ f = 2\omega \sin \theta = \text{Coriolis parameter} \]
\[ \omega = \text{angular velocity of earth’s rotation} \]
\[ \theta = \text{geographic latitude} \]
\[ C = \text{Chezy bed roughness coefficient} \]
\[ v_t = \text{depth-average mean eddy viscosity} \]
\[ \rho_a = \text{density of air (1.292 kg/m}^3) \]
\[ \rho = \text{fluid density} \]
\[ \varepsilon = \text{depth mean eddy viscosity} \]
\[ g = \text{gravitational acceleration} \]
\[ W_x, W_y, W_z = \text{wind velocity in x and y direction} \]
\[ C^* = \text{air-water interface resistance coefficient} \]
\[ H = \text{total water depth} \]
\[ \phi = \text{depth averaged solute concentration} \]
\[ D_{xx}, D_{xy}, D_{yx}, D_{yy} = \text{Dispersion coefficients in x and y directions} \]
\[ k_p = \text{longitudinal depth averaged dispersion constant} \]
\[ k_l = \text{depth averaged turbulent diffusion constant} \]
\[ D_w = \text{wind induced dispersion coefficient} \]
\[ c = \text{Chezy coefficient} \]
\[ \Phi_{\phi} = \text{all other sources and sinks: outfalls, rivers, chemical and biological transformations} \]

3.6. DIVAST hydrodynamic solution scheme

The finite difference solution of the continuity and momentum equations employs an alternating direction implicit (ADI) scheme. The central differences method of expressing derivatives is utilised. The ADI scheme entails the division of each timestep into two half-timesteps; each half-timestep considers implicitly the solution of the water depth and velocity component in one direction whilst the water depth and velocity component on the other direction are considered explicitly, thus avoiding the complexity of solving a two dimensional matrix at each timestep (Falconer et al. 2001). A space staggered orthogonal grid is used as presented in Figure 3.5; the marker and cell method is used (Harlow and Welch 1965) whereby calculations proceed on the basis that water elevations are computed at the centre of each grid cell whilst the velocities and water depths in the x and y directions are solved at the centre of the sides of the grid cells. As shown in Figure 3.5, the model is discretized on the I-J plane, where I axis corresponds to the x direction and J corresponds to the y direction and the origin of the I-J axes is positioned at the upper left extreme of the model domain.
The hydrodynamic equations are solved using Gaussian elimination and back substitution (Falconer 1977). The equations are rearranged with the unknown variables on the left hand side, and the known variables from the previous timestep on the right hand side, as seen in Equations (3.19) and (3.25). The finite difference representations of the continuity equation, the momentum equation and the advection diffusion equation are solved in the x direction for the first half timestep and the y direction for the second timestep. All known data within the formulae are grouped into recursion coefficients $a_i$, $b_i$, $c_i$, $d_i$, $e_i$ and $f_i$ and the formulae reduce to simplified recursion formulae. The equations are arranged in a tri-diagonal matrix to execute simultaneous equations. Application of open boundary conditions initiates the solution of the simultaneous equations via Gaussian elimination and back substitution.

Continuity equation in finite difference format:

$$\zeta_{i,j}^{n+\frac{1}{2}} - \zeta_{i,j}^n + \frac{\Delta t}{2\Delta x} \left[ q_x |_{i+\frac{1}{2},j}^{n+\frac{1}{2}} - q_x |_{i-\frac{1}{2},j}^{n+\frac{1}{2}} + q_y |_{i,j+1/2}^n - q_y |_{i,j-1/2}^n \right] = 0$$

(3.18)

Recursive continuity equation

$$-d_i q_x |_{i-\frac{1}{2},j}^{n+1} + e_i \zeta_{i,j}^{n+\frac{1}{2}} + f_i q_x |_{i+\frac{1}{2},j}^{n+\frac{1}{2}} = A^n_i$$

(3.19)
Numerical Modelling

\[ d_i = f_i = \frac{\Delta t}{2\Delta x} \]  

and

\[ e_i = 1 \]  

and

\[ A_i^n = \zeta_{i,j} - \frac{\Delta t}{2\Delta x} \left[ q_y |_{i+\frac{1}{2}j}^{n+1} - q_y |_{i-\frac{1}{2}j}^{n} \right] \]  

\( X \) direction momentum equation in finite difference format:

\[ q_x |_{i+\frac{1}{2}j}^{n+\frac{1}{2}} = q_x |_{i+\frac{1}{2}j}^{n-\frac{1}{2}} \]

\[ -\frac{\beta \Delta t}{\Delta x} \left[ U' q_x |_{i+1,j}^{n} - U' q_x |_{i,j}^{n} + U' q_y |_{i+\frac{1}{2}j+\frac{1}{2}}^{n} - U' q_y |_{i+\frac{1}{2}j-\frac{1}{2}}^{n} \right] \]

\[ + \Delta t f q_y |_{i+\frac{1}{2}j}^{n} \]

\[ + \frac{\Delta t}{2\Delta x} H^n \left[ \zeta_{i+1,j}^{n+\frac{1}{2}} + \zeta_{i,j}^{n-\frac{1}{2}} - \zeta_{i+1,j}^{n-\frac{1}{2}} - \zeta_{i,j}^{n+\frac{1}{2}} + \rho_a \Delta t C W_x (W_x^2 + W_y^2) \right]^{\frac{1}{2}} \]

\[ \frac{\rho}{2C^2} \left[ U^{n+\frac{1}{2}} + U^{n-\frac{1}{2}} + (U'^2 + V'^2) \right] \]

\[ + \frac{\Delta t}{\Delta x} H^n |_{i+\frac{1}{2}j}^{n} \left[ U'_{i+\frac{1}{2}j} + U'_{i-\frac{1}{2}j} + U'_{i+\frac{1}{2}j+1} + U'_{i+\frac{1}{2}j-1} - 4U'_{i+\frac{1}{2}j} \right] \]  

Solute transport equation in finite difference form

\[ (\phi H) |_{i+\frac{1}{2}j}^{n+\frac{1}{2}} = (\phi H) |_{i+\frac{1}{2}j}^{n-\frac{1}{2}} - \frac{\Delta t}{2\Delta x} \left[ \phi q_{x,i+\frac{1}{2}j}^{n+\frac{1}{2}} - \phi q_{x,i+\frac{1}{2}j}^{n-\frac{1}{2}} + \phi q_{y,i+\frac{1}{2}j}^{n+\frac{1}{2}} - \phi q_{y,i+\frac{1}{2}j}^{n-\frac{1}{2}} \right] \]

\[ + \frac{\Delta t}{2\Delta x} \left[ (HD_{xx})_{i+\frac{1}{2}j}^{n+\frac{1}{2}} \left( \phi_{i+1,j}^{n+\frac{1}{2}} - \phi_{i,j}^{n+\frac{1}{2}} \right) - (HD_{xx})_{i+\frac{1}{2}j}^{n-\frac{1}{2}} \left( \phi_{i+1,j}^{n-\frac{1}{2}} - \phi_{i,j}^{n-\frac{1}{2}} \right) \right] \]

\[ + (HD_{xy})_{i+\frac{1}{2}j}^{n+\frac{1}{2}} \left( \phi_{i+1,j}^{n+\frac{1}{2}} - \phi_{i+\frac{1}{2}j+\frac{1}{2}}^{n+\frac{1}{2}} \right) - (HD_{xy})_{i+\frac{1}{2}j}^{n-\frac{1}{2}} \left( \phi_{i+1,j}^{n-\frac{1}{2}} - \phi_{i+\frac{1}{2}j+\frac{1}{2}}^{n-\frac{1}{2}} \right) \]

\[ + (HD_{yx})_{i+\frac{1}{2}j}^{n+\frac{1}{2}} \left( \phi_{i+\frac{1}{2}j+\frac{1}{2}}^{n+\frac{1}{2}} - \phi_{i,j}^{n+\frac{1}{2}} \right) - (HD_{yx})_{i+\frac{1}{2}j}^{n-\frac{1}{2}} \left( \phi_{i+\frac{1}{2}j+\frac{1}{2}}^{n-\frac{1}{2}} - \phi_{i,j}^{n-\frac{1}{2}} \right) \]

Recursive X direction momentum equation

\[ -a_i \zeta_{i,j}^{n+\frac{1}{2}} + b_i q_x |_{i+\frac{1}{2}j}^{n+\frac{1}{2}} + c_i \zeta_{i+1,j}^{n+\frac{1}{2}} = B_i^n \]

where

\[ a_i = c_i = \frac{g \Delta t}{2\Delta x} H^n |_{i+\frac{1}{2}j}^{n} \]
Numerical Modelling

and

\[ b_i = 1 + \frac{g \Delta t (q_x^2 + q_y^2)}{2 (HC)^2} \]  

(3.27)

and

\[ B_i^n = q_x^n \left|_{i+\frac{1}{2},j} \right. \frac{1}{2} - \frac{\beta \Delta t}{\Delta x} \left( U'q_x |_{i+1,j,\frac{1}{2}} - U'q_x |_{i,j} \right) + \frac{g \Delta t}{2 \Delta x} H^n |_{i+\frac{1}{2},j,\frac{1}{2}} \left[ \zeta_{i+1,j,\frac{1}{2}} - \zeta_{i,j} \right] + \frac{\rho_a \Delta t C W_x (W_x^2 + W_y^2)^{\frac{1}{2}}}{\rho} \]

\[ - \frac{g \Delta t (q_x^2 + q_y^2)}{2 (HC)^2} \left|_{i+\frac{1}{2},j} \right. q_x^{\frac{1}{2}} \]

\[ + \frac{\Delta t}{\Delta x^2} e H^n |_{i+\frac{1}{2},j} \left[ U' |_{i+\frac{1}{2},j,\frac{1}{2}} + U' |_{i-\frac{1}{2},j} + U' |_{i+\frac{1}{2},j+1} + U' |_{i+\frac{1}{2},j-1} - 4U' |_{i+\frac{1}{2},j} \right] \]  

(3.28)

where \( n \) refers to the timestep number and \( i \) and \( j \) subscripts refer to parameters along the \( x \) and \( y \) directions respectively, \( q_x \) and \( q_y \) are the depth integrated volumetric fluxes in the \( x \) and \( y \) direction, \( H \) refers to the water depth, \( U \) and \( V \) are the depth integrated velocities, \( g \) is acceleration due to gravity, \( \zeta \) is the water elevation, \( \Delta t \) refers to the timestep and \( \Delta x \) refers to the grid spacing.

As the terms \( q_x \) and \( \zeta \) and solute \( \phi \) are solved implicitly during the first timestep, terms with a prime such as \( U' \) are expressed explicitly during the first iteration \( n \) as their value from the previous timestep \( n-\frac{1}{2} \) previous values as in Equation (3.29a) and a centrally averaged value for the second iteration, as seen in Equation (3.29b):

\[ U'^{n,1} |_{i+\frac{1}{2},j} = U^{n-\frac{1}{2}} |_{i+\frac{1}{2},j} \]  

(3.29a)

\[ U'^{n,1} |_{i+\frac{1}{2},j} = \frac{1}{2} \left( U^{n+\frac{1}{2}} |_{i+\frac{1}{2},j} + U^{n-\frac{1}{2}} |_{i+\frac{1}{2},j} \right) \]  

(3.29b)
Given the above expressions for the recursion coefficients $a_i, b_i, c_i, d_i, e_i$ and $f_i$ and the terms $A_i^n$ and $B_i^n$, the solution of the equations for each half timestep can be obtained by Gaussian elimination and back substitution. The solution to the system of equations is initiated at the open boundary where the conditions are known either in terms of water surface elevation or velocity distribution.

Figures 3.6 (a) and (b) detail a sample scenario where the open boundary condition highlighted in blue is a water elevation. The water elevation boundary value $\zeta$ is defined along cells (1, 2) to (9, 2). With this knowledge in mind, gaussian elimination and back substitution can be applied at points (1, 2), ($1^{1/2}$, 2) and (2, 2).
Considering Equation (3.25), assuming the value for $\zeta_{i,j}$ is known and rearranging in terms of the discharge $q_x$ at timestep $n^{1/2}$ at point $(i+1/2, j)$ yields an expression of the following form:

$$q_x|_{i+\frac{1}{2},j}^{n+\frac{1}{2}} = -R_i\zeta_{i+\frac{1}{2},j}^{n+\frac{1}{2}} + S_i$$  \hspace{1cm} (3.30)

where

$$R_i = \frac{c_i}{b_i + a_i P_i}$$

$$S_i = \frac{B_i^n + a_i Q_i}{b_i + a_i P_i}$$

The new expression for $q_x$ at timestep $n^{1/2}$ at point $(i+1/2, j)$ can then be inserted into Equation (3.31), which thus reduces to an equation of the following form:

$$\zeta_{i,j}^{n+\frac{1}{2}} = -P_i q_x|_{i+\frac{1}{2},j}^{n+\frac{1}{2}} + Q_i$$  \hspace{1cm} (3.31)

$$P_i = \frac{c_i}{b_i + d_i R_{i-1}}$$

$$Q_i = \frac{A_i^n + d_i S_{i-1}}{e_i + d_i R_{i-1}}$$

Equation (3.31) when considered for the specific case of $i=2$ and $j=2$ can then be reinserted into Equation (3.25) which eliminates the term $\zeta_{2,2}$ at timestep $n^{1/2}$. The upper boundary has no flow across it, hence $q_x$ at point $(12^{1/2}, 2)$ is zero. Hence the iterative procedure is repeated until the closed upper flow boundary can be incorporated into the Equation (3.31) and the solution to the value of $\zeta_{12,2}$ is found using Equation (3.31). The value for water elevation $\zeta$ at point $(12, 2)$ is back substituted into Equation (3.30) hence yielding the solution to $q_x$ at timestep $n^{1/2}$ at point $(11^{1/2}, 2)$. The process is repeated until all elevation and discharge terms are solved for in the area highlighted in Figure 3.6(b). The y direction is solved for the timestep $n+1$ in a similar way.
3.6.1. Stability, Accuracy and Convergence

As a consequence of any numerical model being an approximation of the underlying partial differential equations, stability, accuracy and convergence are important considerations in choosing a numerical model and refining its operation.

Model stability is influenced by round-off errors which are due to calculations being made to a fixed number of decimal places. It is desirable that the round off errors cancel out; otherwise the results may become unstable and the rounding error of each subsequent calculation will be amplified. DIVAST is stable as it is time centred for advective acceleration by iteration and the ADI scheme is naturally stable as a result of the application of Gaussian elimination and back substitution. Accuracy of the DIVAST solution scheme can be maintained by considering the Courant number which in most applications is limited to a maximum value of 8 to maintain accuracy (Falconer et al. 2001).

\[ C_n = \sqrt{gH} \frac{\Delta t}{\Delta x} \] (3.32)

where \( \Delta x \) and \( \Delta t \) refer to the grid spacing and timestep respectively, \( g \) refers to acceleration due to gravity and \( H \) refers to the average water depth below the mean water level throughout the model domain. To maintain accuracy and remain within the maximum prescribed Courant number, the timestep must reduce in tandem with the grid sizing, as in Equation (3.48).

\[ \Delta t \leq 8 \frac{\Delta x}{\sqrt{gH}} \] (3.33)

Convergence of the solution of the difference equations approximating partial differential equations with the exact solution of the partial differential equations is the ultimate goal in numerical modelling. The difference between the two solutions should tend towards zero with ever reducing grid size and timestep (Smith 1985), although there is an associated computation cost to such measures. For example, when considering the numerical approximation \( \bar{U} \) of the velocity \( U \) in the x direction, convergence is considered as:

\[ \lim_{\Delta x \to 0, \Delta t \to 0} (\bar{U} - U) \to 0 \] (3.34)
3.7. Water quality modelling

The equations used in the QUAL2E water quality model (Brown and Barnwell 1985) form the basis for all water quality formulae. Up to ten water quality parameters can be simulated using DIVAST: biochemical oxygen demand, dissolved oxygen, organic, ammonia and nitrate nitrogen, organic phosphorus, orthophosphate, Chlorophyll_a, temperature, salinity. The interaction between the various water quality parameters in DIVAST is presented in Figure 3.7.

Each of the TSAS compliance water quality parameters are simulated within DIVAST as DIN is the sum of ammonia and nitrate and MRP is represented by orthophosphate P. The dissolved oxygen cycle is also included. Chlorophyll_a modelling is carried out to give a proxy measure of overall phytoplankton growth with no explicit consideration of individual species of phytoplankton. The phosphorus and nitrogen nutrient cycle, phytoplankton cycle and dissolved oxygen cycle are fully interactive as can be seen in Equations (3.35) to (3.46) where a number of terms are common. Decay of phytoplankton generates phosphorus and nitrogen by-products.

Figure 3.7 Schematic of the water quality interactions as represented within DIVAST, taken from Nash et al. (2011) [Image removed for copyright reasons]
3.7.1. Phytoplankton kinetics

The Chlorophyll_a growth rate is defined in Equation (3.35), which takes into account growth, decay and the settling velocity of phytoplankton. The growth rate $G_{PI}$ in Equation (3.36) is a lumped term which is a product of individual adjustment or limitation factors which describe the constraint of Chlorophyll_a growth by temperature (Equation 3.37a), available nutrients (Equation 3.37b) and light (Equation 3.37c). Equation (3.37c) describes the light limiting factor $G_L$ which accounts for photoperiod $f$, the light attenuation coefficient $K_A$ and the relative intensity of surface lighting. A site specific light attenuation coefficient for Cork Harbour derived by Hartnett and Nash (2004) is presented in Equation (3.38), which was derived using secchi disk transparency data. It can be seen in equations (3.37c) and (3.38) that intuitively the growth of Chlorophyll_a at depth is hampered by the attenuation of light by Chlorophyll_a the nearer the surface. Thus the light limiting growth factor $G_L$ is inversely related to the concentration of Chlorophyll_a $C_P$.

\[
\frac{\partial C_P}{\partial t} = (G_{PI} - D_{PI} - V_{S4})C_P \tag{3.35}
\]

\[
G_{PI} = (G_{PI})_{max}G_TG_NG_L \tag{3.36}
\]

\[
G_T = (\theta_T)T-20 \tag{3.37a}
\]

\[
G_N = \frac{C_N}{k_N + C_N} \tag{3.37b}
\]

\[
G_L = \left[\log\left(\frac{I_H + I_O}{I_H + I_O - K_AH}\right)\right] \frac{f}{K_AH} \tag{3.37c}
\]

\[
K_A = 0.79012 + 0.004296(C_P) \tag{3.38}
\]

where
- $C_P$ = phytoplankton population (mg m$^{-3}$)
- $G_{PI}$ = growth rate constant (d$^{-1}$)
- $D_{PI}$ = death plus respiration rate constant (d$^{-1}$)
- $V_{s4}$ = settling velocity (m s$^{-1}$)
- $D$ = total water depth (m)
- $(G_{PI})_{max}$ = maximum growth rate under optimum lighting and nutrient conditions (d$^{-1}$)
- $T$ = water temperature
- $\theta_T$ = temperature adjustment coefficient
- $G_T$ = temperature adjustment factor
Numerical Modelling

\[ C_N = \text{nutrient concentration (mg l}^{-1}\text{)} \]
\[ k_N = \text{nutrient half saturation concentration (mg l}^{-1}\text{)} \]
\[ G_N = \text{nutrient limitation factor} \]
\[ G_L = \text{growth limiting factor} \]
\[ I_H = \text{light level which causes growth rate of half of } (G_{PI})_{\text{max}} \text{ (ly d}^{-1}\text{)} \]
\[ I_O = \text{surface light intensity (ly d}^{-1}\text{)} \]
\[ f = \text{fraction of day exposed to sunlight} \]
\[ K_A = \text{light attenuation coefficient (m}^{-1}\text{)} \]

3.7.2. Phosphorus cycle

The simplified phosphorus cycle model that within DIVAST considers phytoplankton P, organic P and orthophosphate P. Equation (3.39) presents the rate of phytoplankton P accumulation, which is directly proportional to the phytoplankton specific growth and decay rates and the phytoplankton settling velocity. Phytoplankton P is lost to the Organic P pool. Organic P is remineralised to orthophosphate, as can be seen by the middle term in Equation (3.40) which is common to Equations (3.40) and (3.41a).

**Phytoplankton P**

\[
\frac{\partial}{\partial t}(C_{P}A_{PC}) = \left( G_{PI} - D_{PI} - \frac{V_{S4}}{D} \right) C_{P}A_{PC} \tag{3.39}
\]

**Organic P**

\[
\frac{\partial}{\partial t}(C_{OP}) = D_{PI}C_{P}A_{PC} - k_{83}G_{P}^{T-20}X_{PRC}C_{OP} - \frac{V_{S3}(1-f_{DB})}{D}C_{OP} \tag{3.40}
\]

**Orthophosphate P**

\[
\frac{\partial}{\partial t}(C_{IP}) = k_{83}G_{P}^{T-20}X_{PRC}C_{OP} - G_{P}C_{P}A_{PC} - \frac{V_{S3}(1-f_{DB})}{D}C_{IP} \tag{3.41a}
\]

\[
X_{PRC} = \frac{C_{P}}{(K_{MPC}+C_{P})} \tag{3.41b}
\]

where

- \( X_{PRC} = \frac{C_{P}}{(K_{MPC}+C_{P})} \)
- \( G_{PI} = \text{phytoplankton growth rate (d}^{-1}\text{)} \)
- \( D_{PI} = \text{phytoplankton loss rate (d}^{-1}\text{)} \)
- \( C_{OP} = \text{Organic phosphorus concentration (mg P l}^{-1}\text{)} \)
- \( C_{IP} = \text{Inorganic phosphorus concentration (mg P l}^{-1}\text{)} \)
- \( A_{PC} = \text{phosphorus to Chlorophyll}_{a} \text{ ratio (mgP mgChla}^{-1}\text{)} \)
- \( k_{83} = \text{dissolved organic P mineralisation rate (d}^{-1}\text{)} \)
3.7.3. Nitrogen cycle

5 forms of Nitrogen are modelled within DIVAST. The simplified nitrogen cycle describing the interaction between the 5 forms is summarised by Equations (3.42) to (3.45). Decay of phytoplankton returns a fraction of its N to the organic N pool. Ammonification of the organic N converts organic N to Ammoniacal N. Ammoniacal N subsequently undergoes transformation to Nitrate N via nitrification in the presence of oxygen. Denitrification of Nitrate N to free Nitrogen gas (N$_2$) is an additional sink for N.

**Phytoplankton N**

\[
\frac{\partial}{\partial t}(C_{pA_{nc}}) = \left( G_{pI} - D_{pI} - \frac{V_{S4}}{D} \right) C_{pA_{nc}} \tag{3.42}
\]

**Organic N**

\[
\frac{\partial}{\partial t}(C_{ON}) = D_{pI} C_{pA_{NC}} f_{ON} - k_{71} \theta_{71}^{T-20} x_{PRC} C_{ON} - \frac{V_{S3}(1 - f_{D7})}{D} C_{ON} \tag{3.43}
\]

**Ammoniacal N**

\[
\frac{\partial}{\partial t}(C_{NH3}) = k_{71} \theta_{71}^{T-20} x_{PRC} C_{ON} - G_{pI} P_{NH3} C_{pA_{NC}} - k_{12} \theta_{12}^{T-20} \left( \frac{C_{DO}}{k_{NO3} + C_{DO}} \right) C_{NH3} + D_{pI} C_{pA_{NC}} (1 - f_{ON}) \tag{3.44}
\]

**Nitrate N**

\[
\frac{\partial}{\partial t}(C_{NO3}) = k_{12} \theta_{12}^{T-20} \left( \frac{C_{DO}}{k_{NO3} + C_{DO}} \right) C_{NH3} - G_{pI}(1 - P_{NH3}) C_{pA_{NC}} - k_{12} \theta_{12}^{T-20} \left( \frac{k_{NO3}}{k_{NO3} + C_{DO}} \right) C_{NH3} \tag{3.45}
\]

where

$A_{nc}$ = nitrogen to Chlorophyll$_a$ ratio (mgN mgChla$^{-1}$)

$C_{DO}$ = dissolved oxygen concentration (mg l$^{-1}$)

θ$_{83}$ = temperature coefficient

$K_{MPc}$ = half saturation constant for phytoplankton limitation of phosphorus recycle (mgP l$^{-1}$)

f$_{D3}$ = fraction of dissolved inorganic P in water column

f$_{D8}$ = fraction of dissolved organic P

$V_{S3}$ = organic matter settling velocity (m s$^{-1}$)

$V_{S4}$ = inorganic sediment settling velocity (m s$^{-1}$)
k_{12} = nitrification rate
CNH3 = Ammoniacal nitrogen concentration (mg l\(^{-1}\))

3.7.4. Dissolved oxygen modelling

Dissolved oxygen modelling must take account for nitrification, the presence of a carbonaceous biological oxygen demand, the consumption and production of dissolved oxygen by phytoplankton and the in situ oxygen concentration. Equation (3.46) accounts for these processes and is utilised in DIVAST for DO modelling.

\[
\frac{\partial}{\partial t} (C_{DO}) = k_a \theta_a T - 20 \left( C_s - C_{DO} \right) - k_d \theta_d T - 20 \left( \frac{C_{DO}}{k_{BOD} + C_{DO}} \right) C_{BOD} - \frac{64}{14} k_{12} \theta_{12} T - 20 \left( \frac{C_{DO}}{k_{NIT} + C_{DO}} \right) C_{NH3} - \frac{SOD}{D} - 64 \frac{14}{14} k_{12} \theta_{12} T - 20 \left( C_{DO} - C_{NH3} \right) + G_{PI} C_P \left( A_{oc} + 48 \frac{14}{14} A_{nc} (1 - P_{NH3}) \right) - A_{oc} k_{ir} \theta_{ir} T - 20 C_P
\]

where
C_s = dissolved oxygen saturation concentration (mg l\(^{-1}\))
C_{BOD} = BOD concentration (mg l\(^{-1}\))
k_a = reaeration rate
k_d = deoxygenation rate
k_{BOD} = half saturation constant for BOD
k_{NIT} = half saturation constant for Nitrogen
SOD = sediment oxygen demand (mg l\(^{-1}\))
D = depth (m)
G_{PI} = phytoplankton growth rate (d\(^{-1}\))
A_{oc} = oxygen to Chlorophyll_a ratio (mgO_2 mgChla\(^{-1}\))
P_{NH3} = Ammoniacal nitrogen preference factor
k_{ir} = rate of dissolved oxygen uptake through phytoplankton respiration and excretion

3.8. Model application and parameters

The DIVAST model of Cork Harbour has been thoroughly calibrated and validated (Hartnett et al. 2011a, Nash et al. 2011). DIVAST has been applied extensively in projects in fields such as aquaculture, water quality and marine renewable energy, and has become an industry standard level hydrodynamic, solute transport and water quality model (Hartnett and Nash 2004).
The detailed model developed by Nash et al. (2011) incorporated monthly averaged input data for a number of parameters, including domestic and industrial effluent discharges at 19 outfalls, as presented in Table 3.1; mean river inflows for 8 rivers including the rivers Lee, Owenacurra, Owenbuidhe and Glashaboy, as presented in Tables 3.2 and 3.3; nutrient influxes at the seaward boundaries; temperature, light intensity and photoperiod for Chlorophyll_a calculations. Monthly averaged nutrient loadings and riverine discharge volumes were taken from Costello et al. (2001).

Table 3.1 Pre-Carrigrennan numerical model industrial and WWTP volumetric flow rates and their licence discharge standards

<table>
<thead>
<tr>
<th>Discharge</th>
<th>Flow (m³/s)</th>
<th>BOD (mg/l)</th>
<th>Org N (mg/l)</th>
<th>NH₃ (mg/l)</th>
<th>NO₃ (mg/l)</th>
<th>Org P (mg/l)</th>
<th>MRP (mg/l)</th>
<th>DO (mg/l)</th>
<th>CHL-a (mg/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Industry + Sewers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cork city north sewer</td>
<td>0.347</td>
<td>370</td>
<td>0</td>
<td>124</td>
<td>15</td>
<td>0</td>
<td>24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cork city south sewer</td>
<td>0.347</td>
<td>370</td>
<td>0</td>
<td>124</td>
<td>15</td>
<td>0</td>
<td>24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Glounthane a</td>
<td>0.011</td>
<td>120</td>
<td>12</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rams head</td>
<td>0.066</td>
<td>257</td>
<td>0</td>
<td>45</td>
<td>58</td>
<td>21</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lough Beg</td>
<td>0.012</td>
<td>108</td>
<td>44</td>
<td>43</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Passage West</td>
<td>0.035</td>
<td>82</td>
<td>0</td>
<td>133</td>
<td>3</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Middleton</td>
<td>0.003</td>
<td>15</td>
<td>0</td>
<td>3</td>
<td>62</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lough Mahon West a</td>
<td>0.015</td>
<td>2850</td>
<td>84</td>
<td>90</td>
<td>4</td>
<td>34</td>
<td>46</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lough Mahon East a</td>
<td>0.002</td>
<td>269</td>
<td>8</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IFI</td>
<td>0.034</td>
<td>25</td>
<td>33</td>
<td>66</td>
<td>10</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Wastewater Treatment Works</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ballincollig</td>
<td>0.041</td>
<td>62</td>
<td>83</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Carrigtwohill</td>
<td>0.005</td>
<td>21</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cloyne</td>
<td>0.001</td>
<td>21</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Crosshaven</td>
<td>0.004</td>
<td>350</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Glanmire a</td>
<td>0.021</td>
<td>350</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Middleton</td>
<td>0.013</td>
<td>350</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Passage/ Monkstown</td>
<td>0.01</td>
<td>350</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tramore a</td>
<td>0.077</td>
<td>350</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cobh</td>
<td>0.03</td>
<td>320</td>
<td>18</td>
<td>29</td>
<td>0</td>
<td>5</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Carrigaline/ Ringaskiddy</td>
<td>0.044</td>
<td>273</td>
<td>15</td>
<td>26</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Carrigrennan</strong></td>
<td>0.568</td>
<td>24</td>
<td>12</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

*Combined into Carrigrennan discharge following construction of treatment plant
Table 3.2 Pre-Carrigrennan numerical model freshwater discharge flow rates and salinity and Chlorophyll_α concentrations for the Lee, Glashaboy, Owenbuidhe and Owenacurra rivers and the three catchments indicated in Figure 3.8

<table>
<thead>
<tr>
<th>River</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sept</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Flow (m³/s)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glashaboy</td>
<td>2.580</td>
<td>2.457</td>
<td>1.761</td>
<td>1.515</td>
<td>0.717</td>
<td>0.307</td>
<td>0.266</td>
<td>0.410</td>
<td>0.369</td>
<td>0.369</td>
<td>1.474</td>
<td>2.007</td>
</tr>
<tr>
<td>Owenbuidhe</td>
<td>8.065</td>
<td>8.553</td>
<td>4.957</td>
<td>4.728</td>
<td>2.298</td>
<td>1.253</td>
<td>0.941</td>
<td>1.004</td>
<td>1.620</td>
<td>1.976</td>
<td>4.247</td>
<td>6.941</td>
</tr>
<tr>
<td>Owenacurra</td>
<td>6.663</td>
<td>6.275</td>
<td>4.589</td>
<td>3.811</td>
<td>1.909</td>
<td>0.947</td>
<td>0.751</td>
<td>0.962</td>
<td>0.816</td>
<td>1.616</td>
<td>3.965</td>
<td>4.966</td>
</tr>
<tr>
<td>Catchment 1</td>
<td>1.978</td>
<td>1.884</td>
<td>1.350</td>
<td>1.162</td>
<td>0.549</td>
<td>0.235</td>
<td>0.204</td>
<td>0.314</td>
<td>0.283</td>
<td>1.099</td>
<td>1.130</td>
<td>1.539</td>
</tr>
<tr>
<td>Catchment 3</td>
<td>1.615</td>
<td>1.538</td>
<td>1.103</td>
<td>0.949</td>
<td>0.449</td>
<td>0.192</td>
<td>0.167</td>
<td>0.256</td>
<td>0.231</td>
<td>0.231</td>
<td>0.923</td>
<td>1.256</td>
</tr>
<tr>
<td><strong>SAL (ppt)</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>CHL-a (mg/m³)</strong></td>
<td>0.00689</td>
<td>0.00689</td>
<td>0.00689</td>
<td>0.00689</td>
<td>0.00689</td>
<td>0.00689</td>
<td>0.00689</td>
<td>0.00689</td>
<td>0.00689</td>
<td>0.00689</td>
<td>0.00689</td>
<td>0.00689</td>
</tr>
</tbody>
</table>

87
Numerical Modelling

Table 3.3 Pre-Carrigrennan numerical model inflow characteristics for rivers and catchments: nutrients, DO and BOD.

<table>
<thead>
<tr>
<th>Inflow characteristics</th>
<th>BOD (mg/l)</th>
<th>Org N (mg/l)</th>
<th>NH$_3$ (mg/l)</th>
<th>NO$_3$ (mg/l)</th>
<th>DO (mg/l)</th>
<th>Org P (mg/l)</th>
<th>MRP (mg/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.213</td>
<td>1.425</td>
<td>0.088</td>
<td>3.106</td>
<td>8.4</td>
<td>0.027</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Figure 3.8 Locations of wastewater discharges, rivers, catchments and monitoring points around the greater Cork harbour area

The model that was used throughout this thesis was extensively calibrated and validated against data that was made available from a) aerial remoting sensing of Chlorophyll-a which provided geo-referenced concentrations b) comprehensive datasets that were collected for the purposes of the 1977 Cork Harbour Pollution Report (O’Sullivan 1977), containing observations on tidal elevation, current velocity measurements, results of dye release surveys and a comprehensive salinity measurement study. Figure 3.9 presents the locations of the tide gauges and current meters for which tidal elevation and velocity were provided in the Cork Harbour Pollution Report.
The hydrodynamic module was forced with a spring tidal range of 4.2m, neap tidal range of 2.1m, a tidal period of 12.5 hours and a spring neap tidal period of 337.5 hours. Figures 3.10 and 3.11 present a comparison of model and field data of spring tide water surface elevations at Passage West and Currabinny. There was good agreement between two datasets confirming that the hydrodynamic module of DIVAST was well calibrated.
Two dye release surveys were carried out as part of the Cork Harbour Pollution Report to determine the circulation patterns within the Cork harbour area. This data was used as a basis for calibration of the solute transport module in DIVAST. Point discharges of dye were specified within the model with the same concentrations and locations. Calibration of the solute transport model involved tuning of the longitudinal dispersion and lateral turbulent diffusion coefficients were adjusted to reduce the spatial errors between simulated and recorded dye plumes. Figures 3.12 and 3.13 show the plots of the model plume and the recorded plume at the same instant, at low water two days after the dye discharge began. Visually, it is clear that there is close agreement between the two plumes in terms of shape and contour values. Table 3.4 contains the final dispersion and diffusion coefficients and the typical values from literature.
Table 3.4 Hydrodynamic constants and coefficients from Falconer et al., 2001 used for calibration (Nash et al. 2011).

<table>
<thead>
<tr>
<th>Calibration Variables</th>
<th>Model Value</th>
<th>Recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hydrodynamic Variables:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- roughness length</td>
<td>30 mm</td>
<td>80 mm</td>
</tr>
<tr>
<td>- kinematic viscosity</td>
<td>$1.41 \times 10^{-6}$ m² s⁻¹</td>
<td>$1.31 \times 10^{\circ}$ (sea water at 20ºC)</td>
</tr>
<tr>
<td>- momentum correction coefficient</td>
<td>1.016</td>
<td>1.016 for seventh law profile</td>
</tr>
<tr>
<td>- coefficient of eddy viscosity</td>
<td>1.00</td>
<td>0.23 – 1.00</td>
</tr>
<tr>
<td><strong>Solute Transport Variables:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- longitudinal dispersion coefficient</td>
<td>13.00</td>
<td>5.93 – 13.00</td>
</tr>
<tr>
<td>- lateral turbulent diffusion coefficient</td>
<td>1.00</td>
<td>0.23 – 1.00</td>
</tr>
</tbody>
</table>

The salinity measurements taken as part of the Cork Harbour Pollution Report were used to validate solute transport prior to initiation of water quality calibration. Two model runs were used, one in July and another in November. Salinity at the tidal boundary to the south and south east of the model domain was specified at 35 ppt, while the salinity values at the River Lee boundary were 15 and 8 ppt for July and November respectively. There was good agreement between model and field data, with the timeseries of salinity values falling between the maximum and minimum values observed at each location in Cork harbour.

During validation of the solute transport module, initial grids were used for salinity in order to assess the potential for the model to reach steady state quicker; every computational cell was given a starting value of 35ppt. From Figure 3.14, it can be seen that the net effect was a reduction of approximately 200 hours in the time taken to reach steady state. The positive results in utilising the initial grids for salinity validation lead to the adoption of initial grids for each of the 9 water quality parameters within the water quality module of the model. Linear regression was used to derive a relationship between salinity and each of the 9 parameters. These relationships were used to derive initial grids for each of the 9 parameters based on the prevailing salinity concentrations.
Figure 3.12: Dye plume predicted by the calibrated model (concentrations in µg/l) (Nash et al. 2011) [Image removed for copyright reasons]

Figure 3.13: Dye plume recorded in the field (Nash et al. 2011) [Image removed for copyright reasons]
Table 3.5 presents the list of final values for the water quality module parameters required to run the water quality module of the model as well as the range of values reported in 4 literature sources. The parameters variously describe rates of algal growth, decay, and mortality, uptake of nutrients, production and mineralisation amongst others. The primary purpose of the model development was to accurately model phytoplankton in the Cork harbour area. Apart from nutrient availability, growth of phytoplankton is a function of temperature and light intensity coupled with photoperiod. The ranges of values for these three parameters as used by Nash et al. (2011) are presented in Figures 3.15 and 3.16.

Previous experience in an Irish context (Brennan et al. 1998, McMahon et al. 1992) indicated that phytoplankton growth was limited more by light than nutrient availability, leading to the adoption of a site specific expression for light attenuation as a function of Secchi disk depth for Wexford harbour (Hartnett and Nash 2004); the same approach was taken by Nash et al. (2011) for Cork harbour. The benefit of the site specific function describing light attenuation resulted in a significant improvement in the accuracy of Chlorophyll a model results, as seen in Figure 3.17.
Figure 3.15: Monthly-averaged waterbody temperature (Nash et al. 2011) [Image removed for copyright reasons]

Figure 3.16: Monthly-averaged light intensity and photoperiod (Nash et al. 2011) [Image removed for copyright reasons]
Figure 3.17: Model predictions of Chlorophyll_a in Cork Harbour using the model default and site specific light attenuation functions. Results shown for Point D1 (Nash et al. 2011) [Image removed for copyright reasons]

The detailed model with site specific light attenuation formulation and monthly riverine and point loadings was run for the duration of September, coinciding with the time of aerial remote sensing; all remote sensing data was collected within a half hour window around spring high water. Figures 3.19 and 3.20 contain the modelled and remote sensed Chlorophyll_a concentrations throughout Cork harbour. The linear regression plot of collocated points in Figure 3.21 indicates a high level of accuracy, particularly in the context of water quality modelling, with an $R^2$ value of 0.78.

Upon visual inspection, the spatial patterns were largely consistent between the two plots. The most pertinent anomaly lies in the area of Lough Mahon, where high concentrations of Chlorophyll_a were not replicated in the remote sensing dataset. This was explained by the inability of the remote sensing to penetrate deeper than 2m below surface level while the model provided a depth averaged concentration of Chlorophyll_a.
Table 3.5 Parameters used in nutrient model formulations and associated values (Nash et al. 2011) [Figure removed for copyright reasons] [Sources: 1-Brown and Barnwell (1985), 2-Bowie et al. (1985), 3-Chapra (1997), 4-Falconer et al. (2001)]
Figure 3.18 Flow chart of DIVAST FORTRAN subroutines
Figure 3.19: Predicted Chlorophyll\textsubscript{a} concentrations for spring high water, September and validation locations (1-25) (Nash et al. 2011) [Image removed for copyright reasons]

Figure 3.20: Remotely sensed Chlorophyll\textsubscript{a} for spring high water, September (Nash et al. 2011) [Image removed for copyright reasons]
Figure 3.21: Linear regression analysis of the modelled and measured Chlorophyll_a data presented in Figures 3.20 and 3.22 respectively at the validation locations shown in Figure 3.20 (Nash et al. 2011) [Image removed for copyright reasons]
Chapter 4. Methodology

4.1. Introduction

Surface water quality monitoring may be carried out with the intention of establishing conditions for assessment of water quality status, although there are also many other reasons for monitoring including detecting temporal trends or spatial patterns in water quality, evaluating the success of management decisions or investigating the cause of non-compliance. A calibrated and validated water quality model may prove invaluable in the same exercises, although monitoring and modelling are rarely carried out with consideration of synergetic interactions between them. Due to the variable nature of estuarine inflows from both point and nonpoint loadings as a result of nutrient and land management and wastewater treatment plant upgrades or changes in wastewater licences, the priority of monitoring points may change over a timespan of months or years leading to inaccuracy of a validated and calibrated water quality model.

In Irish coastal and transitional water bodies, sampling is currently carried out uniformly at each monitoring point for the same water quality parameters although the optimum monitoring location may vary between parameters. Moreover, monitoring networks may vary in spatial and temporal coverage over time which restricts the confidence in overall assessment. Due to changes in freshwater and wastewater discharges to the dynamic estuarine environment, the focus of monitoring may change even though cessation of monitoring at redundant monitoring points may deem data from those monitoring points to be worthless. Optimisation of an existing monitoring network may entail contraction based on budgetary restraint or expansion based on a broadened remit for a monitoring programme.
This chapter outlines the methodology to optimising an existing monitoring network using entropy theory, by prioritising the monitoring network for each parameter. The methodology for spatial optimisation of the existing monitoring network can be extended to the scenario where a monitoring network is designed without consideration of the existing monitoring network by applying the methodology to all of the grid cells in a water quality model.

A modified approach is proposed to take into account the transinformation entropy between repeated measurements throughout the monitoring network. Thus, temporal optimisation of monitoring networks is achieved through the identification of the optimum frequency of sampling for which the transinformation entropy between repeated measurements is minimised.

A methodology is outlined which details how the loss of information due to removing monitoring points may be overcome by applying BME to estimate monitoring data at redundant monitoring locations using monitoring data at the remaining monitoring networks and simulation data from a calibrated and validated high resolution water quality model.

Finally, an overarching methodology is proposed which would coordinate the two aforementioned approaches in order to facilitate ongoing improvement of the accuracy of water quality monitoring and modelling.

4.2. Entropy analysis

Of the various entropy measures described in Chapter 2, marginal entropy and transinformation are the most relevant in the context of monitoring network design; given a PDF at each monitoring location, the marginal entropy provides an indication of the amount of information at each location, while transinformation measures the amount of redundancy between monitoring points. Transinformation may also be used to identify the sampling frequency that minimises the redundant information between successive samples.
4.2.1. Spatial Entropy Analysis

Application of entropy analysis to water quality monitoring data requires firstly the assembly of PDFs for each monitoring point. Figure 4.1 presents the methodology used in the spatial optimization of an existing monitoring network with \( n_m \) monitoring points. In order to complete the first step in the analysis (box 1), discrete PDFs are derived by subdividing the interval between the maximum and minimum data values observed in the entire monitoring network into a fixed number of bins. The PDF at each of the \( n_m \) monitoring points is determined by dividing the number of data points at that monitoring location falling within each bin by the total number of observations at that monitoring location. Hence the area beneath each discrete PDF sums to unity. Equation (2.61) in Chapter 2 is used to determine the marginal entropy of the discrete PDF at each monitoring point. To provide a stable basis for comparison, the same amount of data points should be available for each monitoring location, at approximately the same times. Otherwise, monitoring locations where insufficient data was collected would immediately be appointed a lower priority than would otherwise be appointed with an equal amount of data points described by each monitoring point PDF.

The monitoring locations are ranked from high entropy to low, with the highest entropy station nominated as the priority monitoring station, as presented in box 2 in Figure 4.1. The highest entropy monitoring point is added to the optimum monitoring network; this action is denoted by the hatched line in the top right corner of Figure 4.1, to signify this happens only once. The station with the highest marginal entropy contains the largest range in data observations and the broadest range of probabilities, and thus the greatest amount of information with respect to the water quality parameter of interest. All of the remaining points in the list of candidate monitoring stations are then considered for addition to the optimised network, as denoted by the arrow from box 1 to 3.

With the highest priority monitoring point identified, the transinformation entropy between the point with the largest marginal entropy and each of the remaining \((n_m-1)\) candidate monitoring points is then determined on a one by one basis to determine the second priority monitoring point. This aspect of the analysis is contained in box 4.
in Figure 4.1. The candidate monitoring point with the lowest transinformation entropy is the 2nd point added to the optimised water quality monitoring network. If a monitoring point has high transinformation with the point(s) already in the optimised network, it is returned to the list of candidate monitoring points. Low transinformation entropy between two monitoring points indicate that there is limited redundancy of data between the two monitoring points; two neighbouring monitoring points would be likely to have a high transinformation, as would two monitoring points similarly influenced by the same pollution source.

When the n-th station is considered, the transinformation entropy between the chosen network of n-1 stations and the remaining candidate stations is calculated iteratively and the candidate point with the lowest transinformation is added to the expanding optimised network. The procedure can be applied to pre-existing monitoring networks or any number of points. When an existing monitoring network is being optimised, points are added to the “new” network and removed from the list of candidate monitoring points one by one.
Where the same method is applied to high spatial resolution numerical model data, a PDF is derived at each grid cell in the model domain from snapshot data throughout the domain generated at regular intervals throughout the model simulation period. Marginal entropy is determined for each grid cell. If a single cell has the highest marginal entropy, it is added to the optimised monitoring point with the highest rank. Otherwise, where a collection of monitoring points have the highest marginal entropy, they are all removed from the list of candidate monitoring points, all considered as the highest priority monitoring location and one of the locations is used as a proxy to represent all of the other highest priority monitoring locations in further analysis. The transinformation between the arbitrarily chosen highest priority monitoring location and all other candidate monitoring points is determined. If a single point has the lowest transinformation, it is added to the priority network. Otherwise, if a number of points have the lowest transinformation, they are all removed from the list of candidate points and a single point is selected from those.
Methodology

points to be added to the smaller representative network. The procedure continues iteratively until all points have been used.

4.2.2. Spatial – Temporal Entropy Analysis

The entropy analysis procedure utilised to establish the spatial priority of monitoring points can be extended to identify the sampling frequency which minimises the transinformation between successive sampling events. The procedure is applied after spatial optimisation has been completed. Figure 4.2 presents an expanded algorithm based on Figure 4.1, with the addition of box 5. Starting with the smallest optimum monitoring network comprising 2 highest priority monitoring points, the transinformation between timeseries data sampled at the highest priority monitoring point at the base sampling frequency and the second highest priority monitoring point at multiples of the base sampling frequency is determined. The multiple of the base sampling frequency which delivers the lowest transinformation is the optimum monitoring frequency, which is \( f_1 \) in Figure 4.2 in the case of a monitoring network of 2 points. For each combination of stations, transinformation reaches a local minimum, which is considered the optimum monitoring frequency for that combination of stations. The process can be repeated to determine the optimum sampling frequency with each additional optimum monitoring point, while any base frequency of sampling can be used. Hence, the effect of adding each additional prioritised monitoring point on the optimum sampling frequency can be quantified. Transinformation decreases as the sampling lag increases and autocorrelation decreases.
Figure 4.2 Algorithm describing the procedure to follow to determine the optimum sampling frequency using temporal entropy analysis, based on the optimum monitoring network derived from Figure 4.1.
Methodology

The process of establishing the optimum sampling frequency for each monitoring network size and water quality parameter involves determining the transinformation for different base sampling frequencies and multiples thereof. For each combination of monitoring network size, water quality parameter and base sampling frequency, the transinformation entropy measure will reach a minimum value using a multiple of the base sampling frequency.

4.3. BME analysis

Chapter 2 presents the general theory behind BME estimation. The following section describes the methodology to be followed in general in the process of BME estimation of water quality data, whilst particular attention is drawn to the approach taken in the author’s work.

Figure 4.3 below presents the flowchart of Matlab functions which execute the BME estimation. Later, a rudimentary example is presented of how BME proceeds.

Figure 4.3 Flow chart of computational implementation of BME via MATLAB

108
Methodology

All hard and soft data is prepared for incorporation in the BME estimation procedure; this step is completed by Matlab functions EPA_EPA_DATA, DIVAST_EPA_DATA, EPAselect and DIVASTselect. BME processes interval and probabilistic soft data of many types (Christakos 2000).

The first step in BME analysis is to select monitoring data within the period of assessment. This is followed by the selection of simulated water quality data from the timeseries generated at each monitoring point such that the data coincide with the time of water quality sampling. The water quality monitoring data is then divided into hard data and validation data. In the context of the entropy analysis procedure, hard data comprises water quality monitoring data at the retained monitoring points, while the validation dataset consists of the monitoring data at the locations that have been deemed redundant or below a priority threshold.

Subsequently, soft data is derived for assimilation in the BME estimation procedure. Soft data may be determined in a number of ways. LoBuglio et al. (2007) applied a linear regression to log-transformed BATHTUB box model predictions and Chlorophyll_a sample data to derive the expected value of Chlorophyll_a data given the BATHTUB (Walker 2006) box model predictions and the standard error of estimation, which were both used to derive the shape of the PDF. Money et al. (2009) similarly derived soft log *E. coli* PDFs from collocated data on log transformed *E. coli* and turbidity readings.

Christakos et al. (2004) used a linear regression to describe the empirical correlation between total ozone and tropopause pressure to derive the experimental mean and variance of total ozone values collocated with tropopause pressure readings, such that a unique PDF was available to describe each 5 millibar interval of tropopause pressure.

A similar approach is taken in this work, such that collocated model observations and water quality data will be grouped and the model observations subdivided into intervals for which mean and variance values will be derived and PDFs derived on the basis of the means and variances. Each numerical model datum is described by a PDF summarising the interval of numerical model observations which it lies within. Due to the nonnegative nature of the water quality parameters under consideration,
each soft data point is described by a left truncated (below 0) normal PDF as described by Hald (1952) described in Equation (4.1).

\[
f_{LTN}(x) = \begin{cases} 
0, & -\infty \leq x \leq 0 \\
\int_{0}^{\infty} f(x) \, dx, & 0 \leq x \leq \infty 
\end{cases}
\] (4.1)

After all data is prepared and collated, the spatial and temporal mean trend is established for each monitoring point and day of sampling, respectively. There are two stages to the process. Firstly, the raw mean trend is determined at each location and day of monitoring by averaging the data at each monitoring point over time to determine the spatial trend, and average throughout the spatial domain on each day of sampling to derive the temporal mean trend. Smoothing of the spatial and temporal mean trends is achieved using an exponential smoothing filter. Two parameters are required to smooth the trend, a search radius and smoothing radius, both of which are found by trial and error, although their values will be meaningful as they should not exceed the maximum spatial distance between any two points or the maximum time lag between any two sampling events. By necessity, the smoothing radius is shorter than the search radius, as the smoothed temporal trend must capture all significant spatial and temporal features; if the smoothing radius exceeded the search radius, a flat trend would be determined which is not informative. The ultimate objective is to capture predominating spatial and temporal patterns whilst not capturing the entire data trend by using excessively short search or smoothing radii - this would result in a negligible residual upon removal of the mean trend from the underlying dataset, leading to an uninformative covariance model.

The temporal exponential smoothing filter is defined as (Christakos et al. 2002):

\[
(m_{ts})_j = \frac{\sum_{i=1}^{k} [(m_i)_i \times \exp \left( \frac{t_i}{t_{smooth}} \right)]}{\sum_{i=1}^{k} \exp \left( \frac{t_i}{t_{smooth}} \right)}
\] (4.2)
where \((m_{\text{ts}})_j\) is the smoothed temporal mean trend at time \(j\), \((m_t)_i\) is the raw temporal mean trend at the \(i^{th}\) point of the \(k\) points which lie within the search radius \(t_{\text{search}}\) of time \(j\), \(t_i\) is the temporal lag between times \(j\) and \(i\) and \(t_{\text{smooth}}\) is the exponential smoothing radius to even out the fluctuations in the raw mean trend values. The spatial smoothing filter is analogous. Spatiotemporal mean trend values are derived as the sum of the spatial \((m_s)\) and temporal \((m_t)\) trends minus the average of the temporal trend as presented in Equation (4.3) (Lee et al. 2012):

\[
m(s, t) = m_s(s) + m_t(t) - \overline{m_t}
\]  

where \(m(s, t)\) is the spatiotemporal interpolated mean trend at point \((s, t)\), \(s\) is the spatial coordinate \((x_1, y_1)\) and \(t\) is time.

Figure 4.3 presents a schematic of soft and hard data points in the space-time continuum surrounding a single estimation point, where time is represented by the vertical axis and the two orthogonal horizontal axes represent the \(x\) and \(y\) direction. The hard and soft data to be used for estimation are captured within a search region which is defined considering the range of spatial and temporal covariance observed from the fitted models. The spatial and temporal search radii for estimation are less than the ranges of covariance to take account of autocorrelation between the estimation point and the soft and hard data. Models may be nested such that two covariance structures describe the primary and secondary covariance structure as in Equation (4.4), where \(C_1 >> C_2\) and \(a_{r1}, a_{r2}, a_{t1}\) and \(a_{t2}\) refer to the primary and secondary spatial and temporal ranges fitted to the experimental covariance data.

\[
C = C_1 \exp \left( \frac{-3(s - s')}{a_{r1}} \right) \exp \left( -3 \left[ \frac{t - t'}{a_{t1}} \right]^2 \right) + C_2 \exp \left(-3 \left[ \frac{s - s'}{a_{r2}} \right]^2 \right) \exp \left(-3 \left[ \frac{t - t'}{a_{t2}} \right] \right)
\]  

(4.4)

Figure 4.5 gives a visual overview of the steps to BME estimation. Step (a) involves determining a meaningful spatiotemporal mean trend. Figure 4.4 gives an indication of how the temporal mean trend of a timeseries of sample data might appear. Note that the trend captures the gradual increasing trend towards point 4 with a subsequent drop in values. Upon removal of the mean trend from the timeseries, the residual
Methodology

data in Figure 4.5b) provides the basis for calculating the covariance as presented in Figure 4.5(c), using Equation (4.5).

\[
c_X(r, \tau) = \frac{1}{N(r_{\text{lag}}, \tau_{\text{lag}})} \sum_{i=1}^{N(r_{\text{lag}}, \tau_{\text{lag}})} X_{\text{head},i}X_{\text{tail},i} - m_X^2
\]  

(4.5)

where \(N(r_{\text{lag}}, \tau_{\text{lag}})\) is the number of points separated by the spatial lag \(r_{\text{lag}}\) and the temporal lag \(\tau_{\text{lag}}\), \(X_{\text{head},i}\) and \(X_{\text{tail},i}\) are each pair of points separated by the lag in question, and \(m_X\) is the mean trend at each point.

The expression \(m_X^2\) is zero in this case as the mean trend has already been removed and the values for \(X_{\text{head},i}\) and \(X_{\text{tail},i}\) are the residual data points.

Knowledge of the spatial and temporal coordinates of each data point means that the spatial and temporal distances or “lags” between any two points can be established. Thus, an average of the experimental covariance of the residual data between every pair of points separated by a particular spatial or temporal lag can be determined. The process is repeated to determine the experimental covariance for every spatial and temporal lag that was observed. An optimal covariance model is then fitted to the experimental covariance values as in Figure 4.5(d). The BME estimation procedure depicted in Figure 4.5(e) involves identifying the spatial and temporal lag between every pair of soft data points, hard data points and the estimation point. Referring to the covariance model of best fit derived in Figure 4.5(d), the covariance weights within and between each data type can be determined from the matrix of spatial and temporal lags. Section 4.3.1 below gives a worked example of the estimation procedure.
Figure 4.4 Example of BME estimation using adjacent hard data and soft probabilistic data
Methodology

Figure 4.5 Outline of procedure of BME estimation with the relevant MATLAB functions referenced
4.3.1. Solution of BME mean estimate and estimation error variance - Example

Consider a small basin that is 5km x 5km in plan area. Within the basin, there is a varying DO concentration. 3 samples were extracted and analysed at different locations and times, while a DO dip meter was used at 6 different locations and times. With the data available to us, we would like a probabilistic estimate of the likely range of values at a point k. Table 4.1 below presents the details of 3 hard data points h1-h3, 6 soft data points s1-s6 and the estimation data point k. Let X be the spatiotemporal random field comprising the hard and soft dissolved oxygen data and the unknown value at point k. As noted earlier in chapter 4, realisations of the random field are represented by χ. Let χmap = \[ χk, χ_{\text{data}} \] = \[ χk, [χh, χs] \]. Table 4.1 contains the x and y coordinates and the time of sampling in columns 2, 3 and 4 respectively. The 5th and 6th column contain the values for the mean and variance for each sampling event. Note that the hard data don’t have a variance value as they are considered absolutely certain.

Removing the trend and deriving the covariance model of best fit to the residual data results in a covariance model which is gaussian in the spatial domain and exponential in the temporal domain, with a sill of 0.1251, a spatial range of 3.16 km and temporal range of 2.75 days. The covariance matrix is determined following Equation (A2.88) of the appendices which requires knowledge of the covariance between the soft and hard data along with the covariance matrix of the hard data alone and the soft data alone. Calculation of mean m and variance v at point k requires Equations (4.6) and (4.7).

For brevity, the reader is referred to the 2nd appendix for clarification of the significance of each of the terms in expressions (4.6) and (4.7). Equation (4.6) gives the estimate of the mean of the unknown value at point k given the entire knowledge base which is described by subscript K, while Equation (4.7) gives the error variance of estimation at point k. The vectors Bk|hsh(h) and Bk|hsh(s) are the hard and soft portions respectively of the bayesian update vector which represent the covariance of the estimation point k given the hard and soft data. For example, Bk|hsh(h) = Ck,hsh(h)\( C_{hsh(h),hsh(h)}^{-1} \), i.e. the variance of the estimation point k given the knowledge of the hard portion of the hard and soft data vector is given be the covariance between the point
k and the coordinates of all of the hard and soft data points, divided by the variance of the hard data.

Table 4.1 Hard and soft data for BME and simple kriging estimation

<table>
<thead>
<tr>
<th></th>
<th>x (km)</th>
<th>y (km)</th>
<th>t (days)</th>
<th>mean (mg/l)</th>
<th>variance (mg/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>h₁</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4.8</td>
<td>n/a</td>
</tr>
<tr>
<td>h₂</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4.5</td>
<td>n/a</td>
</tr>
<tr>
<td>h₃</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>n/a</td>
</tr>
<tr>
<td>s₁</td>
<td>2</td>
<td>3</td>
<td>1.5</td>
<td>4</td>
<td>0.0625</td>
</tr>
<tr>
<td>s₂</td>
<td>2</td>
<td>2</td>
<td>1.5</td>
<td>4.2</td>
<td>0.25</td>
</tr>
<tr>
<td>s₃</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2.75</td>
<td>0.01</td>
</tr>
<tr>
<td>s₄</td>
<td>1</td>
<td>2</td>
<td>3.75</td>
<td>1.9</td>
<td>0.04</td>
</tr>
<tr>
<td>s₅</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>0.09</td>
</tr>
<tr>
<td>s₆</td>
<td>3</td>
<td>3</td>
<td>3.5</td>
<td>2.5</td>
<td>0.25</td>
</tr>
<tr>
<td>k</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>m</td>
<td>v</td>
</tr>
</tbody>
</table>

\[
\bar{x}_{k|K} = B_{k|h}(\chi_h) + B_{k|s}(\chi_s) \left( A^{-1} \int d\chi_s f_0(\chi_s) \chi_s \phi(\chi_s; m_{(s|h)}, C_{(s|h)}) \right)
\]  

(4.6)

where

\[
\chi_h = \begin{bmatrix} -0.7173 \\ -0.2811 \\ -0.2539 \end{bmatrix}
\]

\[
B_{k|h} = \begin{bmatrix} 0.1969 & -0.0438 & 0.4823 \\ -0.1091 & -0.0089 & 0.0206 & 0.0026 & -0.0147 \end{bmatrix}
\]

\[
\chi_s = \begin{bmatrix} -0.4612 \\ -0.2935 \\ 0.2009 \\ 0.4286 \\ 0.3412 \end{bmatrix}
\]

\[m_{(s|h)} = 0\] (mean trend removed to be added after estimation)

\[
C_{(s|h)} = \begin{bmatrix} 0.1511 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}
\]

\[
\begin{bmatrix} 0.1016 & 0.1207 \\ 0.0117 & 0.0102 & 0.1474 \\ 0.0063 & 0.0072 & 0.0196 & 0.1569 \end{bmatrix}
\]

\[
\begin{bmatrix} 0.0222 & 0.0082 & 0.0344 & 0.0154 & 0.1571 \\ 0.0119 & 0.0068 & 0.0652 & 0.0264 & 0.0498 & 0.1566 \end{bmatrix}
\]
\[ \phi(x_s; 0, C(s|h)) = \frac{1}{2\pi^{m+1}} \exp \left(-\frac{1}{2} (x_s)^T C(s|h)^{-1} (x_s) \right) \]

\[ A = \int d\chi f_\chi \phi(\chi_s; m(s|h), C(s|h)) \]

\[ \sigma_{k|h}^2 = C(k|h_s) + A^{-1} \int d\chi (B_k|h_s \chi - \bar{x}_{k|h})^2 f_\chi \phi(\chi_s; m(s|h), C(s|h)) \]

(4.7)

The interpolated mean trend (removed prior to covariance analysis as referred to above) is 3.603. \( x_{k|h} = -0.213 \); hence \( m = 3.38 \), whilst \( v = 0.1153 \). The first half of Equations (4.6) (\( B_{k|h}(\chi_h) \)) and (4.7) (\( C_{k|h_s} \)) summarise the simple kriging estimator if soft data is not available. The simple kriging estimate of \( m \) and \( v \) is (3.3697, 0.1193), thus BME returns a narrower estimation variance than simple kriging in this rudimentary application. It is difficult to gauge the accuracy of the estimate at point \( k \) with no value available to compare. In order to assess the accuracy of the BME estimation procedure, each of the hard and soft data points in Table 4.1 were estimated by cross validation whereby each datum was removed and estimated using the remaining points. Table 4.2 presents the results, where it can be seen that in approximately half of the cases, the BME estimate is closer than the kriging estimate to the original data point, while the error variance is consistently lower indicating a narrower posterior PDF, although the error variance of the soft data estimates generally exceeds the soft data variance.

**Table 4.2 Cross validation results in BME estimation example**

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>t</th>
<th>mean</th>
<th>variance</th>
<th>BME estimate</th>
<th>BME error variance</th>
<th>Kriging estimate</th>
<th>Kriging error variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>h1</td>
<td>1</td>
<td>3</td>
<td>1.48</td>
<td>-</td>
<td>4.235</td>
<td>0.14</td>
<td>4.119</td>
<td>0.16</td>
</tr>
<tr>
<td>h2</td>
<td>2</td>
<td>1</td>
<td>4.5</td>
<td>-</td>
<td>4.398</td>
<td>0.14</td>
<td>4.269</td>
<td>0.15</td>
</tr>
<tr>
<td>h3</td>
<td>1</td>
<td>4</td>
<td>2.2</td>
<td>-</td>
<td>3.914</td>
<td>0.32</td>
<td>3.881</td>
<td>0.34</td>
</tr>
<tr>
<td>s1</td>
<td>2</td>
<td>3</td>
<td>1.5</td>
<td>0.0625</td>
<td>4.191</td>
<td>0.11</td>
<td>3.866</td>
<td>0.19</td>
</tr>
<tr>
<td>s2</td>
<td>2</td>
<td>2</td>
<td>1.5</td>
<td>0.25</td>
<td>3.329</td>
<td>0.08</td>
<td>2.887</td>
<td>0.16</td>
</tr>
<tr>
<td>s3</td>
<td>3</td>
<td>2</td>
<td>2.75</td>
<td>0.01</td>
<td>3.031</td>
<td>0.16</td>
<td>2.981</td>
<td>0.18</td>
</tr>
<tr>
<td>s4</td>
<td>1</td>
<td>2</td>
<td>3.75</td>
<td>1.9</td>
<td>2.798</td>
<td>0.18</td>
<td>2.806</td>
<td>0.18</td>
</tr>
<tr>
<td>s5</td>
<td>2</td>
<td>4</td>
<td>3.0</td>
<td>0.09</td>
<td>2.953</td>
<td>0.18</td>
<td>2.945</td>
<td>0.20</td>
</tr>
<tr>
<td>s6</td>
<td>3</td>
<td>3</td>
<td>3.5</td>
<td>2.5</td>
<td>2.151</td>
<td>0.12</td>
<td>2.167</td>
<td>0.14</td>
</tr>
</tbody>
</table>
4.3.2. BME algorithm

Figure 4.6 outlines the procedure that was followed in the application of BME to water quality numerical model results and water quality data, the results of which are presented in Chapter 7. BME estimation reduces to simple kriging estimation as presented in Figure 4.8 as the limiting case if no simulated water quality data is considered for estimation. Figure 4.7 presents a flowchart of how soft data is derived from the available water quality data and simulated water quality data from the numerical model. As the numerical model was calibrated and validated using water quality monitoring data and wastewater and freshwater inflows to Cork Harbour prior to the construction of Carrigrennan WWTP (see Hartnett and Nash (2004)), the first step entails extracting the EPA water quality data which was sampled prior to 2004. Water quality data sampling was carried out by the EPA at high and low tide at each monitoring point during each monitoring excursion. Water quality data was divided into high and low tide datasets and water quality model data was outputted at an equivalent stage of the spring-neap tidal cycle in the same month of the year with the same prevailing environmental conditions, although climatic and hydrological numerical model inputs are averaged on a monthly basis.

The range of numerical model data results is subdivided into intervals within which the mean and standard deviation of the collocated water quality monitoring data is obtained. The mean and standard deviation of the monitoring data for each interval provides the basis for the PDF inputs to BME. Each numerical model datum when used within the BME interpolation procedure is described by the PDF which describes the interval within which the datum lies. As the accuracy of each PDF is dependent upon the accuracy of the numerical model, it is imperative that the numerical model is recalibrated at regular intervals in light of new water quality observations or changes to point and nonpoint sources.

The algorithm in Figure 4.9 presents a novel, iterative monitoring and modelling programme. Monitoring data is used to calibrate and validate the water quality model. Regular comparison of the marginal entropy scores of monitoring and model data at each monitoring point highlights where model inputs require updating, or the model requires recalibration. The updated model is then used to carry out a full entropy analysis which considers every wet grid cell in the model domain as a
potential location for water quality monitoring. If a previously unmonitored location is highlighted by the analysis, a new monitoring point is added at that location. If a previously monitored location is no longer highlighted, that monitoring point can be removed. BME or kriging can be carried out at that location to provide a probabilistic estimate of the likely range of values for the water quality parameter.
Methodology

Figure 4.6 BME algorithm as applied to EPA water quality data and DIVAST simulated water quality datasets via MATLAB
Methodology

Figure 4.7 Data processing algorithm of EPA_EPA_DATA, DIVAST_EPA_DATA, EPAselect and DIVASTselect for BME

1. EPA data
2. Numerical model data
   - Isolate data sampled: high/low tide ±3 hrs
   - Divide range of observed numerical model data into intervals
   - Divide into retained and discarded monitoring points
   - Within each interval, derive mean and standard deviation of all collocated EPA data

Figure 4.8 Simple kriging algorithm as applied to EPA water quality data via MATLAB

1. EPA data (EPA_EPA_DATA and EPAselect)
2. Spatiotemporal mean trend of EPA data (MeanTrendEstEPA_DATA)
3. De-trend EPA data (MeanTrendEstEPA_DATA)
4. Derive experimental covariance data from EPA data (CovarianceEstEPA_DATA)
5. Fit optimal spatiotemporal covariance model in least squares sense (CovModelFit)
6. Derive BME estimates at points of interest (BME_Est_CorkEPA_DATA),
Figure 4.9 Framework for interaction between BME, monitoring network optimisation by entropy analysis and numerical model calibration
4.4. Conclusions

Of the many considerations in the optimisation of water quality monitoring networks, one of the most pertinent questions is: what priority should be attributed to each monitoring location with respect to individual water quality parameters? This chapter has summarised a range of methodologies which can be considered in the optimisation of a water quality monitoring network through the combined use of water quality data and a calibrated and validated water quality model. A framework is presented for the iterative updating of numerical models and water quality monitoring networks in a reciprocal manner. Thus far to the author’s knowledge, no all-encompassing methodology exists which would allow for the design and updating of water quality monitoring networks on an ongoing basis. To counteract the discontinuity of monitoring data which may be caused by monitoring network redesign, an adaptation of the BME methodology which has been applied elsewhere is proposed for the estuarine environment to estimate monitoring data at redundant monitoring locations through combining numerical model data simulations and water quality data. As the accuracy of BME estimation is predicated on the numerical model inputs being up to date, the iterative procedure which is proposed would ensure that any deviation of the numerical model from the monitoring data would be highlighted to point to the requirement for a change of the numerical model inputs.
Chapter 5. Optimisation of water quality monitoring programmes via Entropy methods

5.1. Introduction

Entropy methods provide a measure of the information contained in real or simulated datasets. Entropy methods can be applied equally to measured data or high resolution data produced by numerical models which have been calibrated by measured data, to provide a more accurate insight into entropy from high resolution model output. Time series of data can be used to optimally calibrate and validate the numerical model from which entropy data was initially generated.

At present to the author’s knowledge, there is limited guidance available on the optimum location and frequency of sampling of water quality parameters in estuaries and harbours. The entropy method has been applied to hydrogeology and river catchment basins but not to the coastal environment. Here, Shannon’s entropy theory is applied to water quality and monitoring network optimisation where the Shannon information entropies of PDFs of water quality data are compared at different locations to elucidate the redundancies between monitoring points and prioritise monitoring.

The datasets which are utilised include water quality data provided by the Irish EPA for the test site of Cork Harbour, and simulated water quality data from the high quality finite difference water quality model DIVAST. Here, a two dimensional depth integrated finite difference model was utilised which simulates hydrodynamics, solute transport and the biogeochemical cycle. Entropy analysis was utilised in the prioritisation of locations for sampling of water quality parameters and the identification of the optimum frequency to sample each parameter. The numerical model was run for both winter and summer simulations. High resolution
time series data were outputted at each of the 24 monitoring points within the greater Cork Harbour area.

This chapter details the results derived in the application of entropy methodologies to water quality data collected by the EPA and simulation data generated using DIVAST. Entropy analysis results are discussed and pertinent observations are highlighted for future water quality sampling in estuarine waters.

5.2. Case Study Description

Located at the mouth of the river Lee, Cork Harbour opens onto the Celtic sea and its 190km of coastline make it one of the largest sea inlets in Ireland, see Figure 5.1. The harbour geometry is complex, as a result hydrodynamic circulation patterns in the harbour are also complex. Cork Harbour is comparatively deep and long with a large surface area and drains a large freshwater catchment consisting of the Rivers Lee, Owenacurra, Owenbuidhe and Glashaboy; the River Lee contributes the greatest freshwater discharge. The harbour is macro-tidal with a maximum tidal range of 4.2m and extensive mudflats exposed at low tides. Over 100 large scale industries discharge their effluents into the harbour, including a number of large pharmaceutical plants. Many smaller industries also discharge to the sewers. Nine wastewater treatment plants discharge domestic waste to the harbour and nutrients also enter through freshwater and marine sources. Water quality in Cork Harbour is affected by the large numbers of nutrient sources from rivers and domestic and industrial outfalls. Water quality in Cork Harbour has suffered as a consequence of population growth and heavy industrialisation.

In the trophic status assessment for the period 1995-1999, the Lee Estuary, Lough Mahon and the Owenacurra Estuary were downgraded to eutrophic status. In an evaluation of tidal water quality, monitoring data collected under the National Environmental Monitoring Programme prior to initiation of the WFD monitoring programme were used to evaluate trophic status amongst other parameters (Clabby et al. 2008). Using the trophic status assessment system or TSAS (Toner et al. 2005) as a metric for trophic status assessment, two bodies were downgraded from potentially
eutrophic to eutrophic and three were downgraded from unpolluted to intermediate trophic status.

Lough Mahon in Cork Harbour was the only previously eutrophic body to improve in status due to the Cork main drainage project which was completed in 2004 which included the construction of Carrigrennan WWTP at Little Island in Lough Mahon. The plant provides secondary treatment for sewage in Cork city which previously entered the Lee Estuary and Lough Mahon untreated. A number of improvements were found in other regards in Cork Harbour and environs, including an improvement in the level of oxygen depletion and ammonia concentrations.

The most recent coastal and transitional water quality report (McGarrigle et al. 2010) indicates that a marked improvement has been effected by improved wastewater treatment, lower chlorophyll levels and better oxygen conditions. The percentage of eutrophic coastal or transitional bodies decreased from 18.9 to 10.1%. However, non-compliances were observed in three areas for the DIN environmental quality standard (S.I. No. 272 of 2009) including Cork Harbour and the outer Cork Harbour. MRP non-compliances were observed in four bodies including Lough Mahon at Harper’s Island. Although oxygen conditions were largely healthy, a number of waters were notable for deoxygenation including the lower Lee Estuary.
Figure 5.1 Map showing the monitoring sites in the Cork Harbour area that lie within the numerical model domain
5.3. Comparison of model output and EPA data via marginal entropy

Prior to utilising entropy methods for the design and optimisation of the monitoring network at Cork Harbour, it was considered prudent to compare the marginal entropy of the PDFs describing EPA water quality data at each of the monitoring points with the water quality model simulation data at each monitoring point for the same scenario or period in time. The comparison was made to address the following questions:

- Does the EPA water quality data summarise the range of numerical model simulation data observed at each monitoring point from a representative 3 year model run?
- Alternately does the numerical model adequately capture the range of values captured by monitoring over the simulation period?
- How closely is the underlying PDF captured by EPA sampling?

As set out in Chapter 4, data from each monitoring point was aggregated to generate a PDF at each monitoring point. All PDFs were discretised into 5000 equal width intervals ranging from the highest observed value in the domain to the lowest. The marginal entropy for each of the monitoring locations was derived using Equation (2.53). The priority of stations was ranked from high entropy to low. The aforementioned steps form the initial stage of entropy analysis identified by box 1 in Figure 4.1.

Due to the varying spatial and temporal coverage of EPA monitoring data, transinformation between monitoring points could not be investigated. Some monitoring points were unmonitored for periods of time. Transinformation between PDFs would not be representative unless each PDF was derived from approximately concurrently sampled data. For the same reason, temporal analysis could not be carried out to determine serial dependence through transinformation entropy between data sets at increasing lags as sampling was not carried out on a consistent weekly or monthly basis.

Water quality data was obtained from the Irish EPA for the Cork Harbour area covering the river Lee, Lough Mahon, the North Channel and Cork Harbour. Water
quality data included Chlorophyll_a, MRP and both Ammonia and Total Oxidised Nitrogen which couple to yield DIN. Sampling was carried out at 37 monitoring points from August 1994 to September 2004.

The following observations were made from the analysis of EPA data marginal entropy at each monitoring point from the dataset collected up to 2004 the year that Carrigrennan wastewater treatment plant was commissioned (see Table 5.2):

- Chlorophyll_a lab data entropy suggests that monitoring should focus on the lower Lee Estuary, Lough Mahon and the inner Cork Harbour.
- DIN entropy values are highest in the Lee Estuary and the Owenacurra Estuary.
- MRP entropy values are highest in the upper and lower Lee Estuary reaches and the Glashaboy Estuary.
- BOD entropy values are highest at the Lee and Lough Mahon area.

The areas which were referenced in the aforementioned observations were assessed as eutrophic in at least one of the two EPA estuarine water quality reporting periods (‘95-’99 and ‘99-’03) prior to the commissioning of Carrigrennan WWTP as can be seen in Table 5.1.

<table>
<thead>
<tr>
<th>WFD body No.</th>
<th>Waterbody</th>
<th>EPA reporting period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>'95-’99</td>
</tr>
<tr>
<td>34</td>
<td>Lee Estuary upper</td>
<td>Eut.</td>
</tr>
<tr>
<td>37</td>
<td>Harper's Island</td>
<td>n/a</td>
</tr>
<tr>
<td>40</td>
<td>Glashaboy Estuary</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Eut = Eutrophic; P. Eut = Potentially eutrophic; Int = Intermediate risk of non-compliance; Unp = Unpolluted; n/a = Not available
### Table 5.2 Rank of each of the 39 EPA monitoring stations in Cork Harbour by marginal entropy values

<table>
<thead>
<tr>
<th>Ranking</th>
<th>BOD</th>
<th>MRP</th>
<th>DIN</th>
<th>CHLA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Station</td>
<td>Stn.</td>
<td>Station</td>
<td>Stn.</td>
</tr>
<tr>
<td>2</td>
<td>3.525</td>
<td>LE160</td>
<td>2.842</td>
<td>LE310</td>
</tr>
<tr>
<td>3</td>
<td>3.413</td>
<td>LE130</td>
<td>2.731</td>
<td>LE030</td>
</tr>
<tr>
<td>4</td>
<td>3.249</td>
<td>LE140</td>
<td>2.702</td>
<td>LE020</td>
</tr>
<tr>
<td>5</td>
<td>3.224</td>
<td>LE120</td>
<td>2.572</td>
<td>LE130</td>
</tr>
<tr>
<td>6</td>
<td>3.158</td>
<td>LE330</td>
<td>2.464</td>
<td>LE150</td>
</tr>
<tr>
<td>7</td>
<td>3.145</td>
<td>LE150</td>
<td>2.421</td>
<td>LE220</td>
</tr>
<tr>
<td>8</td>
<td>3.134</td>
<td>LE110</td>
<td>2.392</td>
<td>LE140</td>
</tr>
<tr>
<td>9</td>
<td>3.131</td>
<td>LE310</td>
<td>2.340</td>
<td>LE040</td>
</tr>
<tr>
<td>10</td>
<td>3.113</td>
<td>LE380</td>
<td>2.312</td>
<td>LE120</td>
</tr>
<tr>
<td>11</td>
<td>3.060</td>
<td>LE220</td>
<td>2.202</td>
<td>LE010</td>
</tr>
<tr>
<td>12</td>
<td>2.971</td>
<td>LE030</td>
<td>2.192</td>
<td>LE340</td>
</tr>
<tr>
<td>13</td>
<td>2.963</td>
<td>LE510</td>
<td>2.112</td>
<td>LE110</td>
</tr>
<tr>
<td>14</td>
<td>2.926</td>
<td>LE020</td>
<td>2.107</td>
<td>LE200</td>
</tr>
<tr>
<td>15</td>
<td>2.901</td>
<td>LE180</td>
<td>2.087</td>
<td>LE170</td>
</tr>
<tr>
<td>16</td>
<td>2.872</td>
<td>LE000</td>
<td>2.029</td>
<td>LE330</td>
</tr>
<tr>
<td>17</td>
<td>2.795</td>
<td>LE040</td>
<td>2.020</td>
<td>LE210</td>
</tr>
<tr>
<td>18</td>
<td>2.782</td>
<td>LE170</td>
<td>1.992</td>
<td>LE505</td>
</tr>
<tr>
<td>19</td>
<td>2.623</td>
<td>LE420</td>
<td>1.943</td>
<td>LE520</td>
</tr>
<tr>
<td>20</td>
<td>2.599</td>
<td>LE520</td>
<td>1.887</td>
<td>LE380</td>
</tr>
<tr>
<td>21</td>
<td>2.398</td>
<td>LE610</td>
<td>1.874</td>
<td>LE510</td>
</tr>
<tr>
<td>22</td>
<td>2.210</td>
<td>LE440</td>
<td>1.860</td>
<td>LE360</td>
</tr>
<tr>
<td>23</td>
<td>2.197</td>
<td>LE810</td>
<td>1.829</td>
<td>LE180</td>
</tr>
<tr>
<td>24</td>
<td>2.164</td>
<td>LE500</td>
<td>1.761</td>
<td>LE000</td>
</tr>
<tr>
<td>25</td>
<td>2.164</td>
<td>LE505</td>
<td>1.732</td>
<td>LE350</td>
</tr>
<tr>
<td>26</td>
<td>2.163</td>
<td>LE340</td>
<td>1.673</td>
<td>LE410</td>
</tr>
<tr>
<td>27</td>
<td>2.146</td>
<td>LE350</td>
<td>1.661</td>
<td>LE430</td>
</tr>
<tr>
<td>28</td>
<td>2.138</td>
<td>LE550</td>
<td>1.633</td>
<td>LE500</td>
</tr>
<tr>
<td>29</td>
<td>2.098</td>
<td>LE630</td>
<td>1.509</td>
<td>LE420</td>
</tr>
<tr>
<td>30</td>
<td>2.096</td>
<td>LE360</td>
<td>1.390</td>
<td>LE620</td>
</tr>
<tr>
<td>31</td>
<td>2.025</td>
<td>LE450</td>
<td>1.338</td>
<td>LE630</td>
</tr>
<tr>
<td>32</td>
<td>1.792</td>
<td>LE200</td>
<td>1.324</td>
<td>LE530</td>
</tr>
<tr>
<td>33</td>
<td>1.792</td>
<td>LE530</td>
<td>1.285</td>
<td>LE810</td>
</tr>
<tr>
<td>34</td>
<td>1.699</td>
<td>LE410</td>
<td>1.274</td>
<td>LE610</td>
</tr>
<tr>
<td>35</td>
<td>1.561</td>
<td>LE620</td>
<td>1.251</td>
<td>LE450</td>
</tr>
<tr>
<td>36</td>
<td>1.386</td>
<td>LE430</td>
<td>1.083</td>
<td>LE820</td>
</tr>
<tr>
<td>37</td>
<td>1.099</td>
<td>LE540</td>
<td>1.044</td>
<td>LE550</td>
</tr>
<tr>
<td>38</td>
<td>1.003</td>
<td>LE540</td>
<td>0.340</td>
<td>LE810</td>
</tr>
<tr>
<td>39</td>
<td>0.560</td>
<td>LE440</td>
<td>0.271</td>
<td>LE820</td>
</tr>
</tbody>
</table>
The 39 monitoring points used by the EPA and referenced in Table 5.3 are shown below in Figure 6.2.

Figure 5.2 Map of 39 EPA monitoring locations in the Cork Harbour area

Figures 5.4 to 5.6 present a comparison of the marginal entropies of PDFs derived from EPA data collected pre-Carrigrennan from 1994-2004, compared to water quality data simulated by the calibrated and validated pre-Carrigrennan numerical model at each monitoring point from a 3 year model simulation. The PDFs for each monitoring point and water quality parameter were derived using the upper and lower limits from the combined dataset comprising monitoring and simulated model data, to allow a fair comparison in Figures 5.4 to 5.6.

A general observation from all 3 figures is higher marginal entropy from each dataset in the Lee Estuary and Lough Mahon compared to the North Channel and Cork Harbour. The shape of both the EPA entropy and model entropy curves in Figures 5.5 and 5.6 for MRP and Chlorophyll_a respectively are similar from point LE610 to point LE820 while values are almost convergent at the most seaward monitoring point at LE820 where waters are closest to the tidal boundary condition and flows are tidally dominated.
The profiles of the MRP, Chlorophyll_a and DIN entropy derived from model simulations are similar along the lower Lee Estuary and Lough Mahon as entropy consistently exceeds EPA data entropy. The lower Lee Estuary and Lough Mahon is characterised by a relatively short residence time (Dabrowski 2005) compared to the residence time in the North Channel and Owenacurra Estuary where the model and EPA entropy profiles are closer to convergence. However, an unexpected outcome is observed in the case of DIN entropy, where EPA data entropy exceeded model entropy at all of the points in the North Channel and Owenacurra Estuary.

This discrepancy may be attributed to a longer residence time in the region, as presented in Figure 5.3, or the fact that river and outfall discharges and meteorological data are averaged and inputted to the numerical model on a monthly basis in this region.

Figure 5.3 Residence times in the greater Cork Harbour area (Dabrowski 2005) [Image removed for copyright reasons]

Considering that the model derived marginal entropy at each monitoring point is collated from high resolution timeseries data, one would expect that the model entropy curve would consistently exceed the EPA data entropy at each respective monitoring point, due to the intermittent sampling by the EPA 3 times during the summer and once during winter.
Figure 5.4 Comparison of marginal entropy of PDFs of MRP at each monitoring point, derived from monitoring data and simulated water quality data

Figure 5.5 Comparison of marginal entropy of PDFs of Chlorophyll_a at each monitoring point, derived from monitoring data and simulated water quality data
5.4. Existing monitoring network

5.4.1. Spatial analysis

The order of the priority for each of the existing monitoring points, with respect to each of the four trophic status parameters, was determined by applying the methodology for spatial entropy analysis presented in Chapter 4 to timeseries data generated by a summer and winter model simulation.

Table 5.3 is presented as an example of how spatial optimisation is applied to DIN monitoring for a 1 year numerical model simulation; the methodology was also applied to a 12 month model simulation from January to December, but was excluded from discussion in this chapter because the existing TSAS monitoring regime deals with the summer and winter periods separately.

The first step in the spatial entropy analysis is referred to in box one in Figure 4.1 of Chapter 4. All points are arranged in terms of marginal entropy, from highest to lowest. In this case, the point with the largest marginal entropy is LE170. Thus, the highest priority monitoring station LE170 is ranked number one. The second step is...
to add LE170 to the optimum network on the top left of Table 5.3. The third step in spatial entropy analysis to add each of the remaining 23 points to a list of candidate monitoring points, signified by box three in Figure 4.1. The 23 points are listed along the top row of Table 5.3. The fourth step is iterative and symbolised by box four in Figure 4.1. Taking the second row of Table 5.4, the transinformation between LE170 and each candidate point is determined using Equation (2.60). Examining the values along the row, it can be seen that LE820 has the lowest transinformation with point LE170. Hence, LE820 is added to the optimised network in the second column. The next step entails deriving the transinformation between LE170 and LE820, and each remaining candidate point, using Equations (2.61) and (2.62). LE540 is added in the third row of the table, as the third priority monitoring point.

The transinformation between the optimum monitoring network, and each candidate monitoring point, is determined as the optimum monitoring network grows and the list of candidate points shrinks. The process continues until all points have been used. The addition of each monitoring station contributes to a greater understanding of the water quality parameter under review.

In Table 5.3, the monitoring points ranked three to six are all in close proximity to one another; it should be noted that Table 5.3 presents spatial optimisation for a 12 month simulation. Therefore, the high prioritisation of these monitoring points may indicate that, on a 12 month basis, the data at these locations is different to the extent that there is low transinformation between the adjacent points.

The same process was repeated for each of the four parameters, using timeseries data generated for a summer and winter simulation. Table 5.3 presents the order of spatial priority for monitoring each of the four TSAS trophic status parameters during summer and winter. The model simulation periods were from March to September and September to March respectively, with an extra month allowed at the beginning for the model to reach steady state conditions. With each additional monitoring point that was added to the optimised monitoring network, the transinformation between the points in the optimised network increased, up to the stage when all points had been added.
Table 5.3 Spatial optimisation of the DIN monitoring network from a 1 year numerical model simulation

| Rank | Stn. No. | 150 | 160 | 180 | 220 | 310 | 330 | 340 | 350 | 380 | 410 | 420 | 430 | 450 | 510 | 520 | 530 | 540 | 550 | 610 | 620 | 630 | 810 | 820 |
|------|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1    | 170      | 1.091 | 1.123 | 1.133 | 1.037 | 1.139 | 1.364 | 1.150 | 1.080 | 0.871 | 0.491 | 0.579 | 0.462 | 0.157 | 0.118 | 0.101 | 0.066 | 0.036 | 0.536 | 1.173 | 1.242 | 0.486 | 0.375 | 0.009 |
| 2    | 820      | 1.093 | 1.127 | 1.136 | 1.040 | 1.141 | 1.368 | 1.152 | 1.085 | 0.875 | 0.491 | 0.579 | 0.463 | 0.157 | 0.118 | 0.101 | 0.066 | 0.040 | 0.539 | 1.176 | 1.244 | 0.488 | 0.375 |
| 3    | 540      | 1.215 | 1.251 | 1.253 | 1.158 | 1.257 | 1.483 | 1.261 | 1.207 | 0.985 | 0.502 | 0.586 | 0.487 | 0.242 | 0.118 | 0.102 | 0.069 | 0.603 | 1.178 | 1.292 | 0.577 | 0.412 |
| 4    | 530      | 1.221 | 1.258 | 1.262 | 1.162 | 1.265 | 1.488 | 1.266 | 1.207 | 1.032 | 0.503 | 0.634 | 0.573 | 0.242 | 0.470 | 0.583 | 0.610 | 1.202 | 1.377 | 0.632 | 0.413 |
| 5    | 450      | 1.315 | 1.348 | 1.339 | 1.238 | 1.341 | 1.548 | 1.338 | 1.295 | 1.083 | 0.503 | 0.645 | 0.600 | 0.473 | 0.583 | 0.619 | 1.213 | 1.421 | 0.688 | 0.492 |
| 6    | 510      | 1.333 | 1.367 | 1.363 | 1.258 | 1.365 | 1.569 | 1.358 | 1.303 | 1.106 | 0.504 | 0.716 | 0.683 | 1.081 | 0.620 | 1.215 | 1.445 | 0.730 | 0.493 |
| 7    | 810      | 1.521 | 1.550 | 1.547 | 1.438 | 1.550 | 1.722 | 1.543 | 1.493 | 1.255 | 0.509 | 0.738 | 0.715 | 1.081 | 0.634 | 1.252 | 1.490 | 0.889 |
| 8    | 410      | 1.533 | 1.568 | 1.565 | 1.470 | 1.559 | 1.730 | 1.559 | 1.519 | 1.316 | 0.741 | 0.716 | 1.081 | 0.637 | 1.268 | 1.491 | 0.889 |
| 9    | 550      | 1.593 | 1.633 | 1.616 | 1.525 | 1.608 | 1.774 | 1.606 | 1.588 | 1.365 | 0.748 | 0.725 | 1.082 | 1.284 | 1.544 | 0.915 |
| 10   | 430      | 1.640 | 1.682 | 1.661 | 1.564 | 1.651 | 1.810 | 1.643 | 1.620 | 1.400 | 0.981 | 1.085 | 1.286 | 1.566 | 0.937 |
| 11   | 630      | 1.807 | 1.869 | 1.886 | 1.791 | 1.864 | 2.001 | 1.871 | 1.817 | 1.596 | 1.017 | 1.085 | 1.312 | 1.683 |
| 12   | 420      | 1.831 | 1.896 | 1.927 | 1.828 | 1.905 | 2.037 | 1.908 | 1.840 | 1.601 | 1.089 | 1.313 | 1.685 |
| 13   | 520      | 1.831 | 1.896 | 1.927 | 1.828 | 1.905 | 2.038 | 1.908 | 1.841 | 1.604 | 1.317 | 1.693 |
| 14   | 610      | 1.851 | 1.921 | 1.942 | 1.838 | 1.925 | 2.062 | 1.919 | 1.852 | 1.662 | 1.792 |
| 15   | 380      | 2.126 | 2.208 | 2.311 | 2.217 | 2.309 | 2.397 | 2.313 | 2.125 | 2.034 |
| 16   | 620      | 2.181 | 2.243 | 2.349 | 2.224 | 2.321 | 2.399 | 2.322 | 2.149 | 2.034 |
| 17   | 350      | 2.921 | 3.461 | 3.399 | 3.152 | 3.157 | 2.829 | 2.998 |
| 20   | 220      | 3.587 | 4.292 | 3.762 | 3.764 |
| 21   | 160      | 4.325 | 3.763 | 3.863 |
| 22   | 310      | 4.620 | 4.062 |
| 23   | 340      | 4.787 |
The number of monitoring stations required to achieve 25%, 50% and 75% of the ultimate monitoring network transinformation are highlighted in amber, yellow and green respectively in Figure 5.7 and Table 5.4. The 25, 50 and 75% TST monitoring thresholds indicated in Table 5.4 and Figure 5.7 were identified by dividing the transinformation of each optimal monitoring network, ranging in size from 2 to 24 points, by the total transinformation delivered by all 24 monitoring points. The percentage of total transinformation delivered by each network size from 2 to 24 was rounded down. Where there was no network size with exactly 25, 50 or 75% TST, the rate of change of TST was determined by subtracting the difference between the two percentage values either side of the threshold of interest. The number of stations which delivered the exact threshold percentage was determined to a decimal point, and subsequently rounded, and the threshold percentage of interest was allocated to the resultant value. i.e. number of stations.

**Table 5.4 Summary of monitoring station spatial priority in Cork Harbour**

<table>
<thead>
<tr>
<th>Rank</th>
<th>DIN</th>
<th>MRP</th>
<th>DO</th>
<th>CHLA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Summer</td>
<td>Winter</td>
<td>Summer</td>
<td>Winter</td>
</tr>
<tr>
<td>1</td>
<td>LE170</td>
<td>LE170</td>
<td>LE150</td>
<td>LE150</td>
</tr>
<tr>
<td>2</td>
<td>LE820</td>
<td>LE820</td>
<td>LE220</td>
<td>LE170</td>
</tr>
<tr>
<td>3</td>
<td>LE450</td>
<td>LE540</td>
<td>LE330</td>
<td>LE410</td>
</tr>
<tr>
<td>4</td>
<td>LE550</td>
<td>LE530</td>
<td>LE170</td>
<td>LE820</td>
</tr>
<tr>
<td>5</td>
<td>LE540</td>
<td>LE450</td>
<td>LE530</td>
<td>LE170</td>
</tr>
<tr>
<td>6</td>
<td>LE430</td>
<td>LE520</td>
<td>LE810</td>
<td>LE160</td>
</tr>
<tr>
<td>7</td>
<td>LE410</td>
<td>LE430</td>
<td>LE410</td>
<td>LE540</td>
</tr>
<tr>
<td>8</td>
<td>LE630</td>
<td>LE410</td>
<td>LE310</td>
<td>LE610</td>
</tr>
<tr>
<td>9</td>
<td>LE420</td>
<td>LE550</td>
<td>LE380</td>
<td>LE450</td>
</tr>
<tr>
<td>10</td>
<td>LE510</td>
<td>LE420</td>
<td>LE340</td>
<td>LE630</td>
</tr>
<tr>
<td>11</td>
<td>LE810</td>
<td>LE810</td>
<td>LE160</td>
<td>LE550</td>
</tr>
<tr>
<td>12</td>
<td>LE530</td>
<td>LE510</td>
<td>LE180</td>
<td>LE180</td>
</tr>
<tr>
<td>13</td>
<td>LE510</td>
<td>LE630</td>
<td>LE820</td>
<td>LE430</td>
</tr>
<tr>
<td>14</td>
<td>LE520</td>
<td>LE610</td>
<td>LE610</td>
<td>LE330</td>
</tr>
<tr>
<td>15</td>
<td>LE350</td>
<td>LE220</td>
<td>LE350</td>
<td>LE220</td>
</tr>
<tr>
<td>16</td>
<td>LE620</td>
<td>LE620</td>
<td>LE540</td>
<td>LE380</td>
</tr>
<tr>
<td>17</td>
<td>LE380</td>
<td>LE330</td>
<td>LE630</td>
<td>LE420</td>
</tr>
<tr>
<td>18</td>
<td>LE150</td>
<td>LE150</td>
<td>LE450</td>
<td>LE620</td>
</tr>
<tr>
<td>20</td>
<td>LE220</td>
<td>LE350</td>
<td>LE550</td>
<td>LE310</td>
</tr>
<tr>
<td>21</td>
<td>LE160</td>
<td>LE310</td>
<td>LE420</td>
<td>LE530</td>
</tr>
<tr>
<td>22</td>
<td>LE310</td>
<td>LE160</td>
<td>LE530</td>
<td>LE520</td>
</tr>
<tr>
<td>23</td>
<td>LE340</td>
<td>LE340</td>
<td>LE430</td>
<td>LE510</td>
</tr>
<tr>
<td>24</td>
<td>LE180</td>
<td>LE180</td>
<td>LE520</td>
<td>LE340</td>
</tr>
</tbody>
</table>

| 25% | 50% | 75% |
Figures 5.8 and 5.9 present the spatial distribution of monitoring priority for Chlorophyll_a for summer and winter respectively. The distribution of priority is largely the same with some exceptions; during the summer, the focus of monitoring should be on Lough Mahon and Harper’s Island, whilst in winter, the lower Lee Estuary and the junction between the North Channel and the Owenacurra Estuary is emphasised. Examining the summer Chlorophyll_a priority ranking in Table 5.4, the first 7 monitoring points which contain 25% of the ultimate monitoring network transinformation are focused on the Harper’s Island and Lough Mahon area with the lower Lee Estuary and the Owenacurra Estuary.
Figure 5.8 Monitoring station priority from 1 to 24 for Chlorophyll_a sampling during summer

Figure 5.9 Monitoring station priority from 1 to 24 for Chlorophyll_a sampling during winter
Optimisation of water quality monitoring programmes via Entropy methods

The spatial distribution of MRP monitoring priority in summer in Figure 5.10 highlights the lower Lee Estuary, Lough Mahon and the Glashaboy Estuary. Figure 5.11 presents the priority for MRP during winter; emphasis remains on the lower Lee Estuary, while the North Channel and Inner Harbour are given a greater priority than during summer. The monitoring networks that capture 25% of the summer and winter overall monitoring network transinformation for the summer and winter MRP respectively are unique with respect to their particular areas of coverage. As presented in columns 4 and 5 of Table 5.4, the monitoring points prioritized in the summer are focused on the Lee Estuary and Lough Mahon while the winter phosphate monitoring emphasizes the Lee Estuary, the North Channel and the outer Cork Harbour area. Water quality reporting corresponding to the time period covered by the data used to calibrate and validate the water quality model refer to noncompliance of nutrient criteria in Lough Mahon for that period but fails to specify whether this was from summer or winter assessment. A lack of winter nutrient data for the North Channel is highlighted in the same report.

Figure 5.10 Monitoring station priority from 1 to 24 for MRP sampling during summer
Comparing Figures 5.12 and 5.13, it can be seen that the monitoring priority for DIN is largely the same, with the exception of some minor differences in the continuum from the lower Lee Estuary to Lough Mahon. The Owenacurra Estuary was assessed as eutrophic (Toner et al. 2005) in the period from 2001 to 2003 on the basis of high levels of nitrogen in the river; water quality data from this period was used to calibrate and validate the water quality model. The Owenacurra Estuary is located at a junction between the mouth of the Owenacurra and the North Channel which has a much longer residence time. The high levels of nitrogen entering the estuary, coupled with the restricted flows around the estuary may result in a variation in the DIN at each monitoring point, and hence low transinformation between points. Therefore, the prioritization of the Owenacurra for monitoring during the summer may be reflected in water quality reporting at that time. In the same period, Lough Mahon was classed as Eutrophic as a consequence of elevated Chlorophyll_a and breaches of nutrient thresholds.
Figure 5.12 Monitoring station priority from 1 to 24 for DIN sampling during summer

Figure 5.13 Monitoring station priority from 1 to 24 for DIN sampling during winter

The same observation can be made with DO monitoring priority, presented in Figures 5.14 and 5.15 for summer and winter respectively. There is little difference
between the summer and winter monitoring priority distributions, with the only change in emphasis being a slightly greater priority placed on the lower Lee Estuary and the North Channel during winter. DO Saturation % is assessed for compliance during the summer. The corresponding column 6 of Table 5.4 indicates that the 8 highest priority monitoring points that provide 25% of the total monitoring network transinformation consists of three locations in the lower Lee Estuary, one point from Lough Mahon, one point from the North Channel/Great Island area, two points from Cork Harbour and the outer Cork Harbour area. Water quality reporting for 2001-2003 corresponding to the period from which the model is validated indicated that low oxygen concentrations were noted on the corresponding stretch of the lower Lee Estuary, from the Port of Cork to Blackrock Castle.

Figure 5.14 Monitoring station priority from 1 to 24 for DO sampling during summer
Monitoring station priority for nutrients was considered for both the winter and summer periods. In both summer and winter, half of the 24 monitoring points are required to capture 25% of the total monitoring network transinformation. Of the 12 highest priority monitoring points nominated for each of the assessment periods, there is an overlap of 10 monitoring points. Each of the 2 lists of 12 top priority monitoring points contain locations in the North Channel, Owenacurra Estuary and the outer Cork Harbour area, with the summer list emphasizing the inner Cork Harbour area and the winter list containing all of the monitoring points in the Owenacurra Estuary.

Water quality reporting for the period for which the model was calibrated and validated confirms that there was non-compliance of winter DIN levels in the lower Lee Estuary area. The same report declares the Owenacurra Estuary as eutrophic due to considerable nitrogen levels in the Owenacurra River. A “lack of comprehensive nutrient data” for the North Channel was mentioned in the report, in particular for the winter months. The summer and winter monitoring network which provides the top 25% of the monitoring network transinformation contains all of the areas which were of concern in compliance assessment for the time period that the water quality model was calibrated and validated.
5.4.2. Spatial/Temporal analysis

Figure 4.2 presents the procedure that was followed in completing spatial/temporal entropy analysis. The process was initiated based on the monitoring station ranking for each water quality parameter and assessment period included in Table 5.4. The transinformation entropy was evaluated between sampling at the highest priority monitoring point at the base sampling frequency, and all other points at multiples of the base sampling frequency up to the maximum sampling frequency, using Equation (2.63). Weekly and monthly base sampling frequencies were considered in the derivation of optimum monitoring frequencies for a monitoring network increasing in size from 2 to 24 monitoring stations. Therefore, Tables 5.5 and 5.6 that are presented for discussion of the spatial/temporal analysis exclude a network comprising a single monitoring point.

Starting with a monitoring network comprising the top two monitoring points, the timeseries at the primary monitoring point is sampled at the base sampling frequency (weekly or monthly), while the secondary monitoring point is sampled at each of the multiples of the base frequency. The transinformation between the PDF derived from the primary monitoring point at the highest frequency, and the PDF at the secondary monitoring point, sampled at a multiple of the base frequency, is determined. The process is repeated for the same network size of two points, but for each multiple of the base frequency. The optimum monitoring frequency for the network of size two, for the parameter in question, for the base frequency being used, is the multiple which delivers the lowest transinformation. The process is repeated for each parameter and base sampling frequency, until the full range in monitoring network sizes have been assessed.

From Table 5.5 it can be seen that there is a direct trade-off between an increase in the size of the monitoring network and the sampling frequency which delivers the least redundant information for that network size for Chlorophyll a monitoring over a year long period for the purposes of the WFD. From Table 5.5 it can be seen that there is an increase in the transinformation as each lower priority monitoring station is added to the network, but the optimal sampling frequency for each network size as highlighted in the table is the frequency for which the lowest transinformation is
observed. The sampling frequency or monitoring station which achieves the lowest transinformation or redundant information yields the greatest gain in unique information regarding the water quality parameter in question. For each monitoring network size, the transinformation fluctuates with lowering sampling frequency. However, for the same sampling frequency, there will always be an increase in transinformation when comparing a monitoring network of 2 points to a network of 3 points. The task of finding the optimum monitoring frequency for a given network is thus: identify the lowest transinformation when considering all sampling frequencies.

The observed decrease in the necessary sampling frequency with an increase in the size of the monitoring network, although intuitive, is not necessarily true for all optimal monitoring network configurations whether considering weekly or monthly sampling resolution. In some cases, the unexpected situation emerged, that increased sampling was required with the addition of a monitoring point to the reconfigured network. For example in Table 5.6, the addition of monitoring points 8-11 allows the sampling frequency to be extended from 5 months to 6 months. However, if any more than 11 monitoring points are utilised, a higher sampling frequency of every 3 months is required. Inspecting Table 5.4, the monitoring locations which are 8th to 11th priority for summer monitoring of MRP cover the Lough Mahon and Passage West. The first 11 monitoring points emphasise Lough Mahon and the lower Lee Estuary whilst the monitoring points ranked from 12 to 24 mainly cover the North Channel, Owenacurra Estuary and the inner Cork Harbour. The reason for the requirement of sampling at a higher frequency when using more than 11 monitoring points may be that MRP dynamics occur at a different timescale in the North Channel and Owenacurra Estuary compared to the Lough Mahon and lower Lee Estuary. There is a slower turnover of water in the North Channel and Owenacurra which is reflected in a longer residence time in the region of 50-80 days, compared to less than 30 days in the Lough Mahon and lower Lee Estuary (Dabrowski 2005)
Table 5.5 Transinformation between sampling Chlorophyll-a at LE330 on a monthly basis for a one year period, and each additional monitoring point from LE610 to LE530 at multiples of the monthly frequency

<table>
<thead>
<tr>
<th>Network size</th>
<th>Stations</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>LE330, LE610</td>
<td>0.41</td>
<td>0.51</td>
<td>0.48</td>
<td>0.42</td>
<td>0.51</td>
<td>0.53</td>
<td>0.59</td>
<td>0.77</td>
<td>0.49</td>
<td>0.63</td>
<td>0.44</td>
</tr>
<tr>
<td>3</td>
<td>LE160</td>
<td>0.42</td>
<td>0.72</td>
<td>1.08</td>
<td>1.17</td>
<td>0.67</td>
<td>0.53</td>
<td>0.59</td>
<td>1.00</td>
<td>0.86</td>
<td>0.87</td>
<td>0.55</td>
</tr>
<tr>
<td>4</td>
<td>LE410</td>
<td>0.42</td>
<td>1.16</td>
<td>1.12</td>
<td>1.43</td>
<td>0.70</td>
<td>0.61</td>
<td>1.05</td>
<td>1.01</td>
<td>0.86</td>
<td>0.88</td>
<td>1.31</td>
</tr>
<tr>
<td>5</td>
<td>LE820</td>
<td>0.43</td>
<td>1.80</td>
<td>1.12</td>
<td>1.74</td>
<td>1.26</td>
<td>0.68</td>
<td>1.72</td>
<td>1.19</td>
<td>1.30</td>
<td>0.98</td>
<td>1.38</td>
</tr>
<tr>
<td>6</td>
<td>LE170</td>
<td>0.87</td>
<td>1.91</td>
<td>1.17</td>
<td>2.41</td>
<td>1.65</td>
<td>1.13</td>
<td>1.90</td>
<td>1.27</td>
<td>1.38</td>
<td>2.06</td>
<td>1.48</td>
</tr>
<tr>
<td>7</td>
<td>LE810</td>
<td>1.46</td>
<td>2.18</td>
<td>1.44</td>
<td>2.45</td>
<td>1.82</td>
<td>1.57</td>
<td>1.92</td>
<td>1.72</td>
<td>1.77</td>
<td>2.10</td>
<td>1.92</td>
</tr>
<tr>
<td>8</td>
<td>LE350</td>
<td>1.46</td>
<td>2.19</td>
<td>1.45</td>
<td>2.48</td>
<td>2.39</td>
<td>1.57</td>
<td>2.10</td>
<td>1.79</td>
<td>2.34</td>
<td>2.28</td>
<td>1.92</td>
</tr>
<tr>
<td>9</td>
<td>LE150</td>
<td>1.46</td>
<td>2.37</td>
<td>1.46</td>
<td>2.52</td>
<td>2.55</td>
<td>1.59</td>
<td>2.16</td>
<td>1.80</td>
<td>2.34</td>
<td>2.60</td>
<td>2.07</td>
</tr>
<tr>
<td>10</td>
<td>LE180</td>
<td>1.73</td>
<td>2.37</td>
<td>1.47</td>
<td>2.52</td>
<td>2.55</td>
<td>1.63</td>
<td>2.17</td>
<td>1.90</td>
<td>2.43</td>
<td>2.63</td>
<td>2.07</td>
</tr>
<tr>
<td>11</td>
<td>LE510</td>
<td>2.22</td>
<td>2.41</td>
<td>1.85</td>
<td>2.52</td>
<td>2.93</td>
<td>2.01</td>
<td>2.88</td>
<td>1.90</td>
<td>2.43</td>
<td>2.70</td>
<td>2.27</td>
</tr>
<tr>
<td>12</td>
<td>LE310</td>
<td>2.29</td>
<td>2.55</td>
<td>1.92</td>
<td>2.55</td>
<td>3.04</td>
<td>2.13</td>
<td>2.93</td>
<td>1.96</td>
<td>2.53</td>
<td>2.70</td>
<td>2.54</td>
</tr>
<tr>
<td>13</td>
<td>LE340</td>
<td>2.45</td>
<td>3.11</td>
<td>2.00</td>
<td>2.75</td>
<td>3.19</td>
<td>2.19</td>
<td>2.93</td>
<td>2.02</td>
<td>2.85</td>
<td>2.71</td>
<td>2.54</td>
</tr>
<tr>
<td>14</td>
<td>LE380</td>
<td>2.53</td>
<td>3.23</td>
<td>2.54</td>
<td>3.14</td>
<td>3.21</td>
<td>2.21</td>
<td>2.96</td>
<td>2.26</td>
<td>2.88</td>
<td>2.76</td>
<td>2.60</td>
</tr>
<tr>
<td>15</td>
<td>LE540</td>
<td>2.60</td>
<td>3.65</td>
<td>2.57</td>
<td>3.27</td>
<td>3.35</td>
<td>2.32</td>
<td>3.00</td>
<td>2.32</td>
<td>2.99</td>
<td>2.76</td>
<td>3.20</td>
</tr>
<tr>
<td>16</td>
<td>LE220</td>
<td>2.60</td>
<td>4.01</td>
<td>2.80</td>
<td>3.56</td>
<td>3.54</td>
<td>2.34</td>
<td>3.35</td>
<td>2.35</td>
<td>3.07</td>
<td>3.10</td>
<td>3.24</td>
</tr>
<tr>
<td>17</td>
<td>LE630</td>
<td>2.61</td>
<td>4.10</td>
<td>3.45</td>
<td>3.69</td>
<td>3.55</td>
<td>2.35</td>
<td>3.35</td>
<td>2.38</td>
<td>3.12</td>
<td>3.10</td>
<td>3.64</td>
</tr>
<tr>
<td>18</td>
<td>LE420</td>
<td>2.68</td>
<td>4.11</td>
<td>3.66</td>
<td>3.74</td>
<td>3.72</td>
<td>2.42</td>
<td>3.41</td>
<td>2.63</td>
<td>3.43</td>
<td>3.15</td>
<td>3.67</td>
</tr>
<tr>
<td>19</td>
<td>LE450</td>
<td>3.13</td>
<td>4.31</td>
<td>3.70</td>
<td>4.10</td>
<td>3.81</td>
<td>2.49</td>
<td>3.41</td>
<td>2.65</td>
<td>3.43</td>
<td>3.59</td>
<td>3.72</td>
</tr>
<tr>
<td>20</td>
<td>LE620</td>
<td>3.57</td>
<td>4.60</td>
<td>3.83</td>
<td>4.10</td>
<td>3.81</td>
<td>3.45</td>
<td>3.66</td>
<td>3.37</td>
<td>4.95</td>
<td>3.79</td>
<td>3.94</td>
</tr>
<tr>
<td>21</td>
<td>LE550</td>
<td>3.57</td>
<td>4.64</td>
<td>3.85</td>
<td>4.15</td>
<td>4.03</td>
<td>3.46</td>
<td>3.67</td>
<td>3.37</td>
<td>5.01</td>
<td>3.88</td>
<td>4.04</td>
</tr>
<tr>
<td>22</td>
<td>LE520</td>
<td>3.59</td>
<td>4.65</td>
<td>4.15</td>
<td>4.18</td>
<td>5.08</td>
<td>3.52</td>
<td>3.81</td>
<td>3.39</td>
<td>5.39</td>
<td>4.19</td>
<td>4.04</td>
</tr>
<tr>
<td>23</td>
<td>LE430</td>
<td>3.62</td>
<td>4.70</td>
<td>4.16</td>
<td>4.23</td>
<td>5.08</td>
<td>3.61</td>
<td>4.60</td>
<td>3.56</td>
<td>5.43</td>
<td>4.27</td>
<td>4.07</td>
</tr>
<tr>
<td>24</td>
<td>LE530</td>
<td>4.01</td>
<td>4.72</td>
<td>4.18</td>
<td>5.93</td>
<td>5.08</td>
<td>3.67</td>
<td>4.60</td>
<td>3.60</td>
<td>6.25</td>
<td>4.34</td>
<td>4.16</td>
</tr>
</tbody>
</table>
Table 5.6 Transinformation between sampling MRP at LE150 on a monthly basis during summer, and each additional monitoring point from LE220 to LE520 at multiples of the monthly frequency

<table>
<thead>
<tr>
<th>Network size</th>
<th>Stations</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>LE150, LE220</td>
<td>0.01</td>
<td>0.22</td>
<td>0.01</td>
<td>0.21</td>
<td>0.14</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>LE330</td>
<td>0.27</td>
<td>0.25</td>
<td>0.41</td>
<td>0.32</td>
<td>0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>4</td>
<td>LE170</td>
<td>0.29</td>
<td>0.34</td>
<td>0.42</td>
<td>0.33</td>
<td>0.17</td>
<td>0.47</td>
</tr>
<tr>
<td>5</td>
<td>LE510</td>
<td>0.30</td>
<td>0.43</td>
<td>0.52</td>
<td>0.34</td>
<td>0.20</td>
<td>0.65</td>
</tr>
<tr>
<td>6</td>
<td>LE810</td>
<td>0.40</td>
<td>0.62</td>
<td>0.83</td>
<td>0.43</td>
<td>0.21</td>
<td>0.70</td>
</tr>
<tr>
<td>7</td>
<td>LE410</td>
<td>0.56</td>
<td>0.66</td>
<td>0.98</td>
<td>0.95</td>
<td>0.28</td>
<td>0.70</td>
</tr>
<tr>
<td>8</td>
<td>LE310</td>
<td>0.88</td>
<td>0.87</td>
<td>1.10</td>
<td>0.99</td>
<td>0.80</td>
<td>0.70</td>
</tr>
<tr>
<td>9</td>
<td>LE380</td>
<td>0.92</td>
<td>0.90</td>
<td>1.10</td>
<td>1.01</td>
<td>0.84</td>
<td>0.70</td>
</tr>
<tr>
<td>10</td>
<td>LE340</td>
<td>0.94</td>
<td>0.99</td>
<td>1.15</td>
<td>1.01</td>
<td>0.86</td>
<td>0.76</td>
</tr>
<tr>
<td>11</td>
<td>LE160</td>
<td>1.10</td>
<td>1.25</td>
<td>1.24</td>
<td>1.25</td>
<td>1.25</td>
<td>1.05</td>
</tr>
<tr>
<td>12</td>
<td>LE180</td>
<td>1.45</td>
<td>1.41</td>
<td>1.38</td>
<td>1.39</td>
<td>1.40</td>
<td>1.51</td>
</tr>
<tr>
<td>13</td>
<td>LE820</td>
<td>1.48</td>
<td>1.44</td>
<td>1.40</td>
<td>1.41</td>
<td>1.42</td>
<td>1.52</td>
</tr>
<tr>
<td>14</td>
<td>LE610</td>
<td>1.61</td>
<td>1.57</td>
<td>1.52</td>
<td>1.53</td>
<td>1.55</td>
<td>1.53</td>
</tr>
<tr>
<td>15</td>
<td>LE350</td>
<td>1.76</td>
<td>1.71</td>
<td>1.64</td>
<td>1.65</td>
<td>1.67</td>
<td>1.64</td>
</tr>
<tr>
<td>16</td>
<td>LE540</td>
<td>1.90</td>
<td>1.83</td>
<td>1.75</td>
<td>1.76</td>
<td>1.78</td>
<td>1.82</td>
</tr>
<tr>
<td>17</td>
<td>LE630</td>
<td>2.03</td>
<td>1.92</td>
<td>1.82</td>
<td>1.84</td>
<td>1.85</td>
<td>1.99</td>
</tr>
<tr>
<td>18</td>
<td>LE450</td>
<td>2.27</td>
<td>2.16</td>
<td>2.05</td>
<td>2.07</td>
<td>2.09</td>
<td>2.06</td>
</tr>
<tr>
<td>19</td>
<td>LE620</td>
<td>2.31</td>
<td>2.19</td>
<td>2.07</td>
<td>2.10</td>
<td>2.12</td>
<td>2.22</td>
</tr>
<tr>
<td>20</td>
<td>LE550</td>
<td>2.40</td>
<td>2.27</td>
<td>2.12</td>
<td>2.14</td>
<td>2.17</td>
<td>2.23</td>
</tr>
<tr>
<td>21</td>
<td>LE420</td>
<td>2.49</td>
<td>2.31</td>
<td>2.14</td>
<td>2.17</td>
<td>2.19</td>
<td>2.29</td>
</tr>
<tr>
<td>22</td>
<td>LE530</td>
<td>2.57</td>
<td>2.39</td>
<td>2.21</td>
<td>2.24</td>
<td>2.27</td>
<td>2.50</td>
</tr>
<tr>
<td>23</td>
<td>LE430</td>
<td>2.71</td>
<td>2.51</td>
<td>2.32</td>
<td>2.35</td>
<td>2.38</td>
<td>2.69</td>
</tr>
<tr>
<td>24</td>
<td>LE520</td>
<td>3.04</td>
<td>2.83</td>
<td>2.62</td>
<td>2.65</td>
<td>2.68</td>
<td>2.87</td>
</tr>
</tbody>
</table>

A summary of the optimum sampling frequencies required to capture 25%, 50%, 75% and 100% of the total monitoring network transinformation for the summer and winter TSAS assessment periods and the monitoring in any given year for the purposes of the WFD is contained in Table 5.7 which summarises the outcome of each spatial/temporal analysis. For each of the parameters evaluated as part of TSAS assessment, multiples of weekly and monthly resolutions were evaluated. The size and locations of the monitoring network which contains 25-100% of transinformation is known in advance as summarised in Table 5.4 and the sampling frequency which minimised transinformation for that network size was identified in tables such as Table 5.5. Each value in Table 5.7 is the frequency required for sampling, within the 6 month summer and winter period. The values in Table 5.7...
represent the time lag between subsequent sampling events, with the lag in weeks or months, depending upon the base resolution.

### Table 5.7 The optimum sampling frequency for all management options

<table>
<thead>
<tr>
<th>Sampling Frequency</th>
<th>Weeks</th>
<th>Months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIN</td>
<td>MRP</td>
</tr>
<tr>
<td><strong>Summer</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>50%</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>75%</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>100%</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td><strong>Winter</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>50%</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>75%</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>100%</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

TSAS trophic assessment is carried out based on four parameters: DIN, MRP, DO and Chlorophyll_a monitoring data during summer and winter. Nutrient concentrations are evaluated on the basis of summer or winter median values whilst only summer monitoring data is used in assessment of oxygenation conditions and Chlorophyll_a. Summer monitoring data is collected during three monitoring campaigns at approximately monthly intervals during the June to August period and winter sampling is completed once at all monitoring stations between December and January. Considering the 24 monitoring stations in the greater Cork Harbour area, TSAS monitoring entails a sampling effort of 72 grab samples during the summer assessment period and 24 during the winter period. This is based on sampling thrice during the summer (24 x 3) and once during the winter (24 x 1) respectively.

Under the WFD, the minimum sampling frequency required for chlorophyll is every 6 months whilst nutrient and oxygenation status sampling must be carried out every 3 months which equates to 48 (24 x (12/6)) and 96 (24 x (12/3)) grab samples required respectively throughout the Cork Harbour monitoring network. Table 5.8 contains values which were determined as the number of times during summer and winter that the waters could be sampled by honouring the optimum sampling frequencies declared in Table 5.7. Therefore, where the sampling frequency was every 2 months in a 6 month period, the number of sampling events required per site would be 4, assuming that samples were taken from the very beginning of the 6 month period. However, where the optimum sampling frequency was every 4 or 5
months in a 6 month period, the optimum number of sampling events at each site would be 2. Table 5.9 summarises the monitoring network sizes which capture 25-100% of the total monitoring network transinformation contained in Table 5.4. Table 5.10 presents the monitoring effort for each management option; monitoring effort refers to the product of the number of sites required to be sampled (see Table 5.9) and the number of sampling events required per site (see Table 5.8).

Table 5.8 The optimum number of sampling events required per monitoring location for all management options

<table>
<thead>
<tr>
<th></th>
<th>Weeks</th>
<th></th>
<th>Months</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIN</td>
<td>MRP</td>
<td>DO</td>
<td>CHLA</td>
</tr>
<tr>
<td>Summer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>50%</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>75%</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>100%</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Winter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>3</td>
<td>2</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td>50%</td>
<td>3</td>
<td>26</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>75%</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>100%</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 5.9 The optimum number of monitoring locations for all management options

<table>
<thead>
<tr>
<th></th>
<th>Weeks</th>
<th></th>
<th>Months</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIN</td>
<td>MRP</td>
<td>DO</td>
<td>CHLA</td>
</tr>
<tr>
<td>Summer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>12</td>
<td>6</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>50%</td>
<td>17</td>
<td>13</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>75%</td>
<td>21</td>
<td>19</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>100%</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Winter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>12</td>
<td>5</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>50%</td>
<td>15</td>
<td>9</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>75%</td>
<td>20</td>
<td>16</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>100%</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
</tbody>
</table>

Of the 64 summer and winter TSAS monitoring network configurations presented in Table 5.10, 28 require the current sampling and lab analysis effort or less. Of the 32 monthly TSAS monitoring network configurations, 17 require less than the current effort, while only 11 of the 32 weekly TSAS monitoring network configurations require less than the current effort. In order to capture 100% of the monitoring network transinformation in a winter model simulation, 2.75 times the monitoring effort of present monitoring would be necessary if using monthly
sampling, while 4.5 times the monitoring effort of present monitoring would be necessary if using weekly sampling. These values are determined assuming that current winter monitoring effort involves sampling at the 24 monitoring points, and 1 lab analysis procedure for each of the 4 water quality parameters, which amounts to an effort of 96. If this figure is divided into the sum of the first 4, or last 4 values in the bottom row of Table 5.10, the values of 4.5 and 2.75 respectively are returned.

When water quality monitoring frequency is considered at weekly resolution, a considerable monitoring effort is required, especially in the cases of DO Saturation % and Chlorophyll_a monitoring. The ultimate transinformation captured in the optimized monitoring network at multiples of monthly sampling intervals is marginally greater than the transinformation captured in the optimized monitoring programme that considers weekly sampling intervals in the case of summer and winter nutrient TSAS assessment, as seen in Table 5.11. Sampling at weekly resolution delivers higher ultimate transinformation for all DO Saturation % and Chlorophyll_a monitoring in every assessment period.

Table 5.10 The monitoring effort required to implement each optimum monitoring network size

<table>
<thead>
<tr>
<th></th>
<th>Weeks</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Monitoring Effort (No. of Samples x Monitoring Network Size)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DIN</td>
<td>MRP</td>
<td>DO</td>
<td>CHLA</td>
<td>DIN</td>
<td>MRP</td>
<td>DO</td>
</tr>
<tr>
<td>Summer</td>
<td></td>
<td>25%</td>
<td>36</td>
<td>24</td>
<td>40</td>
<td>28</td>
<td>48</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50%</td>
<td>51</td>
<td>26</td>
<td>80</td>
<td>60</td>
<td>68</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75%</td>
<td>84</td>
<td>38</td>
<td>100</td>
<td>120</td>
<td>84</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100%</td>
<td>96</td>
<td>48</td>
<td>96</td>
<td>144</td>
<td>96</td>
<td>72</td>
</tr>
<tr>
<td>Winter</td>
<td></td>
<td>25%</td>
<td>36</td>
<td>10</td>
<td>182</td>
<td>16</td>
<td>48</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50%</td>
<td>51</td>
<td>26</td>
<td>65</td>
<td>48</td>
<td>60</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75%</td>
<td>80</td>
<td>128</td>
<td>95</td>
<td>72</td>
<td>80</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100%</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>144</td>
<td>96</td>
<td>72</td>
</tr>
</tbody>
</table>

Table 5.11 Comparison of ultimate monitoring network transinformation at weekly and monthly resolution

<table>
<thead>
<tr>
<th></th>
<th>DIN</th>
<th>MRP</th>
<th>DO</th>
<th>CHLA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Summer</td>
<td>Winter</td>
<td>Summer</td>
<td>Winter</td>
</tr>
<tr>
<td>Weekly</td>
<td>1.765</td>
<td>1.533</td>
<td>2.559</td>
<td>2.218</td>
</tr>
<tr>
<td>Monthly</td>
<td>2.012</td>
<td>2.831</td>
<td>2.616</td>
<td>2.624</td>
</tr>
</tbody>
</table>
Table 5.12 presents the compliance level achieved using the combination of spatially optimised monitoring networks from 25-100% TST, coupled with the optimum sampling frequencies at weekly or monthly resolution. The compliance level returned using the existing monitoring regime is also presented for comparison, which entails sampling 3 times during the summer at monthly intervals and once in the winter.

The same compliance level was observed in Lough Mahon, the Owenacurra Estuary and Cork Harbour Outer by sampling the dataset with the spatially and temporally optimised monitoring programme as would be returned if the existing monitoring programme were followed.

Of the 7 waterbodies included in Table 5.12 which were assessed in the 1999 to 2003 period, Lough Mahon, Owenacurra Estuary and Cork Harbour Outer achieved the same trophic status using the 3 year numerical model simulation and the sampling regime at that time. All waterbodies with the exception of Cork Harbour Outer were assessed as eutrophic using the 3 year numerical model simulation.
Table 5.12 Comparison of the level of compliance of the proposed monitoring regime versus the existing monitoring regime; sampling is carried out on data generated from a 3 year model simulation. The corresponding EPA reported compliance status is also presented for comparison.

<table>
<thead>
<tr>
<th>Waterbody</th>
<th>% TST</th>
<th>DIVAST Weekly</th>
<th>Existing</th>
<th>DIVAST Monthly</th>
<th>EPA Report '99-'03</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Lee</td>
<td>25</td>
<td>Eut</td>
<td>Eut</td>
<td>Eut</td>
<td>Eut</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>Eut</td>
<td>Eut</td>
<td>Eut</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>Eut</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glashaboy</td>
<td>25</td>
<td>Int</td>
<td>Eut</td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lough Mahon</td>
<td>25</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Channel - Great Island</td>
<td>25</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owenacurra estuary</td>
<td>25</td>
<td>Int</td>
<td>Int</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>Int</td>
<td>Int</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>Int</td>
<td>Int</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cork harbour Inner</td>
<td>25</td>
<td>Int</td>
<td>Int</td>
<td></td>
<td>Eutra</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td>Inta</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>Eut</td>
<td>Eut</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cork harbour Outer</td>
<td>25</td>
<td>Int</td>
<td>Int</td>
<td></td>
<td>Inta</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>Int</td>
<td>Int</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>Int</td>
<td>Int</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>Int</td>
<td>Int</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Eut = Eutrophic; P. Eut = Potentially eutrophic; Int = Intermediate risk of non-compliance; Unp = Unpolluted; n/a = Not available

a Cork harbour was assessed in it's entirety in the 99-03 period and subsequently assessed as inner and outer sections.
5.5. Ab initio spatial monitoring design

Previous monitoring optimisation in this chapter has been restricted to the 24 existing EPA sanctioned monitoring points that fall within the model domain. The following section details the results of a full assessment of the suitability of each of the 90,233 30 x 30m wet grid cells in the model domain; firstly considering the waters by salinity class and secondly by dividing waters by salinity class and trophic status.

In the following sections, the pool of points to be ranked was limited based firstly on salinity, and then based on salinity and trophic status. The methodology that was adopted is outlined in Chapter 4. The model was run for a 12 month period and snapshots were produced at every spring high tide resulting in a PDF of each of the four TSAS parameters at each wet grid cell in the model domain. In each instance, the pool of points were initially organised in terms of their respective marginal entropies from high to low. The point or points with the highest entropy were removed from the pool, and assuming all points were identical, the transinformation between one of the highest entropy points, and each of the remaining points in the pool, was determined. The point or points with the lowest transinformation were all removed from the pool, and appointed the second priority. The process was repeated iteratively until all candidate points were used.

5.5.1. Monitoring design by salinity class

3 salinity classes were used as in the TSAS assessment protocol with freshwaters categorised as having a salinity less than 5 psu, intermediate salinity waters covering the range between 5 and 30 psu and coastal saline waters describing all waters with average salinity in excess of 30 psu. Figure 5.16 presents the three salinity classes. Freshwaters in the model domain extend up the Owenbuidhe Estuary. Intermediate salinity waters span from the lower end of the lower Lee Estuary, Glashaboy Estuary, Lough Mahon, Harper’s Island, Cork Harbour inner, and Owenacurra Estuary. Fully saline waters extend seaward from Roches’ point.
5.5.1.1. Freshwater

Figures 5.17 to 5.20 present the monitoring rank of Chlorophyll\textsubscript{a}, DIN, MRP and DO Saturation %. A similar pattern of priority is revealed in the monitoring priority for Chlorophyll\textsubscript{a}, MRP and DO saturation; either end of the Owenbuidhe Estuary are highlighted as distinct monitoring locations. 3 clear zones for monitoring are revealed in Figure 5.18 for DIN; the highest and lowest priority monitoring zones lie at the outer and inner extremes of the Owenbuidhe Estuary.
Figure 5.17 Priority of potential monitoring locations for Chlorophyll_a in freshwater

Figure 5.18 Priority of potential monitoring locations for DIN in freshwater
Figure 5.19 Priority of potential monitoring locations for MRP in freshwater

Figure 5.20 Priority of potential monitoring locations for DO sat% in freshwater
5.5.1.2. Intermediate salinity

Intermediate salinity waters extend from the lower Lee Estuary through to Cork Harbour Inner thus covering all 7 of the WFD waterbodies in the greater Cork Harbour area except Cork Harbour outer.

Four distinct zones can be identified from Figure 5.21 below; the lower Lee and Lough Mahon area (LE150-LE340), Whitegate (adjacent to Catchment 3), the North Channel (LE420 and LE430) and the Owenacurra Estuary (LE450-LE550). There is no monitoring point within the waters adjacent to Whitegate; this area is potentially a site for additional Chlorophyll_a monitoring due to the absence of any monitoring points there at present, although it would be a low priority based on the colour of the map in the area.

In Figure 5.22 below, the lower Lee Estuary, Lough Mahon and the west end of inner Cork Harbour are high priority and well sampled for DIN. The remainder of inner Cork Harbour has a lower rank of 10-25 and contains two regions which are unmonitored, the Whitegate Bay area and the waters between East Ferry and

Figure 5.21 Priority of potential monitoring locations for Chlorophyll_a in brackish waters
Optimisation of water quality monitoring programmes via Entropy methods

Aghada. The entire North Channel and Owenacurra Estuary area is well sampled within the EPA monitoring network and the only other distinct zone which is unmonitored lies to the east of Harper’s Island.

In relation to MRP monitoring priority in Figure 5.23, the western end of the North Channel area is distinctly identified by higher priority compared to the Lee Estuary and Lough Mahon as well as the Harper’s Island area and the Owenacurra Estuary. The Slatty waters area of Harper’s Island is distinct with regard to both MRP and DIN entropy rank in Figures 5.22 and 5.23.

The entire intermediate salinity region of the Cork Harbour area is of approximately equal priority in Figure 5.24 with respect to DO Saturation % monitoring with the exception of the upper end of the lower Lee Estuary between monitoring points LE150 and LE160.

Figure 5.22 Priority of potential monitoring locations for DIN in brackish waters
Optimisation of water quality monitoring programmes via Entropy methods

Figure 5.23 Priority of potential monitoring locations for MRP in brackish waters

Figure 5.24 Priority of potential monitoring locations for DO sat% in brackish waters
5.5.1.3. Seawater

The fully saline coastal waters which extend from the mouth of Cork Harbour towards the tidal boundary contain Cork Harbour outer as described under the WFD. There is a clear pattern of highest priority sampling for Chlorophyll_a, MRP and DIN to the east of Roches’ point, with a gradient of lower priority monitoring towards Minane in Figures 5.25 to 5.27. The DO Saturation % monitoring entropy rank is much more dispersed as seen in Figure 5.28.

Figure 5.25 Priority of potential monitoring locations for Chlorophyll_a in seawater
Figure 5.26 Priority of potential monitoring locations for DIN in seawater

Figure 5.27 Priority of potential monitoring locations for MRP in seawater
5.5.2. Monitoring design by TSAS compliance level

The TSAS compliance criteria in Table 5.1 were applied to the timeseries data at each of the wet grid cells in each salinity class to determine the compliance level before determining the monitoring priority for each compliance level in each salinity class. The range in priority values apportioned to different areas in this section is distinct from the previous section due to the restriction of the number of grid cells to be prioritised. The pool of points to be chosen from is narrower due to prior stratification of regions based on salinity class and trophic status. The entire freshwater area of the Owenbuidhe Estuary was classified as eutrophic and due to the lack of different trophic risk levels in the area, the entropy rank for the four TSAS parameters were the same as in section 5.5.1.1.
5.5.2.1. Intermediate waters

5.5.2.1.1. Eutrophic

Eutrophic waters in the intermediate salinity class in the greater Cork Harbour area extend to cover the lower end of lower Lee Estuary, Glashaboy Estuary, Lough Mahon, Harper’s Island, majority of Cork Harbour inner and the Owenacurra Estuary.

The region in the vicinity of Slatty waters in Harper’s Island is highlighted as a distinct monitoring priority location in Figures 5.29 to 5.31 for Chlorophyll_a, MRP and DIN whilst the general Harper’s Island area is a monitoring priority highlighted in Figure 5.32 for DO Saturation %. This region was assessed under the TSAS assessment scheme for the period spanning from 2002 to 2006 but was not assessed prior to that period. In the first assessment period the waters were assessed as eutrophic (see Table 5.3) as is the case here with the application of the TSAS assessment criteria to the three year numerical model simulation data.

Both Lough Mahon and the south end of Cork Harbour inner differ in Figure 5.29 with Lough Mahon appointed the highest priority. A gradient can be observed from the east end of the North Channel towards the Owenacurra Estuary as is the case with MRP and DIN in Figures 5.30 and 5.31.

Regarding prioritisation of DIN monitoring, Lough Mahon, the lower Lee Estuary, the Glashaboy Estuary and the majority of Cork Harbour inner are all ranked the highest. There is a steep gradient of monitoring priority in the Owenacurra Estuary from the North Channel at LE450 towards LE500.

In Figure 5.31, a higher ranked zone is widespread throughout Lough Mahon and North of Cork Harbour. A gradient is observed in the direction of Harper’s Island and the Owenacurra Estuary. Although Harper’s Island is monitored at point LE350, the darker shaded region between LE340 and LE350 remains unmonitored although it is distinct from its neighbouring waters.

There are fewer grades of priority for DO saturation monitoring in the eutrophic intermediate salinity waters in Figure 5.32. The area adjacent to Harper’s Island is unique as compared to the rest of the domain as the whole area is entirely 2nd and
3rd grade monitoring priority. Although the east end of Harper’s Island adjacent to Slatty waters is a priority zone for nutrients and Chlorophyll \_a, the west end of Harper’s Island around monitoring point LE350 is a priority for DO saturation monitoring.

Figure 5.29 Priority of potential monitoring locations for Chlorophyll \_a in brackish eutrophic waters
Figure 5.30 Ranking of potential monitoring locations for DIN in brackish eutrophic waters

Figure 5.31 Priority of potential monitoring locations for MRP in brackish eutrophic waters
5.5.2.1.2. Potentially eutrophic

The potentially eutrophic intermediate salinity waters cover the North Channel and the perimeter around Whitegate/Aghada as can be seen in Figures 5.33 to 5.36.

There are 3 clear bands of monitoring priority for Chlorophyll_a in the North Channel as seen in Figure 5.33. One zone is unsampled at present between monitoring points LE420 and LE430. Similarly in the case of DIN in Figure 5.34, there is an unmonitored stretch between LE430 and LE450 which requires sampling. Monitoring points LE420 and LE430 both lie within the highest priority ranking for DIN and thus there is a redundancy between both locations for DIN monitoring. The same reach is unmonitored for MRP although it is the highest priority region for MRP monitoring. The gradient of monitoring priority is unique for Chlorophyll_a, MRP and DIN in the North Channel; the existing monitoring network insufficiently covers the gradients for each of the 3 TSAS parameters in the North Channel and the location could be moved eastwards or an additional monitoring point added between LE430 and LE450. DO Saturation % monitoring in the North Channel is well
monitored at present with both the first and second priority zones that are highlighted in Figure 5.36 sufficiently covered.

Figure 5.33 Ranking of potential monitoring locations for Chlorophyll_a in brackish potentially eutrophic waters

Figure 5.34 Ranking of potential monitoring locations for DIN in brackish potentially eutrophic waters
Figure 5.35 Priority of potential monitoring locations for MRP in brackish potentially eutrophic waters

Figure 5.36 Ranking of potential monitoring locations for DO sat\% in brackish potentially eutrophic waters
5.5.2.1.3. Intermediate trophic status

This region covers the Whitegate Bay area and the upper end of the Lower Lee Estuary. Chlorophyll_a and DO Saturation % monitoring priority is broadly dispersed as can be seen in Figures 5.37 and 5.40.

The lowest ranked monitoring zone is near Whitegate in the case of DIN and MRP in Figures 5.37 and 5.38 respectively, whilst the highest priority monitoring is sufficiently covered by waters south-southwest of LE610 and LE620 for DIN and MRP.

![Map of monitoring locations](image)

*Figure 5.37 Priority of potential monitoring locations for Chlorophyll_a in brackish intermediate risk waters*
Figure 5.38 Priority of potential monitoring locations for DIN in brackish intermediate risk waters

Figure 5.39 Priority of potential monitoring locations for MRP in brackish intermediate risk waters
5.5.2.2. Full salinity waters

5.5.2.2.1. Eutrophic

The eutrophic region within the full salinity waters extend from the mouth of Cork Harbour towards the south-south east open sea boundary.

The highest priority region for Chlorophyll_a monitoring lies east of Roche’s point as presented in Figure 5.41. The location for DIN sampling highlighted in Figure 5.42 is irrelevant in eutrophic waters as all grid cells are of equal ranking. MRP sampling priority in Figure 5.43 is approximately the same as DIN. DO saturation sampling priority is largely dispersed in Figure 5.44 with the highest priority at Roche’s point.
Figure 5.41 Priority of potential monitoring locations for Chlorophyll_a in eutrophic seawaters

Figure 5.42 Priority of potential monitoring locations for DIN in eutrophic seawaters
Figure 5.43 Priority of potential monitoring locations for MRP in eutrophic seawaters

Figure 5.44 Priority of potential monitoring locations for DO sat% in eutrophic seawaters
5.5.2.2.2. Potentially eutrophic

Potentially eutrophic waters extend from south of Roche’s point towards the Minane area, as can be seen in Figures 5.45 to 5.48. The highest priority monitoring zones for Chlorophyll_a in Figure 5.45 lie in the region either side of point LE820. The region to the east of monitoring point LE820 is the highest priority monitoring zone for DIN in Figure 5.46. The region centred on monitoring point LE820 is the highest priority monitoring zone for MRP in Figure 5.47, whilst the monitoring priority for DO Saturation % sampling is broadly dispersed and inconclusive in Figure 5.48.

Figure 5.45 Priority of potential monitoring locations for Chlorophyll_a in potentially eutrophic seawaters
Figure 5.46 Priority of potential monitoring locations for DIN in potentially eutrophic seawaters

Figure 5.47 Priority of potential monitoring locations for MRP in potentially eutrophic seawaters
5.5.2.2.3. Intermediate trophic status

Intermediate trophic status full salinity waters extend towards Minane from monitoring point LE820. Monitoring priority ranks for nutrients and Chlorophyll_a follow a gradient towards Minane as can be seen in Figures 5.49 to 5.52, with the highest rank at the seaward end of the zone. As observed in the case of each other trophic status in the fully saline waters, there is no conclusive pattern to monitoring priority for DO Saturation % saturation in this region.
Figure 5.49 Priority of potential monitoring locations for Chlorophyll_a in intermediate risk eutrophic seawaters

Figure 5.50 Priority of potential monitoring locations for DIN in intermediate risk eutrophic seawaters
Figure 5.51 Priority of potential monitoring locations for MRP in intermediate risk eutrophic seawaters

Figure 5.52 Priority of potential monitoring locations for DO sat% in intermediate risk eutrophic seawaters
5.6. Conclusions

Surface water quality monitoring may be carried out with the basic requirement of establishing average conditions for assessment of water quality status although other reasons for monitoring include detecting temporal trends or spatial patterns in water quality, evaluating the success of management decisions or investigating the cause of non-compliance. Thus water quality monitoring and numerical modelling of water quality are clearly complementary in terms of their uses and their end goals. Locations for monitoring in Irish waters have previously been chosen based on expert opinion to find a representative location (EPA 2006). There is no strict methodology outside of expert knowledge for sampling at the selected locations. The mandatory minimum sampling frequency set out in the WFD may be overruled based on expert knowledge but no prescriptive procedure exists to determine what frequency should be selected. A large proportion of monitoring guidance is aimed towards monitoring within the river basin network in lakes, streams and rivers, with little reference to the highly dynamic estuarine environment which is the ultimate sink for watershed nutrients. The work presented here has utilised a calibrated and validated two-dimensional depth integrated hydrodynamic, solute transport and water quality model which is capable of fully incorporating the aforementioned factors. Although a full range of water quality parameters were available within the numerical model, only the parameters included in TSAS assessment were considered in the analysis, namely MRP, DIN, Chlorophyll_a and Dissolved Oxygen.

A comparison of the marginal entropy derived from EPA monitoring data versus model simulated water quality data at each of the 24 monitoring points in the model domain yielded some interesting insights. Marginal entropy is proportional to the variance or standard deviation of the underlying dataset. By virtue of the fact that EPA water quality data was sampled discretely four times annually whereas the model data was a continuous timeseries, one would expect the model marginal entropy to be greater than or equal to the EPA monitoring data entropy at each monitoring point; this was the pattern observed in the case of Chlorophyll_a and MRP. In the case of DIN, at a number of points in the North Channel and Owenacurra Estuary, EPA monitoring data marginal entropy exceeded model marginal entropy. There may be a number of causes for this discrepancy such as
insufficient or inaccurate data on inflows from the River Owenacurra and the associated catchment loads or the wastewater discharges to the area. These results nonetheless emphasise the diagnostic features of entropy analysis whereby inconsistencies between model and data could suggest errors in data or model inputs.

The 24 EPA monitoring stations within the greater Cork Harbour waters were considered for prioritisation, in order to determine which monitoring stations inform the end user of the greatest amount of information about the water quality parameter in question. With knowledge of the ultimate monitoring network transinformation were all monitoring points to be used, the required network size to achieve 25%, 50%, 75% or 100% of the total transinformation was determined for each parameter and sampling resolution. The rank of each monitoring station was used as a basis for determining the optimum sampling frequency for monitoring as the reconfigured monitoring scheme expands from 2 to 24 monitoring points.

Weekly and monthly sampling resolution was considered. The optimum frequency at weekly or monthly resolution yielded the same TSAS trophic status as the existing monitoring regime for all waterbodies. In all waterbodies except the Lee and Glashaboy estuaries, the optimised monitoring network and sampling regime that provide 25% through to 100% of total system transinformation returned the same TSAS status whether a weekly or monthly resolution was considered. The lower Lee Estuary, the Glashaboy Estuary, Lough Mahon, the North Channel and Cork Harbour outer returned the same TSAS status with the optimised monthly resolution regimes as the existing monitoring programme. The existing monitoring regime, when applied to sample model simulated nutrient, DO and Chlorophyll_a data for TSAS compliance over a 3 year period returned identical status values for Lough Mahon, the Owenacurra Estuary, and Cork Harbour outer as the EPA reporting for the pre-Carrigrennan WWTP 1999-2003 period.

The comparison between these results should be cautioned as the EPA reporting period in that instance was 5 years as opposed to the 3 year period which is usually used for TSAS assessment reporting and thus was used in replicating the existing monitoring regime with numerical model simulated data. Also, reporting in the '99-'03 period returned trophic status for Cork Harbour as a whole, thus comparison between the existing monitoring regime and the reporting period should be
interpreted cautiously.

The spatial optimisation analysis which considered each of the 90,233 wet 30 x 30m grid cells as potential monitoring locations yielded some interesting insights into the monitoring requirements prior to commissioning of Carrigrennan WWTP. Depending upon the water quality parameter under consideration, extra monitoring points were suggested at locations in the North Channel, Harper’s Island, East Ferry, Whitegate Bay and Cork Harbour outer. The entire model domain fell under at least intermediate TSAS status when the TSAS criteria were applied on a cell by cell basis although the limit values are intended to be applied to areal averages from all of the monitoring points in each waterbody. The author suggests that an alternative limit value could be applied to each cell in determining an optimum monitoring network, such as a long term 95 percentile being exceeded for each water quality parameter. With a suitable limit value adopted, the optimal monitoring network would identify regions with persistent issues for each respective water quality parameter. The 95th percentile limit values would decrease with increasing salinity as is the case at present with the current TSAS system.

The results presented here suggest a number of potential uses for the proposed methodology. In previously unmonitored waters, all locations may initially be considered for monitoring and ranked based on marginal entropy scores. A calibrated and validated water quality model was utilized in generating high resolution time series at 24 monitoring locations in the Cork Harbour area. Spatial comparison of marginal entropy and transinformation entropy prioritized the locations and water reaches which were non-compliant areas of concern in water quality reporting in the periods from 1995 to 1999 and 1999 to 2003. The proposed methodology could potentially be used to determine the optimum location and frequency for monitoring under different water quality management scenarios such as increased discharge of point source pollution from wastewater treatment plants or a reduction in the riverine nutrient loading. Scaled down monitoring networks to capture 25% through to 100% of the total monitoring network transinformation have been presented.
To summarise, the main findings of this chapter are as follows:

- Optimal expansion and contraction of the trophic status monitoring network was investigated using Cork Harbour and environs as a case study.
- At a preliminary stage, the marginal entropy of the PDFs of monitoring data and model simulation data were compared to establish how closely the current monitoring regime would capture the underlying water quality characteristics. The underlying water quality characteristics were assumed to be summarised by the equivalent PDF derived from water quality data generated by the water quality model. The model MRP and Chlorophyll_a marginal entropy consistently exceeded monitoring data marginal entropy with the largest difference at monitoring points in well flushed waters with a short residence time and the smallest difference in the poorly flushed waters in the North Channel. Where numerical model marginal entropy was lower than the EPA monitoring data marginal entropy in the case of DIN, the author suggests that this may have been due to inaccurate DIN loadings from the Owenacurra Estuary or wastewater effluent discharges in the North Channel. The exercise could prove beneficial in an iterative cycle of model recalibration and modification of numerical model inputs in light of ongoing EPA water quality monitoring.
- The proposed methodology is adaptive to other legislation and directives.
- Contraction of the monitoring network had a negligible effect on outcome of water compliance assessment. Of the 7 waterbodies in the greater Cork harbour area, 4 returned the same level of compliance irrespective of the level of contraction. Of the remaining 3, 2 waterbodies (Glashaboy Estuary and Cork harbour Inner) achieved the same level of compliance with the 50% TST monitoring network or greater, as would be attained with the existing monitoring network. The Owenacurra Estuary required the existing monitoring points to be retained to achieve the same level of compliance, with any contraction resulting in an apparent improvement in status from eutrophic to intermediate.
- The existing monthly sampling resolution and weekly resolution were
considered in the process of sampling frequency optimisation; the process was a progression from the spatial monitoring optimisation. For each spatially optimised monitoring network, an optimum multiple of weekly or monthly resolution was derived. There was no benefit from sampling at weekly frequency or multiples thereof for compliance purposes.

- The spread of monitoring priority varies depending upon the water quality parameter of concern, and thus, using the same monitoring network for each water quality parameter to determine trophic status may be inefficient and unnecessary.

- The monitoring network used by the EPA was to a great extent replicated by the high resolution monitoring design procedure. The locations which were prioritised for monitoring each water quality parameter were areas of concern in national reporting.

The opportunity to redesign a monitoring network at regular intervals, by scaling down or expanding the monitoring network, or reallocating monitoring points, introduces the prospect of monitoring data discontinuity due to cessation of sampling at certain monitoring locations, whilst a greater volume of monitoring data may be made available elsewhere if the monitoring network is expanded and redesigned at the same time. A valid, robust estimate is desirable where data discontinuities occur, in order to counteract the temporary shortage of monitoring data. The following chapter considers the range of options available to estimate monitoring data at redundant monitoring points and the trophic status assessment outcomes derived from using various estimates to represent redundant monitoring points.
Chapter 6. BME and geospatial estimation

6.1. Introduction

Monitoring programme optimisation, as discussed in the previous chapter, either contraction of the monitoring network or relocation of monitoring points within a network, will lead to discontinuity at certain locations and uncertainty as to the validity of water quality assessments over the assessment period.

In Chapter 6 it was concluded that a reduction in the number of monitoring points in Cork Harbour would have a negligible influence on the outcome of water quality compliance assessment on the basis of fewer monitoring points or a change in sampling frequency; this chapter considers the veracity of estimating the concentration of the water quality parameters included under the EPA TSAS scheme by utilising BME to combine numerical model soft data and EPA water quality monitoring data at retained monitoring points to estimate the water quality parameters at the removed monitoring points.

Estimation of water quality in scaled back monitoring networks is carried out on the basis of covariance models fitted to monitoring data from the retained monitoring points within the network. A separate covariance model is derived for each of the 7 water bodies in Cork Harbour defined for the purposes of WFD monitoring and assessment.

Estimation of water quality parameters at data gaps is then carried out by assuming quasi-stationarity (Journel and Huijbregts 1978); covariograms are estimated centred on each monitoring point within a moving window, thus covariograms are unique to the neighbourhood around the estimation location. A similar approach cannot be taken when estimating at redundant monitoring points as in theory the point no longer exists for monitoring purposes and the covariogram should reflect the underlying dataset.
In most previous applications of BME to geospatial datasets, stationarity was assumed; whilst a number of publications assume stationarity on a subscale (Christakos et al. 2004, Li et al. 2013), which is assumed in this chapter when estimating monitoring data at redundant monitoring points. Other publications have assumed quasi-stationarity of covariance centred on each estimation location which proved highly successful in each case (Akita et al. 2012). To investigate likelihood of quasi-stationarity, quasi-stationary covariance models are used in the leave-one-out cross validation of BME and kriging as a proposed improvement on standard BME estimation.

6.2. EPA monitoring data – ‘Hard Data’

Cork Harbour and environs as defined in this study includes the lower Lee Estuary, Lough Mahon, North Channel Great Island, the Owenacurra Estuary, Glashaboy Estuary, Harpers Island and Cork Harbour inner and outer as shown in Figure 5.1. In total, Cork Harbour covers approximately 81km² and is monitored by 24 monitoring points, resulting in an average density of monitoring points of 0.3 monitoring points per km² or each monitoring point on average describes the surrounding 3.38km².

The EPA monitoring data which forms the basis of this study was collected from 1994 onwards with coverage varying in the intervening years. For the purposes of this study, only the data spanning from 1994 to 2003 was included, corresponding to the model validation period.

Water quality sampling was carried out in accordance with the procedures set out in O'Boyle et al. (2015) and is summarised as follows:

- Sampling was carried out once in winter and three times in summer
- Samples were collected at high and low tide to capture the influence of tidal variation
- Chlorophyll_a, nutrients and biochemical oxygen demand were sampled 0.5 m above the seabed using a 2 L Hydrobios Ruttner bottle and surface sampling was undertaken using a plastic bucket.
• Chlorophyll_a samples were filtered using 1.2μm Whatman GF/C filters. Chlorophyll_a pigments following the procedures set out in Standing Committee of Analysts (1980).

• Total ammonia, total oxidized nitrogen and MRP were measured following American Public Health Association (2005) guidelines.

For the purposes of this research, data collected on each day of monitoring at each monitoring point at high and low tide were averaged to give mean daily values. DO Saturation % was not considered here due to a shortage of monitoring information.

BME does not limit the underlying distribution of data. Nonetheless, the analysis that follows in this chapter will consider applying simple kriging and BME to both standard and log-transformed datasets to establish the suitability of adopting log transformed datasets for each of the water quality datasets in question.

The main steps involved in deriving BME estimates are as follows (see Figure 4.5):

1. Log transformation is applied to the dataset where necessary.

2. The spatial and temporal mean trend are determined and subsequently removed from the dataset (section 6.2.1). A homogenous and stationary trend is necessary to allow covariance values to be expressed in terms of spatial and temporal lag but not spatial and temporal coordinate. This condition is indicated by a normal residual upon removal of the spatiotemporal mean trend from data.

3. Raw covariance values are calculated by applying Equation (4.5) to the residual data after removal of the mean trend. The range of spatial and temporal lags between all monitoring data points are determined and subsequently divided into subgroupings based on spatial and temporal lags. Spatial and temporal lag classes are approximated to regular intervals up to the maximum lag values. A tolerance of half the regular interval is used either side of each lag class such that no data point is unused. For each combination of spatial and temporal lag, the raw covariance is determined by summing the product of the residual data for each pair of data points in a subgrouping. The average of the sum of the product is the average raw covariance for the respective lag.
4. A covariance model is fitted to the raw covariance data. The covariance model fitted for complex spatiotemporal datasets would feature at least one nonseparable covariance model or alternatively a primary and secondary covariance model. In such cases where two models are used, the primary covariance model describes the spatial and temporal dissipation of the majority of the variance while the secondary model describes minor far field dissipation of covariance.

5. Soft data is derived as outlined in section 6.3. The spatiotemporal mean trend is linearly interpolated to each soft datum and the mean trend is removed from that data point.

6. BME estimators are generated using the fitted covariance model to derive the matrix describing covariance between all hard data points, all soft data points and between all hard and soft data points.

7. Application of BME estimation to log-transformed datasets requires the data to be natural log-transformed at the outset, with back transformation by deriving the exponent of the ultimate BME posterior probability density function. Back transformation of lognormal posterior BME mean estimate yields the median estimator.

Table 6.1 summarises the normality of each of the three water quality datasets when considered in terms of skewness and kurtosis. Positive skewness indicates that the histogram or probability density function of the dataset has an asymmetrical distribution with a longer tail to the right hand side which would be a common feature of water quality data where extreme values occur on occasion. Kurtosis provides a measure of the nature of the peak of the dataset with values greater than 3 indicating a sharp peak while datasets with a broader flatter distribution would be indicated by values less than 3. Comparing the three datasets in standard form and after log transformation in terms of skewness and kurtosis values in Table 6.1, and observing the distribution shapes in Figures 6.1-6.3, it can be seen that log transformation yields the greatest improvement in the case of chorophyll a. Log transformation of MRP and DIN shifts the datasets from right skewed to slightly left skewed with flatter shapes.
Table 6.1 Comparison of skewness and kurtosis for water quality parameters before and after log-transformation

<table>
<thead>
<tr>
<th>Data</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll a</td>
<td>3.48</td>
<td>22.45</td>
</tr>
<tr>
<td>Log-Chlorophyll a</td>
<td>0.25</td>
<td>3.13</td>
</tr>
<tr>
<td>MRP</td>
<td>1.58</td>
<td>5.62</td>
</tr>
<tr>
<td>Log-MRP</td>
<td>-0.27</td>
<td>2.29</td>
</tr>
<tr>
<td>DIN</td>
<td>1.72</td>
<td>5.17</td>
</tr>
<tr>
<td>Log-DIN</td>
<td>-0.18</td>
<td>1.88</td>
</tr>
<tr>
<td>Reference value</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 6.1 Histograms of Chlorophyll_a data and log transformed Chlorophyll_a data
Figure 6.2 Histograms of MRP data and log transformed MRP data

Figure 6.3 Histograms of DIN data and log transformed DIN data
6.2.1. Spatial and temporal data trends

Raw spatial and temporal mean trend values are determined by averaging over the spatial and temporal search radii respectively. To determine the spatial mean trend at a monitoring point, all monitoring data within the spatial search radius is averaged and the temporal mean trend is captured by averaging all monitoring data throughout the monitoring network within the temporal search radii. The final spatial and temporal trend values are determined by applying an exponential smoothing filter to raw spatial and temporal average values over spatial and temporal smoothing radii such as presented in Equation 4.2. Nazelle et al. (2010) suggests that the search and smoothing radii should be selected such that the resultant mean trend is physically meaningful whereby all of the obvious trends in the raw data are approximated by the trend data that is extracted.

Figure 6.4 presents a cyclical trend in Chlorophyll_a values with the most notable peak occurring in summer ’97 in Lough Mahon and the adjacent waters in the inner harbour. In later years, a downward trend is observed in Chlorophyll_a values in Lough Mahon with an increasing trend in the North Channel and the adjacent Owenacurra Estuary.

Peak MRP values coincide with summertime in Figure 6.5, with the highest values occurring in ’95 in the Lee and North Channel followed by a downward trend thereafter. The peak observed in ’95 in the Lee Estuary is not matched with a similar peak in the downstream waters of Lough Mahon which indicates MRP consumption by primary production as seen in Figure 6.4. There is no notable MRP trend elsewhere in the extended Cork Harbour area thereafter.

The Owenacurra Estuary features the highest observed trend in DIN values in Figure 6.6 in summer ’96 and ’98 with a downward trend observed thereafter. DIN values in Lough Mahon peak in summer ’01. Otherwise, only the expected cyclical fluctuations are observed elsewhere in Cork Harbour.

The spatial mean trends are presented in Figures 6.7, 6.8 and 6.9. The maps present the average of each water quality parameter for the duration of monitoring data considered. The gradual variation in the spatial trend is a result of applying an exponential spatial smoothing filter. Two areas are highlighted in the spatial mean
trend plots. Chlorophyll\textsubscript{a} and DIN are consistently high in the Owenacurra Estuary while Chlorophyll\textsubscript{a} and MRP are consistently high in the lower Lee Estuary.

Figure 6.4 Temporal trend of standard and log transformed Chlorophyll\textsubscript{a} within each waterbody and across Cork Harbour
Figure 6.5 Temporal trend of standard and log transformed MRP within each waterbody and across Cork Harbour

Figure 6.6 Temporal trend of standard and log transformed DIN within each waterbody and across Cork Harbour
Figure 6.7 Spatial trend of Chlorophyll \_a throughout Cork Harbour

Figure 6.8 Spatial trend of MRP throughout Cork Harbour
Figure 6.9 Spatial trend of DIN throughout Cork Harbour

The spatial and temporal search and smoothing radii were manually modified to capture the most salient spatial and temporal trends in water quality data over the 8 years in question. The residual datasets that remain after removal of the spatial and temporal trend are summarised by the histograms in Figures 6.10 to 6.12; the associated skewness and kurtosis characteristics are summarised in Table 6.2. The optimised monitoring networks derived and presented in Chapter 6 are used in this chapter to assess the potential to adopt a downscaled monitoring in Cork Harbour with different sampling regimes adopted for each parameter.
The log residual refers to the data that remains upon removal of the mean trend of the log transformed data from the log transformed data. Referring to Table 6.2 and Figure 6.10, log transformation of the Chlorophyll_a dataset in each of the three optimised monitoring networks results in the greatest improvement in normality of the residual histogram, and in particular between Figures 6.10(b) and (e), and Figures 6.10(c) and (f). The values for skewness and kurtosis of log transformed Chlorophyll_a in Table 6.2 corroborate the apparent improvement in normality in Figure 6.10. The same cannot be said for MRP in Figure 6.11, where log transformation results in a switch from left to right skewed in the case of the smallest monitoring network in Figures 6.11(a) and (d), while the transformation has the opposite effect in the case of the two larger monitoring networks in Figures 6.11(b) and (e) and Figures 6.11 (c) and (f), although kurtosis is markedly reduced in each case. Log transformation of DIN as presented in Figure 6.12 results in residual histograms characterised by skewness and kurtosis values of 0-0.1 and 3.38-3.64 respectively, which are close to the ideal values of 0 and 3 respectively.

<table>
<thead>
<tr>
<th>% TST</th>
<th>Chl a</th>
<th>Log Chl a</th>
<th>MRP</th>
<th>Log MRP</th>
<th>DIN</th>
<th>Log DIN</th>
<th>Ref. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.54</td>
<td>0.24</td>
<td>-0.03</td>
<td>0.64</td>
<td>0.55</td>
<td>0.10</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>4.05</td>
<td>0.05</td>
<td>1.15</td>
<td>-0.69</td>
<td>1.90</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>4.11</td>
<td>0.06</td>
<td>1.14</td>
<td>-0.37</td>
<td>1.79</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>4.25</td>
<td>2.72</td>
<td>3.09</td>
<td>3.59</td>
<td>6.14</td>
<td>3.38</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>37.56</td>
<td>2.94</td>
<td>8.02</td>
<td>4.04</td>
<td>6.74</td>
<td>3.66</td>
<td>3</td>
</tr>
<tr>
<td>75</td>
<td>38.75</td>
<td>2.98</td>
<td>10.37</td>
<td>3.16</td>
<td>5.82</td>
<td>3.64</td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.10 Histograms of residual data after removal of the spatial and temporal mean trend data for standard and log transformed Chlorophyll_a

Figure 6.11 Histograms of residual data after removal of the spatial and temporal mean trend data for standard and log transformed MRP
Figure 6.12 Histograms of residual data after removal of the spatial and temporal mean trend data for standard and log transformed DIN

Figure 6.13 provides a general example of the form and characteristics of three single covariance models; the sill, which refers to the covariance value at zero spatial and temporal lag, increases with the increasing size of the underlying dataset and thus the increasing inherent variation in the underlying data, while the range refers to the full extent of the influence of autocorrelation between any 2 points.
Figures 6.14 and 6.15 present the covariance models that were fitted to raw covariance data from the standard MRP data and log transformed Chlorophyll_a data from their respective 25% and 75% TST monitoring networks and Figure 6.15 is also distinct as it presents the covariance models for the residuals of a log transformed dataset. The sill is much lower for all waterbodies due to the smaller residual values. Also, there is greater consistency in the shape of covariance models for adjacent waterbodies for the covariance of log transformed data. Note that the sill for the covariance model describing all of Cork Harbour has the highest sill as it must describe the greatest volume of data. The spatial range indicates the area represented by a single measurement while the temporal range represents the length of time after which zero correlation occurs.
Figure 6.14 Covariance models derived from the residual data from optimised MRP monitoring network with the lowest TST of 25%

Figure 6.15 Covariance models derived from the residual data from the log-transformed Chlorophyll_a monitoring data the optimised monitoring network with the TST of 75%
From Figures 6.14 and 6.15, it can be seen that adjacent waterbodies such as the Lee Estuary and Lough Mahon or the inner and outer Cork Harbour area possess unique covariance patterns. The spatial covariance model for the Lee estuary and Lough Mahon in Figure 6.14 are similar in shape but differing in proportions. The Lee estuary has a shorter range but a higher sill, indicating that there is greater variation in MRP within the Lee estuary when compared to Lough Mahon.

The covariance models that were considered included any combination of exponential, gaussian, spherical and cosine. In each case, an effort was made by least squares error to fit two nonseparable models describing the primary and secondary covariance. In the case of log transformed MRP, for monitoring data captured by the 25, 50 and 75% TST monitoring network, only a primary model was fitted which was spherical in space and exponential in time as in Equation (6.1) with parameters defined as in Table 6.3.

\[
C_{r,t} = C \left[ 1 - \frac{3r}{2a_r} + \frac{1}{2} \left( \frac{r}{a_r} \right)^3 e^{-\frac{3t}{a_t}} \right]
\]  

(6.1)

where \(C_{r,t}\) refers to the covariance value at spatial and temporal lags of \(r\) and \(t\) respectively, \(a_r\) is the spatial range and \(a_t\) is the temporal range and \(C\) is the sill, as presented in Figure 6.14.

Table 6.3 Covariance model parameters for log transformed MRP for each of the 3 monitoring networks

<table>
<thead>
<tr>
<th>MRP</th>
<th>(C)</th>
<th>(a_r)</th>
<th>(a_t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.011</td>
<td>2462</td>
<td>42.671</td>
</tr>
<tr>
<td>50</td>
<td>0.022</td>
<td>2462</td>
<td>49.353</td>
</tr>
<tr>
<td>75</td>
<td>0.022</td>
<td>2462</td>
<td>49.353</td>
</tr>
</tbody>
</table>
6.3. Derivation of soft data

In the context of water quality monitoring, BME possesses the potential to combine measured and probabilistic water quality data to estimate posterior PDFs of the likely range of values for any water quality parameters.

To support the use of BME for geostatistical estimation of water quality in Cork Harbour, the dataset of simulated water quality outputs generated by the model must be expressed in terms of PDFs of the associated monitoring data for the parameter in question.

The model was run for the duration of coverage of the monitoring data collected prior to the commissioning of Carrigrennan wastewater treatment plant from June 1994 to October 2002. Timeseries were output at each of the monitoring points in Cork Harbour and each ‘hard’ monitoring datum was matched with the equivalent water quality data simulation for ammoniacal and nitrate nitrogen, orthophosphate and Chlorophyll_a.

As a consequence of considering EPA monitoring data both in its standard format and after log transformation, the derivation of probabilistic soft data for BME from simulated water quality data was considered on both scales. Figures 6.16 – 6.18 present the scatter plots of collocated monitoring and modelled data in standard format for Chlorophyll_a, MRP and DIN respectively. The figures indicate that the majority of monitoring data in the period was clustered to the lower end of the range of values; a similar pattern was observed in the dispersal of simulated data.

Figures 6.19 – 6.21 present the scatter plots of collocated monitoring and modelled data after log transformation of simulated numerical model water quality data and monitoring data provided by the EPA for Chlorophyll_a, MRP and DIN respectively. Log transformation results in a more homogenous spread of data points in the regression plots in Figures 6.19 – 6.21, in particular in the case of MRP in Figure 6.21. The distributions of collocated monitoring and model data are both right skewed with a long tale of extreme values in Figures 6.16 – 6.18, thus precluding the possibility of a representative linear regression of data in its standard format. Figures 6.19 – 6.21 present log transformed data for which linear regression is much easier to
implement. The monitoring data are approximately normally distributed vertically across the linear regression line.

A number of approaches have been adopted in deriving probabilistic soft data for BME; in the work presented here, soft data was derived for data in standard format after Christakos et al. (2004), while soft data derived on the log scale followed the approach of Money et al. (2009).

In the case of data in standard format, the range of model observations in each case was divided into neighbouring intervals. Within each interval, the mean and standard deviation of monitoring observations was determined and thus for each model output, a corresponding mean and standard deviation could be appointed.

In the case of log transformed data, a linear regression model of the log transformed model observations against log transformed water quality data was derived with the associated equation describing the expected value of the log transformed monitoring data, with Chlorophyll\textsubscript{a}, MRP and DIN described by:

\[
\begin{align*}
\ln(\text{Chl}_a)_{\text{EPA}} &= 2.19 [\ln(\text{Chl}_a)_{\text{MODEL}}] + 0.027 \\
\ln(\text{MRP})_{\text{EPA}} &= 2.47 [\ln(\text{MRP})_{\text{MODEL}}] + 0.23 \\
\ln(\text{DIN})_{\text{EPA}} &= -2.12 [\ln(\text{DIN})_{\text{MODEL}}] + 0.55
\end{align*}
\]

Consequently, for each model observation, the linear regression model was used to determine the expected value of the collocated log transformed monitoring data. The error between the regression model estimate and the actual observation was used in each case to determine the standard error of estimation for adjoining intervals. The same range of intervals was used for standard and log transformed data. The intervals were chosen from 0 to 2000 in the relevant units, with higher resolution closer to 0 because of the higher amount of data points in the lower end of the data range.
Figure 6.16 Chlorophyll-a monitoring data versus model Chlorophyll-a data

Figure 6.17 MRP monitoring data versus model MRP data
Figure 6.18 DIN monitoring data versus model DIN data

Figure 6.19 Log transformed Chlorophyll_a monitoring data versus log transformed model Chlorophyll_a data
Figure 6.20 Log transformed MRP monitoring data versus log transformed model MRP data

Figure 6.21 Log transformed DIN monitoring data versus log transformed model DIN data
6.4. BME and kriging estimation

BME is innovative due to its ability to process probabilistic “soft” data in the same manner as certain “hard” data which is perceived as having a single value with probability of 1. The mathematical expressions for the BME mean estimate and variance estimate reduce to their kriging equivalents when only “hard” data is used. Therefore, in the following section both BME and kriging will be compared with respect to their abilities to estimate water quality data at the monitoring points which were deemed surplus to requirements in Chapter 5. For each parameter, three monitoring networks are proposed with each network containing a selection of the existing monitoring points. Depending on the parameter in question, different locations are identified. The network with the lowest spatial coverage is the network with 25% total monitoring station transinformation, with the coverage of the 50% and 75% network increasing gradually towards the maximum number of monitoring points used in the monitoring network at present.

BME and kriging are applied to both standard data and log transformed data. As mentioned previously, monitoring was carried out at each monitoring point near high and low tide. All estimates made within this chapter are estimates of daily averages. Estimation was carried out following the summary of the procedure outlined in section 6.2. Spatial and temporal interpolation of the mean trend values to the estimation points required a minimum of 3 points; in some cases there were insufficient monitoring points within the waterbody to interpolate to the estimation location. For each water quality parameter and monitoring network size, BME and kriging estimates were generated using two covariance model types: one model derived from all monitoring data throughout Cork Harbour and a second model derived using monitoring data which was specific to each waterbody. Due to a shortage of monitoring points for the 25% and 50% TST monitoring networks, waterbody covariance models could not be derived and so discussion here focuses on the ability of BME and kriging to estimate water quality data from redundant monitoring points using a single covariance model derived for the whole of Cork Harbour.
The following section investigates the feasibility of utilising BME or kriging to estimate water quality parameters at redundant monitoring points. Covariance is considered assuming covariance is stationary within Cork Harbour and alternatively stationary within each water body in Cork Harbour. Later in this chapter, the veracity of assuming quasi-stationarity is assessed for each water quality parameter by completing leave-one-out cross validation with separate covariance models derived for each monitoring point using either all data points from Cork Harbour or specifically monitoring data from the water body in question.

6.4.1. Standard estimation

Standard BME and kriging estimates were generated at each redundant monitoring point that was removed from the existing monitoring network to derive the 25, 50 and 75% TST monitoring networks for Chlorophyll\_a, MRP and DIN.

Data limitations induced by reducing the number of monitoring points in the 25 and 50% TST optimised monitoring networks precluded the possibility of either deriving covariance models to describe each waterbody or interpolating the spatial mean trend to estimation points. Consequently, Chlorophyll\_a was the only parameter for which BME and kriging could be carried out at waterbody level for the North Channel at monitoring point LE430 using the 75% TST optimised monitoring network.

Figure 6.22 presents the timeseries estimated using BME interpolation along with 68, 95 and 99% confidence intervals. At each data point, the confidence intervals prove somewhat informative with a number of data points falling within the 68% confidence interval. The confidence intervals are nonetheless broad, reaching as high as 40mg/l at one point in the timeseries. This is a direct consequence of the variance of Chlorophyll\_a monitoring data collocated with Chlorophyll\_a simulated data in Figure 6.16.

Figure 6.23 presents, for comparison purposes, the same timeseries of BME estimates but in this case using a covariance model derived to describe Chlorophyll\_a data autocorrelation throughout the Cork Harbour area. It is apparent upon visual inspection that the PDFs are non-informative at almost each data point. There is no consistency in over or under estimation of Chlorophyll\_a in Figure 6.23.
By contrast, in Figure 6.22 a number of BME mean estimates lie very close to Chlorophyll_\_a data points.

Figures 6.24 and 6.25 compare the BME estimates derived from data provided by the 25% and 50% TST optimised monitoring network for DIN in the lower Lee Estuary, with spatiotemporal covariance models derived from the entire water quality dataset for Cork Harbour; BME consistently underestimates each of the cross validation data points, with the linear regression line in each plot lying closer to the x axis representing the removed data points than the y axis representing the estimates. Table 6.4 and 6.5 also reflect the underestimation, of both BME and kriging, with the majority of mean error values falling below zero indicating underestimation. The cross-validation scatter plot in Figure 6.25 shows less dispersion of estimates when compared to Figure 6.24, thus indicating a greater trend for underestimation when more DIN monitoring points are available in the lower Lee Estuary for BME estimation. This is borne out in the figures for ME in Table 6.5, where the mean error reduces from -0.656 to -0.627 mg/l.

Tables 6.4 to 6.11 summarise the main cross validation statistics for BME and kriging estimates of Chlorophyll_\_a, DIN and MRP monitoring data for their respective 25%, 50% and 75% TST monitoring networks. Tables 6.4, 6.6, 6.8 and 6.10 present the kriging cross validation statistics of mean error, mean absolute error, root mean squared error and R^2. The four tables provide a baseline or control upon which BME estimates can be compared so that the benefit of incorporating probabilistic representations of numerical model water quality data simulations can be assessed. Mean error gives an indication of whether estimates on average overestimate or underestimate.

Comparing Table 6.4 to 6.5, BME estimation of redundant monitoring points using the 25% TST optimised monitoring network yields a considerable reduction in mean error of Chlorophyll_\_a estimation in all waterbodies except the Glashaboy Estuary and the Cork inner harbour area. BME estimation reduces the mean error consistently compared to kriging estimation for each of the 3 monitoring network options in the North Channel with respect to Chlorophyll_\_a. BME estimation of MRP results in a deterioration in accuracy compared to kriging estimation in almost all waterbodies and monitoring network options. BME estimation of DIN provides
mixed results based on mean error, with a reduction in mean error for the smallest monitoring network in the lower Lee Estuary, the Glashaboy Estuary and Lough Mahon.

Tables 6.6 and 6.7 present the mean absolute error of kriging and BME estimation throughout Cork Harbour. With the exception of a few instances, BME increases the mean absolute error for all water quality parameters, waterbodies and monitoring network sizes.

Root mean squared error provides a measure of the standard deviation of estimation errors. Tables 6.8 and 6.9 present the RMSE for kriging and BME respectively. BME estimation reduces the RMSE for a third of all waterbody, network and water quality parameters. Similarly, in a number of waterbodies, BME results in an improvement in $R^2$ values for Chlorophyll_a and MRP.

Comparing BME against kriging in the preceding paragraphs and in Tables 6.4 to 6.11, the most salient observation is the inconsistency in the improvement yielded by BME. This is due mainly to the broad shape of the soft data PDFs that are processed in the derivation of the posterior PDF describing each estimation point. The broad shape of the soft PDFs are a consequence of the procedure followed to derive probabilistic expressions for soft data describing the PDF of monitoring data given the modelling data in each subinterval. These soft PDFs are non-informative due to their broad shapes and thus describe a large range of water quality data; were the numerical model recalibrated and validated on an ongoing basis to improve the accuracy based on new monitoring data, the variance of the soft PDFs would be narrower and thus more informative and the posterior would be more accurate. The derivation of soft data was also not regionalised for each waterbody. Were a separate group of soft PDFs derived for each waterbody and subsequently used for BME estimation, accuracy may be improved. The gaps which are present in Tables 6.4 to 6.11 are mainly due to a lack of redundant monitoring points in some waterbodies once the larger optimised monitoring networks are evaluated. Some gaps, however, are due to a shortage of monitoring data in adjacent priority monitoring stations. Consequently, BME estimates are possible on the basis of probabilistic data describing numerical model water quality data.
Figure 6.22 Estimated timeseries of Chlorophyll_a at monitoring point LE430 using standard BME and monitoring data from the 75% TST monitoring network for Chlorophyll_a using a covariance model derived for the North Channel.

Figure 6.23 Estimated timeseries of Chlorophyll_a at monitoring point LE430 using standard BME and monitoring data from the 75% TST monitoring network for Chlorophyll_a using a covariance model derived for the whole of Cork Harbour using all available data
Figure 6.24 standard BME estimation of DIN in the lower Lee Estuary using 25% TST monitoring network

$y = 0.1531x + 0.62484$

Figure 6.25 standard BME estimation of DIN in the lower Lee Estuary using 50% TST monitoring network

$y = 0.13181x + 0.69596$
Table 6.4 ME of kriging cross validation of monitoring data for 3 monitoring scenarios

<table>
<thead>
<tr>
<th>Chlorophyll a (mg/m³)</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average M.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-2.973</td>
<td>-15.421</td>
<td>-1.655</td>
<td>-2.080</td>
<td>-5.532</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-6.531</td>
<td>-14.762</td>
<td>-</td>
<td>-</td>
<td>-10.647</td>
</tr>
<tr>
<td>MRP (µg/l)</td>
<td>-65.936</td>
<td>-32.867</td>
<td>30.330</td>
<td>0.448</td>
<td>33.568</td>
<td>30.779</td>
<td>-0.613</td>
<td>13.114</td>
</tr>
<tr>
<td>25%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.780</td>
<td>-15.060</td>
<td>0.358</td>
<td>-</td>
<td>13.114</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-10.287</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-4.653</td>
<td>-16.899</td>
<td>-</td>
<td>-</td>
<td>-10.776</td>
</tr>
<tr>
<td>DIN (mg/l)</td>
<td>0.346</td>
<td>0.571</td>
<td>0.519</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.479</td>
</tr>
<tr>
<td>25%</td>
<td>-0.855</td>
<td>-2.873</td>
<td>-1.413</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1.226</td>
<td>-</td>
</tr>
<tr>
<td>50%</td>
<td>0.346</td>
<td>0.571</td>
<td>0.519</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>1.582</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1.544</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 6.5 ME of BME cross validation of monitoring data for 3 monitoring scenarios

<table>
<thead>
<tr>
<th>Chlorophyll a (mg/m³)</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average M.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>3.788</td>
<td>3.536</td>
<td>-0.251</td>
<td>-3.396</td>
<td>2.589</td>
<td>4.025</td>
<td>0.044</td>
<td>1.476</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.780</td>
<td>-15.060</td>
<td>0.358</td>
<td>3.803</td>
<td>-2.530</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-4.653</td>
<td>-16.899</td>
<td>-</td>
<td>-</td>
<td>-10.776</td>
</tr>
<tr>
<td>MRP (µg/l)</td>
<td>-118.226</td>
<td>-4.155</td>
<td>29.088</td>
<td>18.324</td>
<td>49.225</td>
<td>19.003</td>
<td>-1.124</td>
<td>-</td>
</tr>
<tr>
<td>25%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-5.970</td>
<td>9.227</td>
<td>-9.062</td>
<td>1.958</td>
<td>-0.961</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-11.388</td>
<td>-12.346</td>
<td>-</td>
<td>-</td>
<td>-11.867</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DIN (mg/l)</td>
<td>0.455</td>
<td>-</td>
<td>0.495</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1.439</td>
<td>0.495</td>
</tr>
<tr>
<td>25%</td>
<td>-0.656</td>
<td>-2.757</td>
<td>-1.273</td>
<td>-3.489</td>
<td>0.456</td>
<td>-</td>
<td>-1.544</td>
<td>-</td>
</tr>
<tr>
<td>50%</td>
<td>-0.627</td>
<td>-2.553</td>
<td>-1.136</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1.439</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>0.495</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.495</td>
</tr>
</tbody>
</table>
Table 6.6 MAE of kriging cross validation of monitoring data for 3 monitoring scenarios

<table>
<thead>
<tr>
<th></th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average M.A.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll a (mg/m³)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9.892</td>
<td>15.655</td>
<td>5.737</td>
<td>4.822</td>
<td>9.026</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.745</td>
<td>17.362</td>
<td>-</td>
<td>-</td>
<td>13.054</td>
</tr>
<tr>
<td>MRP (µg/l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14.250</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14.250</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>15.211</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DIN (mg/l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>0.873</td>
<td>2.873</td>
<td>1.508</td>
<td>-</td>
<td>-</td>
<td>0.319</td>
<td>-</td>
<td>1.393</td>
</tr>
<tr>
<td>50%</td>
<td>0.435</td>
<td>0.571</td>
<td>0.519</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.508</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>1.582</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.582</td>
</tr>
</tbody>
</table>

Table 6.7 MAE of BME cross validation of monitoring data for 3 monitoring scenarios

<table>
<thead>
<tr>
<th></th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average M.A.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll a (mg/m³)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11.258</td>
<td>16.536</td>
<td>7.688</td>
<td>8.310</td>
<td>10.948</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9.596</td>
<td>17.631</td>
<td>-</td>
<td>-</td>
<td>13.614</td>
</tr>
<tr>
<td>MRP (µg/l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10.240</td>
<td>13.320</td>
<td>10.022</td>
<td>4.626</td>
<td>9.552</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>19.755</td>
<td>16.995</td>
<td>-</td>
<td>-</td>
<td>18.375</td>
</tr>
<tr>
<td>DIN (mg/l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>0.738</td>
<td>2.757</td>
<td>1.461</td>
<td>-</td>
<td>3.489</td>
<td>0.505</td>
<td>-</td>
<td>1.790</td>
</tr>
<tr>
<td>50%</td>
<td>0.636</td>
<td>2.553</td>
<td>1.479</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.556</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>0.788</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.788</td>
</tr>
</tbody>
</table>
Table 6.8 RMSE of kriging cross validation of monitoring data for 3 monitoring scenarios

<table>
<thead>
<tr>
<th></th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average R.M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chlorophyll</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a (mg/m³)</td>
<td>25%</td>
<td>8.559</td>
<td>7.423</td>
<td>15.387</td>
<td>15.103</td>
<td>23.602</td>
<td>8.742</td>
<td>8.052</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>13.286</td>
<td>25.467</td>
<td>7.457</td>
<td>6.112</td>
<td>13.081</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>11.808</td>
<td>26.597</td>
<td>-</td>
<td>-</td>
<td>19.202</td>
</tr>
<tr>
<td><strong>MRP (µg/l)</strong></td>
<td>25%</td>
<td>77.701</td>
<td>-</td>
<td>47.352</td>
<td>45.755</td>
<td>27.430</td>
<td>41.612</td>
<td>35.284</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>15.372</td>
<td>-</td>
<td>-</td>
<td>15.372</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>18.363</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>DIN (mg/l)</strong></td>
<td>25%</td>
<td>1.038</td>
<td>3.243</td>
<td>2.168</td>
<td>-</td>
<td>-</td>
<td>0.478</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-</td>
<td>0.516</td>
<td>0.636</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>-</td>
<td>2.079</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.079</td>
</tr>
</tbody>
</table>

Table 6.9 RMSE of BME cross validation of monitoring data for 3 monitoring scenarios

<table>
<thead>
<tr>
<th></th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average R.M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chlorophyll</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>14.961</td>
<td>25.974</td>
<td>9.557</td>
<td>10.083</td>
<td>15.144</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>12.948</td>
<td>27.967</td>
<td>-</td>
<td>-</td>
<td>20.457</td>
</tr>
<tr>
<td><strong>MRP (µg/l)</strong></td>
<td>25%</td>
<td>118.655</td>
<td>-</td>
<td>45.881</td>
<td>51.561</td>
<td>30.376</td>
<td>49.284</td>
<td>29.273</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>11.853</td>
<td>15.762</td>
<td>12.063</td>
<td>4.874</td>
<td>11.138</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>41.382</td>
<td>23.599</td>
<td>-</td>
<td>-</td>
<td>32.491</td>
</tr>
<tr>
<td><strong>DIN (mg/l)</strong></td>
<td>25%</td>
<td>0.912</td>
<td>3.126</td>
<td>2.094</td>
<td>-</td>
<td>3.529</td>
<td>0.658</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>2.831</td>
<td>2.831</td>
<td>0.801</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>-</td>
<td>0.952</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.952</td>
</tr>
</tbody>
</table>
Table 6.10 $R^2$ values of kriging cross validation of monitoring data for 3 monitoring scenarios

<table>
<thead>
<tr>
<th></th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll a</td>
<td>25% 0.236</td>
<td>0.004</td>
<td>0.028</td>
<td>0.015</td>
<td>0.157</td>
<td>0.035</td>
<td>0.260</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>50% 0.003</td>
<td>-</td>
<td>-</td>
<td>0.022</td>
<td>0.023</td>
<td>0.002</td>
<td>0.063</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>75% 0.076</td>
<td>0.004</td>
<td>0.000</td>
<td>0.296</td>
<td>0.042</td>
<td>0.041</td>
<td>0.120</td>
<td>0.064</td>
</tr>
<tr>
<td>MRP</td>
<td>25% 0.128</td>
<td>0.083</td>
<td>0.139</td>
<td>0.000</td>
<td>0.146</td>
<td>0.007</td>
<td>0.367</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>50% 0.012</td>
<td>-</td>
<td>-</td>
<td>0.001</td>
<td>0.007</td>
<td>0.006</td>
<td>0.006</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>75% 0.120</td>
<td>-</td>
<td>0.120</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.120</td>
</tr>
</tbody>
</table>

Table 6.11 $R^2$ values of BME cross validation of monitoring data for 3 monitoring scenarios

<table>
<thead>
<tr>
<th></th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll a</td>
<td>25% 0.021</td>
<td>0.081</td>
<td>0.071</td>
<td>0.059</td>
<td>0.585</td>
<td>0.033</td>
<td>0.242</td>
<td>0.156</td>
</tr>
<tr>
<td></td>
<td>50% -</td>
<td>-</td>
<td>-</td>
<td>0.100</td>
<td>0.093</td>
<td>0.000</td>
<td>0.146</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td>75% -</td>
<td>-</td>
<td>-</td>
<td>0.009</td>
<td>0.183</td>
<td>-</td>
<td>-</td>
<td>0.096</td>
</tr>
<tr>
<td>MRP</td>
<td>25% 0.920</td>
<td>-</td>
<td>0.003</td>
<td>0.082</td>
<td>0.049</td>
<td>0.776</td>
<td>0.052</td>
<td>0.314</td>
</tr>
<tr>
<td></td>
<td>50% -</td>
<td>-</td>
<td>-</td>
<td>0.912</td>
<td>0.426</td>
<td>0.374</td>
<td>0.007</td>
<td>0.430</td>
</tr>
<tr>
<td></td>
<td>75% -</td>
<td>-</td>
<td>-</td>
<td>0.012</td>
<td>0.001</td>
<td>-</td>
<td>-</td>
<td>0.006</td>
</tr>
<tr>
<td>DIN</td>
<td>25% 0.019</td>
<td>0.038</td>
<td>0.012</td>
<td>-</td>
<td>0.920</td>
<td>0.001</td>
<td>-</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>50% 0.033</td>
<td>0.029</td>
<td>0.033</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>75% -</td>
<td>-</td>
<td>0.120</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.120</td>
</tr>
</tbody>
</table>
6.4.2. Log estimation

Log-kriging and log-BME estimates were derived in the same manner as standard kriging and BME estimates as above. The ultimate resultant posterior PDF generated in logarithms is back transformed by deriving the exponent of the estimate. As a result, log estimation naturally causes right skewed posterior PDF after back transformation.

Figures 6.26 and 6.27 present a comparison of the standard and log transformed BME estimates at monitoring point LE 380 at the mouth of the main channel to the west of Cork Harbour. In this example, the 68% confidence intervals for the log transformed BME estimates capture the monitoring data points more frequently.

Log estimation results in a notable reduction in mean error when compared to standard estimation, when comparing Tables 6.12 and 6.13 with Tables 6.4 and 6.5. Log-BME estimates of MRP have the lowest mean error when compared to log-kriging, standard kriging and standard BME. Log-kriging and log-BME don’t result in an improvement in mean absolute error to any great extent when comparing results in Table 6.14 and 6.15 to the corresponding results for standard estimation however.

Log-BME estimation results in a notable reduction in RMSE for Chlorophyll_a and MRP in Table 6.17 when compared to the corresponding results in Table 6.16 for log-kriging. However, there are no notable gains with respect to $R^2$ by adopting log-BME estimation instead of log-kriging, although the standard deviation of estimation errors is reduced, as evidenced by a reduction in the RMSE.
Figure 6.26 Estimated timeseries of Chlorophyll_a at monitoring point LE380 using log-kriging and monitoring data from the 25% TST monitoring network for Chlorophyll_a using a single covariance model derived for Cork inner harbour

Figure 6.27 Estimated timeseries of Chlorophyll_a at monitoring point LE380 using standard kriging and monitoring data from the 25% TST monitoring network for Chlorophyll_a using a single covariance model derived for Cork inner harbour
Table 6.12 Mean Error (M.E.) of log-kriging cross validation of EPA monitoring data for 3 different monitoring scenarios

<table>
<thead>
<tr>
<th></th>
<th>Chlorophyll a (mg/m³)</th>
<th>MRP (µg/l)</th>
<th>DIN (mg/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lee</td>
<td>Glashaboy</td>
<td>Lough Mahon</td>
</tr>
<tr>
<td>25%</td>
<td>-4.473</td>
<td>-1.465</td>
<td>-5.878</td>
</tr>
<tr>
<td>25%</td>
<td>-79.843</td>
<td>-1.542</td>
<td>-3.945</td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td>-10.146</td>
<td>-8.337</td>
</tr>
<tr>
<td>75%</td>
<td>-9.374</td>
<td>-13.097</td>
<td>-8.089</td>
</tr>
<tr>
<td>25%</td>
<td>-1.542</td>
<td>-2.982</td>
<td>-3.960</td>
</tr>
<tr>
<td>50%</td>
<td>-1.469</td>
<td>-2.514</td>
<td>-3.960</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 6.13 M.E. of log-BME cross validation of EPA monitoring data for 3 different monitoring scenarios

<table>
<thead>
<tr>
<th></th>
<th>Chlorophyll a (mg/m³)</th>
<th>MRP (µg/l)</th>
<th>DIN (mg/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lee</td>
<td>Glashaboy</td>
<td>Lough Mahon</td>
</tr>
<tr>
<td>25%</td>
<td>-</td>
<td>1.350</td>
<td>-</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-5.934</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-13.097</td>
</tr>
<tr>
<td>25%</td>
<td>-</td>
<td>-</td>
<td>32.395</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>0.135</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-7.841</td>
</tr>
<tr>
<td>50%</td>
<td>-3.023</td>
<td>-3.960</td>
<td>-3.960</td>
</tr>
<tr>
<td>75%</td>
<td>-3.960</td>
<td>-3.960</td>
<td>-3.960</td>
</tr>
</tbody>
</table>

219
### Table 6.14 Mean Absolute Error (M.A.E.) of log-kriging cross validation of EPA monitoring data for 3 different monitoring scenarios

<table>
<thead>
<tr>
<th>Chlorophyll a (mg/m³)</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average M.A.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13.204</td>
<td>16.897</td>
<td>1.902</td>
<td>8.178</td>
<td>10.045</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>16.909</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>16.909</td>
</tr>
<tr>
<td>MRP (µg/l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>22.566</td>
<td>39.160</td>
<td>8.639</td>
<td>72.237</td>
<td>35.650</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>36.916</td>
<td>99.138</td>
<td>-</td>
<td>-</td>
<td>68.027</td>
</tr>
<tr>
<td>DIN (mg/l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>1.542</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.542</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>0.993</td>
<td>2.545</td>
<td>1.650</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.729</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>0.523</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.523</td>
</tr>
</tbody>
</table>

### Table 6.15 M.A.E. of log-BME cross validation of EPA monitoring data for 3 different monitoring scenarios

<table>
<thead>
<tr>
<th>Chlorophyll a (mg/m³)</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average M.A.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>-</td>
<td>6.413</td>
<td>-</td>
<td>12.205</td>
<td>15.772</td>
<td>-</td>
<td>-</td>
<td>11.463</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.066</td>
<td>14.935</td>
<td>5.176</td>
<td>4.974</td>
<td>8.288</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.191</td>
<td>15.879</td>
<td>-</td>
<td>-</td>
<td>12.035</td>
</tr>
<tr>
<td>MRP (µg/l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>34.806</td>
<td>13.866</td>
<td>29.750</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>15.356</td>
<td>10.626</td>
<td>10.398</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>15.543</td>
<td>18.043</td>
<td>-</td>
<td>-</td>
<td>16.793</td>
</tr>
<tr>
<td>DIN (mg/l)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>-</td>
<td>2.993</td>
<td>-</td>
<td>-</td>
<td>3.960</td>
<td>1.011</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>50%</td>
<td>1.248</td>
<td>3.023</td>
<td>1.804</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.025</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### Table 6.16 Root Mean Squared Error (R.M.S.E.) of log-kriging cross validation of EPA monitoring data for 3 different monitoring scenarios

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average R.M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>17.569</td>
<td>25.616</td>
<td>2.623</td>
<td>10.478</td>
<td>14.071</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>25.906</td>
<td>-</td>
<td>-</td>
<td>25.906</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average R.M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>92.953</td>
<td>-</td>
<td>47.593</td>
<td>39.221</td>
<td>16.831</td>
<td>45.402</td>
<td>48.486</td>
<td>48.415</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>33.607</td>
<td>58.938</td>
<td>10.872</td>
<td>115.993</td>
<td>54.852</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>56.869</td>
<td>144.489</td>
<td>-</td>
<td>100.679</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average R.M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.624</td>
<td>-</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.129</td>
<td>-</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.829</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table 6.17 R.M.S.E. of log-BME cross validation of EPA monitoring data for 3 different monitoring scenarios

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average R.M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>10.969</td>
<td>-</td>
<td>-</td>
<td>17.395</td>
<td>23.826</td>
<td>-</td>
<td>-</td>
<td>17.397</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11.204</td>
<td>23.853</td>
<td>6.806</td>
<td>5.762</td>
<td>11.906</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>11.481</td>
<td>25.341</td>
<td>-</td>
<td>-</td>
<td>18.411</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average R.M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>49.297</td>
<td>16.223</td>
<td>43.360</td>
<td>-</td>
<td>36.293</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>29.242</td>
<td>15.955</td>
<td>16.686</td>
<td>-</td>
<td>20.628</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>32.910</td>
<td>22.925</td>
<td>-</td>
<td>-</td>
<td>27.918</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average R.M.S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>3.275</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.204</td>
</tr>
<tr>
<td>50%</td>
<td>1.328</td>
<td>3.289</td>
<td>2.368</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.328</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### Table 6.18 $R^2$ values of log-kriging cross validation of EPA monitoring data for 3 different monitoring scenarios

<table>
<thead>
<tr>
<th></th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chlorophyll a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>0.013</td>
<td>0.002</td>
<td>0.018</td>
<td>0.028</td>
<td>0.002</td>
<td>0.016</td>
<td>0.165</td>
<td>0.035</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.037</td>
<td>0.614</td>
<td>0.929</td>
<td>0.119</td>
<td>0.425</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.531</td>
<td>-</td>
<td>-</td>
<td>0.531</td>
</tr>
<tr>
<td><strong>MRP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>0.063</td>
<td>-</td>
<td>0.005</td>
<td>0.823</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.149</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.039</td>
<td>0.741</td>
<td>0.000</td>
<td>0.003</td>
<td>0.196</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.081</td>
<td>0.180</td>
<td>-</td>
<td>-</td>
<td>0.131</td>
</tr>
<tr>
<td><strong>DIN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>0.128</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.128</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>0.002</td>
<td>0.019</td>
<td>0.034</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.019</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>0.466</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.466</td>
</tr>
</tbody>
</table>

### Table 6.19 $R^2$ values of log-BME cross validation of EPA monitoring data for 3 different monitoring scenarios

<table>
<thead>
<tr>
<th></th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
<th>Average $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chlorophyll a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>-</td>
<td>0.009</td>
<td>0.000</td>
<td>0.040</td>
<td>0.020</td>
<td>0.002</td>
<td>-</td>
<td>0.014</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.024</td>
<td>0.169</td>
<td>0.037</td>
<td>0.274</td>
<td>0.126</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.054</td>
<td>0.808</td>
<td>-</td>
<td>-</td>
<td>0.431</td>
</tr>
<tr>
<td><strong>MRP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.188</td>
<td>0.001</td>
<td>0.001</td>
<td>0.009</td>
<td>0.050</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.032</td>
<td>0.387</td>
<td>0.002</td>
<td>0.054</td>
<td>0.118</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.050</td>
<td>0.154</td>
<td>-</td>
<td>-</td>
<td>0.102</td>
</tr>
<tr>
<td><strong>DIN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td>-</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
<td>0.019</td>
<td>0.035</td>
<td>-</td>
<td>0.018</td>
</tr>
<tr>
<td>50%</td>
<td>0.013</td>
<td>0.000</td>
<td>0.006</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.006</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
6.4.3. Comparison of overall accuracy of lognormal and standard interpolation

In order to assess the ability of kriging and BME to replicate the original dataset, the skewness and kurtosis of standard and log estimates of both geostatistical estimates were determined and are contained in Table 6.20, with reference values for both parameters given as those for the original dataset. The histograms of the original dataset and the 12 estimates of the original dataset are given in Figures 6.28 to 6.30 for Chlorophyll_a, MRP and DIN. In any of the three figures, a blank plot indicates either insufficient data for estimation or insufficient data available to derive a covariance model.

For each water quality parameter, the histogram of estimates combines the monitoring data from the retained monitoring points in the optimised monitoring network, with the estimates derived at the redundant monitoring points. It would be expected that the histogram of the 75% TST optimised monitoring networks would bear the greatest similarity to the original dataset, due to the greater volume of original monitoring data contained within it, and fewer geostatistical estimates derived from a larger amount of data. Upon visual inspection of Figure 6.28, the Chlorophyll_a dataset is closely replicated by kriging, BME and log-kriging. The histogram of log-BME estimates does not appear similar to the original, and this is reinforced by associated skewness and kurtosis values in Table 6.20.

Each of the 4 estimation configurations for the 75% TST optimised monitoring network for MRP result in histograms which closely match the skewness and kurtosis of the original dataset. 3 of the 4 estimation configurations for DIN result in a dataset or histogram which closely matches the original DIN dataset.

The larger optimised monitoring networks provided by the 75% TST network provide sufficient data for the kriging and BME estimation of redundant monitoring points without the estimates skewing the overall characteristics of the dataset. Taking the estimates underpinned by the 25% and 50% TST networks in Figures 6.28 to 6.30, it can be seen that there is more inconsistency in the number of lower value data points that are correctly estimated. Observing the leftmost bar of the histograms in Figure 6.28, in the original dataset the amount of data points falling into the lowest bin was approximately 130, whereas in the second and third rows of Figure
6.28 indicate that the amount of data points falling into the lowest bin either increased significantly to in excess of 170, or dropped to approximately 100. The same pattern is evident in Figure 6.29. There is no emergent trend in the 25 and 50% TST network DIN estimates presented in Figure 6.30, with even greater variation in the shape of the datasets.

R² values rarely exceed 0.5 in the results presented earlier. Due to the low proportion of estimation data points that contribute to the overall dataset, the moderate success of kriging and BME in estimating data at redundant monitoring points has little impact on the overall shape of the dataset.

Table 6.21 summarises the optimum estimation configuration for each waterbody and monitoring network size based on R² values. In summary, of the 33 estimation configurations considered in Table 6.21, kriging and log kriging provide the best estimate for 17 whilst BME and log-BME provides the best estimate for 16.

Based on the reference skewness and kurtosis values, BME and log-kriging best recreated the original dataset for the 75% TST monitoring network monitoring data for Chlorophyll_a and MRP. The original DIN dataset is well replicated by kriging, log-kriging and BME for the 75% TST monitoring network.
Figure 6.28 Comparison of Chlorophyll_a estimation datasets for all monitoring optimisation schemes and geostatistical estimation combinations

Figure 6.29 Comparison of MRP estimation datasets for all monitoring optimisation schemes and geostatistical estimation combinations
Figure 6.30 Comparison of DIN estimation datasets for all monitoring optimisation schemes and geostatistical estimation combinations
Table 6.20 Skewness and kurtosis parameters for combinations of kriging and BME of standard and log transformed monitoring data

<table>
<thead>
<tr>
<th>Chlorophyll a</th>
<th>% TST</th>
<th>Kriging</th>
<th>BME</th>
<th>Log Kriging</th>
<th>Log BME</th>
<th>Ref. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness</td>
<td>25</td>
<td>-</td>
<td>2.39</td>
<td>2.02</td>
<td>16.22</td>
<td>3.48</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-</td>
<td>4.36</td>
<td>4.06</td>
<td>4.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>4.16</td>
<td>4.07</td>
<td>4.03</td>
<td>4.22</td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td>25</td>
<td>-</td>
<td>10.79</td>
<td>8.07</td>
<td>263.98</td>
<td>22.45</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-</td>
<td>34.24</td>
<td>28.98</td>
<td>38.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>32.19</td>
<td>31.16</td>
<td>30.07</td>
<td>32.85</td>
<td></td>
</tr>
<tr>
<td>MRP</td>
<td>% TST</td>
<td>Kriging</td>
<td>BME</td>
<td>Log Kriging</td>
<td>Log BME</td>
<td>Ref. Value</td>
</tr>
<tr>
<td>Skewness</td>
<td>25</td>
<td>-</td>
<td>0.93</td>
<td>1.76</td>
<td>16.37</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1.31</td>
<td>1.29</td>
<td>1.43</td>
<td>16.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>1.53</td>
<td>1.58</td>
<td>1.54</td>
<td>1.62</td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td>25</td>
<td>-</td>
<td>5.40</td>
<td>7.20</td>
<td>269.00</td>
<td>5.62</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>4.99</td>
<td>4.86</td>
<td>5.28</td>
<td>268.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>5.51</td>
<td>5.70</td>
<td>5.55</td>
<td>5.84</td>
<td></td>
</tr>
<tr>
<td>DIN</td>
<td>% TST</td>
<td>Kriging</td>
<td>BME</td>
<td>Log Kriging</td>
<td>Log BME</td>
<td>Ref. Value</td>
</tr>
<tr>
<td>Skewness</td>
<td>25</td>
<td>2.67</td>
<td>1.14</td>
<td>2.99</td>
<td>-</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1.53</td>
<td>1.52</td>
<td>1.52</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>9.70</td>
<td>3.45</td>
<td>11.69</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td>25</td>
<td>4.68</td>
<td>4.63</td>
<td>4.46</td>
<td>-</td>
<td>5.17</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.21 Comparison of the optimum geostatistical interpolation technique for each water quality parameter, waterbody and monitoring network size

<table>
<thead>
<tr>
<th>Chlorophyll a</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>Krig</td>
<td>BME</td>
<td>BME</td>
<td>BME</td>
<td>Log-BME</td>
<td>Krig</td>
<td>Krig</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Log-BME</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Krig</td>
<td>Log-BME</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MRP</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>BME</td>
<td>-</td>
<td>Log-Krig</td>
<td>Log-Krig</td>
<td>Krig</td>
<td>BME</td>
<td>BME</td>
</tr>
<tr>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>BME</td>
<td>BME</td>
<td>BME</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Krig</td>
<td>Log-Krig</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIN</th>
<th>Lee</th>
<th>Glashaboy</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>Log-Krig</td>
<td>Krig</td>
<td>BME</td>
<td>-</td>
<td>BME</td>
<td>Log-BME</td>
<td>-</td>
</tr>
<tr>
<td>50%</td>
<td>Krig</td>
<td>Krig</td>
<td>Krig</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>75%</td>
<td>-</td>
<td>-</td>
<td>Log-Krig</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
6.4.4. Comparison of water quality assessment outcomes

The ultimate objective of geostatistical interpolation of the three water quality parameters analysed in the preceding sections is to utilise the estimates in combination with the retained data from the priority monitoring points retained in each of the proposed optimum monitoring networks to determine the trophic status of each waterbody. Tables 6.22 to 6.28 compare the trophic status outcomes from each combination of kriging or BME estimation of standard or log transformed data for each of the 3 proposed monitoring networks with the outcome were the entire original dataset to be used. Where the trophic status returned is the same as EPA reporting, the status is highlighted in bold font. The emerging trend from the collection of tables indicates that the most accurate match of the trophic status returned by the entire dataset is achieved through using the standard dataset, the largest monitoring network given by the 75% TST optimum network and either kriging or BME.

Table 6.22 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout the lower Lee Estuary

<table>
<thead>
<tr>
<th></th>
<th>95-99</th>
<th></th>
<th>99-03</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EPA</td>
<td>KRG</td>
<td>BME</td>
<td>EPA</td>
<td>KRG</td>
</tr>
<tr>
<td>Std.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.23 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout the Glashaboy Estuary

<table>
<thead>
<tr>
<th></th>
<th>95-99</th>
<th></th>
<th>99-03</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EPA</td>
<td>KRG</td>
<td>BME</td>
<td>EPA</td>
<td>KRG</td>
</tr>
</tbody>
</table>
Table 6.24 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout the North Channel

<table>
<thead>
<tr>
<th></th>
<th>95-99</th>
<th>EPA DATA</th>
<th>KRING</th>
<th>BME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std.</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>Log</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>99-03</th>
<th>EPA DATA</th>
<th>KRING</th>
<th>BME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std.</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>Log</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
</tbody>
</table>

Table 6.25 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout Lough Mahon

<table>
<thead>
<tr>
<th></th>
<th>95-99</th>
<th>EPA DATA</th>
<th>KRING</th>
<th>BME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std.</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>Log</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>99-03</th>
<th>EPA DATA</th>
<th>KRING</th>
<th>BME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std.</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>Log</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
</tbody>
</table>

Table 6.26 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout the Owenacurra Estuary

<table>
<thead>
<tr>
<th></th>
<th>95-99</th>
<th>EPA DATA</th>
<th>KRING</th>
<th>BME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std.</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>Log</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>99-03</th>
<th>EPA DATA</th>
<th>KRING</th>
<th>BME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std.</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>Log</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
</tbody>
</table>

Table 6.27 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout Cork inner harbour

<table>
<thead>
<tr>
<th></th>
<th>95-99</th>
<th>EPA DATA</th>
<th>KRING</th>
<th>BME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std.</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>Log</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>99-03</th>
<th>EPA DATA</th>
<th>KRING</th>
<th>BME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Std.</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
<tr>
<td></td>
<td>Log</td>
<td>Int</td>
<td>Int</td>
<td>Int</td>
</tr>
</tbody>
</table>
Table 6.28 Comparison of the TSAS trophic status returned using standard and log variants of kriging and BME estimators to estimate water quality at redundant monitoring points throughout Cork outer harbour

<table>
<thead>
<tr>
<th></th>
<th>95-99</th>
<th>99-03</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EPA</td>
<td>KRI</td>
</tr>
<tr>
<td></td>
<td>DATA</td>
<td></td>
</tr>
<tr>
<td>Std.</td>
<td>Int</td>
<td>Int</td>
</tr>
<tr>
<td>Log</td>
<td>Int</td>
<td>Int</td>
</tr>
</tbody>
</table>

6.5. Quasi-stationary BME and kriging

The preceding sections assessed the potential of BME and kriging in the marine environment where estimation was underpinned by either a single covariance model derived from the entire monitoring dataset or individual covariance models for each waterbody which were derived solely from monitoring data within that waterbody; thus far stationarity was assumed either at the scale of the entire monitoring network for Cork Harbour or within each waterbody.

The following section pursues the possibility that covariance is quasi-stationary such that modelling of covariance centred around each monitoring point results in a spatial gradient of covariance sill and spatial and temporal range as per Akita et al. (2012). The quasi-stationary covariance models thus derived are used to underpin both kriging and BME estimates of standard and log transformed water quality data.

Figure 6.31 presents three quasi-stationary covariance models of log transformed MRP within the lower Lee Estuary. Upon visual inspection, it is clear that the covariance models are similar with respect to shape, sill and range. The monitoring points are named such that the lower valued labels are positioned furthest upstream with the higher labelled values falling further downstream such that LE150, LE160 and LE170 are ordered upstream to downstream. The spatial covariance models of LE150 and LE160 are quite similar with respect to shape, while LE160 and LE170 have near identical sill values. Figure 6.32 contains all of the spatial and temporal covariance models for log transformed MRP. Collectively, Figures 6.31 and 6.32 give credence to the possibility that covariance within Cork Harbour is quasi-stationary.
Leave-one-out cross correlation was completed by removing each daily averaged monitoring value at each monitoring point for Chlorophyll_a, MRP and DIN and using all of the remaining daily averaged data points for the parameter in question to estimate the spatiotemporal mean trend and covariance models and subsequently derive kriging and BME estimates of that data point.

Figures 6.33 to 6.37 present scatter plots for a number of leave-one-out cross validation configurations. The most notable feature of each figure is that the line of best fit is much closer to the line representing direct proportionality which would indicate 100% accuracy of estimation.

Tables 6.28 to 6.31 summarise the $R^2$ values of the leave-one-out cross validation of BME estimation of the three water quality parameters in standard and log-transformed format. Comparing Table 6.28 and 6.29, it is clear that BME estimation underpinned by quasi-stationary covariance models describing each water quality parameter within each waterbody results in a far greater accuracy of BME. The same conclusion can be drawn from BME estimation of log transformed water quality parameters as indicated by the increased $R^2$ values in Table 6.31 compared to Table 6.30.

Tables 6.32 and 6.33 indicate the optimum interpolation configuration to consider for each waterbody and water quality parameter using quasi-stationary covariance models centred on each monitoring point describing water quality variation throughout Cork Harbour and within each waterbody respectively.

Tables 6.35 to 6.38 have been included to summarise the ability of different geostatistical estimation configurations to capture cross validation data within the 68, 95 and 99% confidence intervals of the posterior PDFs; for each waterbody, the percentage of monitoring data values which fell within the confidence intervals is given. In the vast majority of cases contained in the four tables, assuming quasi-stationary covariance within individual waterbodies proved favourable as opposed to assuming quasi-stationary covariance within Cork Harbour, with a greater percentage of monitoring data points contained in the three confidence interval sizes in most cases. From Table 6.35, it can be seen that leave-one-out cross validation of kriging resulted in a number of MRP posterior PDF confidence intervals not
capturing a single monitoring data point when quasi-stationarity was assumed within Cork Harbour.

Log kriging and Log BME leave-one-out cross validation of the three water parameters results in some confounding results in Tables 6.37 and 6.38. While in the case of standard kriging and BME cross validation, more monitoring data points were captured by the posterior PDF confidence intervals when quasi-stationarity was assumed within each waterbody, in the case of log kriging and log BME, the confidence intervals contain more data points when covariance is assumed quasi-stationary throughout Cork Harbour. This is a consequence of the log estimates having larger confidence intervals after back transformation, and the noise of distant soft PDFs is assimilated in the posterior PDF and scaled up due to back transformation. Thus, the validity of the log kriging and log BME should be questioned where quasi-stationarity is assumed on the scale of Cork harbour.

Figures 6.38 and 6.39 contain the histograms of the 4 interpolation configurations for interpolation assuming quasi-stationarity throughout Cork Harbour and within each waterbody respectively. The skewness and kurtosis statistics are contained in Tables 6.38 and 6.39. Clearly from the histograms and the associated statistics, standard kriging and BME come closest to replicating the original dataset through the cross validation procedure.

![Covariance models derived from log-transformed MRP residual data in the lower Lee Estuary](image)

Figure 6.31 Covariance models derived from log-transformed MRP residual data in the lower Lee Estuary
Figure 6.32 Covariance models derived from log-transformed MRP residual data throughout Cork Harbour.

Figure 6.33 Leave-one-out BME estimation of Chlorophyll_a in Cork Inner harbour using quasi-stationary covariance models for estimation centred around each monitoring point.
BME and geospatial estimation

Figure 6.34 Leave-one-out BME estimation of Chlorophyll_a in the lower Lee Estuary using quasi-stationary covariance models for estimation centred around each monitoring point.

Figure 6.35 Leave-one-out log-BME estimation of DIN in the lower Lee Estuary using quasi-stationary covariance models derived from data for all of Cork Harbour.
Figure 6.36 Leave-one-out log-kriging estimates of DIN in Cork inner harbour using quasi-stationary covariance models for estimation centred on each monitoring point.

\[ y = 0.5856x + 0.552894 \]

Figure 6.37 Comparison of standard kriging estimates of DIN in Cork inner harbour using quasi-stationary covariance models for estimation centred on each monitoring point.

\[ y = 0.8279x + 0.040392 \]
### Table 6.29: R² values of leave-one-out BME cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from the entire monitoring dataset

<table>
<thead>
<tr>
<th></th>
<th>BME of EPA data</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lee</td>
<td>North Channel</td>
<td>Lough Mahon</td>
<td>Owenacurra</td>
<td>Inner harbour</td>
<td>Outer harbour</td>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorophyll a</td>
<td>0.522</td>
<td>0.047</td>
<td>0.145</td>
<td>0.846</td>
<td>0.417</td>
<td>0.132</td>
<td>0.351</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRP</td>
<td>0.209</td>
<td>0.001</td>
<td>0.049</td>
<td>0.005</td>
<td>0.114</td>
<td>0.053</td>
<td>0.072</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIN</td>
<td>0.222</td>
<td>0.036</td>
<td>0.075</td>
<td>0.917</td>
<td>0.099</td>
<td>0.031</td>
<td>0.230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 6.30: R² values of leave-one-out BME cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from monitoring data for each waterbody

|                | BME of EPA data |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|----------------|-----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|          |
|                | Lee             | North Channel | Lough Mahon | Owenacurra | Inner harbour | Outer harbour | Average |
| Chlorophyll a  | 0.757           | 0.303     | 0.833    | 0.248    | 0.709     | 0.340     | 0.532    |
| MRP            | 0.308           | 0.534     | 0.734    | 0.846    | 0.842     | 0.486     | 0.625    |
| DIN            | 0.054           | 0.107     | 0.261    | 0.689    | 0.013     | 0.921     | 0.341    |
Table 6.31 $R^2$ values of leave-one-out log-BME cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from the entire monitoring dataset

<table>
<thead>
<tr>
<th></th>
<th>Chlorophyll a</th>
<th>MRP</th>
<th>DIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee</td>
<td>0.009</td>
<td>-</td>
<td>0.128</td>
</tr>
<tr>
<td>North Channel</td>
<td>0.001</td>
<td>0.121</td>
<td>0.000</td>
</tr>
<tr>
<td>Lough Mahon</td>
<td>0.033</td>
<td>0.226</td>
<td>0.000</td>
</tr>
<tr>
<td>Owenacurra</td>
<td>0.104</td>
<td>0.011</td>
<td>0.319</td>
</tr>
<tr>
<td>Inner harbour</td>
<td>-</td>
<td>0.355</td>
<td>0.139</td>
</tr>
<tr>
<td>Outer harbour</td>
<td>-</td>
<td>0.002</td>
<td>0.018</td>
</tr>
<tr>
<td>Average</td>
<td>0.078</td>
<td>0.143</td>
<td>0.101</td>
</tr>
</tbody>
</table>

Table 6.32 $R^2$ values of leave-one-out log-BME cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from monitoring data for each waterbody

<table>
<thead>
<tr>
<th></th>
<th>Chlorophyll a</th>
<th>MRP</th>
<th>DIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee</td>
<td>0.853</td>
<td>0.821</td>
<td>0.555</td>
</tr>
<tr>
<td>North Channel</td>
<td>0.495</td>
<td>0.915</td>
<td>0.002</td>
</tr>
<tr>
<td>Lough Mahon</td>
<td>0.898</td>
<td>-</td>
<td>0.006</td>
</tr>
<tr>
<td>Owenacurra</td>
<td>0.905</td>
<td>0.457</td>
<td>0.210</td>
</tr>
<tr>
<td>Inner harbour</td>
<td>0.062</td>
<td>-</td>
<td>0.004</td>
</tr>
<tr>
<td>Outer harbour</td>
<td>0.704</td>
<td>0.428</td>
<td>0.043</td>
</tr>
<tr>
<td>Average</td>
<td>0.653</td>
<td>0.655</td>
<td>0.137</td>
</tr>
</tbody>
</table>

Table 6.33 Summary of the optimum estimation procedures of leave-one-out cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from the entire monitoring dataset

<table>
<thead>
<tr>
<th></th>
<th>Chlorophyll a</th>
<th>MRP</th>
<th>DIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee</td>
<td>BME</td>
<td>Krig</td>
<td>Log-Krig</td>
</tr>
<tr>
<td>North Channel</td>
<td>Log-Krig</td>
<td>Krig</td>
<td>BME</td>
</tr>
<tr>
<td>Lough Mahon</td>
<td>Log-Krig</td>
<td>Krig</td>
<td>BME</td>
</tr>
<tr>
<td>Owenacurra</td>
<td>Log-Krig</td>
<td>Krig</td>
<td>BME</td>
</tr>
<tr>
<td>Inner harbour</td>
<td>BME</td>
<td>Log-Krig</td>
<td>Log-Krig</td>
</tr>
<tr>
<td>Outer harbour</td>
<td>Log-Krig</td>
<td>Log-Krig</td>
<td>Log-Krig</td>
</tr>
</tbody>
</table>
Table 6.34 Summary of the optimum estimation procedures of leave-one-out cross validation of the monitoring dataset using quasi-stationary covariance models centred on each monitoring point derived from monitoring data for each waterbody

<table>
<thead>
<tr>
<th></th>
<th>Lee</th>
<th>North Channel</th>
<th>Lough Mahon</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chlorophyll a</strong></td>
<td>Krig</td>
<td>Log-Krig</td>
<td>Log-Krig</td>
<td>Krig</td>
<td>Log-Krig</td>
<td>Krig</td>
</tr>
<tr>
<td><strong>MRP</strong></td>
<td>Log-Krig</td>
<td>Log-Krig</td>
<td>Log-Krig</td>
<td>Krig</td>
<td>Krig</td>
<td>Log-Krig</td>
</tr>
<tr>
<td><strong>DIN</strong></td>
<td>Krig</td>
<td>Krig</td>
<td>Log-Krig</td>
<td>BME</td>
<td>Krig</td>
<td>BME</td>
</tr>
</tbody>
</table>
### Table 6.35 A comparison of the percentage of monitoring data in each waterbody falling within the 68%, 95% and 99% when estimated using standard kriging and quasi-stationary covariance models derived from the entire dataset or from the respective waterbody.

<table>
<thead>
<tr>
<th>C.I. %</th>
<th>Std Kriging</th>
<th>Lee</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>40% 66% 71%</td>
<td>29% 71% 82%</td>
<td>23% 42% 54%</td>
<td>11% 42% 68%</td>
<td>10% 51% 69%</td>
<td>39% 67% 76%</td>
<td></td>
</tr>
<tr>
<td>Waterbody</td>
<td>60% 86% 89%</td>
<td>48% 73% 83%</td>
<td>31% 54% -</td>
<td>0% 0% 50%</td>
<td>53% 71% 82%</td>
<td>65% 84% 89%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C.I. %</th>
<th>MRP</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0% 0% 0%</td>
<td>29% 71% 71%</td>
<td>0% 0% 0%</td>
<td>41% 69% 82%</td>
<td>66% 87% 89%</td>
</tr>
<tr>
<td>Waterbody</td>
<td>54% 77% 85%</td>
<td>38% 38% 46%</td>
<td>8% 36% -</td>
<td>30% 60% 80%</td>
<td>33% 71% 86%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C.I. %</th>
<th>DIN</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0% 0% 9%</td>
<td>7% 35% 56%</td>
<td>61% 67% 78%</td>
</tr>
<tr>
<td>Waterbody</td>
<td>50% 62% 65%</td>
<td>47% 69% 73%</td>
<td>17% 17% -</td>
</tr>
</tbody>
</table>

### Table 6.36 A comparison of the percentage of monitoring data in each waterbody falling within the 68%, 95% and 99% when estimated using standard BME and quasi-stationary covariance models derived from the entire dataset or from the respective waterbody.

<table>
<thead>
<tr>
<th>C.I. %</th>
<th>Std BME</th>
<th>Lee</th>
<th>Lough Mahon</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>17% 38% 48%</td>
<td>26% 51% 60%</td>
<td>23% 32% 36%</td>
<td>0% 28% 56%</td>
<td>8% 40% 54%</td>
<td>36% 64% 69%</td>
<td></td>
</tr>
<tr>
<td>Waterbody</td>
<td>50% 62% 65%</td>
<td>47% 69% 73%</td>
<td>17% 17% -</td>
<td>75% 75% 81%</td>
<td>45% 73% 82%</td>
<td>43% 57% 57%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C.I. %</th>
<th>MRP</th>
<th>North Channel</th>
<th>Owenacurra</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0% 0% 8%</td>
<td>18% 27% 32%</td>
<td>9% 18% 18%</td>
<td>22% 33% 44%</td>
<td>13% 27% 32%</td>
</tr>
<tr>
<td>Waterbody</td>
<td>35% 59% 59%</td>
<td>17% 27% 43%</td>
<td>10% 15% -</td>
<td>50% 54% 75%</td>
<td>33% 70% 87%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C.I. %</th>
<th>DIN</th>
<th>Inner harbour</th>
<th>Outer harbour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0% 0% 10%</td>
<td>8% 38% 55%</td>
<td>39% 39% 48%</td>
</tr>
<tr>
<td>Waterbody</td>
<td>72% 76% 76%</td>
<td>71% 96% 98%</td>
<td>0% 21% -</td>
</tr>
</tbody>
</table>
### BME and geospatial estimation

Table 6.37 A comparison of the percentage of monitoring data in each waterbody falling within the 68%, 95% and 99% when estimated using log-kriging and quasi-stationary covariance models derived from the entire dataset or from the respective waterbody.

<table>
<thead>
<tr>
<th></th>
<th>Log Kriging</th>
<th>C.I. %</th>
<th>Lee</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
<th>Lough Mahon</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
<th>North Channel</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
<th>Owenacurra</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
<th>Inner harbour</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
<th>Outer harbour</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorophyll a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.I. %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterbody</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.I. %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterbody</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.I. %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterbody</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.38 A comparison of the percentage of monitoring data in each waterbody falling within the 68%, 95% and 99% when estimated using log-BME and quasi-stationary covariance models derived from the entire dataset or from the respective waterbody.

<table>
<thead>
<tr>
<th></th>
<th>Log BME</th>
<th>C.I. %</th>
<th>Lee</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
<th>Lough Mahon</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
<th>North Channel</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
<th>Owenacurra</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
<th>Inner harbour</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
<th>Outer harbour</th>
<th>68%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorophyll a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.I. %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterbody</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.I. %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterbody</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.I. %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waterbody</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

240
Figure 6.38 Comparison of histograms of datasets generated by carrying out leave-one-out cross validation following 4 estimation procedures using quasi-stationary covariance models derived from data for all of Cork Harbour.

Figure 6.39 Comparison of histograms of datasets generated by carrying out leave-one-out cross validation following estimation procedures using separate quasi-stationary covariance models for each waterbody in Cork Harbour.
Table 6.39 Skewness and kurtosis values for the datasets estimated by kriging and BME using standard and log transformed data using quasi-stationary covariance models centred around each monitoring point derived using monitoring data from all over Cork Harbour

<table>
<thead>
<tr>
<th>Data</th>
<th>Ref. Value</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Kriging</td>
<td>BME</td>
</tr>
<tr>
<td>Chlorophyll a</td>
<td>3.48</td>
<td>5.25</td>
<td>3.41</td>
</tr>
<tr>
<td>MRP</td>
<td>1.58</td>
<td>2.43</td>
<td>2.11</td>
</tr>
<tr>
<td>DIN</td>
<td>1.72</td>
<td>2.02</td>
<td>1.13</td>
</tr>
</tbody>
</table>

Table 6.40 Skewness and kurtosis values for the datasets estimated by kriging and BME using standard and log transformed data using quasi-stationary covariance models centred around each monitoring point derived using monitoring data restricted to the waterbody in which the monitoring point is located

<table>
<thead>
<tr>
<th>Data</th>
<th>Ref. Value</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Kriging</td>
<td>BME</td>
</tr>
<tr>
<td>Chlorophyll a</td>
<td>3.48</td>
<td>3.77</td>
<td>4.20</td>
</tr>
<tr>
<td>MRP</td>
<td>1.58</td>
<td>2.01</td>
<td>1.81</td>
</tr>
<tr>
<td>DIN</td>
<td>1.72</td>
<td>1.71</td>
<td>0.78</td>
</tr>
</tbody>
</table>

6.6. Summary and conclusions

In Chapter 5, entropy methods were applied to the simulated datasets generated at each of the existing monitoring points in Cork Harbour using a calibrated and validated water quality model. The process yielded 3 grades of optimised monitoring networks for Chlorophyll a, DIN and MRP, resulting in a total of 9 proposed monitoring network configurations. The proposed monitoring networks were graded with respect to their ability to reproduce the total monitoring network transinformation that would be provided were the entire existing monitoring network used; the 3 grades were 25, 50 and 75% with the increasing percentage of total monitoring network transinformation indicating the inclusion of more of the original monitoring points. Of the 24 monitoring points in Cork Harbour, the number of monitoring points retained in each of the 9 proposed downsized monitoring networks ranged from 4 to 21.
In order to counteract a lack of information from the reduced number of monitoring points, the feasibility of two geospatial interpolation techniques was investigated, namely kriging and BME. BME is a relatively new development in geostatistics, first presented by Christakos (1990) which allows probabilistic data to be incorporated into the traditional interpolation environment. BME presents a unique opportunity to incorporate the same calibrated and validated detailed water quality model which was used previously in Chapter 5 into the probabilistic estimation of water quality parameters.

The research contained in this chapter involved the evaluation of the suitability of BME in the estuarine environment for geostatistical estimation of 3 parameters used by the Irish EPA as part of trophic status assessment. The matter of identifying the scale of covariance stationarity was considered in addition to the possible benefit of log transformation of skewed datasets.

The potential benefit of interpolation in water quality monitoring is clear in the context of monitoring network optimisation. In this chapter, kriging interpolation of the monitoring data at the retained monitoring points was used to estimate water quality data at the removed monitoring points. Subsequently, BME was used to estimate the same monitoring data. The monitoring data at the retained monitoring points was supplemented with probabilistic data derived from the regression of numerical model simulation data against EPA monitoring data.

Estimation was carried out using the standard dataset and the log transformed dataset. Covariance was considered and modelled on 3 bases: assuming autocorrelation was stationary at the scale of Cork Harbour, at the scale of each of the constituent waterbodies, and stationary centred on each monitoring point with covariance characteristics largely the same but gradually varying.

As a control scenario upon which to compare the benefit to BME estimation, kriging estimation proved to be reasonably successful in terms of estimating water quality parameters. BME estimates of the monitoring data at the removed monitoring points were more accurate than kriging in a selection of combinations of waterbody, monitoring network and data format. There was no pattern of suitability of BME over kriging however. Comparing the overall shape of the dataset with respect to
range, skewness and kurtosis, BME estimates of the three water quality parameters using the standard dataset and the largest of the three potential optimized monitoring networks resulted in dataset parameters most closely matching that of the original dataset.

The resulting datasets from a collection of estimation configurations were used to assess the trophic status of the component waterbodies in Cork Harbour. The status returned by the datasets indicated that BME estimation underpinned by the monitoring data in standard format from the largest proposed optimised monitoring network with 75% total monitoring network transinformation consistently returned the same water quality status that would be returned if the original dataset was used in its entirety.

To conclude the assessment of BME and kriging in the estuarine environment, the effect of assuming quasi-stationary covariance was considered, such that covariance parameters such as the sill and spatial and temporal range were gradually changing from one monitoring point to another but consistent within waterbodies. In order to test the veracity of this approach, a leave-one-out cross validation of BME and kriging estimation of the entire dataset of daily averaged monitoring data was carried out, yielding promising results. The dataset consisted of 266 Chlorophyll_a points, 244 DIN data points and 271 MRP points. As before, the datasets were processed in their standard and log transformed format.

Covariance models were derived assuming quasi-stationarity centred on each monitoring point but two cases were assessed, where monitoring data used for covariance modelling was taken from the entire Cork Harbour area or secondly from the specific waterbody where the monitoring point was located. There was a considerable increase in the accuracy of the cross validation estimates derived assuming quasi-stationarity, with the highest $R^2$ value returned by the 4 estimation configurations exceeding 0.5 in the majority of cases with some values close to unity. The leave-one-out cross validation process is incompatible with the optimised monitoring networks yielded by Chapter 6 due to the requirement of ongoing monitoring data to derive the covariance models centred on each monitoring point, which would not be the case if a monitoring point was removed due to redundancy.
In summary, the results of the leave-one-out cross validation supports the potential for application of BME and kriging to the estimation of water quality between sampling events. BME and kriging facilitate the estimation of water quality parameters outside of the monitoring network or between sampling events, providing a probabilistic representation of the likely range of values.

The covariance characteristics in Cork harbour were deemed to be quasi-stationary, and using covariance models derived from datasets restricted to zones adjacent to each monitoring point returned the most accurate results, hence corroborating the assertion that Cork harbour is quasi-stationary.

The accuracy of BME and kriging estimates was lower when stationarity was assumed at the scale of individual waterbodies. The most accurate results in the configuration were derived by using the largest of the three downsized networks given by the 75% TST optimum network. The moderate contraction of the existing monitoring network, with a tailored monitoring network for each parameter, yielded satisfactory results, returning the same trophic status class that would be obtained by using the existing monitoring regime employed by the EPA.

Merging the underlying monitoring data that was used in BME and kriging estimation with the estimates produced in each of the various combinations of monitoring network size and estimation type reproduced the original monitoring dataset to a reasonable extent when data was processed without log transformation, considering the characteristics of skewness, kurtosis and maximum and minimum values.

BME and kriging have the potential to complement water quality monitoring in the estuarine environment, by assimilating monitoring data and simulated water quality data generated by a model to generate probabilistic estimates of water quality parameters at a spatiotemporal location. The greatest results are generated in the context of maintaining a pre-existing monitoring network, although promising results were returned by geostatistical estimates derived from a contracted monitoring network. Greater accuracy of BME would be expected with further improvement of the accuracy of the water quality model, which would result in narrower variance values for soft data PDFs which are processed by BME.
Chapter 7. Summary and conclusions

7.1. Summary

This thesis addresses shortcomings in the analysis and design of water quality monitoring networks in tidal waters. Guidelines for the design of water quality monitoring networks have traditionally focused on groundwater and inland surface waters with little consideration given to the design of water quality monitoring programmes for transitional and coastal waters which are the end point for point and diffuse pollution from anthropogenic sources, in addition to marine nutrient fluxes from neighbouring coastal waters. Guidance presented on water quality monitoring network design and optimisation in coastal and transitional waters is generic and qualitative. No single methodology exists which facilitates the design of a monitoring network in coastal and transitional waters for optimising the locations for monitoring and the sampling frequency. Throughout the research, model simulated water quality datasets and monitoring data provided by the Irish EPA for Cork Harbour were used side by side or in combination in order to optimise monitoring networks and explore the effects of adopting optimised estuarine water quality monitoring networks on the outcome of water quality status assessments.

Chapter 2 set the context for the research described later in the thesis. A review is presented of water quality monitoring activities carried out in Ireland and the elsewhere in Europe under the WFD and the OSPAR agreement. The established statistical methods for designing monitoring networks were described, without reference to the marine environment, as well as the qualitative guidance on monitoring network design in general and that which applies specifically to the coastal environment. A number of novel methods of interrogating existing monitoring networks and designing monitoring networks were outlined. A review of the potential for integrating water quality monitoring and modelling was also given. The theory behind kriging and BME geostatistical interpolation is described; also the equations which govern BME and the three stages of BME estimation are explained. Detailed derivations of the solutions to BME are presented in Appendix 2.
Chapter 3 describes the theory behind the hydrodynamic modelling of coastal waters in 0D, 1D, 2D and 3D models as well as the modelling of phytoplankton, dissolved oxygen modelling and the nitrogen and phosphorus cycles which relate to the trophic status of water quality. The solution scheme which is used by the numerical model is described in detail and a short discussion of stability, accuracy and convergence is given.

Chapter 4 contains the methodology for application of entropy methods and BME to the transitional and coastal water quality datasets and model outputs; the framework which ties together the various aspects of research contained in this thesis in order to iteratively update both monitoring regimes and the accuracy of model results in a reciprocal manner is presented here.

In Chapter 5, entropy analysis methods were applied to both water quality sampling data and water quality datasets simulated by a calibrated and validated water quality model. A comparison between the marginal entropy scores derived for each monitoring point from each type of data proved to be a good diagnostic tool for identifying monitoring points where further monitoring or modelling was required. In the cases of MRP and Chlorophyll_a, the marginal entropy score of the high resolution simulated dataset consistently exceeded the score derived from the associated monitoring data which was sampled on 4 days of the year. This result was expected, because the marginal entropy score of a probability density function is proportional to the variance of the dataset. The difference in the scores indicates the PDFs of simulated data described a much larger variation in water quality conditions. When the same comparison was applied to DIN data, the entropy score of the monitoring data exceeded that of the simulated data, but the unusual observation was limited to the North Channel area. This observation corroborated EPA reporting prior to the commissioning of Carrigrennan WWTP where it was concluded that insufficient data on the DIN inputs to the North Channel from the Owenacurra Estuary hindered a full understanding of water quality. The entropy scores converged the closer each monitoring point was relative to the coastal waters.

Entropy methods were also considered as a means of establishing the level of redundancy between existing monitoring points such that the monitoring points in the existing monitoring network could be prioritised for each water quality...
parameter. Entropy methods provided a means of ranking both the existing monitoring network and all potential monitoring points spatially, and the sampling frequency for each optimum monitoring network that minimised the redundant information between consecutive samples was established.

The number of monitoring points required to capture 25, 50 or 75% of the total monitoring network transinformation varied depending upon the parameter and whether data was used from a summer or winter simulation. The spatial priority for each water quality parameter varied, although the optimum monitoring networks per each parameter were largely the same irrespective of whether a summer or winter simulation was used. Less monitoring effort was required for approximately half of the spatially and temporally optimised monitoring networks. Trophic status assessment was carried out using a simulated dataset for Chlorophyll_a, DIN and MRP using the existing monitoring regime and each of the proposed monitoring regimes. Of the 7 waterbodies assessed, 5 waterbodies returned the same trophic status with the optimised monitoring network that was returned when the existing monitoring regime used.

Chapter 6 set out to address perceived shortcomings in modifying the coverage of the water quality monitoring network. The second area of research that was pursued as part of this thesis was the assessment of the feasibility of BME and kriging interpolation to estimate water quality parameters at each of the removed monitoring points that were deemed redundant by the entropy methods in Chapter 5. Covariance gives a measure of the change of autocorrelation in space or time; both BME and kriging required an understanding of spatial and temporal autocorrelations, which are captured by a covariogram which is fitted to raw covariance data at a range of spatial and temporal lags. The two factors which were explored in the assessment of the suitability of BME and kriging were the effect of log transformation of each dataset and the scale and type of covariance stationarity.

The prominent discussion point in Chapter 6 was whether covariance could be assumed stationary on the scale of the extent of Cork Harbour, or its constituent waterbodies, or whether covariance characteristics were not entirely stationary but very gradually varying spatially, a characteristic known here as “quasi-stationary”.

The number of monitoring points that were retained by the 3 grades of optimised
monitoring networks in chapter 5 varied from 5 to 22 out of a total of 24. As a consequence, for each parameter there were sufficient monitoring points to assemble a covariance model that assumed stationarity on the scale of the entire Cork Harbour, but an insufficient number of monitoring points to derive covariance models assuming stationarity on the scale of individual waterbodies. The only exception was the largest optimised monitoring network (75%) which provided sufficient monitoring points to develop a covariance model describing Chlorophyll_a in the North Channel. The results in this case proved promising; assuming covariance was stationary within waterbodies resulted in BME estimate PDFs which were more informative than kriging, due to the PDFs consistently capturing the cross validation data points as well as the temporal trend. BME and kriging proved to be reasonably successful in estimating timeseries of water quality data at the redundant monitoring points. Although different levels of covariance stationarity were valid to the extent that BME and kriging estimates were reasonably accurate, assuming that covariance was quasi-stationary resulted in the greatest accuracy in BME and kriging cross validation.

7.2. Critique of research output

There is no methodology by which a monitoring network can be designed for the transitional and coastal environment without input from experts at the outset on the initial location of monitoring points.

The greater Cork harbour area was selected as a case study for the research in this thesis, due to its complex geometry with regions of restricted exchange and long residence times. The waters of the area historically suffered from high nutrient loading and subsequent issues with eutrophication manifesting as phytoplankton blooms. The waters which received the nutrient-enriched riverine inputs would initially present with high nutrient concentrations whilst further downstream, primary production would lead to uptake of these nutrients and a resultant drop in dissolved inorganic phosphorus and nitrogen. The inner harbour downstream of the phytoplankton bloom hence was lower in nutrients and phytoplankton and the outer harbour would have better trophic status again. Thus, the challenge in establishing an optimum monitoring network for each trophic status parameter in the Cork harbour
Summary and conclusions

area lay in, firstly, identifying the discrete sub regions within Cork harbour where a particular parameter was an issue, without clustering monitoring points in each problem area as additional points were added to the network. As the monitoring network would grow, additional points may be added to eventually create a clustering effect, but not before each problematic area was included first.

7.2.1. Entropy analysis

Applying entropy analysis to high resolution simulated water quality datasets prioritised the existing monitoring network, placing a different priority on each of the water quality parameters of concern. This indicates the possibility that using the same monitoring network and sampling regime to determine trophic status may be inefficient and unnecessary. The prioritised areas and monitoring points were corroborated by reporting on water quality for the assessment periods in question. Detailed entropy analysis in Cork Harbour using simulated water quality datasets to rank each numerical model grid cell in the entire domain supported the locations of the existing monitoring network; “hot spots” were highlighted around existing monitoring points. Thus it can be concluded that applying entropy methods to simulated datasets from a calibrated and validated water quality model could identify good monitoring locations.

Using the optimised monitoring network and sampling frequency for each water quality parameter resulted in the same trophic status category in 5 out of the 7 waterbodies in Cork Harbour, with the exceptions being the Owenacurra Estuary and the Inner Cork Harbour. In the case of the Owenacurra Estuary, all of the existing monitoring points were required to match the trophic status obtained using the existing monitoring network, while in the case of the Inner Cork harbour, the smallest of the four proposed optimum monitoring networks was the only one which did not return the trophic status determined by the existing monitoring network.

The existing monitoring regime applied for Cork harbour involves sampling at three times during the summer and once during the winter. Spatially, monitoring was carried out at 24 locations within the numerical model domain. Amalgamating the monitoring points nominated by the 25% TST networks for each parameter together on a season by season basis, sampling locations LE180, LE340 and LE620 do not
feature in either the summer or winter networks. In addition, LE310, LE330, LE350 and LE630 are unused in any of the 25% TST winter networks. All 24 monitoring points feature in at least one of the summer 50% TST monitoring networks, whilst LE330 and LE620 are unused in any of the 50% TST winter networks. All 24 monitoring points feature in at least 1 of the 75% TST summer and winter networks.

There was no change in the trophic status returned by sampling at weekly resolution, hence justifying the present sampling frequency which is approximately monthly during the peak season for primary production. Using the optimum network in question would still represent less overall monitoring effort for the entire year. The ability of BME and kriging to estimate water quality data at the redundant monitoring points indicates that the retained monitoring network contains sufficient information for estimation elsewhere in Cork Harbour.

A comparison of the marginal entropy scores from simulated water quality model datasets and monitoring data at each monitoring point indicated that the marginal entropy derived from DIN monitoring data exceeded the marginal entropy of the simulated DIN data from the numerical model in the North Channel and Owenacurra Estuary. This result indicated that the monitoring data captured more information and variation in DIN concentrations than the modelled dataset. Reporting at the time of monitoring queried the accuracy of knowledge on DIN loadings to that waterbody at the time. Comparison between the marginal entropy scores of monitoring data and numerical model data could prove beneficial as a means of identifying potential exceedances of licence loadings in the case where the model is forced with accurately determined riverine loadings and tidal boundary fluxes, and outfalls which are assumed to emit at most the licence discharge limit values. Hence, at locations where discrepancies were observed, further investigations could be carried out.

7.2.2. BME and kriging estimation

The BME estimation procedure honours the value of sampled water quality data, which always takes precedent over model output as it is appointed a probability of 1, while supporting assessment where data is sparse by incorporating probabilistic representations of model outputs in the form of probability density functions. Thus,
Summary and conclusions

BME introduces the possibility that a well validated water quality model could be incorporated in operational trophic status assessment.

Although not considered in this work, BME has the potential to assimilate a variety of different PDF shapes and types. For example, interval data could be described as a flat PDF with a uniform probability density between the upper and lower limits of an interval. Therefore, BME could assimilate data that would be otherwise redundant due to laboratory detection limits by using such a flat distribution.

Both of the interpolation techniques proved useful in the estimation of water quality data at removed monitoring locations. Kriging estimates provided a benchmark upon which to compare BME estimation, which was used to incorporate the data generated by a validated high resolution water quality model that was expressed in terms of PDFs to describe the likely range of EPA monitoring data. In a number of cases, BME presented an improvement upon the kriging estimate. The accuracy of BME estimates was hindered by dispersed PDF representations of model data in terms of EPA monitoring data in both the standard and log transformed regression procedures.

Nonetheless, the posterior PDFs generated by BME at each cross validation data point at redundant monitoring points proved informative, with the 68% confidence intervals capturing the redundant data points in most instances and the mean value of the posterior PDFs often falling very close to the cross validation data point.

There was a variation in the accuracy of BME and kriging, which was a result of the linear regression of model data against the water quality data; the width of the variance of the BME soft data PDFs introduces noise which assists in cases of data sparsity, while kriging is more accurate with abundant water quality data. Improved model accuracy would increase the veracity of BME by reducing the noise that is introduced to the PDFs. Kriging would be undermined by insufficient hard data points. Repeated monitoring would increase the accuracy of both BME and kriging in the short term, while long term improvement in model accuracy would allow greater emphasis on BME for water quality estimation, supplemented by an optimised monitoring network.
The number of data points in each leave-one-out cross validation exercise varied in the 250-300 range; 1 point at a time was estimated, with the process collating the remaining 250-300 monitoring data points and the additional probabilistic representations of water quality data derived from the model. The accuracy of the leave-one-out cross validation of BME suggests that BME and kriging could be worthwhile for operational water quality data assimilation; the maximum values of $R^2$ from the range of configurations for each waterbody consistently exceeded 0.5 and in many cases fell in the range of 0.8 to 1. BME could be used to derive additional estimates throughout Cork Harbour away from existing monitoring points whilst retaining the current un-optimised monitoring network. This would be predicated on ongoing updating of the Cork Harbour model inputs.

The manner in which kriging and log kriging estimates of DIN for the three optimal DIN monitoring networks outperformed their BME equivalents is noteworthy; this is borne out in particular by the histograms comparing the original dataset with the combined dataset consisting of estimates and retained monitoring data. One of the main causes for this may be the inaccurate DIN loadings in the model highlighted in the comparison of marginal entropy values in Chapter 5. Hence, the soft data derived from linear regression of model and monitoring data resulted in erroneous soft data which lead to a deterioration in the quality of BME estimates relative to kriging, irrespective of whether data was in standard or log form.

Quasi-stationary BME could be used to assimilate water quality monitoring data and numerical model simulated data on an ongoing basis in light of new monitoring data, whether or not the existing monitoring network was scaled back; at each monitoring point, the centred covariogram for each monitoring data would allow for probabilistic estimation of water quality data between sampling events.

The success of BME and kriging interpolation was based on accurate characterisation of spatial and temporal autocorrelation as described by covariance models describing raw covariance data. The spatial range of covariance varied between water quality parameters and between waterbodies. The range of covariance for each parameter and waterbody is the limit of autocorrelation of each parameter; consequently, its value can be interpreted as a proxy of the radius of the circular region represented by each monitoring point.
Summary and conclusions

A detailed analysis of marginal and transinformation entropy between monitoring points and monitoring events can then be carried out on simulated datasets from the improved numerical model. The monitoring regime can thus be modified in light of the results of the new entropy analysis, and where monitoring points have been removed, BME estimates can be derived at the redundant monitoring locations, either for compliance purposes if the network is redesigned during a multiannual assessment period, or for trend analysis. Increased veracity of the model through repeated recalibration and revalidation hence increases the credibility of the optimised monitoring network and the BME estimates, as the soft data that is utilised in the BME estimation procedure will have a narrower variance and subsequently less noise will be introduced into the BME estimate.

It has been shown that using entropy methods, it can be useful to initiate monitoring on a broad basis and subsequently contract the monitoring network in light of the results of monitoring optimisation. At locations where monitoring might be suspended because of the outcome of monitoring network optimisation, BME estimates of water quality data timeseries may provide an indication of the probable range of values at that location.

The research presented in this thesis introduces the possibility of a reciprocal improvement in the representativeness of the water quality monitoring locations leading to increased accuracy of the water quality model, the results of which would underpin further monitoring network optimisation and long term efficiencies.

7.3. Conclusions

The research detailed in this thesis has a number of implications for the design and optimisation of water quality monitoring networks in transitional and coastal waterbodies. A number of noteworthy observations and findings were identified in this thesis and are detailed below.

7.3.1. Entropy methods

The application of entropy methods to the task of water quality monitoring network optimisation resulted in a number of conclusions:
Summary and conclusions

- A comparison of the marginal entropy scores of the PDFs of water quality monitoring data and numerical model water quality simulation data yielded an insight into how well the current EPA monitoring regime captured the underlying water quality characteristics in Cork Harbour. Comparison between the marginal entropy scores of monitoring data and model data could prove beneficial in an iterative cycle of model recalibration and revalidation in light of ongoing EPA water quality monitoring. The process would entail initial comparison of marginal entropy scores, followed by investigation into monitoring points where there is a large difference between marginal entropy scores. Investigation might involve querying of monitoring data and lab methods, the river discharge values in the model or the nutrient loadings from the rivers or outfalls. The model loadings would be then modified and the model rerun. Model tuning parameters related to hydrodynamics or biogeochemistry may subsequently require modification. The comparison has the potential to highlight the requirement for further monitoring, query model accuracy, or highlight regions where updated numerical model inputs are required.

- The measurement of transinformation entropy between monitoring points for each parameter proved an effective means of ranking all monitoring points. A different priority was placed on each monitoring point depending upon the water quality parameter. This indicates the possibility that using the same monitoring network and sampling regime to determine trophic status may be inefficient and unnecessary. Therefore, the optimum theoretical approach to consider in monitoring for all water quality parameters would entail monitoring at the locations which are common to the 75% TST networks for each parameter. The optimum sampling frequency for each season would be chosen from the range of optimum sampling frequencies nominated for each parameter. Those monitoring points which are unique to a water quality parameter could be estimated using BME or kriging based on all monitoring data collected from the 75% TST monitoring.

- The locations which were prioritised for monitoring each water quality parameter from model analysis were areas of concern in national reporting. It
Summary and conclusions

is concluded that applying entropy methods to simulated datasets from a calibrated and validated operational water quality model can identify the locations which require monitoring.

- Additional transinformation analysis could be used to nominate a monitoring network from the ranking of grid cells. The process would entail determining the marginal entropy of all grid cells of a particular rank; to distinguish between each of the grid cells of equal rank, a higher resolution timeseries may be necessary for each cell. The location in the range of equally ranked cells with the highest marginal entropy would be added to the monitoring network. The transinformation between that point and all other cells would then be determined, and the cells with all equally low transinformation with the point added to the network would then be included in a more detailed transinformation analysis. The process would ultimately be computationally intensive. In the case of Cork Harbour, over 90,000 cells would be iteratively considered, and a cut-off number of monitoring points would be necessary to limit computational work.

- Contraction of the monitoring network for each of the three water quality parameters to the 25%, 50% or 75% grade of optimised monitoring networks had a negligible effect on the outcome of water compliance assessment, assuming that each of the 24 existing monitoring points were representative of water quality in the surrounding waters. As a conservative approach, reducing the monitoring network for each parameter to the respective 75% TST network would be adequate in the short term. With repeated refinement of the model with respect to loadings and calibration, it would be expected that the 25% TST network would be sufficient.

- Adopting the 75% TST monitoring networks both summer and winter would involve 19% less sampling throughout the year, with less sampling in summer and more in winter; an attendant financial saving of 19% would be made assuming that sampling and analysis costs are equal irrespective of the water quality parameter.

- Entropy analysis methods enable the identification of the sampling frequency at which the transinformation or data redundancy is lowest. The most suitable
frequencies at weekly and monthly resolution for each parameter and monitoring network size were identified. Although the current monitoring regime in Cork Harbour entails sampling three times during the summer at approximately monthly frequency and once during the winter when primary production is lowest, there would be no benefit from sampling at weekly frequency or multiples thereof for compliance purposes.

One of the main objectives of this thesis was to establish a methodology for monitoring in estuarine environments. In light of the above points, the results in Chapter 6 suggest that the monitoring methodology proposed in Chapter 5 is suited to the nonhomogeneous, dynamic, marine environment. The monitoring methodology was able to prioritise existing monitoring points, establish the veracity of the numerical model in advance of applying the numerical model to the task of detailed spatial and temporal optimisation of the existing monitoring network, by prioritising the monitoring points spatially and identifying the optimum monitoring frequency as the monitoring network contracted. A detailed high resolution study of all likely monitoring locations in Cork harbour largely corroborated the existing monitoring network, with additional monitoring locations suggested only in the North Channel and the east end of Inner Cork Harbour.

7.3.2. BME

BME was adapted to the Cork Harbour area, and the monitoring data at the redundant monitoring points nominated by entropy methods were cross validated. The following observations and conclusions were drawn from the results of BME cross validation of three of the four water quality parameters related to the main trophic status parameters:

- BME and kriging facilitated the estimation of water quality parameters outside of the optimised monitoring network, providing a probabilistic representation of the likely range of values. $R^2$ values for BME and kriging estimation of water quality varied from almost zero to unity. No pattern emerged with respect to which estimation method was more suited to each waterbody. Nonetheless, in a number of combinations of monitoring network
size, parameter and waterbody, BME offered an improved estimate of water quality data at removed monitoring points. In a number of cases where insufficient data precluded kriging from estimation, BME was able to generate a probabilistic estimate using model data only. Kriging provides a good proxy estimate were there insufficient soft data to underpin BME estimation, although kriging itself is limited by insufficient hard data.

- BME and kriging reproduced the original monitoring dataset based upon monitoring data from the prioritised retained monitoring points and water quality model simulated data. BME and kriging estimation using the largest optimised with 75% of total monitoring network transinformation both consistently performed better than the monitoring networks with 25% or 50% of total monitoring network transinformation. Using the optimum network in question would still represent less overall monitoring effort for the entire year, while the success of BME and kriging at estimating water quality data at the redundant monitoring points removed from the network indicates that the retained monitoring network contains sufficient information for estimation elsewhere in Cork Harbour.

- Covariance as a measure of spatial dependency was considered on 3 different scales in the application of kriging and BME. Where sufficient data allowed, the covariance models of optimal fit for each waterbody varied, thus bringing into doubt the validity of assuming stationarity at the scale of the entire Cork Harbour domain.

- BME and kriging estimates were most accurate when covariance throughout Cork Harbour was assumed to be quasi-stationary, and in particular by assuming quasi-stationarity at the scale of each waterbody. Adopting an assumption of quasi-stationary covariance is not feasible in the case of estimating water quality parameters at redundant monitoring points as data at the monitoring point is a basic requirement of quasi-stationary covariance modelling where covariance calculations are centred spatially on each monitoring point. Quasi-stationary BME and kriging estimates would be best utilised to derive probabilistic estimates of water quality between sampling events.
The trophic status assessment criteria that are applied to water quality monitoring data every 3 years by the Irish EPA were applied to the combinations of estimation types (kriging or BME), optimum monitoring network size (25, 50 and 75%) and data type (standard or log transformed) for the assessment periods 1995-1999 and 1999-2003. Irrespective of the data type, using the largest optimum monitoring network size (75%) for each water quality parameter to estimate at the redundant monitoring point using BME and kriging returned the same trophic status result that would be returned if the TSAS criteria were applied to the entire water quality dataset without BME or kriging estimates. Therefore, even with a modest contraction of the existing monitoring network, the same trophic status was returned when interpolation was used.

The second objective of this thesis was to establish a reliable estimate of water quality parameters at unmonitored locations, to complement the optimum monitoring network which would be nominated as a consequence of meeting the first objective.

The monitoring data provided by each of the three levels of optimum monitoring networks derived as part of the first objective were assimilated by BME and kriging to derive probabilistic estimates of water quality at the redundant monitoring points. The results generated in order to meet each objective provide a basis for the application of both methodologies in unison within the proposed framework in order to build ongoing synergies between monitoring and modelling activities.

7.4. Future Work

The research detailed in this thesis validates two complementary methods, the first for optimal water quality monitoring network design, and the second for probabilistic estimation of water quality data. Entropy methods provide a means of designing a water quality monitoring network for transitional and coastal waters, utilising water quality data and simulated datasets from a calibrated and validated water quality model. The perceived loss of information in relocating or removing monitoring points within monitoring networks is offset by the ability of BME to process water quality data at retained monitoring points and probabilistic data.
Summary and conclusions

derived from regression of collocated water quality data and model simulated data to produce PDFs to estimate timeseries of monitoring data at the redundant monitoring points.

The author recommends the following points for further consideration to build upon the research presented in this thesis:

- Frequent recalibration of the water quality model of Cork Harbour with up to date water quality datasets to improve the accuracy of the probabilistic data used for BME and the application of entropy methods to monitoring network optimisation. This would allow implementation of the methodology outlined in this thesis on a trial basis.

- Inclusion of water quality sensor data in the BME estimation process as an additional source of soft data, instead of restricting the BME estimation to water quality monitoring data and model data.

- Incorporate live sensor data and SCADA information from wastewater treatment plants into the model to better calibrate and validate the ambient response to end-of-pipe emissions.

- The spatial and temporal range of covariance provides a proxy measure of the spatial extent and the length of time for which a single sampling event is representative. The influence of the rates of growth and decay of primary production and residence time on the spatial and temporal range of covariance of each trophic status parameter has not yet been elucidated. Were these questions addressed, additional monitoring points could be added to the monitoring network at times when hydrological or tidal conditions significantly influenced residence time, and hence, the spatial or temporal zone represented by a single point. An increased growth rate for primary production may lead to high consumption of nutrients, resulting in proliferation of phytoplankton and gradients in nutrient concentrations due to consumption which may shorten the spatial range represented by a monitoring point.

- Cork Harbour has a complex geometry and zones of restricted exchanges of water. Consequently, the relatively high variations in covariance characteristics in Cork Harbour might be expected. BME could prove to be
Summary and conclusions

even more accurate in open bays such as Dublin Bay, Galway Bay or the Shannon Estuary. In such cases where covariograms might represent larger areas due to simpler geometry, there may be greater scope to optimise the monitoring network, reduce the effort required to monitor through sampling at a lower frequency or at fewer monitoring points, and increase the benefit of water quality models.

- Bayesian parameter estimation would enable quicker model calibration and validation (Arhonditsis et al. 2008, Borsuk and Stow 2000, Gronewold et al. 2009). The approach involves construction of joint probability functions of parameters conditioned on observed data. As a result, marginal PDFs can be derived to determine the best parameter values. For example, in the case of Borsuk and Stow (2000), a first order model of BOD decay was assumed at first, but through the process of bayesian parameter estimation, it was established that a mixed order model with an exponent between 1.8 and 4 was more suitable, depending upon conditions. A first order BOD decay model would be adopted for simplicity, but without any logic or regard for the underlying process.
Appendix 1. DIVAST Solution

The continuity equation in Equation (1.37) considers that the mass entering a finite space must be equal to the mass leaving that space. The $X$ or $Y$ direction momentum equations as presented in Equations (1.3!8a) and (1.38b) contain the momentum terms on the left hand side from local and advective accelerations, whilst the terms on the right hand side represent the momentum of the causative forces which consist of the Coriolis affect, pressure gradients, wind shear stress, bed resistance and turbulence. Reynolds stresses in the vertical plane are approximated after Boussinesq.

The finite difference solution of the continuity and momentum equations employs an alternating direction implicit (ADI) scheme. The central differences method of expressing derivatives is utilised. The ADI scheme entails the division of each timestep into two half-timesteps. Each half-timestep considers implicitly the solution of the water depth and velocity component in one direction whilst the water depth and velocity component on the other direction are considered explicitly, thus avoiding the complexity of solving a two dimensional matrix at each timestep.

The grid system which is used is space staggered orthogonal. Water elevations are computed at the centre of each grid cell whilst the velocities and water depths in the $x$ and $y$ directions are solved at the centre of the sides of the grid cells.

The hydrodynamic equations are solved using Gaussian elimination and back substitution (Falconer 1977). The equations are rearranged with the unknown variables on the left hand side, and the known variables from the previous timestep on the right hand side. All known data within the formulae are grouped into recursion coefficients and the formulae reduce to simplified recursion formulae.

The finite difference representations of the continuity equation, the momentum equation and the advection diffusion equation are solved in the $x$ direction for the first half timestep and the $y$ direction for the second timestep.
**Table of parameters**

- $u, v = x & y$ direction velocity
- $U, V = x & y$ direction depth integrated velocity
- $\zeta = \text{water elevation above/below mean water level}$
- $q_x, q_y = \text{depth integrated volumetric flux in x and y direction} \ (q_x = UH, q_y = VH)$
- $\beta = \text{momentum correction factor for non-uniform vertical velocity profile}$
- $f = \text{Coriolis parameter}$
- $C = \text{Chezy bed roughness coefficient}$
- $v_t = \text{depth-average mean eddy viscosity}$
- $\rho_a = \text{density of air (1.292 kg/m}^3\text{)}$
- $\rho = \text{fluid density}$
- $\varepsilon = \text{depth mean eddy viscosity}$
- $g = \text{gravitational acceleration}$
- $W_x, W_y = \text{wind velocity in x and y direction}$
- $C^* = \text{air-water interface coefficient}$
- $H = \text{total water depth} \ (H = h + \zeta)$
- $\phi = \text{...}$
- $D_{xx}, D_{xy}, D_{yx}, D_{yy} = \text{x and y direction dispersion coefficients}$
- $\Phi_{\phi} = \text{...}$
Continuity equation

\[
\frac{\partial \zeta}{\partial t} + \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} = 0
\]  

(A1.1)

Continuity equation in finite difference format:

\[
\zeta_{i,j}^{n+1} - \zeta_{i,j}^{n} + \frac{\Delta t}{2\Delta x} \left[ q_x|_{i+\frac{1}{2},j}^{n+1} - q_x|_{i-\frac{1}{2},j}^{n+1} + q_y|_{i,j+\frac{1}{2}}^{n+1} - q_y|_{i,j-\frac{1}{2}}^{n} \right] = 0
\]  

(A1.2)

Recursive continuity equation

\[
-d_i q_x|_{i-\frac{1}{2},j}^{n+1} + e_i \zeta_{i,j}^{n+1} + f_i q_x|_{i+\frac{1}{2},j}^{n+1} = A_i^n
\]  

(A1.3)

where

\[
d_i = f_i = \frac{\Delta t}{2\Delta x}
\]  

(A1.4)

and

\[e_i = 1\]

(A1.5)

and

\[A_i^n = \zeta_{i,j}^n - \frac{\Delta t}{2\Delta x} \left[ q_y|_{i,j+\frac{1}{2}}^{n+1} - q_y|_{i,j-\frac{1}{2}}^{n} \right]
\]  

(A1.6)

X direction momentum equation

\[
\frac{\partial q_x}{\partial t} + \beta \left[ \frac{\partial U q_x}{\partial x} + \frac{\partial U q_y}{\partial y} \right] = f q_y - g H \frac{\partial \zeta}{\partial x} + \rho_a C W_x \left( W_x^2 + W_y^2 \right)^{\frac{1}{2}} + \frac{\Delta t}{\Delta x} \left[ U' q_x q_x'_{i+1,j}^{n} - U' q_x q_x'_{i,j}^{n} + U' q_y q_y'_{i+1,j+1}^{n} - U' q_y q_y'_{i+1,j-1}^{n} \right]
\]  

(A1.7)

X direction momentum equation in finite difference format:

\[
q_x|_{i+\frac{1}{2},j}^{n+1} = q_x|_{i,\frac{1}{2},j}^{n} - \frac{\Delta t}{\Delta x} \left[ U' q_x q_x'_{i+1,j}^{n} - U' q_x q_x'_{i,j}^{n} + U' q_y q_y'_{i+1,j+1}^{n} - U' q_y q_y'_{i+1,j-1}^{n} \right]
\]
\[\begin{align*}
+\Delta tf q_y^n|_{i+\frac{1}{2},j} - \frac{g\Delta t}{2\Delta x} H^n_{i+\frac{1}{2},j} \left[ \zeta^{n+\frac{1}{2}}_{i+1,j} - \zeta^{n+\frac{1}{2}}_{i-j,1} - \zeta^{n+\frac{1}{2}}_{i,j} - \zeta^{n+\frac{1}{2}}_{i-j,1} \right] \\
+ \rho_a \Delta t C W_x \left( W_x^2 + W_y^2 \right)^{\frac{1}{2}} \\
\rho \\
- \frac{g\Delta t}{2} \left( U^{n+\frac{1}{2}} + U^{n-\frac{1}{2}} \right) (U'^2 + V'^2)^{\frac{1}{2}}|_{i+\frac{1}{2},j} \\
+ \frac{\Delta t}{\Delta x^2} \epsilon H^n_{i+\frac{1}{2},j} \left[ U'_{i+\frac{1}{2},j} + U'_{i-\frac{1}{2},j} + U'_{i+\frac{1}{2},j+1} + U'_{i+\frac{1}{2},j-1} - 4U'_{i+\frac{1}{2},j} \right]
\end{align*}\]

Recursive X direction momentum equation

\[\begin{align*}
- a_i \zeta^{n+\frac{1}{2}}_{i,j} + b_i q_x^{n+\frac{1}{2}}_{i+\frac{1}{2},j} + c_i \zeta^{n+\frac{1}{2}}_{i+1,j} = B^n_i
\end{align*}\]

where

\[a_i = c_i = \frac{g\Delta t}{2\Delta x} H^n_{i+\frac{1}{2},j}\]

and

\[b_i = 1 + \frac{g\Delta t}{2} \left( q'^2_x + q'^2_y \right)^{\frac{1}{2}}|_{i+\frac{1}{2},j}\]

and

\[B^n_i = q_x^n_{i+\frac{1}{2},j} - \frac{\beta \Delta t}{\Delta x} \left[ U' q_x^n_{i+1,j} - U' q_x^n_{i+\frac{1}{2},j} \right] + \rho_a \Delta t C W_x \left( W_x^2 + W_y^2 \right)^{\frac{1}{2}} \]

\[+ \Delta t f q_y^n_{i+\frac{1}{2},j} - \frac{g\Delta t}{2\Delta x} H^n_{i+\frac{1}{2},j} \left[ \zeta^{n+\frac{1}{2}}_{i+1,j} - \zeta^{n+\frac{1}{2}}_{i-j,1} - \zeta^{n+\frac{1}{2}}_{i,j} - \zeta^{n+\frac{1}{2}}_{i-j,1} \right] + \rho_a \Delta t C W_x \left( W_x^2 + W_y^2 \right)^{\frac{1}{2}} \]

\[- \frac{g\Delta t}{2} \left( q'^2_x + q'^2_y \right)^{\frac{1}{2}}|_{i+\frac{1}{2},j} \]

\[+ \frac{\Delta t}{\Delta x^2} \epsilon H^n_{i+\frac{1}{2},j} \left[ U'_{i+\frac{1}{2},j} + U'_{i-\frac{1}{2},j} + U'_{i+\frac{1}{2},j+1} + U'_{i+\frac{1}{2},j-1} - 4U'_{i+\frac{1}{2},j} \right]\]
U’ is expressed explicitly as the previous U for the first iteration and a centrally averaged value for the second iteration.

\[
U'_{i+\frac{1}{2},j}^n = U_{i+\frac{1}{2},j}^{n-\frac{1}{2}}
\]

\[
U'_{i+\frac{1}{2},j}^n = \frac{1}{2} \left( U_{i+\frac{1}{2},j}^{n+\frac{1}{2}} + U_{i+\frac{1}{2},j}^{n-\frac{1}{2}} \right)
\]

Advection diffusion equation

\[
\frac{\partial H \phi}{\partial t} + \frac{\partial H U \phi}{\partial x} + \frac{\partial H V \phi}{\partial y} = \frac{\partial}{\partial x} \left[ D_{xx} \frac{\partial \phi}{\partial x} + D_{xy} \frac{\partial \phi}{\partial y} \right] + \frac{\partial}{\partial y} \left[ D_{yx} \frac{\partial \phi}{\partial x} + D_{yy} \frac{\partial \phi}{\partial y} \right] + \Phi
\]

Finite difference representation of advection diffusion equation

\[
(\phi H)_{i,j}^{n+\frac{1}{2}} + \frac{\Delta t}{4\Delta x} \left[ q_x |_{i+\frac{1}{2},j}^{n+\frac{1}{2}} \left( \phi |_{i+1,j}^{n+\frac{1}{2}} + \phi |_{i,j}^{n+\frac{1}{2}} \right) - q_x |_{i-\frac{1}{2},j}^{n+\frac{1}{2}} \left( \phi |_{i,j}^{n+\frac{1}{2}} + \phi |_{i-1,j}^{n+\frac{1}{2}} \right) \right]
\]

\[
- \frac{\Delta t}{2\Delta x^2} \left[ (HD_{xx}) |_{i+\frac{1}{2},j}^{n+\frac{1}{2}} \left( \phi |_{i+1,j}^{n+\frac{1}{2}} + \phi |_{i,j}^{n+\frac{1}{2}} \right) - (HD_{xx}) |_{i-\frac{1}{2},j}^{n+\frac{1}{2}} \left( \phi |_{i,j}^{n+\frac{1}{2}} + \phi |_{i-1,j}^{n+\frac{1}{2}} \right) \right]
\]

\[
= (\phi H)_{i,j}^{n} + \frac{\Delta t}{4\Delta x} \left[ q_y |_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} \left( \phi |_{i,j+1}^{n+\frac{1}{2}} + \phi |_{i,j}^{n+\frac{1}{2}} \right) - q_y |_{i,j-\frac{1}{2}}^{n+\frac{1}{2}} \left( \phi |_{i,j}^{n+\frac{1}{2}} + \phi |_{i,j-1}^{n+\frac{1}{2}} \right) \right]
\]

\[
+ \frac{\Delta t}{2\Delta x^2} \left[ (HD_{yy}) |_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} \left( \phi |_{i,j+1}^{n+\frac{1}{2}} - \phi |_{i,j}^{n+\frac{1}{2}} \right) - (HD_{yy}) |_{i,j-\frac{1}{2}}^{n+\frac{1}{2}} \left( \phi |_{i,j}^{n+\frac{1}{2}} - \phi |_{i,j-1}^{n+\frac{1}{2}} \right) \right]
\]

Solute transport

\[
(\phi H)_{i,j}^{n+\frac{1}{2}} = (\phi H)_{i,j}^{n} - \frac{\Delta t}{2\Delta x} \left[ \phi q_x |_{i+\frac{1}{2},j}^{n+\frac{1}{2}} - \phi q_x |_{i-\frac{1}{2},j}^{n+\frac{1}{2}} + \phi q_y |_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} - \phi q_y |_{i,j-\frac{1}{2}}^{n+\frac{1}{2}} \right]
\]

\[
+ \frac{\Delta t}{2\Delta x} \left[ (HD_{xx}) |_{i+\frac{1}{2},j}^{n+\frac{1}{2}} \left( \phi |_{i+1,j}^{n+\frac{1}{2}} - \phi |_{i,j}^{n+\frac{1}{2}} \right) - (HD_{xx}) |_{i-\frac{1}{2},j}^{n+\frac{1}{2}} \left( \phi |_{i,j}^{n+\frac{1}{2}} - \phi |_{i-1,j}^{n+\frac{1}{2}} \right) \right]
\]
\begin{align*}
+ (HD_{yy})_{i,j}^n \left( \phi_{i,j+1}^n - \phi_{i,j}^n \right) - (HD_{yy})_{i,j-1}^n \left( \phi_{i,j}^{n+\frac{1}{2}} - \phi_{i,j-1}^{n+\frac{1}{2}} \right) \\
+ (HD_{xy})_{i+\frac{1}{2},j}^{n+\frac{1}{2}} \left( \phi_{i+\frac{1}{2},j+\frac{1}{2}}^n - \phi_{i+\frac{1}{2},j-\frac{1}{2}}^n \right) - (HD_{xy})_{i-\frac{1}{2},j}^{n+\frac{1}{2}} \left( \phi_{i-\frac{1}{2},j+\frac{1}{2}}^n - \phi_{i-\frac{1}{2},j-\frac{1}{2}}^n \right) \\
+ (HD_{yx})_{i,j+\frac{1}{2}}^n \left( \phi_{i+\frac{1}{2},j+\frac{1}{2}}^n - \phi_{i-\frac{1}{2},j+\frac{1}{2}}^n \right) - (HD_{yx})_{i,j-\frac{1}{2}}^n \left( \phi_{i+\frac{1}{2},j-\frac{1}{2}}^n - \phi_{i-\frac{1}{2},j-\frac{1}{2}}^n \right)
\end{align*}
Appendix 2. Kriging and BME derivation

The following derivation is taken with the kind permission of Marc L. Serre from his Ph.D thesis (Serre 1999). Derivations have been expanded and explained in further detail.

In the derivation of kriging and BME estimators assumes the following nomenclature:

- The spatiotemporal random field is known as $X(p)$, where $p = (s, t)$
- $x_i$ refers to the random variable $x$ at point $i$ ; $i = 1, 2, ..., m$
- $\chi_i$ refers to a realisation of the random variable $x$ at point $i$
- $x = [x_1 x_2 ... x_m]^T$ refers to the random variables at the $m$ points in the domain
- $X = [\chi_1 \chi_2 ... \chi_m]^T$ is vector of realisations at the $m$ points in the domain
  - $X_{\text{data}} = [X_{\text{hard}} X_{\text{soft}}]^T$ is the vector of realisations of hard data with a probability of 1 and soft data with associated probability distributions
  - $X_{\text{hard}} = [\chi_1 \chi_2 ... \chi_m]^T$ refers to the realisations at the $m_h$ hard data locations
  - $X_{\text{soft}} = [\chi_{m_h+1} ... \chi_m]^T$ refers to the realisations at the $|m - m_h|$ soft data locations
  - $X_{\text{map}} = [X_{\text{data}} \chi_k]^T$ is the vector of known realisations at all data points at the unknown realisation which is to be estimated at point $k$

Simple Kriging

Simple kriging, as established by Krige and developed by Matheron (Krige 1953, Matheron 1963), is based on a number of assumptions:
Kriging and BME derivation

- The generated estimate is a linearly weighted combination of the $m_h$ pre-existing observed values,
- Unbiasedness of the estimate is assumed. The expected value of the estimate is assumed to be equal to the expected value of the expression for the estimator,
- Mean squared error is minimised,
- The mean trend $m_x(p)$ and the estimated covariance model $c_x(p, p')$ are both known,
- Only hard, certain data is available ($P=1$ for each data point).

The estimate of the realisation $\chi_k$ of the random variable $x_k$ is assumed to be described by the following expression, where $\lambda_0$ and $\lambda$ are the linear kriging weights which must be determined, hence satisfying the assumption of linear weighting:

$$\chi_k^* = \lambda_0 + \lambda^T \chi_{hard}$$  \hspace{1cm} (A2.1)  
$$\lambda^T = [\lambda_1 \lambda_2 \ldots \lambda_{m_h}]^T$$  \hspace{1cm} (A2.2)

To satisfy the unbiasedness condition:

$$\overline{\chi_k^*} = \overline{\chi_k}$$  \hspace{1cm} (A2.3)

Applying to equation (A2.1):

$$\overline{\chi_k^*} = (\overline{\lambda_0} + \overline{\lambda^T \chi_{hard}}) = \lambda_0 + \lambda^T \overline{\chi_{hard}}$$  \hspace{1cm} (A2.4)

Rearranging in terms of $\lambda_0$:

$$\lambda_0 = \overline{\chi_k} - \lambda^T \overline{\chi_{hard}}$$  \hspace{1cm} (A2.5)

Hence reinserting equation (A2.5) into (A2.1):

$$\chi_k^* = (\overline{\chi_k} - \lambda^T \overline{\chi_{hard}}) + \lambda^T \chi_{hard}$$  \hspace{1cm} (A2.6)

\[270\]
Note that: \(\overline{x}_k = m_x(p_k)\) i.e. the mean trend interpolated to the estimation point, and \(\overline{x}_{\text{hard}} = m_{\text{hard}} = [m_1 \ldots m_{m_h}]\) which is the vector of mean trend values at all hard data points.

To satisfy the requirement of the estimator being unbiased, the expected value of the estimation error must equal to zero.

\[
e_k = x_k - x_k^* = (x_k - \overline{x}_k) - \lambda^T (x_{\text{hard}} - \overline{x}_{\text{hard}})
\]  
(A2.7)

\[
\overline{e}_k = x_k - x_k^* = 0
\]  
(A2.8)

As \(\overline{e}_k = 0\), the mean square error or variance of estimation error is given by:

\[
\sigma_e^2 = (e_k - \overline{e}_k)^2 = e_k^2
\]  
(A2.9)

Inserting equation (A2.7):

\[
\sigma_e^2 = e_k^2 = (x_k - \overline{x}_k)^2 - 2\lambda(x_k - \overline{x}_k)(x_{\text{hard}} - \overline{x}_{\text{hard}}) + \lambda^T(x_{\text{hard}} - \overline{x}_{\text{hard}})^2
\]  
(A2.10)

Note that:

\[
C_{k,k} = c_x(p_k, p_k) = (x_k - \overline{x}_k)^2
\]  
(A2.11a)

\[
C_{h,h} = c_x(p_h, p_h) = (x_{\text{hard}} - \overline{x}_{\text{hard}})^2
\]  
(A2.11b)

\[
C_{k,h} = c_x(p_k, p_h) = (x_k - \overline{x}_k)(x_{\text{hard}} - \overline{x}_{\text{hard}})
\]  
(A2.11c)

\[
C_{h,h} = (x_{\text{hard}} - \overline{x}_{\text{hard}})^2
c_{x}(p_{m_h}, p_{m_h})
\]

(A2.12)

\[
C_{k,h} = (x_k - \overline{x}_k)(x_{\text{hard}} - \overline{x}_{\text{hard}}) = [c_x(p_k, p_1) \ldots c_x(p_k, p_{m_h})]
\]  
(A2.13)

Hence:

\[
\sigma_e^2 = C_{k,k} - 2C_{k,h}\lambda + \lambda^T C_{h,h}\lambda
\]  
(A2.14)
To minimise mean square error, which is one of the main assumptions of kriging, the partial derivative of mean square error with respect to the kriging weights $\lambda_i$ is determined and equated to zero:

$$\frac{\partial \sigma_e^2}{\partial \lambda_i} = \frac{\partial}{\partial \lambda_i} [C_{k,k} - 2C_{k,h} + \lambda^T C_{h,h} \lambda] = 0 \quad (i = 1, 2, \ldots, m_h)$$ (A2.15)

$$\frac{\partial \sigma_e^2}{\partial \lambda_i} = -2C_{k,h} + 2C_{h,h} \lambda = 0$$ (A2.16)

Solving for $\lambda$:

$$\lambda = C_{h,h}^{-1}C_{h,k}$$ (A2.17)

Thus:

$$\lambda^T = C_{k,h}C_{h,h}^{-1}$$ (A2.18)

Inserting into the original kriging equations:

$$x_k^* = (\overline{x}_k - \lambda^T \overline{x}_{hard}) + \lambda^T \chi_{hard}$$ (A2.19)

$$= \overline{x}_k + C_{h,h} C_{k,h}^{-1}(\chi_{hard} - \overline{x}_{hard})$$ (A2.20)

From equation (above A2.20), the solution to the simple kriging estimation problem at point $k$ is the summation of the mean trend at the estimation point and the covariance matrix for the hard points multiplied by the inverse of the covariance matrix between the estimation point and the hard data points, multiplied by the residual at the hard data points once the mean trend is removed at those points.

Returning to earlier expression (A2.14) for estimation error, and in light of expressions (A2.17) and (A2.18), the expression expands to:

$$\sigma_e^2 = C_{k,k} - 2C_{k,h} (C_{h,h}^{-1}C_{h,k}) + (C_{k,h}C_{h,h}^{-1})C_{h,h}(C_{h,h}^{-1}C_{h,k})$$ (A2.21)
This reduces to:

\[ \sigma_e^2 = C_{k,k} - C_{k,h}C_{h,h}^{-1}C_{h,k} \]  

(A2.22)

**BME**

**Regular Entropy functional S**

The regular Shannon information entropy measure of the unknown PDF \( f_G(\chi_{map}) \) is defined as equation (A2.27):

\[ S(f_x) = \text{Inf}(\chi_{map}) = - \int d\chi_{map} f_G(\chi_{map}) \ln f_G(\chi_{map}) \]  

(A2.27)

The expected value of a range of known statistical moments \( \overline{g_\alpha} \) are defined in equation (A2.28):

\[ \overline{g_\alpha} = \int d\chi_{map} f_G(\chi_{map}) g_\alpha(\chi_{map}); \] \( \alpha = 1, \ldots, n_{con} \)  

(A2.28)

**Modified Entropy functional S_C**

Combining the Shannon information entropy functional with the known statistical moments \( g_\alpha \) and the unknown Lagrange multipliers \( \mu_\alpha \) results in the modified entropy functional A2.29 which must be solved to determine the shape of the maximised entropy PDF \( f_G(\chi_{map}) \) which describes the data domain:

\[ S_C(f_x) = \text{Inf}(\chi_{map}) \]  

(A2.29)

\[ = - \int d\chi_{map} f_G(\chi_{map}) \ln f_G(\chi_{map}) + \sum_{\alpha=0}^{N_\alpha} \mu_\alpha \overline{g_\alpha} [f_G(\chi_{map})] \]  

Entropy is defined as:

\[ \epsilon(f_x) = - \int_{R^{m+1}} d\chi_{map} f_X(\chi_{map}) \ln f_X(\chi_{map}) \]  

(A2.30)
We wish to maximise the expression for Shannon’s information entropy as defined in equation (A2.30) but constrained by utilising known parameters \( g_q \) i.e. mean trend, covariance. Equation (A2.31) describes the set of constraints which are applied. The expected values of the constraints which are known on the left hand side are equated to the expression for the expected value of the statistical moments in a general sense on the right hand side:

\[
\overline{g}_\alpha = - \int_{\mathbb{R}^{m+1}} dX_{\text{map}} f_X(X_{\text{map}}) g_\alpha(X_{\text{map}}), \alpha = 1, \ldots, n_{\text{con}} \tag{A2.31}
\]

Hence the modified entropy functional \( S_c(f_X) \) defined above is developed and renamed \( F[f_X] \) here:

\[
F[f_X] = - \int_{\mathbb{R}^{m+1}} dX_{\text{map}} f_X(X_{\text{map}}) \ln(f_X(X_{\text{map}}))
- \sum_{\alpha=1}^{n_{\text{con}}} \mu_\alpha \left[ \int_{\mathbb{R}^{m+1}} dX_{\text{map}} f_X(X_{\text{map}}) g_\alpha(X_{\text{map}}) - \overline{g}_\alpha \right] \tag{A2.32}
\]

This can be rearranged as:

\[
F[f_X] = - \int_{\mathbb{R}^{m+1}} dX_{\text{map}} f_X(X_{\text{map}}) \ln(f_X(X_{\text{map}}))
- \int_{\mathbb{R}^{m+1}} dX_{\text{map}} f_X(X_{\text{map}}) \sum_{\alpha=1}^{n_{\text{con}}} \mu_\alpha [g_\alpha(X_{\text{map}}) - \overline{g}_\alpha] \tag{A2.33}
\]

Introducing the term \( f_Y(X_{\text{map}}) \) results in the functional reducing to:

\[
F[f_X] = - \int_{\mathbb{R}^{m+1}} dX_{\text{map}} f_X(X_{\text{map}}) \ln(f_X(X_{\text{map}}))
- \int_{\mathbb{R}^{m+1}} dX_{\text{map}} f_X(X_{\text{map}}) \ln(f_Y(X_{\text{map}})) \tag{A2.34}
\]

where \( f_Y(X_{\text{map}}) \) is defined as follows and assumed to be valid:

\[
f_Y(X_{\text{map}}) = \frac{\exp \sum_{\alpha=1}^{n_{\text{con}}} \mu_\alpha g_\alpha(X_{\text{map}})}{\exp \sum_{\alpha=1}^{n_{\text{con}}} \mu_\alpha \overline{g}_\alpha} \tag{A2.35}
\]
Papoulis proves that the following is true (Papoulis 2002):

\[- \int_{R^{m+1}}^{R^{m+1}} dX_{\text{map}} f_X(X_{\text{map}}) \ln \left( f_X(X_{\text{map}}) \right) \leq \int_{R^{m+1}}^{R^{m+1}} dX_{\text{map}} f_X(X_{\text{map}}) \ln \left( f_Y(X_{\text{map}}) \right) \]

(A2.36)

Considering equations (A2.34) and (A2.36), we know that the functional is always less than or equal to zero.

\[F[f_X] \leq 0 \]

(A2.37)

The maximum occurs when \( f_X(X_{\text{map}}) = f_Y(X_{\text{map}}) \) hence:

\[f_X(X_{\text{map}}) = \frac{\exp \left( \sum_{\alpha=1}^{n_{\text{con}}} \mu_\alpha g_\alpha(X_{\text{map}}) \right)}{\exp \left( \sum_{\alpha=1}^{n_{\text{con}}} \mu_\alpha g_\alpha \right)} \]

(A2.38)

Hence it can be shown that:

\[f_G(X_{\text{map}}) = Z^{-1} \exp \left( Y_G(X_{\text{map}}) \right) \]

\[= Z^{-1} \exp \left( \sum_{\alpha=0}^{N_\alpha} \mu_\alpha (p_{\text{map}}) g_\alpha(X_{\text{map}}) \right) \]

(A2.39)

\[Z = \int dX_{\text{map}} \exp \left( \sum_{\alpha=0}^{N_\alpha} \mu_\alpha (p_{\text{map}}) g_\alpha(X_{\text{map}}) \right) \]

(A2.40)

Note above that the integral in the \( Z \) term in expression (A2.40) indicates the calculation of the expected value of the term to the right, hence tying in with the denominator in equation (A2.38). Equation (A2.39) yields the shape of the general solution to the prior PDF in BME. This PDF has maximum entropy a priori considering all available data in advance of estimation a posteriori.
Implementing the mean trend and covariance constraints yields the detailed solution to the PDF:

- Mean trend constraint (for $i = 1, 2, \ldots, m, k$)
  \[ g_i(\chi_{map}) = \chi_i; g_i(\chi_{map}) = m_i \]

- Centred covariance constraint (for $i = 1, 2, \ldots, m, k$)
  \[ g_{ij}(\chi_{map}) = (\chi_i - m_i)(\chi_j - m_j); g_i(\chi_{map}) = c_{ij} \]

As a result:

\[
\begin{align*}
  f_G(\chi_{map}) &= Z^{-1} \exp \left( Y_G(\chi_{map}) \right) \\
  &= Z^{-1} \exp \left( \sum_{i=1}^{m,k} \mu_i \chi_i \right)_{\text{Mean Trend}} \\
  &\quad + \sum_{i=1}^{m,k} \mu_{ij}(\chi_i - m_i)(\chi_j - m_j)_{\text{Covariance}} \tag{A2.41}
\end{align*}
\]

Solving for $\mu_i$ first by applying the mean trend constraint as $m_i$ is known. The known mean trend is equated to the expected value term on the right, which is the integral of the term $\chi_i$ multiplied by the general solution to the prior PDF.

\[
m_i = \int d\chi_{map}(\chi_i) \left[ Z^{-1} \exp \left( \sum_{i=1}^{m,k} \mu_i \chi_i + \sum_{i=1}^{m,k} \mu_{ij}(\chi_i - m_i)(\chi_j - m_j) \right) \right] \tag{A2.42}
\]

Obtaining the partial differential with respect to the constraints $\mu_i$ and equating to zero

\[
Z^{-1}(\chi_i) \left[ \frac{\partial}{\partial \mu_i} \int d\chi_{map} \left[ \exp \left( \sum_{i=1}^{m,k} \mu_i \chi_i + \sum_{i=1}^{m,k} \mu_{ij}(\chi_i - m_i)(\chi_j - m_j) \right) \right] \right] = 0 \tag{A2.43}
\]

The term $\chi_i$ can be removed, giving in shortened form:

\[
Z^{-1} \left[ \frac{\partial Z}{\partial \mu_i} \right] = \frac{\partial \ln Z}{\partial \mu_i} \tag{A2.44}
\]
A new coefficient $\lambda_{ij}$ is introduced as a function of $\mu_{ij}$.

$$\lambda_{ij} = -2\mu_{ij} \quad (A2.45)$$

Describing in matrix format, the prior PDF $f_\mathcal{G}(\chi_{\text{map}})$ is given by:

$$f_\mathcal{G}(\chi_{\text{map}}) = Z^{-1}\exp \left( \sum_{i=1}^{m,k} \mu_i \chi_i - \frac{1}{2} \sum_{i=1}^{m,k} \lambda_{ij} (\chi_i - m_i) (\chi_j - m_j) \right) \quad (A2.46)$$

$$= Z^{-1}\exp \left( \mu^T \chi_{\text{map}} - \frac{1}{2} (\chi_{\text{map}} - m_{\text{map}})^T \lambda (\chi_{\text{map}} - m_{\text{map}}) \right) \quad (A2.47)$$

Integrating on both sides yields, we know that the integral of any PDF over it’s domain yields 1:

$$1 = \int d\chi_{\text{map}} \left[ Z^{-1}\exp \left( \sum_{i=1}^{m,k} \mu_i \chi_i - \frac{1}{2} \sum_{i=1}^{m,k} \lambda_{ij} (\chi_i - m_i) (\chi_j - m_j) \right) \right] \quad (A2.48)$$

Multiplying across by $Z$, describing $\psi_{\text{map}} = \chi_{\text{map}} - m_{\text{map}}$, and $\chi_{\text{map}} = \psi_{\text{map}} + m_{\text{map}}$:

$$Z = \int d\chi_{\text{map}} \exp \left( \mu^T (\psi_{\text{map}} + m_{\text{map}}) - \frac{1}{2} (\psi_{\text{map}})^T \lambda (\psi_{\text{map}}) \right) \quad (A2.49)$$

$$= \exp (\mu^T m_{\text{map}}) \int d\chi_{\text{map}} \exp \left( \mu^T \psi_{\text{map}} - \frac{1}{2} (\psi_{\text{map}})^T \lambda (\psi_{\text{map}}) \right)$$

Expanding the latter term in the exponential:

$$-\frac{1}{2} (\psi_{\text{map}})^T \lambda (\psi_{\text{map}})$$

$$= -\frac{1}{2} (\psi_{\text{map}} - \lambda^{-1}\mu)^T \lambda (\psi_{\text{map}} - \lambda^{-1}\mu) + \frac{1}{2} \mu^T \lambda^{-1}\mu \quad (A2.50)$$

Hence:

$$Z = \exp (\mu^T m_{\text{map}}) \int d\chi_{\text{map}} \exp \left( \mu^T \psi_{\text{map}} - \frac{1}{2} (\psi_{\text{map}} - \lambda^{-1}\mu)^T \lambda (\psi_{\text{map}} - \lambda^{-1}\mu) + \frac{1}{2} \mu^T \lambda^{-1}\mu \right) \quad (A2.51)$$
Kriging and BME derivation

Multivariate normal distribution

One of the properties of a multivariate normal distribution is that the complete integral of a multivariate normal distribution is 1. This represents the fact that the sum of the probabilities of all of the likelihoods covered is 1.

From earlier, looking at the integral of \( Z \), it is known that the integral of a multivariate normal distribution such as \( Z \) is equal to 1 as shown above.

\[
\int d\chi_{\text{map}} \exp\left( \mu^T \psi_{\text{map}} - \frac{1}{2} (\psi_{\text{map}} - \lambda^{-1} \mu)^T \lambda (\psi_{\text{map}} - \lambda^{-1} \mu) \right) \\
= \exp\left( -\frac{1}{2} \mu^T (\lambda^{-1} \mu - 2m_{\text{map}}) \right)
\]

Applying the partial derivative of equation (A2.51) with respect to \( \mu_i \):

\[
\frac{\partial}{\partial \mu_i} \left[ 2\pi^{\frac{m+1}{2}} |\lambda|^{-\frac{1}{2}} \exp\left( -\frac{1}{2} \mu^T (\lambda^{-1} \mu - 2m_{\text{map}}) \right) \right] = 0
\]  

(A2.53)

\[
\frac{\partial}{\partial \mu_i} \left( -\frac{1}{2} \mu^T (\lambda^{-1} \mu - 2m_{\text{map}}) \right) = \frac{\partial}{\partial \mu_i} \left( -\frac{1}{2} \mu^2 \lambda^{-1} + \mu^T m_{\text{map}} \right) = 0
\]

Hence:

\[
\mu \lambda^{-1} + m_{\text{map}} = 0
\]  

(A2.55)

As \( m_{\text{map}} \) is assumed to be equal to zero because the mean trend has been removed, \( \mu = 0 \) or \( \lambda^{-1} = 0 \).

Until the covariance constraint is applied, no conclusions can be made regarding the correct solution to the mean trend constraint.

Covariance constraint:

\[
\lambda_{ij} = -2\mu_{ij} \rightarrow \mu_{ij} = \frac{1}{2} \lambda_{ij}
\]

(A2.56)
Solving for $\lambda_{ij}$ and hence $\mu_{ij}$, and assuming that $\mu_i = 0$ from application of the mean trend constraint:

$$c_{ij} = \int d\chi \{ (\chi_i - m_i)(\chi_j - m_j) f_0(\chi) \}$$

(A2.57)

$$c_{ij} = \frac{1}{2\pi^{m+1}} \int d\chi \{ (\chi_i - m_{map})^T(\chi_j - m_{map}) \}$$

(A2.58)

$$\times \exp \left( -\frac{1}{2} (\chi_i - m_{map})^T \lambda (\chi_j - m_{map}) \right)$$

As before:

$$\psi_{map} = \chi_{map} - m_{map}$$

(A2.59)

$$c_{ij} = \frac{1}{2\pi^{m+1}} \int d\chi \{ \psi_{map}^T \psi_{map} \} \exp \left( -\frac{1}{2} (\psi_{map})^T \lambda (\psi_{map}) \right)$$

(A2.60)

The following transformation is used:

$$A = B\psi_{map} \rightarrow \psi_{map} = B^{-1}A$$

(A2.61)

$$d\psi_{map} = \frac{dA_{map}}{|B|} = \frac{dA_{map}}{|\lambda|^2}$$

(A2.62)

The covariance constraint is now described with new transformation:

$$c_{ij} = \frac{1}{2\pi^{m+1}} \int dA_{map} (B^{-1}A)^T(B^{-1}A) \exp \left( -\frac{1}{2} (\psi_{map})^T B^T B \psi_{map} \right)$$

(A2.63)

Given that $A = B\psi_{map}, A^T = (\psi_{map})^T (B)^T$, the constraint reduces to:

$$c_{ij} = \frac{1}{2\pi^{m+1}} \int dA_{map} (B^{-1}A)^T(B^{-1}A) \exp \left( -\frac{1}{2} A^T A \right)$$
Kriging and BME derivation

\[
\begin{align*}
B^{-1}B^{-1\top} = & \frac{B^{-1}B^{-1\top}}{2\pi\frac{m+1}{2}} \int dA_{\text{map}}A^{\top}A \exp \left( -\frac{1}{2}A^{\top}A \right) \\
= & \frac{B^{-1}B^{-1\top}}{2\pi\frac{m+1}{2}} \left[ \delta \left( 2\pi\frac{m+1}{2} \right) \right] = B^{-1}B^{-1\top} = (BB^{\top})^{-1} = \lambda^{-1}
\end{align*}
\]  

(A2.64)

As shown above, applying the covariance constraint to determine the weights \( \lambda_{ij} / \mu_{ij} \), the constraint is outlined and the integral of the covariance is completed without differentiating w.r.t. \( \mu_{ij} \). The integral collapses, resulting in the definite solution:

\[ C = \lambda^{-1} ; \lambda = C^{-1} \]  

(Hence confirming the earlier solution for the mean trend constraint that \( \mu_i = 0 \), and rejecting the initial potential solution \( \lambda^{-1} = 0 \).

As a result, the prior multivariate normal distribution, generally referred to as \( f_G(\chi_{\text{map}}) \), is described by the following parameterised equation which incorporates collective hard and soft data values in the mapping area, the mean trend and covariance:

\[
\phi(\chi_{\text{map}}; m_{\text{map}}, C_{\text{map}}) = \frac{1}{2\pi\frac{m+1}{2}|C|^2} \exp \left( -\frac{1}{2}(\chi_{\text{map}} - m_{\text{map}})^{\top}C^{-1}(\chi_{\text{map}} - m_{\text{map}}) \right)
\]  

(A2.66)

Obtaining the posterior PDF requires two considerations:

- Firstly, the adaptation of the prior gaussian multivariate PDF by multiplying the prior PDF by the multivariate PDF describing all the soft data which is being included. The resulting multivariate PDF describes the entire mapped area \( \chi_{\text{map}} = [\chi_{\text{hard}} \chi_{\text{soft}}] \) including the soft data, which is not explicitly included at the prior stage. The prior PDF treats all data as certain, hard data with probability 1.

\[
f_K(\chi_{\text{map}}) = f_S(\chi_{\text{soft}}) \times f_G(\chi_{\text{map}})
\]  

(A2.67)
where the subscript K indicates full knowledge, the subscript S indicates soft uncertain knowledge and subscript G represents general knowledge of mean and covariance.

- Secondly, to obtain the univariate posterior PDF $f_K(\chi_k)$ describing the prediction point $p_k$, the multivariate PDF describing the mapped area $\chi_{map} = [\chi_k \chi_{hard} \chi_{soft}]$ is divided by the marginal multivariate PDF describing only the data locations $\chi_{data} = [\chi_{hard} \chi_{soft}]$. The posterior PDF $f_K(\chi_k)$ can thus be considered as:

$$f_K(\chi_k) = \frac{f_K(\chi_{map})}{f_K(\chi_{data})} = f_K([\chi_k \chi_{hard} \chi_{soft}] | [\chi_{hard} \chi_{soft}])$$  \hspace{1cm} (A2.68)

**Detailed application of BME to probabilistic soft data $\chi_{soft}$ and hard data $\chi_{hard}$**

Following on from the above remarks, the posterior PDF can be written generally as

$$f_K(\chi_k) = A^{-1} \int d\chi_{soft} f_s(\chi_{soft}) f_G(\chi_{map})$$  \hspace{1cm} (A2.69)

$A^{-1}$ refers to the Bayesian aspect of the procedure, where the posterior PDF is conditioned on the knowledge of the PDF describing the data points alone. Note in the following expression that the general knowledge prior PDF $f_G$ describes the data points alone:

$$A = \int d\chi_{soft} f_s(\chi_{soft}) f_G(\chi_{data})$$  \hspace{1cm} (A2.70)

Using the previous notation for the parameterised prior PDF yields the following expression:

$$f_K(\chi_k) = A^{-1} \int d\chi_{soft} f_s(\chi_{soft}) [\phi(\chi_{map}; m_{map}, c_{map})]$$  \hspace{1cm} (A2.71)
And $A$ can be considered as

$$A = \int d\chi_{soft} f_s(\chi_{soft}) \left[ \phi(\chi_{data}; m_{data}, C_{data}) \right] \quad (A2.72)$$

The multivariate gaussian prior PDF $\phi(\chi_{map}; m_{map}, C_{map})$ is disaggregated into three normal prior PDFs through the conditional probabilities rule:

$$f_k(\chi_k, \chi_{hard}, \chi_{soft}) = f(\chi_{hard}) \times f(\chi_k | \chi_{hard}) \times f(\chi_{soft} | \chi_k, \chi_{hard}) \quad (A2.73)$$

As the PDF of the hard data is a constant value, the chain of prior PDFs is represented by:

$$\phi(\chi_{[h,s]}, m_{[h,s]}, C_{[h,s]}) = \phi(\chi_k; m_{[k|h]}, C_{[k|h]}) \times \phi(\chi_{soft}; m_{[s|kh]}, C_{[s|kh]}) \quad (A2.74)$$

The PDF $f_k(\chi_{data})$ in the expression for $f_k(\chi_k)$ is required to implement the bayesian conditional rule, whereby the knowledge of the estimation point is described by the PDF of the mapped area, conditional on the knowledge of the hard and soft data. The expression for $f_k(\chi_{data})$ is the marginal PDF of $\chi_{data}$ derived from the integration of $f_K(\chi_{map})$ with respect to $\chi_k$, to remove the description of the characteristics of $\chi_k$:

$$f_k(\chi_{hard}, \chi_{soft}) = \int d\chi_k f_k(\chi_k, \chi_{hard}, \chi_{soft})$$

$$= \int d\chi_k \left[ f(\chi_{hard}) \times f(\chi_k | \chi_{hard}) \times f(\chi_{soft} | \chi_k, \chi_{hard}) \right]$$

$$= f(\chi_{soft} | \chi_{hard}) \quad (A2.75)$$

$$f(\chi_{soft} | \chi_{hard}) = \phi(\chi_{[h,s]}, m_{[h,s]}, C_{[h,s]}) = \phi(\chi_{soft}; m_{[s|kh]}, C_{[s|kh]}) \quad (A2.76)$$
Hence the posterior PDF describing knowledge of the estimation point \( p_k \) is described by:

\[
f_K(\chi_k) = A^{-1} \int d\chi_{soft} f_s(\chi_{soft}) \left[ \phi(\chi_k; m_{(k|h)}, C_{(k|h)}) \times \phi(\chi_{soft}; m_{(s|kh)}, C_{(s|kh)}) \right]
\]  

(A2.77)

Rearranging for integration:

\[
f_K(\chi_k) = A^{-1} \phi(\chi_k; m_{(k|h)}, C_{(k|h)}) \int d\chi_{soft} f_s(\chi_{soft}) \phi(\chi_{soft}; m_{(s|kh)}, C_{(s|kh)})
\]  

(A2.78)

And

\[
A = \int d\chi_{soft} f_s(\chi_{soft}) \left[ \phi(\chi_{soft}; m_{(s|h)}, C_{(s|h)}) \right]
\]  

(A2.79)

The conditional mean and covariance data is obtained as follows:

\[
m_{(k|h)} = B_{(k|h)}\chi_{hard}
\]  

(A2.80)

\[
B_{(k|h)} = C_{k,h}C_{h,h}^{-1}
\]  

(A2.81)

\[
C_{(k|h)} = C_{k,k} - B_{(k|h)}C_{h,k}
\]  

(A2.82)

\[
m_{(s|kh)} = B_{(s|kh)}\chi_{hard}
\]  

(A2.83)

\[
B_{(s|kh)} = C_{s,kh}C_{kh,kh}^{-1}
\]  

(A2.84)

\[
C_{(s|kh)} = C_{s,s} - B_{(s|kh)}C_{kh,s}
\]  

(A2.85)

\[
m_{(s|h)} = B_{(s|h)}\chi_{hard}
\]  

(A2.86)

\[
B_{(s|h)} = C_{s,h}C_{h,h}^{-1}
\]  

(A2.87)

\[
C_{(s|h)} = C_{s,s} - B_{(s|h)}C_{h,s}
\]  

(A2.88)

The resultant equation forms the basis for the calculation of different measures on the posterior probability distribution for point \( p_k \), such as mean, median and mode estimates, standard deviation, percentiles and quartiles.
The BME mode is identified as the peak of the posterior PDF describing the estimation point, and is obtained via differentiation of the posterior PDF with respect to \( \chi_k \) and equating to zero. Hence, the most probable value is identified when the differential is solved for \( (\hat{\chi}_k)_{\text{mode}} \). The derivation of the BME mode estimate is outlined in the following section, firstly in general terms and subsequently in terms of the conditional means and covariances as implemented in Matlab.

\[
f_K(\chi_k) = A^{-1} \int d\chi_{\text{soft}} f_{\text{soft}}(\chi_{\text{soft}}) f_G(\chi_{\text{map}})
\]

\[
\frac{\partial}{\partial \chi_k} f_K(\chi_k) = A^{-1} \int d\chi_{\text{soft}} f_{\text{soft}}(\chi_{\text{soft}}) \frac{\partial f_G(\chi_{\text{map}})}{\partial \chi_k} = 0
\]

\[
\frac{\partial f_G(\chi_{\text{map}})}{\partial \chi_k} = \frac{\partial}{\partial \chi_k} \left[ Z^{-1} \exp \left( -\frac{1}{2} (\chi_{\text{map}} - m_{\text{map}})^T C^{-1} (\chi_{\text{map}} - m_{\text{map}}) \right) \right]
\]

\[
= f_G(\chi_{\text{map}}) \frac{\partial}{\partial \chi_k} \left[ -\frac{1}{2} (\chi_{\text{map}} - m_{\text{map}})^T C_{\text{map,map}}^{-1} (\chi_{\text{map}} - m_{\text{map}}) \right]
\]

\[
= f_G(\chi_{\text{map}}) \left[ -(\chi_{\text{map}} - m_{\text{map}})^T C_{\text{map,k}}^{-1} \right]
\]

Reinserting into the original equation:

\[
\frac{\partial}{\partial \chi_k} f_K(\chi_k) = A^{-1} \int d\chi_{\text{soft}} f_{\text{soft}}(\chi_{\text{soft}}) f_G(\chi_{\text{map}}) \left[ -(\chi_{\text{map}} - m_{\text{map}})^T C_{\text{map,k}}^{-1} \right] = 0
\]

\[
\left[ -(\chi_{\text{map}} - m_{\text{map}})^T C_{\text{map,k}}^{-1} \right] = 0
\]
This can be re-expressed in terms of the \((m_h)\) hard data points, \((m_h + 1: m)\) soft data points and the single estimate data at \(k\):

\[
- \sum_{i=1}^{m,k} (\chi_i - m_i)C_{i,k}^{-1} = 0
\]  
(A2.94)

The matrices in question are divided up into their hard, soft and \(k\) components:

\[
- \sum_{i=1}^{m_h} (\chi_i - m_i)C_{i,k}^{-1} - \sum_{i=m_h+1}^{m} (\chi_i - m_i)C_{i,k}^{-1} - (\chi_k - m_k)C_{k,k}^{-1} = 0
\]  
(A2.95)

Hence:

\[
\hat{\chi}_k - m_k = - \frac{1}{C_{k,k}^{-1}} \left[ \sum_{i=1}^{m_h} (\chi_i - m_i)C_{i,k}^{-1} + \sum_{i=m_h+1}^{m} (\chi_i - m_i)C_{i,k}^{-1} \right]
\]  
(A2.96)

Within the expression above, the expected values \(\chi_i\) of the soft data locations are required for the solution of \(\hat{\chi}_k\). The expected value is determined as follows:

\[
\chi_i - m_i = \left( \int d\chi_{soft} f_s(\chi_{soft}) f_G(\chi_{map}) \right)^{-1} \int d\chi_{soft} f_s(\chi_{soft}) (\chi_i - m_i) f_G(\chi_{map})
\]  
(A2.97)

As the entire solution procedure assumes zero mean trend, \(m_i\) is excluded.

\[
\chi_i = \left( \int d\chi_{soft} f_s(\chi_{soft}) f_G(\chi_{map}) \right)^{-1} \int d\chi_{soft} f_s(\chi_{soft}) (\chi_i) f_G(\chi_{map})
\]  
(A2.98)

As before:

\[
f_G(\chi_{map}) = f_G(\chi_{hard}) \times f_G(\chi_k | \chi_{hard}) \times f_G(\chi_{soft} | \chi_{hard})
\]
\[
= \phi(\chi_k; m_{(k|h)}, C_{(k|h)}) \times \phi(\chi_{soft}; m_{(s|kh)}, C_{(s|kh)})
\]  
(A2.99)
Inserting into the solution for $\overline{\chi}_i$:

$$\overline{\chi}_i = \left( \int d\chi \phi(\chi; m(\chi|\ell), C(\chi|\ell)) \right) \times \left( \int d\chi \phi(\chi; m(\ell|\chi), C(\chi|\ell)) \right)^{-1} \int d\chi \phi(\chi; m(\chi|\ell), C(\chi|\ell))$$

(A2.100)

The BME mean estimate

The BME mean is observed as the value which minimises the mean square error.

The solution is derived as follows:

$$\overline{\chi}_{k|K} = \int d\chi_k \chi_k A^{-1} \int d\chi \phi(\chi; m(\chi|\ell), C(\chi|\ell))$$

(A2.102)

where, as per usual:

$$A = \int d\chi \phi(\chi; m(\chi|\ell), C(\chi|\ell))$$

(A2.103)

Expanding with the conditional probabilities rule:

$$\overline{\chi}_{k|K} = \int d\chi_k \chi_k A^{-1} \int d\chi \phi(\chi; m(\chi|\ell), C(\chi|\ell))$$

(A2.104)
\[ \chi_k = B_{k|hs} \chi_{hs} \]  

(A2.105)

where \( \chi_{hs} \) is the concatenated vector of \( \chi_h \) and \( \chi_s \).

\[
\bar{x}_{k|K} = \int d\chi_k A^{-1} B_{k|hs} \phi(\chi_k; m_{k|h}, C_{k|h}) \int d\chi_{soft} f_s(\chi_{soft}) \phi(\chi_{soft}; m_{s|kh}, C_{s|kh})
\]  

(A2.106)

Note the integral:

\[
\int d\chi_k \phi(\chi_k; m_{k|h}, C_{k|h}) \phi(\chi_{soft}; m_{s|kh}, C_{s|kh}) = \phi(\chi_{soft}; m_{s|h}, C_{s|h})
\]  

(A2.107)

Hence:

\[
\bar{x}_{k|K} = A^{-1} B_{k|hs} \chi_{hs} \int d\chi_{soft} f_s(\chi_{soft}) \phi(\chi_{soft}; m_{s|h}, C_{s|h})
\]  

(A2.108)

Also:

\[
A = \int d\chi_{soft} f_s(\chi_{soft}) \phi(\chi_{soft}; m_{s|h}, C_{s|h})
\]  

(A2.109)

Cancelling A above and below yields:

\[
\bar{x}_{k|k} = B_{k|hs} \chi_{hs} = B_{k|hs(h)} \chi_h + B_{k|hs(s)} \chi_s
\]  

(A2.110)

The expected value \( \bar{\chi}_s \) of the soft data \( \chi_s \) is used in the above equation. The expected value is given by:

\[
\bar{\chi}_s = A^{-1} \int d\chi_{soft} f_s(\chi_{soft}) \chi_s f_G(\chi_{map})
\]  

(A2.111)

Keeping in mind the integration of \( \int d\chi_k f_G(\chi_{map}) \) earlier on still applies, \( f_G(\chi_{map}) \) can be considered as \( \phi(\chi_{soft}; m_{s|h}, C_{s|h}) \).

\[
\bar{\chi}_s = A^{-1} \int d\chi_{soft} f_s(\chi_{soft}) \chi_s \phi(\chi_{soft}; m_{s|h}, C_{s|h})
\]  

(A2.112)

287
And:

\[ A = \int d\chi_{\text{soft}} f_s(\chi_{\text{soft}}) \phi(\chi_{\text{soft}}; m(s|h), C(s|h)) \]  (A2.113)

Conditional mean and covariance values remain as identified before.

Hence:

\[ \overline{x}_k = B_{k|hs} x_{hs} \]
\[ = B_{k|hs(h)} x_h \]
\[ + B_{k|hs(s)} \left( A^{-1} \int d\chi_{\text{soft}} f_s(\chi_{\text{soft}}) \chi_s \phi(\chi_{\text{soft}}; m(s|h), C(s|h)) \right) \]  (A2.114)

**BME posterior PDF - variance and standard deviation**

\[ \sigma_k^2 = \int d\chi_k (\chi_k - \overline{x}_k)^2 \left[ A^{-1} \int d\chi_{\text{soft}} f_s(\chi_{\text{soft}}) \phi(\chi_{\text{soft}}; m(k|h), C(k|h)) \right. \]
\[ \left. \times \phi(\chi_{\text{soft}}; m(s|kh), C(s|kh)) \right] \]  (A2.115)

Applying the integral:

\[ \int d\chi_k \phi(\chi_k; m(k|h), C(k|h)) \times \phi(\chi_{\text{soft}}; m(s|kh), C(s|kh)) = \phi(\chi_{\text{soft}}; m(s|h), C(s|h)) \]  (A2.116)

\[ \chi_k = B_{k|hs} x_{hs} \]  (A2.117)
\[ \sigma_k^2 = A^{-1} \int d\chi_{\text{soft}} \left( B_{k|hs} x_{hs} - \overline{x}_k \right)^2 f_s(\chi_{\text{soft}}) \phi(\chi_{\text{soft}}; m(s|h), C(s|h)) \]  (A2.118)

\( C_{k|hs} \) must be also included to describe the variation in k already described by hard and soft data.

Hence:

\[ \sigma_k^2 = C_{k|hs} + A^{-1} \int d\chi_{\text{soft}} \left( B_{k|hs} x_{hs} - \overline{x}_k \right)^2 f_s(\chi_{\text{soft}}) \phi(\chi_{\text{soft}}; m(s|h), C(s|h)) \]  (A2.119)


Geographical Information System International Group and International Center of Coastal and Ocean Policy Studies. Genova, Italy.


EPA (2013) EPA report under Article 29(1)(b) of the European Communities (Good Agricultural Practice for Protection of Waters) Regulations 2010.
European Commission (2003a) Guidance document No. 6 Towards a guidance on establishment of the intercalibration network and the process on the intercalibration exercise.


Nash, S. (2010) DEVELOPMENT OF AN EFFICIENT DYNAMICALLY-NESTED MODEL FOR TIDAL HYDRAULICS AND SOLUTE TRANSPORT, NUI GALWAY.
O'Boyle, S., Wilkes, R. and Tomasz, D. (2011) Ireland's report on 'distance to target'; a quantitative estimate of the amount of nutrient reduction required to reach non-problem area status with regard to eutrophication, OSPAR Commission.