

# An Improved Water Quality Model for the Onkaparinga Catchment

Petra Kuhnert, Dan Pagendam Jim Cox, Nigel Fleming Ying He,  
Clive Jenkins Leon van der Linden



Goyder Institute for Water Research  
Technical Report Series No. 15/6



[www.goyderinstitute.org](http://www.goyderinstitute.org)

**Goyder Institute for Water Research Technical Report Series ISSN: 1839-2725**

The Goyder Institute for Water Research is a partnership between the South Australian Government through the Department of Environment, Water and Natural Resources, CSIRO, Flinders University, the University of Adelaide and the University of South Australia. The Institute will enhance the South Australian Government's capacity to develop and deliver science-based policy solutions in water management. It brings together the best scientists and researchers across Australia to provide expert and independent scientific advice to inform good government water policy and identify future threats and opportunities to water security.



The following Associate organisations contributed to this report:



Enquires should be addressed to: Goyder Institute for Water Research  
Level 1, Torrens Building  
220 Victoria Square, Adelaide, SA, 5000  
tel: 08-8303 8952  
e-mail: [enquiries@goyderinstitute.org](mailto:enquiries@goyderinstitute.org)

**Citation**

Kuhnert, P.M., Pagendam, D., Cox, J., Fleming, N., He, Y., Thomas, S., Jenkins, C.I., van der Linden, L., 2015, *An improved water quality model for the Onkaparinga Catchment*, Goyder Institute for Water Research Technical Report Series No. 15/6, Adelaide, South Australia

**Copyright**

© 2015 CSIRO. To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of CSIRO.

**Disclaimer**

The Participants advise that the information contained in this publication comprises general statements based on scientific research and does not warrant or represent the completeness of any information or material in this publication.

## Table of Contents

<b>List of Figures .....</b>	<b>5</b>
<b>List of Tables .....</b>	<b>11</b>
<b>Executive Summary .....</b>	<b>13</b>
<b>Acknowledgements .....</b>	<b>16</b>
<b>Introduction .....</b>	<b>17</b>
<b>Study Region .....</b>	<b>18</b>
<b>Activity 1: Rainfall-Runoff Model Calibration &amp; Error Quantification .....</b>	<b>25</b>
<b>Motivation.....</b>	<b>25</b>
<b>Overview of Statistical Modelling of Hydrographic Data .....</b>	<b>25</b>
<b>Scott Creek (A5030502) .....</b>	<b>28</b>
<b>Onkaparinga River at Houlgraves (A5030504).....</b>	<b>31</b>
<b>Echunga Creek (A5030506) .....</b>	<b>34</b>
<b>Lenswood Creek (A5030507) .....</b>	<b>37</b>
<b>Aldgate Creek (A5030509) .....</b>	<b>40</b>
<b>Cox Creek at Uraidla (A5030526).....</b>	<b>43</b>
<b>Activity 2: Development of Site-Based Statistical Models .....</b>	<b>47</b>
<b>Motivation.....</b>	<b>47</b>
<b>Overview of Statistical Modelling Framework .....</b>	<b>47</b>
<b>Scott Creek (A5030502) .....</b>	<b>48</b>
<b>Onkaparinga River at Houlgraves (A5030504).....</b>	<b>57</b>
<b>Echunga Creek (A5030506) .....</b>	<b>66</b>
<b>Lenswood Creek (A5030507) .....</b>	<b>74</b>
<b>Aldgate Creek (A5030509) .....</b>	<b>83</b>
<b>Cox Creek at Uraidla (A5030526).....</b>	<b>92</b>
<b>Activity 3: Spatio-Temporal Models to Investigate Scenarios.....</b>	<b>101</b>
<b>Motivation.....</b>	<b>101</b>
<b>Scenario 1: Sale of SA Water Land Holdings (Scott Creek sub-catchment).....</b>	<b>101</b>
<b>Scenario 2: Impact of Continued expansion of Perennial Horticulture (Cox Creek sub-catchment)</b> <b>.....</b>	<b>102</b>
<b>Overview of Statistical Modelling Framework .....</b>	<b>103</b>
<b>Scenario 1 Results .....</b>	<b>104</b>
TSS.....	104
TN.....	106
TP .....	109
<b>Scenario 2 Results .....</b>	<b>111</b>
TSS.....	111
TN.....	113
TP .....	114
<b>Scenario 3 Results .....</b>	<b>116</b>

TSS.....	116
TN.....	119
TP .....	121
<b>Discussion and Conclusions .....</b>	<b>123</b>
<b>References .....</b>	<b>125</b>
<b>Appendix A – Summary of Covariates used in Modelling .....</b>	<b>128</b>
<b>Appendix B - Diagnostic Plots for Site Based Models .....</b>	<b>129</b>
Scott Creek.....	129
Onkaparinga River, Houlgraves.....	130
Echunga Creek.....	131
Lenswood Creek.....	132
Aldgate Creek.....	133
Cox Creek.....	134
<b>Appendix C - Land-use Compositions Upstream from Onkaparinga Gauges .....</b>	<b>135</b>
Scott Creek (A5030502) .....	135
Onkaparinga River at Houlgraves (A5030504).....	135
Echunga Creek (A5030506) .....	136
Lenswood Creek (A5030507) .....	136
Aldgate Creek (A5030509) .....	137
Cox Creek at Uraidla (A5030526).....	137

## List of Figures

Figure 1: Maps showing (a) the Mt Lofty catchment and contributing areas and (b) the Onkaparinga catchment with sites that were investigated as part of this project overlaid. ....	19
Figure 2: Raw flow data for the six sites in the Onkaparinga catchment. ....	20
Figure 3: Plots of the raw data showing samples of TSS for the six sites in the Onkaparinga catchment. ....	21
Figure 4: Plots of the raw data showing samples of TN for the six sites in the Onkaparinga catchment. ....	22
Figure 5: Plots of the raw data showing samples of TP for the six sites in the Onkaparinga catchment. ....	23
Figure 6: Summary of TSS samples stratified by flow and represented through a histogram for (a) Cox Creek (A5030526), (b) Scott Creek (A5030502), (c) Lenswood Creek (A5030507), (d) Echunga Creek (A5030506), (e) Aldgate Creek (A5030509); and (f) Onkaparinga River at Houlgraves (A5030504). ....	24
Figure 7: Rainfall-runoff model calibration results for Scott Creek (A5030502) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	30
Figure 8: Rainfall-runoff model calibration results for Scott Creek (A5030502) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	31
Figure 9: Rainfall-runoff model calibration results for the Onkaparinga River at Houlgraves (A5030504) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	33
Figure 10: Rainfall-runoff model calibration results for the Onkaparinga River at Houlgraves (A5030504) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	34
Figure 11: Rainfall-runoff model calibration results for Echunga Creek (A5030506) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	36
Figure 12: Rainfall-runoff model calibration results for Echunga Creek (A5030506) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	37
Figure 13: Rainfall-runoff model calibration results for Lenswood Creek (A5030507) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	39
Figure 14: Rainfall-runoff model calibration results for Lenswood Creek (A5030507) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	40
Figure 15: Rainfall-runoff model calibration results for Aldgate Creek (A5030509) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	42
Figure 16: Rainfall-runoff model calibration results for Aldgate Creek (A5030509) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	43
Figure 17: Rainfall-runoff model calibration results for Cox Creek at Uraidla (A5030526) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. ....	45

Figure 18: Rainfall-runoff model calibration results for Cox Creek at Uraidla (A5030526) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM. .... 46

Figure 19: Raw flow and concentration data presented on the log-scale for TSS at the Scott Creek site (A5030526)..... 49

Figure 20: Raw flow and concentration data presented on the log-scale for TN at the Scott Creek site (A5030526)..... 49

Figure 21: Raw flow and concentration data presented on the log-scale for TP at the Scott Creek site (A5030526)..... 50

Figure 22: Smooth terms from the generalised additive models fit for Scott Creek (A5030502) for each of the three constituents. TSS: (a) past sum of flow; TN: (b) seasonal term, (c) past sum of flow (d) discounted flow ( $d=0.75$ ); TP: (e) seasonal term..... 52

Figure 23: Predictions from the generalised additive model fit to constituent data at Scott Creek (A5030502) for (a) TSS, (b) TN and (c) TP. .... 54

Figure 28: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for Scott Creek (A5030502) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP. .... 55

Figure 29: Raw flow and concentration data presented on the log-scale for TSS at the Onkaparinga River, Houlgraves site (A5030504)..... 58

Figure 29: Raw flow and concentration data presented on the log-scale for TN at the Onkaparinga River, Houlgraves site (A5030504)..... 58

Figure 29: Raw flow and concentration data presented on the log-scale for TP at the Onkaparinga River, Houlgraves site (A5030504)..... 59

Figure 30: Smooth terms from the generalised additive model for Onkaparinga River at Houlgraves (A5030504) showing (a) the characteristics of season, (b) historical flow, (c) discounting term ( $d0.25$ ), (d) discounting term ( $d0.5$ ) and discounting term ( $d0.95$ ) in relation to TSS..... 61

Figure 31: Smooth terms from the generalised additive model for Onkaparinga River at Houlgraves (A5030504) showing the characteristics of (a) season, and (b) the discounted flow ( $d0.75$ ) in relation to TN..... 62

Figure 32: Smooth terms from the generalised additive model for Onkaparinga River at Houlgraves (A5030504) showing the characteristics of (a) season, (b) the past sum of flow, (c) discounted flow ( $d0.5$ ) and (d) discounted flow ( $d0.1$ ) in relation to TP. .... 63

Figure 33: Predictions from the generalised additive model fit to constituent data at Onkaparinga River, Houlgraves Site (A5030504) for (a) TSS, (b) TN and (c) TP. .... 64

Figure 34: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for Onkaparinga River, Houlgraves Site (A5030504) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP..... 65

Figure 35: Raw flow and concentration data presented on the log-scale for TSS at the Echunga Creek site (A5030506)..... 68

Figure 35: Raw flow and concentration data presented on the log-scale for TN at the Echunga Creek site (A5030506)..... 68

Figure 35: Raw flow and concentration data presented on the log-scale for TP at the Echunga Creek site (A5030506)..... 69

Figure 36: Smooth terms from the generalised additive models fit for Echunga Creek (A5030506) for TSS and TN. TSS: (a) seasonal term, (b) past sum of flow (c) discounted flow ( $d=0.5$ ); TN: (d) seasonal term, (e) past sum of flow (f) discounted flow ( $d=0.5$ )..... 70

Figure 37: Smooth terms from the generalised additive model for Echunga Creek (A5030506) showing the (a) seasonal term, (b) past sum of flow (c) discounting term ( $d0.95$ ) and (d) discounting term ( $d0.75$ ) in relation to TP. .... 71

Figure 38: Predictions from the generalised additive model fit to constituent data at Echunga Creek Site (A5030506) for (a) TSS, (b) TN and (c) TP.....	72
Figure 39: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for the Echunga Creek Site (A5030506) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP. ....	73
Figure 40: Raw flow and concentration data presented on the log-scale for TSS at the Lenswood Creek site (A5030507).....	75
Figure 40: Raw flow and concentration data presented on the log-scale for TN at the Lenswood Creek site (A5030507).....	76
Figure 40: Raw flow and concentration data presented on the log-scale for TP at the Lenswood Creek site (A5030507).....	76
Figure 41: Smooth terms from the generalised additive model for Lenswood Creek (A5030507) showing the characteristics of (a) the seasonal term in the model, (b) the past sum of flow, (c) discounting term d0.1, (d) discounting term d0.25, (e) discounting term d0.5 and (f) discounting term d0.75 in relation to TSS.....	78
Figure 42: Smooth terms from the generalised additive model for Lenswood Creek (A5030507) showing the characteristics of (a) the seasonal term in the model, (b) the past sum of flow, (c) discounting term d0.1, (d) discounting term d0.25 and (e) discounting term d0.95 in relation to TN. ....	79
Figure 43: Smooth terms from the generalised additive model for Lenswood Creek (A5030507) showing the characteristics of (a) the seasonal term in the model, (b) the past sum of flow, (c) discounting term d0.1, (d) discounting term d0.25 and (e) discounting term d0.95 in relation to TP. 80	
Figure 44: Predictions from the generalised additive model fit to constituent data at Lenswood Creek Site (A5030507) for (a) TSS, (b) TN and (c) TP.....	81
Figure 45: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for the Lenswood Creek Site (A5030507) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP.....	82
Figure 46: Raw flow and concentration data presented on the log-scale for TSS at the Aldgate Creek site (A5030509).....	86
Figure 46: Raw flow and concentration data presented on the log-scale for TN at the Aldgate Creek site (A5030509).....	86
Figure 46: Raw flow and concentration data presented on the log-scale for TP at the Aldgate Creek site (A5030509).....	87
Figure 47: Smooth terms from the generalised additive models fit for Aldgate Creek (A5030509) for TSS and TN. TSS: (a) seasonal term, (b) past sum of flow, (c) discounted flow (d0.1), (d) discounted flow (d0.25), (e) discounted flow (d0.75); TN: (f) past sum of flow.....	88
Figure 48: Smooth terms from the generalised additive model for Aldgate Creek (A5030509) showing the characteristics of (a) the seasonal term in the model, (b) the past sum of flow, (c) discounting term d0.95 and (d) discounting term d0.75 in relation to TP.....	89
Figure 49: Predictions from the generalised additive model fit to constituent data at Aldgate Creek Site (A5030509) for (a) TSS, (b) TN and (c) TP.....	90
Figure 50: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for the Aldgate Creek Site (A5030509) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP. ....	91
Figure 51: Marginal effect of Brooks Bridge sedimentation basin on TSS (a) prior to being operational and operational post 30 November 2006 and TP (a) prior to being operations and operational past 30 November 2007. Dotted lines represent 95% confidence intervals around the marginal effect. ..	92
Figure 52: Raw flow and concentration data presented on the log-scale for TSS at the Cox Creek site (A5030526).....	94

Figure 52: Raw flow and concentration data presented on the log-scale for TN at the Cox Creek site (A5030526)..... 94

Figure 52: Raw flow and concentration data presented on the log-scale for TP at the Cox Creek site (A5030526)..... 95

Figure 53: Smooth terms from the generalised additive models fit for Cox Creek (A5030526) for TSS and TP. TSS: (a) seasonal term; TP: (f) seasonal term..... 97

Figure 54: Smooth terms from the generalised additive model for Cox Creek (A5030526) showing the characteristics of (a) season, (b) the accumulation of flow, (c) discounted flow (d0.25) and (d) discounted flow (d0.75) in relation to TN..... 98

Figure 55: Predictions from the generalised additive model fit to constituent data at Cox Creek Site (A5030526) for (a) TSS, (b) TN and (c) TP. .... 99

Figure 56: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for the Cox Creek Site (A5030526) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP. .... 100

Figure 57: Variable importance ranking for TSS based on two criteria (left) Percent MSE and (right) node purity..... 104

Figure 58: Estimates of loads and total flow for each financial year for Scott Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue. .... 105

Figure 59: Estimates of the flow weighted concentrations for Scott Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue. .... 106

Figure 60: Variable importance ranking for TN based on two criteria (left) Percent MSE and (right) node purity..... 107

Figure 61: Estimates of TN loads and total flow for each financial year for Scott Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue. .... 108

Figure 62: Estimates of the flow weighted TN concentrations for Scott Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue. .... 108

Figure 63: Variable importance ranking for TP based on two criteria (left) Percent MSE and (right) node purity..... 109

Figure 64: Estimates of TP loads and total flow for each financial year for Scott Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue. .... 110

Figure 65: Estimates of the flow weighted TP concentrations for Scott Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue. .... 110

Figure 66: Variable importance ranking for TSS based on two criteria (left) Percent MSE and (right) node purity..... 111

Figure 67: Estimates of loads and total flow for each financial year for Cox Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue. .... 112

Figure 68: Estimates of the flow weighted concentrations for Cox Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue. .... 112

Figure 69: Variable importance ranking for TN based on two criteria (left) Percent MSE and (right) node purity..... 113

Figure 70: Estimates of TN loads and total flow for each financial year for Cox Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue. .... 114

Figure 71: Estimates of the flow weighted TN concentrations for Cox Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue. .... 114

Figure 72: Variable importance ranking for TP based on two criteria (left) Percent MSE and (right) node purity..... 115

Figure 73: Estimates of TP loads and total flow for each financial year for Cox Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue. .... 115

Figure 74: Estimates of the flow weighted TP concentrations for Cox Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue. .... 116

Figure 75: Variable importance ranking for TSS based on two criteria (left) Percent MSE and (right) node purity..... 117

Figure 76: Estimates of loads and total flow for each financial year for Aldgate Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue. .... 118

Figure 77: Estimates of the flow weighted concentrations for Aldgate Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue. .... 118

Figure 78: Variable importance ranking for TN based on two criteria (left) Percent MSE and (right) node purity..... 119

Figure 79: Estimates of TN loads and total flow for each financial year for Cox Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue. .... 120

Figure 80: Estimates of the flow weighted TN concentrations for Cox Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue. .... 120

Figure 81: Variable importance ranking for TP based on two criterion (left) Percent MSE and (right) node purity..... 121

Figure 82: Estimates of TP loads and total flow for each financial year for Aldgate Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue..... 122

Figure 83: Estimates of the flow weighted TP concentrations for Aldgate Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue. .... 122

Figure 84: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row 2) and TP (row3) data collected at Scott Creek (A5030502) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals..... 129

Figure 85: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row 2), TP (row3) data collected at Onkaparinga River at Houlgraves (A5030504) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals. .... 130

Figure 86: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row2), TP (row 3) data collected at Echunga Creek (A5030506) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals..... 131

Figure 87: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row2), TP (row3) data collected at Lenswood Creek (A5030507) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals..... 132



Figure 88: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row2), TP (row3) data collected at Aldgate Creek (A5030509) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals. .... 133

Figure 89: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row 2) and TP (row3) data collected at Cox Creek (A503026) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals..... 134

## List of Tables

Table 1: Summary of sites used in this study that span the Onkaparinga catchment with numbers of observations and years when TN, TP and TSS were collected.....	19
Table 2: Summary of errors in flow for each of the 7 sites in the Onkaparinga catchment. Note, as there is no information about the potential error in the gauge positioning, we have borrowed from GBR studies that suggest this is around 10%.....	27
Table 3: The nine parameters in the SIMHYD rainfall-runoff model. ....	28
Table 4: Prior distributions and posterior distribution summary statistics for Scott Creek. ....	28
Table 5: Prior distributions and posterior distribution summary statistics for the Onkaparinga River at Houlgraves. ....	32
Table 6: Prior distributions and posterior distribution summary statistics Echunga Creek. ....	35
Table 7: Prior distributions and posterior distribution summary statistics Lenswood Creek. ....	38
Table 8: Prior distributions and posterior distribution summary statistics Aldgate Creek. ....	41
Table 9: Prior distributions and posterior distribution summary statistics Cox Creek at Uraidla. ....	44
Table 10: Summary of parameter estimates from generalised additive model fit to TSS from Scott Creek (A5030526) that explains 50.1% of the variation in the data.....	51
Table 11: Summary of parameter estimates from generalised additive model fit to TN from Scott Creek (A5030526) that explains 48.7% of the variation in the data.....	51
Table 12: Summary of parameter estimates from generalised additive model fit to TP from Scott Creek (A5030526) that explains 41.1% of the variation in the data.....	51
Table 13: Summary of parameter estimates from generalised additive model fit to TSS from Onkaparinga River at Houlgraves (A5030504) that explains 50.3% of the variation in the data. ....	60
Table 14: Summary of parameter estimates from generalised additive model fit to TN from Onkaparinga River at Houlgraves (A5030504) that explains 73.5% of the variation in the data. ....	60
Table 15: Summary of parameter estimates from generalised additive model fit to TP from Onkaparinga River at Houlgraves (A5030504) that explains 57.4% of the variation in the data. ....	60
Table 16: Summary of parameter estimates from generalised additive model fit to SS from Echunga Creek (A5030506) that explains 45.5% of the variation in the data.....	67
Table 17: Summary of parameter estimates from generalised additive model fit to TN from Echunga Creek (A5030506) that explains 62.7% of the variation in the data.....	67
Table 18: Summary of parameter estimates from generalised additive model fit to TP from Echunga Creek (A5030506) that explains 64.4% of the variation in the data.....	69
Table 19: Summary of parameter estimates from generalised additive model fit to SS from Lenswood Creek (A5030507) that explains 54.2% of the variation in the data.....	77
Table 20: Summary of parameter estimates from generalised additive model fit to TN from Lenswood Creek (A5030507) that explains 65.6% of the variation in the data. ....	77
Table 21: Summary of parameter estimates from generalised additive model fit to TP from Lenswood Creek (A5030507) that explains 57.1% of the variation in the data.....	77
Table 22: Summary of parameter estimates from generalised additive model fit to SS from Aldgate Creek (A5030509) that explains 44.9% of the variation in the data.....	85
Table 23: Summary of parameter estimates from generalised additive model fit to TN from Aldgate Creek (A5030509) that explains 50.6% of the variation in the data.....	85
Table 24: Summary of parameter estimates from generalised additive model fit to TP from Aldgate Creek (A5030509) that explains 47.1% of the variation in the data.....	85
Table 25: Summary of parameter estimates from generalised additive model fit to TSS from Cox Creek (A5030526) that explains 56.9% of the variation in the data.....	96

Table 26: Summary of parameter estimates from generalised additive model fit to TN from Cox Creek (A5030526) that explains 58.9% of the variation in the data.....	96
Table 27: Summary of parameter estimates from generalised additive model fit to TP from Cox Creek (A5030526) that explains 57% of the variation in the data.....	97
Table 28: Land-use changes imposed in scenario 1.....	101
Table 29: Land-use changes imposed in scenario 2.....	102
Table 30: Land-use changes imposed in scenario 3.....	102
Table 31: Covariates used in site-based statistical models.....	128
Table 32: Land-uses in the catchment area contributing to the Scott Creek site. ....	135
Table 33: Land-uses in the catchment area contributing to the Houlgraves site.....	135
Table 34: Land-uses in the catchment area contributing to the Echunga Creek site. ....	136
Table 35: Land-uses in the catchment area contributing to the Lenswood Creek site. ....	136
Table 36: Land-uses in the catchment area contributing to the Aldgate Creek site. ....	137
Table 37: Land-uses in the catchment area contributing to the Cox Creek site.....	137

## Executive Summary

The investment into the quality of water supply arising from catchments of the Mount Lofty Ranges (MLR) watershed in South Australia is critical to ensure a safe and reliable water supply to the city of Adelaide. The MLR watershed occupies an area of 1,640 km<sup>2</sup> and houses a range of land-uses that include agricultural, urban and conservation areas. Soils in the region vary from sandy loam to clay and rock and rainfall across the region ranges between 600 to 1200 mm per year. In 1996, a composite sampler network was established in the MLR watershed to investigate the impact of particular land uses on water quality. A number of constituents have been studied since this time to monitor potential impacts on the water supply and the health of aquatic ecosystems.

The Mt Lofty Ranges watershed is comprised of a number of catchments consisting of the Torrens and Little Para catchments in the north and the Onkaparinga and Myponga catchments in the south. The focus of this study is the Onkaparinga catchment and contains a number of sub-catchment sites that can export very high nutrient loads during periods of intense runoff. The monitoring sites of interest in this report were chosen during a workshop with SA Water, SA EPA and SARDI for their importance in the catchment under study. These sites are summarised below.

Location	Site ID	TN		TP		TSS	
		#	Years	#	Years	#	Years
Scott Creek	A5030502	605	1996-2009	682	1996-2013	565	1999-2013
Onkaparinga River at Houlgraves	A5030504	194	2004-2009	677	1996-2013	560	1999-2013
Echunga Creek	A5030506	193	2004-2009	621	1996-2011	508	1999-2011
Lenswood Creek	A5030507	549	1996-2010	594	1996-2012	501	1999-2012
Aldgate Creek	A5030509	163	2004-2009	544	1996-2012	465	1999-2012
Cox Creek at Uraidla	A5030526	666	1996-2012	689	1996-2012	574	1999-2012

Flow data and water quality data is collected at gauges. Because the composite sampling network was established in 1996, historical data for flow date back further than for water quality. Furthermore, flow is measured at regular intervals (every 5 minutes) and is easily aggregated to obtain measurements of daily flow volumes. Composite water quality sampling results in flow-weighted samples of various constituents that are composited for collection every two to four weeks.

This report focuses on three key pollutants, namely total suspended sediment (TSS), total nitrogen (TN) and total phosphorous (TP). Using statistical models, we study the processes that drive hydrology and water quality in the Onkaparinga catchment and apply these in land-use change scenario modelling. Specifically, this report focuses on:

1. Applying a Bayesian calibration approach to calibrate the SIMHYD rainfall-runoff model for use in the Onkaparinga catchment and quantify potential sources of uncertainty in the hydrology.
2. Developing statistical models (site based models) for sites monitored in the Onkaparinga catchment in the MLR watershed for the purpose of quantifying constituent loads with an estimation of uncertainty.
3. Using statistical models to investigate three scenarios of land-use change and whether there are changes in loads and the uncertainty around loads.

Statistical models employed to address the above points, consisted of generalised additive models (GAMs) and generalised additive mixed models (GAMMs) through the Loads Regression Estimator (LRE) package that was developed for the quantification of loads for the Great Barrier Reef catchments. Site based models for the six sites studied in the Onkaparinga catchment used a variety of hydrological variables as covariates for understanding the variation in the data measured for each site. Specifically, these hydrological variables included flow, decomposed into baseflow and runoff as well as flow discounting terms that took into account past characteristics of the hydrograph. This could consist of a total accumulation of flow from the start of sampling to the short-term flow record prior to the current constituent sampled. Models were fit using the LRE package using the R statistical programming language.

Three scenarios that were explored as part of this report consisted of:

1. Investigating the sale of SA Water land holdings in Scott Creek sub-catchment
2. Quantifying the impact of continued expansion of perennial horticulture in the Cox Creek sub-catchment.
3. Quantifying the impact on water quality of infill within township boundaries of Aldgate Creek Railway Station.

These scenarios were determined at meetings with SA Water and SA EPA and were structured around the statistical modelling approach used to evaluate each scenario. A Random Forests modelling approach was used to develop a spatio-temporal model for each constituent across the six sites of interest in the Onkaparinga catchment. The model is non-parametric and popular in the machine learning and is based on decision tree methodology. The approach can take a large number of potential covariates as predictors to develop an ensemble of decision trees on bootstrap samples of the data. Variable importance rankings can assist in identifying important variables.

A summary of the findings from the statistical modelling performed in this report is provided below along with some suggestions in relation to the data collected, models fitted and interpretations from each type of model that can be taken forward into the future.

Item	Summary of Findings
<i>Site Based Statistical Models</i>	
1	The statistical models presented in this report should be regarded as a first (preliminary) investigation into the water quality of sites in the Onkaparinga catchment in the Mt Lofty Ranges. These models require some detailed investigations into their interpretation and the prediction of the loads as presented in this report.
2	The impact of the sedimentation pond at Brooks Bridge (upstream of the Cox Creek monitoring site) is not conclusive and highlights a complex relationship between the pond (once operational) and it's interaction with flow and (potentially) other factors. These models need careful interpretation and investigation to ensure all relationships are captured in the model.
3	While the site-based models for each constituent explained a large proportion of variation in the data, there were some difficulties noted for some sites and some constituents when extremes (high and low values of the constituent) were predicted. This may be due to the nature of sampling (i.e. composite sampling) and may require a more dedicated focus on capturing samples at those extremes.
4	A large proportion of the variation explained by the site based models for each constituent is hydrological (i.e. can be explained by patterns in the recent flows) rather than seasonal.
<i>Constituent and Flow Data</i>	
5	Measurements of flow and concentration for all constituents need to be carefully

	<p>examined for outliers and highly influential values. While every effort was made in this report to use reliable data, additional (new) data would be required if improved modelling outputs are considered necessary. Flow in particular for some sites exhibited some unusual patterns. While it appears that a dry spell may have contributed to low flow events, it would be useful to confirm that the data provided is accurate to ensure the predictions resulting from the statistical models is appropriate.</p>
<p><i>Scenarios</i></p>	
6	<p>Scenarios investigated through the Random Forest methodology were preliminary and we suggest that the specific scenarios that were implemented in this report be revisited for their suitability as there was considerable discussion over the duration of the project in relation to the land-uses considered.</p>
7	<p>Scenarios conducted within the statistical framework presented in this report allows for an assessment of confidence around the changes in loads observed. This is an advantage of the statistical modelling approach when compared to deterministic modelling approaches such as dynamic Sednet.</p>

## Acknowledgements

We wish to acknowledge Peta Hansen from the South Australian Environmental Protection Agency who provided the flow and rating curve data necessary for the statistical modelling work conducted as part of this project. We would also like to acknowledge the preliminary exploratory work conducted by Jess Ford (CSIRO) who provided collated water quality data that was utilised in the site based analysis and spatio-temporal analysis and Shaun Thomas for initial discussions had with him regarding the project. We also acknowledge the discussions and assistance of eWater throughout the dynamic Sednet development and the support of CSIRO in making this component of the work a planned investment for the future. Thanks are also due to Josh Bowden and Ondrej Hlinka for technical support for LibBi on the CSIRO GPU cluster for undertaking the rainfall-runoff calibrations. Finally, we wish to thank Sue Cuddy for her guidance and support throughout the final stages of this project and for reviewing this work in the lead up to the final report.

## Introduction

The investment into the quality of water supply arising from catchments of the Mount Lofty Ranges (MLR) watershed in South Australia is critical to ensure a safe and reliable water supply to the city of Adelaide. The MLR watershed occupies an area of 1,640 km<sup>2</sup> and houses a range of land-uses that include agricultural, urban and conservation areas. Soils in the region vary from sandy loam to clay and rock and rainfall across the region ranges between 600 to 1200 mm per year. Monitoring of the quality and quantity of water in MLR watershed has been ongoing since 1996. A number of constituents have been studied since this time through composite sampling measures to monitor potential impacts on the water supply and the health of aquatic ecosystems. This report focuses on three key pollutants, namely total suspended sediment (TSS), total nitrogen (TN) and total phosphorous (TP) and the potential mitigation strategies that could be implemented to reduce the impact of these constituents on waterways. Note we do not distinguish between the different species of nitrogen that comprise TN in the analyses that follow.

There have been a number of investigations into the quantification of constituent loads in the MLR watershed with the aim of assessing the impact to the Adelaide water supply (Anonymous, 2012, Cox et al., 2000, Cox et al. 2011, Dougherty et al., 2004, Fleming et al. 2001, Fleming et al. 2012, Fleming et al. 2010, Kirkby et al. 1997, Stevens et al. 1999). These studies have focussed on the Cox Creek and Onkaparinga River where the former of these has been identified as having poor water quality and requiring specialised treatment measures to ensure the water is safe to use. In recent years, there has been a focus towards constructing a sediment and nutrient budget through a catchment modelling tool such as Source (Welsh et al. 2013) that aims to spatially represent the movement of constituents in the catchment through a hydrological network. Flow is generated through a rainfall-runoff model that incorporates rainfall and potential evapotranspiration (PET) to generate flow based on one or more deterministic relationships. Constituent generation is based on a physical-process model (Source and earlier representations of catchment processes such as Sednet (Wilkinson et al. 2009) and CMSS (Davis and Farley, 1997) that identifies major sources, sinks and loads of sediments at a daily time step (Wilkinson et al. 2014). A more recent focus is the quantification of event mean concentrations (EMC) and dry weather concentrations (DWC) for calibrating the Source model for different land-uses within the MLR watershed (Fleming et al., 2010, Thomas et al., 2010). Calibration of a hydrological model implemented in Source was also investigated using the Parameter Estimation tool or PEST (Fleming et al. 2012).

While there has been considerable effort in applying these models in the MLR watershed, it was noted in Thomas et al. (2010) that a considerable improvement in modelling the TN, TP and TSS processes is required to ensure the model is applicable for the MLR watershed and has the capacity to support natural management policy and planning initiatives. As such, this report has focussed on three activities (outlined below) to assist in the delivery of a methodology that can assist in managing the water quality and quantity in the MLR watershed.

### Activity 1.

In the first activity, we applied a Bayesian calibration approach similar to that recently developed by Pagendam et al. (2014) to quantify uncertainties in flow data and obtain calibrations that acknowledge these uncertainties. This calibration involved the specification of a Bayesian Hierarchical Model (BHM) through three components: (i) a parameter model; (ii) a process model; and (iii) a data model. Each of these component models is used to formulate our scientific understanding about the relationship between rainfall and runoff and account for potential sources of uncertainty. The process model used in this activity was based upon the SIMHYD rainfall-runoff model (Chiew et al., 2002), which is a popular rainfall-runoff model in Source. The data model used

in this activity was based on a characterisation of error in the rating curve by comparing it to gaugings. Under the BHM formulation, the parameters of the SIMHYD rainfall-runoff model were estimated (with measures of uncertainty also provided) and the calibrated model was visually compared to the observed stream flow records.

#### Activity 2.

This second activity revolved around developing statistical models (site-based models and spatio-temporal models) for sites monitored in the Onkaparinga catchment in the MLR watershed for predicting concentrations of TSS, TN and TP. This work was based on methods developed in Great Barrier Reef (GBR) catchments (Kroon et al., 2011; Kuhnert et al., 2012) which rely on the use of Generalized Additive Models (GAMs) and Generalized Additive Mixed Models (GAMMs) (Wood, 2006) for the purpose of estimating constituent loads and the uncertainty around these estimates. The work conducted in Activity 1 was incorporated into these site-based models to provide an estimate of the error in flow rates. The GAMs and GAMMs used in this activity made use of a number of predictors based on important characteristics of flow in addition to flexible nonparametric spline terms.

#### Activity 3.

Activity three evaluated three land-use change scenarios within a statistical modelling framework, to investigate whether these resulted in changes in loads. The statistical models adopted in this activity were Random Forests (Breiman, 2001), which use decision trees constructed on the predictor variables to partition the observed data into homogenous groups and then apply simple prediction models within each group. Individual trees are created on bootstrap samples of the data and with random feature selection (random selection of predictors), with the Random Forest itself then constructed as an ensemble of trees (either regression or classification based), which when averaged, lead to more accurate predictions. The Random Forests were built using a variety of predictor variables including important characteristics of flow (as in Activity 2) as well as the proportions of the catchment in different land-use categories. Once these models were constructed from the observed data for the existing monitoring sites, predictions were made using modified land-use variables. For two of the scenarios, an increase in urbanisation was considered and the rainfall-runoff calibrations from Activity 1 were used to generate synthetic time series of stream flow by increasing SIMHYD's pervious fraction parameter to be in line with the new proportion of urban land-use in the catchment.

## Study Region

The Mt Lofty Ranges Watershed (Figure 1 (a)) is comprised of a number of catchments consisting of the Torrens and Little Para catchments in the north and the Onkaparinga and Myponga catchments in the south. The focus of this study is the Onkaparinga catchment (see Figure 1 (b)) and contains a number of sub-catchment sites that can export very high nutrient loads during periods of intense runoff. The monitoring sites of interest in this study are outlined in Table 1 and were identified at a workshop with SA Water, SA EPA and SARDI staff because they offered good spatial coverage of the Onkaparinga catchment. Daily measurements of flow are available from gauges at water quality monitoring sites over a number of decades (Figure 2). Within the catchment, flow and water quality data are collected at different temporal resolutions. Flow data is captured at regular intervals (daily), whereas water quality sampling conducted in the Onkaparinga catchment has concentrated on capturing data during low flow periods with large flows being captured a small proportion of the time. Events are typically measured using composite water quality samples, whereby a broad range of samples might be collected over the event. Figure 6 highlights the sampling distribution for TSS

stratified by percentiles of flow for the sites monitored in Table 1. It is apparent from these histograms that water quality samples are more representative of the lower percentiles of flow (baseflow) than the higher percentiles of flow. Similar plots arise for TN and TP. Figures 3 – 5 summarise the raw TSS, TN and TP data for the six sites investigated.

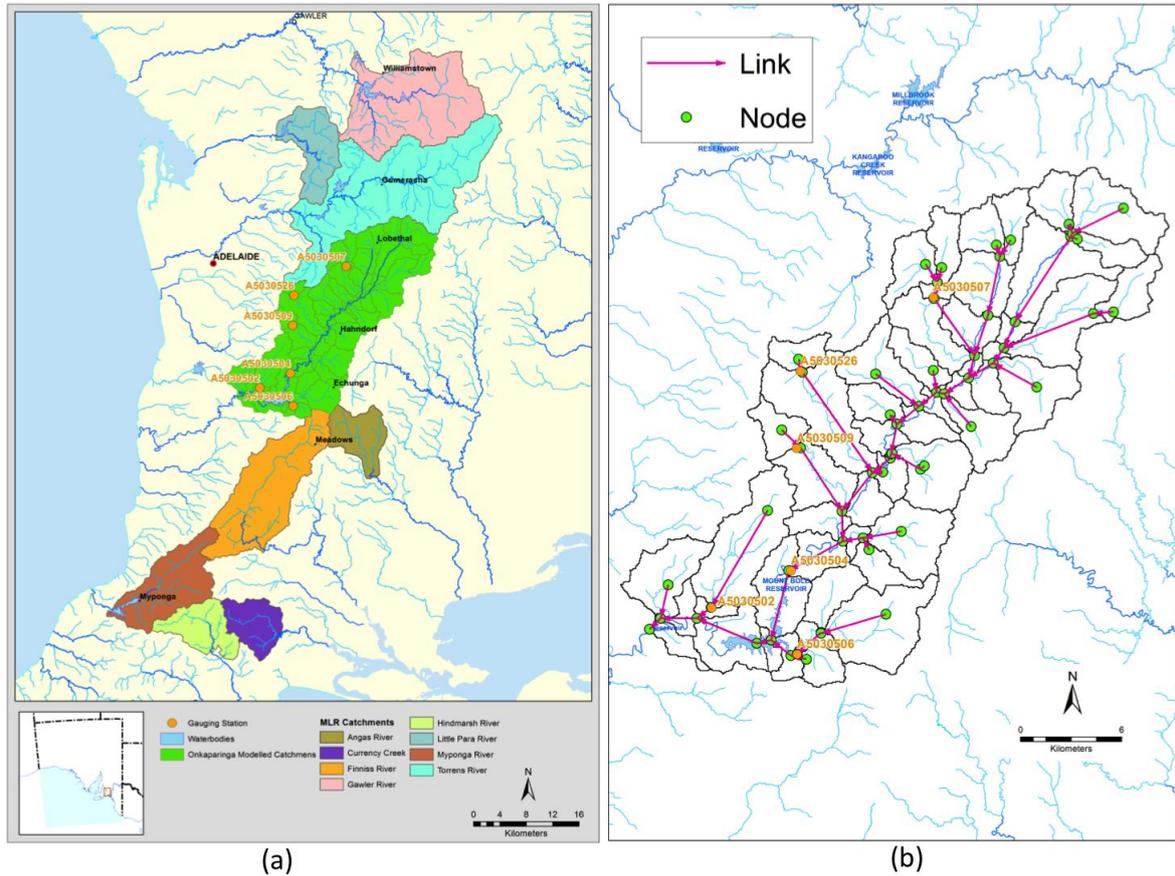


Figure 1: Maps showing (a) the Mt Lofty catchment and contributing areas and (b) the Onkaparinga catchment with sites that were investigated as part of this project overlaid.

Table 1: Summary of sites used in this study that span the Onkaparinga catchment with numbers of observations and years when TN, TP and TSS were collected.

Location	Site ID	TN		TP		TSS	
		#	Years	#	Years	#	Years
Scott Creek	A5030502	605	1996-2009	682	1996-2013	565	1999-2013
Onkaparinga River at Houlgraves	A5030504	194	2004-2009	677	1996-2013	560	1999-2013
Echunga Creek	A5030506	193	2004-2009	621	1996-2011	508	1999-2011
Lenswood Creek	A5030507	549	1996-2010	594	1996-2012	501	1999-2012
Aldgate Creek	A5030509	163	2004-2009	544	1996-2012	465	1999-2012
Cox Creek at Uraidla	A5030526	666	1996-2012	689	1996-2012	574	1999-2012

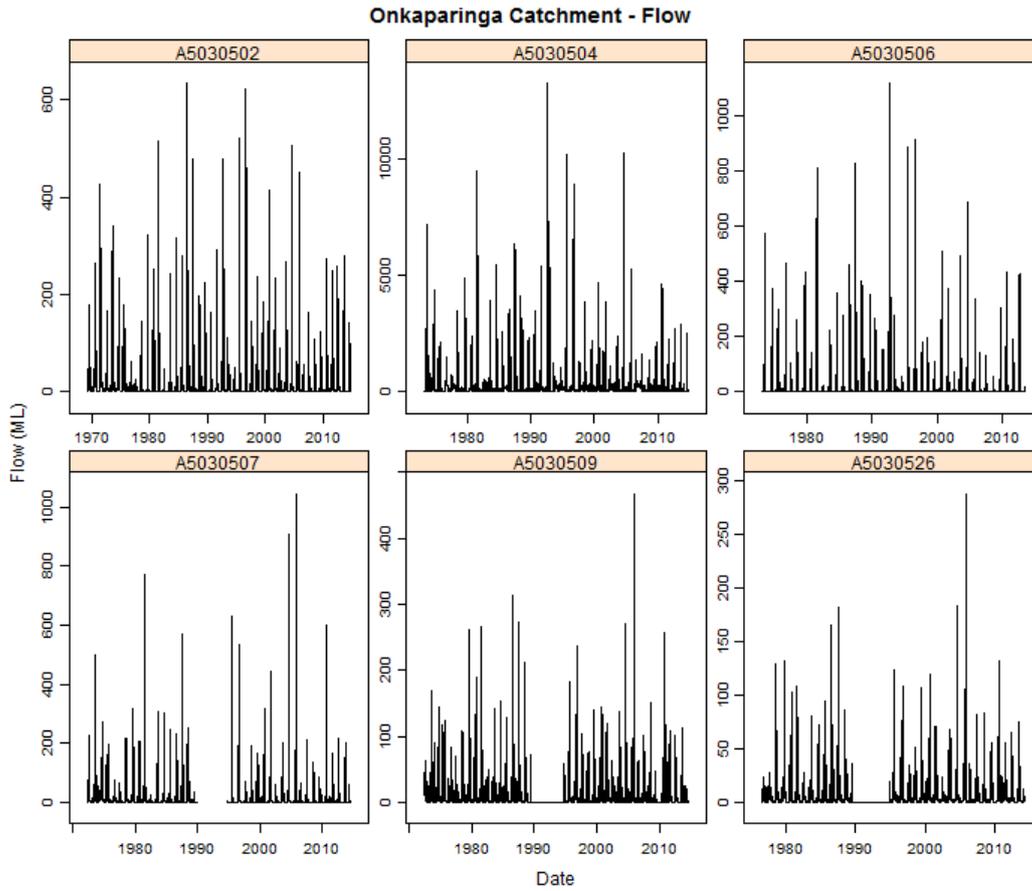


Figure 2: Raw flow data for the six sites in the Onkaparinga catchment.

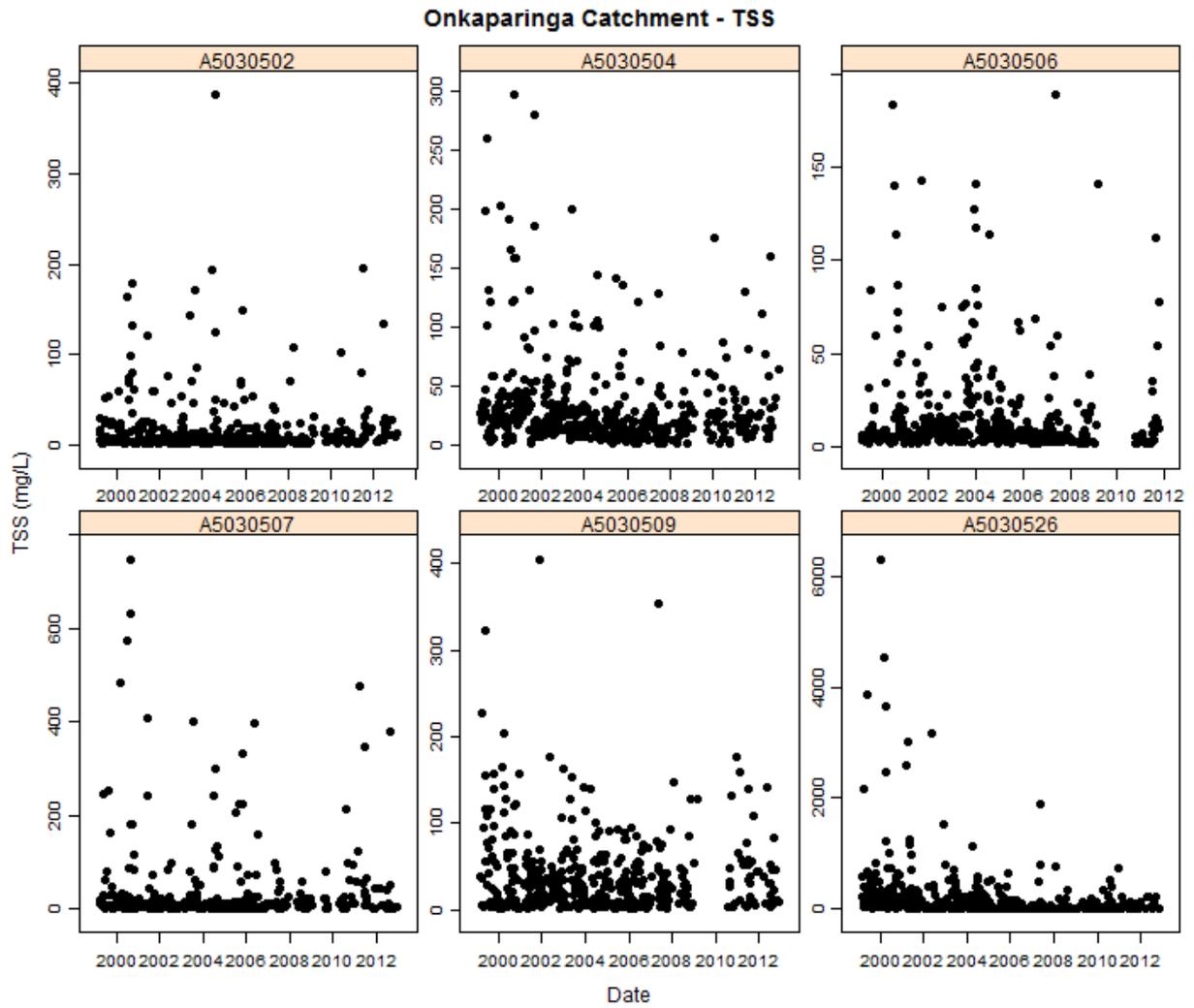


Figure 3: Plots of the raw data showing samples of TSS for the six sites in the Onkaparinga catchment.

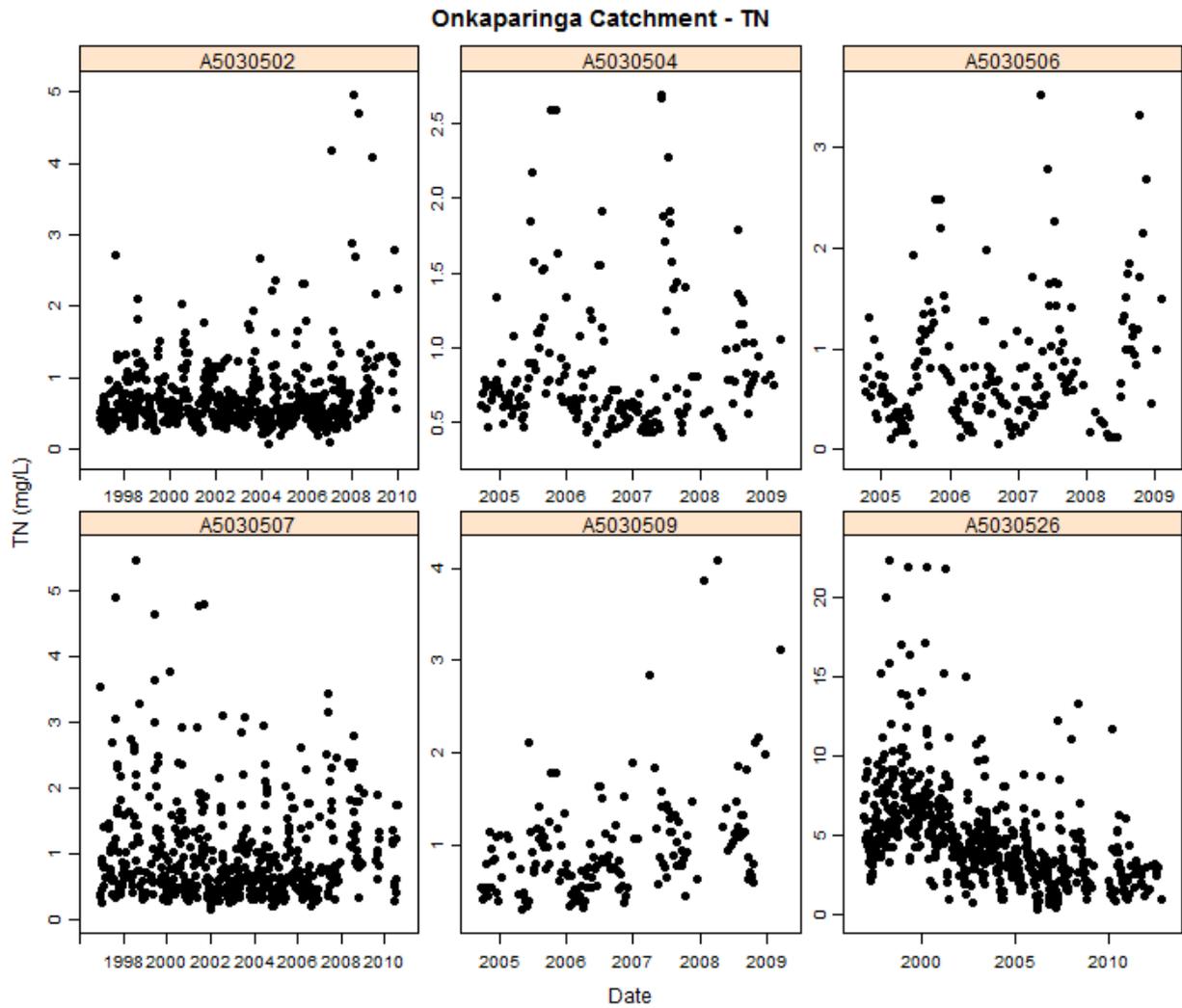


Figure 4: Plots of the raw data showing samples of TN for the six sites in the Onkaparinga catchment.

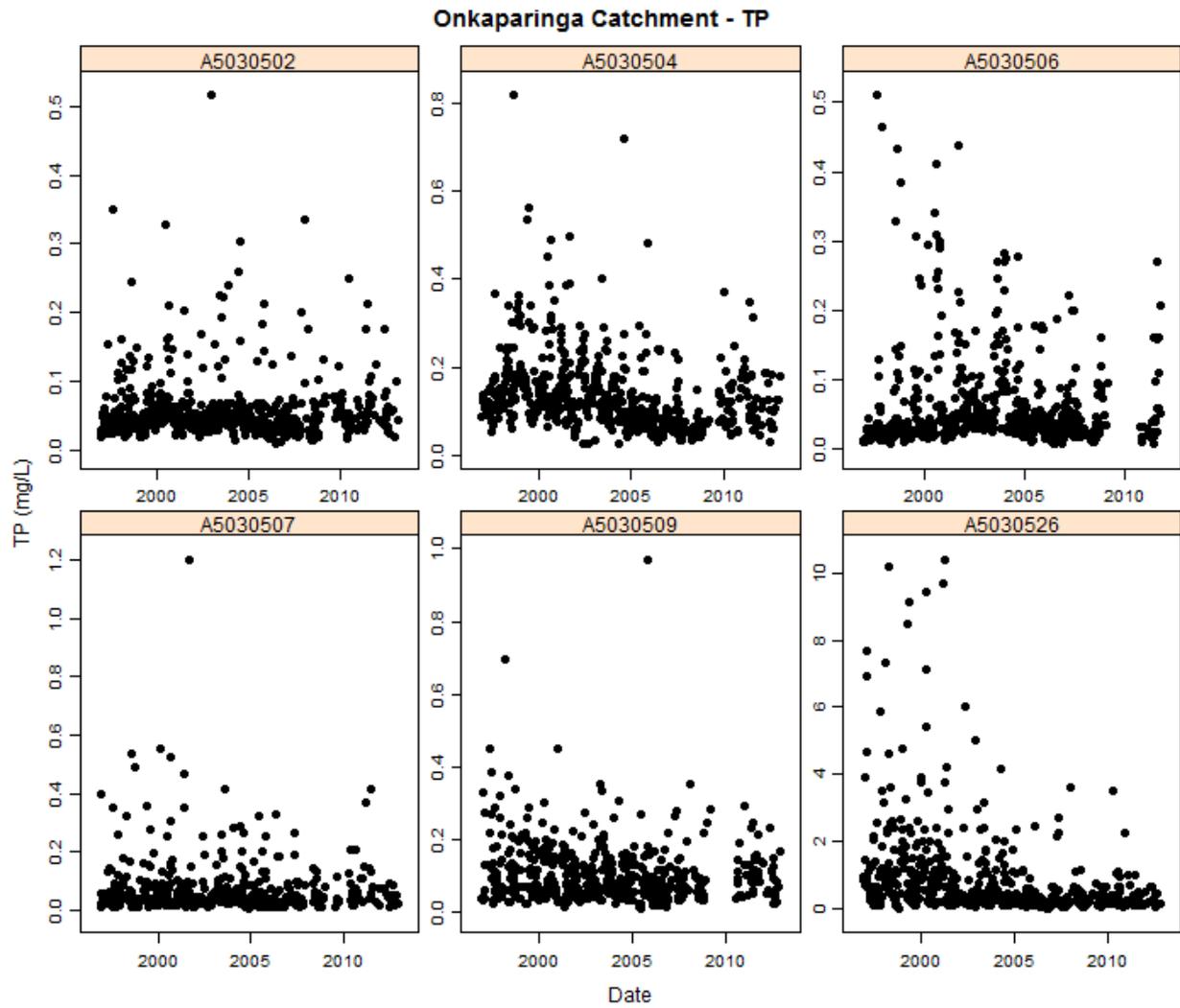


Figure 5: Plots of the raw data showing samples of TP for the six sites in the Onkaparinga catchment.

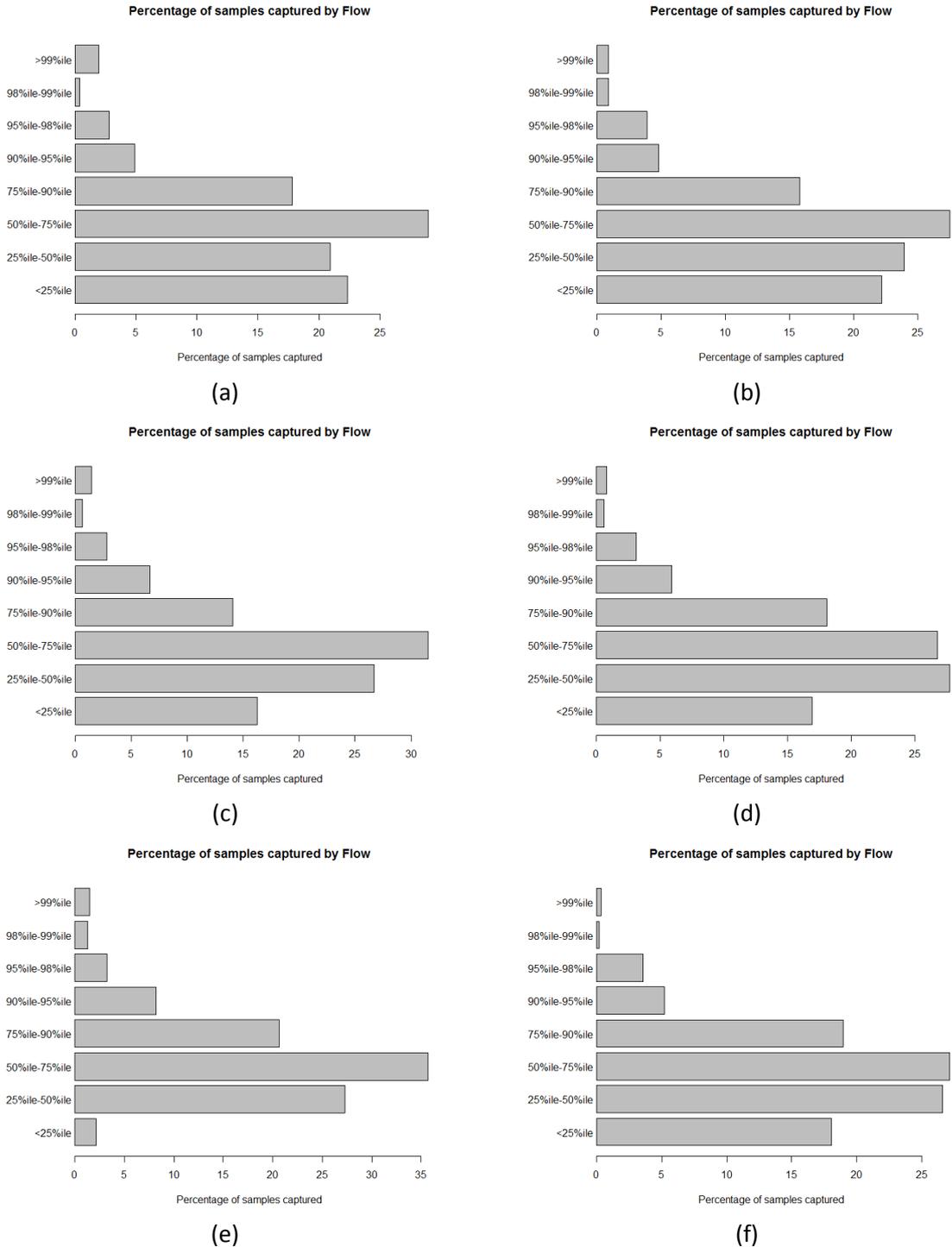


Figure 6: Summary of TSS samples stratified by flow and represented through a histogram for (a) Cox Creek (A5030526), (b) Scott Creek (A5030502), (c) Lenswood Creek (A5030507), (d) Echunga Creek (A5030506), (e) Aldgate Creek (A5030509); and (f) Onkaparinga River at Houlgraves (A5030504).

## Activity 1: Rainfall-Runoff Model Calibration & Error Quantification

### Motivation

As stated in the introduction, this report focussed on three key activities for assisting in managing the water quality and quantity in the MLR watershed. The first activity was to obtain rainfall-runoff calibrations that acknowledge uncertainties in the observed flow data and acknowledge the existence of model structural error (i.e. that the rainfall-runoff model itself is imperfect). Rainfall-runoff models are widely used in hydrology and allow stream flow to be predicted from time-series of rainfall and potential evapotranspiration. These models are a core component of catchment models such as Source and require careful calibration to observed flow data. Once calibrated, a rainfall runoff model can be used for a variety of purposes, including infilling periods of missing flow data in hydrographic records and predicting flow records in ungauged catchments.

This section demonstrates how a statistical modelling framework known as Bayesian Hierarchical Modelling (BHM) can be used to calibrate rainfall-runoff models whilst acknowledging uncertainty in the observed flow data and structural errors in the rainfall-runoff model. Specifically, this is demonstrated for six gauged sites in the Onkaparinga catchment with the SIMHYD rainfall-runoff model (Chiew et al., 2002). Using the BHM approach outlined in the subsequent section, we provide calibrated parameter sets for each of the six gauges so that these might be used in future modelling activities. These calibrations are in fact used in the section “Scenarios for the Onkaparinga Catchment”, where we investigate the likely water quality impacts from a number of scenarios.

### Overview of Statistical Modelling of Hydrographic Data

In order to use catchment models such as Source for studying the potential impacts of land-use change scenarios, reliable calibration of rainfall-runoff models are required. Typical calibrations rely on maximising a suitable objective function (usually some measure of fit between the model predictions and observed data). However, the modeller’s choice of this objective function typically subjective and often does not acknowledge various sources of error or uncertainty that exist in the data and the model itself. For example, hydrographic data is understood to have observation errors that one can quantify by comparing data from gaugings to rating curves. In addition, we cannot treat a rainfall-runoff model as being a perfect representation of the true stream flow, there are structural errors in these deterministic models that should be acknowledged and quantified. One could also argue that the rainfall time series that drives the rainfall-runoff model contains errors that could also be acknowledged, but these can also be accounted for to some degree through model structural error.

In this section, we undertake statistical analyses that estimate the most appropriate parameters for rainfall-runoff models, given the observed hydrographic data, whilst acknowledging: (i) uncertainty in the observed data; and (ii) model structural error. These analyses build on the methods used in the recent work of Pagendam et al. (2014). The statistical approach employed for these calibrations belongs to the Bayesian Hierarchical Modelling framework and, more specifically, is known as Bayesian State-Space Modelling (see Cressie and Wikle, 2011). In recent years the approach has

been steadily gaining in the hydrology literature (see Kuczera et al, 2006; Vrugt et al. 2008; Wu et al, 2010; Schmelter et al. 2011). Bayesian statistical methods typically rely on a computational approach known as Markov chain Monte Carlo (MCMC). For the work undertaken herein, the MCMC algorithm used was the Particle Marginal Metropolis-Hastings (PMMH) algorithm and this was implemented using the LibBi modelling language ([www.libbi.org](http://www.libbi.org)), on CSIRO's Bragg GPU cluster. For a thorough account of Bayesian Hierarchical Modelling for spatio-temporal statistics, we direct the reader to the book by Cressie and Wikle (2011). We outline the approach adopted in this activity below.

A Bayesian Hierarchical Model (BHM) of the type outlined by (Cressie and Wikle, 2011), consists of three component models: (i) a parameter model; (ii) a process model; and (iii) a data model. The parameter model summarises our prior beliefs about parameter values given previous studies or scientific knowledge from expert opinions, but can also be uninformative if there is little prior information to draw upon. The process model is constructed by conditioning on the parameter values and is usually based on some stochastic analogue of a deterministic process model, this is often achieved by adding Gaussian noise (or noise from some other distribution) to the process model. The data model is formulated as being conditional on a set of parameters and a realisation of the underlying process and is simply the likelihood of the data given the parameters and the true underlying process.

Statistical analyses were undertaken for six gauged sites within the Onkaparinga catchment in the Mount Lofty ranges watershed. The process model chosen to represent the relationship between rainfall and runoff was the SIMHYD model (Chiew et al., 2002), which is a popular model choice within the Source community. For each of these analyses, slightly different parameter models were used for each of the gauges (these are outlined in the section for each of the modelled gauges) and each calibration was performed assuming that the contributing area to each gauge behaved as a single homogeneous unit.

The process model employed in the BHM was:

$$Y_t \sim TN(q_t, \sigma_Y^2),$$

where  $Y_t$  is the true (latent) stream flow at time  $t$ ,  $TN(\mu, \sigma^2)$  represents a truncated-normal distribution with mean  $\mu$  and variance  $\sigma^2$  (truncated at zero, so that it's domain is the non-negative real line),  $y_t = (q_t, g_t, s_t)$  is the state of the SIMHYD rainfall runoff model at time  $t$ , with  $q_t$  denoting the total runoff,  $g_t$  the groundwater store and  $s_t$  the soil moisture store. From one time step to the next, the state of the SIMHYD model is propagated forward as,  $y_t = f(y_{t-1}, r_t, p_t)$ , where  $f(\cdot)$  denotes the operation performed by SIMHYD to propagate the state vector forward one day using the rainfall  $r_t$  and potential evapotranspiration (PET)  $p_t$ .

The data model used in the BHM was:

$$Z_t \sim TN(Y_t, (\gamma Y_t)^2),$$

where  $Z_t$  is the observed stream flow in  $\text{m}^3/\text{s}$  and  $\gamma$  is the coefficient of variation. The coefficient of variation was quantified offline by examining the deviations between derived flow from rating curves and flow measured during gaugings by hydrographers. These deviations were quantified for each of the six stations and the results are summarised in Table 2. The coefficients of variations are assumed known and fixed at these values for each of the analyses in the subsequent sections. Table 2 also provides values for the coefficient of variation in stream cross-section measurements that are

used in the statistical models discussed in section titled “Development of Site-Based Statistical Models”.

*Table 2: Summary of errors in flow for each of the 7 sites in the Onkaparinga catchment. Note, as there is no information about the potential error in the gauge positioning, we have borrowed from GBR studies that suggest this is around 10%.*

Location	Site ID	Rating Curve Variance (log scale)	Measurement Error (CV)	Cross-sectional Error (CV)
Scott Creek	A5030502	$(0.2830)^2$	0.2888	0.1
Onkaparinga River at Houlgraves	A5030504	$(0.0838)^2$	0.0839	0.1
Echunga Creek	A5030506	$(0.1280)^2$	0.1285	0.1
Lenswood Creek	A5030507	$(0.1389)^2$	0.1396	0.1
Aldgate Creek	A5030509	$(0.0326)^2$	0.0326	0.1
Cox Creek at Uraidla	A5030526	$(0.0594)^2$	0.0594	0.1

The output of the BHM analysis was a posterior probability distribution over the rainfall-runoff model parameters and the model structural error variance. These posterior distributions represented our understanding of these drivers of the system behaviour after having observed the hydrographic data. The primary motivation for these analyses was to obtain useful calibrations of the SIMHYD rainfall-runoff model at the six gauges in the Onkaparinga catchment. SIMHYD has nine parameters (see Table 3), however, we assumed that for each gauge, the pervious fraction was equal to the fraction that was not classified as “dense urban” land-use. The remaining eight parameters are “conceptual” parameters that do not correspond to actual measurable attributes and therefore do not have units associated with them. Uninformative prior distributions were placed on these parameters using uniform distributions spanning the range of parameters allowed by the SIMHYD model. The posterior distributions over the remaining eight parameters, provided probability distributions showing the likely values that these parameters should take, acknowledging the structural error in the model and the uncertainties in the hydrographic data. Following the statistical analysis, summary statistics of the posterior distributions (mean and standard deviation) for each of the calibrations (one for each gauging station) were reported and, using the posterior mean as the calibrated parameter value, the quality of the calibrated model was examined visually by plotting the modelled hydrographs over the observed flow data. This was carried out for periods both early and late in the gauging station’s historical record. The subsequent sections summarise these SIMHYD calibrations for each of the gauges, so that these parameter values might be used in future modelling activities.

Traditionally in the “calibration” of hydrological modelling, plots of modelled daily flow data against observed daily flow data have been used as a way of assessing goodness-of-fit. Under the BHM framework, we intentionally steer away from the use of these plots for two reasons: (i) the traditional approach treats discrepancies between the observed data and the model as arising because of error in the model and does not acknowledge that the data also contain error; and (ii) the BHM approach, results in a distribution for the flow on each day, which is difficult to plot in this way. We have therefore opted for plotting the estimated flow time series (samples from the posterior distribution) under the calibrated model over the observed data. This provides a visual depiction of the agreement between the data and the estimated flows and is common in modelling papers employing Bayesian Hierarchical models in surface water hydrology (see Vrugt et al. 2008; Pagendam et al. 2014). In our analyses we provide estimates of the error in stream flow data

derived from rating-curve information and posterior estimates of the error in the SIMHYD model. Both of these variances summarise the error between what we consider to be the true flow and the observations and model outputs respectively and are preferred under the BHM framework to other measures of goodness-of-fit (e.g. mean square error between model output and observed data) that might be used in simpler “model calibration” activities.

*Table 3: The nine parameters in the SIMHYD rainfall-runoff model.*

<b>SIMHYD Parameter</b>	<b>Parameter Description</b>
BFC	Baseflow Coefficient
ImpT	Impervious Threshold
InfC	Infiltration Coefficient
InfS	Infiltration Shape
IntC	Interflow Coefficient
RISC	Rainfall Interception Store Capacity
RC	Recharge Coefficient
SMSC	Soil Moisture Store Capacity
PF	Pervious Fraction

It should be noted, that our BHM used for identifying appropriate parameters for SIMHYD did not attempt to model the uncertainty around the forcing variables for rainfall and PET. Because these forcing variables are based on a complex pre processing of weather station data, it is difficult to formulate (by necessity) informative priors on these model inputs. Since error in rainfall and PET manifests itself as an error in flow, we effectively account for errors in these forcing variables through the model structural error.

## Scott Creek (A5030502)

The BHM constructed for stream flow at Scott Creek used rainfall, PET and stream flow data collected between 1/1/1980 and 9/9/2014. Rainfall and PET data were extracted from a Source model for the Onkaparinga as time series of spatially averaged SILO data over the contributing area to the gauge. Table 4 summarises the prior and marginal posterior distributions for each of the parameters in the model. Parameters denoted as “fixed” were assumed known when constructing the BHM and therefore have no summary statistics for the posterior distribution noted as NA (not applicable).

*Table 4: Prior distributions and posterior distribution summary statistics for Scott Creek.*

<b>Parameter</b>	<b>Prior</b>	<b>Posterior Mean</b>	<b>Posterior Std. Dev.</b>
BFC	Uniform(0.0, 1.0)	0.5095	0.01653
ImpT	Uniform(0.0, 5.0)	4.515	0.06926
InfC	Uniform(0.0, 400.0)	368.8	6.9786
InfS	Uniform(0.0, 10.0)	2.541	0.1711
IntC	Uniform(0.0, 1.0)	0.01044	0.009041
RISC	Uniform(0.0, 5.0)	2.981	0.08707
RC	Uniform(0.0, 1.0)	0.5372	0.03487
SMSC	Uniform(1.0, 500.0)	441.86	5.688

PF	Fixed at 0.9882	NA	NA
$\sigma_Y^2$	Uniform(0.0, 50.0)	0.08196	0.003346
$\gamma$	Fixed at 0.568	NA	NA

Modelled hydrographs were obtained by running SIMHYD using parameter values equal to the posterior means in Table 4. All of the parameters estimated had very tight posterior distributions compared to the diffuse prior distributions, indicating that the data was highly informative about all of the parameters and that there was a high degree of sensitivity to all of the parameters. Figures 7 and 8 overlay the modelled flow from SIMHYD with the observed flow data for a two-year period early in the time series (1985-1987) and later (2005-2007). These figures also show the trajectories of stream flow sampled from the posterior distribution of the BHM using MCMC in grey. These samples included a noise term to capture model structural error. For the Scott Creek site, there is very little difference between the BHM samples and the SIMHYD model, suggesting that the structural error contribution was small (this is confirmed by the fact that the posterior mean for  $\sigma_Y^2$  is small relative to flows).

Discrepancies between SIMHYD and observations in Figures 7(b) and 8(b) are attributable to a combination of model structural error and observation error in the flows. In modelling the latter error we assumed a constant coefficient of variation, resulting in larger absolute errors at higher flows than at lower flows. This explains the apparent close agreement at lower flows, with more obvious deviations at higher flows. A major difficulty in using deterministic rainfall-runoff models such as SIMHYD to mimic observed flows is that the rainfall input data can be error prone and that catchments can respond differently to the spatial distribution of rainfall. This source of error is most likely one of the main sources of discrepancy between the observed and the modelled series.

Whilst the calibration results obtained through the BHM estimation are satisfactory, there does appear to be a tendency for the SIMHYD model to have lower peaks than the observed series in high-flow events and higher peaks in some of the smaller events at this site.

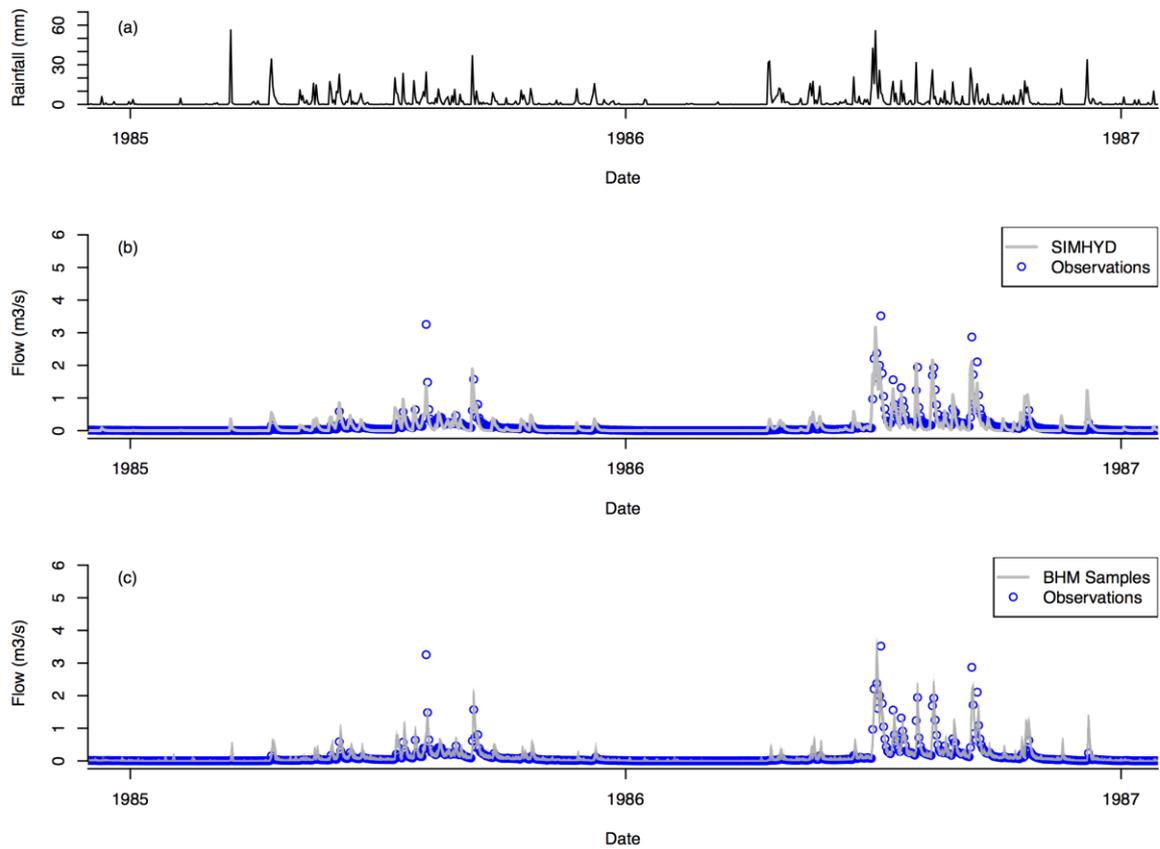


Figure 7: Rainfall-runoff model calibration results for Scott Creek (A5030502) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

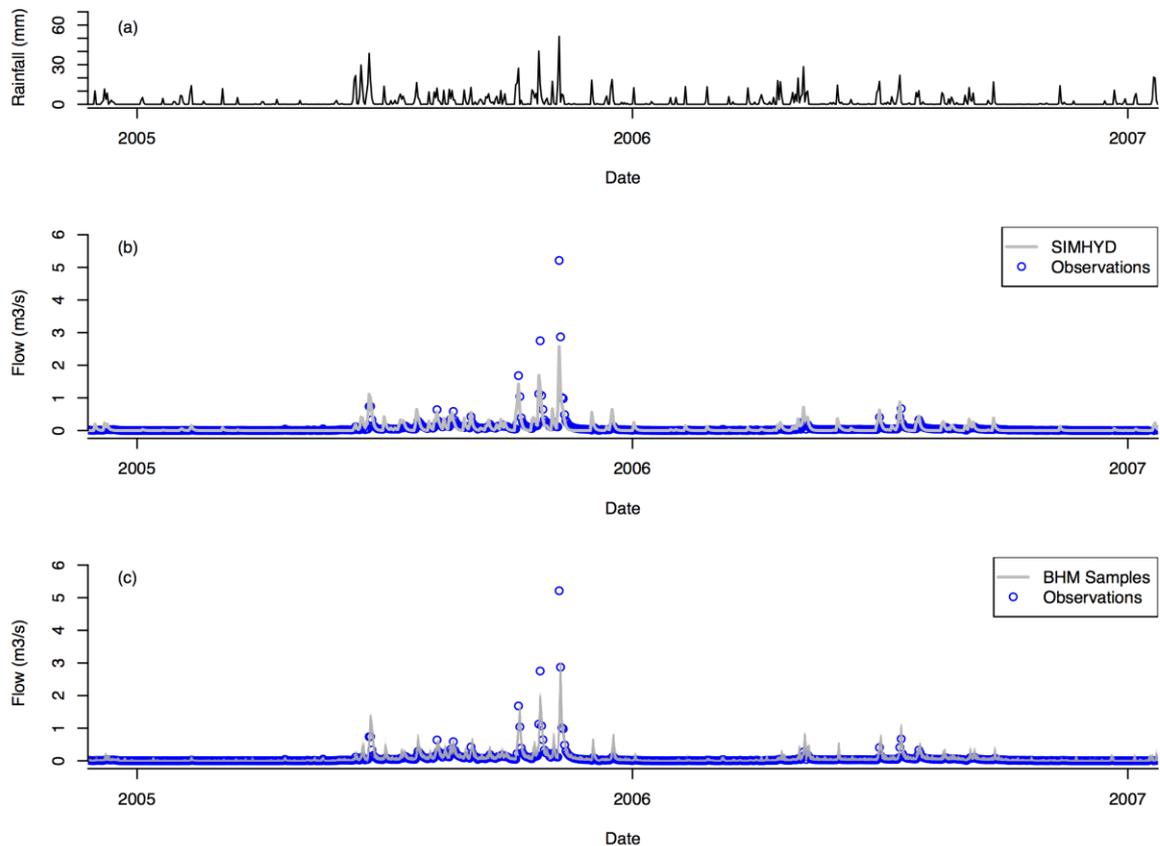


Figure 8: Rainfall-runoff model calibration results for Scott Creek (A5030502) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

## Onkaparinga River at Houlgraves (A5030504)

The BHM constructed for stream flow at Houlgraves Weir used rainfall, PET and stream flow data collected between 1/1/1980 and 9/9/2014. Rainfall and PET data were extracted from a Source model for the Onkaparinga as time series of spatially averaged SILO data over the contributing area to the gauge. Table 5 summarises the prior and marginal posterior distributions for each of the parameters in the model. Parameters denoted as “fixed” were assumed known when constructing the BHM and therefore have summary statistics of the posterior distribution noted as NA (not applicable).

Unlike the other sites in this study, the gauged flows at Onkaparinga are the sum of the natural flow from the contributing area and water that has been diverted from the Murray River and released upstream at Hahndorf Creek. The former component (i.e. natural flow) is what is modelled by a rainfall-runoff model and therefore to calibrate the SIMHYD model, it was first necessary to correct the observed flow data by subtracting the time series of diversions from the time series of flow at this site.

Modelled hydrographs were obtained by running SIMHYD using parameter values equal to the posterior means in Table 5. All of the parameters estimated had very tight posterior distributions compared to the diffuse prior distributions, indicating that the data was highly informative about all of the parameters and that there was a high degree of sensitivity to all of the parameters. Figures 9 and 10 overlay the modelled flow from SIMHYD with the observed flow data for a two-year period early in the time series (1985-1987) and later (2005-2007). These figures also show the trajectories of stream flow sampled from the posterior distribution of the BHM using MCMC in grey. These samples included a noise term to capture model structural error. For the Houlgraves site, there are some noticeable differences between the SIMHYD model and the BHM samples, particularly at lower flows. The structural error in the BHM samples provides a mechanism for the stochastic model to provide better agreement with the data. We note however, that the discrepancies at lower flows may not have been due to model structural error, but possibly an artefact of this flow data having been corrected for the effects of water diverted from the Murray River and entering at Hahndorf.

*Table 5: Prior distributions and posterior distribution summary statistics for the Onkaparinga River at Houlgraves.*

Parameter	Prior	Posterior Mean	Posterior Std. Dev.
BFC	Uniform(0.0, 1.0)	0.2208	0.01482
ImpT	Uniform(0.0, 5.0)	1.527	0.05130
InfC	Uniform(0.0, 400.0)	310.8	2.541
InfS	Uniform(0.0, 10.0)	7.757	0.07199
IntC	Uniform(0.0, 1.0)	0.5938	0.02017
RISC	Uniform(0.0, 5.0)	4.559	0.06048
RC	Uniform(0.0, 1.0)	0.9152	0.003087
SMSC	Uniform(1.0, 500.0)	480.7	0.2356
PF	Fixed at 0.9446	NA	NA
$\sigma_Y^2$	Uniform(0.0, 50.0)	3.921	0.01338
$\gamma$	Fixed at 0.296	NA	NA

Figures 9(b) and 10(b) show reasonable agreement between the SIMHYD modelled and observed flow data. The SIMHYD flow contains a number of small “phantom” events not present in the observed data. This is either as a result of: (i) spurious peaks in the rainfall time series that drive the SIMHYD model (possibly rainfall that fell in the catchment but did not lead to events at the gauge); or (ii) the erroneous removal of small events from the observation series in correcting for diversions at Hahndorf.

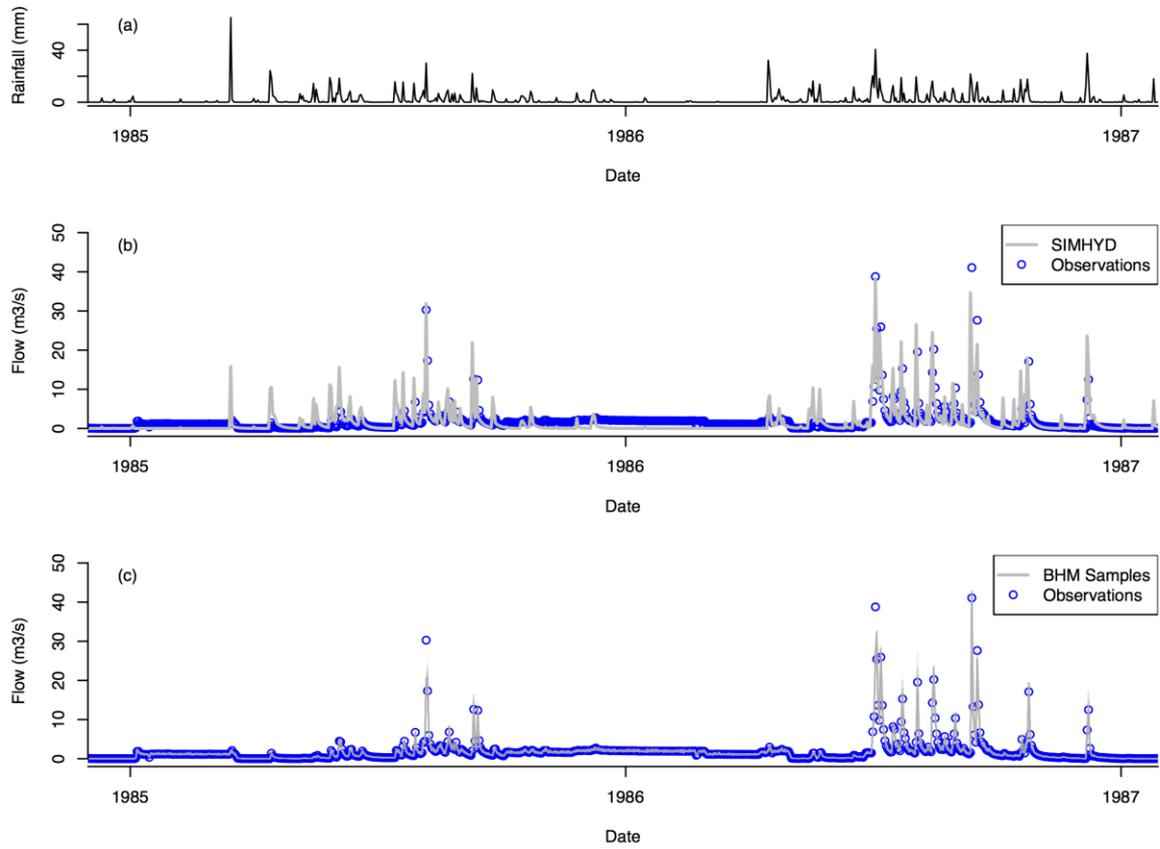


Figure 9: Rainfall-runoff model calibration results for the Onkaparinga River at Houlgraves (A5030504) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

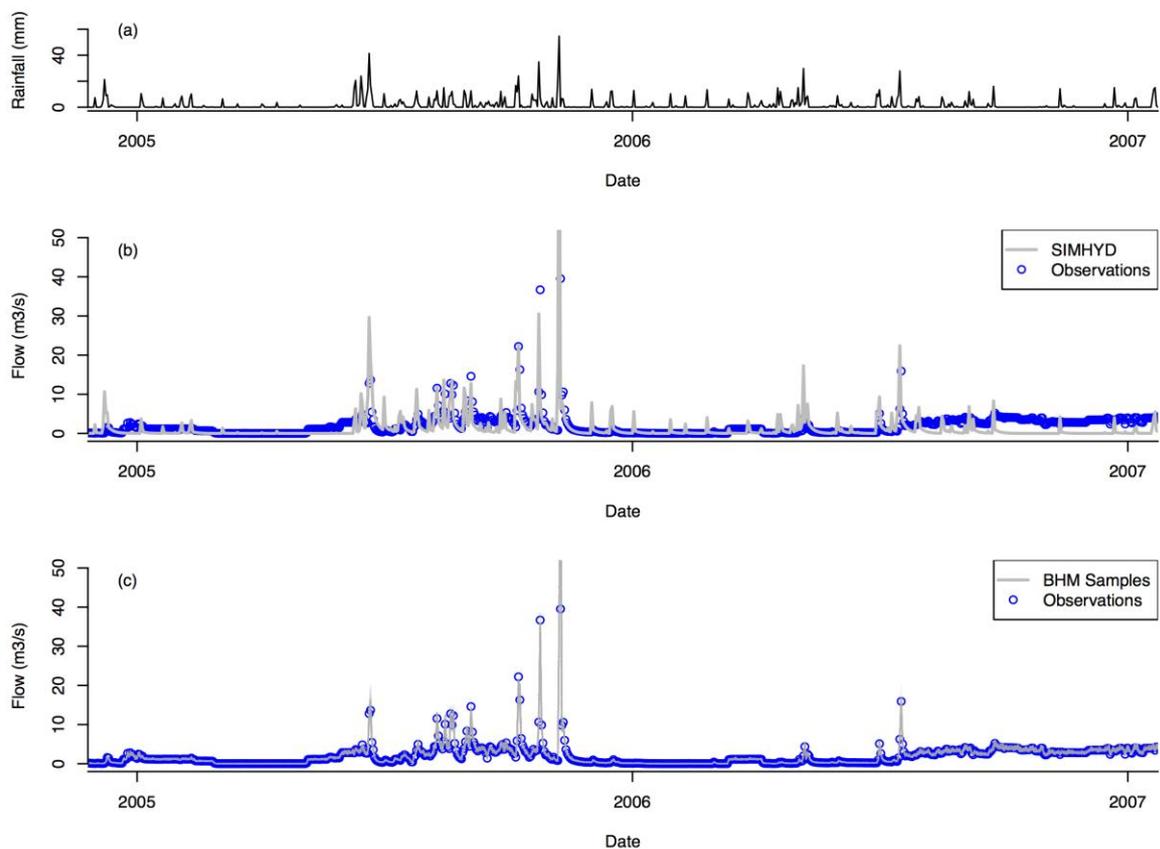


Figure 10: Rainfall-runoff model calibration results for the Onkaparinga River at Houlgraves (A5030504) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

## Echunga Creek (A5030506)

The BHM constructed for stream flow at Echunga Creek used rainfall, PET and stream flow data collected between 1/1/1980 and 21/6/2014. Rainfall and PET data were extracted from a Source model for the Onkaparinga as time series of spatially averaged SILO data over the contributing area to the gauge. Table 6 summarises the prior and marginal posterior distributions for each of the parameters in the model. Parameters denoted as “fixed” were assumed known when constructing the BHM and therefore have summary statistics of the posterior distribution noted as NA (not applicable).

Modelled hydrographs were obtained by running SIMHYD using parameter values equal to the posterior means in Table 6. All of the parameters estimated had very tight posterior distributions compared to the diffuse prior distributions, indicating that the data was highly informative about all of the parameters and that there was a high degree of sensitivity to all of the parameters. Figures 11 and 12 overlay the modelled flow from SIMHYD with the observed flow data for a two-year

period early in the time series (1985-1987) and later (2005-2007). These figures also show the trajectories of stream flow sampled from the posterior distribution of the BHM using MCMC in grey. These samples included a noise term to capture model structural error. There is minimal difference between the BHM samples and SIMHYD, suggesting that the estimated model structural error was small. Most notably, the BHM has corrected for the amplitude of small “phantom” events appearing in the SIMHYD output, but not in the observed flow data (see for example the event in Figures 11(b) and (c) just prior to 1987).

*Table 6: Prior distributions and posterior distribution summary statistics Echunga Creek.*

Parameter	Prior	Posterior Mean	Posterior Std. Dev.
BFC	Uniform(0.0, 1.0)	0.2631	0.009358
ImpT	Uniform(0.0, 5.0)	4.600	0.04618
InfC	Uniform(0.0, 400.0)	263.6	2.233
InfS	Uniform(0.0, 10.0)	3.689	0.08585
IntC	Uniform(0.0, 1.0)	0.1931	0.005312
RISC	Uniform(0.0, 5.0)	2.155	0.03049
RC	Uniform(0.0, 1.0)	0.4143	0.01648
SMSC	Uniform(1.0, 500.0)	311.5	7.237
PF	Fixed at 0.9880	NA	NA
$\sigma_Y^2$	Uniform(0.0, 50.0)	0.3265	0.008075
$\gamma$	Fixed at 0.369	NA	NA

Overall, there is reasonable agreement between the SIMHYD output and the observed data in Figures 11(b) and 12(b). The SIMHYD calibration does not always appear to have peak flows that agree with observed data during events, but as with Scott Creek, we need to acknowledge that the calibration accounts for potential error in the observed data and that this has been modelled as having a constant coefficient of variation. We therefore allow for greater absolute error between model and observed data at higher flows than at lower flows.

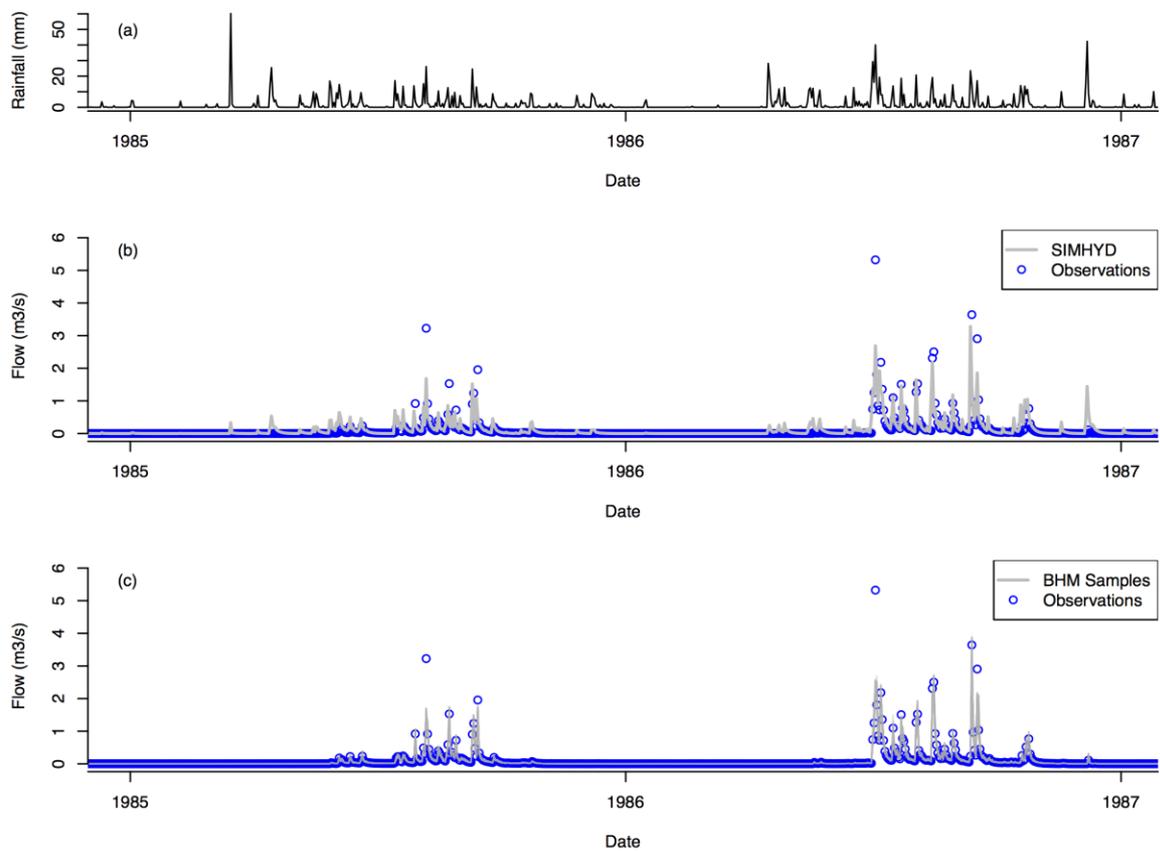


Figure 11: Rainfall-runoff model calibration results for Echung Creek (A5030506) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

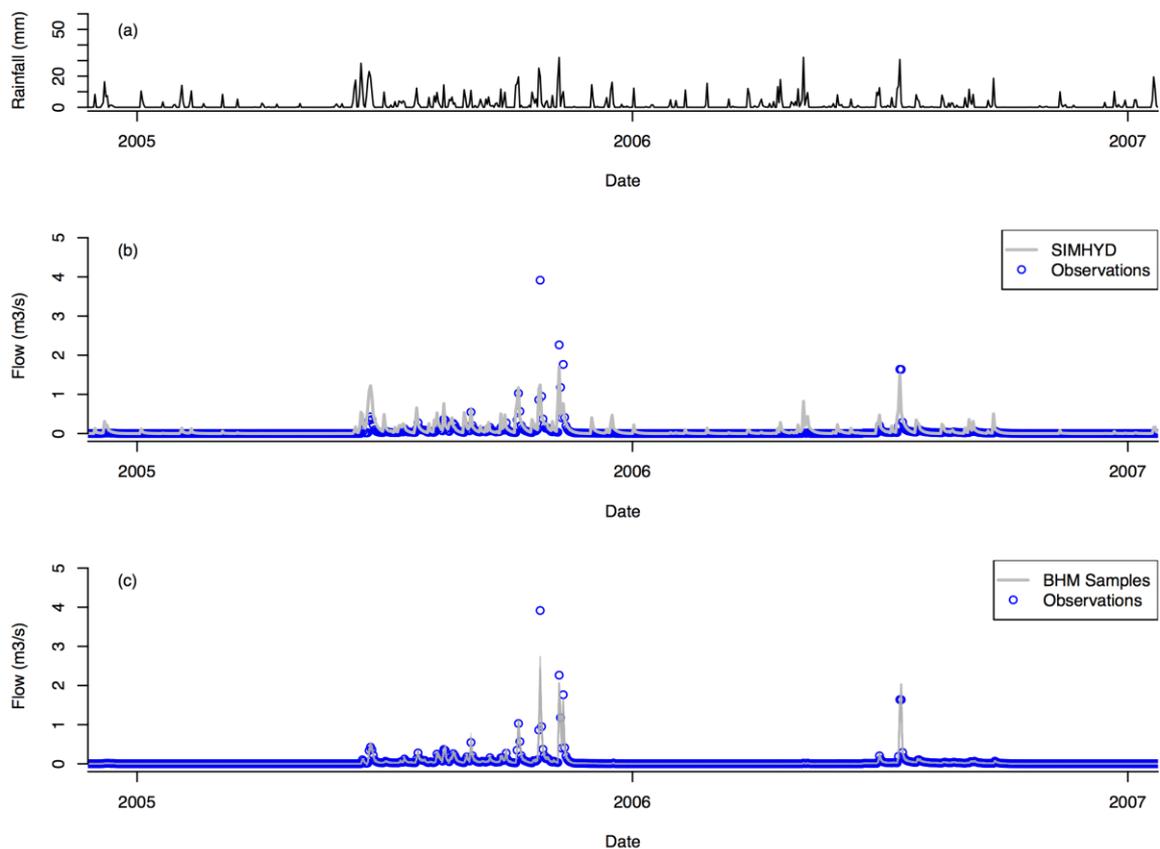


Figure 12: Rainfall-runoff model calibration results for Echunga Creek (A5030506) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

## Lenswood Creek (A5030507)

The BHM constructed for stream flow at Lenswood Creek used rainfall, PET and stream flow data collected between 1/1/1980 and 21/6/2014. Rainfall and PET data were extracted from a Source model for the Onkaparinga as time series of spatially averaged SILO data over the contributing area to the gauge. Table 7 summarises the prior and marginal posterior distributions for each of the parameters in the model. Parameters denoted as “fixed” were assumed known when constructing the BHM and therefore have summary statistics of the posterior distribution noted as NA (not applicable).

Modelled hydrographs were obtained by running SIMHYD using parameter values equal to the posterior means in Table 7. All of the parameters estimated had very tight posterior distributions compared to the diffuse prior distributions, indicating that the data was highly informative about all of the parameters and that there was a high degree of sensitivity to all of the parameters. Figures 13 and 14 overlay the modelled flow from SIMHYD with the observed flow data for a two-year period early in the time series (1985-1987) and later (2005-2007). These figures also show the trajectories of stream flow sampled from the posterior distribution of the BHM using MCMC in grey.

These samples included a noise term to capture model structural error. Unlike the previous examples, in Figures 13(c) and 14(c) there are a number of gaps in the observation record which have been stochastically infilled by the BHM. The noise in these stochastic trajectories reflects the model structural error in contrast to the deterministic infilling by SIMHYD seen in Figures 13(b) and 14(b).

*Table 7: Prior distributions and posterior distribution summary statistics Lenswood Creek.*

Parameter	Prior	Posterior Mean	Posterior Std. Dev.
BFC	Uniform(0.0, 1.0)	0.7684	0.01128
ImpT	Uniform(0.0, 5.0)	3.607	0.04588
InfC	Uniform(0.0, 400.0)	271.8	1.861
InfS	Uniform(0.0, 10.0)	1.939	0.1014
IntC	Uniform(0.0, 1.0)	0.1112	0.03103
RISC	Uniform(0.0, 5.0)	2.742	0.03875
RC	Uniform(0.0, 1.0)	0.7989	0.02408
SMSC	Uniform(1.0, 500.0)	495.5	4.621
PF	Fixed at 0.9872	NA	NA
$\sigma_Y^2$	Uniform(0.0, 50.0)	0.5469	0.003965
$\gamma$	Fixed at 0.386	NA	NA

In general, there is satisfactory agreement between the SIMHYD calibrated flow and the observed data during both high and low flows. One criticism of the calibration might be that the recession of flow post-event appears to occur more rapidly in the SIMHYD calibration than in the observed data. This mismatch has been corrected for in the BHM samples through the model structural error term. As with all deterministic rainfall-runoff models, there is also the afore-mentioned problem of large rainfall events that are not realised as events in the flow record. An example of such rainfall event can be seen following the main flow events in late 2005.

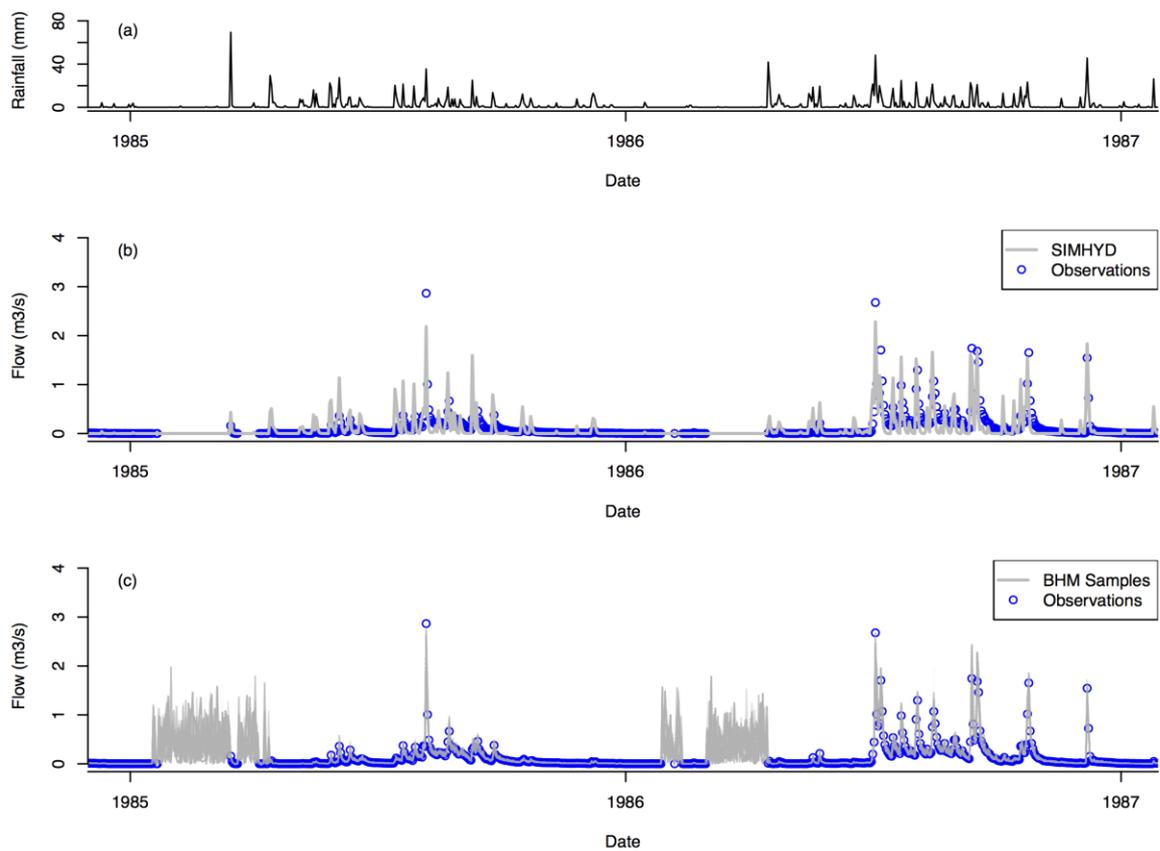


Figure 13: Rainfall-runoff model calibration results for Lenswood Creek (A5030507) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

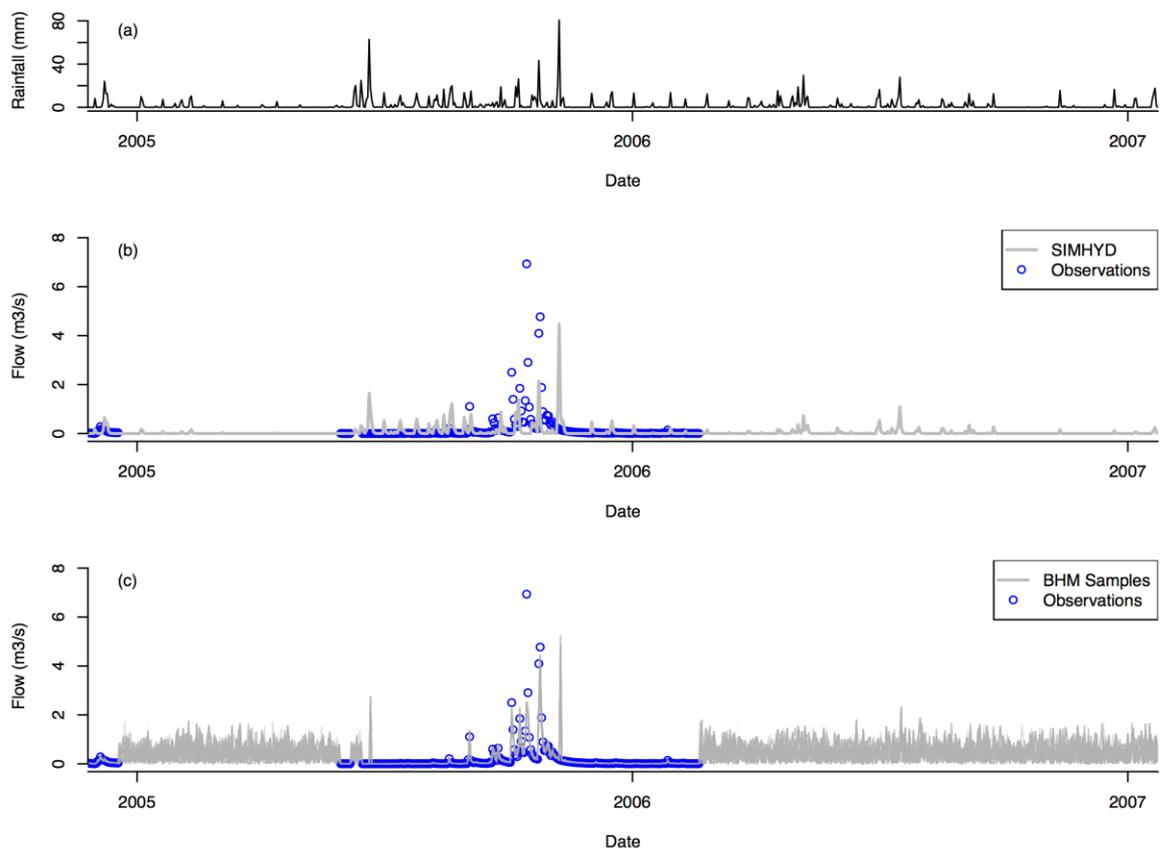


Figure 14: Rainfall-runoff model calibration results for Lenswood Creek (A5030507) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

## Aldgate Creek (A5030509)

The BHM constructed for stream flow at Aldgate Creek used rainfall, PET and stream flow data collected between 1/1/1980 and 11/6/2014. Rainfall and PET data were extracted from a Source model for the Onkaparinga as time series of spatially averaged SILO data over the contributing area to the gauge. Table 8 summarises the prior and marginal posterior distributions for each of the parameters in the model. Parameters denoted as “fixed” were assumed known when constructing the BHM and therefore have summary statistics of the posterior distribution noted as NA (not applicable).

Modelled hydrographs were obtained by running SIMHYD using parameter values equal to the posterior means in Table 8. All of the parameters estimated had very tight posterior distributions compared to the diffuse prior distributions, indicating that the data was highly informative about all of the parameters and that there was a high degree of sensitivity to all of the parameters. Figures 15 and 16 overlay the modelled flow from SIMHYD with the observed flow data for a two-year period early in the time series (1985-1987) and later (2005-2007). These figures also show the trajectories of stream flow sampled from the posterior distribution of the BHM using MCMC in grey.

These samples included a noise term to capture model structural error. Whilst the sampled BHM trajectories in Figures 15(c) and 16(c) agree well with the observed flow, there is much larger discrepancy for the SIMHYD calibration shown in Figures 15(b) and 16(b). Most notably, there is a tendency for SIMHYD to overestimate flows during most events as well as a large number of “phantom events” in the modelled hydrographs. The latter appears to be due in part to a rainfall series that is either: (i) not representative of the rainfall that generates flow in the catchment; or (ii) contains errors.

*Table 8: Prior distributions and posterior distribution summary statistics Aldgate Creek.*

Parameter	Prior	Posterior Mean	Posterior Std. Dev.
BFC	Uniform(0.0, 1.0)	0.2775	0.003277
ImpT	Uniform(0.0, 5.0)	1.491	0.01671
InfC	Uniform(0.0, 400.0)	385.4	1.049
InfS	Uniform(0.0, 10.0)	4.878	0.02044
IntC	Uniform(0.0, 1.0)	0.9252	0.002706
RISC	Uniform(0.0, 5.0)	3.136	0.01111
RC	Uniform(0.0, 1.0)	0.1423	0.003133
SMSC	Uniform(1.0, 500.0)	368.2	1.765
PF	Fixed at 0.4562	NA	NA
$\sigma_{\gamma}^2$	Uniform(0.0, 50.0)	0.8228	0.001927
$\gamma$	Fixed at 0.182	NA	NA

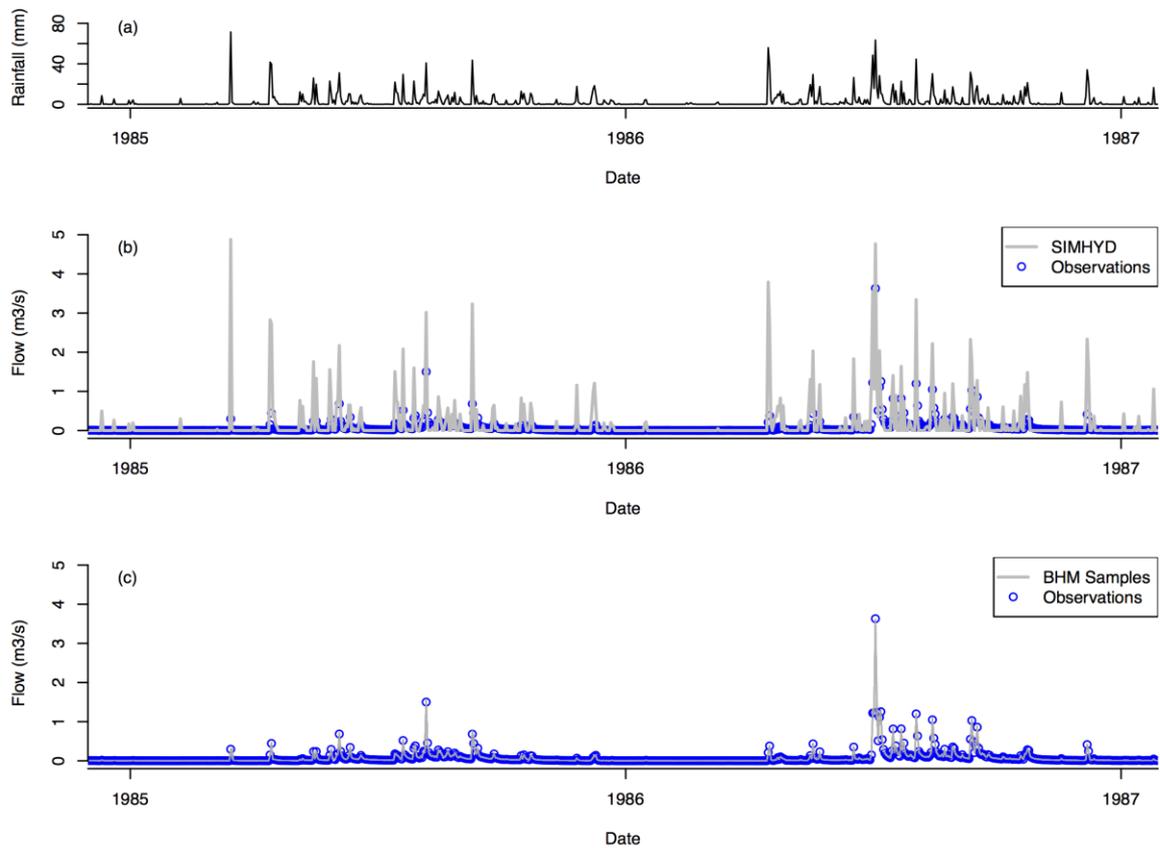


Figure 15: Rainfall-runoff model calibration results for Aldgate Creek (A5030509) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

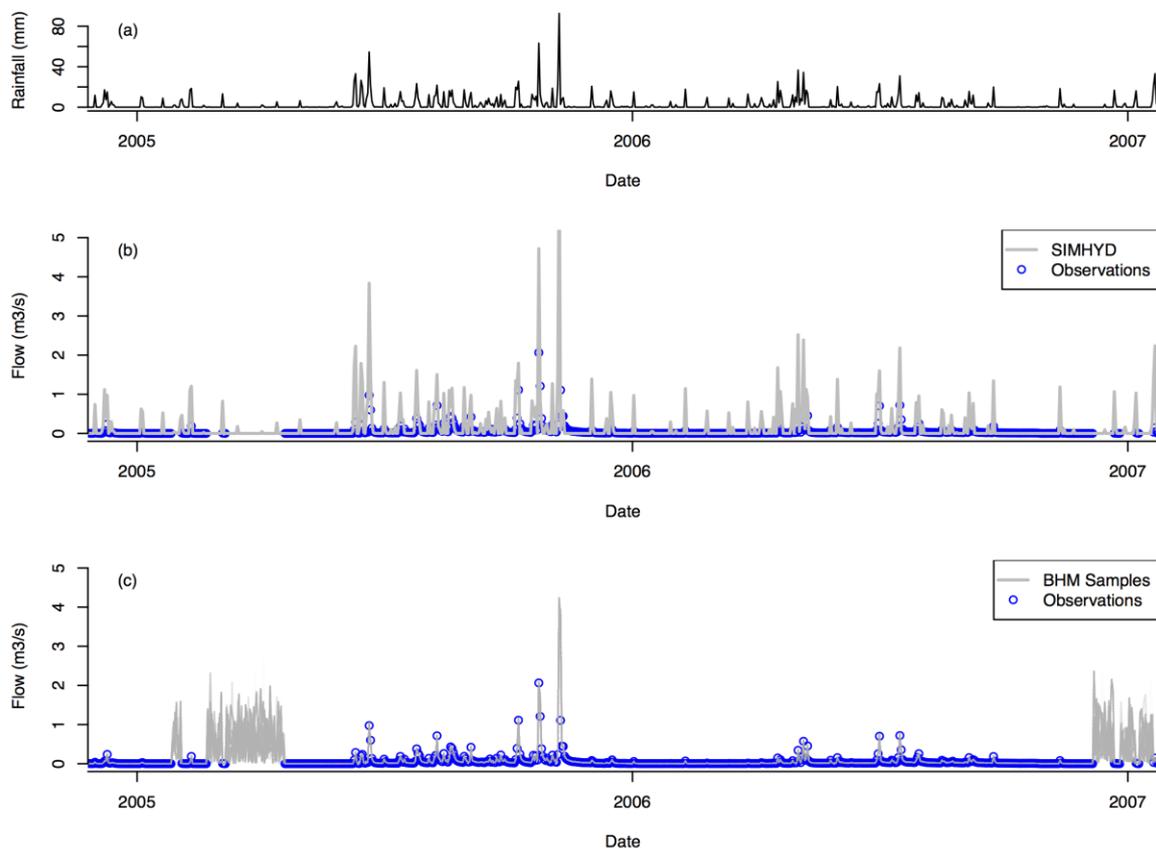


Figure 16: Rainfall-runoff model calibration results for Aldgate Creek (A5030509) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

## Cox Creek at Uraidla (A5030526)

The BHM constructed for stream flow at Cox Creek used rainfall, PET and stream flow data collected between 1/1/1980 and 16/6/2014. Rainfall and PET data were extracted from a Source model for the Onkaparinga as time series of spatially averaged SILO data over the contributing area to the gauge. Table 9 summarises the prior and marginal posterior distributions for each of the parameters in the model. Parameters denoted as “fixed” were assumed known when constructing the BHM and therefore have summary statistics of the posterior distribution noted as NA (not applicable).

Modelled hydrographs were obtained by running SIMHYD using parameter values equal to the posterior means in Table 9. All of the parameters estimated had very tight posterior distributions compared to the diffuse prior distributions, indicating that the data was highly informative about all of the parameters and that there was a high degree of sensitivity to all of the parameters. Figures 17 and 18 overlay the modelled flow from SIMHYD with the observed flow data for a two-year period early in the time series (1985-1987) and later (2005-2007). These figures also show the trajectories of stream flow sampled from the posterior distribution of the BHM using MCMC in grey.

These samples included a noise term to capture model structural error, the influence of which can be seen in Figure 17(c), where a small gap in the observed data has been stochastically in-filled just prior to 1987. In general, there is good agreement between the SIMHYD and the observed data at both high and low flows and minimal difference between the deterministic SIMHYD modelled flows and the flows sampled from the BHM. The latter point highlights that there was little structural error estimated through the BHM, which is supported by the small value of  $\sigma_{\gamma}^2$  (equal to 0.135) relative to flows.

*Table 9: Prior distributions and posterior distribution summary statistics Cox Creek at Uraidla.*

Parameter	Prior	Posterior Mean	Posterior Std. Dev.
BFC	Uniform(0.0, 1.0)	0.2261	0.005003
ImpT	Uniform(0.0, 5.0)	2.742	0.01957
InfC	Uniform(0.0, 400.0)	370.5	1.548
InfS	Uniform(0.0, 10.0)	2.436	0.06078
IntC	Uniform(0.0, 1.0)	0.1361	0.009319
RISC	Uniform(0.0, 5.0)	0.2541	0.06880
RC	Uniform(0.0, 1.0)	0.4402	0.01305
SMSC	Uniform(1.0, 500.0)	307.4	1.936
PF	Fixed at 0.9106	NA	NA
$\sigma_{\gamma}^2$	Uniform(0.0, 50.0)	0.1354	0.004492
$\gamma$	Fixed at 0.247	NA	NA

In stark contrast to many of the other sites considered, Cox Creek did not appear to suffer from the presence of spurious “phantom” events appearing in the SIMHYD output. This suggested that the rainfall series for this site was fairly reliable and representative of the rainfall that actually drove the flow observed in the field.

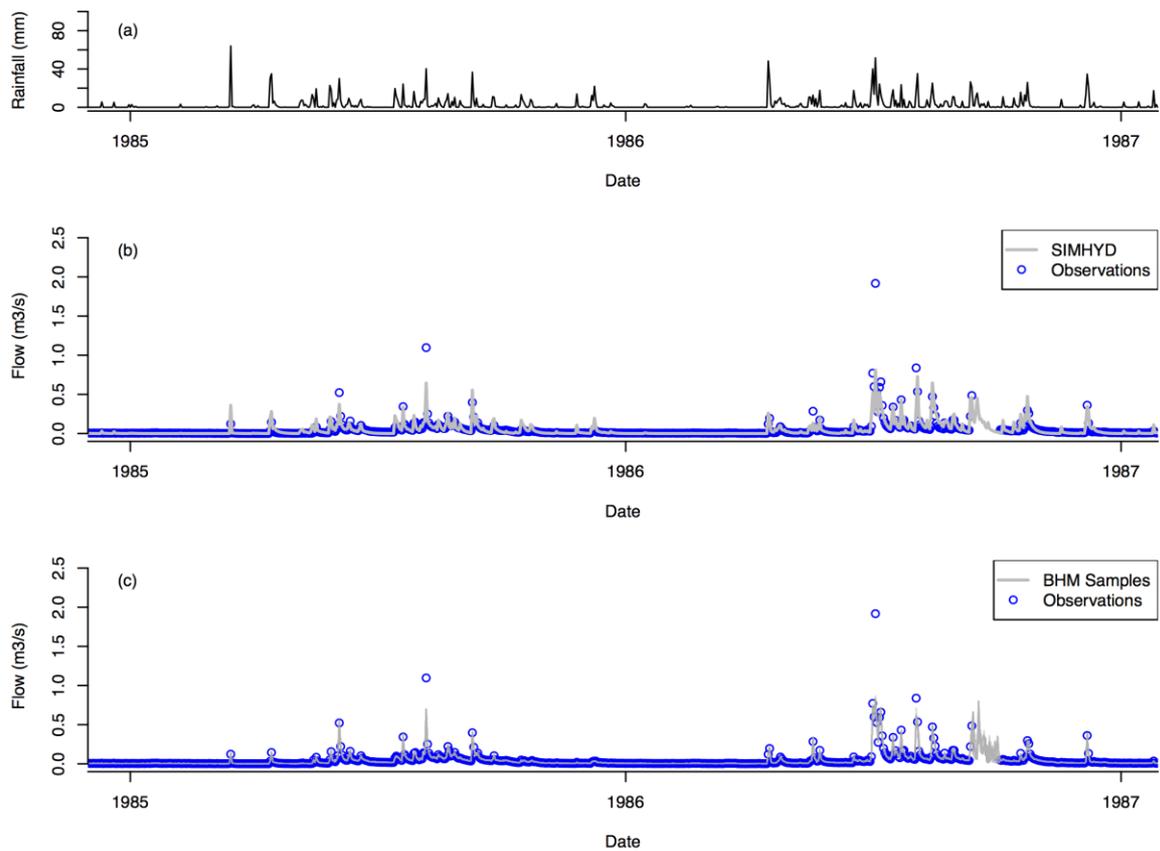


Figure 17: Rainfall-runoff model calibration results for Cox Creek at Uraidla (A5030526) between 1985 and 1987 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

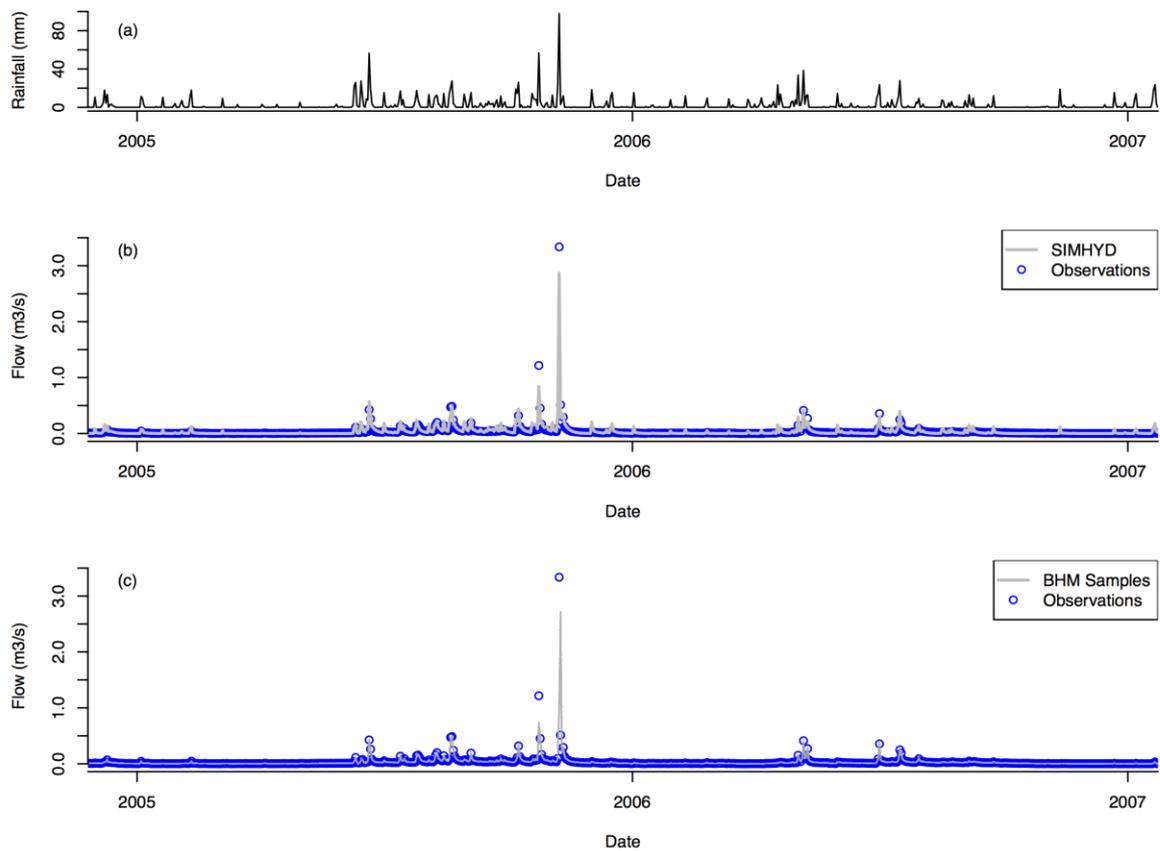


Figure 18: Rainfall-runoff model calibration results for Cox Creek at Uraidla (A5030526) between 2005 and 2007 showing: (a) rainfall input data; (b) SIMHYD Calibration using parameters equal to the mean of the posterior distribution; and (c) posterior samples from the BHM.

## Activity 2: Development of Site-Based Statistical Models

### Motivation

As outlined in the introduction, one of the activities undertaken for this report was to develop statistical models (site based models and spatio-temporal models) for sites monitored in the Onkaparinga catchment in the MLR watershed for the purpose of quantifying constituent loads with an associated estimate of uncertainty. It was hoped that the statistical relationships identified in this activity, might help identify important processes that determine the concentrations of total nitrogen (TN), total phosphorous (TP) and total suspended sediment (TSS). Processes and important factors that were identified in this activity could be used in future projects for the development of improved constituent generation models in Source. In the following sections, an overview of the statistical modelling approach used for identifying these important processes is presented and then the statistical models developed for each of the six water quality monitoring sites are presented and discussed.

### Overview of Statistical Modelling Framework

The statistical modelling framework considered for the site based statistical analysis makes use of Generalised Additive Models (GAM) and Generalised Additive Mixed Models (GAMM) (Wood, 2006) through the Loads Regression Estimator (LRE) package (for the R statistical computing environment), that was developed by Kuhnert et al. (2012) for modelling end of catchment sites in the Great Barrier Reef catchments (Bainbridge et al., 2014; Kroon et al., 2011; Kuhnert et al., 2012; Lewis et al., 2013; Robson and Dourdet, 2015). GAMs and GAMMs are essentially Generalized Linear Models (McCullagh and Nelder, 1989) where additional smooth nonparametric terms can also be additively included as predictors in the model. As outlined in Kuhnert et al. (2012), LRE is constructed from a four step process that consists of (1) estimation of stream flow using smoothing spline interpolation, (2) a predictive model for concentration (given the flow), (3) the estimation of the load and (4) the quantification of the errors in the load that incorporates errors in the flow rates provided in Table 2. LRE does require that flow is measured at regular intervals and as a result, step 1 of the approach checks for regularity in the flow data and constructs an interpolation spline (smooth relationship of flow through time) that can be used to infill the gaps in the flow record. While this results in a regularised flow record, it is important to note that this interpolation approach is not ideal for large gaps in flow records and other methods, such as that outlined in Pagendam et al. (2014) and using a more sophisticated deterministic rainfall-runoff model. For all of the sites under study, we were able to utilise the daily flow records provided.

The predictive model considered for each constituent is comprised of a GAM, which houses terms that either enter linearly into the model or are flexible, smooth spline functions driven by the data (Kuhnert et al., 2012). Covariates that were considered in the characterisation of each constituent consisted of flow (linear and quadratic terms) in addition to hydrological terms that mimicked certain features of the hydrology of each river or creek system. Features of many river systems that were considered in each analysis consisted of: a rising falling limb term that was incorporated as a categorical variable that captured flow appearing on the rise (+1), fall (-1) or the flat (0) of an event; discounted flow terms that considered lags in flow to take into account the impact that past flow events have on current patterns; and sinusoidal terms that factored in seasonal changes. As a substantial proportion of water quality samples occurred during baseflow conditions, we also

deconstructed the observed flow data into quick flow and baseflow components using a digital filter outlined in Nathan and McMahon (2010). Appendix A outlines the suite of covariates considered as part of the analyses conducted in this report. While every effort was made to include terms that had some potential hydrological relationship to the dynamics of constituents, it was sometimes not possible to explain all of the variability in the data. As such, a correlation term was considered when temporal dependence existed in the residuals of the model and no other terms could account for this variation. Inclusion of the correlation term had the effect of widening the confidence intervals of the loads estimates.

### Scott Creek (A5030502)

Constituent data in the form of TSS, TN and TP along with flow data for the Scott Creek catchment is shown in Figures 19-21. Flow data spanned the period from 1969 through to 2014, while the constituent data was collected between 1996 and 2014, with some constituents sampled less frequently than others. All measurements are presented on the log scale.

A GAM was fit to each of the three constituents with the results shown in Tables 10-12. The amount of variation explained by each model ranged between 41% and 50%. The primary terms fit across all models consisted of flow (baseflow and quickflow), discounted terms and a seasonal term. To accommodate the additional temporal dependencies in the data, an autoregressive structure was also fit in each model.

Results from fitting a GAMM to the TSS data collected for the Scott Creek catchment is shown in Table 10. This model explained approximately 50% of the variation. The linear terms included in the final model consisted of base and quick flow and indicated an increase of 0.198 (log-scale) in TSS when quickflow increased and a decrease in TSS when baseflow increased. A series of lagged flow or discounted flow terms were also included to represent the impact of historical flows on TSS generation. One of these was a smooth term relating to the accumulation of flow from the start of the time series where TSS was collected. These terms were significant and represented a complex hydrological representation of flow in Scott Creek that leads to generations of TSS represented in this catchment. Figure 22(a) shows the TSS and the accumulation of flow as it increases. A correlation term (Autoregressive) was also fit in this model to account for temporal dependencies that could not be accounted for by the terms in the model. Diagnostic plots are shown in Appendix B and indicate a reasonable fit given the data.

Like TSS, a GAMM was fit to total nitrogen data collected in the Scott Creek catchment, which explains approximately 49% of the variation in the data. Hydrological terms (flow comprised of baseflow and quick flow; discounted flow terms and accumulation of flow) were mainly fit in this model in addition to a seasonal term (month). An autoregressive term was also included to account for additional temporal dependence in the model. Diagnostic plots examining the fit of the model are shown in Appendix B. Figures 22(b – d) summarise the smooth terms represented in the model, namely the seasonal term, the accumulation of flow term (csQ) and the discounted flow term (d0.75). The seasonal term indicates an increase in TN from March through to May, which decreases for months beyond May. The accumulation of flow in Figure 22(c) indicates a constant generation of TN until large flows are accumulated over time, which results in an increase in TN. There also appears to be an increase in TN as the build up of past flows increases. Figure 22(e) shows the seasonal term fit in the model for TP and like TN, indicates increases in TP between March-May with a decline in TP concentration for months beyond May. The discounting terms indicate some complex

hydrological relationships occurring depending on the lag. More recent flows (d0.25) appear to be associated with a decrease in TP, while past flows indicate an increase (Table 12).

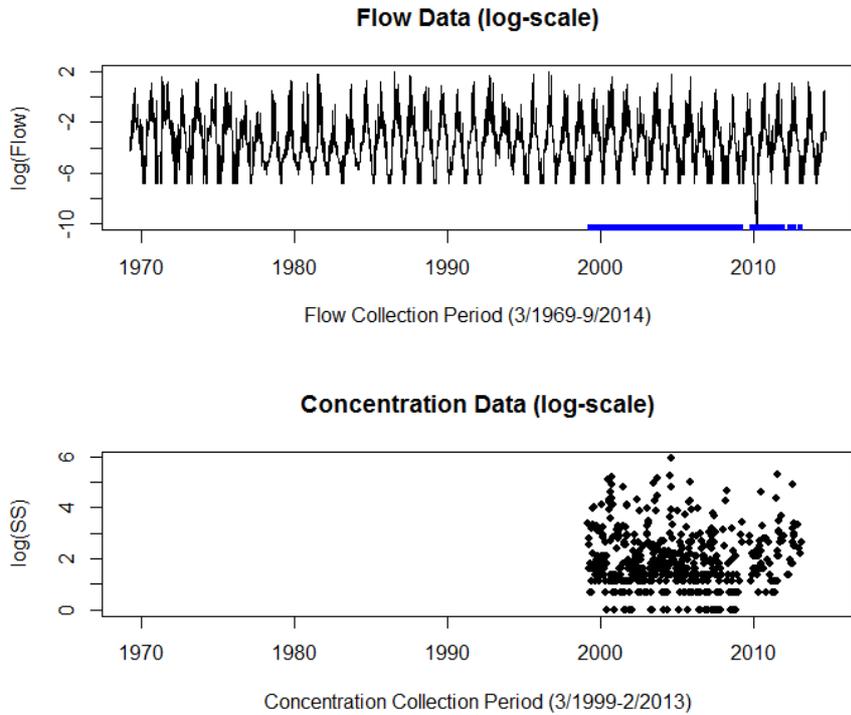


Figure 19: Raw flow and concentration data presented on the log-scale for TSS at the Scott Creek site (A5030526).

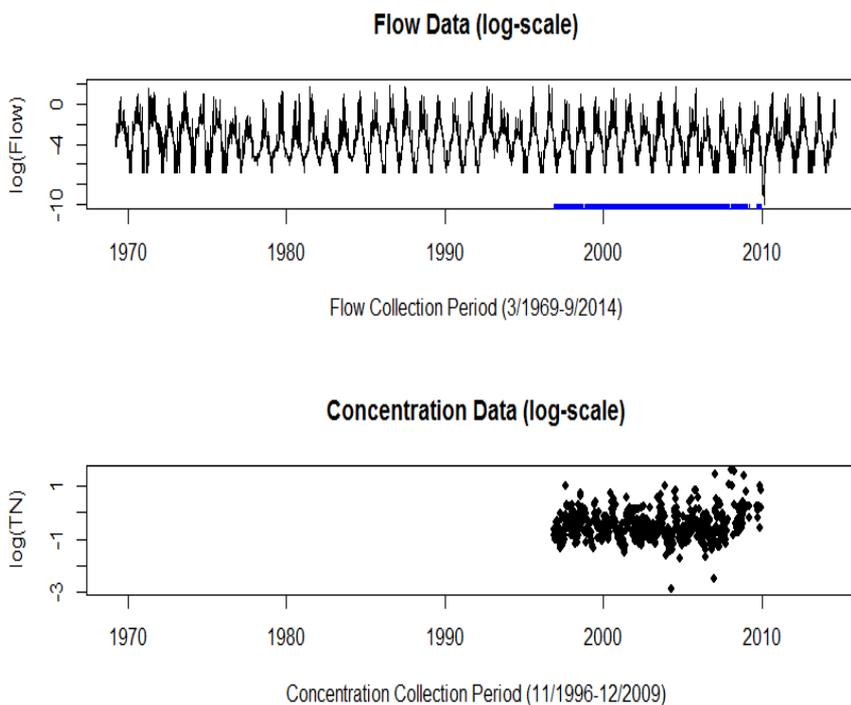


Figure 20: Raw flow and concentration data presented on the log-scale for TN at the Scott Creek site (A5030526).

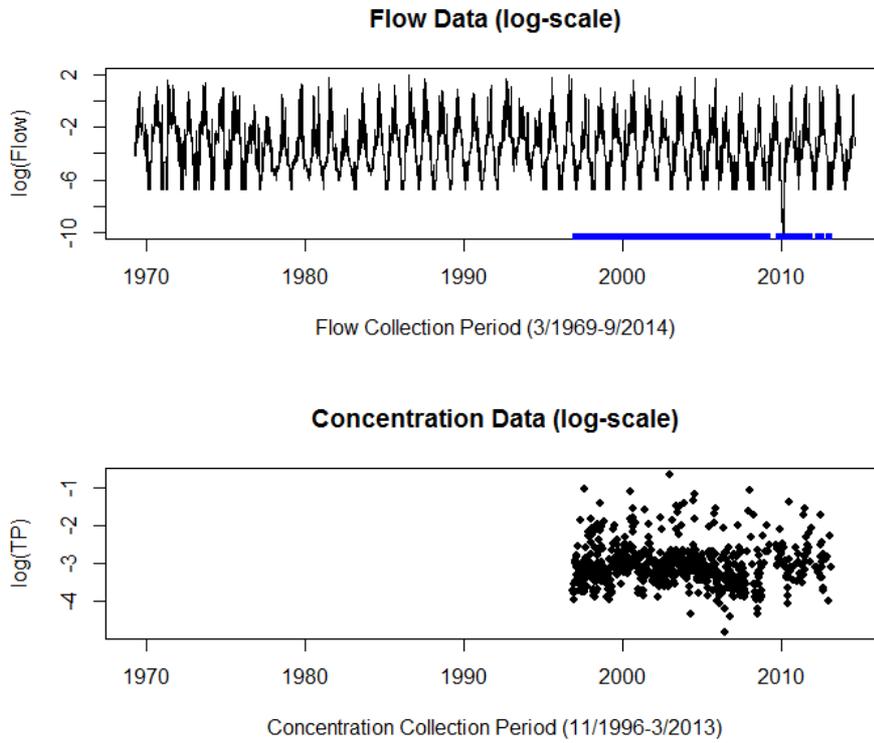


Figure 21: Raw flow and concentration data presented on the log-scale for TP at the Scott Creek site (A5030526).

Table 10: Summary of parameter estimates from generalised additive model fit to TSS from Scott Creek (A5030526) that explains 50.1% of the variation in the data.

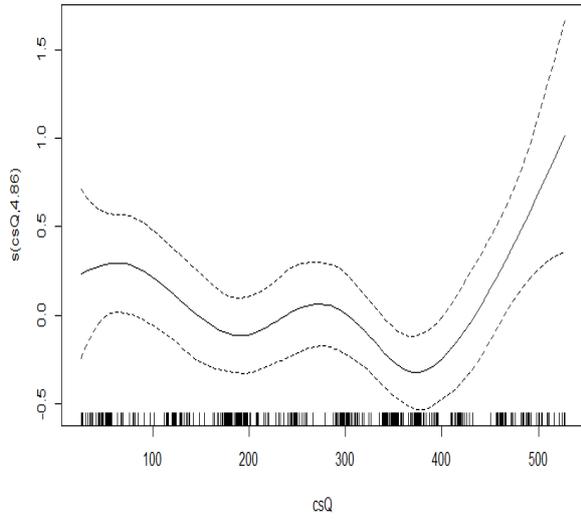
Parameter	Estimate	SE	p-value
Intercept	3.060	1.21	<0.001
Flow			
- log(Baseflow)	-0.225	0.08	0.003
- log(Quickflow)	0.198	0.02	<0.001
Discounted Flow Terms			
- 0.25	-1.871	0.37	<0.001
- 0.50	3.086	0.54	<0.001
- 0.75	-1.058	0.24	<0.001
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(cumulative sum flow)	4.863	3.639	0.003
<b>Correlation Term – AR(1)</b>	<b>Estimate</b>	<b>95% CI</b>	
$\phi$	0.383	[0.30,0.46]	

Table 11: Summary of parameter estimates from generalised additive model fit to TN from Scott Creek (A5030526) that explains 48.7% of the variation in the data.

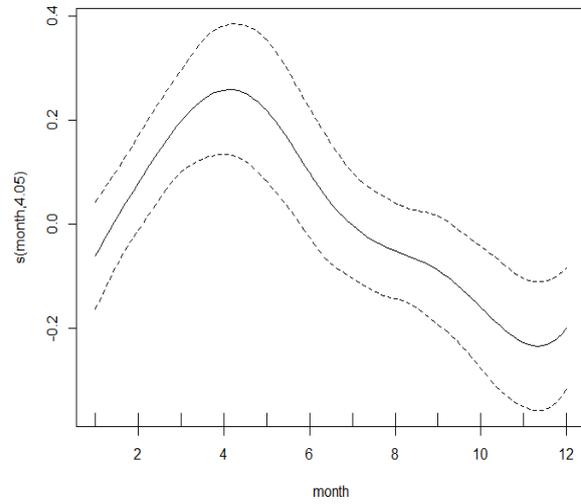
Parameter	Estimate	SE	p-value
Intercept	-1.020	0.26	<0.001
Flow			
- log(Baseflow)	0.006	0.03	0.859
- log(Quickflow)	0.060	0.01	<0.001
Discounted Flow Terms			
- 0.50	-0.242	0.08	0.002
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	4.048	3.094	<0.001
s(cumulative sum flow)	5.001	9.554	<0.001
Discounted Flow Terms			
- 0.75	5.001	18.462	<0.001
<b>Correlation Term – AR(1)</b>	<b>Estimate</b>	<b>95% CIs</b>	
$\phi$	0.319	[0.24,0.40]	

Table 12: Summary of parameter estimates from generalised additive model fit to TP from Scott Creek (A5030526) that explains 41.1% of the variation in the data.

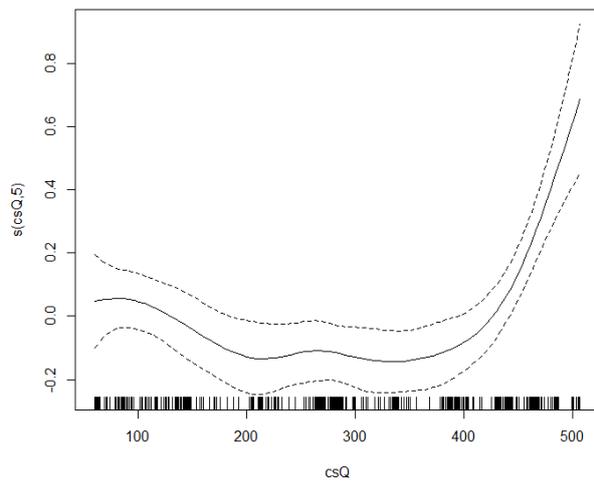
Parameter	Estimate	SE	p-value
Intercept	-2.038	0.097	<0.001
Flow			
- log(Baseflow)	-0.045	0.03	0.180
- log(Quickflow)	0.083	0.01	<0.001
Discounted Flow Terms			
- 0.25	-0.356	0.11	0.002
- 0.50	0.519	0.11	<0.001
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	4.101	3.843	<0.001
<b>Correlation Term – AR(1)</b>	<b>Estimate</b>	<b>95% CI</b>	
$\phi$	0.417	[0.34,0.48]	



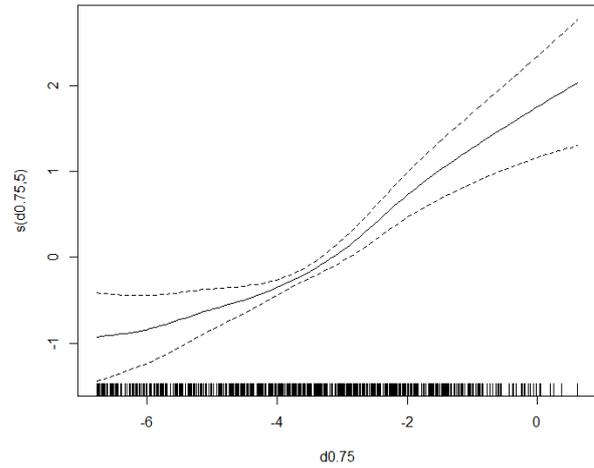
(a)



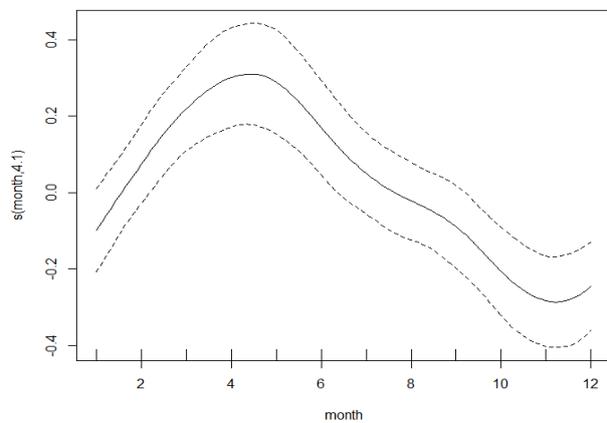
(b)



(c)



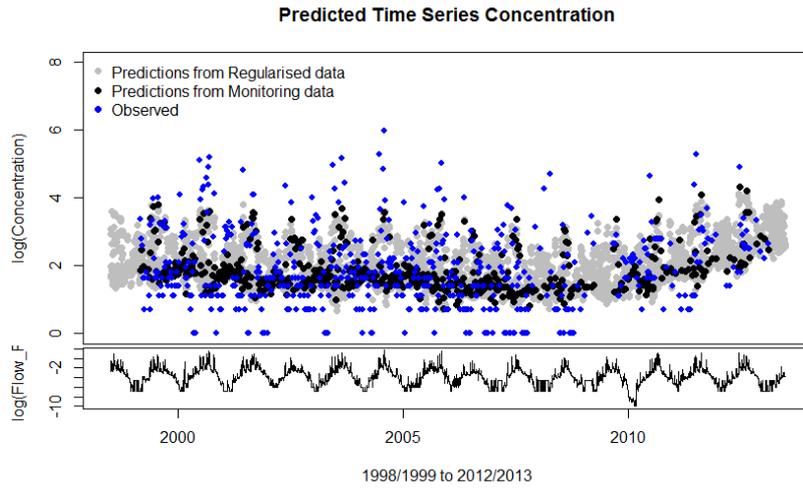
(d)



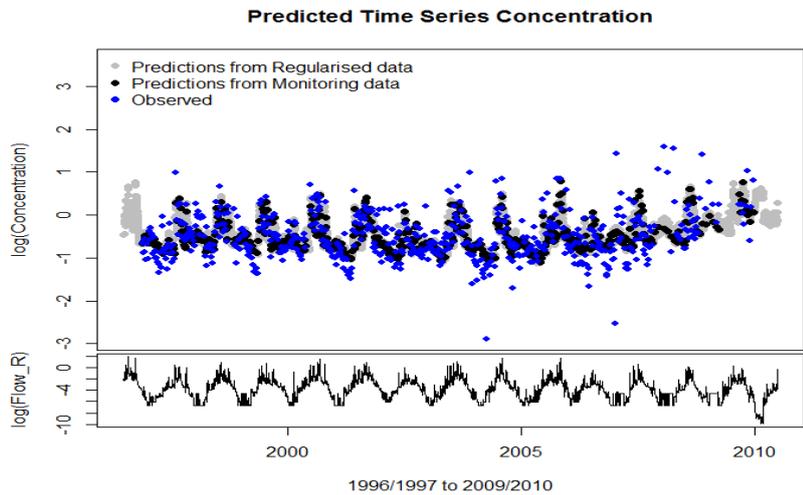
(e)

Figure 22: Smooth terms from the generalised additive models fit for Scott Creek (A5030502) for each of the three constituents. TSS: (a) past sum of flow; TN: (b) seasonal term, (c) past sum of flow (d) discounted flow ( $d=0.75$ ); TP: (e) seasonal term.

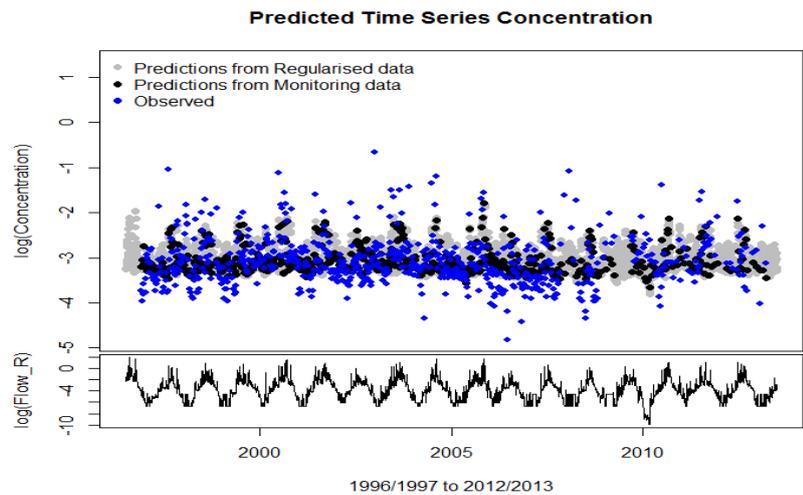
The predicted time series for each constituent is shown in Figure 23 and shows the predicted constituent values from each model (black) with the observed (blue) measurements overlaid. We also include the predicted constituent based on the entire flow record, i.e. where flow was represented at regular daily intervals. The plot below this figure shows the flow (log-scale). In Figure 23(a), the GAMM appears to characterise the pattern of TSS through time, although it does have some difficulty predicting at the extremes. Predictions of TN and TP shown in Figure 23(b) and 23(c) also show a reasonably good match with the data apart from some extreme points. The difficulty in predicting extremes could be due to the sample design used to monitor each constituent in addition to the ability to characterise those extremes with the covariates available.



(a)



(b)



(c)

Figure 23: Predictions from the generalised additive model fit to constituent data at Scott Creek (A5030502) for (a) TSS, (b) TN and (c) TP.

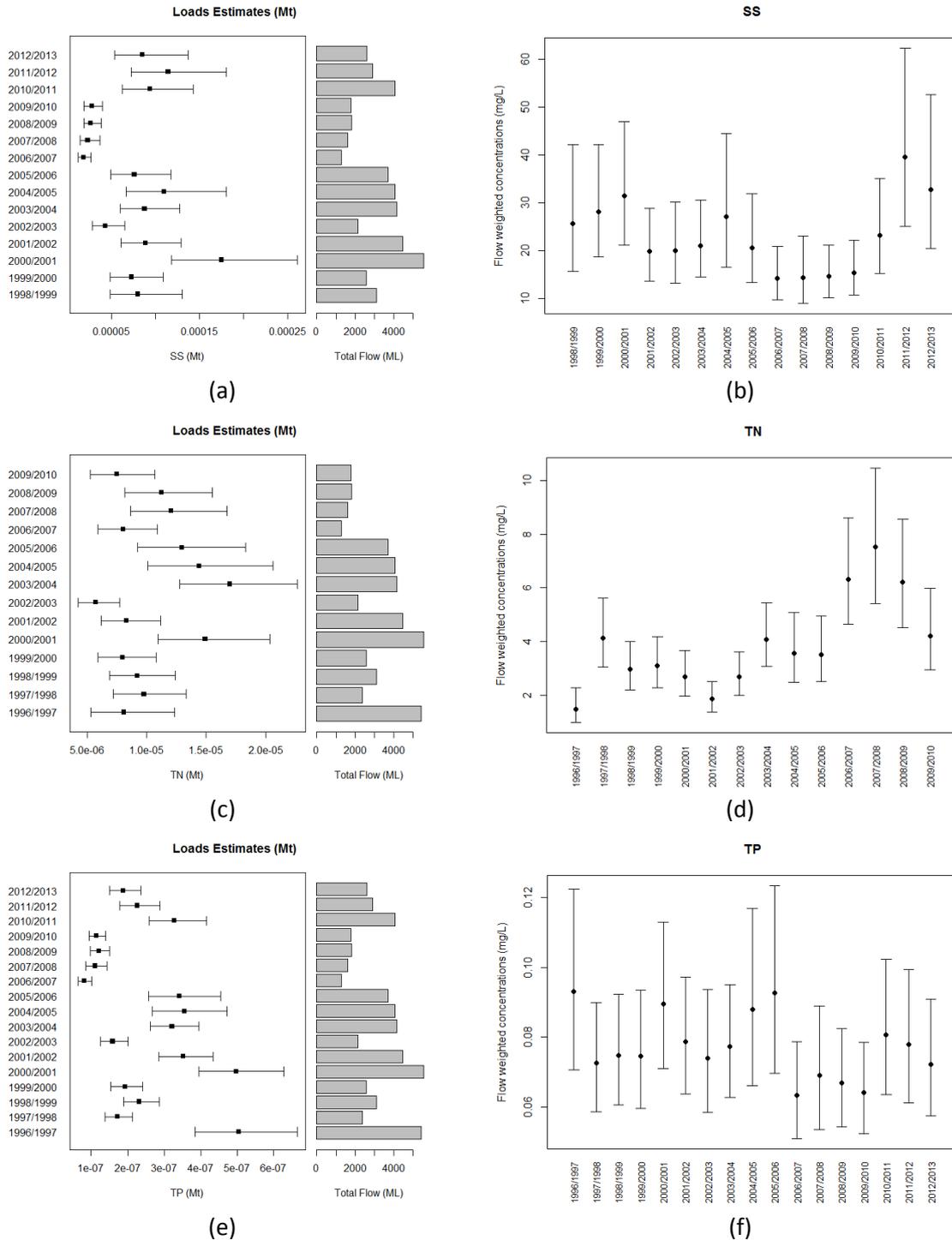


Figure 24: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for Scott Creek (A5030502) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP.

Load estimates produced for each constituent are shown in Figure 24 and are accompanied by 80% confidence intervals to convey the variability in the estimates. These are estimated for each financial year and are accompanied by the total volume of flow across the years sampled. We also present the flow weighted concentrations to showcase the patterns in flow across years. The flow weighted TSS concentrations in Figure 24(b) show some variation in the loads estimates. Fluctuations in TSS appear to occur across years but it is difficult to make any conclusive statements regarding trends due to the wide confidence intervals. The flow weighted TN concentrations are shown in Figure 24(d) and like TSS, show some variability in the flow estimates. This figure does show some increases in TSS in the later years, although the uncertainty around these estimates widens. Figure 24(f) exhibits substantial variation in the estimates and a corresponding lack of confidence in predicting TP through time. This may be a combination of the sample design and the ability of the model to characterise the complex TP relationships.

## Onkaparinga River at Houlgraves (A5030504)

Constituent data and flow were obtained for the Onkaparinga River at Houlgraves and are presented in Figure 25 - 27 (log scale). Constituent data were collected between 1996 and 2013 depending on the constituent measured. Flow measured at this site was also captured but required adjustment for inflows from a Murray River diversion entering the system at the Hahndorf Dissipator. The low flows exhibited in 2007 and 2008 correspond to a period of drought.

A GAM was used to predict TSS and identify important predictors. The model which explains approximately 50% of the variation is summarised in Table 13. Apart from flow, which was partitioned into baseflow and quickflow, a smooth seasonal term was fit in the model along with the accumulation of flow (csQ) and discounting terms that captured a range of flows at short and long lags. The smooth terms fit in the GAM are shown in Figure 28. The seasonal term (Figure 28(a)) identifies increases in TSS from February through to May before a decline is noted. The accumulation of flow through time (Figure 28(b)) shows increases and declines throughout the sampling period that may reflect long-term trends. The discounting terms that comprise the remainder of the plots in Figure 28 show varying patterns in the data, either increases or decreases in TSS depending on how the past flows were represented. Figure 31(a) displays the predicted TSS concentrations through time with the corresponding flow (log-scale) at the base of the figure. The predictions are represented by black points, while the grey represents predictions made on a regularised time series of flow. Observations are overlaid in blue and show that the predictions through time are reasonable, despite some extreme points. The predicted annual loads and flow weighted concentrations are shown in Figure 32(a-b). Both figures show a variable pattern in TSS through time which are accompanied by wide 80% credible intervals for some years.

Data collected for total nitrogen at the Houlgraves site were fairly limited, with only data captured between 2004 and 2009 as shown in Figure 26. Although quite limited in terms of sampling, a GAM fit to the data explained approximately 74% of the variation using flow and variants of flow (accumulation of flow and discounting terms) in addition to a seasonal term. Figure 29 summarises the smooth terms represented in the model. The seasonal term shows increases in total nitrogen in the later months of the year (August-November). The discounted flow terms show increases in TN that appear to occur at long lags compared to shorter lags investigated.

Concentration predictions are shown in Figure 31(b) and illustrate a close match with the data. Loads estimation methods produced TN loads that are shown in Figure 32(c). The flow-weighted concentrations that are shown in Figure 32(d) show some variation in loads and highlight a decrease in 2006/2007.

Samples of total phosphorous were collected between 1996 and 2013 for the Onkaparinga catchment at Houlgraves as shown in Figure 27. A GAMM was fit to the data to accommodate temporal dependencies that could not be accounted for by other terms in the model. Terms fit in this model explained nearly 58% of the variation and consisted of hydrological terms and a seasonal term. Results are displayed in Table 15 and highlight that quick flow is an important predictor, suggesting that as quickflow (or runoff) increases, we are likely to see an increase in total phosphorous. All smooth terms used in the model are shown in Figure 30. The seasonal term shows a steady increase in TP through the months before declining.

The accumulation of flow (csQ) highlights a reduction in the generation of TP during large build ups of flow, particularly later in the data record. The discounted flow also highlights a relationship

between flow and TP, but at different lags than were tested in the model. The predictions from the model are shown in Figure 31(c) and show a reasonable match with the original data that was used to develop the model. Loads estimates are shown in Figure 32(e) and show considerable variation in the data from year to year. The plot of the flow-weighted concentrations (Figure 32(f)) illustrates some years where there were declines.

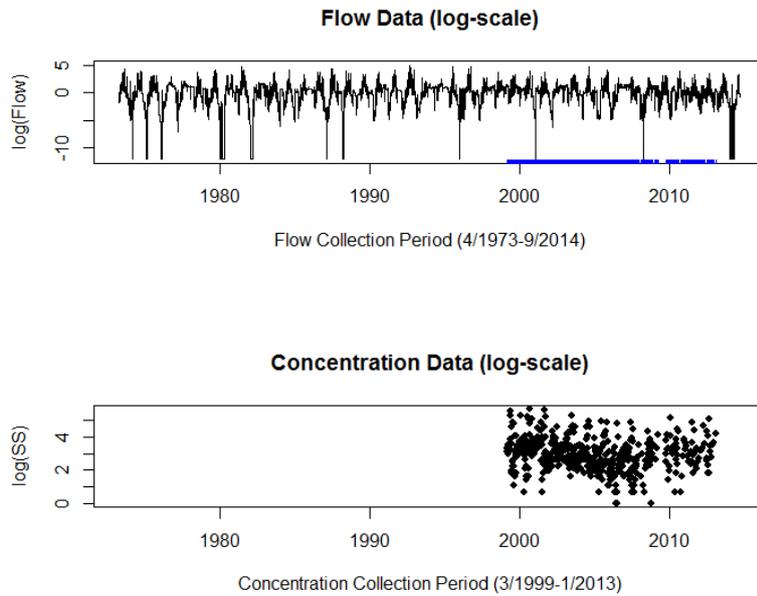


Figure 25: Raw flow and concentration data presented on the log-scale for TSS at the Onkaparinga River, Houlgraves site (A5030504).

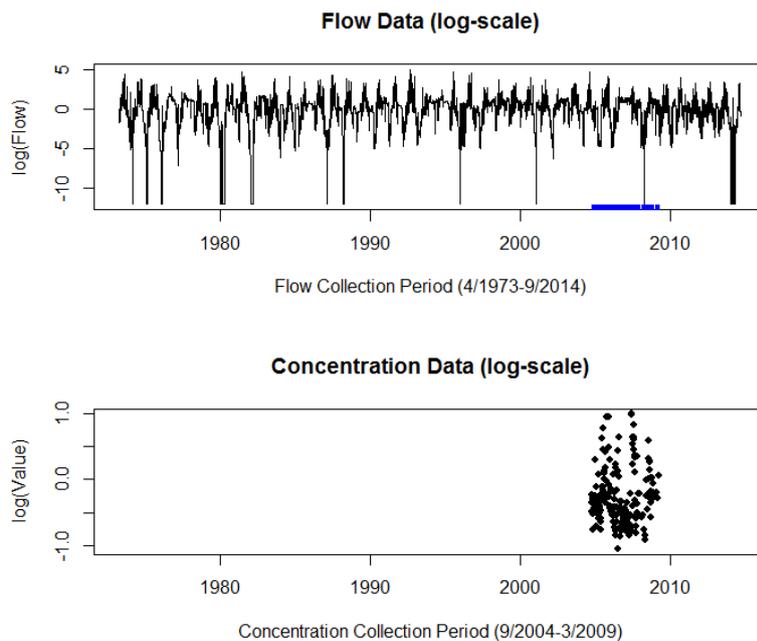


Figure 26: Raw flow and concentration data presented on the log-scale for TN at the Onkaparinga River, Houlgraves site (A5030504).

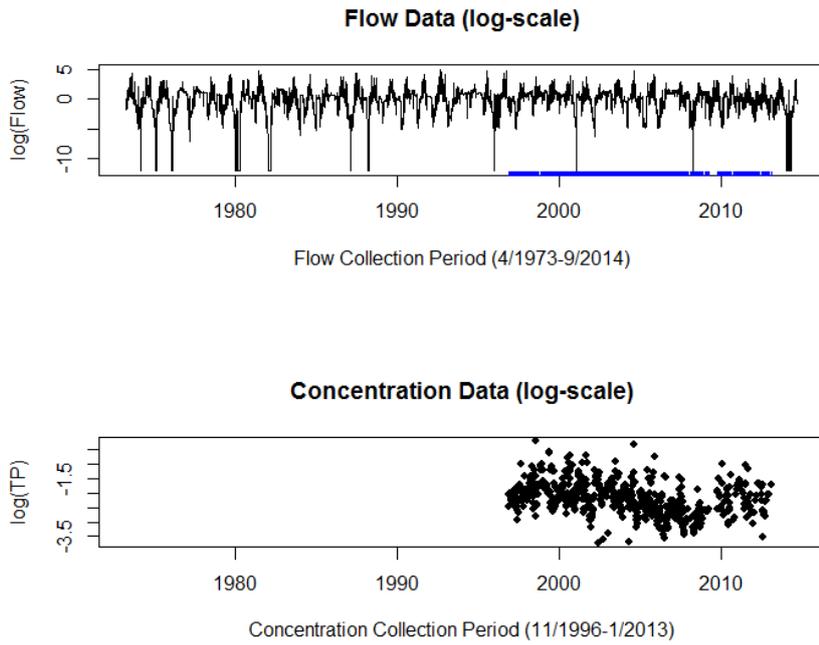


Figure 27: Raw flow and concentration data presented on the log-scale for TP at the Onkaparinga River, Houlgraves site (A5030504).

Table 13: Summary of parameter estimates from generalised additive model fit to TSS from Onkaparinga River at Houlgraves (A5030504) that explains 50.3% of the variation in the data.

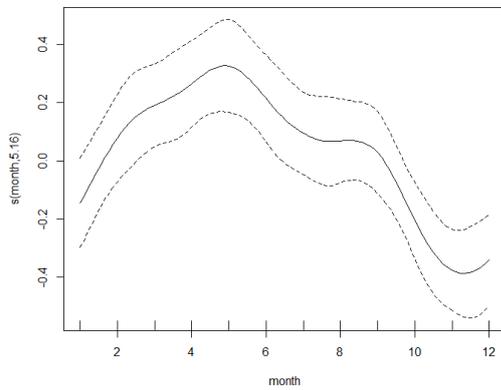
Parameter	Estimate	SE	p-value
Intercept	3.307	0.07	<0.001
Flow			
log(Baseflow)	0.075	0.07	0.267
log(Quickflow)	0.091	0.02	<0.001
Smooth Terms	EDF	F-statistic	p-value
s(month)	5.161	5.124	<0.001
csQ	7.622	14.092	<0.001
Discounted Flow terms			
- 0.95	2.665	3.170	0.02
- 0.50	2.653	15.374	<0.001
- 0.25	5.001	5.507	<0.001

Table 14: Summary of parameter estimates from generalised additive model fit to TN from Onkaparinga River at Houlgraves (A5030504) that explains 73.5% of the variation in the data.

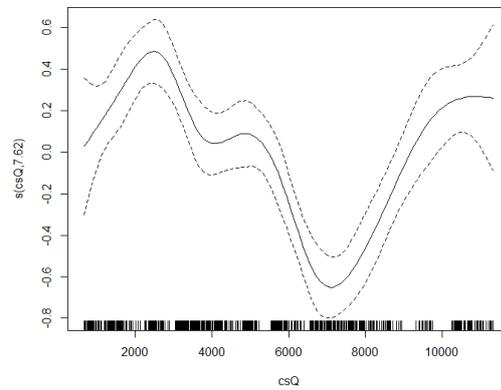
Parameter	Estimate	SE	p-value
Intercept	-0.185	0.05	0.0001
Flow			
- log(Baseflow)	0.038	0.05	0.428
- log(Quickflow)	0.015	0.01	0.0215
Smooth Terms	EDF	F-statistic	p-value
s(month)	6.978	7.979	<0.001
s(cumulative sum flow)	8.027	4.555	<0.001
Discounted Flow Terms			
- 0.10	8.621	3.659	0.0004
- 0.50	8.691	4.211	<0.001
- 0.75	2.929	6.286	0.0002

Table 15: Summary of parameter estimates from generalised additive model fit to TP from Onkaparinga River at Houlgraves (A5030504) that explains 57.4% of the variation in the data.

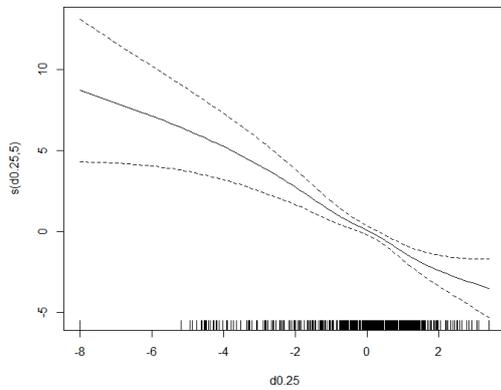
Parameter	Estimate	SE	p-value
Intercept	-2.073	0.04	<0.001
Flow			
- log(Baseflow)	0.029	0.03	0.351
- log(Quickflow)	0.025	0.01	<0.001
Smooth Terms	EDF	F-statistic	p-value
s(month)	4.403	5.635	<0.001
s(cumulative sum of flow)	6.784	18.726	<0.001
Discounted Flow terms			
- d0.5	3.187	23.816	<0.001
- d0.1	4.964	6.553	<0.001
Correlation Term – AR(1)	Estimate	95% CI	
$\phi$	0.436	[0.36,0.51]	



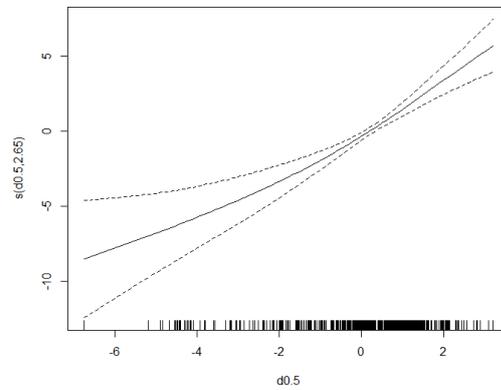
(a)



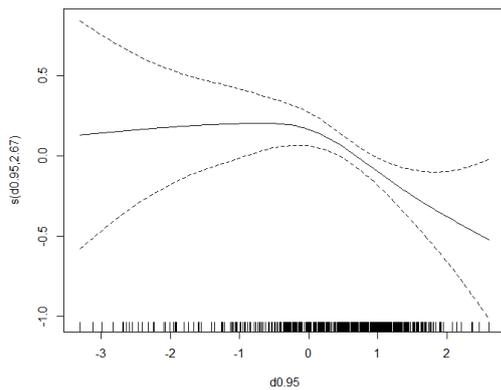
(b)



(c)

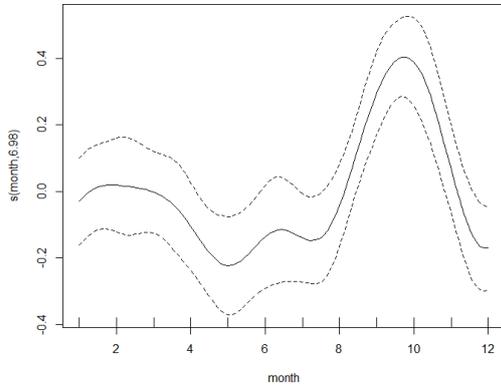


(d)

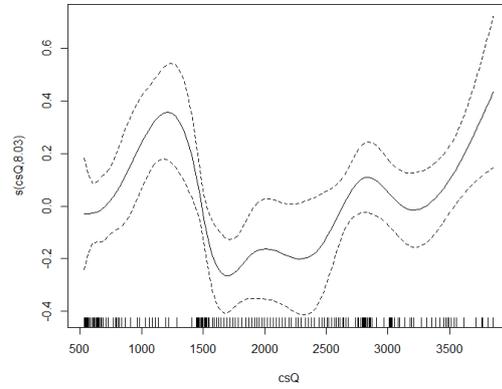


(e)

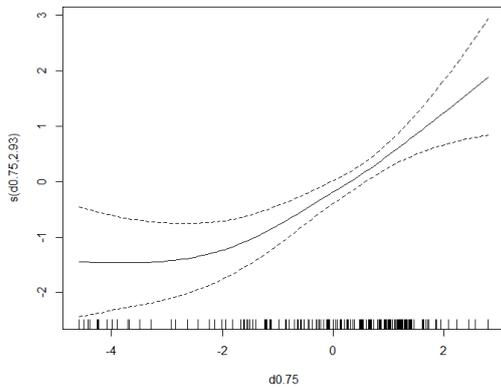
Figure 28: Smooth terms from the generalised additive model for Onkaparinga River at Houlgraves (A5030504) showing (a) the characteristics of season, (b) historical flow, (c) discounting term ( $d0.25$ ), (d) discounting term ( $d0.5$ ) and discounting term ( $d0.95$ ) in relation to TSS.



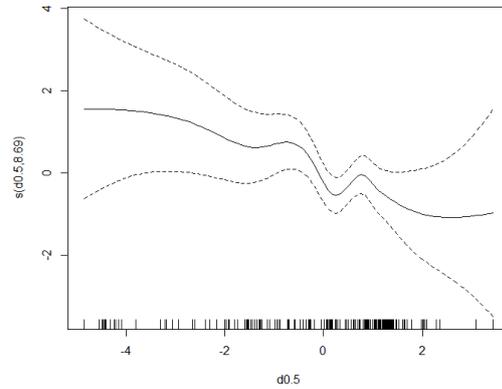
(a)



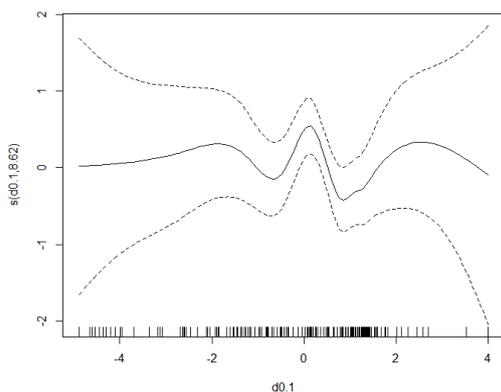
(b)



(c)

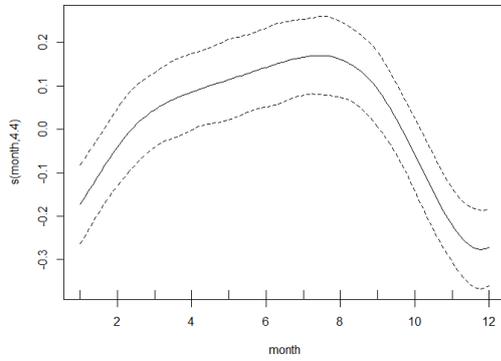


(d)

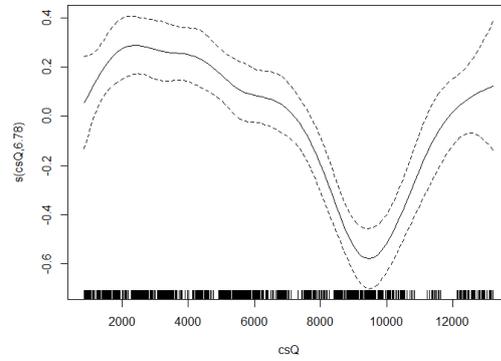


(e)

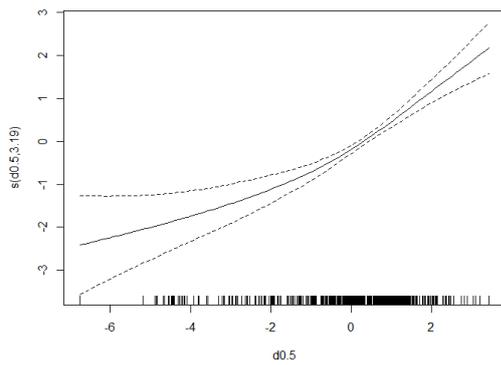
Figure 29: Smooth terms from the generalised additive model for Onkaparinga River at Houlgraves (A5030504) showing the characteristics of (a) season, and (b) the discounted flow ( $d0.75$ ) in relation to TN.



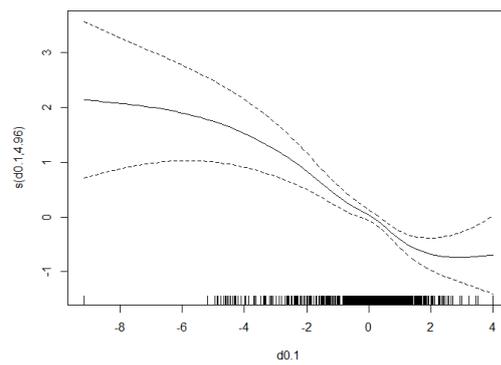
(a)



(b)

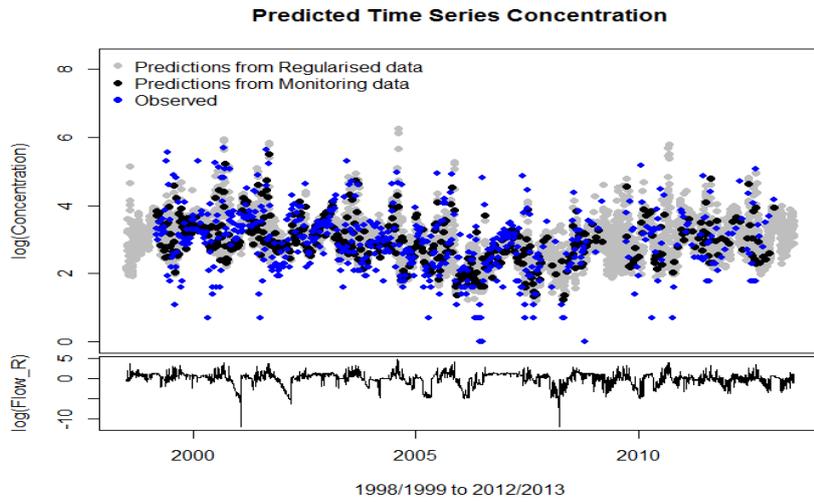


(c)

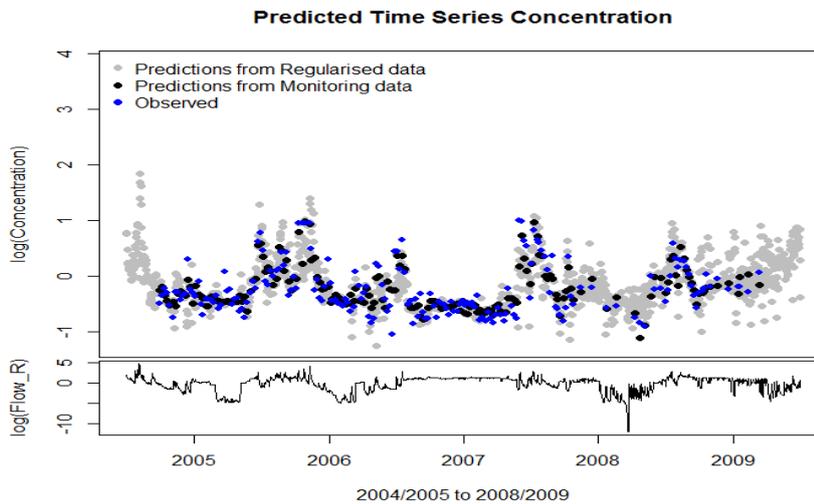


(d)

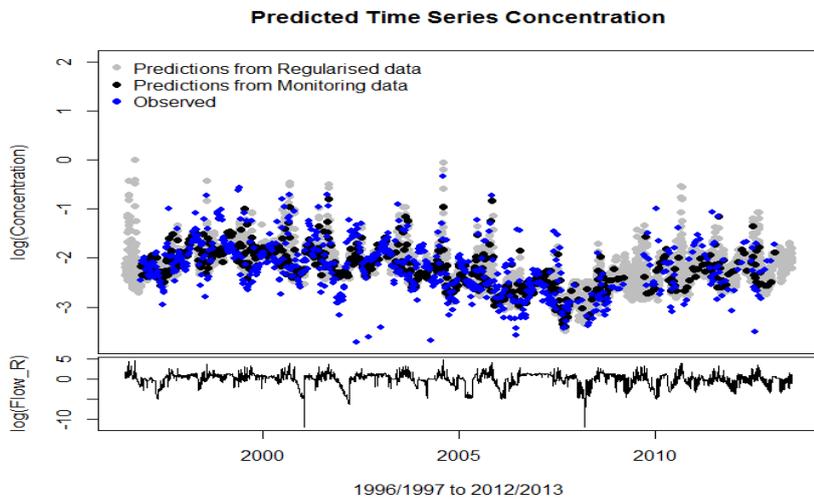
Figure 30: Smooth terms from the generalised additive model for Onkaparinga River at Houlgraves (A5030504) showing the characteristics of (a) season, (b) the past sum of flow, (c) discounted flow (d0.5) and (d) discounted flow (d0.1) in relation to TP.



(a)



(b)



(c)

Figure 31: Predictions from the generalised additive model fit to constituent data at Onkaparinga River, Houlgraves Site (A5030504) for (a) TSS, (b) TN and (c) TP.

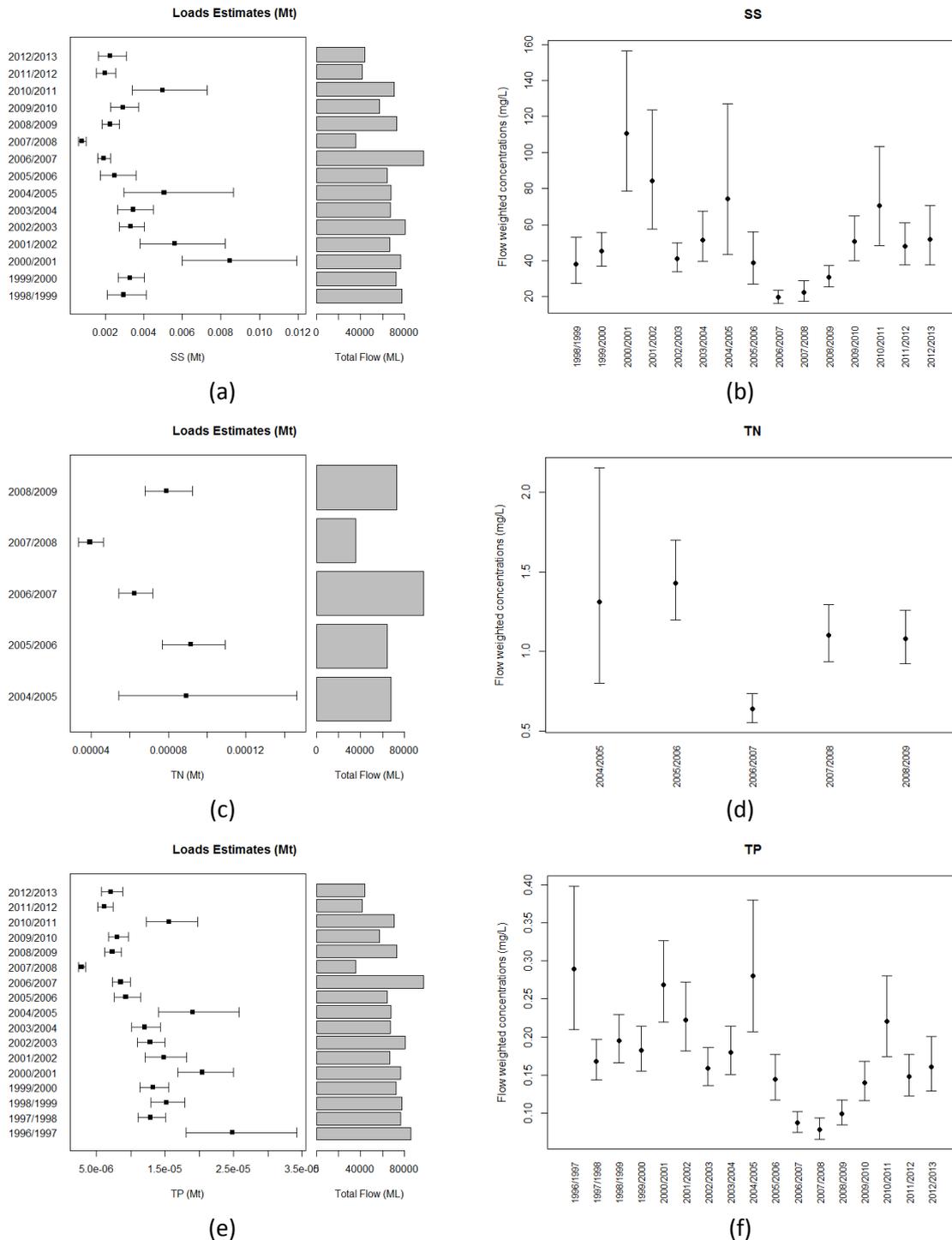


Figure 32: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for Onkaparinga River, Houlgraves Site (A5030504) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP.

## Echunga Creek (A5030506)

Constituent measurements, namely TSS, TN and TP along with daily flow were obtained for the Echunga Creek site and are shown in Figures 33-35.

TSS data was collected between 1999 and 2011. A GAMM was fit to the TSS data for Echunga Creek. Important terms fit in the model are shown in Table 16 and include hydrological terms such as flow, discounted flow and the accumulation of flow. A seasonal term was also fitted and shows higher TSS predicted for February through to May before tapering off (Figure 36(a)). The accumulation term (csQ) shows a slight increase in TSS as flow accumulates towards 200 cumecs but then decreases beyond this amount (Figure 36(b)). The discounting terms indicate a complex hydrological process through which the TSS measurements are influenced by the volume of past discharges (see Table 16 and Figure 36(c)). An autocorrelation term was also included in this model to accommodate additional temporal dependency that could not be captured by the hydrological terms in the model. This model explained about 45% of the variation in the data. Predictions from the model are shown in Figure 38(a) (black points) with observed values shown in blue and predictions made on regularised flow data shown in grey. Apart from some extreme values at the low end, the predictions from the model appear to mimic the data. Estimates of the loads are shown in Figure 39(a) and show some variability in the estimates across years, particularly early on.

Total nitrogen was measured at the Echunga Creek site between 2004 and 2009 (Figure 34). A GAM was fit to the total nitrogen data measured at the Echunga Creek site. Results are displayed in Table 17 and incorporate hydrological terms and a seasonal effect that explains approximately 63 percent of the variation in the data. Figures 36(d-f) summarises the smooth terms represented in the model in Table 17. The seasonal term shows increases in TN between February and May. The hydrological terms that are comprised of an accumulation of flow and discounting past flows indicate some increases in TN as flows increase. Predictions of TN are shown in Figure 38(b) and show how the model characterises the patterns in TN through time at this site. Load estimates for the 5 years over which measurements were taken are shown in Figures 39(c)-39(d). Increases in TN are observed through time, although the confidence intervals are reasonably wide and overlap.

Total phosphorous measured at the Echunga Creek site is shown in Figure 35 along with daily flow. A GAMM was fit to TP that included flow terms (quick and base flow) along with a seasonal term and hydrological terms included discounted flow terms and an accumulation of flow through time (Table 18). An AR1 term was included in the model to account for additional temporal correlation that could not be removed by fitting covariates in the model. This model resulted in approximately 64% of the variation being explained. The smooth terms fit in the model are shown in Figure 37 and shows the relationship between TP and each of the terms fit in the model. Summer and autumn periods showed an increase in TP compared to spring and winter periods (Figure 37(a)). The hydrological terms shown in Figures 37(b-d) indicate some complex flow relationships with TP. Predicted TP concentrations are shown in Figure 38 which shows a time series of TP predicted between 1996 and 2011. Predictions are shown in black while observations are shown in blue. Predictions of TP based on flow at regular time intervals (e.g. daily) are shown in grey. The hydrograph (log-scale) for the period is shown beneath the plot. Predicted TP appears to mimic the original data and seasonal pattern highlighted by the flow and seasonal term in the model.

Load estimates for TP are shown in Figure 39(e) and show some variability between estimates for different years. After adjusting for flow and producing the flow weighted TP concentrations shown

in Figure 39(f) illustrates a variable set of predictions that fluctuate from year to year. It is difficult to conclude from this figure whether there are any trends in the data due to the overlapping intervals.

Table 16: Summary of parameter estimates from generalised additive model fit to SS from Echunga Creek (A5030506) that explains 45.5% of the variation in the data.

Parameter	Estimate	SE	p-value
Intercept	-0.909	1.02	0.3755
Flow			
- log(Baseflow)	0.184	0.10	0.058
- log(Quickflow)	0.134	0.03	<0.002
Discounted Flow terms			
- 0.25	-0.988	0.24	<0.001
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	3.691	3.978	<0.001
s(cumulative flow)	3.023	2.821	0.038
Discounted Flow terms			
- 0.50	3.823	16.302	<0.001
<b>Correlation Term – AR(1)</b>	<b>Estimate</b>	<b>95%CI</b>	
$\phi$	0.405	[0.31,0.49]	

Table 17: Summary of parameter estimates from generalised additive model fit to TN from Echunga Creek (A5030506) that explains 62.7% of the variation in the data.

Parameter	Estimate	SE	p-value
Intercept	-0.660	0.19	<0.001
Flow			
- log(Quickflow)	-0.020	0.02	0.327
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	6.713	3.131	0.0005
s(cumulative flow)	5.294	3.920	0.001
Discounted Flow terms			
- 0.50	8.800	10.218	<0.001

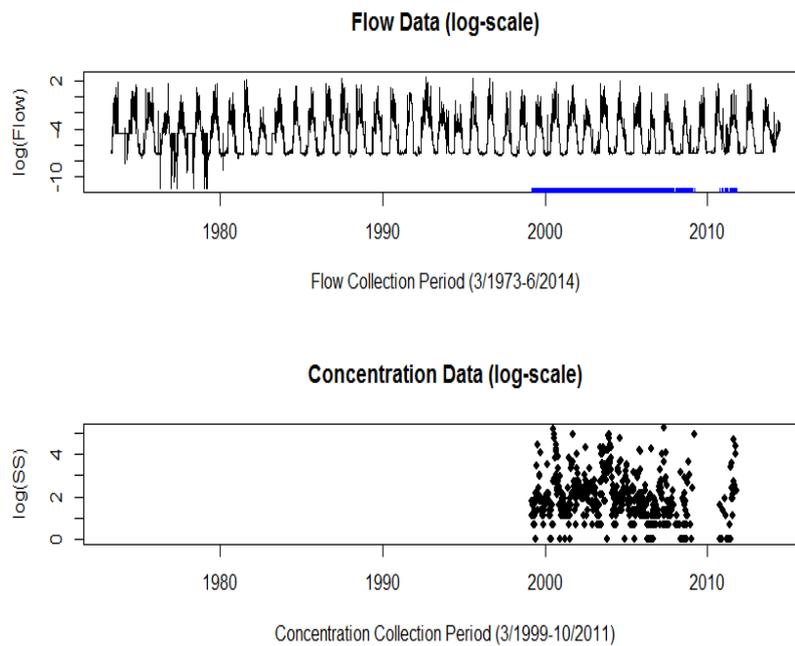


Figure 33: Raw flow and concentration data presented on the log-scale for TSS at the Echunga Creek site (A5030506).

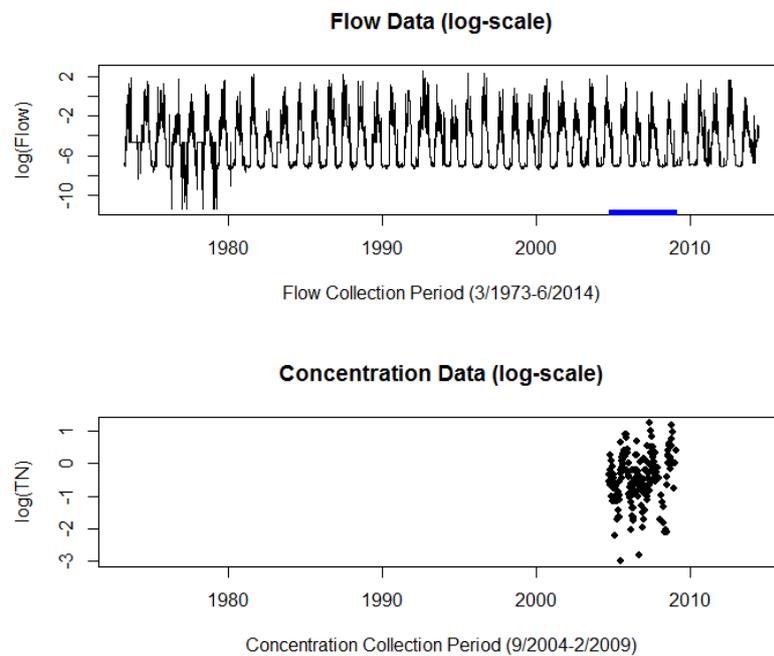


Figure 34: Raw flow and concentration data presented on the log-scale for TN at the Echunga Creek site (A5030506).

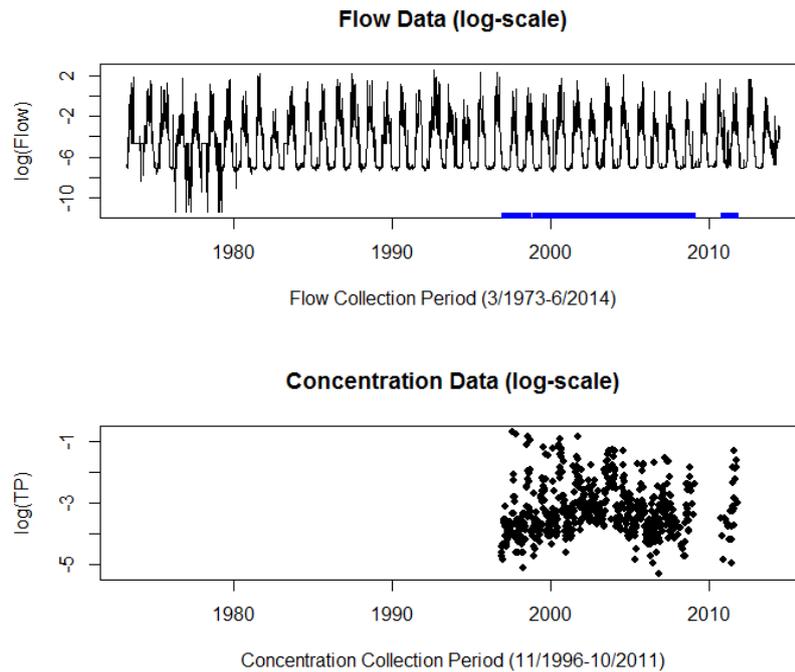


Figure 35: Raw flow and concentration data presented on the log-scale for TP at the Echunga Creek site (A5030506).

Table 18: Summary of parameter estimates from generalised additive model fit to TP from Echunga Creek (A5030506) that explains 64.4% of the variation in the data.

Parameter	Estimate	SE	p-value
Intercept	-2.640	0.214	<0.001
Flow			
- log(Baseflow)	0.048	0.05	0.297
- log(Quickflow)	0.034	0.01	<0.001
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	3.465	3.254	<0.001
s(cumulative flow)	3.410	5.163	0.001
Discounted Flow terms			
- 0.75	5.515	4.919	0.006
- 0.95	2.225	32.272	<0.001
<b>Correlation Term – AR(1)</b>	<b>Estimate</b>	<b>95%CI</b>	
$\phi$	0.463	[0.38,0.54]	

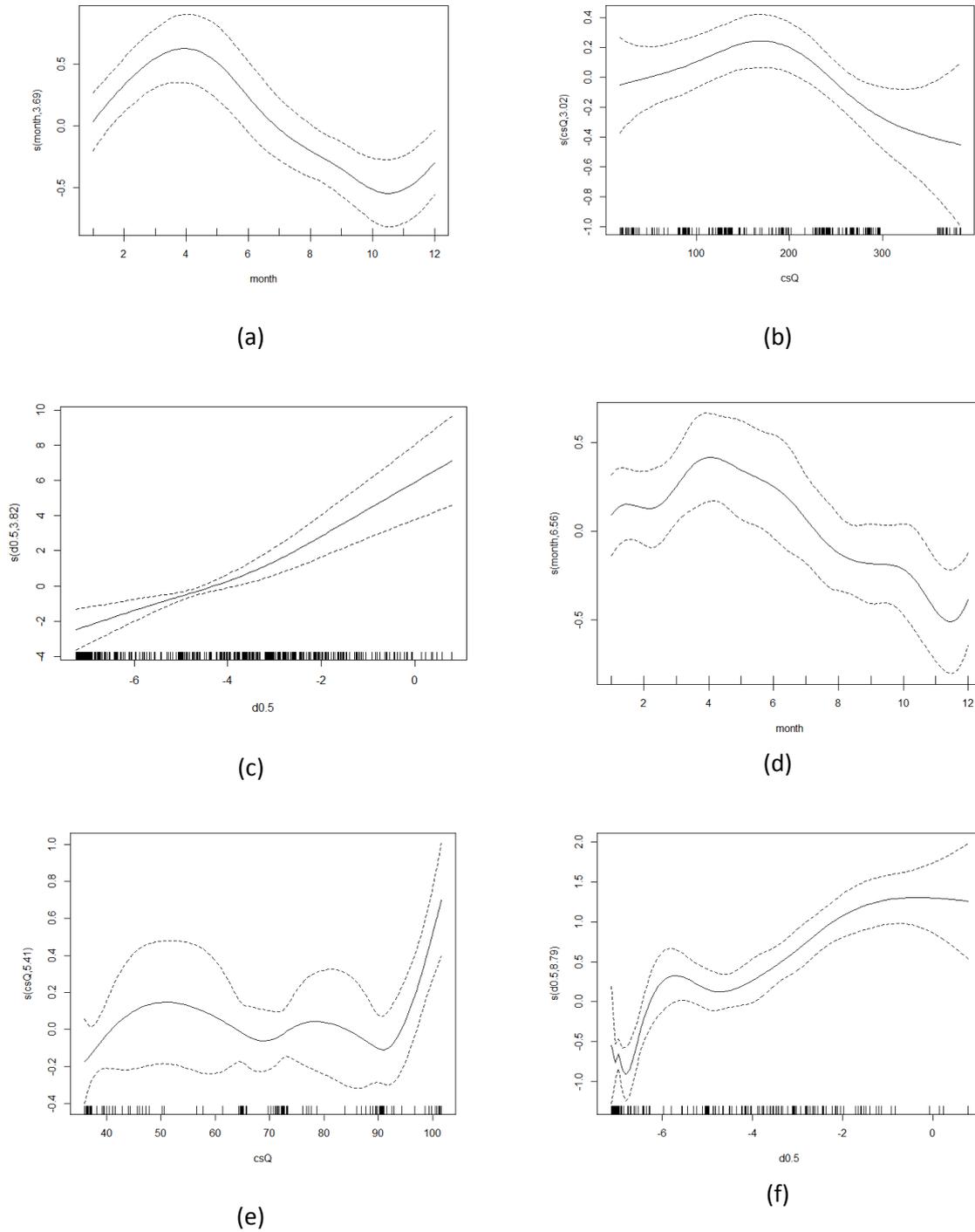
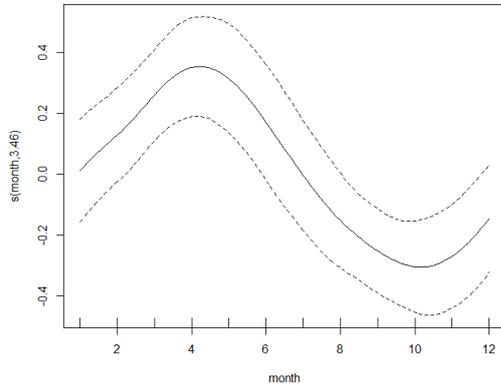
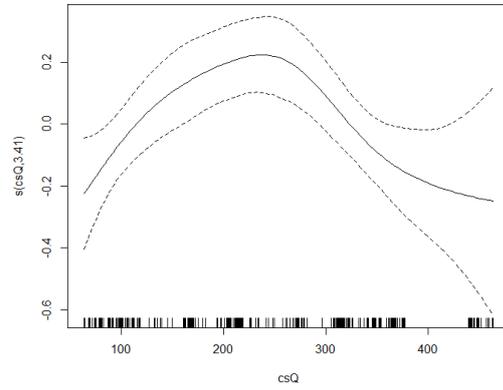


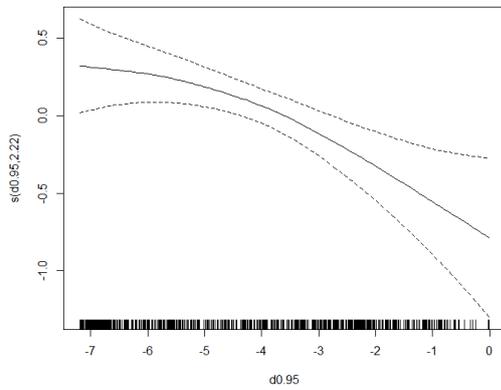
Figure 36: Smooth terms from the generalised additive models fit for Echunga Creek (A5030506) for TSS and TN. TSS: (a) seasonal term, (b) past sum of flow (c) discounted flow ( $d=0.5$ ); TN: (d) seasonal term, (e) past sum of flow (f) discounted flow ( $d=0.5$ ).



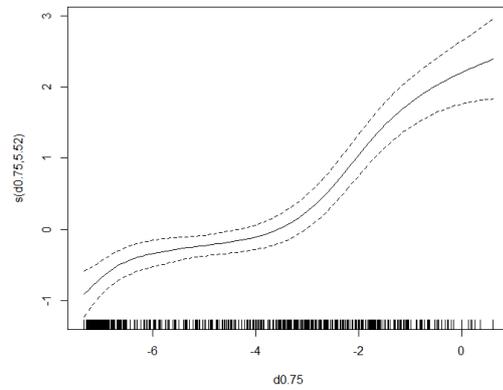
(a)



(b)



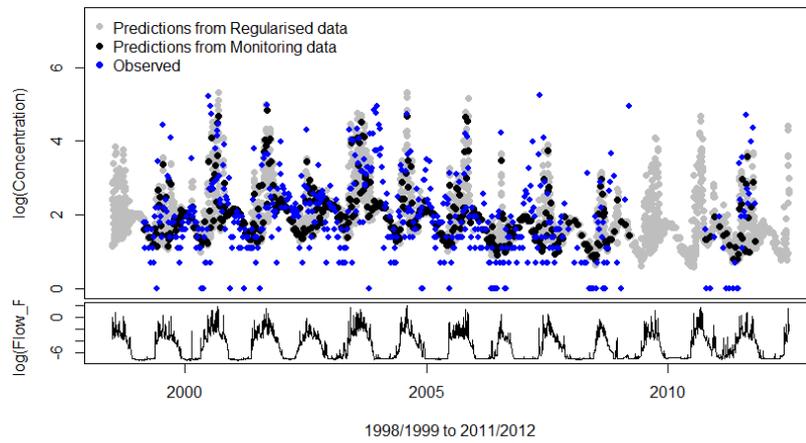
(c)



(d)

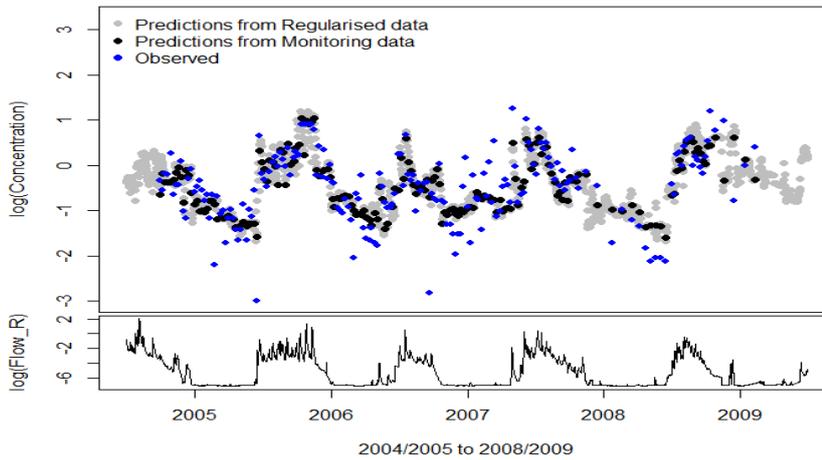
Figure 37: Smooth terms from the generalised additive model for Echunga Creek (A5030506) showing the (a) seasonal term, (b) past sum of flow (c) discounting term (d0.95) and (d) discounting term (d0.75) in relation to TP.

**Predicted Time Series Concentration**



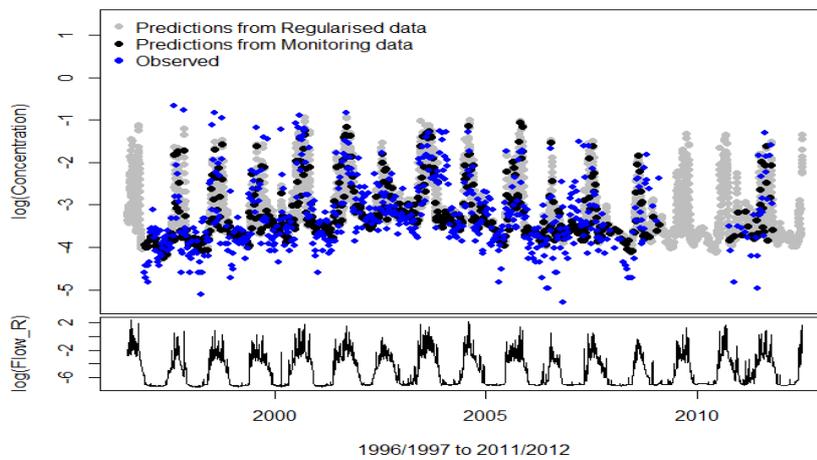
(a)

**Predicted Time Series Concentration**



(b)

**Predicted Time Series Concentration**



(c)

Figure 38: Predictions from the generalised additive model fit to constituent data at Echunga Creek Site (A5030506) for (a) TSS, (b) TN and (c) TP.

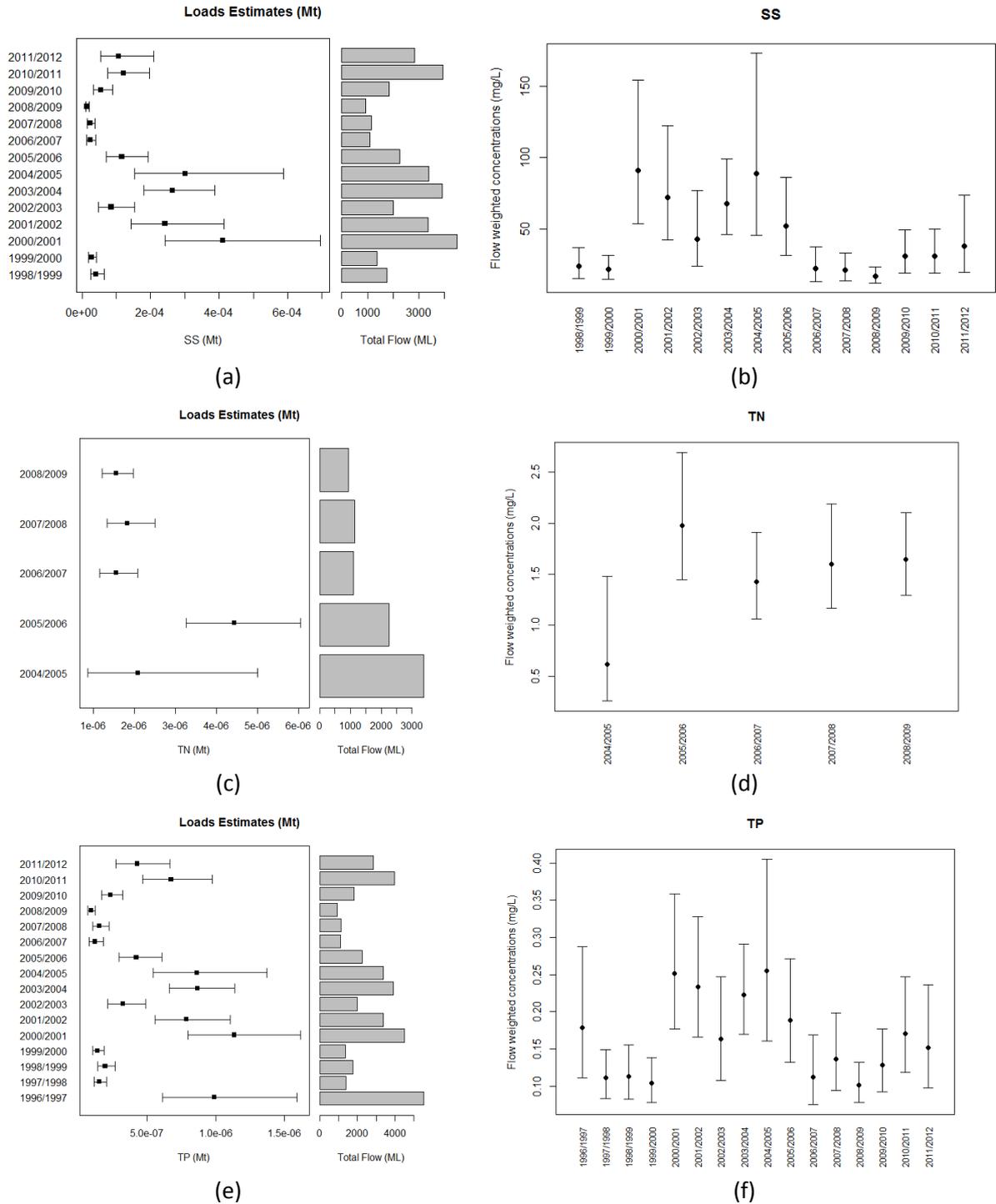


Figure 39: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for the Echunga Creek Site (A5030506) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP.

## Lenswood Creek (A5030507)

Data for Lenswood Creek is summarised in Figure 40-42, showing the time series for TSS, TN and TP and the accompanying flow, all of which are shown on the log-scale. Note that in 2009-2010 there is a period of flow that is fairly constant and close to, if not zero, indicating little or no flow. Upon checking the time series record, it is confirmed that the data on the raw scale has this feature. It is suggested that this flow record be checked for accuracy to ensure the low flow data is accurate.

Total suspended sediment was captured for the Lenswood Creek site between 1999 and 2012. A GAM was fit to the TSS data and is displayed in Table 19. The model uses a number of hydrological terms and also includes a seasonal term. This model explains nearly 55% of the variation in the data. The smooth terms represented in the model in Table 19 are shown in Figure 43. The seasonal term shows increases in TSS for the summer and autumn months, while the hydrological terms illustrate some complex interactions with the data as can be seen in Figure 43(b-f). These plots strongly illustrate the importance of past flows in predicting TSS for this site. Predictions from the model are shown in Figure 46 and indicate a reasonable fit to the data. Black points represent the predictions, blue points represent the observed data and the grey points represent predictions derived from daily flow. This is also supported by a series of diagnostic plots that are presented in Appendix B. Load estimates are shown in Figure 47 which shows the total load and accompanying annual flow (Figure 47(a)) and the flow weighted concentrations (Figure 47(b)). Load estimates for some years have quite large confidence intervals and this appears to be due to the size of flow produced for those years (e.g. 2010/11).

Total nitrogen data was measured for the period between 1996 and 2010. A GAM was fit to TN data that explained about 66% of the variation in the data. The model included flow and discounting terms along with an accumulation of flow. A seasonal term was also fit to the data and shows a peak in the summer and autumn months suggesting increases in TN for these periods (Figure 44(a)). Rises in TN are also noted for increases in quick flow and baseflow with a slightly sharper rise noted for quickflow compared to baseflow. The four discounting terms and an accumulation term highlight the complex interactions between flow and TN concentrations occurring at the site that explain the bulk of the variation in the data. Predictions for TN data at Lenswood Creek are shown in Figure 46 and indicate a reasonable fit to the data. This is also confirmed through a series of diagnostic plots in Appendix B. The black points in this figure represent the predictions while the blue points represent the observations. Predictions based on daily (regularised) flow data are shown in grey. Flow (log-scale) is shown beneath each plot for clarity. Load estimates based on the model in Table 20 are shown in Figure 46(c) and 46(d) and show variable TN loads through time with the exception to the first year where data was limited and the flow was large. The flow weighted concentrations show slight increases in 2005/06 through to 2008/2009 but with overlapping 80% confidence intervals.

Total phosphorous data for the Lenswood Creek site were obtained between 1996 through to 2012 and exhibit substantial variability over these years. A GAM was fit to the data to explore what parameters explained the bulk of the variation in the data (Table 21). This model explains approximately 57% of the variation in the data. Both baseflow and quickflow appeared important in the fit of the model with increases in TP noted for both, but a little higher for baseflow. Hydrological terms in the model included a series of discounting terms (short lags and longer lags) along with an accumulation of flow term, all of which are summarised graphically in Figure 45. The seasonal term identifies larger TP values during the summer and autumn months (Figure 45(a)) while the accumulated flow term presents a cyclical pattern that may be associated with certain types of rainfall events. Predicted TP is shown in 46 and along with the diagnostic plots in Appendix B, shows

good agreement with the data. Black points represent the predicted TP while blue points represent the data and grey points represent predictions from a regular (daily) flow record. The flow (log-scale) is shown beneath the plot. Estimated loads are shown in 47, where the total load and volume of flow is shown (Figure 47(e)) in addition to the flow weighted concentrations shown in Figure 47(f). The adjusted flow estimates shown in 47(f) show substantial variability in the width of confidence intervals. Based on these estimates and model, it is difficult to determine any significant trends in the data.

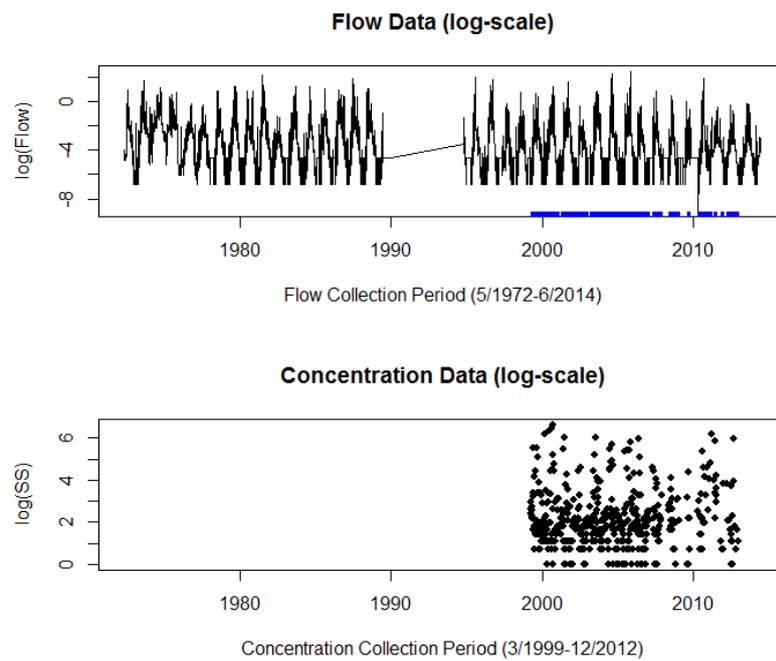


Figure 40: Raw flow and concentration data presented on the log-scale for TSS at the Lenswood Creek site (A5030507).

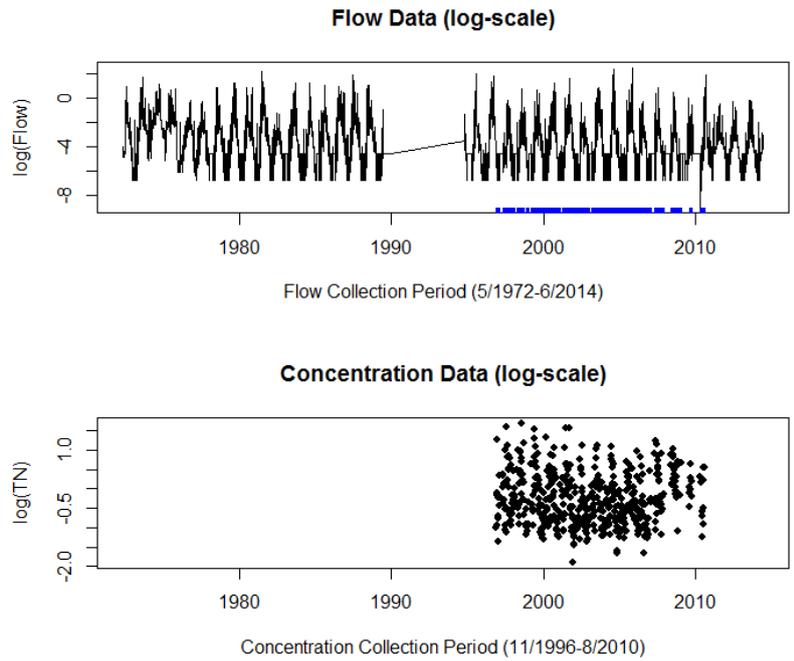


Figure 41: Raw flow and concentration data presented on the log-scale for TN at the Lenswood Creek site (A5030507).

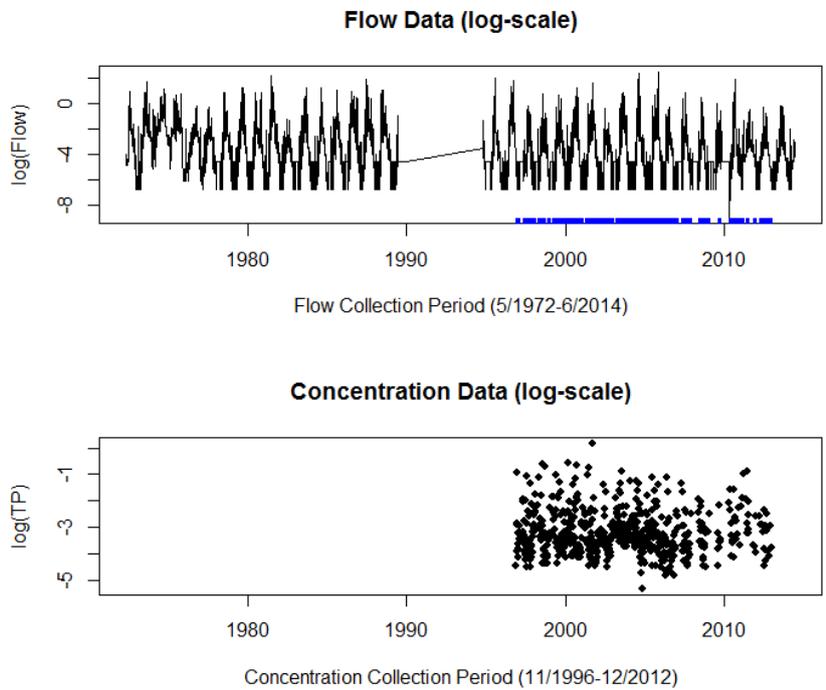


Figure 42: Raw flow and concentration data presented on the log-scale for TP at the Lenswood Creek site (A5030507).

Table 19: Summary of parameter estimates from generalised additive model fit to SS from Lenswood Creek (A5030507) that explains 54.2% of the variation in the data.

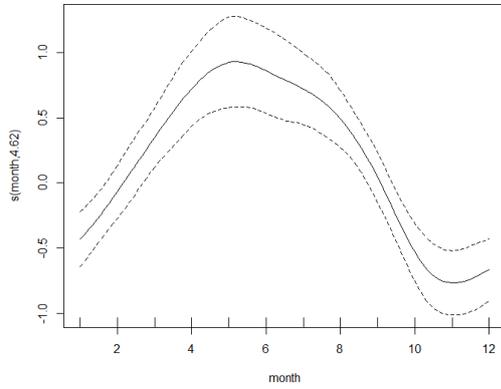
Parameter	Estimate	SE	p-value
Intercept	4.161	0.37	<0.001
Flow			
- log(Baseflow)	0.208	0.08	0.010
- log(Quickflow)	0.175	0.03	<0.001
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	4.621	6.598	<0.001
s(cumulative flow)	7.146	4.925	<0.001
Discounted Flow terms			
- 0.10	7.315	4.210	<0.001
- 0.25	9.000	3.585	<0.001
- 0.50	5.830	4.398	<0.001
- 0.75	1.344	10.994	<0.001

Table 20: Summary of parameter estimates from generalised additive model fit to TN from Lenswood Creek (A5030507) that explains 65.6% of the variation in the data.

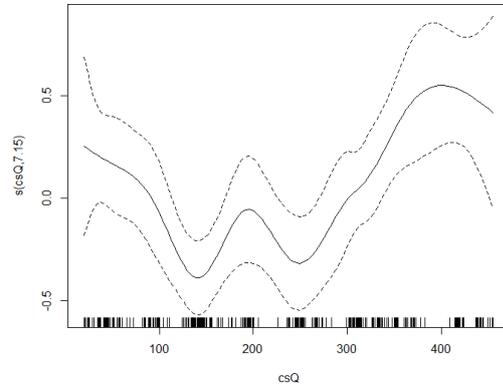
Parameter	Estimate	SE	p-value
Intercept	1.879	0.31	<0.001
Flow			
- log(Baseflow)	0.110	0.03	<0.001
- log(Quickflow)	0.074	0.02	<0.001
Discounted Flow Terms			
- 0.75	0.329	0.08	<0.001
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	5.907	10.410	<0.001
s(cumulative flow)	8.116	14.858	<0.001
Discounted Flow terms			
- 0.10	6.1396	2.159	0.035
- 0.25	7.954	2.094	0.032
- 0.95	6.188	7.319	<0.001

Table 21: Summary of parameter estimates from generalised additive model fit to TP from Lenswood Creek (A5030507) that explains 57.1% of the variation in the data.

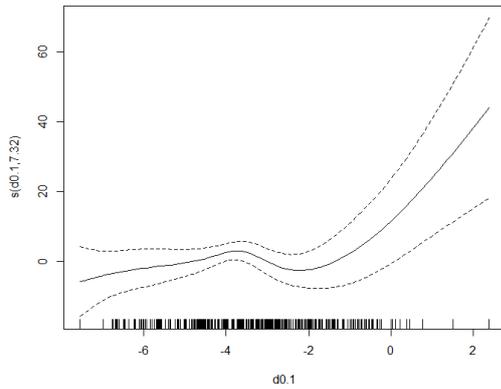
Parameter	Estimate	SE	p-value
Intercept	1.436	1.06	0.177
Flow			
- log(Baseflow)	0.208	0.04	<0.001
- log(Quickflow)	0.157	0.02	<0.001
Discounted Flow terms			
- d0.5	0.734	0.27	0.007
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	5.003	10.383	<0.001
s(cumulative flow)	7.031	4.498	<0.001
Discounted Flow terms			
- 0.10	5.919	2.658	0.0103
- 0.25	7.848	2.280	0.0193
- 0.95	3.748	10.570	<0.001



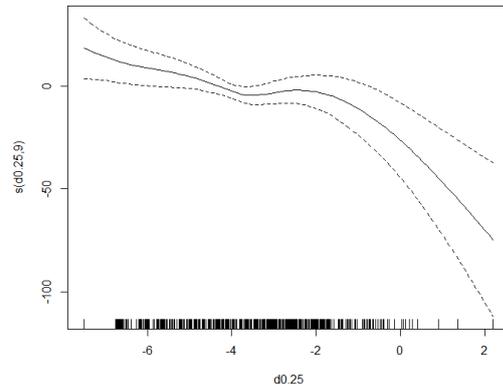
(a)



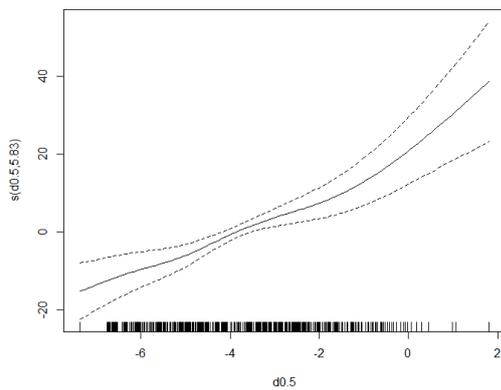
(b)



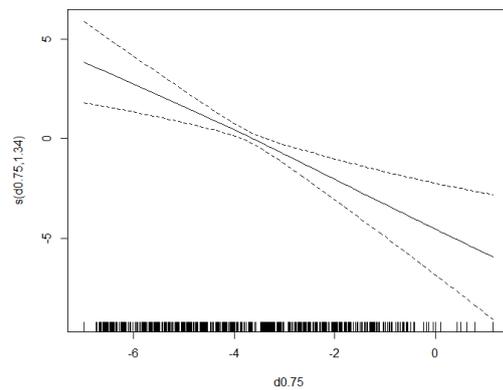
(c)



(d)

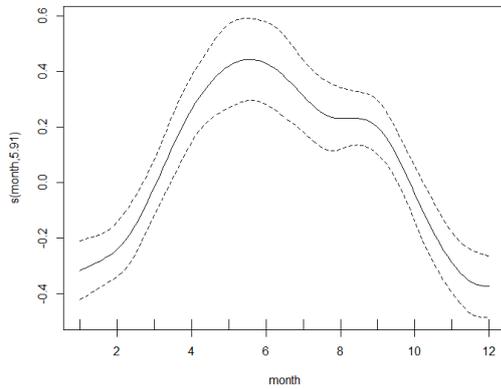


(e)

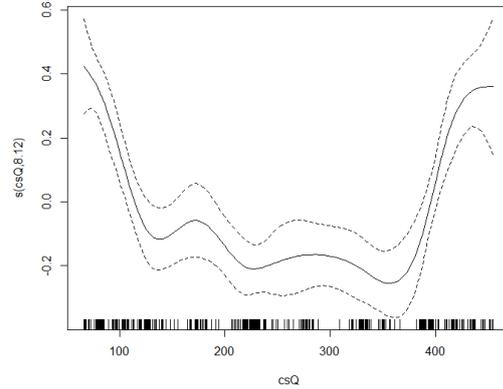


(f)

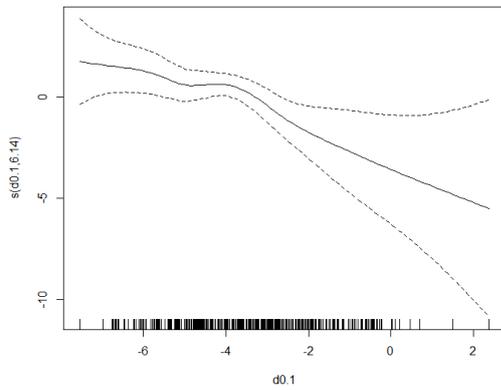
Figure 43: Smooth terms from the generalised additive model for Lenswood Creek (A5030507) showing the characteristics of (a) the seasonal term in the model, (b) the past sum of flow, (c) discounting term  $d0.1$ , (d) discounting term  $d0.25$ , (e) discounting term  $d0.5$  and (f) discounting term  $d0.75$  in relation to TSS.



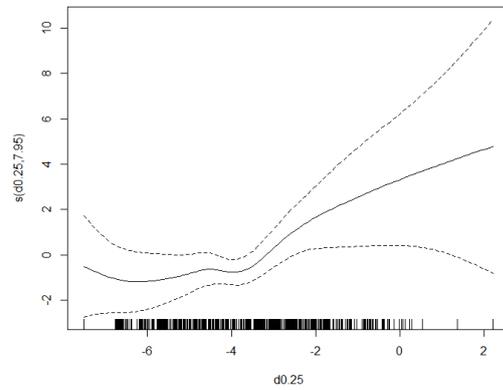
(a)



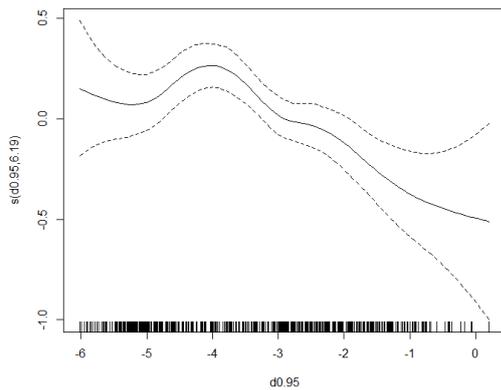
(b)



(c)

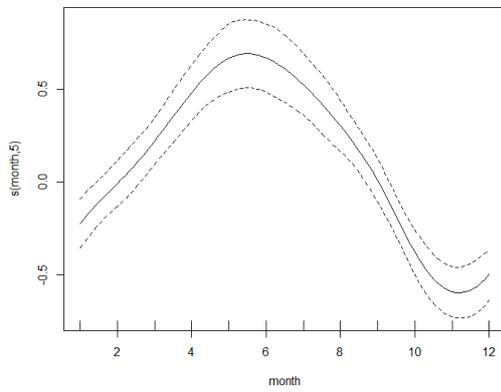


(d)

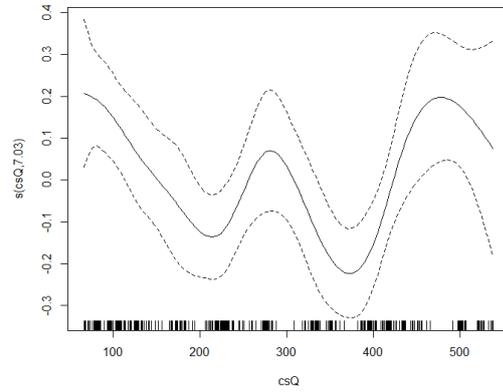


(e)

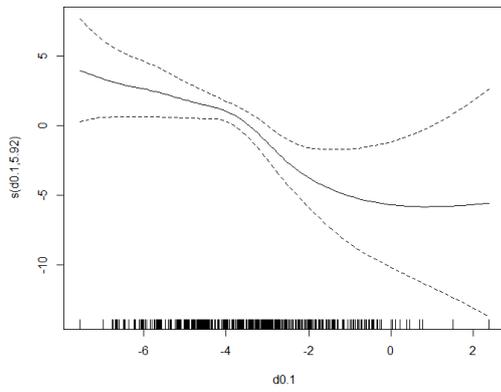
Figure 44: Smooth terms from the generalised additive model for Lenswood Creek (A5030507) showing the characteristics of (a) the seasonal term in the model, (b) the past sum of flow, (c) discounting term  $d0.1$ , (d) discounting term  $d0.25$  and (e) discounting term  $d0.95$  in relation to TN.



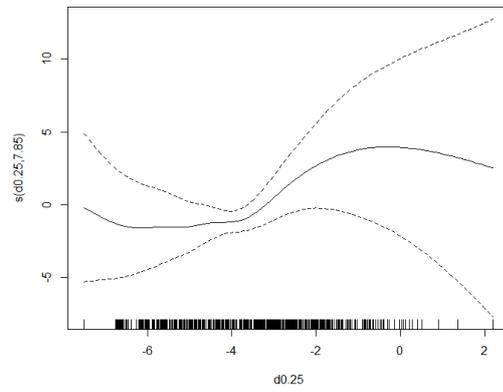
(a)



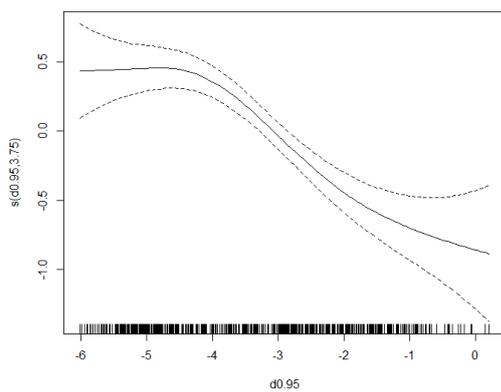
(b)



(c)

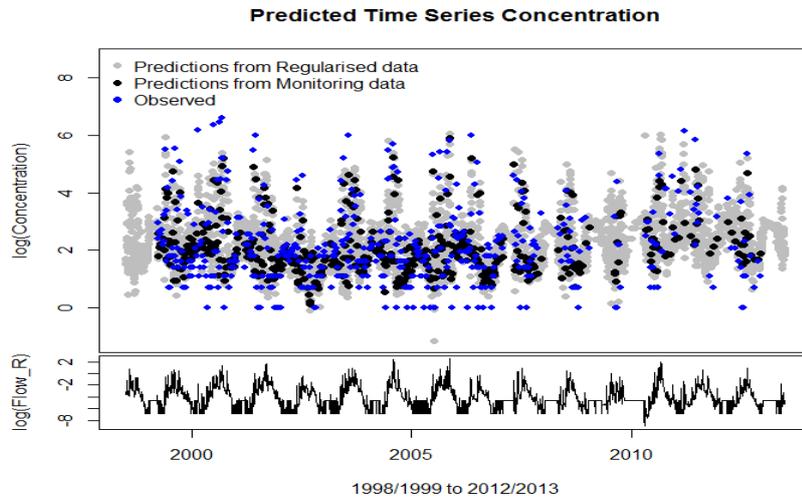


(d)

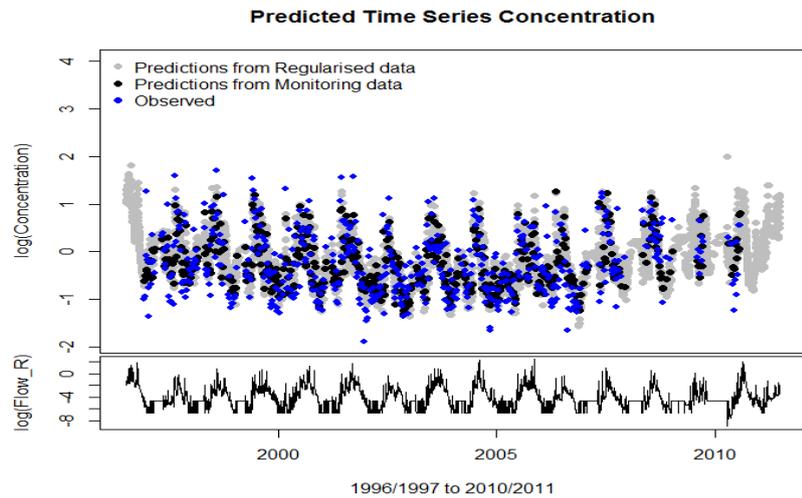


(e)

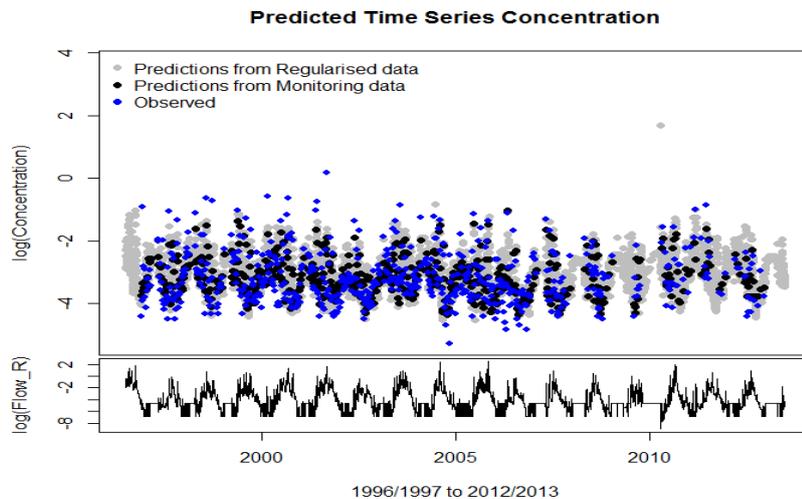
Figure 45: Smooth terms from the generalised additive model for Lenswood Creek (A5030507) showing the characteristics of (a) the seasonal term in the model, (b) the past sum of flow, (c) discounting term  $d0.1$ , (d) discounting term  $d0.25$  and (e) discounting term  $d0.95$  in relation to TP.



(a)



(b)



(c)

Figure 46: Predictions from the generalised additive model fit to constituent data at Lenswood Creek Site (A5030507) for (a) TSS, (b) TN and (c) TP.

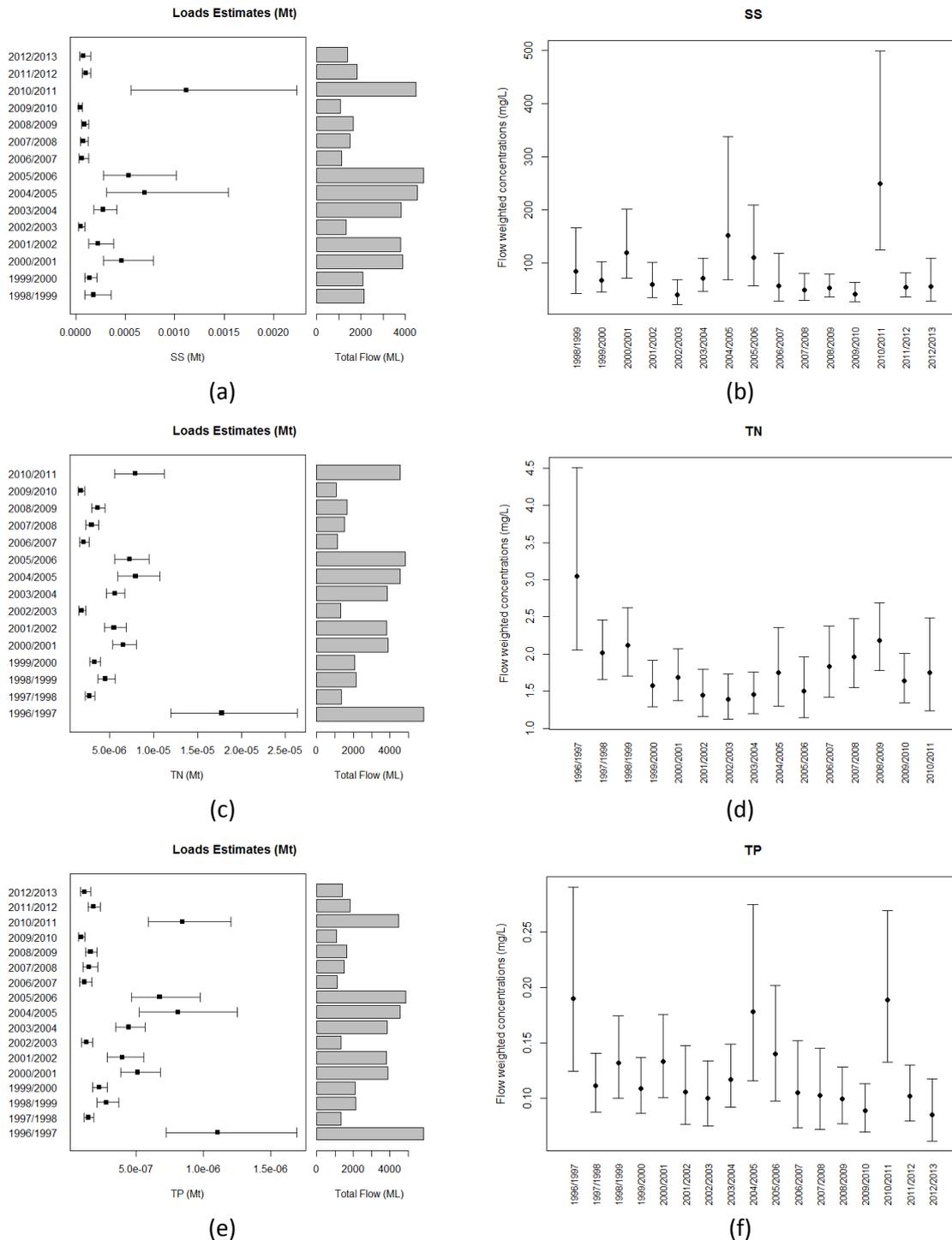


Figure 47: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for the Lenswood Creek Site (A5030507) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP.

## Aldgate Creek (A5030509)

Constituent data was collected for the Aldgate Creek catchment between 1996 and 2013 as shown in Figures 48 - 50. The corresponding flow captured by the gauge is also shown in this figure (log-scale) and highlights a gap in the data. As for the Lenswood Creek site, flow between 2009 and 2010 indicates limited or no flows occurring during that period. It is suggested that this data be checked to determine whether this pattern is accurate.

A GAM was fit to the TSS data (log-scale) that included a seasonal term and a suite of hydrological terms that included baseflow and quickflow, an accumulation of flow term and discounting terms to capture historical flows at this site. The results are displayed in Table 22 and Figure 51. Table 22 highlights that both baseflow and quickflow are important predictors of TSS, with an increase in baseflow leading to an increase in sediment. Increases in quickflow are also associated with increases in sediment, however the effect is not as large as for baseflow as the coefficient in the model is smaller. The seasonal term fit in the model is shown in Figure 51(a) and shows increases in TSS between March and August compared with other months of the year. The hydrological terms fit in the model that include the accumulation of flow and the discounting terms indicate a complex hydrological relationship that leads to the generation of TSS (Figures 51(b-e)). Predictions from the model are shown in Figure 53(a) and together with the diagnostic plots shown in Appendix B suggest a reasonable fit to the data. Predictions from the model are shown in black while the observed data is shown in blue. Grey points represent predictions made on a regular time series of flow. Estimates of the annual loads and flow-weighted concentrations are shown in Figure 54. In Figure 54(b), all years (with the exception of the first year) show little or no trend in the data and are accompanied by wide credible intervals.

Total nitrogen collected for the Aldgate Creek catchment is shown in Figure 50 along with the daily time series of flow. TN data span the period 2004 through to 2009 and show a possible slight increase in TN as we progress through the years. A GAM was fit to the data (summarised in Table 23) that explained approximately 51% of the variation in the data. Table 23 highlights that quickflow is important, suggesting an increase in TN as runoff increases. Baseflow is not significant and therefore does not contribute to TN generation. Two discounted flow terms and an accumulation of flow term were fitted. In this model, a seasonal term was not required. In this model the accumulation of past flows tends to result in an increase in TN as can be seen in Figure 51(f) and Table 23. Predictions from the model are displayed in Figure 53(b) and show a reasonable fit with the data. The results suggest an increasing trend as flows increase. Figure 54 shows the loads estimates and 80% confidence interval of the load. Figure 54 presents the loads estimates and flow-weighted concentrations. Results appear to suggest an increase in TN and this is particularly evident for Figure 54(d).

Total phosphorous was collected for the Aldgate Creek site between 1999 and 2011. A GAM was fit to the data that explained 47% of the variation in the data and consisted of both baseflow and runoff terms in the model. Both flow types were associated with an increase in TP as flow increased. The accumulation and discounting terms in this model provided a mechanism for capturing complex hydrological relationships (both linear and non-linear) between flow and phosphorous concentrations. The seasonal term showed a similar pattern to the smooth plots for other constituents and other sites suggesting that higher TP values occur between March and June. These figures are displayed in Figure 52. Predictions for this model are shown in Figure 53(c) and



together with the diagnostic plots appearing in Appendix B indicate a good fit to the data. Load estimates and flow-weighted concentrations are shown in Figure 54 and after adjusting for flow, Figure 54(e) illustrates substantial variability from year to year in the load estimates.

Table 22: Summary of parameter estimates from generalised additive model fit to SS from Aldgate Creek (A5030509) that explains 44.9% of the variation in the data.

Parameter	Estimate	SE	p-value
Intercept	13.287	2.60	<0.001
Flow			
- log(Baseflow)	0.424	0.07	<0.001
- log(Quickflow)	0.065	0.03	0.023
Discounted Flow Terms			
- d0.50	2.433	0.64	0.0001
- d0.95	-0.294	0.14	0.034
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	6.717	4.962	<0.001
s(cumulative flow)	6.203	2.450	0.017
Discounted Flow terms			
- 0.10	6.106	4.312	0.0001
- 0.25	6.590	4.168	0.0001
- 0.75	3.094	4.016	0.003

Table 23: Summary of parameter estimates from generalised additive model fit to TN from Aldgate Creek (A5030509) that explains 50.6% of the variation in the data.

Parameter	Estimate	SE	p-value
Intercept	0.490	0.11	<0.001
Flow			
- log(Baseflow)	0.013	0.39	0.696
- log(Quickflow)	0.046	2.11	0.037
Discounted Flow Terms			
- d0.10	-0.397	0.06	<0.001
- d0.50	0.499	0.07	<0.001
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(cumulative flow)	5.242	6.104	<0.001

Table 24: Summary of parameter estimates from generalised additive model fit to TP from Aldgate Creek (A5030509) that explains 47.1% of the variation in the data.

Parameter	Estimate	SE	p-value
Intercept	-0.382	0.34	0.262
Flow			
- log(Baseflow)	0.128	0.03	<0.001
- log(Quickflow)	0.049	0.01	<0.001
Discounted Flow Terms			
- 0.25	-0.743	0.12	<0.001
- 0.50	1.088	0.19	<0.001
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	4.314	4.872	<0.001
s(cumulative flow)	5.692	7.313	<0.001
Discounted Flow terms			
- 0.75	6.187	2.162	<0.001
- 0.95	4.881	6.092	0.035

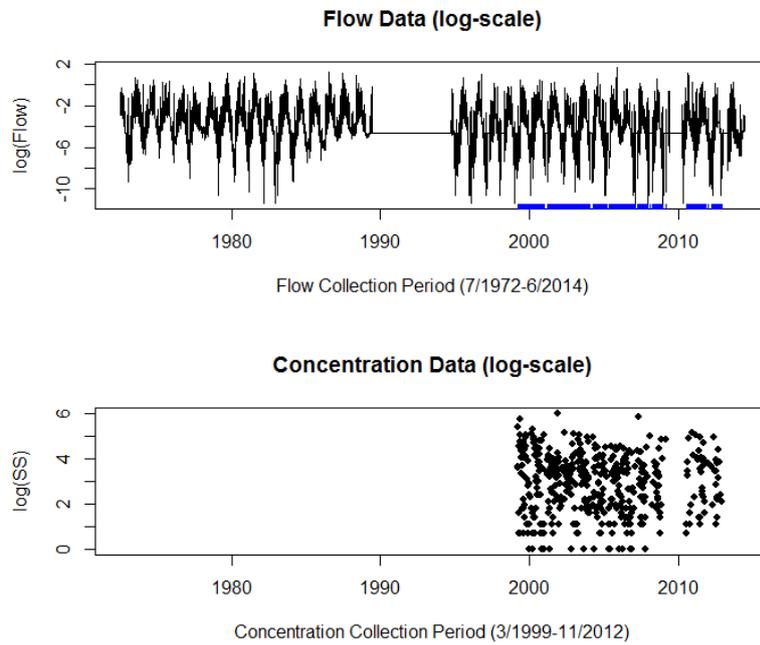


Figure 48: Raw flow and concentration data presented on the log-scale for TSS at the Aldgate Creek site (A5030509).

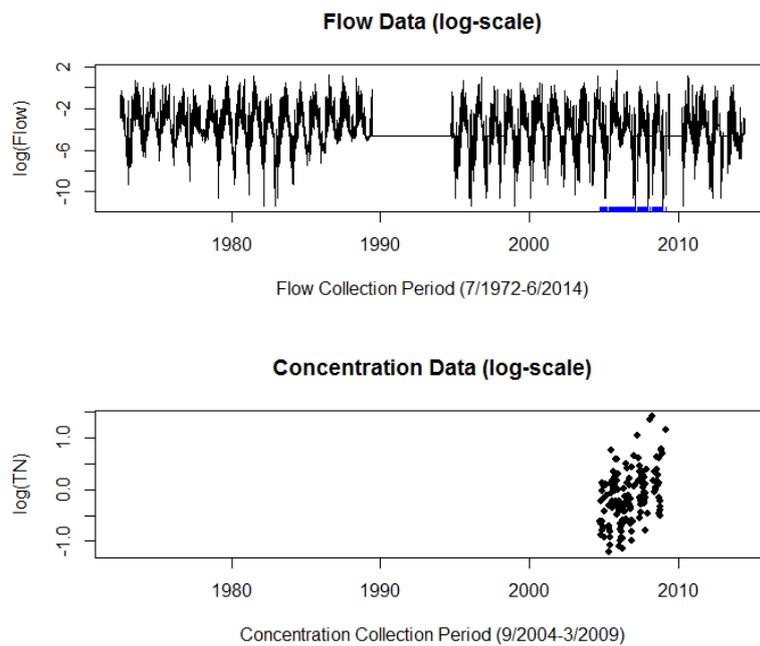


Figure 49: Raw flow and concentration data presented on the log-scale for TN at the Aldgate Creek site (A5030509).

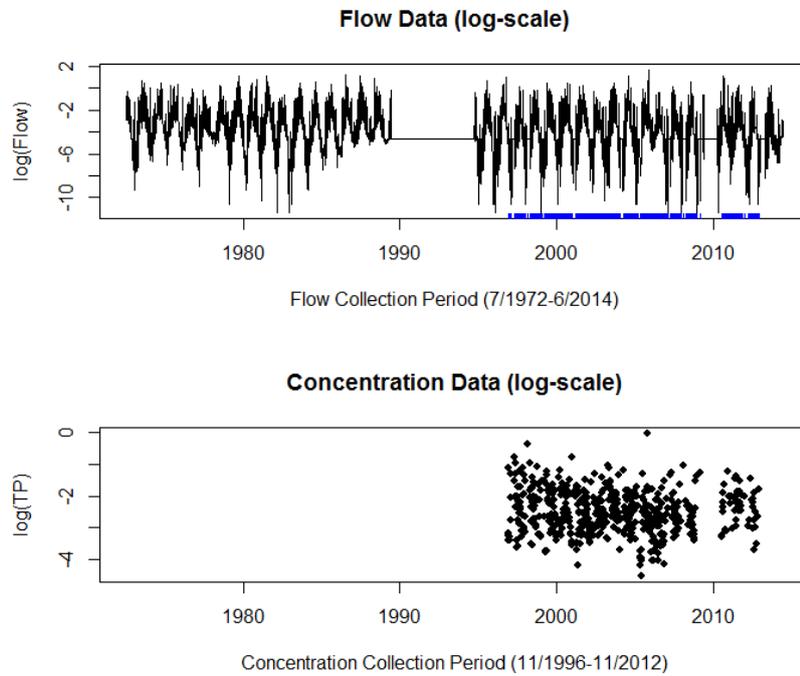
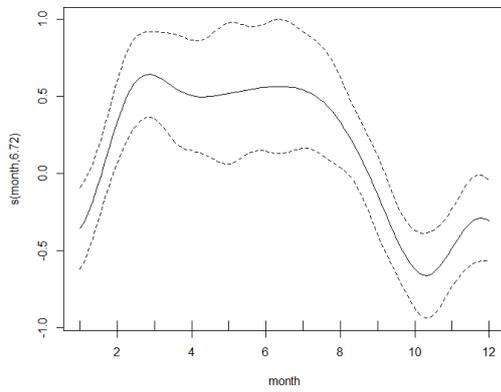
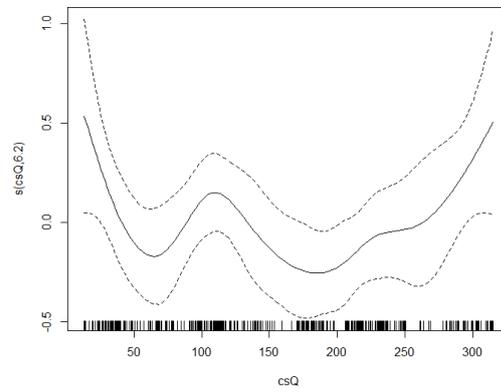


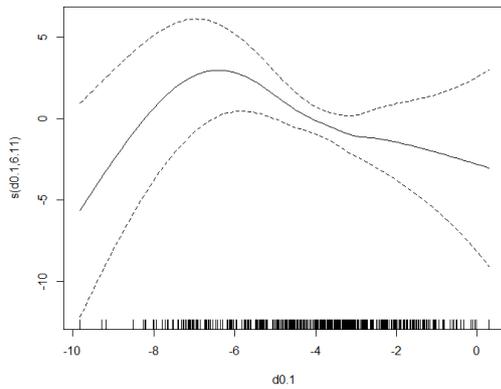
Figure 50: Raw flow and concentration data presented on the log-scale for TP at the Aldgate Creek site (A5030509).



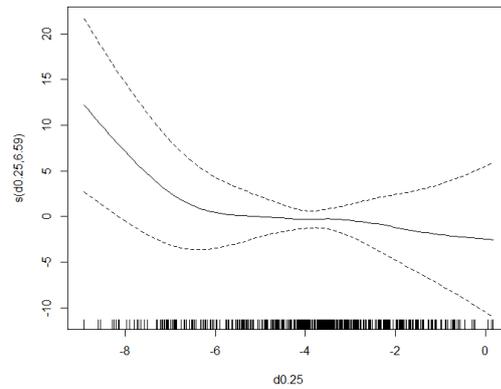
(a)



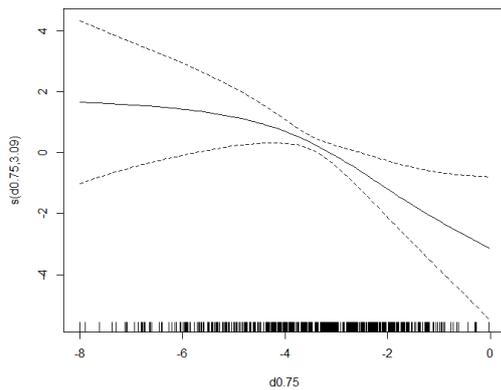
(b)



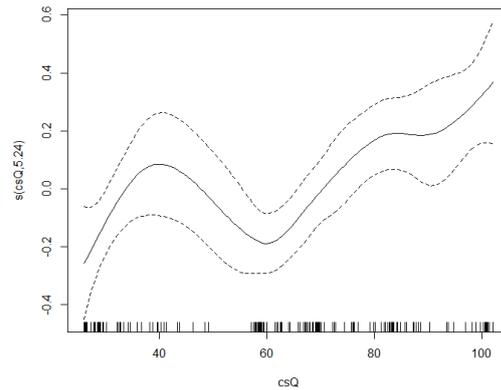
(c)



(d)

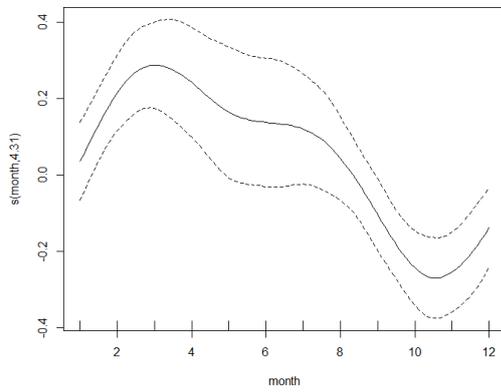


(e)

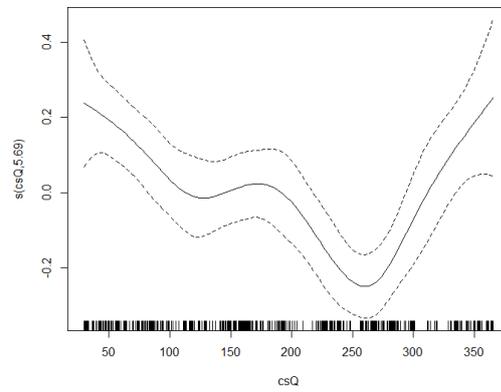


(f)

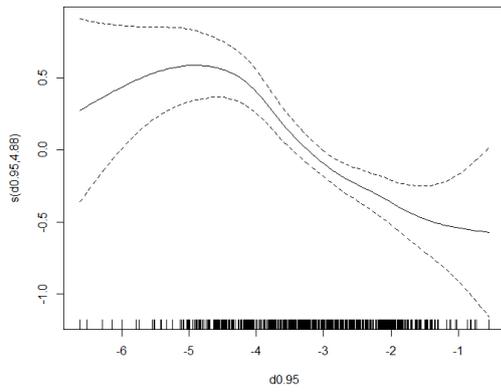
Figure 51: Smooth terms from the generalised additive models fit for Aldgate Creek (A5030509) for TSS and TN. TSS: (a) seasonal term, (b) past sum of flow, (c) discounted flow (d0.1), (d) discounted flow (d0.25), (e) discounted flow (d0.75); TN: (f) past sum of flow.



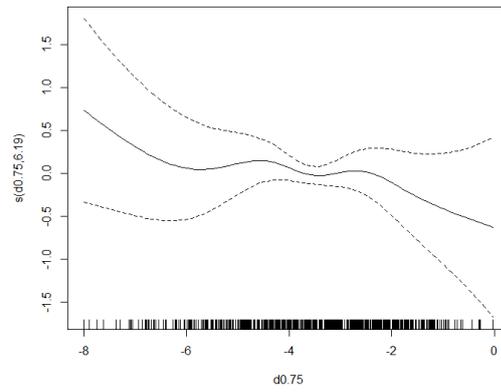
(a)



(b)



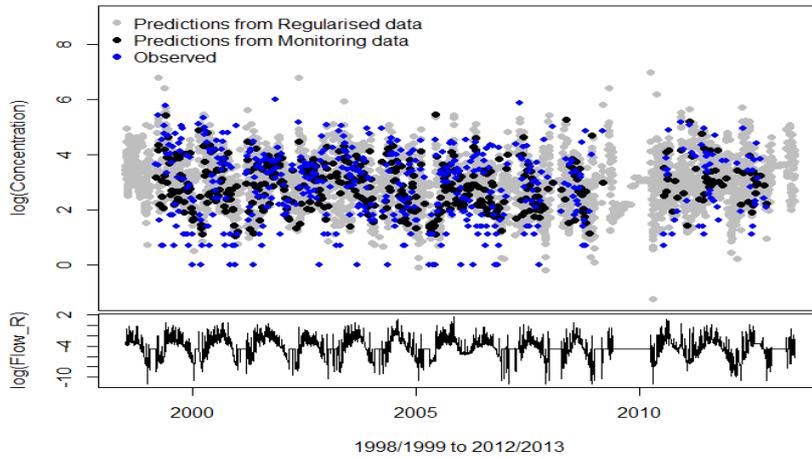
(c)



(d)

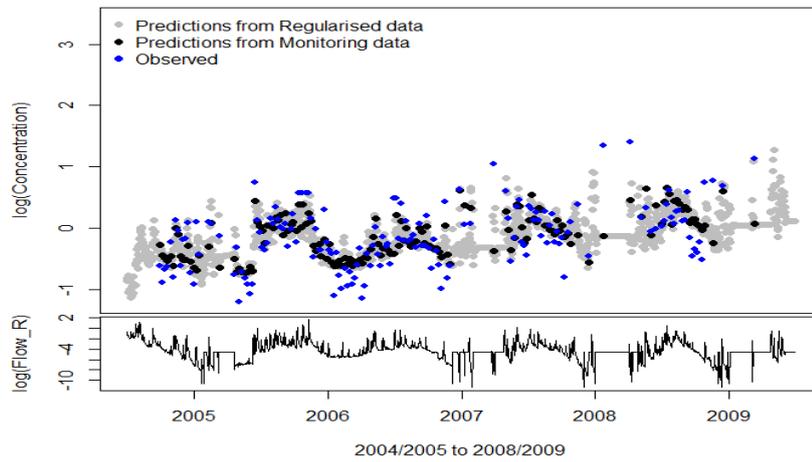
Figure 52: Smooth terms from the generalised additive model for Aldgate Creek (A5030509) showing the characteristics of (a) the seasonal term in the model, (b) the past sum of flow, (c) discounting term  $d0.95$  and (d) discounting term  $d0.75$  in relation to TP.

**Predicted Time Series Concentration**



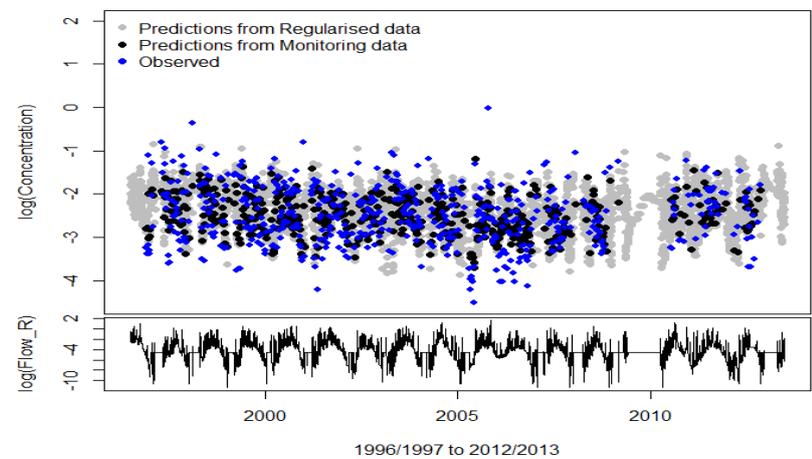
(a)

**Predicted Time Series Concentration**



(b)

**Predicted Time Series Concentration**



(c)

Figure 53: Predictions from the generalised additive model fit to constituent data at Aldgate Creek Site (A5030509) for (a) TSS, (b) TN and (c) TP.

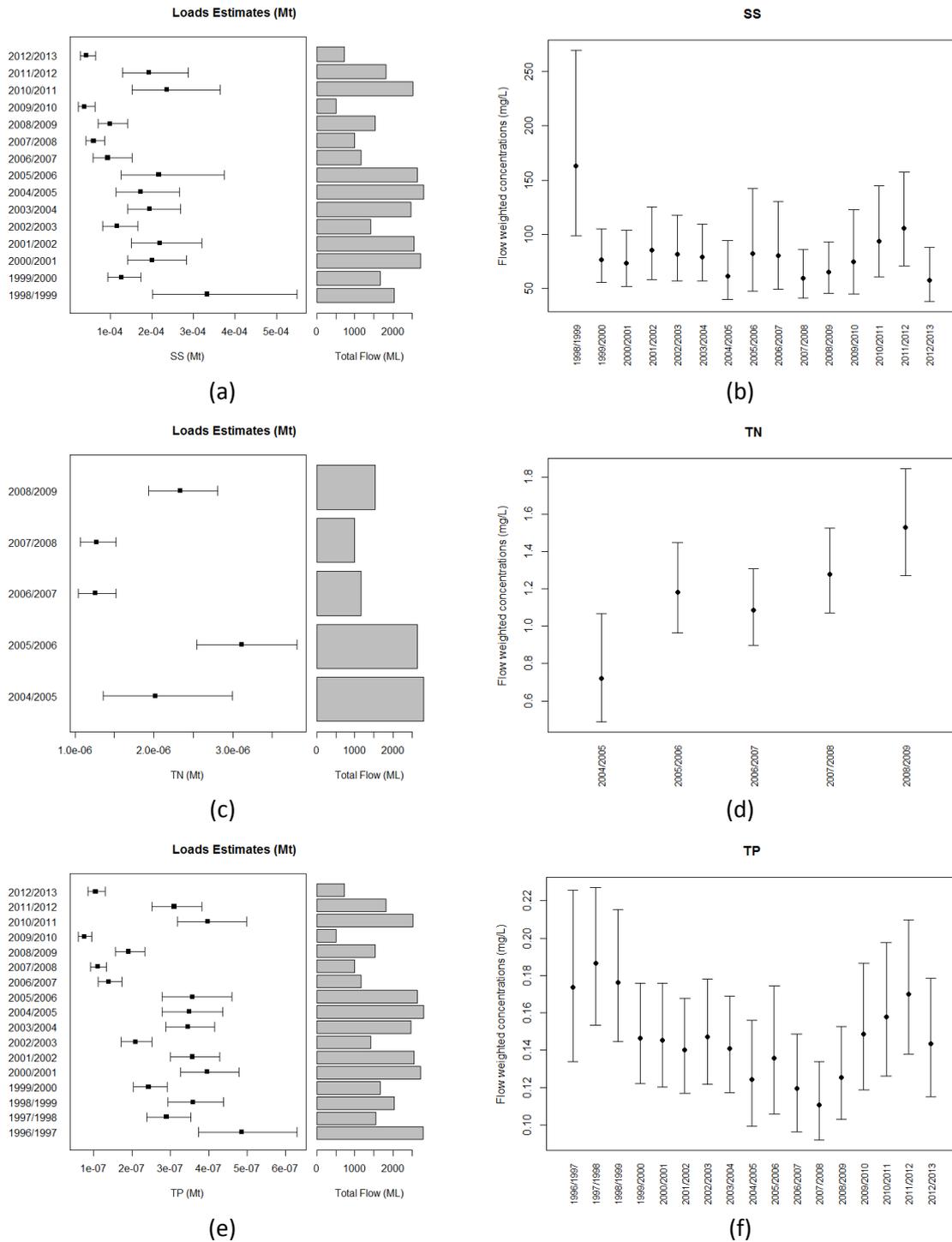


Figure 54: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for the Aldgate Creek Site (A5030509) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP.

## Cox Creek at Uraidla (A5030526)

Constituent and flow data was collected at the Cox Creek, Uraidla site to investigate the primary drivers of water quality. Figures 56 - 58 show the time series of flow (log-scale) and samples of TSS, TN and TP respectively. Where constituent data was collected, daily flow was measured regularly through time. In addition to the typical hydrological relationships that we have explored for the previous five sites in the Onkaparinga catchment, we also explored the potential influence of the Brooks Bridge sedimentation basin, hereafter referred to as the “sedimentation pond”, which is located on Swamp road, downstream of the township of Uraidla. The installation of the sedimentation pond on the 30 November 2006 was intended to reduce sediment and nutrient loads from the upper catchment (these loads were considered to be unusually high for a relatively small catchment) and therefore limit the number of algal blooms at this site (Frizenschaf and Vial, 2012). We constructed a binary covariate, where a value of 1 indicated the pond was operational (i.e. any date after the 30 November 2006). In all three analyses of the constituents, the influence of the sedimentation pond was difficult to disentangle as there appeared to be an important interaction with flow and in particular, the historical build up of flow throughout the sampling period. As such, the results below need to be interpreted with some caution as further investigation would be required to conclusively state what the impact of the sedimentation pond has been on water quality.

Table 25 provides the results from a GAM fit to the TSS data at the Cox Creek site. In addition to the hydrological terms used in the model, we also investigated the impact of the sedimentation pond and noted a strong association with the accumulation of flow (csQ) through time and fit that interaction accordingly. The interpretation of the sedimentation pond and flow term are depicted in Figure 55(a) that shows the marginal effects of the pond as flow accumulates through the subcatchment for operational and non-operational periods. The figure suggests that when not operational, as flow accumulates, the amount of TSS decreases linearly. However, when it became operational, the effect was the opposite, resulting in an increase in TSS, although the “increase” is slight due to the negative values shown on the y-axis (log-scale) that equate to very small increases in TSS.

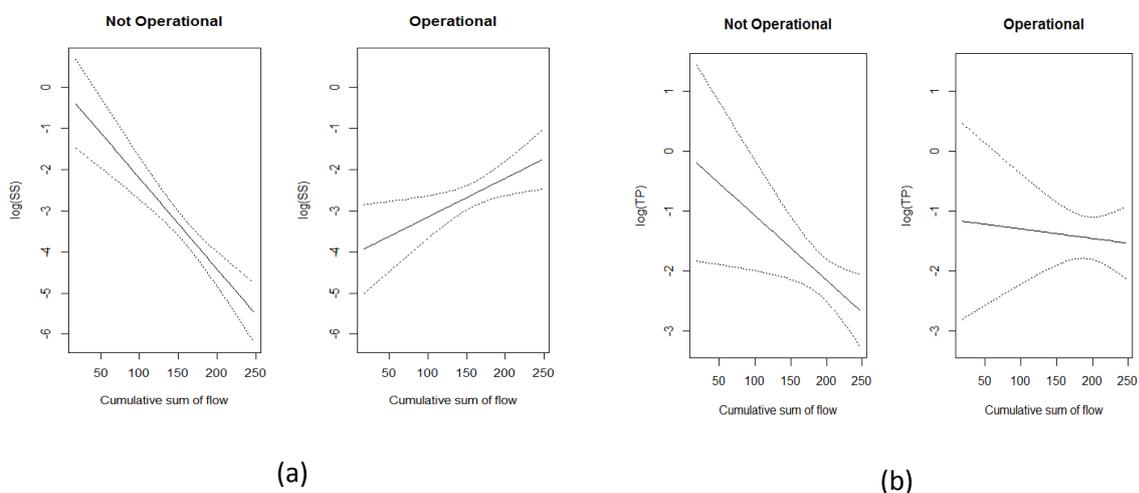


Figure 55: Marginal effect of Brooks Bridge sedimentation basin on TSS (a) prior to being operational and operational post 30 November 2006 and TP (a) prior to being operations and operational past 30 November 2007. Dotted lines represent 95% confidence intervals around the marginal effect.

Table 26 displays the parameter estimates from the GAMM fit to the TN data. In this model it was very difficult to fit a “sedimentation pond effect” with flow as the residuals from the model suggested that additional information was needed to explain some strong associations between TN samples collected. After incorporating an AR1 term to account for the association between sampling times, the interaction between the sedimentation pond and flow was no longer significant and a sedimentation pond effect on its own was suggesting a positive increase in TN when the pond was operational. As such, we decided to exclude this term from the model and only included the hydrological relationships to explain the variation in TN through time.

A GAMM was fit to TP data collected for the Cox Creek site at Uraidla and included an interaction term between the sedimentation pond and the accumulation of flow along with a complex suite of flow variables to mimic the hydrological relationships at this site. Figure 55(b) shows the marginal contribution of the sedimentation pond term with respect to flow as it accumulates through the catchment. For this constituent it appears that both trends exhibited are negative but the stronger downward trend in TP is noted for when the sedimentation pond was not operational.

Figures 59 and 60 summarise the smooth terms fit in each of the models. The seasonal term consistently shows a peak during February and March, suggesting an increase in constituents during these months over and above other months of the year. The hydrological terms for each model exhibit some of the complex hydrological relationships existing at this site.

Figure 61(a-c) summarise the predictions for TSS, TN and TP respectively based on the models shown in Tables 25-27 and showing a reasonable fit when compared with the observations. As in previous models, there is some difficulty predicting extremes. As suggested earlier, this could in part be due to the sampling design and in particular, due to the composite sampling approach used which does not focus on extremes occurring at this site. There may also be other terms (hydrological or other) not explored that could better explain the dynamics. Loads estimates are shown in Figures 62(a-c) for each of the three constituents and demonstrate a gradual decline in the constituents since around 2000. The gradual decrease with overlapping confidence intervals for total nitrogen may explain why we could not detect a significant effect of the sedimentation pond with the correlation term in the model. Overall, the results of the analysis suggest a decline through time and this is most likely because of one or more of: (i) a targeted campaign of landholder education; (ii) the installation of the sedimentation pond; and (iii) lower flows during the drought period from 2006 to 2009.

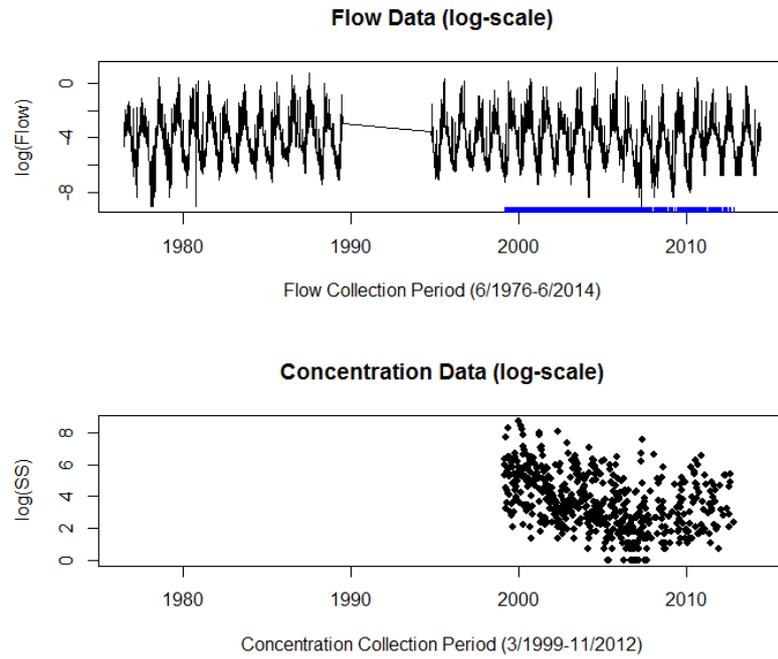


Figure 56: Raw flow and concentration data presented on the log-scale for TSS at the Cox Creek site (A5030526).

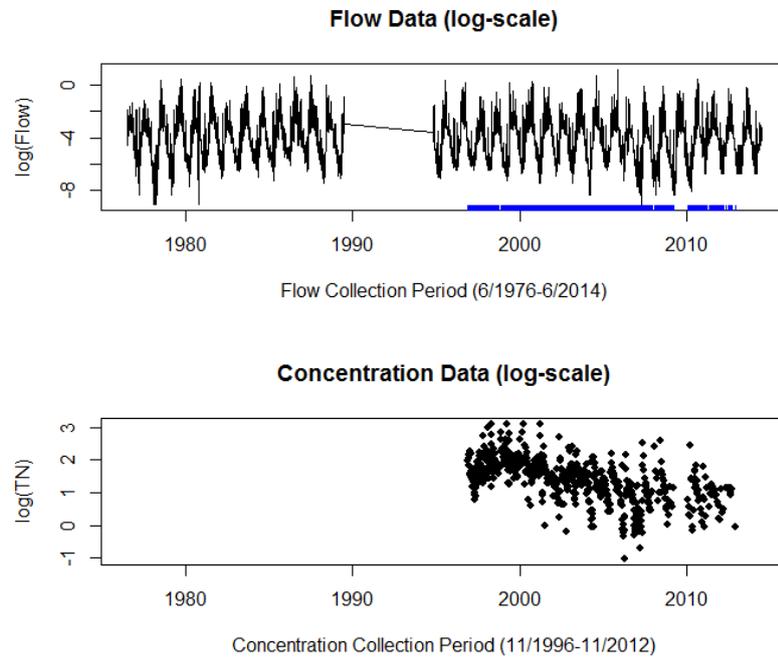


Figure 57: Raw flow and concentration data presented on the log-scale for TN at the Cox Creek site (A5030526).

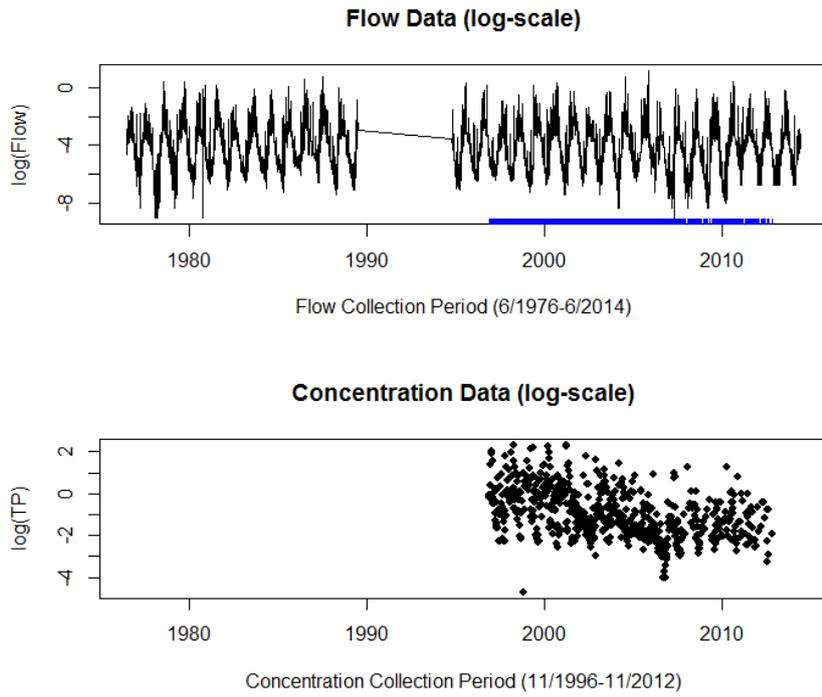


Figure 58: Raw flow and concentration data presented on the log-scale for TP at the Cox Creek site (A5030526).

Table 25: Summary of parameter estimates from generalised additive model fit to TSS from Cox Creek (A5030526) that explains 56.9% of the variation in the data.

Parameter	Estimate	SE	p-value
Intercept	9.280	0.35	<0.001
Flow			
<i>log(Baseflow)</i>	0.157	0.11	0.154
<i>log(Quickflow)</i>	0.197	0.04	<0.001
Influence of sedimentation pond with flow			
<i>Pond Operational</i>	-4.098	0.72	<0.001
<i>Cumulative sum flow</i>	-0.022	0.002	<0.001
<i>Interaction</i>	0.032	0.004	<0.001
Discounted Flow Terms			
- 0.95	-0.577	0.13	<0.001
- 0.50	3.237	0.36	<0.001
- 0.25	-2.262	0.34	<0.001
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	4.17	13.64	<0.001

Table 26: Summary of parameter estimates from generalised additive model fit to TN from Cox Creek (A5030526) that explains 58.9% of the variation in the data.

Parameter	Estimate	SE	p-value
Intercept	1.485	0.14	<0.001
Flow			
- <i>log(Baseflow)</i>	0.012	0.03	0.70
- <i>log(Quickflow)</i>	0.007	0.01	0.40
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	5.301	8.358	<0.001
s(cumulative sum flow)	2.114	65.909	<0.001
Discounted Flow Terms			
- 0.25	4.005	4.581	0.001
- 0.75	6.270	13.771	<0.001
<b>Correlation Term – AR(1)</b>	<b>Estimate</b>	<b>95% CI</b>	
$\phi$	0.548	[0.47,0.62]	

Table 27: Summary of parameter estimates from generalised additive model fit to TP from Cox Creek (A5030526) that explains 57% of the variation in the data.

Parameter	Estimate	SE	p-value
Intercept	2.032	0.25	<0.001
Flow			
- log(Baseflow)	0.118	0.06	0.057
- log(Quickflow)	0.091	0.02	<0.001
Discounted Flow Terms			
- 0.25	-1.572	0.26	<0.001
- 0.50	2.503	0.43	<0.001
- 0.75	-0.382	0.26	0.142
- 0.95	-0.433	0.11	0.0001
Influence of sedimentation pond with flow			
<i>Pond Operational</i>	-1.138	0.66	0.085
<i>Cumulative sum flow</i>	-0.011	0.001	<0.001
<i>Interaction</i>	0.009	0.003	0.006
<b>Smooth Terms</b>	<b>EDF</b>	<b>F-statistic</b>	<b>p-value</b>
s(month)	5.273	15.3	<0.001
<b>Correlation Term – AR(1)</b>	<b>Estimate</b>	<b>95% CI</b>	
$\phi$	0.261	[0.18,0.34]	

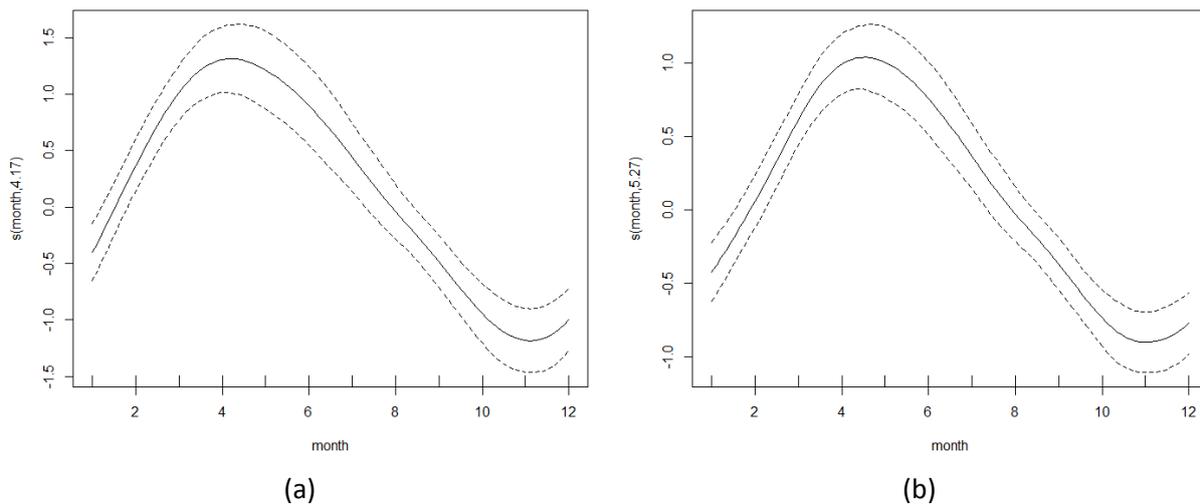
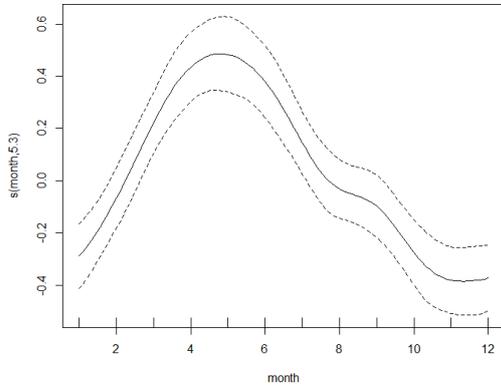
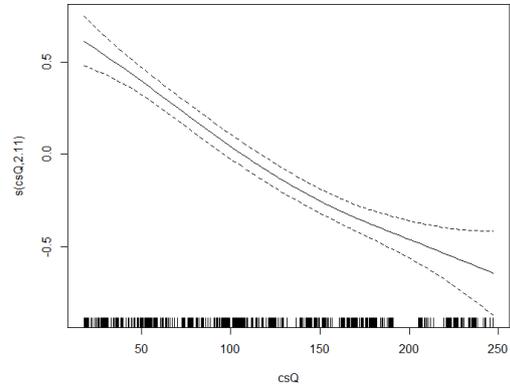


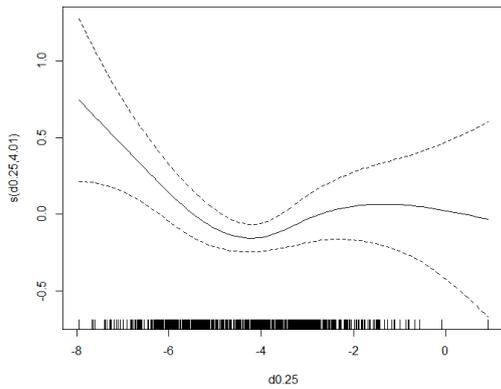
Figure 59: Smooth terms from the generalised additive models fit for Cox Creek (A5030526) for TSS and TP. TSS: (a) seasonal term; TP: (b) seasonal term.



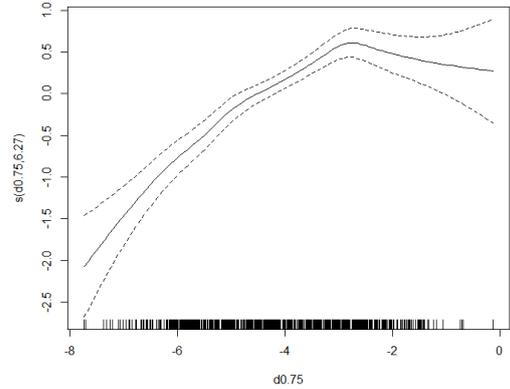
(a)



(b)

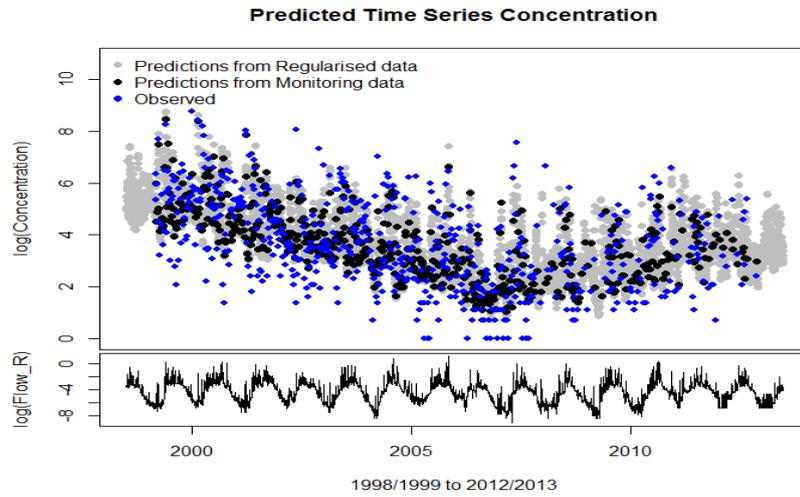


(c)

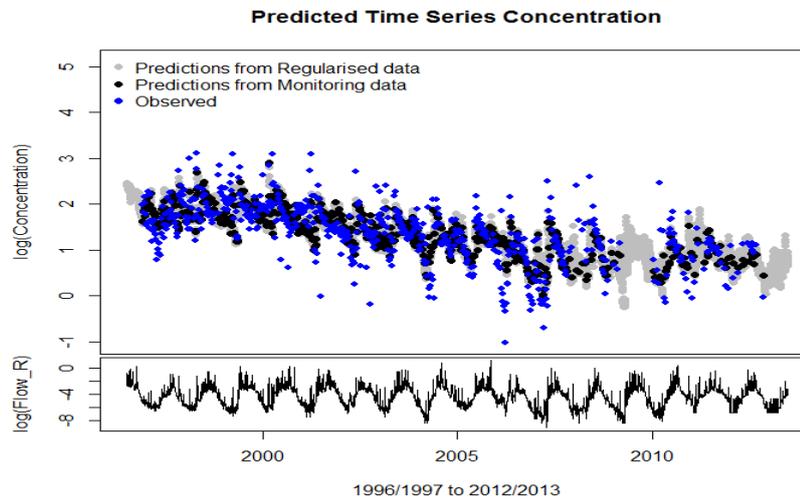


(d)

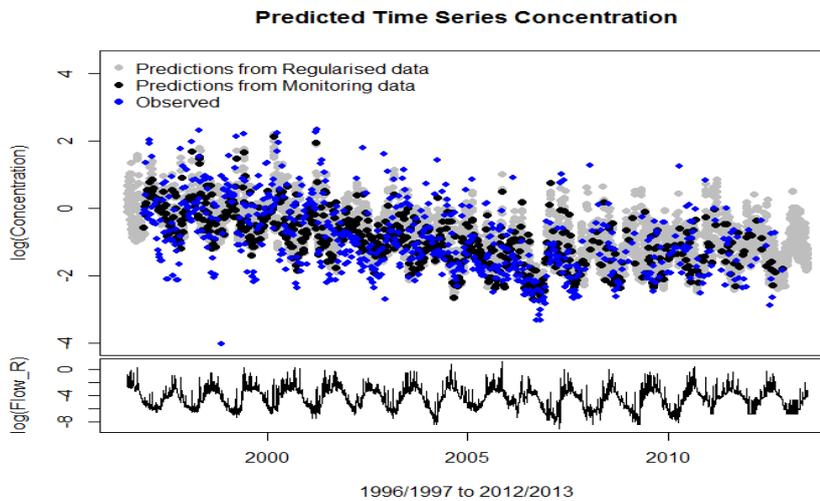
Figure 60: Smooth terms from the generalised additive model for Cox Creek (A5030526) showing the characteristics of (a) season, (b) the accumulation of flow, (c) discounted flow ( $d0.25$ ) and (d) discounted flow ( $d0.75$ ) in relation to TN.



(a)



(b)



(c)

Figure 61: Predictions from the generalised additive model fit to constituent data at Cox Creek Site (A5030526) for (a) TSS, (b) TN and (c) TP.

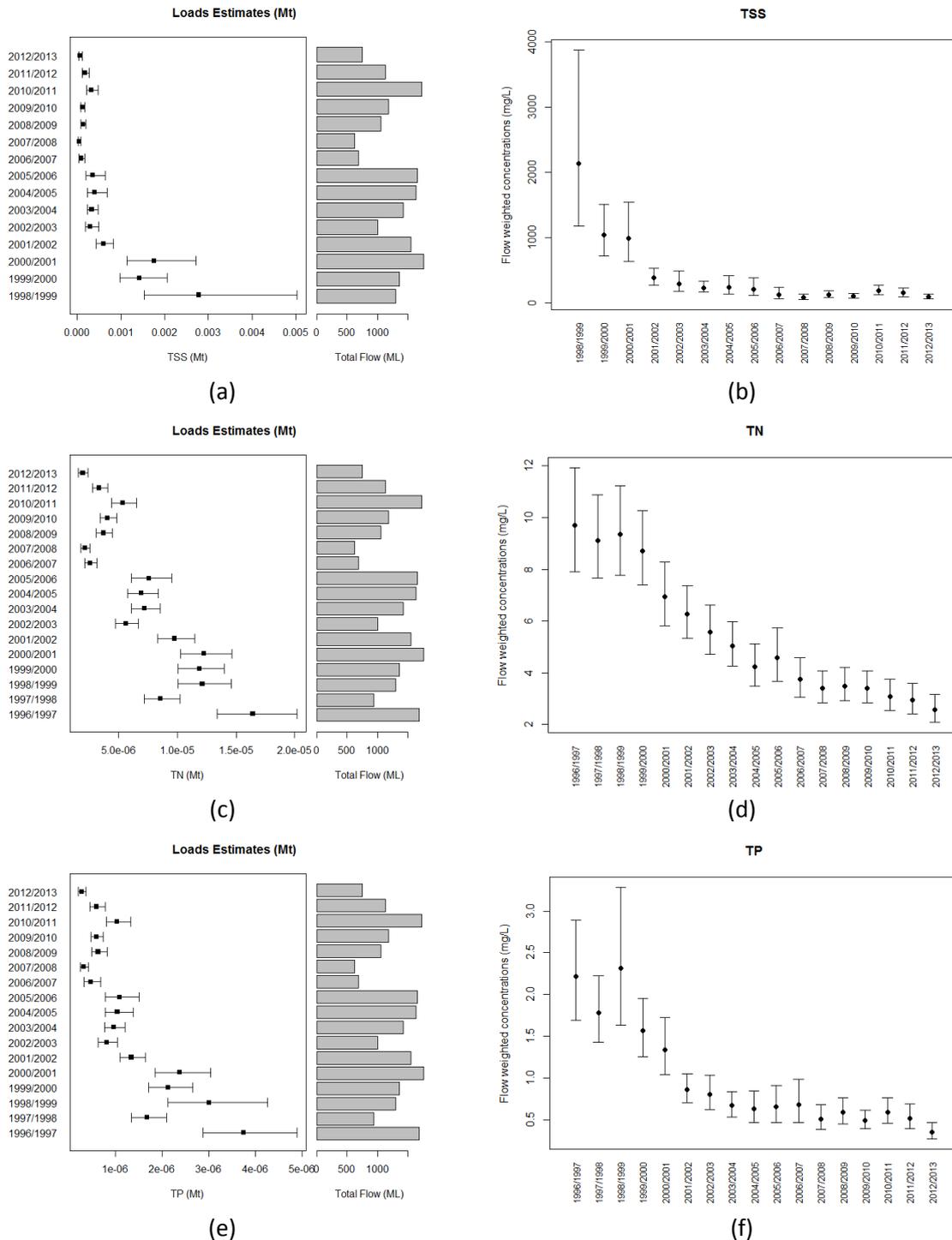


Figure 62: Estimate of the annual loads (Mt), flows (ML) and flow weighted concentrations (mg/L) for the Cox Creek Site (A5030526) accompanied by 80% confidence intervals: (a)-(b) TSS, (c)-(d) TN and (e)-(f) TP.

## Activity 3: Spatio-Temporal Models to Investigate Scenarios

### Motivation

As discussed in the introduction, the third modelling activity that was carried out for this report was the evaluation of land-use change scenarios to investigate whether there are changes in loads. Three scenarios were investigated for the Onkaparinga catchment in an attempt to determine how the loads in TSS, TN and TP might change and whether any changes identified could be considered significant. Due to the time frame of this project, only simple scenarios involving land-use were considered, where the data required could be readily incorporated into a statistical analysis. In constructing these scenarios for the statistical analysis, it is also envisaged that they be trialled in the dynamic Sednet model application when finalised and made available. The scenarios were developed by SA Water in consultation with CSIRO to ensure they could be examined and tested through the statistical modelling approaches investigated. As such, the scenarios consisted of

1. Investigating the sale of SA Water land holdings in Scott Creek sub-catchment
2. Quantifying the impact of continued expansion of perennial horticulture in the Cox Creek sub-catchment.
3. Quantifying the impact on water quality of infill within township boundaries of Aldgate Creek Railway Station.

Each of these scenarios will be discussed in more detail in the sections below.

Land-use summaries upstream from each catchment monitoring site is summarised in Appendix C in terms of a more fine scaled water quality functional unit (FU) and broad scale FU. As well as the area in square kilometres assigned to each FU, a percent FU is also presented.

### Scenario 1: Sale of SA Water Land Holdings (Scott Creek sub-catchment)

SA Water owns land in the sub-catchment of Scott Creek comprising 4.05km<sup>2</sup> (8.69%) of the total area above gauge A5030502. This drains into the Onkaparinga River downstream of Mount Bold Reservoir but upstream of Clarendon Weir. As a result, the quality of water transferred to Happy Valley Reservoir is potentially directly influenced by changes in the quality of water sourced from this catchment. An investigation of the conversion of this proportion of the land-use to alternatives may influence the decision to retain or sell and is therefore of interest as a scenario to SA Water. Initially we investigated converting this proportion of land-use to alternative land-uses but after some discussion with SA Water, the scenario involved taking half of the land set aside for conservation use at Scott Creek and transferring this across to an urban land-use. We transferred 22.85% of the total land-use in the sub-catchment to the “Dense Urban” land-use and investigated the impact on to the loads. See Appendix C for specific details. We also generated an updated flow record for the Aldgate Creek sub-catchment using SIMHYD to reflect a change in land-use from non-urban to urban and the corresponding decrease in the pervious fraction of the catchment.

*Table 28: Land-use changes imposed in scenario 1.*

Land-use	From	To
Conservation	45.7%	22.85%
Dense Urban	1.18%	24.65%

## Scenario 2: Impact of Continued expansion of Perennial Horticulture (Cox Creek sub-catchment)

Recent decades has seen the progressive expansion of perennial horticulture at the expense of annual intensive horticulture and grazing. The majority of this expansion is through the establishment of vineyards. We investigated for each constituent a change in perennial horticulture at the Cox Creek sub-catchment by halving the existing percentage under this functional unit and transferring this across to annual horticulture. This resulted in a change to the following functional unit percentages for Cox Creek:

*Table 29: Land-use changes imposed in scenario 2.*

Land-use	From	To
Broad-scale annual horticulture	28.01%	14.01%
Broad-scale perennial horticulture	23.14%	37.15%

For each model constructed under this scenario, we included the suite of covariates outlined in Appendix A in addition to land-use and specifically, included broad-scale annual horticulture and broad-scale perennial horticulture as predictors in the model. Note, only these two land-use functional units were included in the spatio-temporal models for the purpose of investigating an increase in perennial horticulture practices from annual horticulture practices.

## Scenario 3: Impact on Water Quality of Infill within Township Boundaries (Aldgate Creek sub-catchment)

This scenario investigates a major change in land-use from non-urban uses to urban uses for the Aldgate Creek sub-catchment. This results in a change in land-use as described below:

*Table 30: Land-use changes imposed in scenario 3.*

Land-use	From	To
Non-Urban	11.90%	0.00%
Urban	88.75%	99.84
Water	0.08%	0.08%

For each model constructed under this scenario, we included the suite of covariates outlined in Appendix A in addition to the broad land-use categories outlined in Appendix C. We also generated an updated flow record for the Aldgate Creek sub-catchment using SIMHYD to reflect a change in land-use from non-urban to urban and the corresponding decrease in the pervious fraction of the catchment.

## Overview of Statistical Modelling Framework

We chose to adopt a non-parametric approach to modelling the six sites in the Onkaparinga catchment as this facilitated investigation of multiple interactions between variables and sites that were difficult to fit using a GAM framework. Random Forests (Breiman, 2001) is a statistical approach that constructs an ensemble of decision trees (either regression or classification based), which when averaged, lead to more accurate predictions. Trees are created using a process known as “bagging” or “bootstrap aggregation” of the training data and then at each splitting point in a decision tree, random feature selection (i.e. a random selection of a subset of the predictors) is used, yielding an overall model with the ability to identify the most important predictors but also with the lowest predictive error rate. Predictions from the model are obtained by averaging across many trees, typically 500. Random Forests are seen as being superior to individual decision trees, because they do not overfit to the training set, which results in superior predictive capability when applied to new data. The way in which Random Forests are created, makes the interpretability of the model difficult, partial dependence plots can be constructed from the model to investigate the marginal contributions of each variable. Partial dependence plots can be likened to a smoothing spline from a GAM, however, the representations are not smooth due to the non-parametric nature of the model. They do provide an indication of how a variable relates to the prediction formed from the model (i.e. whether this relationship is linear or non-linear) that can be beneficial for interpretation. For a detailed mathematical explanation of the Random Forest methodology, we refer the interested reader to (Breiman, 2001). The analyses implemented in this activity were performed using the “randomForest” package for R ([cran.r-project.org](http://cran.r-project.org)).

We explored each of the scenarios below using the Random Forest modelling framework and examined predictions from the best model with 80% confidence intervals where a change in land-use was of interest. Specifically, these scenario investigations involved:

1. Developing a Random Forest model for the constituent of interest using hydrological and spatial (land-use) data;
2. Predicting the constituent and quantifying the load and flow weighted concentrations with 80% confidence intervals for the original data; and
3. Predicting the constituent for a change in land-use based on one of three scenarios examined.

Where confidence intervals overlapped between the scenario predictions and the predictions from the original data, we loosely concluded that there was no significant difference between the predictions. Where the confidence intervals did not overlap, we note the differences in load estimates quantified for the original data when compared with the scenario investigated. Note, as the scenarios need some further clarification with regards to changes in land-uses, we suggest that the scenarios be formally finalised prior to formal testing for these graphical differences for each financial year using a non-parametric paired test such as the Wilcoxon signed rank test (Bauer, 1972).

Diagnostic plots (not presented here due to space limitations) were examined to assess the performance of each model. These included standard residual plots, plots of fitted values versus actual observations, autocorrelation functions to examine temporal dependence and variograms produced from the residuals for each fitted model to examine spatial dependence. For each scenario investigated, the Random Forest model that included the suite of covariates and land-use variables outlined in Appendix A accommodated spatial and temporal dependence. Furthermore, the standard diagnostic plots of residuals and fitted values did not exhibit any unusual predictions or residuals that were worthwhile exploring.

## Scenario 1 Results

For each model constructed under this scenario, we included the suite of covariates outlined in Appendix A in addition to the land-use categories outlined in Appendix C. We also generated an updated flow record for the Scott Creek sub-catchment to reflect a change in land-use from non-urban to urban. The new flow series was generated by running SIMHYD with the parameters (posterior means) provided in Table 4, but with a pervious fraction of 0.7535 instead of 0.9882 to reflect the changes in land-use.

### TSS

A Random Forest model consisting of 500 trees was fit to the TSS measurements collected for the 6 sites investigated. This model resulted in a mean squared error of 0.889 and a percent variance explained of approximately 58%. The variable importance ranking based on the suite of covariates included in the model is shown in Figure 63 for two variable importance criterion, the mean square error and node purity when each variable is excluded from the model. Variables are plotted in order of importance with the most important variables shown at the top of each figure. It is clear from this figure that the most important variables for predicting TSS across the six sites are the hydrological variables with the past summation of flow (csQ) being the most important. Land-use in the form of broad-scale annual horticulture and broad-scale perennial horticulture ranked low in comparison.

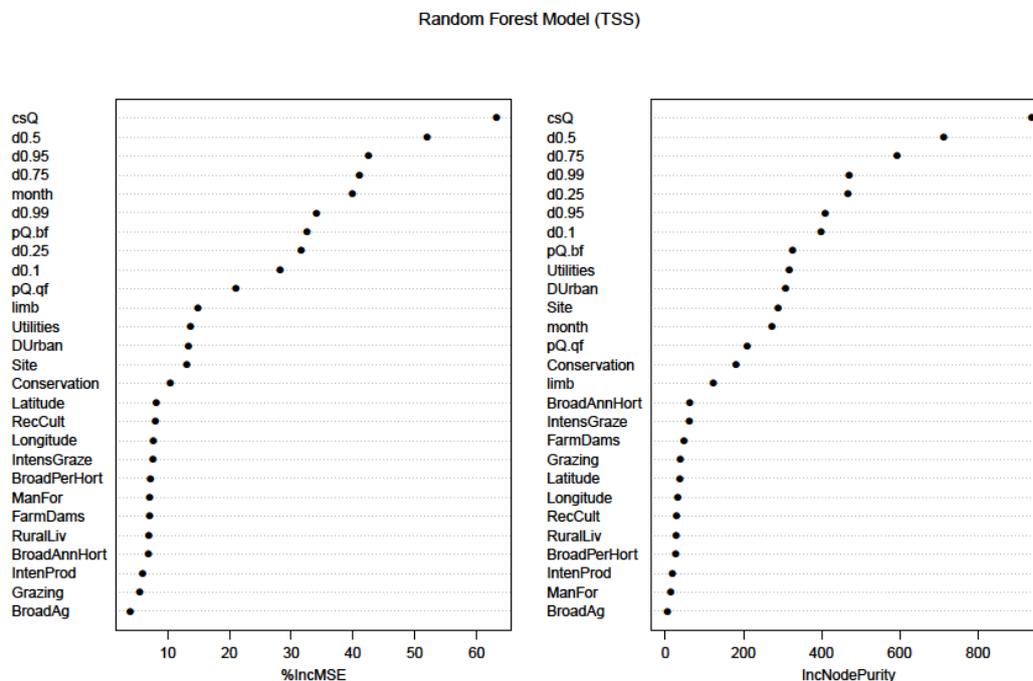


Figure 63: Variable importance ranking for TSS based on two criteria (left) Percent MSE and (right) node purity.

Figures 64 and 65 compare the TSS loads and flow weighted concentrations for the Scott Creek site under the scenario investigated. Estimates based on the original data are shown in black (with 80%

confidence intervals) while those based on the scenario examined are shown in blue. It is clear across all financial years that the model predicts a slight increase in the loads and flow-weighted concentrations under the scenario of increased urbanisation. It should be noted that there is also substantial uncertainty around the predictions made from the scenario modelling compared to the original data. Given that the 80% intervals for the data and the scenario show substantial overlap, we conclude that there is not strong evidence to suggest that the increased level of urbanisation would increase the modelled quantities, but that there is also substantial uncertainty around how much TSS loads might increase by.

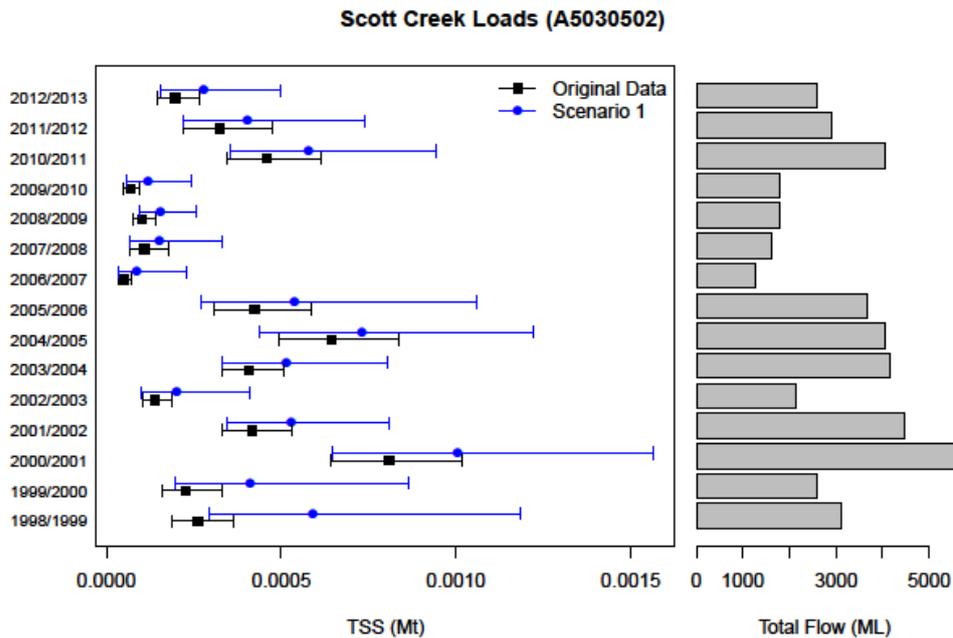


Figure 64: Estimates of loads and total flow for each financial year for Scott Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue.

**Scott Creek (A5030502) – Average Concentrations**

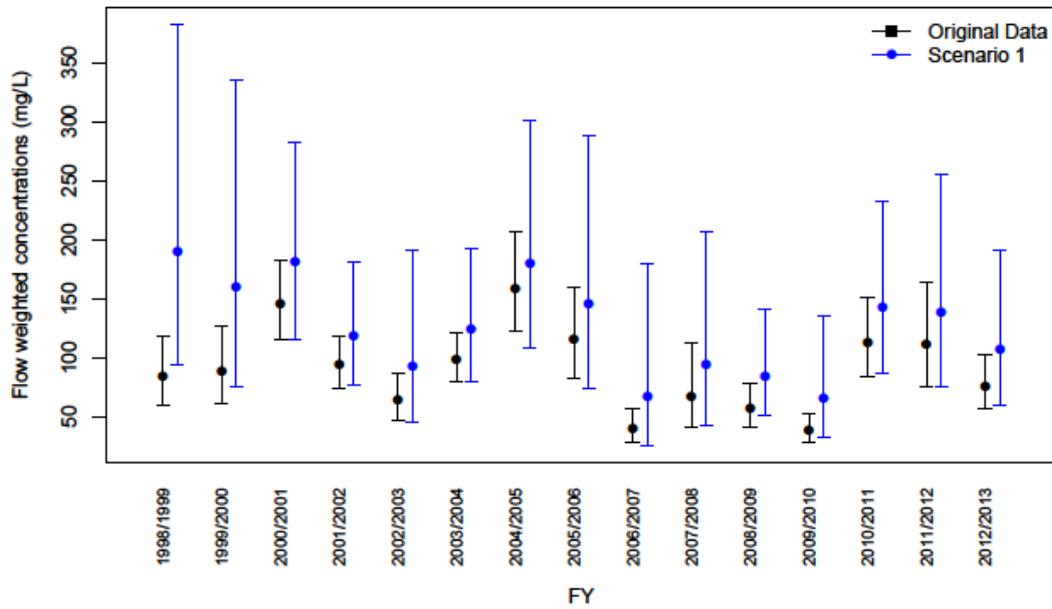


Figure 65: Estimates of the flow weighted concentrations for Scott Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue.

**TN**

A Random Forest model was fit to total nitrogen data collected at the six sites in the Onkaparinga catchment. The results that follow are based on 500 trees created on bootstrap samples of the data with a random feature selection. The mean square error of residuals resulted in 0.153, while the percent variation explained was nearly 85%.

Random Forest Model (TN)

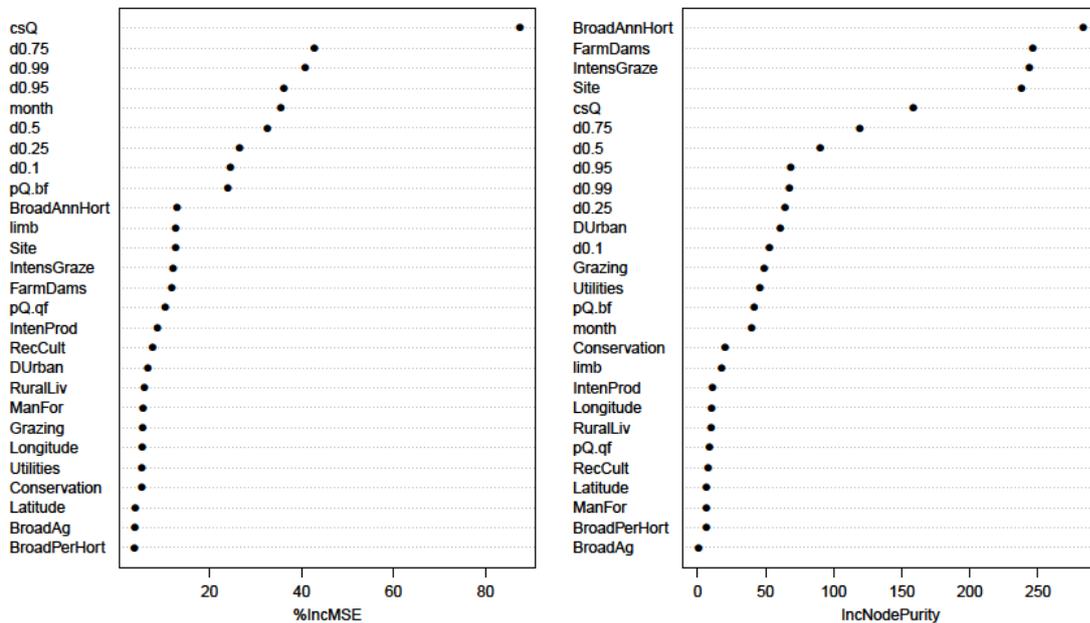


Figure 66: Variable importance ranking for TN based on two criteria (left) Percent MSE and (right) node purity.

The variable importance ranking is shown in Figure 66 for the two criteria examined and highlights the importance of land-use for TN, particularly when node purity is used as the criterion. This importance is reflected in the TN loads estimates for Scott Creek that is shown in Figures 67 and 68 where we compared the loads predicted from the Random Forest model with an increase in urbanisation. In both of these plots we can see slight increases in the predicted values for all financial years under the urbanisation scenario (blue) compared to the original data (black). We also see a small increase in the width of the 80% intervals for the scenario. There is substantial overlap between the 80% intervals for the scenario and the original data for both the loads and flow-weighted concentrations. This suggested little strong evidence for an increase in the total nitrogen load and flow-weighted concentration for nitrogen under this increased level of urbanisation.

**Scott Creek Loads (A5030502)**

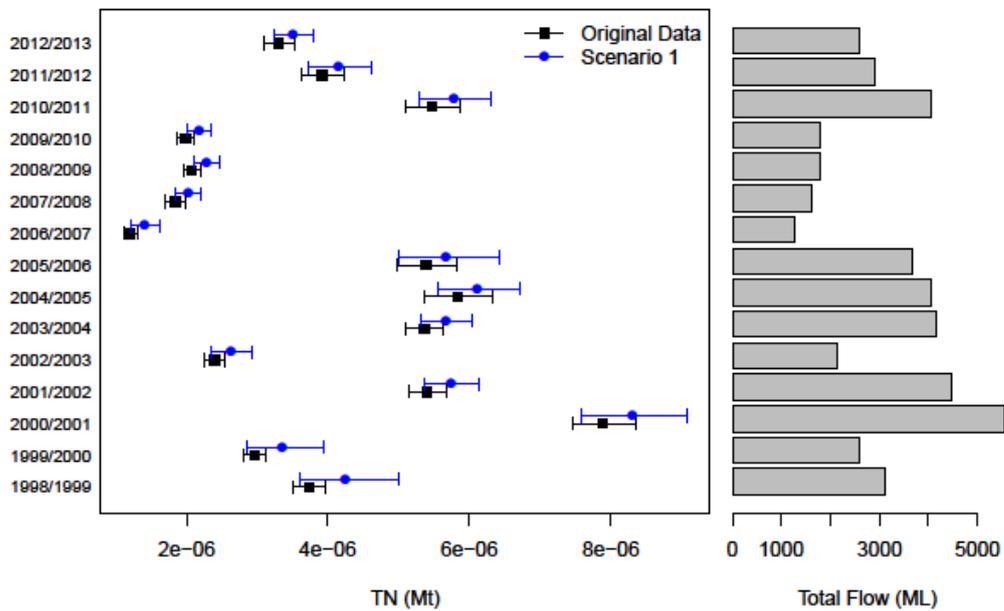


Figure 67: Estimates of TN loads and total flow for each financial year for Scott Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue.

**Scott Creek (A5030502) – Average Concentrations**

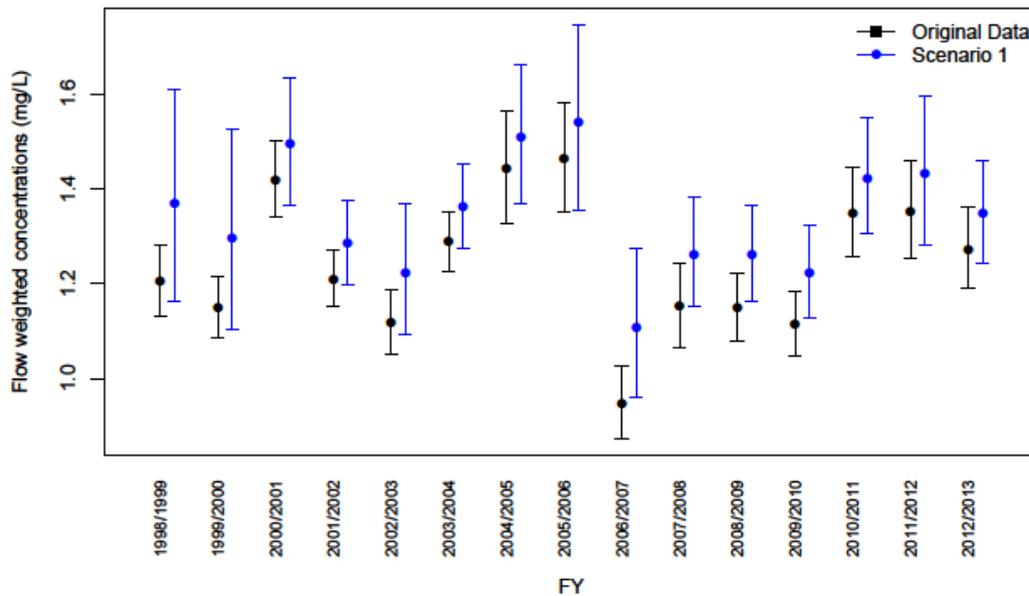


Figure 68: Estimates of the flow weighted TN concentrations for Scott Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue.

TP

The Random Forests modelling approach was used to investigate scenario 1 for TP loads and see whether increased urbanisation resulted in any significant changes. The result is displayed in the figures below, where Random Forests was used to build 500 trees to TP concentrations collected for the six sites in the Onkaparinga catchment. This model resulted in a mean square error of 0.275 and a percent variation explained of close to 78%. The variable importance ranking shown in Figure 69 also highlights land-use as being potentially important along with site and many of the discounted flow variables.

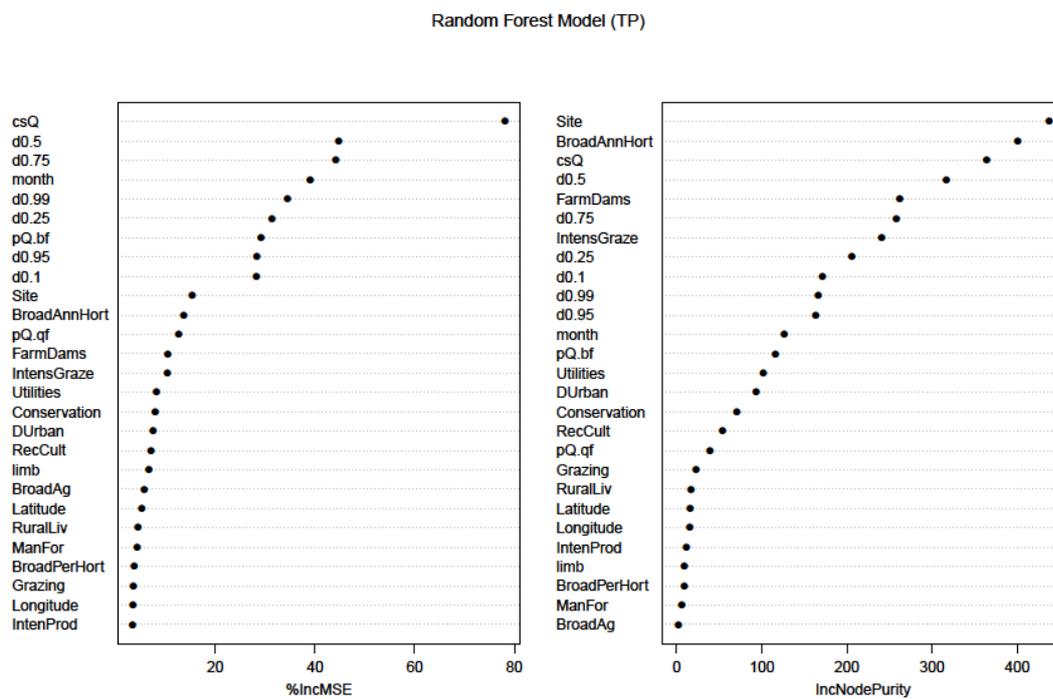


Figure 69: Variable importance ranking for TP based on two criteria (left) Percent MSE and (right) node purity.

Total phosphorous load estimates based on the Random Forest model for the original data and the scenario outlined for Cox Creek are shown in Figures 70 and 71. The model suggests a slight increase in total phosphorous might be expected under increased urbanisation, but 80% confidence intervals overlap in all financial years for both the total load and the flow-weighted concentration. There is no clear evidence to suggest an increase in these quantities under the scenario of increased urbanisation, but the modelling also highlights that there is substantial uncertainty around the predictions for the scenario relative to the uncertainty around the loads from the observed data.

**Scott Creek Loads (A5030502)**

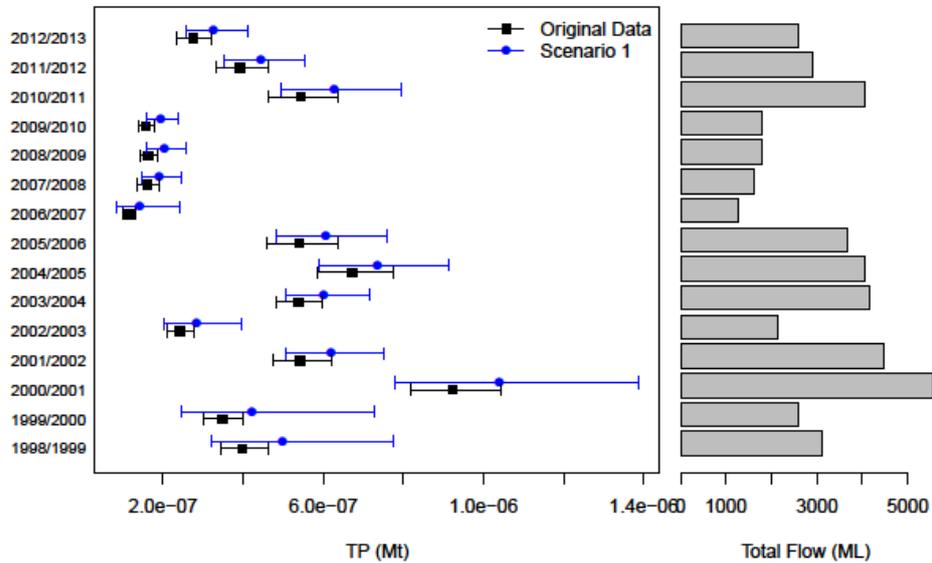


Figure 70: Estimates of TP loads and total flow for each financial year for Scott Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue.

**Scott Creek (A5030502) – Average Concentrations**

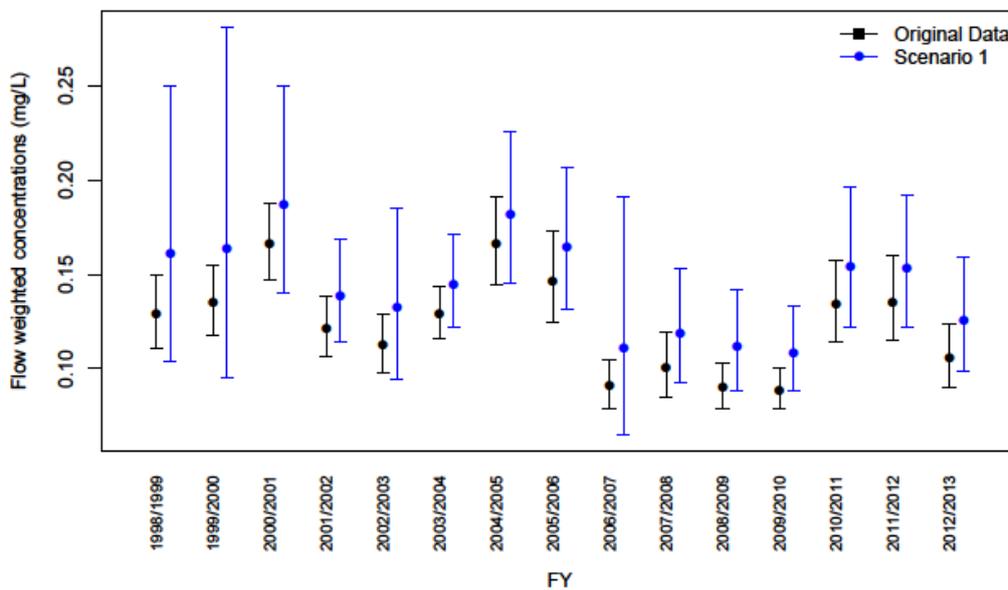


Figure 71: Estimates of the flow weighted TP concentrations for Scott Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue.

## Scenario 2 Results

### TSS

A Random Forest model consisting of 500 trees was fit to the TSS measurements collected for the 6 sites investigated. This model resulted in a mean squared error of 0.889 and a percent variance explained of approximately 58%. The variable importance ranking based on the suite of covariates included in the model is shown in Figure 72 for two variable importance criterion, the mean square error and node purity when each variable is excluded from the model. Variables are plotted in order of importance with the most important variables shown at the top of each figure. It is clear from this figure that the most important variables for predicting TSS across the six sites are the hydrological variables with the past summation of flow (csQ) being the most important. Land-use in the form of broad-scale annual horticulture and broad-scale perennial horticulture ranked low in comparison.

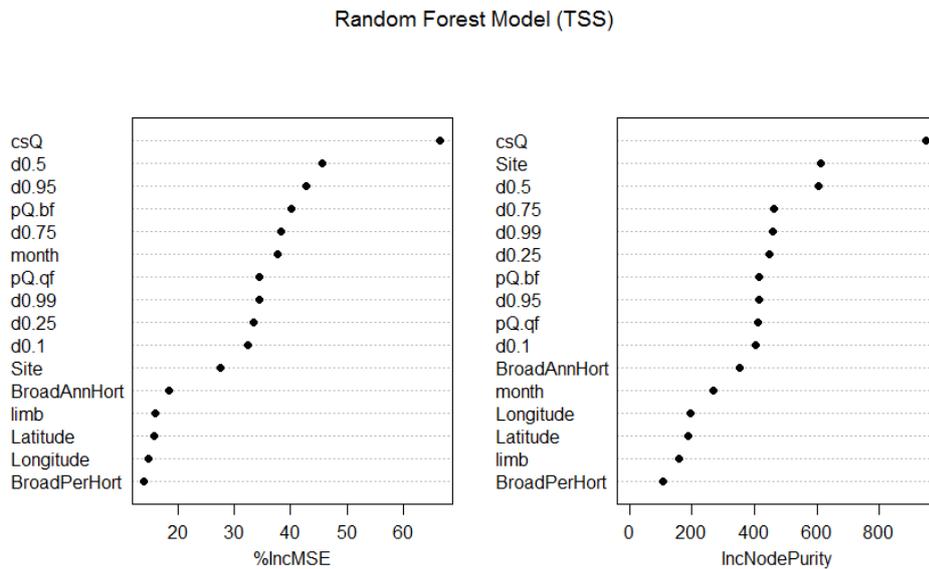


Figure 72: Variable importance ranking for TSS based on two criteria (left) Percent MSE and (right) node purity.

Figures 73 and 74 compare the TSS loads and flow weighted concentrations for the Cox Creek site under the scenario investigated. Estimates based on the original data are shown in black (with 80% confidence intervals) while those based on the scenario examined are shown in blue. It is clear across all financial years that there is little difference between the original TSS load estimates and those estimated based on the scenario as the 80% intervals overlap. This is also evident for the flow-weighted concentrations.

### Cox Creek Loads (A5030526)

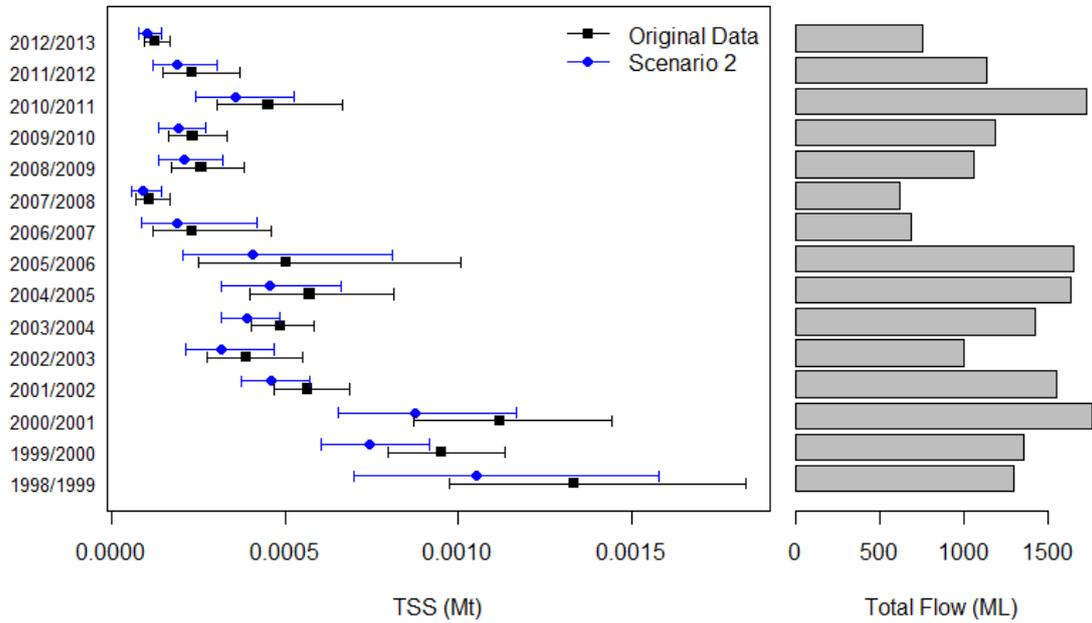


Figure 73: Estimates of loads and total flow for each financial year for Cox Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue.

### Cox Creek (A5030526) - Average Concentrations

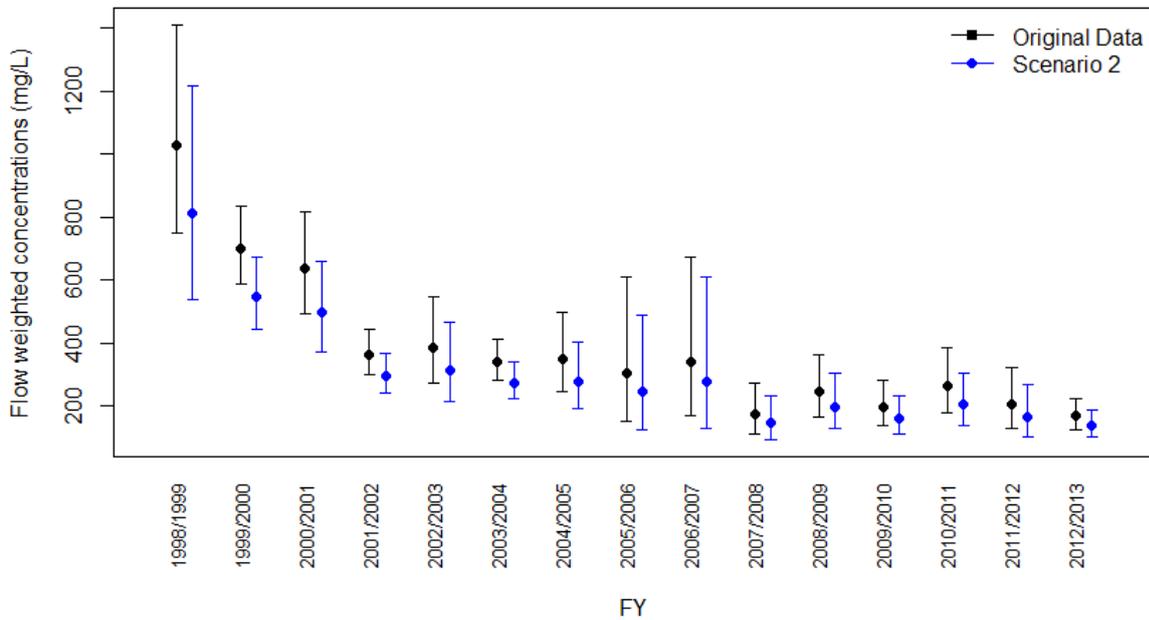


Figure 74: Estimates of the flow weighted concentrations for Cox Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue.

TN

A Random Forest model was fit to total nitrogen data collected at the six sites in the Onkaparinga catchment. The results that follow are based on 500 trees created on bootstrap samples of the data with a random feature selection. The mean square error of residuals resulted in 0.153, while the percent variation explained was nearly 85%.

Random Forest Model (TN)

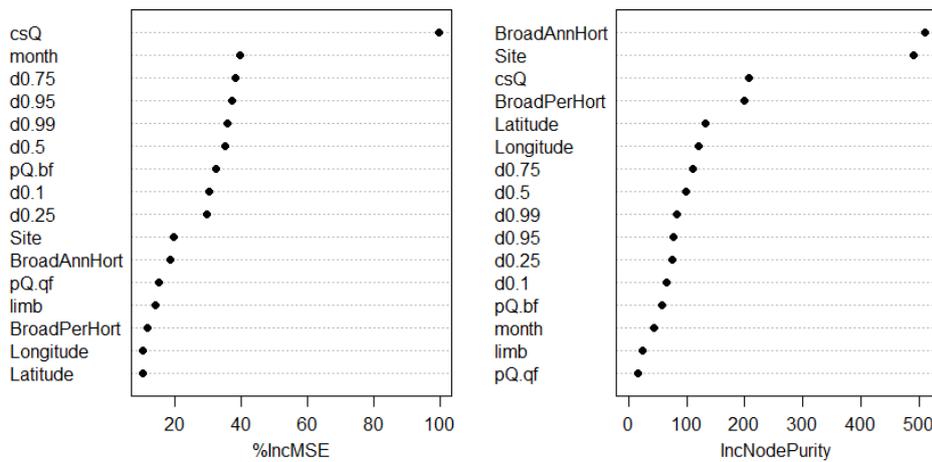


Figure 75: Variable importance ranking for TN based on two criteria (left) Percent MSE and (right) node purity.

The variable importance ranking is shown in Figure 75 for the two criteria examined and highlights the importance of land-use for TN, particularly when node purity is used as the criterion. This importance is reflected in the TN loads estimates for Cox Creek that is shown in Figures 76 and 77 where we compared the loads predicted from the Random Forest model with a change in land-use as we can see distinct differences between the original data (black) and the scenario (blue), where the scenario highlights a decrease in loads when annual horticulture is changed to perennial horticulture.

### Cox Creek Loads (A5030526)

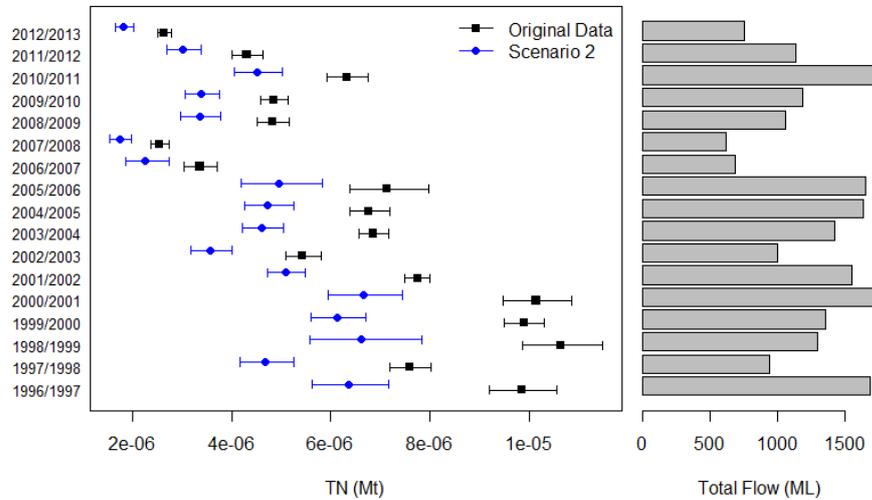


Figure 76: Estimates of TN loads and total flow for each financial year for Cox Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue.

### Cox Creek (A5030526) - Average Concentrations

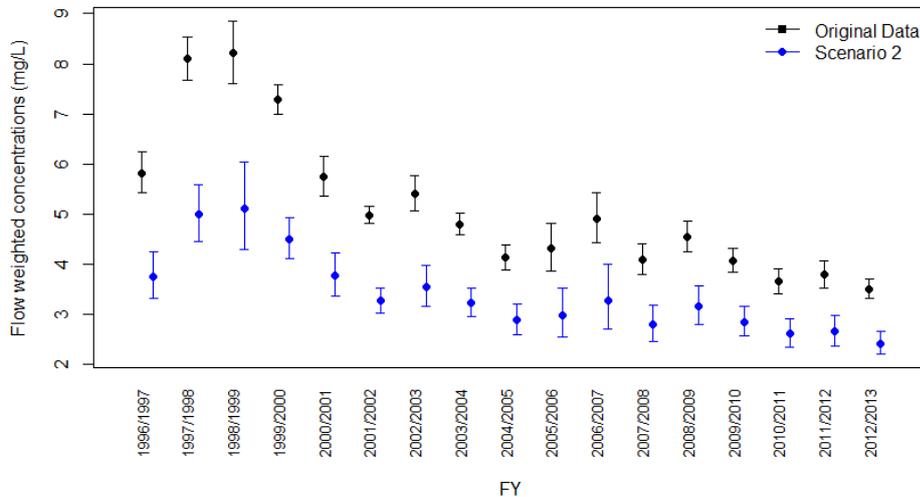


Figure 77: Estimates of the flow weighted TN concentrations for Cox Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue.

## TP

The Random Forests modelling approach was used to investigate scenario 2 for TP loads and see whether a change in horticulture practices resulted in a change in loads. The result is displayed in the figures below, where Random Forests was used to build 500 trees to TP concentrations collected for the six sites in the Onkaparinga catchment. This model resulted in a mean square error of 0.275 and a percent variation explained of close to 78%. The variable importance ranking shown in Figure 78 also highlights land-use as being potentially important along with site and many of the discounted flow variables.

Random Forest Model (TP)

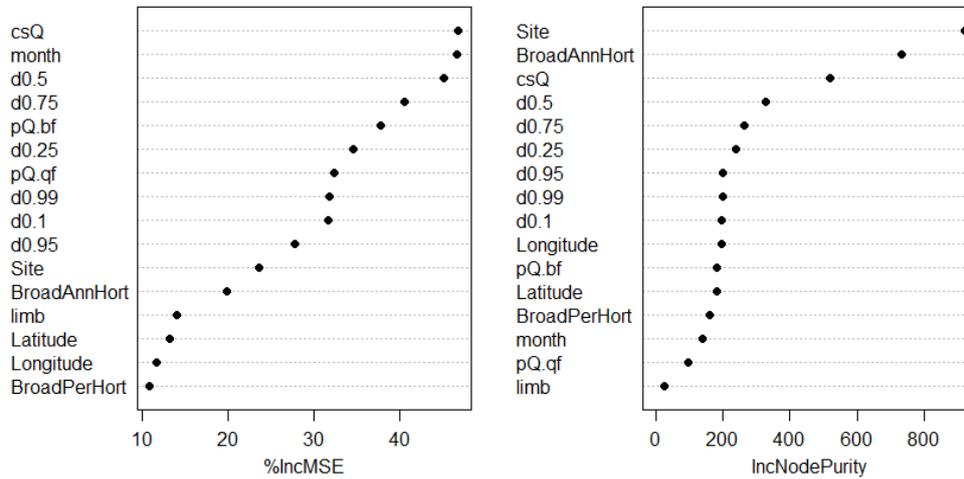


Figure 78: Variable importance ranking for TP based on two criteria (left) Percent MSE and (right) node purity.

Total Phosphorous load estimates based on the Random Forest model for the original data and the scenario outlined for Cox Creek are shown in Figures 79 and 80. For some financial years there is some evidence supporting a decline in the load and flow-weighted concentrations (based on non-overlapping 80% confidence intervals) under the scenario of increased perennial horticulture.

Cox Creek Loads (A5030526)

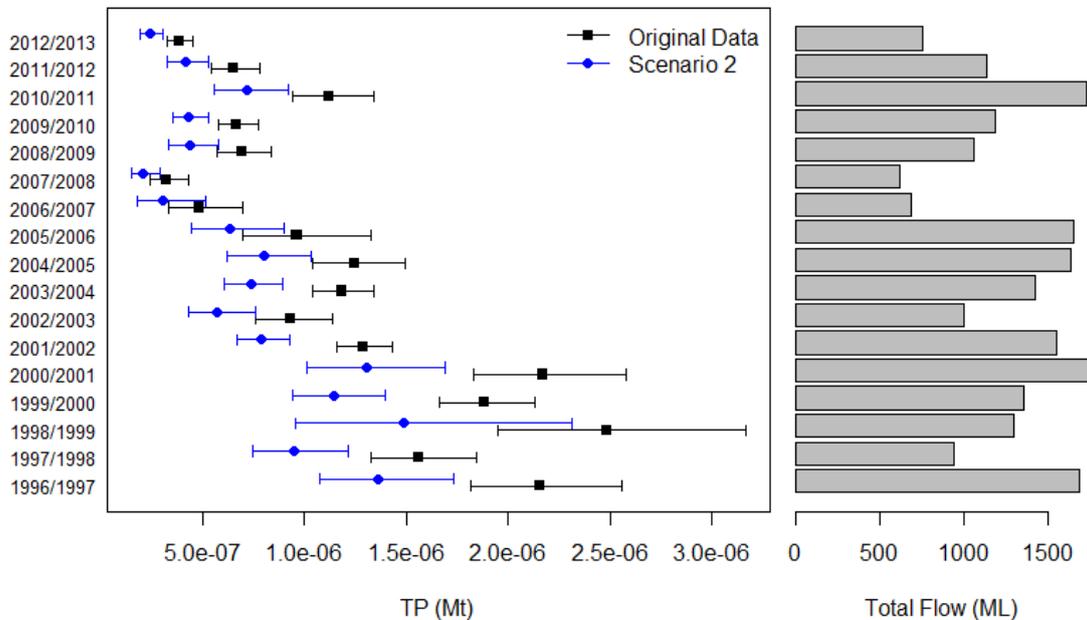


Figure 79: Estimates of TP loads and total flow for each financial year for Cox Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue.

### Cox Creek (A5030526) - Average Concentrations

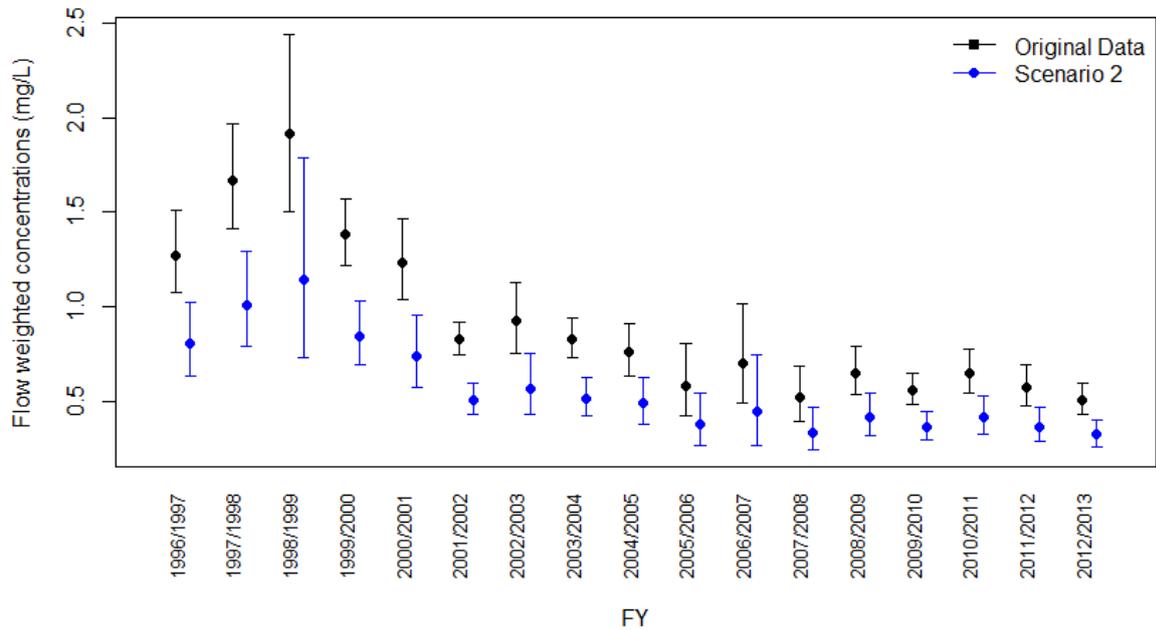


Figure 80: Estimates of the flow weighted TP concentrations for Cox Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue.

### Scenario 3 Results

For each model constructed under this scenario, we included the suite of covariates outlined in Appendix A in addition to the land-use categories outlined in Appendix C. We also generated an updated flow record for the Aldgate Creek sub-catchment to reflect increased levels of urbanisation. The new flow series was generated by running SIMHYD with the parameters shown in Table 4 (posterior means), but with a pervious fraction of 0.3372 instead of 0.4562 to reflect the changes in land-use.

### TSS

A Random Forest model consisting of 500 trees was fit to the TSS measurements collected for the 6 sites investigated. This model resulted in a mean squared error of 0.889 and a percent variance explained of approximately 58%. The variable importance ranking based on the suite of covariates included in the model is shown in Figure 81 for two variable importance criterion, the mean square error and node purity when each variable is excluded from the model. Variables are plotted in order of importance with the most important variables shown at the top of each figure. It is clear from this figure that the most important variables for predicting TSS across the six sites are the hydrological variables with the past summation of flow (csQ) being the most important. Variables such as the percent of urban and non-urban land-use in the catchment ranked low in comparison, suggesting these had minor explanatory power.

Random Forest Model (TSS)

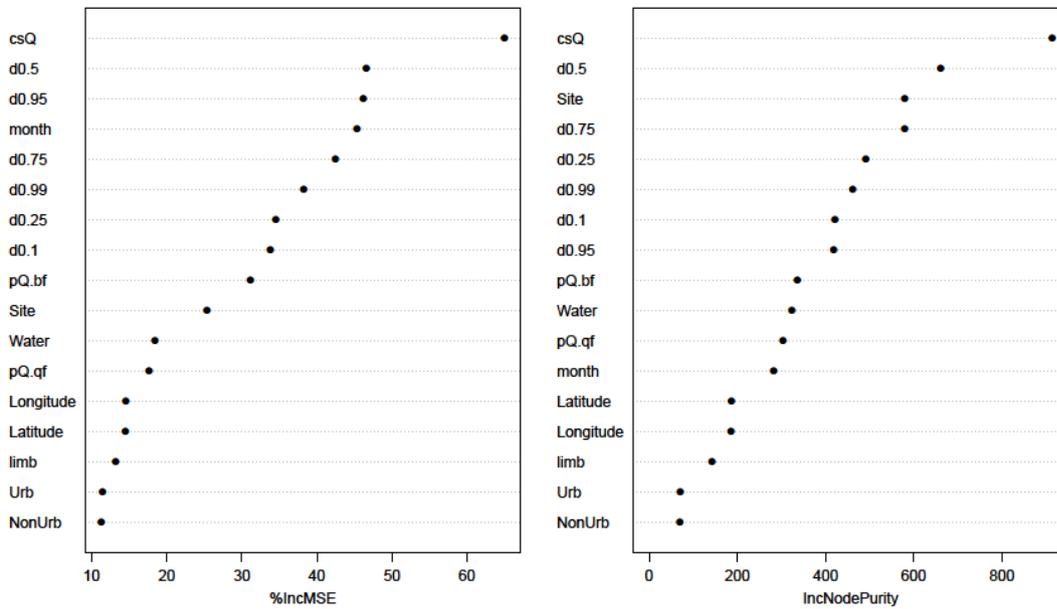


Figure 81: Variable importance ranking for TSS based on two criteria (left) Percent MSE and (right) node purity.

Figures 82 and 83 compare the TSS loads and flow weighted concentrations for the Aldgate Creek site under the scenario investigated. Estimates based on the original data are shown in black (with 80% confidence intervals) while those based on the scenario examined are shown in blue. It is clear across all financial years that the model predicts a negligible change in the loads and flow-weighted concentrations under the scenario of increased urbanisation. It should be noted that there is also substantial uncertainty around the predictions made from the original data and from the scenario modelling. Given that the 80% intervals for the data and the scenario show substantial overlap, this analysis provides no strong evidence to suggest that the increased level of urbanisation would increase the modelled quantities.

**Aldgate Creek Loads (A5030509)**

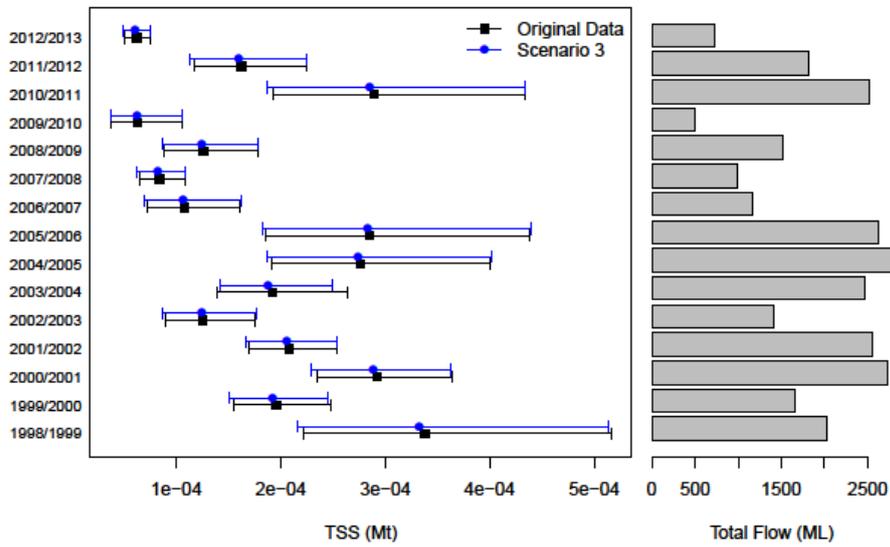


Figure 82: Estimates of loads and total flow for each financial year for Aldgate Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue.

**Aldgate Creek (A5030509) – Average Concentrations**

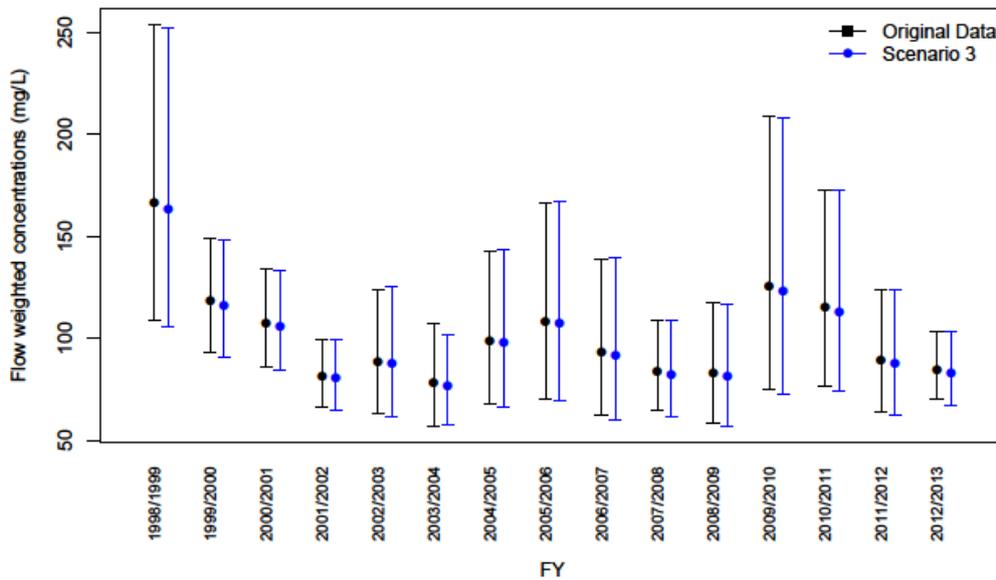


Figure 83: Estimates of the flow weighted concentrations for Aldgate Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue.

## TN

A Random Forest model was fit to total nitrogen data collected at the six sites in the Onkaparinga catchment. The results that follow are based on 500 trees created on bootstrap samples of the data with a random feature selection. The mean square error of residuals resulted in 0.153, while the percent variation explained was nearly 85%.

The variable importance ranking is shown in Figure 84 for the two criteria examined and highlights the importance of land-use for TN, particularly when node purity is used as the criterion. This importance is reflected in the TN loads estimates for Aldgate Creek that is shown in Figures 85 and 86 where we compared the loads predicted from the Random Forest model with an increase in urbanisation. In both of these plots we can see decreases in the predicted values for all financial years under the urbanisation scenario (blue) compared to the original data (black). The widths of the 80% intervals for the scenario and the observed data are almost identical. There is some overlap between the 80% intervals for the scenario and the original data for both the loads and flow-weighted concentrations over most years, but the overlap is not substantial. This provided some preliminary evidence that there may be a change in the nitrogen load and flow-weighted concentration for nitrogen under the increased level of urbanisation in this scenario.

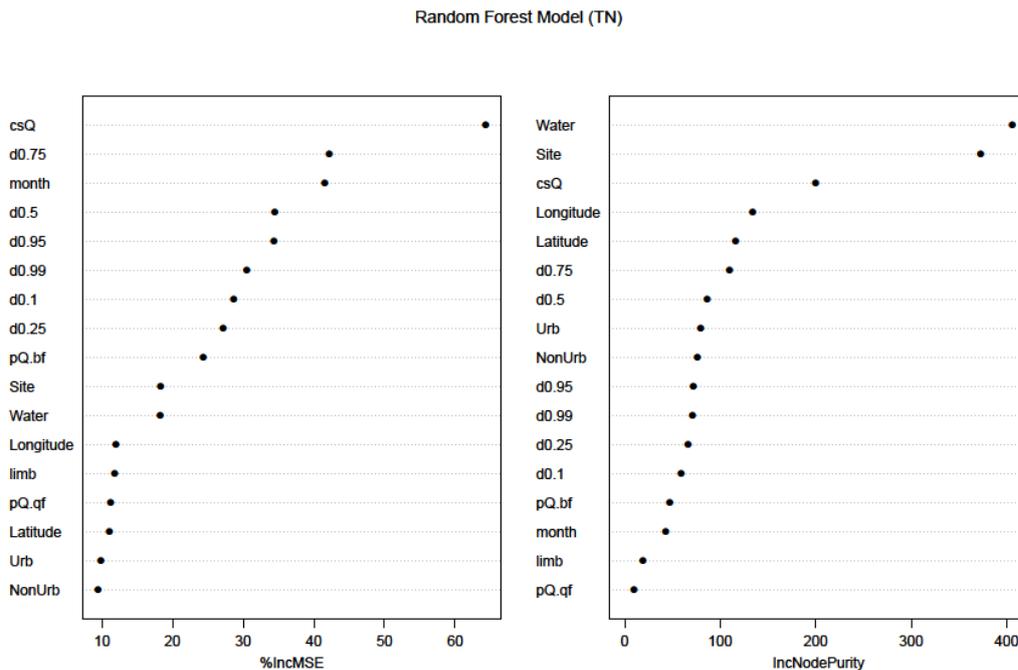


Figure 84: Variable importance ranking for TN based on two criteria (left) Percent MSE and (right) node purity.

### Aldgate Creek Loads (A5030509)

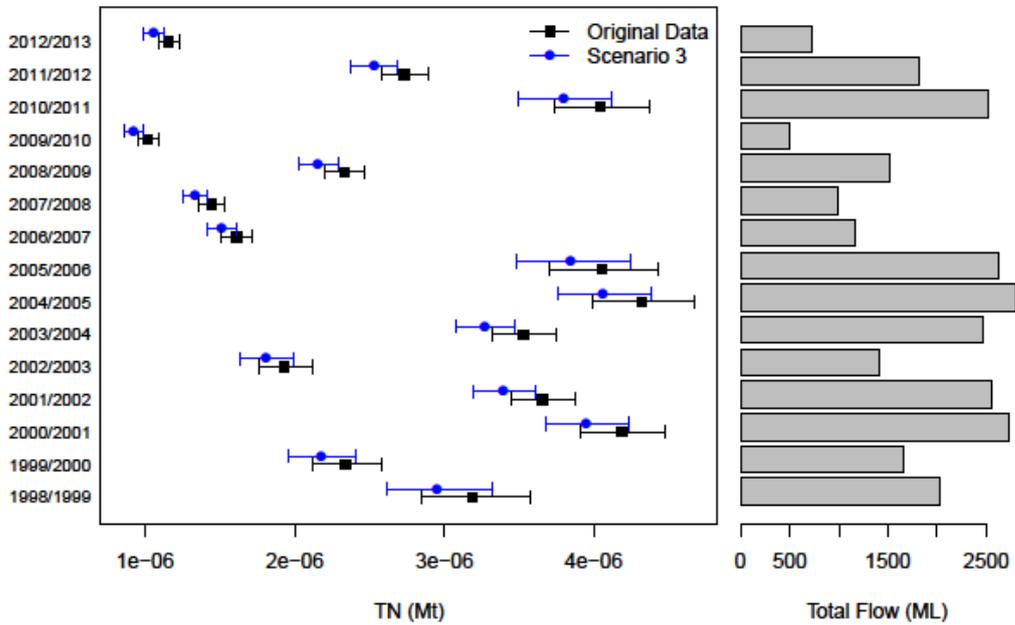


Figure 85: Estimates of TN loads and total flow for each financial year for Cox Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue.

### Aldgate Creek (A5030509) – Average Concentrations

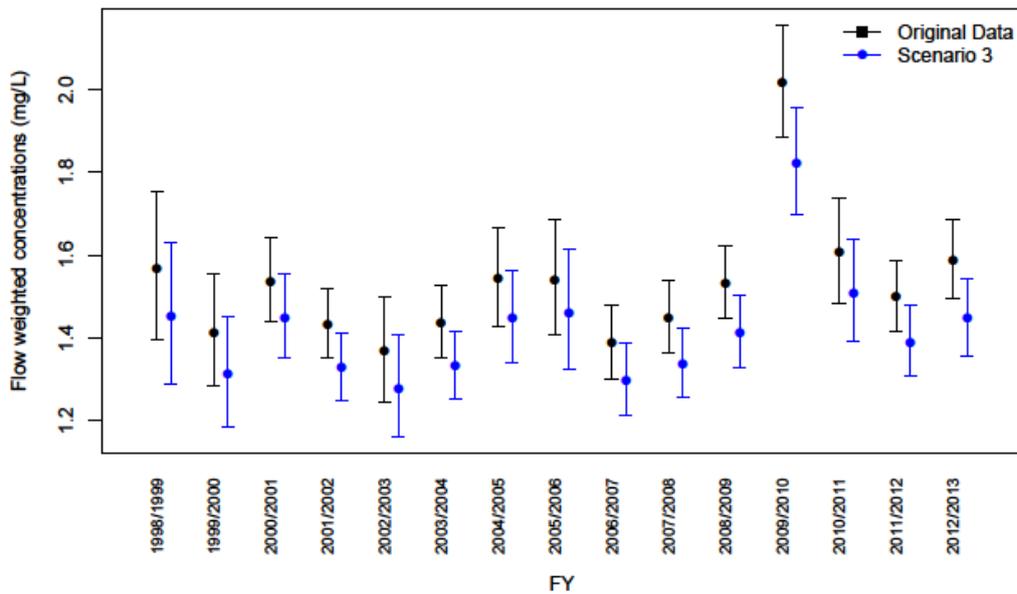


Figure 86: Estimates of the flow weighted TN concentrations for Cox Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue.

TP

The Random Forests modelling approach was used to investigate scenario 3 for TP loads and to see whether increased urbanisation resulted in any changes. The result is displayed in the figures below, where Random Forests was used to build 500 trees to TP concentrations collected for the six sites in the Onkaparinga catchment. This model resulted in a mean square error of 0.275 and a percent variation explained of close to 78%. The variable importance ranking shown in Figure 87 also highlights land-use as being potentially important along with site and many of the discounted flow variables.

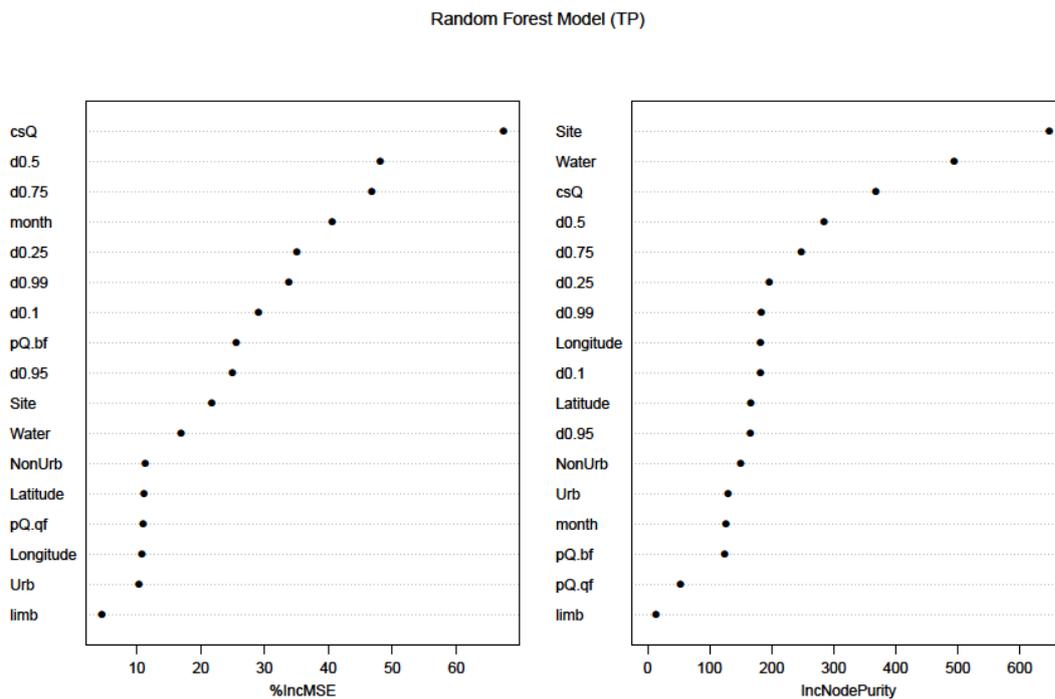


Figure 87: Variable importance ranking for TP based on two criterion (left) Percent MSE and (right) node purity.

Total phosphorous load estimates based on the Random Forest model for the original data and the scenario outlined for Aldgate Creek are shown in Figures 88 and 89. The model tends to predict very slight increases in total phosphorous loads might be expected under increased urbanisation, but such changes are very small relative to the width of the 80% confidence intervals and these intervals overlap in all financial years for both the total load and the flow-weighted concentration. We therefore conclude that there is no strong evidence to suggest a change in these quantities under the scenario of increased urbanisation for Aldgate Creek.

### Aldgate Creek Loads (A5030509)

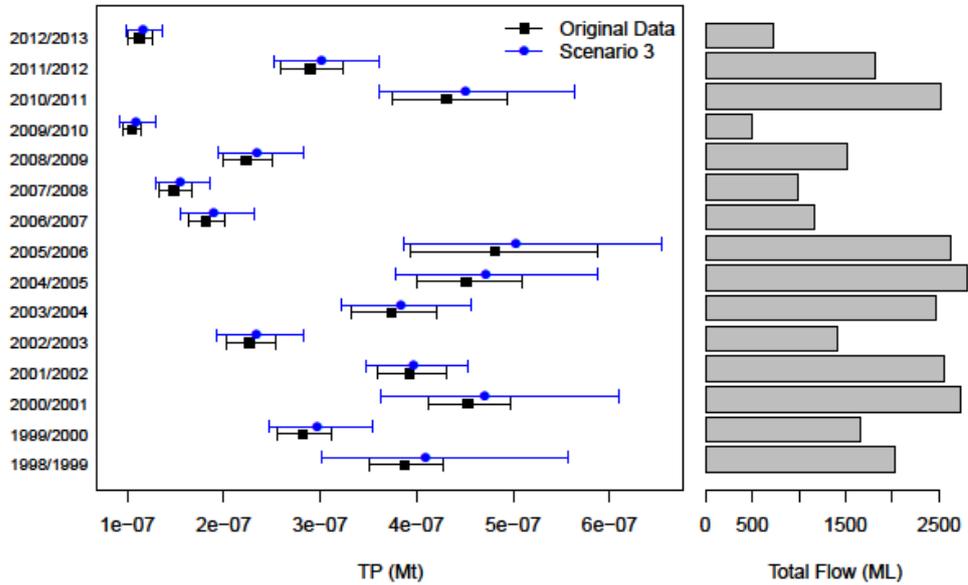


Figure 88: Estimates of TP loads and total flow for each financial year for Aldgate Creek. Estimates for the original data are shown in black while for the scenario of land-use change, estimates are shown in blue.

### Aldgate Creek (A5030509) – Average Concentrations

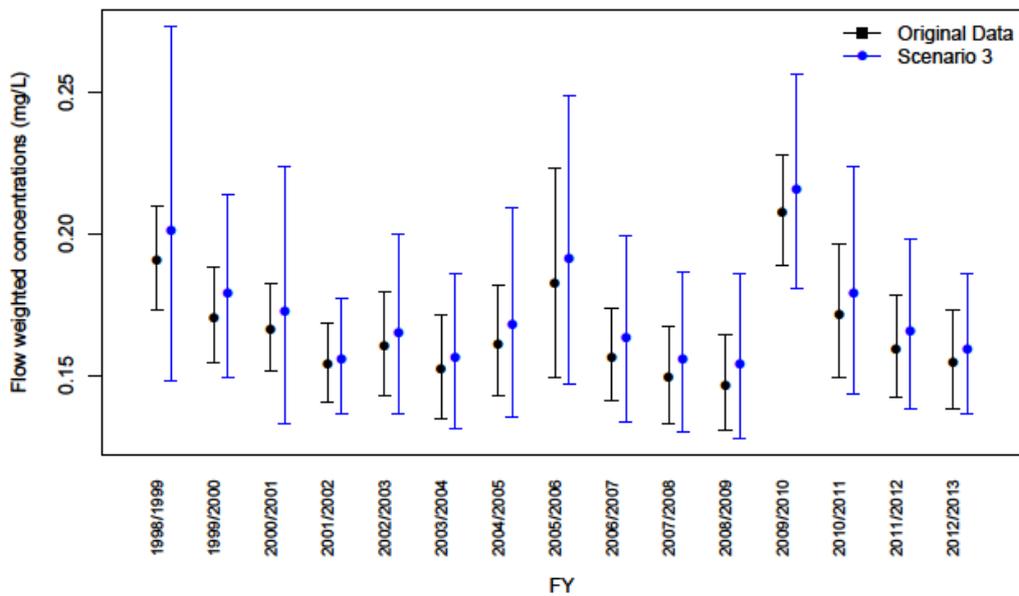


Figure 89: Estimates of the flow weighted TP concentrations for Aldgate Creek for each financial year. Estimates for the original data are shown in black while estimates for the scenario are shown in blue.

## Discussion and Conclusions

The statistical analyses presented in this report offer an initial insight into the water quality data collected for the Onkaparinga catchment in the Mt Lofty Ranges. For the site-based models in particular, these should not be considered as “final” as they require some detailed investigations into their interpretation and the prediction of each constituent and loads. While preliminary however, these models do provide some interesting interpretations and build on the basic constituent generation models developed previously for these sites.

The impact of the sedimentation pond at Cox Creek is an interesting analysis, highlighting some complex relationships between the operational use of the pond and flow that requires some further investigation. It is clear that the sedimentation pond has had some impact but in the current suite of models presented in this report, it has been difficult to disentangle what this relationship is in the time frame set aside, particularly since this was not a primary investigation of the project. We encourage SA Water to revisit this data and modelling exercise to tease out the impact of the sedimentation on the sediment and nutrients collected.

It is important to note that the TSS, TN and TP data collected for the six sites investigated is collected through composite sampling methods and is particularly variable. The composite sampling regime may have resulted in the model’s inability to predict well at the extremes. Sampling that focuses on actual events may improve this prediction in addition to other covariates (hydrological or otherwise) that have not been considered here. The variable nature of the samples collected is not a surprise and adds to the difficulty in modelling these complex relationships. Overall however, the GAM and GAMM models fit to the data performed quite well at explaining the variation in the data with most models explaining 50% or more of the variation in the data. In fact, a large proportion of the variation explained in these models was due to hydrological terms, suggesting that complex hydrological relationships are driving these systems.

While every effort was made to correct for any issues relating to the flow and constituent data, we suggest checking the flow and constituent data for any unusual or highly influential values as can be noted in some of the exploratory plots and predictions presented throughout the report. Flow for some sites exhibited some unusual patterns that seemed to mimic low or no flow situations. This appeared to be in the raw data provided rather the result of any “regularisation” performed as part of the LRE package that was used to implement the GAM and GAMM models. We encourage SA Water to check their data (both flow and constituent data) for any erroneous observations or outliers prior to further investigations in relation to the models presented here.

While the scenarios investigated were preliminary, the analyses presented here using the Random Forest approach demonstrated the power of a statistical framework for detecting significant changes. While these can also be investigated within the dynamic Sednet modelling framework, it should be noted that it will be difficult to determine significance since Sednet does not quantify the error in the loads that are estimated. So, whilst declines may be generated through Sednet scenarios, these may be small compared to the uncertainty surrounding these estimates, making it



difficult to make conclusions with any real degree of confidence. Throughout the duration of the project, there was considerable discussion regarding the scenarios and changes to land-use considered. The results presented here represent a demonstration of how these scenarios can be explored within a statistical framework and we suggest that the scenarios be revisited to ensure they meet the expectations of SA Water, particularly with reference to land-use changes that are realistic.

## References

- ANONYMOUS 2012. Cox Creek subcatchment barrier assessment report. Adelaide, South Australia: SA Water.
- BAINBRIDGE, Z. T., LEWIS, S. E., SMITHERS, S. G., KUHNERT, P. M., HENDERSON, B. & BRODIE, J. E. 2014. Fine suspended sediment and water budgets for a large, seasonally-dry tropical catchment: Burdekin River catchment, Queensland Australia. *Water Resources Research*, In Press.
- BAUER, D. F. 1972. Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, 67, 687-690.
- BREIMAN, L. (2001), Random Forests, *Machine Learning* 45(1), 5-32.
- CHIEW, F., PEEL, M., WESTERN, A. 2002. Application and testing of the simple rainfall-runoff model SIMHYD. In *Mathematical Models of Small Watershed Hydrology and Applications*, Singh V, Frevert D (eds). Water Resources Publications: Littleton, Colorado, USA; 335–367.
- COX, J. W., KIRKBY, C. A., CHITTLEBOROUGH, D. J., FLEMING, N. K. & SMYTHE, L. J. 2000. Mobility of phosphorus through intact soil cores collected from the Adelaide Hills, South Australia. *Australian Journal of Soil Research*, 38, 973-990.
- COX, J. W., OLIVER, D. P., FLEMING, N. K. & ANDERSON, J. S. 2011. Off-site transport of nutrients and sediment from three main land-uses in the Mt Lofty Ranges, South Australia. *Agricultural Water Management*, 106, 50-59.
- CRESSIE, N., WIKLE, C. 2011. *Statistics for Spatio-Temporal Data*. Wiley: Hoboken, NJ.
- DAVIS, J. R. & FARLEY, T. F. N. 1997. CMSS: policy analysis software for catchment managers. *Environmental Modelling and Software*, 12, 197-210.
- DOUGHERTY, W., FLEMING, N. K., COX, J. W. & CHITTLEBOROUGH, D. J. 2004. Phosphorus transfer in surface runoff from intensive pasture systems at various scales. *Journal of Environmental Quality*, 33, 1973-1988.
- FLEMING, N. K., COX, J. W., CHITTLEBOROUGH, D. J. & DYSON, P. 2001. Prediction of chemical loads and forms in overland flow from dairy pastures on duplex soils in South Australia. *Hydrological Processes*, 15, 393-405.
- FLEMING, N. K., COX, J. W., HE, Y. & THOMAS, S. 2012. Source hydrological calibration in the Mount Lofty Ranges using parameter estimation tool (PEST). eWater CRC Technical Report.
- FLEMING, N. K., COX, J. W., HE, Y., THOMAS, S. & FRIZENSCHAF, J. 2010. Analysis of Total Suspended Sediment and Total Nutrient Concentration data in the Mount Lofty Ranges to derive event mean concentrations. eWater CRC Technical Report.
- FRIZENSCHAF, J. & VIAL, H. 2012. Constructed Wetlands in Water Supply Catchments – Nutrient Mitigation to Support Reservoir Management, Unpublished Report, SA Water, Adelaide Australia.
- KIRKBY, C. A., SMYTHE, L. J., COX, J. W. & CHITTLEBOROUGH, D. J. 1997. Phosphorus movement down a toposequence from a landscape with texture-contrast soils. *Australian Journal of Soil Research*, 35, 399-417.
- KROON, F. J., KUHNERT, P. M., HENDERSON, B. L., WILKINSON, S. N., KINSEY-HENDERSON, A., ABBOTT, B., BRODIE, J. E. & TURNER, R. D. R. 2011. River loads of suspended solids, nitrogen,

phosphorus and herbicides delivered to the Great Barrier Reef lagoon. *Marine Pollution Bulletin*, doi:10.1016/j.marpolbul.2011.10.018.

KUCZERA, G., KAVETSKI, D., FRANKS, S., THYER, M. 2006. Towards a Bayesian total error analysis of conceptual rainfall-runoff models: characterising model error using storm-dependent parameters. *Journal of Hydrology* 331: 161–177.

KUHNERT, P. M., HENDERSON, B. L., LEWIS, S., BAINBRIDGE, Z., WILKINSON, S. N. & BRODIE, J. 2012. Quantifying total suspended sediment export from the Burdekin River catchment using the loads regression estimator tool. *Water Resources Research*, 48.

LEWIS, S. E., BAINBRIDGE, Z. T., KUHNERT, P. M., SHERMAN, B. S., HENDERSON, B., DOUGALL, C., COOPER, M. & BRODIE, J. E. 2013. Calculating sediment trapping efficiencies for reservoirs in tropical settings: a case study from the Burdekin Falls Dam, NE Australia. *Water Resources Research* 49, 1017-1029.

MCCULLAGH, P. & NELDER, J. 1989. *Generalized Linear Models* (second edition), Boca Raton, USA, Chapman and Hall/CRC.

NATHAN, R. J. & MCMAHON, T. A. 2010. Evaluation of automated techniques for base flow and recession analyses. *Water Resources Research*, 26, 1465-1473.

PAGENDAM, D. E., KUHNERT, P. M., LEEDS, W. B., WIKLE, C. K., BARTLEY, R. & PETERSON, E. E. 2014. Assimilating catchment processes with monitoring data to estimate sediment loads to the Great Barrier Reef. *Environmetrics*, 25, 214-229.

ROBSON, B. J. & DOURDET, V. 2015. Prediction of sediment, particulate nutrient and dissolved nutrient concentrations in a dry tropical river to provide input to a mechanistic coastal water quality model. *Environmental Modelling and Software*, 63, 97-108.

SCHMELTER, M.L., HOOTEN, M.B., STEVENS, D.K. 2011. Bayesian sediment transport model for unisize bed load. *Water Resources Research* 47: W11514.

STEVENS, D. P., COX, J. W. & CHITTLEBOROUGH, D. J. 1999. Pathways of phosphorus, nitrogen and carbon movement over and through texturally differentiated soils, South Australia. *Australian Journal of Soil Research*, 37, 679-693.

THOMAS, S., HE, Y. & FLEMING, N. K. 2010. Progress Report on the Mount Lofty Ranges Source Catchments Application Project. eWater CRC Technical Report.

VRUGT, J., TER BRAK, C., CLARK, M., HYMAN, J. 2008. Treatment of input uncertainty in hydrologic modelling: doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research* 44: W00B09.

WELSH, W. D., VAZE, J., DUTTA, D., RASSAM, D., RAHMAN, J. M., JOLLY, I. D., WALLBRINK, P., PODGER, G. M., BETHUNE, M., HARDY, M. J., TENG, J. & LERAT, J. 2013. An integrated modelling framework for regulated river systems. *Environmental Modelling and Software*, 39, 81-102.

WILKINSON, S. N., DOUGALL, C., KINSEY-HENDERSON, A. E., SEARLE, R. D., ELLIS, R. J. & BARTLEY, R. 2014. Development of a time-stepping sediment budget model for assessing land-use impacts in large river basins. *Science of the Total Environment*, 468-469, 1210-1224.

WILKINSON, S. N., PROSSER, I. P., RUSTOMJI, P. & READ, A. M. 2009. Modelling and testing spatially distributed sediment budgets to relate erosion processes to sediment yields. *Environmental Modelling and Software*, 24, 489-501.

WOOD, S. N. 2006. *Generalized additive models: an introduction with R*, Boca Raton, FL, Chapman and Hall.



WU, W., CLARK, J., VOSE, J. 2010. Assimilating multi-source uncertainties of a parsimonious conceptual hydrological model using hierarchical Bayesian modeling. *Journal of Hydrology* 394: 436–446.

## Appendix A – Summary of Covariates used in Modelling

Table 31: Covariates used in site-based statistical models.

Predictor Type	Predictor Name	Description
Historical Flows	csQ	Past sum of flow
	d0.1	Discounted flow – 10%
	d0.25	Discounted flow – 25%
	d0.5	Discounted flow – 50%
	d0.75	Discounted flow – 75%
	d0.95	Discounted flow – 95%
	d0.99	Discounted flow – 99%
	Flow	pQ.bf
pQ.qf		Quick flow or runoff
limb		Rising or Falling Limb (-1 = fall, 0 = flat, 1 = rise)
Seasonal	month	Month of the year
Spatial	Site	Name of Site in the Onkaparinga catchment
	Longitude	Longitude
	Latitude	Latitude
	Land-use	As per Appendix C

## Appendix B - Diagnostic Plots for Site Based Models

### Scott Creek

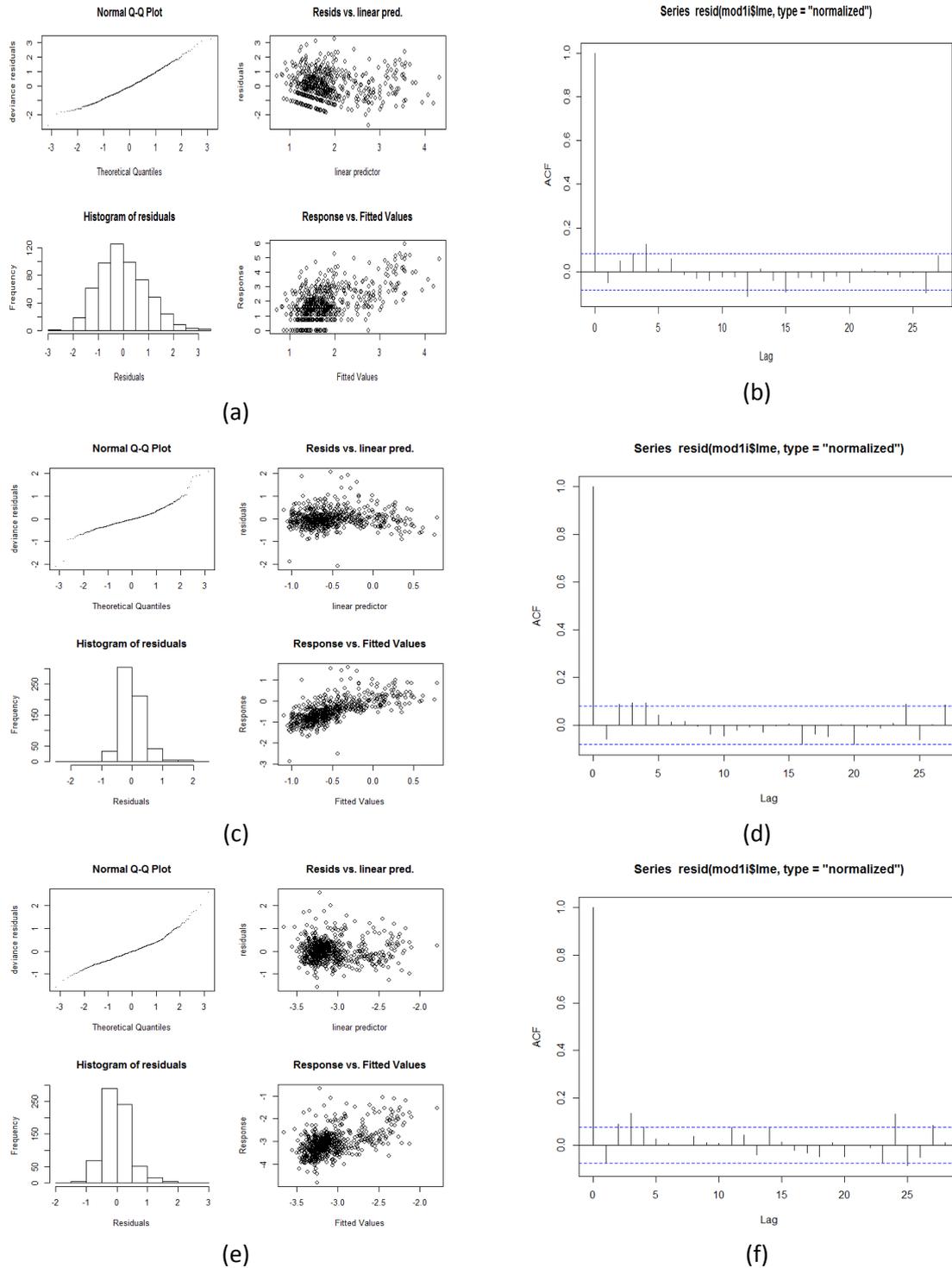


Figure 90: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row 2) and TP (row 3) data collected at Scott Creek (A5030502) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals.

## Onkaparinga River, Houlgraves

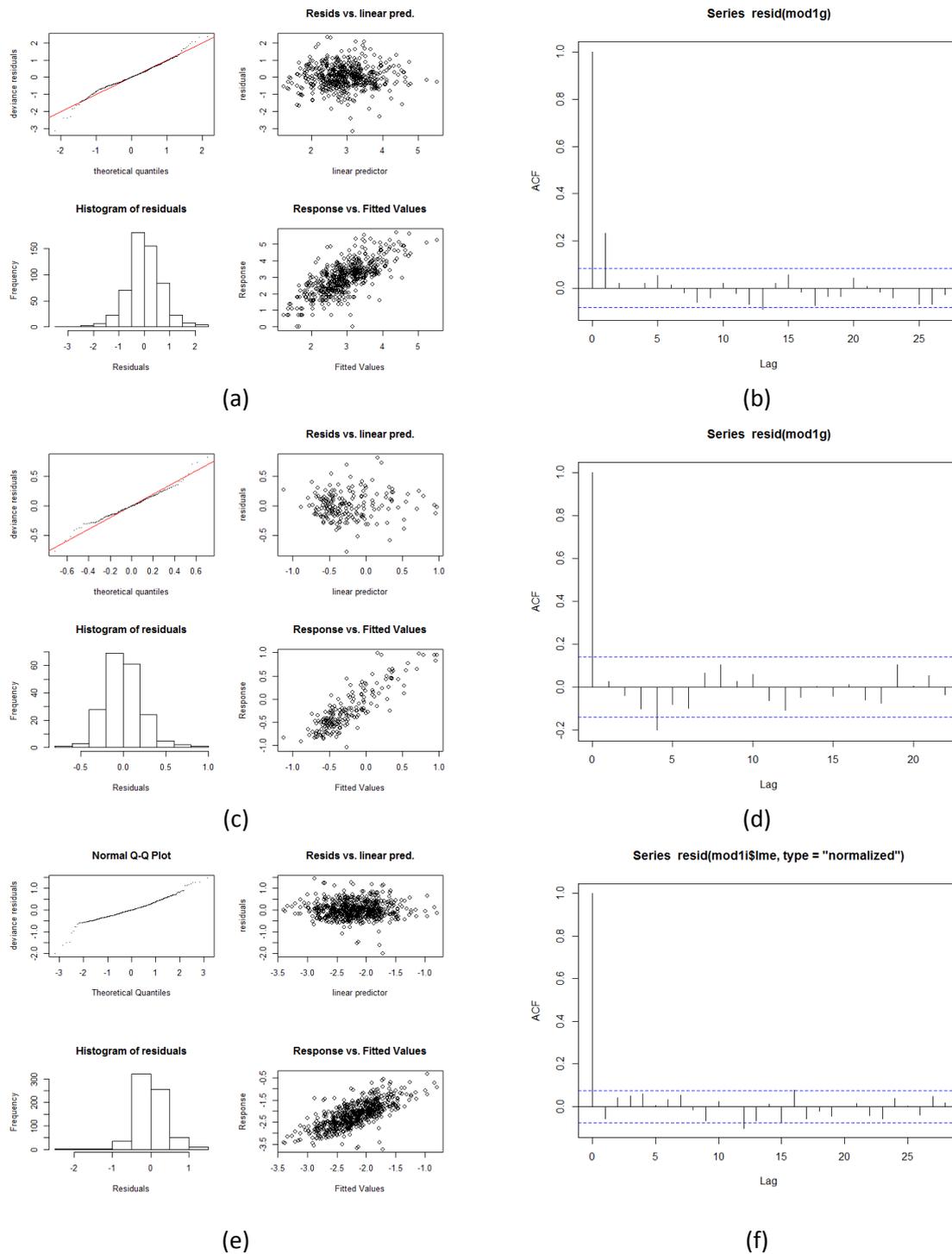


Figure 91: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row 2), TP (row 3) data collected at Onkaparinga River at Houlgraves (A5030504) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals.

## Echunga Creek

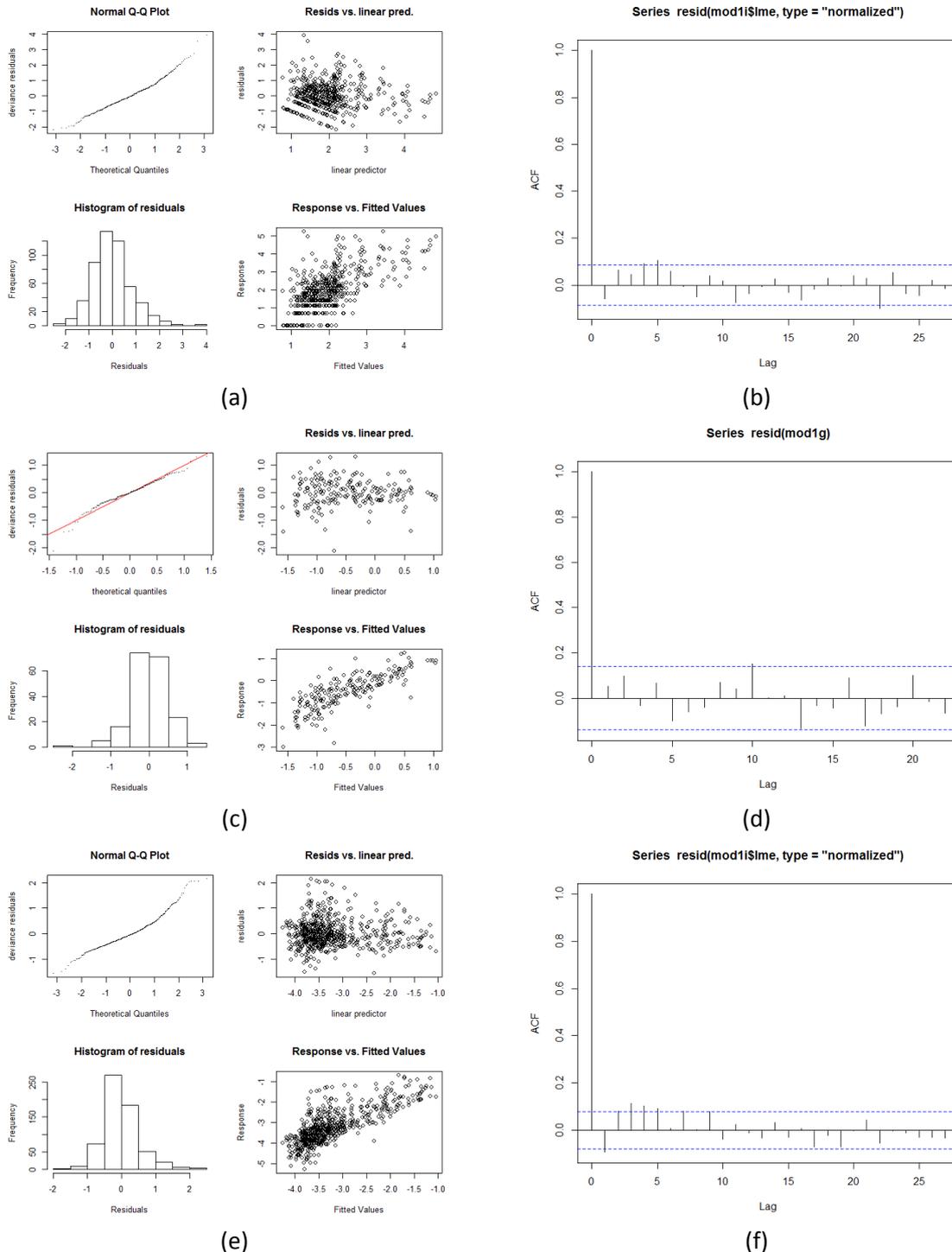


Figure 92: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row 2), TP (row 3) data collected at Echunga Creek (A5030506) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals.

## Lenswood Creek

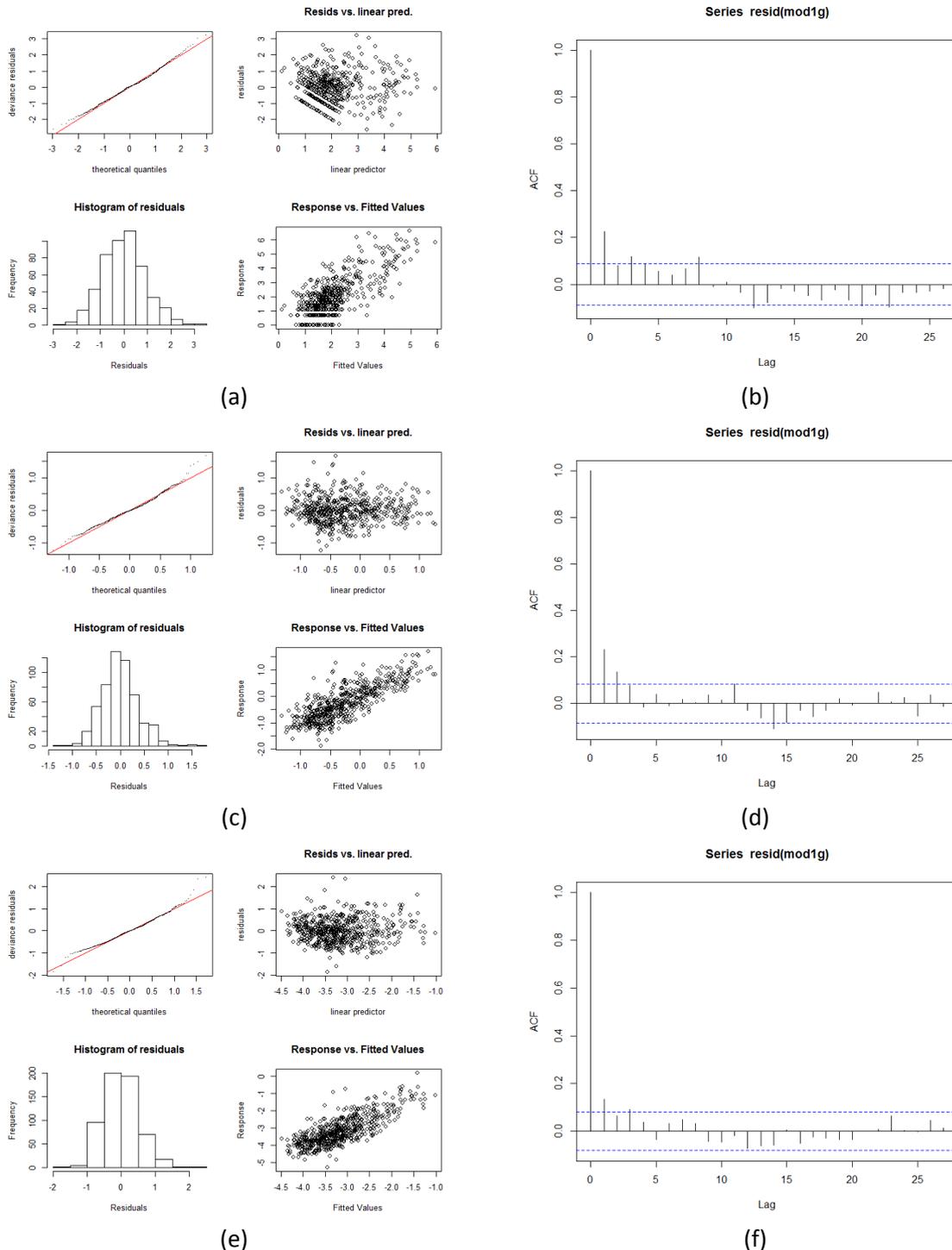


Figure 93: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row2), TP (row3) data collected at Lenswood Creek (A5030507) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals.

## Aldgate Creek

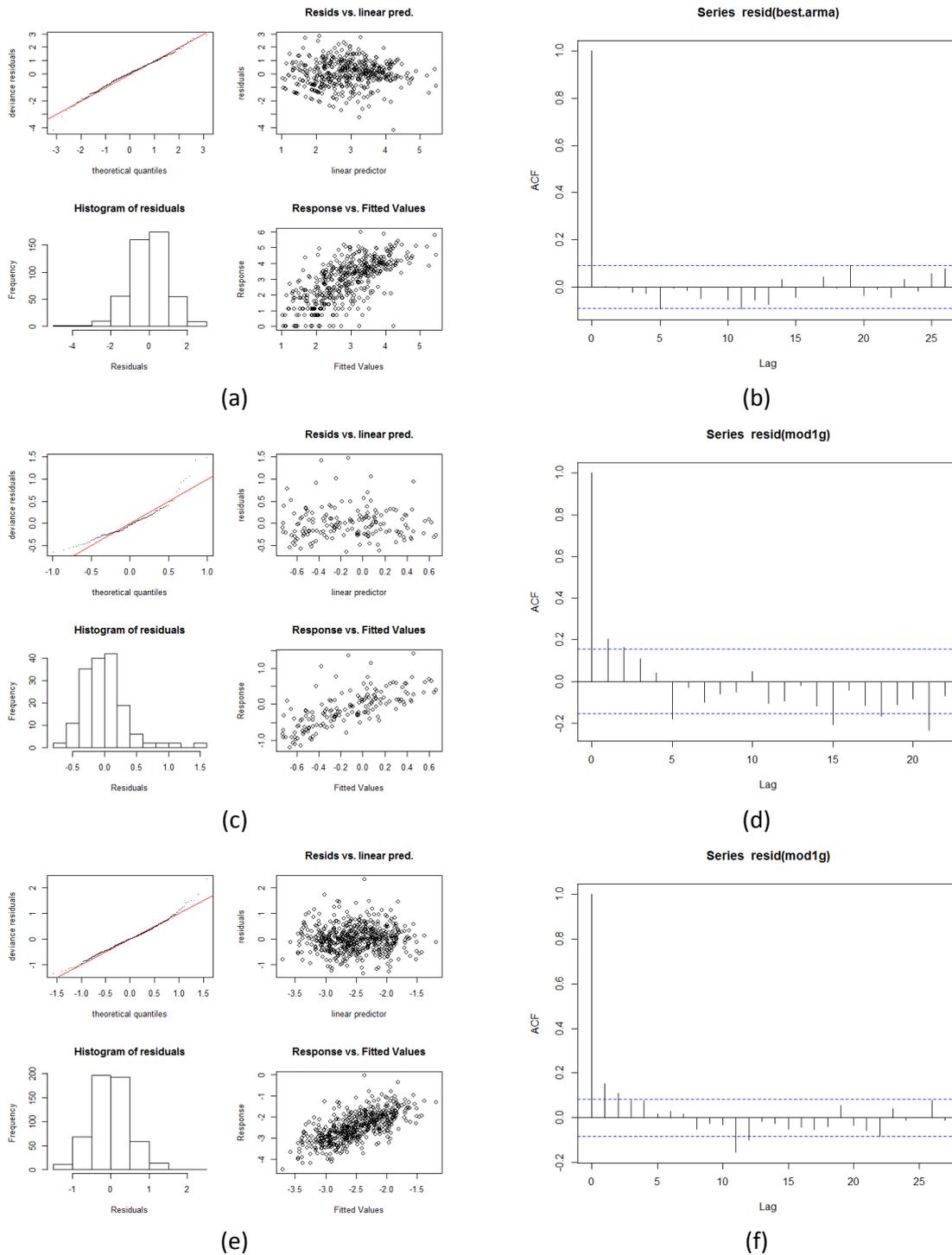


Figure 94: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row2), TP (row3) data collected at Aldgate Creek (A5030509) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals.

## Cox Creek

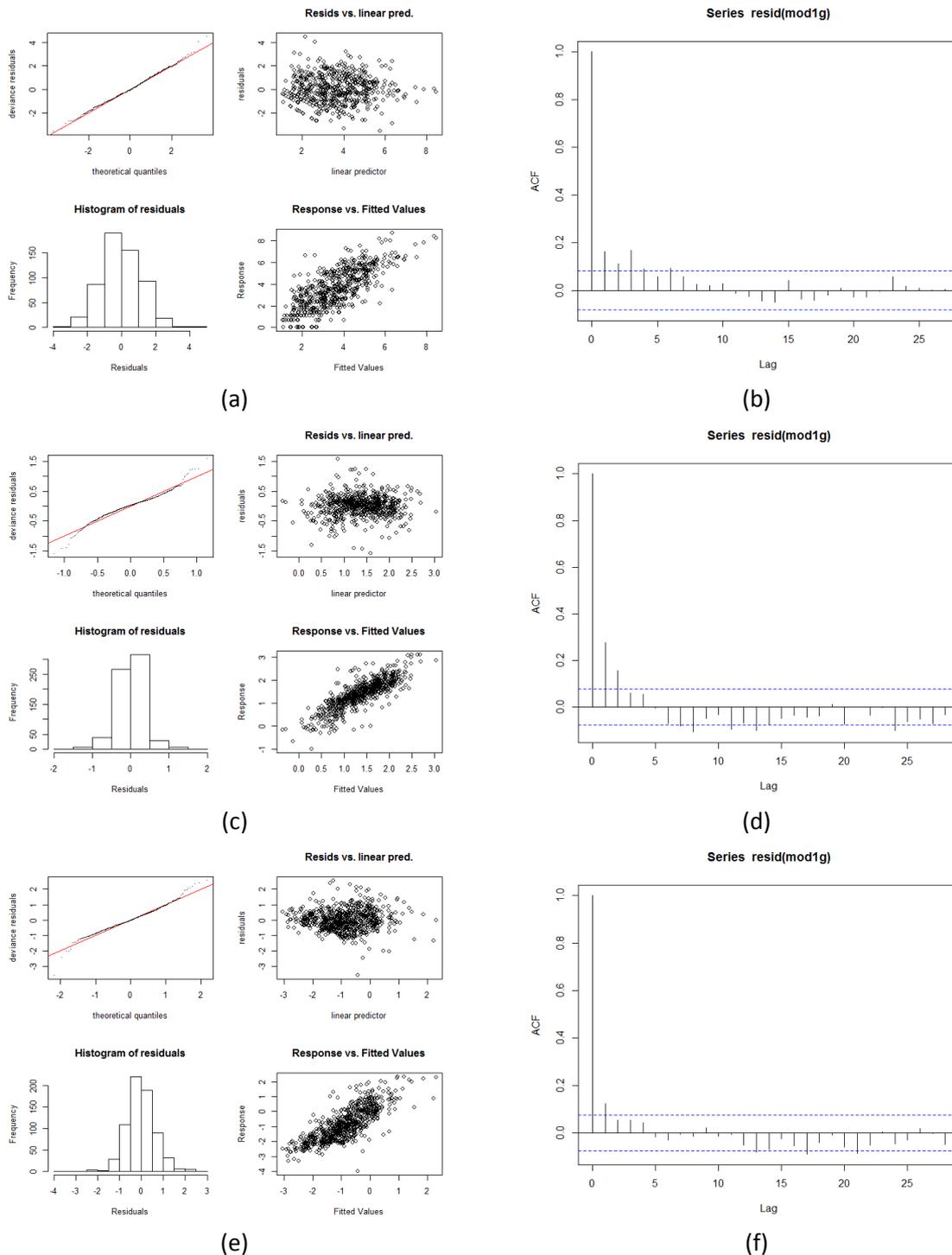


Figure 95: Diagnostic plots for the generalised additive model fit to TSS (row 1), TN (row 2) and TP (row 3) data collected at Cox Creek (A503026) showing (a) standard residual plots testing for normality and (b) the autocorrelation function of the residuals.

## Appendix C - Land-use Compositions Upstream from Onkaparinga Gauges

### Scott Creek (A5030502)

Table 32: Land-uses in the catchment area contributing to the Scott Creek site.

Land-use Functional Units	Broad Land-use Categorisations	Functional Unit Area (ha)	% Functional Unit
Broad-scale annual horticulture	non-urban	3.18	0.12%
Broad-scale perennial horticulture	non-urban	24.16	0.91%
Conservation area	non-urban	1216.20	45.70%
Dense urban	urban	31.50	1.18%
Farm dams	water	4.34	0.16%
Grazing	non-urban	430.76	16.19%
Intensive grazing	non-urban	19.52	0.73%
Intensive production	non-urban	21.04	0.79%
Recreation and culture	non-urban	10.88	0.41%
Rural living	urban	787.17	29.58%
Utilities	urban	112.31	4.22%

### Onkaparinga River at Houlgraves (A5030504)

Table 33: Land-uses in the catchment area contributing to the Houlgraves site.

Land-use Functional Units	Broad Land-use Categorisations	Functional Unit Area (ha)	% Functional Unit
Broad-scale annual horticulture	non-urban	1721.42	5.38%
Broad-scale perennial horticulture	non-urban	361.17	1.13%
Conservation area	non-urban	4074.14	12.74%
Dense urban	urban	4223.41	13.21%
Farm dams	water	1770.30	5.54%
Grazing	non-urban	319.72	1.00%
Intensive grazing	non-urban	12174.07	38.08%
Intensive production	non-urban	1211.86	3.79%
Recreation and culture	non-urban	39.80	0.12%
Rural living	urban	180.96	0.57%
Utilities	urban	510.17	1.60%

## Echunga Creek (A5030506)

Table 34: Land-uses in the catchment area contributing to the Echunga Creek site.

Land-use Functional Units	Broad Land-use Categorisations	Functional Unit Area (ha)	% Functional Unit
Broad-scale annual horticulture	non-urban	141.61	4.18%
Broad-scale perennial horticulture	non-urban	28.20	0.83%
Conservation area	non-urban	158.28	4.67%
Dense urban	urban	580.84	17.13%
Farm dams	water	40.79	1.20%
Grazing	non-urban	49.38	1.46%
Intensive grazing	non-urban	1625.92	47.95%
Intensive production	non-urban	205.15	6.05%
Recreation and culture	non-urban	121.08	3.57%
Rural living	urban	38.83	1.15%
Utilities	urban	299.93	8.85%

## Lenswood Creek (A5030507)

Table 35: Land-uses in the catchment area contributing to the Lenswood Creek site.

Land-use Functional Units	Broad Land-use Categorisations	Functional Unit Area (ha)	% Functional Unit
Broad-scale annual horticulture	non-urban	9.80	0.59%
Broad-scale perennial horticulture	non-urban	799.76	48.13%
Conservation area	non-urban	318.72	19.18%
Dense urban	urban	21.19	1.28%
Farm dams	water	24.50	1.47%
Grazing	non-urban	318.22	19.15%
Intensive grazing	non-urban	1.09	0.07%
Intensive production	non-urban	10.73	0.65%
Recreation and culture	non-urban	6.70	0.40%
Rural living	urban	78.90	4.75%
Utilities	urban	71.94	4.33%

## Aldgate Creek (A5030509)

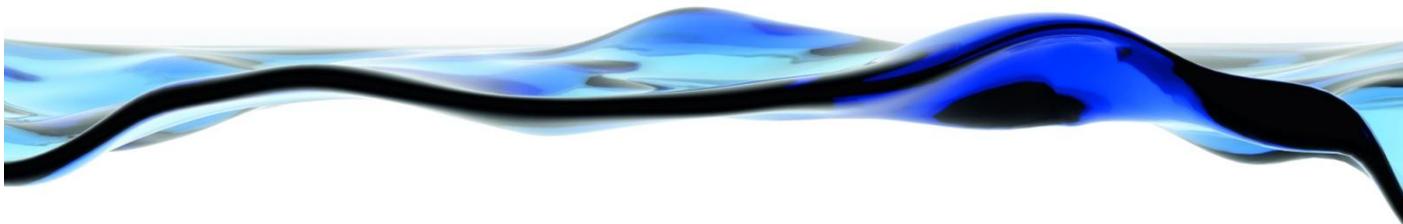
Table 36: Land-uses in the catchment area contributing to the Aldgate Creek site.

Land-use Functional Units	Broad Land-use Categorisations	Functional Unit Area (ha)	% Functional Unit
Broad-scale annual horticulture	non-urban	0.00	0.00%
Broad-scale perennial horticulture	non-urban	0.00	0.00%
Conservation area	non-urban	35.39	5.40%
Dense urban	urban	356.70	54.38%
Farm dams	water	1.06	0.16%
Grazing	non-urban	0.00	0.00%
Intensive grazing	non-urban	0.17	0.03%
Intensive production	non-urban	0.00	0.00%
Recreation and culture	non-urban	37.19	5.67%
Rural living	urban	95.82	14.61%
Utilities	urban	129.67	19.77%

## Cox Creek at Uraidla (A5030526)

Table 37: Land-uses in the catchment area contributing to the Cox Creek site.

Land-use Functional Units	Broad Land-use Categorisations	Functional Unit Area (ha)	% Functional Unit
Broad-scale annual horticulture	non-urban	149.82	28.01%
Broad-scale perennial horticulture	non-urban	123.76	23.14%
Conservation area	non-urban	71.93	13.45%
Dense urban	urban	47.82	8.94%
Farm dams	water	0.84	0.16%
Grazing	non-urban	11.14	2.08%
Intensive grazing	non-urban	3.68	0.69%
Intensive production	non-urban	1.11	0.21%
Recreation and culture	non-urban	5.44	1.02%
Rural living	urban	86.73	16.22%
Utilities	urban	32.60	6.10%



Government  
of South Australia

Department of Environment,  
Water and Natural Resources



THE UNIVERSITY  
of ADELAIDE



University of  
South Australia



Flinders  
UNIVERSITY  
ADELAIDE • AUSTRALIA

The Goyder Institute for Water Research is a partnership between the South Australian Government through the Department of Environment, Water and Natural Resources, CSIRO, Flinders University, the University of Adelaide and the University of South Australia.