

Modelling macronutrient dynamics in the Hampshire Avon river: A Bayesian approach to estimate effect of storm events

Monica Pirani^{a,*}, Anouska Panton^b, Duncan A. Purdie^c, Sujit K. Sahu^d

^a*MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, W2 1PG London, UK*

^b*School of Earth and Environmental Sciences, University of Portsmouth, PO1 3QL Portsmouth, UK*

^c*Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, SO14 3ZH Southampton, UK*

^d*Southampton Statistical Sciences Research Institute, University of Southampton, SO17 1BJ Southampton, UK*

Abstract

The macronutrients nitrate and phosphate are aquatic pollutants that arise naturally, however, in excess concentrations they can be harmful to human health and ecosystems. These pollutants are driven by river currents and show dynamics that are heavily affected by storms. As a result, the nutrient budget in the receiving estuaries and coasts change suddenly and seasonally, causing ecological damage to resident wildlife and fish populations. In this paper, we propose a statistical change point model that interacts between seasons and storm events, described by abrupt changes in river flow, to capture the macronutrient dynamics. It also accounts for the nonlinear effect of water quality properties via nonparametric penalised splines. This model enables us to estimate the riverine macronutrient fluxes and their annual deposited levels. In particular, we present a study on macronutrient dynamics on the Hampshire Avon River, which flows to the southern coast of the UK through the Christchurch Harbour estuary. We model daily data for more than a year during 2013-14 in which period there were multiple severe storm events leading to localised flooding. Adopting a hierarchical model within a Bayesian inference framework, we have quantified riverine macronutrient fluxes and the model also allows us to estimate past fluxes based on input river flow values. Out of sample empirical validation methods justify our approach, which captures also the dependencies of macronutrient levels with water body characteristics.

Keywords: Change point analysis, Hierarchical models, Macronutrients, Fluxes, Storms, Water quality properties.

1. Introduction

2 River ecosystems are experiencing rapid transformations in response to anthropogenic and
3 climatological stressors, which impact on macronutrient pollution, water quality characteristics,
4 biodiversity and ultimately on the ecological health of the rivers (Whitehead et al., 2009). In
5 particular, macronutrients nitrate and phosphate occur naturally in freshwater bodies, but when
6 present in excessive amounts can be harmful not only for aquatic life but also for human health,
7 reducing drinking water quality (Whitehead and Crossman, 2012). Many sources can contribute
8 to macronutrient over enrichment (eutrophication) from human activities, including runoff from
9 fertilised fields, discharge from sewage treatment, burning of fossil fuels and food production (e.g.
10 Conley et al., 2009; Paerl, 2009; Withers et al., 2014).

11 In addition to these disturbances, natural features of the environment and in particular climate
12 changes can compromise macronutrient cycles in fresh waters (Woodward et al., 2010; Whitehead
13 and Crossman, 2012). Climate changes are likely to bring an increasing number of extreme events
14 including increased frequency and intensity of storms, leading to high winds and rainfall.

15 In this paper, we are concerned with the dynamics of riverine nitrate and phosphate concentra-
16 tions driven by sudden and stochastic storm events, while accounting for the effect of water quality
17 properties, such as temperature, conductivity, dissolved oxygen and turbidity. Storms can greatly
18 impact on the macronutrient runoff and budget, but their ecological importance in riverine ecosys-
19 tems, is not well documented as noted by Leigh et al. (2014). Our understanding of the effect of
20 storms in riverine macronutrient dynamics will provide new insight into how the macronutrient
21 fluxes are transferred from river to estuarine waters and ultimately to coasts and ocean system.

22 Our contribution addresses the analysis using a non-standard regression approach, within a
23 Bayesian hierarchical structure (Berliner, 1996), which results in a generalized additive model that
24 is able to: (i) differentiate the potential effect of storm events on nitrate and phosphate according to
25 the time of year in which they occur, and (ii) capture the complex nonlinear relationships among

*Corresponding author: monica.pirani@imperial.ac.uk, MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, W2 1PG London, UK

Email addresses: monica.pirani@imperial.ac.uk (Monica Pirani), anouska.panton@port.ac.uk
(Anouska Panton), duncan.purdie@noc.soton.ac.uk (Duncan A. Purdie), S.K.Sahu@soton.ac.uk (Sujit K. Sahu)

26 macronutrients with water quality properties through unspecified smooth functions of these prop-
27 erties. The resulting model allows us to generate an annual budget, with quantified uncertainties,
28 for the aqueous input of macronutrients from river body to the downstream estuary.

29 The detection of impact of storm events is performed by way of the identification of significant
30 abrupt changes in the time-series of river flow. Indeed, flow alteration is an environmental factor to
31 which riverine ecosystems respond considerably (Poff and Zimmerman, 2010; Rolls et al., 2012).
32 We consider that these changes in river flow rate, that we call change thresholds following the
33 terminology adopted by Alameddine et al. (2011), can be different according to the period of the
34 year in which they occur, thus we accommodate a seasonal window introducing a possible shift in
35 time, that we simply call switchpoint, to distinguish it from the terminology used to refer to abrupt
36 changes in river flow. Therefore, we call the whole statistical model identifying the switchpoint in
37 time and the change thresholds in river flow simply as change point structures.

38 Change point analysis has become a popular tool in ecological statistics and in the simplest
39 form it detects distributional changes within observations that are ordered in time. Nevertheless,
40 its use in water quality models is still limited to only a few contributions. For example, Fortin
41 et al. (2004) reformulated the shifting-level model to show that it belongs to the class of hidden
42 Markov models, and developed Bayesian methods for evaluating its parameter uncertainty and
43 demonstrated its utility for forecasting of streamflows. Alameddine et al. (2011) used a change
44 point and threshold model to analyse the relationship between nitrogen concentrations and flow
45 regimes during a long period of 29 years, quantifying changes in this relationship across time.

46 In this paper, we adopt a Bayesian formulation of the model, which allows us to properly
47 account for the uncertainty associated with the detection of the switchpoint in time and the change
48 thresholds in river flows, modelling their locations as unknown. However, our formulation of
49 the change point analysis is different, as we use interaction terms between time and river flows
50 to capture a seasonal behavior in freshwater, that is known to be an important determinant when
51 considering macronutrient loadings (Sigleo and Frick, 2003; Laud and Ibrahim, 2008).

52 The modelling approach here presented is built on ad-hoc measurements collected within
53 the UK *Macronutrient Cycles Programme* founded by Natural Environmental Research Council
54 (NERC). In particular, we use macronutrient and river flow data, as well as water quality proper-

ties, from the Hampshire Avon river which flows to the South coast of the UK and feeds into the Christchurch Harbour estuary. We model daily concentrations of nitrate and phosphate for more than a year during 2013-14, a period in which the UK experienced a highly unusual number of storm events (Muchan et al., 2015), with series of destructive floods across the country.

59 2. Methods

60 2.1. Study area

61 The Hampshire Avon is one of the most biodiverse chalk rivers in the UK, providing a habitat
62 for a very rich flora and fauna. Much of the Hampshire Avon river has been designated as *Sites of*
63 *Special Scientific Interest* or as a *Special Area of Conservation*, and its water has been used for a
64 number of purposes including general agriculture, spray irrigation and fish farming, as well as for
65 public and private water supplies (Environment Agency, 2012).

66 The sampling site for this study is located at the lowest water flow gauging station on Hamp-
67 shire Avon at Knapp Mill (latitude: 50.74807, longitude: -1.77968), encompassing a catchment
68 area of 1706 km². Fig. 1 provides a map of the study area.

69 2.2. Macronutrient and water quality samples

70 Sampling at Knapp Mill was carried out between 22 November 2013 and 19 December 2014.
71 Water quality properties, including temperature, conductivity, dissolved oxygen, turbidity and
72 chlorophyll concentration, were measured *in situ* every 10 minutes using an EXO2 multiparam-
73 eter sonde (Xylem, UK). Samples for macronutrient analysis were collected every 8 to 15 hours
74 with an ISCO automated water sampler (RS Hydro, UK). Water samples were fixed immediately
75 with 0.015M mercuric chloride (750 µL in 150 mL) and later filtered through a glass fibre filter
76 upon return to the laboratory. Concentrations of inorganic macronutrients were determined at the
77 University of Portsmouth using a QuAAstro segmented flow nutrient analyser (SEAL Analytical,
78 UK). River flow data were obtained from the UK Environment Agency. To regularise the sampling
79 intervals between measurements, the 24-hour (daily) means were calculated and used for further
80 analyses.

81 *2.3. Exploratory analysis*

82 Table 1 provides the descriptive statistics for all the data collected at the Knapp Mill station
83 and also for the daily river flow data. The large difference between the mean and median daily
84 river flow clearly highlights the severe impacts of storm events that the UK experienced during
85 the 2013-2014 winter months. Time-series plots of these data are given in Figs. 2 and 3. A
86 visual inspection of the plots shows considerable variation in the daily levels of the data, with a
87 winter/summer seasonal pattern. The time-series for nitrate exhibits lower concentrations during
88 the winter months, from December 2013 to March 2014, while phosphate does not show a specific
89 trend but does display lower concentrations during the months of February and March 2014. In
90 general, there is greater overall variability in nitrate than seen in phosphate concentrations. From
91 Fig. 2 it is also apparent that river flow rates are at the highest during the winter months 2013-2014
92 with levels that gradually decline towards summer. Among the water quality properties (Fig. 3),
93 we observe, as expected, a seasonal temperature pattern, and higher level of turbidity during winter
94 months, consistent with altered flow regimes. Nitrate concentrations show a trend consistent with
95 changes in conductivity.

96 Table 1 shows the Spearman rank correlation coefficients between macronutrients and water
97 quality properties. Only temperature and conductivity have a strong positive correlation (>0.90),
98 while moderate correlations are found for dissolved oxygen % saturation with conductivity (0.50)
99 and turbidity (-0.62). Fig. 4 shows the relationship between macronutrient concentrations and
100 water quality properties, with these scatter plots revealing generally nonlinear relationships.

101 *2.3.1. Data pre-processing*

102 The various measured water quality properties have a range of different units, therefore for
103 modelling purposes these are standardised to have zero mean and unit variance. This procedure
104 makes the magnitude of the coefficients comparable. Macronutrient concentrations and river flow
105 data are modelled on logarithmic scale to stabilise their variance. Moreover, logarithmic transfor-
106 mation of the data is convenient for macronutrients, as they are nonnegative and their distributions
107 are often skewed to the right.

108 2.3.2. Water quality property selection

109 Before embarking on the task of modelling the data, we carefully examined the possibility of
110 issues arising from multicollinearity among the water quality properties that may compromise the
111 estimation of the regression coefficients and thus affect their interpretation. To mitigate this, we
112 applied a covariate selection procedure based on knowledge of riverine ecosystems as well as on a
113 conventional statistical methods such as Lasso (Least Absolute Shrinkage and Selection Operator;
114 Tibshirani (1996)) that allows identification of the water quality properties that have the strongest
115 association with variation in the macronutrient concentrations. Lasso is a method that is used in the
116 context of regression analysis, and it can simultaneously perform variable selection and shrinkage
117 of the vector of regression coefficients toward zero.

118 We use a Bayesian formulation of Lasso regression (Park and Casella, 2008; Hans, 2009;
119 O'Hara and Sillanpää, 2009) that is constructed by specifying a Laplace distribution as a prior
120 distribution for the model coefficients. We standardised all regressor variables and implemented
121 the Bayesian Lasso regression technique described by Lykou and Ntzoufras (2011, 2013). This
122 Lasso technique revealed temperature, conductivity, dissolved oxygen, and turbidity as the most
123 important water quality properties for modelling nitrate data on the log scale and our subsequent
124 analysis proceeds with these only. For modelling phosphate data, the Lasso technique showed
125 temperature and dissolved oxygen as the two most important covariates, followed by chlorophyll,
126 turbidity and conductivity. In this instance, however, we exclude chlorophyll from the main anal-
127 ysis on phosphate. In fact, although chlorophyll is important in analysing data sets from estuarine
128 and coastal waters, we find that chlorophyll is less important in explaining macronutrient dynamics
129 within riverine systems, where it is more likely the result from storm runoff and not a predictor.

130 **3. Model set-up**

131 The discussion in the previous section leads us to consider a regime switching model for
132 macronutrients according to both season and river flow that is able to adjust for nonlinear ef-
133 fects of the chosen water quality properties. The two macronutrients, nitrate and phosphate, are
134 here modelled separately, although it is possible to model them jointly. Joint modelling of the two
135 macronutrients nitrate and phosphate, is not of interest here since our objective is not to study their

136 inter-relationships, which is seen to be rather weak (correlation -0.16 in Table 1), but to predict
 137 their individual daily and annual fluxes into the estuary.

138 The model is developed for data y_t , which denotes the natural logarithm of the observed
 139 macronutrient concentration at day t , for $t = 1, \dots, T = 393$. At the first stage of the modelling
 140 hierarchy, we assume an independent Gaussian measurement error model:

$$y_t \sim \text{Normal}(\mu_t, \sigma^2), \quad t = 1, \dots, 393 \quad (1)$$

141 where μ_t denotes the time varying mean and σ^2 is the variance assumed to be constant at all time
 142 points. We do not consider time varying variances as we do not have replicated data at each time
 143 point to estimate them. Rather, our effort is dedicated to finding the best model for the mean
 144 concentration μ_t at time t in the next stage of modelling hierarchy.

145 The second stage of the hierarchy defines the model for μ_t . To incorporate nonlinear effects
 146 of each of the p water quality properties, we incorporate a nonparametric smoothing function
 147 $g_j(x_{tj})$ of x_{tj} at each time point t , where x_{tj} denotes the value of the j th water quality property at
 148 t th time point. The choice of the $g_j(\cdot)$ functions ranges from linear to nonparametric penalised
 149 splines (Eilers and Marx, 1996; Ruppert et al., 2003) which are well-known to be very flexible.
 150 In our implementation, following Crainiceanu et al. (2005), we construct the splines using radial
 151 basis functions (for details, see Bishop (1995)), which provides a more stable fit than traditional
 152 truncated linear basis. By denoting x to be a generic covariate, we define a set of K knots, $k_1 <$
 153 $k_2 < \dots < k_K$ taken to be equally spaced over the range of x . We consider a linear mixed model
 154 formulation of a penalised spline model given by:

$$g(x) = \beta_0 + \beta_1 x + \sum_{k=1}^K b_k (x - k_k)_+^d \quad (2)$$

155 where we treat β_0 and β_1 to be fixed but unknown parameters and assume $\mathbf{b} = (b_1, \dots, b_K)'$ to be
 156 the vector of random parameters corresponding to the set of basis functions $(x - k_k)_+^d$, that is equal
 157 to $(x - k_k)^d$ if $(x - k_k)^d > 0$ and zero otherwise, and d is the degree of the spline. Each component
 158 of \mathbf{b} is assigned an independent normal prior distribution with mean zero and unknown variance,
 159 σ_b^2 , to be estimated from the model.

160 Model (2), assumed for the j th covariate at t th time point, x_{tj} , is given by:

$$g_j(x_{tj}) = \beta_{0j} + \beta_{1j}x_{tj} + \sum_{k=1}^K b_{kj}(x_{tj} - k_{kj})_+^d, \quad j = 1, \dots, p, t = 1, \dots, T. \quad (3)$$

Here, we consider a model with the same set of knots and the same degree for the splines for all the covariates that have been normalised already, see Section 2.3.1. Assuming an additive model, we obtain the total contribution:

$$\sum_{j=1}^p g_j(x_{tj}) = \sum_{j=1}^p \beta_{0j} + \sum_{j=1}^p \beta_{1j}x_{tj} + \sum_{j=1}^p \sum_{k=1}^K b_{tkj}(x_{tj} - k_{kj})_+^d, \quad t = 1, \dots, T.$$

161 However, the p separate intercept terms will not be identified and hence we only take one global
162 intercept β_0 in place of the sum $\sum_{j=1}^p \beta_{0j}$. For ease of notation we shall write $\beta_j = \beta_{1j}$ for $j =$
163 $1, \dots, p$. Now, each b_{kj} for $k = 1, \dots, K$ and $j = 1, \dots, p$ is given an independent normal prior
164 distribution with mean zero and unknown variance, σ_b^2 as mentioned above.

165 Ruppert (2002) and Crainiceanu et al. (2005) recommends a number of knots that is large
166 enough to ensure flexibility. In our application we choose the number of knots to be 5 for the cubic
167 splines, (i.e. $d = 3$), which is judged to be sufficient for model fitting and prediction purposes. The
168 knots are chosen at equal spaced quantiles of each water quality variable.

169 Now we turn to modelling the step changes in nitrate and phosphate concentrations due to
170 changes in seasons and river flow. The exploratory analysis in Section 2.3 has made it clear that
171 the nutrient concentration is severely impacted upon by not only river flow but also seasonality.
172 However, it is likely that variations in river flow will have different effects on concentration in
173 different seasons. Moreover, natural rain fall, and hence river flow, does not strictly adhere to the
174 calendar dates. That is why, we let the parameter τ denote the switchpoint in time that serves as
175 the unknown boundary between the end of the winter high flow season and the start of the low
176 flow season spanning the rest of the year. Since τ is unknown we estimate it along with all other
177 parameters. To allow for interactions between seasons and river flow levels we imagine that there
178 are two change thresholds in river flow which occur once during the winter and the other during
179 the rest of the year. Let φ_1 and φ_2 , denote these flow threshold parameters. Hence, we introduce
180 the following four terms in the model:

- 181 1. $\delta_1 I(t < \tau)I(f_t < \varphi_1)(f_t - \varphi_1)$ describing the effect of incremental flow less than φ_1 before the
 182 switchpoint in time,
- 183 2. $\delta_2 I(t < \tau)I(f_t \geq \varphi_1)(f_t - \varphi_1)$ describing the effect of incremental flow greater than φ_1 before
 184 the switchpoint in time,
- 185 3. $\delta_3 I(t \geq \tau)I(f_t < \varphi_2)(f_t - \varphi_2)$ describing the effect of incremental flow less than φ_2 after the
 186 switchpoint in time,
- 187 4. $\delta_4 I(t \geq \tau)I(f_t \geq \varphi_2)(f_t - \varphi_2)$ describing the effect of incremental flow greater than φ_2 after
 188 the switchpoint in time,

189 where $I(A) = 1$ if A is true and 0 otherwise. For model identifiability reasons, we set $\delta_3 = 0$ so
 190 that the three remaining parameters, δ_1, δ_2 and δ_4 measure incremental slope relative to the one for
 191 low river flow after the switchpoint in time.

192 Putting the above discussions together, we arrive at the following model for μ_t :

$$\mu_t = \beta_0 + \sum_{j=1}^p \beta_j x_{tj} + \sum_{j=1}^p \sum_{k=1}^K b_{kj}(x_{tj} - k_{kj})_+^d + \sum_{h=1}^4 \delta_h v_{th} \quad (4)$$

193 where v_{th} denotes the product of the two indicator functions and the incremental river flow corre-
 194 sponding to δ_h for $h = 1, \dots, 4$. In subsequent discussion we denote this general model by M1.
 195 We compare this model with the following sub-models of interests:

- 196 • M2. A linear regression model for the water quality properties, but no change point struc-
 197 tures:

$$\mu_t = \beta_0 + \sum_{j=1}^p \beta_j x_{tj} \quad (5)$$

- 198 • M3. A linear regression model for the water quality properties, with only a switchpoint in
 199 time:

$$\mu_t = \beta_0 + \sum_{j=1}^p \beta_j x_{tj} + \delta_1 I(t \geq \tau) \quad (6)$$

- 200 • M4. A linear regression model for the water quality properties, with only a change threshold
 201 in river flow:

$$\mu_t = \beta_0 + \sum_{j=1}^p \beta_j x_{tj} + \delta_1 f_t + \delta_2 I(f_t \geq \varphi)(f_t - \varphi) \quad (7)$$

- 202 • M5. A linear regression model for the water quality properties, with change point structures
 203 for time and river flow:

$$\mu_t = \beta_0 + \sum_{j=1}^p \beta_j x_{tj} + \sum_{h=1}^4 \delta_h v_{th} \quad (8)$$

- 204 • M6. A regression model via penalised splines for the water quality properties, but no change
 205 point structures:

$$\mu_t = \beta_0 + \sum_{j=1}^p \beta_j x_{tj} + \sum_{j=1}^p \sum_{k=1}^K b_{kj}(x_{tj} - k_{kj})_+^d \quad (9)$$

206 To account for temporal dependence that is expected to occur between measurements collected
 207 on consecutive days, we also evaluated the additional inclusion in (4) of a random intercept, mod-
 208 elled as a linear stationary first-order autoregressive process, η_t , which is a very popular choice in
 209 time series analyses. Thus, the model for η_t assumes the form: $\eta_t = \rho \eta_{t-1} + u_t$, where the error u_t
 210 is white noise, that is normally distributed with mean 0 and variance σ_η^2 , and the parameter ρ is
 211 assumed be in the interval $[-1, 1]$. However, we were not able to fit this model to the data due to
 212 lack of identifiability.

213 The Bayesiasn model is completed by assuming prior distributions for all the unknown pa-
 214 rameters, denoted by θ . We assume that the switchpoint in time, τ , is uniformly distributed on
 215 $[1, 2, \dots, T]$. Note that $\tau = 1$ and $\tau = T$ does not imply any change. We also assess a discrete
 216 uniform prior for the switchpoint in time, that showed lead to a better convergence for the phos-
 217 phate model, though requiring a higher computational effort. Similarly, we adopt uniform prior
 218 distributions for the two change thresholds on river flow φ_1 and φ_2 , in the interval $[1.995, 4.631]$,
 219 which are the minimum and maximum values of the river flow on the logarithmic scales. The
 220 precision parameters (i.e. inverse of the variance parameters) specific for each macronutrient, σ^{-2} ,
 221 are assumed to follow a Gamma distribution $Ga(a, b)$ independently, with shape parameter, a , and
 222 expectation, a/b . In particular, we assume a proper prior specification by taking $a = 2$ and $b = 1$
 223 for these parameters. We assume normal prior distributions for β_0 and the fixed effect parameters

224 β specified as $\text{Normal}(0, 10^4)$. Moreover, as previously mentioned, an independent normal prior
 225 distribution, centered at zero, is chosen for the random effects parameters \mathbf{b} associated with the
 226 penalised splines for the water quality properties. For σ_b^2 , which controls the amount of smooth-
 227 ness of the water quality properties, we consider two different prior distributions: (i) a Gamma
 228 distribution for the precision parameter, $\sigma_b^{-2} \sim \text{Ga}(a_b, b_b)$, with $a_b = 1$ and $b_b = 0.001$, and (ii) a
 229 half-Cauchy distribution for the standard deviation parameter, $\sigma_b \sim \text{half-Cauchy}(A)$, with $A = 25$,
 230 as suggested by Marley and Wand (2010). By comparing model fits under both of these prior dis-
 231 tributions, we adopt the first parameterization for the nitrate model and the latter for the phosphate
 232 model. Finally, we assume a normal prior distribution for the δ parameters associated with the
 233 change point structures.

234 Fig. 5 presents the Directed Acyclic Graph (DAG) of our model, that is a simplified graphical
 235 representation of the hierarchical modelling structure. In this diagram each quantity is represented
 236 by a node and links between nodes show direct dependence. The ellipses represent stochastic
 237 nodes (that is, variables that have a probability distribution) or logical nodes (that is, deterministic
 238 relationships or functions). The small squares identify nodes that are constants. Stochastic depen-
 239 dence and functional dependence are denoted by solid and dashed arrows, respectively. Finally,
 240 the large square plates represent repetitive structures (i.e. the 'loop' from $t = 1$ to T).

241 To compare the quality of the model fit of the proposed modelling approach in comparison
 242 to the above described simpler statistical models, we adopt the predictive model choice criterion
 243 (PMCC; Laud and Ibrahim (1995); Gelfand and Ghosh (1995)) defined by:

$$PMCC = \sum_{t=1}^T \left\{ y_t^{\text{obs}} - E(y_t^{\text{rep}}) \right\}^2 + \sum_{t=1}^T \text{Var}(y_t^{\text{rep}}) \quad (10)$$

244 where y_t^{rep} denote the future replicate of the observed macronutrient concentrations y_t^{obs} . The
 245 PMCC essentially quantifies the fit of the model by comparing the posterior predictive distribution
 246 obtained from the assumed model $p(y_t^{\text{rep}}|y_t^{\text{obs}})$ with the observed data. The first term of (10) gives a
 247 goodness of fit measure (G) which will decrease with increasing model complexity and the second
 248 term of (10) is a penalty term (P) which tends to be larger for complex models. The model with
 249 the smallest value of PMCC is the preferred model.

250 *3.1. Computation*

251 Our Bayesian model fitting and computations are based on Markov chain Monte Carlo (MCMC)
252 methods (e.g. Gilks et al., 1996). In particular, using MCMC, we obtain a sample of the model
253 parameters from the target posterior distribution. MCMC samples are used to obtain summaries
254 of the posterior distributions, such as mean, median and quantiles which were used to construct
255 the 95% credible intervals (CI).

256 The implementation of the models has been performed using the freely available software
257 package WinBUGS (version 1.4.3; Lunn et al. (2000)), that was executed in batch mode using
258 the R library R2WinBUGS (version 2.1-19; Sturtz et al. (2005)). WinBUGS code for the model
259 M1 is available in the Supplementary material. We have run two parallel MCMC chains indepen-
260 dently starting at two very different initial values for 50,000 iterations with 20,000 burn-in, and
261 we thinned the Markov chains by a factor of 10, resulting in samples of size 6,000 to estimate the
262 posterior distributions for the parameters of interest. Convergence was assessed by checking the
263 trace plots of the samples of the posterior distribution and the estimated autocorrelation functions
264 and the Monte Carlo standard errors.

265 *3.1.1. Estimation and prediction of macronutrient fluxes*

266 To assess the quality of the probabilistic predictions of macronutrient concentrations, which
267 can be obtained using the proposed model, we use out-of-sample validation techniques. Here,
268 we remove a set of consecutive observations from the sample and then use the remaining data
269 to fit the models. Using the fitted model we predict the set aside data based on their posterior
270 predictive distributions. These predictions are compared with the actual observations to validate
271 the model. In particular, we remove the last 20 days (from 30/11/2014 to 19/12/2014) data from
272 the macronutrient time-series and compare these set aside samples with model based predictions.

273 The Bayesian methods allow us to estimate the daily total deposit (mass flux) of each macronu-
274 trient as follows. Note that macronutrient flux is defined as the product of concentration times river
275 flow rate (Sigleo and Frick, 2003; Quillb   et al., 2006), measured in *Kg/day*, i.e. flux at day t , de-
276 noted by ξ_t is $\mu_t \times f_t$ where μ_t is converted to be measured in *Kg/m³* and river flow is converted in
277 *m³/day*. We estimate ξ_t and its uncertainty by using $\xi_t^\ell = \mu_t^\ell f_t$ where $\ell = 1, \dots, 6000$ indexes the

278 thinned MCMC iterates.

279 4. Results

280 Values of the PMCC are reported in Table 3 from which we are able to judge the worth of
281 each of the modelling strategies: change point structures, penalised splines and linear regression
282 model for the water quality properties. Model M5 based on the linear regression model for the
283 water quality properties provides almost equal performance but shows a worse goodness-of-fit
284 as expected, since the spline based models are more flexible. Henceforth, we proceed to make
285 inference using the best model M1.

286 To assess the adequacy of the chosen model M1 for the macronutrients data, we have checked
287 the residuals plots. Fig. 6 illustrates the median of the posterior distributions of the standardised
288 residuals plotted against the time period for nitrate and phosphate. No discernible patterns is
289 present for nitrate, with a random scattering of points. For phosphate, the residuals scatter around
290 zero reasonably with a few exceptions. This result supports an overall adequacy of the model for
291 the data.

292 4.1. Parameter and flux estimates for the macronutrients

293 Parameter estimates for the chosen model M1 are presented in Table 4. The switchpoint in
294 time for nitrate, estimated to occur on 08/03/2014 (95%CI: 05/03/2014, 13/03/2014), identifies
295 essentially two seasonal periods that are, clearly, winter and summer times. The change thresh-
296 olds before and after this switchpoint captures two regimens in river flow, occurring at $27.87\text{ m}^3/\text{s}$
297 (95%CI: 16.26, 43.64) and $10.64\text{ m}^3/\text{s}$ (95%CI: 7.64, 13.41) in winter and summer times respec-
298 tively. Taking low flow conditions in summer as reference category, the results suggest that a
299 higher level of river flow in winter, as well as in summer, is associated with increased concen-
300 trations of nitrate, such that a difference of 1 in river flow corresponds, on original scale, to an
301 increase in nitrate of about 1.17 mg/L in winter, and about 1.22 mg/L in summer.

302 Phosphate shows a considerable different change point structure, with a no clearly identifiable
303 seasonal variation. The switchpoint in time for phosphate is estimated to occur on 24/01/2014
304 (95%CI: 22/01/2014, 28/01/2014). Because of this early identification of the switchpoint in time,

305 during which the Hampshire Avon is still experiencing extremely high flow levels, the associated
306 estimation of the change thresholds in river flow lacks of precision. This is clearly showed in a
307 larger uncertainty in the estimation of the change threshold parameters, occurring at $66.35\text{ m}^3/\text{s}$
308 (95%CI: 7.89, 100.28) before the switchpoint in time and at $34.29\text{ m}^3/\text{s}$ (95%CI: 29.19, 40.32)
309 after the switchpoint in time. The increase in phosphate before the switchpoint in time, associated
310 with high river flow is not significant, however after the switchpoint in time, a higher level of river
311 flow seems associated with a dilution of phosphate of about 0.29 mg/L .

312 Figs. 7 and 8 show graphically the relationship between river flow with nitrate and phosphate
313 concentrations respectively, according to the estimated parameters for the change point structures.

314 The fixed effects for the water quality properties in the model for nitrate show a negative
315 relationship with temperature and a positive relationship for conductivity and turbidity. A nega-
316 tive fixed effect of dissolved oxygen is estimated for phosphate. However these relationships are
317 nonlinear as confirmed by the estimated four standard deviations of penalised splines, that are
318 non-zero. The estimates of the measurement error variance are higher in magnitude for phosphate
319 than for nitrate.

320 Fig. 9 shows the daily time-series of macronutrient fluxes (Kg/Day) based on the measured
321 data (black dots) and estimated by the model (black solid lines; shaded area represent 95% CI),
322 along with the fluxes predicted by the model assuming the observed data from 30/11/2014 to
323 19/12/2014 as unknown (here plotted within the red rectangle). The 95% CI for the predicted
324 fluxes include the actual 20 observed fluxes for the macronutrient data, although these intervals
325 are more conservative (that is, wider) for phosphate in comparison to nitrate. In general, these
326 analyses show that the posterior prediction concentrations reasonably agree with the observed
327 data.

328 Finally, we estimate the total macronutrient fluxes from the complete model, according to the
329 estimated parameters for the change point structures. We find strong seasonal effects in the riverine
330 nitrate fluxes as shown in Fig. 10. For example, in winter time low-flow conditions (that is, before
331 the 08/03/2014) the mean of the daily observed nitrate fluxes is $5,552\text{ Kg}/\text{Day}$ (that correspond
332 to an estimated daily posterior mean of $5,531\text{ Kg}/\text{Day}$ from our model), while in winter time
333 high-flow conditions, the mean increases to $31,696\text{ Kg}/\text{Day}$ (that correspond to an estimated daily

334 posterior mean of 31,668 Kg/Day from our model). The seasonal structure is not so clear in the
335 model for phosphate. From Fig. 11 we can see that most of the days (no. 260), occurring after
336 the 24/01/2014, are classified as low-flow conditions. However, we can still estimate the effect
337 of high flow caused by extreme rainfall events in the model for phosphate. For example, before
338 24/01/2014, the mean of the daily observed nitrate fluxes in low-flow conditions is 197.18 Kg/Day
339 (that corresponds to an estimated daily posterior mean of 208.68 Kg/Day from our model), and the
340 mean in high-flow conditions is definitively higher, being equal to 605.7 Kg/Day (that corresponds
341 to an estimated daily posterior mean of 573.0 Kg/Day from our model).

342 5. Discussion and conclusion

343 This paper proposes a flexible model for studying the dynamics of different macronutrient
344 species in response to changes in river flow, which are largely driven by storm events. This is
345 accomplished by using an interaction between time and river flow via change point structures. It
346 also investigates the role of the water quality properties of the catchment area, without imposing
347 any parametric model (e.g. linear) in their relationships with macronutrient concentrations. In
348 the application considered in the paper, we observe that a nonparametric spline based model out-
349 performs the multiple linear regression model. This is coherent with Walther (2010), who also
350 noted that the relationships among the components of ecological systems are complex and that
351 interactions and feed-back mechanisms can lead to nonlinear and abrupt changes.

352 An important feature of our model is that it allows the predictions of macronutrient concen-
353 trations and also the quantification of riverine input fluxes to the estuary. We illustrate this by
354 providing estimates of daily fluxes along with their uncertainties. These daily fluxes can be ag-
355 gregated to seasonal levels, as demonstrated in our application, where we found that the amount
356 of macronutrients delivered to the estuary can change dramatically according to the period of the
357 year in which storm events, associated with high river flow, occur. This is particularly evident for
358 nitrate which shows a clear seasonal pattern, while flux estimates for phosphate present a weaker
359 seasonal structure, that lead to a higher uncertainty in our modelling approach. Moreover, we can
360 estimate the macronutrient fluxes for historical time periods if the model inputs, i.e. river low
361 levels and water quality properties are available.

362 The Bayesian modelling framework adopted here can be extended in various ways by including
363 more relevant covariates, e.g. wind speed and direction so that a better estimate of change of
364 concentrations and fluxes can be obtained. Multivariate modelling for both the aquatic pollutants
365 and for data from multiple sites may also lead to fruitful research.

366 Computationally, the proposed model has been implemented in publicly available software,
367 which makes it an ideal candidate for similar data modelling and analysis problems for estuarine
368 pollution.

369 **6. Acknowledgement**

370 This study is supported by the Natural Environmental Research Council (NERC) through *The*
371 *Christchurch Harbour Macronutrients Project* (grant number: NE/J012238/1), which is one of the
372 four consortium projects of the *Macronutrients Cycles Programme*.

373 **7. Conflict of interest**

374 The authors declare that there are no conflicts of interest.

375 **References**

- 376 Alameddine I, Qian SS, Reckhow KH. A Bayesian changepoint-threshold model to examine the effect of TMDL
377 implementation on the flow-nitrogen concentration relationship in the Neuse River basin. Water Research
378 2011;45:51–62.
- 379 Berliner LM. Hierarchical bayesian time series models. In: Hanson K, Silver R, editors. Maximum Entropy and
380 Bayesian Methods. Kluwer Academic Publishers; 1996. p. 15–22.
- 381 Bishop CM. Neural Networks for Pattern Recognition. Oxford: Oxford University Press, 1995.
- 382 Conley DJ, Paerl HW, Howarth RW, Boesch DF, Seitzinger SP, Havens KE, Lancelot C, Likens GE. Controlling
383 eutrophication: nitrogen and phosphorus. Science 2009;323:1014–5.
- 384 Crainiceanu C, Ruppert D, Wand MP. Bayesian analysis for penalized spline regression using WinBUGS. Journal of
385 Statistical Software 2005;14:1–24.
- 386 Eilers PHC, Marx BD. Flexible smoothing with b-splines and penalties. Statistical Science 1996;11:89–121.
- 387 Environment Agency . Hampshire Avon WFD Management Area Abstraction Licensing Strategy. Environment
388 Agency, 2012.

- 389 Fortin V, Perreault L, Salas J. Retrospective analysis and forecasting of streamflows using a shifting level model.
390 Journal of Hydrology 2004;296:135–63.
- 391 Gelfand A, Ghosh S. Model choice: A minimum posterior predictive loss approach. Biometrika 1995;85:1–11.
- 392 Gilks WR, Richardson S, Spiegelhalter DJ. Markov Chain Monte Carlo in practice. Chapman & Hall/CRC, 1996.
- 393 Hans C. Bayesian lasso regression. Biometrika 2009;69:835–45.
- 394 Laud P, Ibrahim J. Predictive model selection. Journal of the Royal Statistical Society, Series B 1995;57:247–62.
- 395 Laud P, Ibrahim J. Seasonal variation of riverine nutrient inputs in the northern Bay of Biscay (France), and patterns
396 of marine phytoplankton response. Journal of Marine Systems 2008;72:309–19.
- 397 Leigh C, Bush A, Harrison ET, Ho SS, Luke L, Rolls RJ, Ledger ME. Ecological effects of extreme climatic events
398 on riverine ecosystems: insights from Australia. Freshwater Biology 2014;60:2620–2638.
- 399 Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modelling framework: concepts, structure, and
400 extensibility. Statistics and Computing 2000;10:325–37.
- 401 Lykou A, Ntzoufras I. WinBUGS: a tutorial. Wiley Interdisciplinary Reviews: Computational Statistics 2011;3:385–
402 96.
- 403 Lykou A, Ntzoufras I. On Bayesian lasso variable selection and the specification of the shrinkage parameter. Statistics
404 and Computing 2013;23:361–90.
- 405 Marley JK, Wand MP. Non-standard semiparametric regression via BRugs. Journal of Statistical Software 2010;37:1–
406 30.
- 407 Muchan K, Lewis M, Hannaford J, Parry S. The winter storm of 2013/2014 in the UK: hydrological responses and
408 impacts. Weather 2015;70:55–61.
- 409 O’Hara RB, Sillanpää MJ. A review of Bayesian variable selection methods: what, how and which. Bayesian Analysis
410 2009;4(1):85–117.
- 411 Paerl HW. Controlling eutrophication along the freshwater-marine continuum: dual nutrient (N and P) reductions are
412 essential. Estuaries and Coasts 2009;32:593–601.
- 413 Park T, Casella G. The Bayesian Lasso. Journal of the American Statistical Association 2008;103:681–6.
- 414 Poff NL, Zimmerman JKH. Ecological responses to altered flow regimes: a literature review to inform the science
415 and management of environmental flows. Freshwater Biology 2010;55:194–205.
- 416 Quilbé R, Rousseau AN, Duchemin M, Poulin A, Gangbazo G, Villeneuve JP. Selecting a calculation method to
417 estimate sediment and nutrient loads in streams: Application to the Beaurivage river (Québec, Canada). Journal of
418 Hydrology 2006;326:295–310.
- 419 Rolls RJ, Leigh C, Sheldon F. Mechanistic effects of low-flow hydrology on riverine ecosystems: ecological principles
420 and consequences of alteration. Freshwater Science 2012;31:1163–86.
- 421 Ruppert D. Selecting the number of knots for penalized splines. Journal of Computational and Graphical Statistics
422 2002;11:735–57.

- 423 Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press, 2003.
- 424 Sigleo AC, Frick WE. Seasonal variations in river flow and nutrient concentrations in a northwestern USA watershed.
- 425 In: Proceedings of the first interagency on research in the watersheds. USDA, Benson, Arizona; 2003. p. 370–6.
- 426 Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software* 2005;12:1–16.
- 428 Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 1996;58:267–88.
- 430 Walther GR. Community and ecosystem responses to recent climate change. *Philosophical Transactions of the Royal Society B* 2010;365:2019–24.
- 432 Whitehead P, Crossman J. Macronutrient cycles and climate change: Key science areas and an international perspective. *Science of the Total Environment* 2012;434:13–7.
- 434 Whitehead PG, Wilby RL, Battarbee RW, Kernan M, Wade AJ. A review of the potential impacts of climate change on surface water quality. *Hydrological Sciences Journal* 2009;54:101–23.
- 436 Withers PJA, Neal C, Jarvie HP, Doody DG. Agriculture and eutrophication: where do we go from here? *Sustainability* 2014;6:5853–75.
- 438 Woodward G, Perkins DM, Brown LE. Climate change and freshwater ecosystems: impacts across multiple levels of organization. *Philosophical Transactions of the Royal Society B* 2010;365:2093–106.

Table 1: Summary statistics for macronutrients, physical and chemical properties of water and river flow. Hampshire Avon at Knapp Mill, 22/11/2013 to 19/12/2014.

	Min	1 st Q	Mean	Median	3 rd Q	Max
<i>Macronutrients</i>						
Nitrate (mg/L)	2.48	4.83	5.33	5.21	5.93	7.10
Phosphate (mg/L)	0.01	0.06	0.07	0.07	0.09	0.39
<i>Water properties</i>						
Temperature (°C)	4.71	8.11	12.26	12.21	16.24	21.96
Conductivity ($\mu\text{S}/\text{cm}$)	200.96	340.06	384.60	381.24	439.65	501.94
Dissolved oxygen (%)	77.44	90.04	96.33	94.55	103.40	119.18
Turbidity (NTU)	1.04	2.27	5.96	4.29	7.90	42.95
Chlorophyll ($\mu\text{g}/\text{L}$)	0.89	1.58	2.71	2.40	3.41	8.73
<i>River Flow</i>						
Flow (m^3/s)	7.35	9.99	30.19	19.82	35.79	102.64

Table 2: Correlation coefficients between pairs of measured data.

	Nitrate	Phosphate	Temperature	Conductivity	Dissolved oxygen %	Turbidity	Chlorophyll
Nitrate	1						
Phosphate	-0.16	1					
Temperature	0.06	0.23	1				
Conductivity	0.31	0.20	0.92	1			
Dissolved oxygen %	0.24	-0.39	0.44	0.50	1		
Turbidity	0.16	0.30	-0.13	-0.20	-0.62	1	
Chlorophyll	0.02	-0.30	-0.11	-0.27	-0.02	0.25	1

Table 3: Goodness of fit term (G), penalty term (P) and overall predictive model choice criterion (G+P).

Models	Nitrate			Phosphate		
	G	P	G+P	G	P	G+P
M1. Penalised spline for water quality data + change point structures	1.31	3.76	5.07	33.84	42.61	76.45
M2. Linear model for water quality data, no change point structures	3.16	5.33	8.49	66.79	71.42	138.20
M3. Linear model for water quality data + a switchpoint in time	3.01	5.14	8.16	73.50	78.58	152.08
M4. Linear model for water quality data + a change threshold in flows	2.55	4.70	7.26	67.72	72.42	140.14
M5. Linear model for water quality data + change point structures	1.55	3.98	5.53	44.24	50.67	94.91
M6. Penalised splines for water quality data, no change point structures	2.35	4.75	7.09	50.24	57.91	108.15

Table 4: Parameter estimations.

Parameters	Nitrate		Phosphate	
	Median	95%CI	Median	95%CI
<i>Change point structures</i>				
τ (Switchpoint in time, occurring in the year 2014)	08/03	(05/03, 13/03)	24/01	(22/01, 28/01)
φ_1 (Change threshold in flow before switchpoint in time)	27.87	(16.26, 43.64)	4.19	(2.07, 4.61)
φ_2 (Change threshold in flow after switchpoint in time)	10.64	(7.65, 13.41)	3.54	(3.37, 3.70)
δ_1 (Slope for low flow before switchpoint in time)	-0.80	(-0.63, -0.88)	0.07	(-0.63, 1.26)
δ_2 (Slope for high flow before switchpoint in time)	1.17	(1.09, 1.32)	0.06	(-0.74, 1.94)
δ_4 (Slope for high flow after switchpoint in time)	1.21	(1.16, 1.27)	-1.25	(-1.59, -0.97)
<i>Penalised splines</i>				
β_0 (Global intercept)	4.74	(4.31, 5.18)	0.05	(0.04, 0.09)
β_1 (Fixed effect for temperature)	-0.18	(-0.29, -0.07)	-0.01	(-0.41, 1.25)
β_2 (Fixed effect for conductivity)	0.29	(0.17, 0.40)	0.37	(-0.14, 0.87)
β_3 (Fixed effect for dissolved oxygen)	0.06	(-0.03, 0.22)	-0.47	(-1.13, -0.07)
β_4 (Fixed effect for turbidity)	0.10	(0.05, 0.16)	0.15	(-0.09, 0.51)
σ_{b_1} (Standard deviation for spline on temperature)	0.02	(0.01, 0.71)	0.12	(0.03, 0.71)
σ_{b_2} (Standard deviation for spline on conductivity)	0.03	(0.02, 0.06)	0.14	(0.05, 0.45)
σ_{b_3} (Standard deviation for spline on dissolved oxygen)	0.03	(0.02, 0.08)	0.17	(0.04, 0.61)
σ_{b_4} (Standard deviation for spline on turbidity)	0.03	(0.01, 0.07)	0.07	(0.01, 0.44)
<i>Other</i>				
σ^2 (Measurement error variance)	0.01	(0.00, 0.01)	0.10	(0.08, 0.12)

Figure 1: Map of the study area.

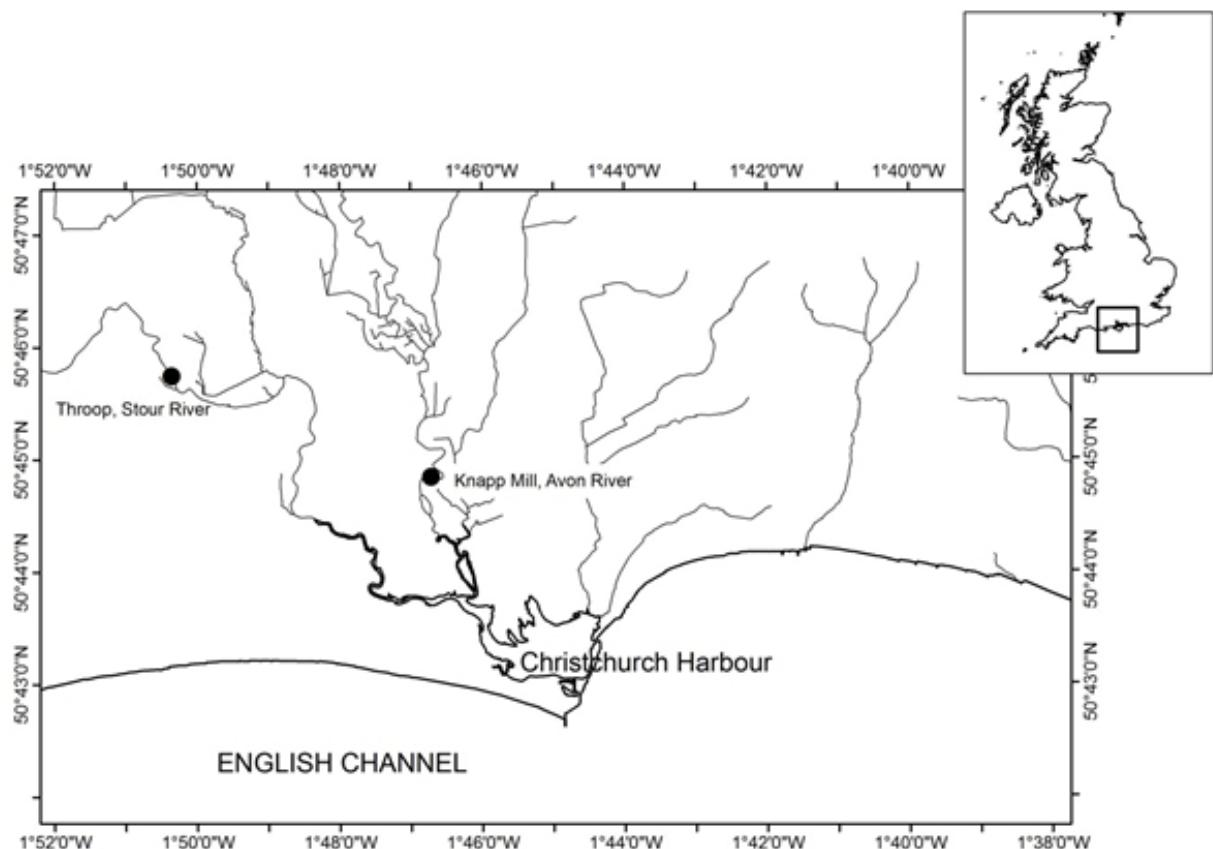


Figure 2: Daily macronutrient and river flow data (22/11/2013 to 19/12/2014). Data are plotted on original scale: nitrate (solid line) in mg/L , phosphate (dashed line) in mg/L and river flow (dotted line) in m^3/s .

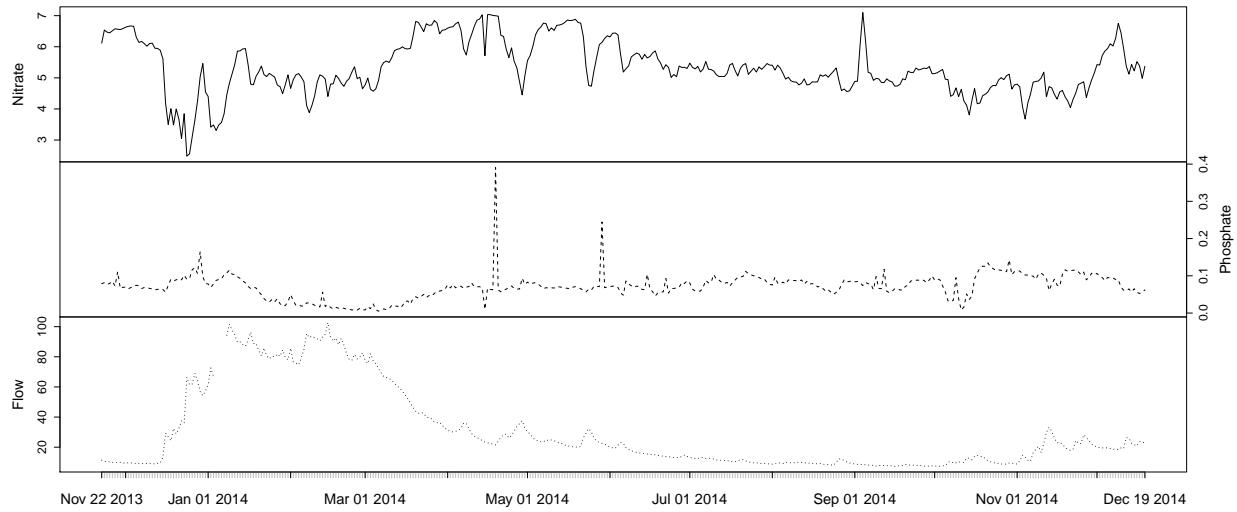


Figure 3: Daily water quality data values (22/11/2013 to 19/12/2014). Data are plotted on original scale: temperature (solid line) in $^{\circ}C$, conductivity (dashed line) in $\mu S/cm$, dissolved oxygen saturation (dotted line) in %, turbidity (dotdash line) in NTU and chlorophyll (longdash line) in $\mu g/L$.

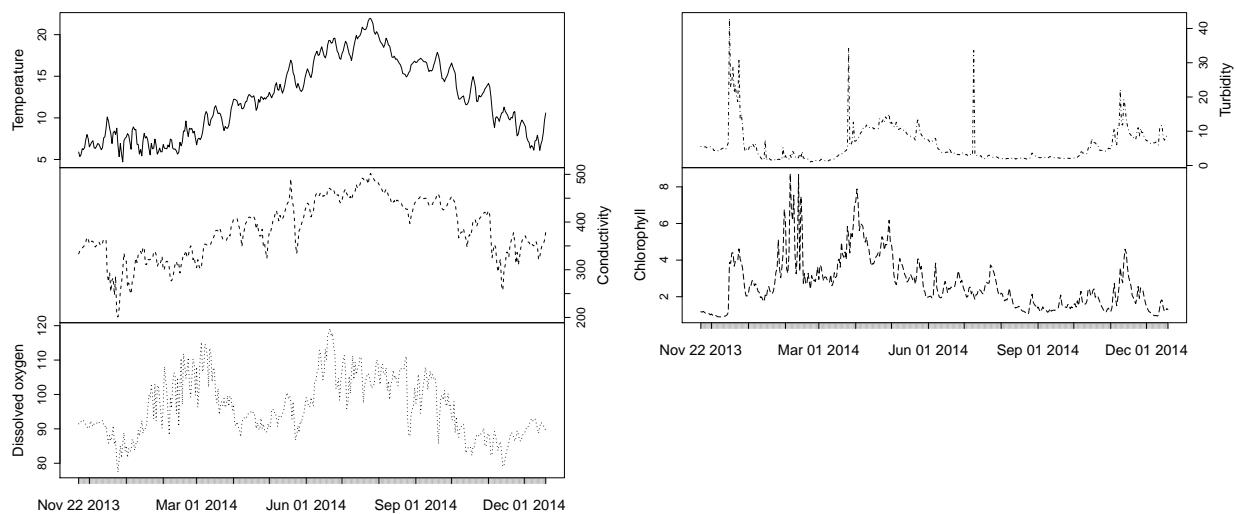


Figure 4: Scatterplot of macronutrient data *versus* water quality properties. Data are plotted on original scale: nitrate in mg/L , phosphate in mg/L , temperature in $^{\circ}C$, conductivity in $\mu S/cm$, dissolved oxygen (DO) in %, turbidity in NTU and chlorophyll in $\mu g/L$.

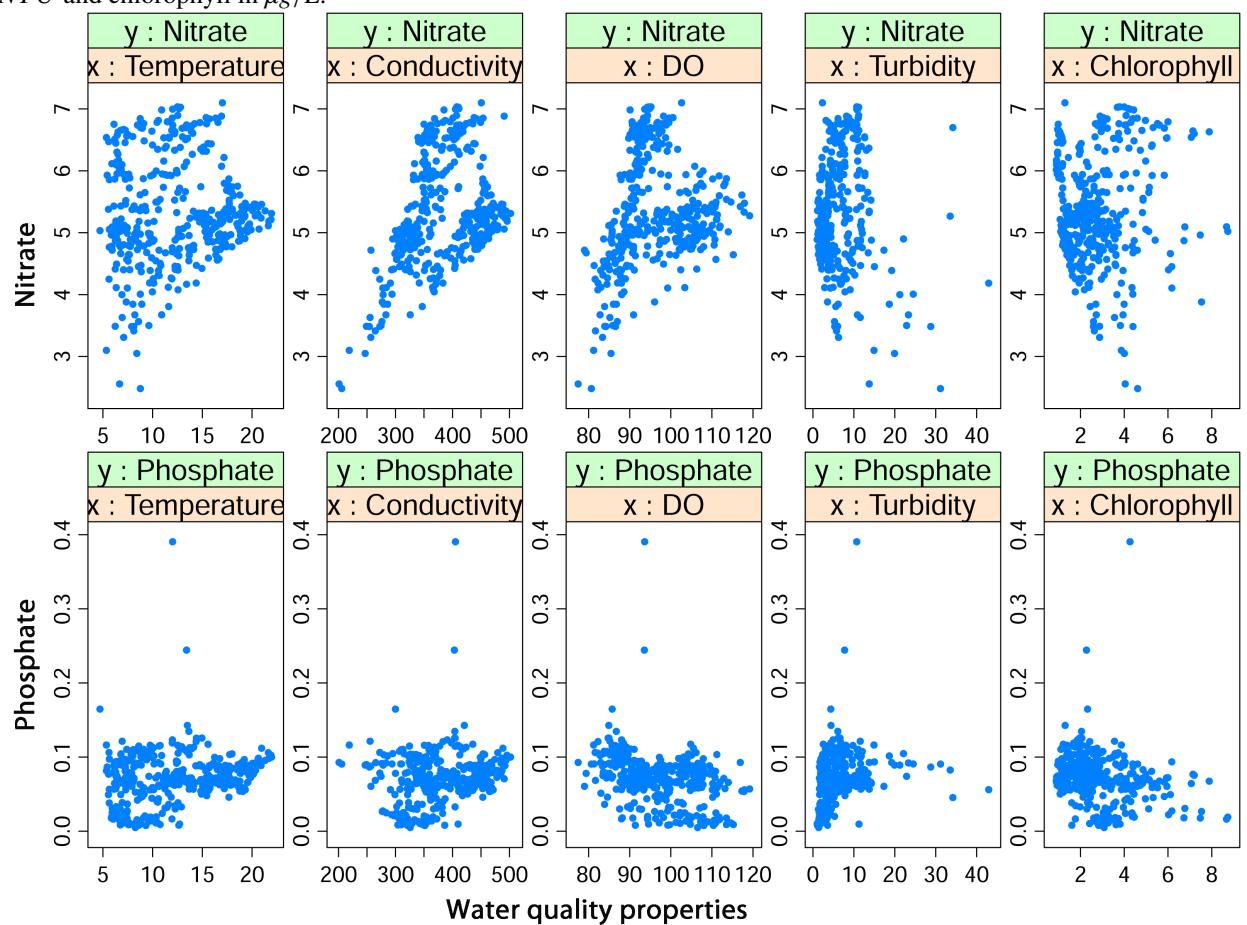


Figure 5: Directed Acyclic Graph (DAG) for the model M1 (Note that $z_{tjk} = (x_{tj} - k_{kj})_+^d$).

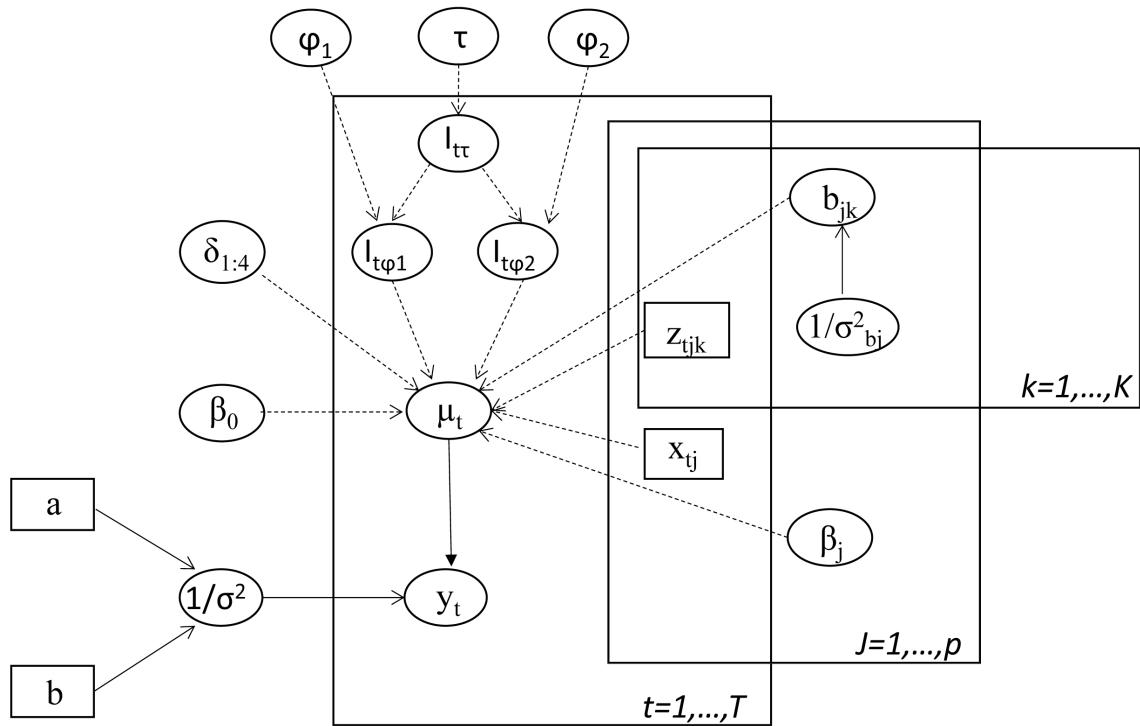


Figure 6: Time-series plot of the residuals for (a) nitrate and (b) phosphate.

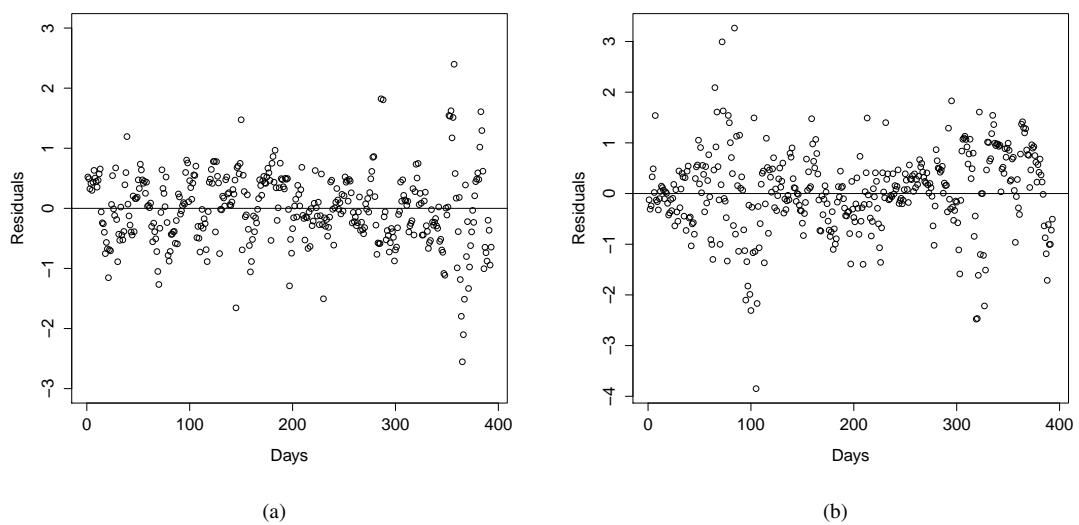


Figure 7: Relationship between nitrate and river flow using the estimated parameters for the change point structures. Data are plotted on original scale: nitrate in mg/L , and river flow in m^3/s .

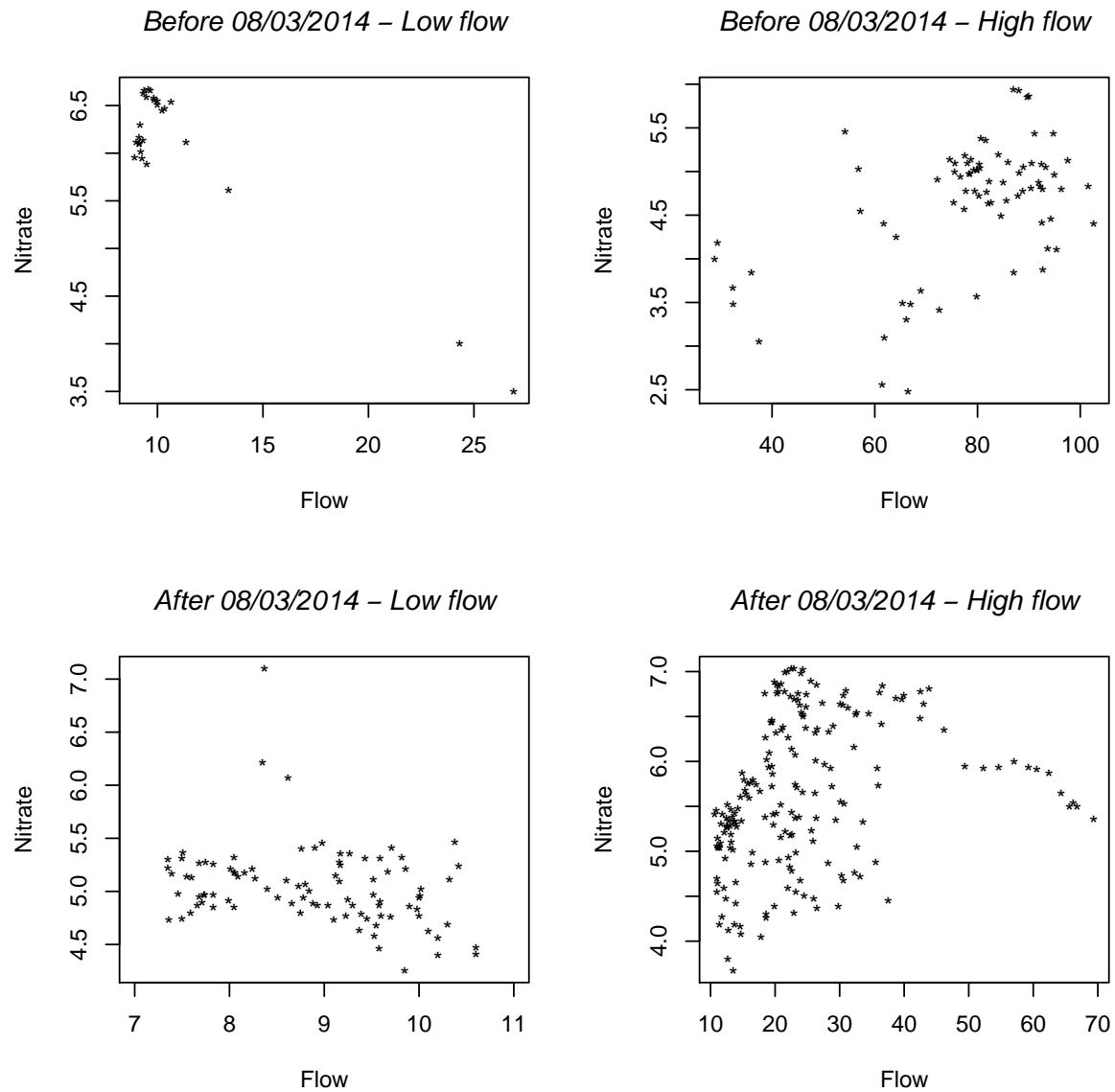


Figure 8: Relationship between phosphate and river flow using the estimated parameters for the change point structures. Data are plotted on original scale: phosphate in mg/L , and river flow in m^3/s .

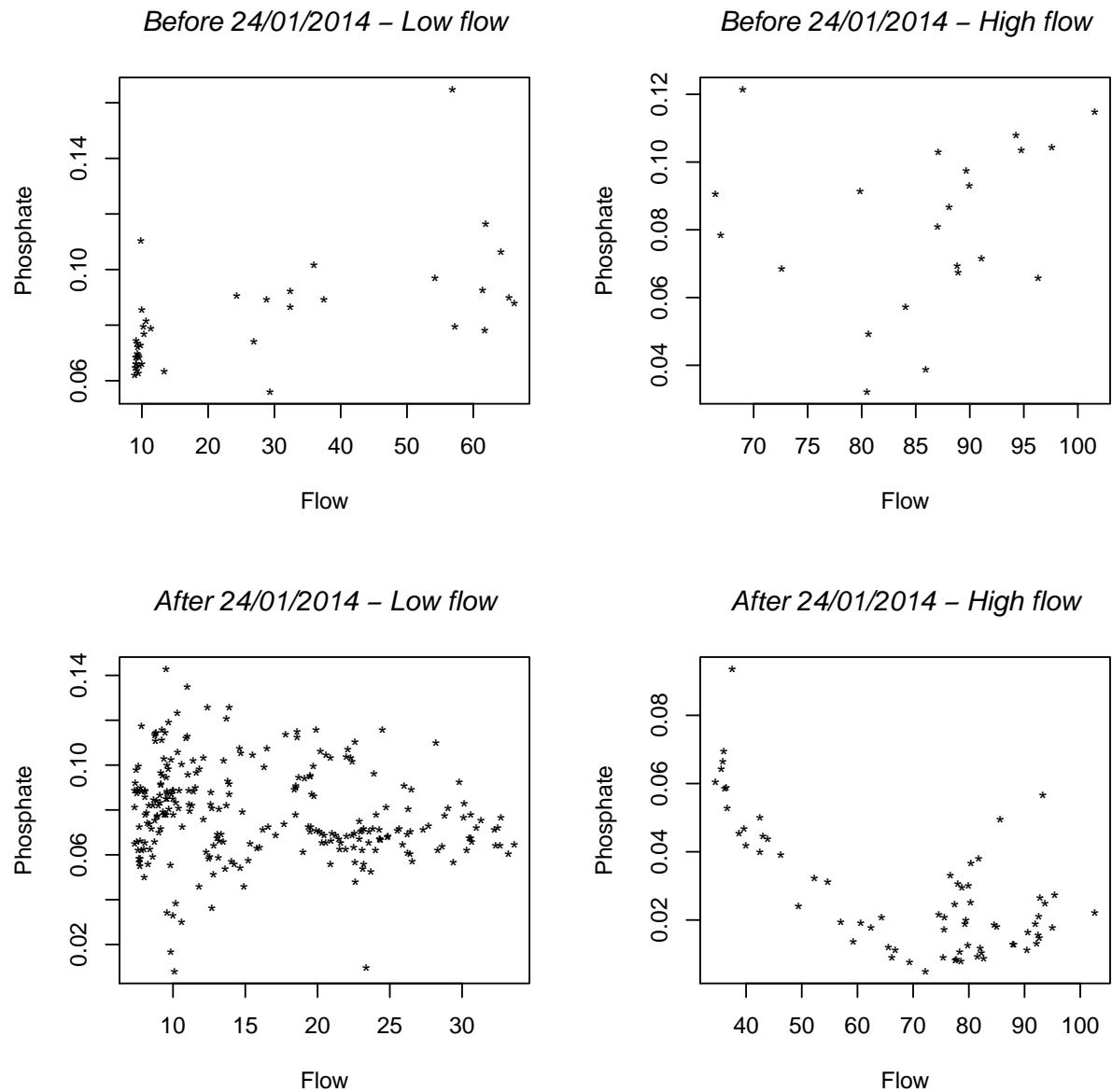


Figure 9: Plots of daily fluxes (Kg/Day) for (a) nitrate and (b) phosphate. Red rectangle on the right-hand of the panel contains the predicted fluxes from the model M1 for the period 30/11/2014 to 19/12/2014.

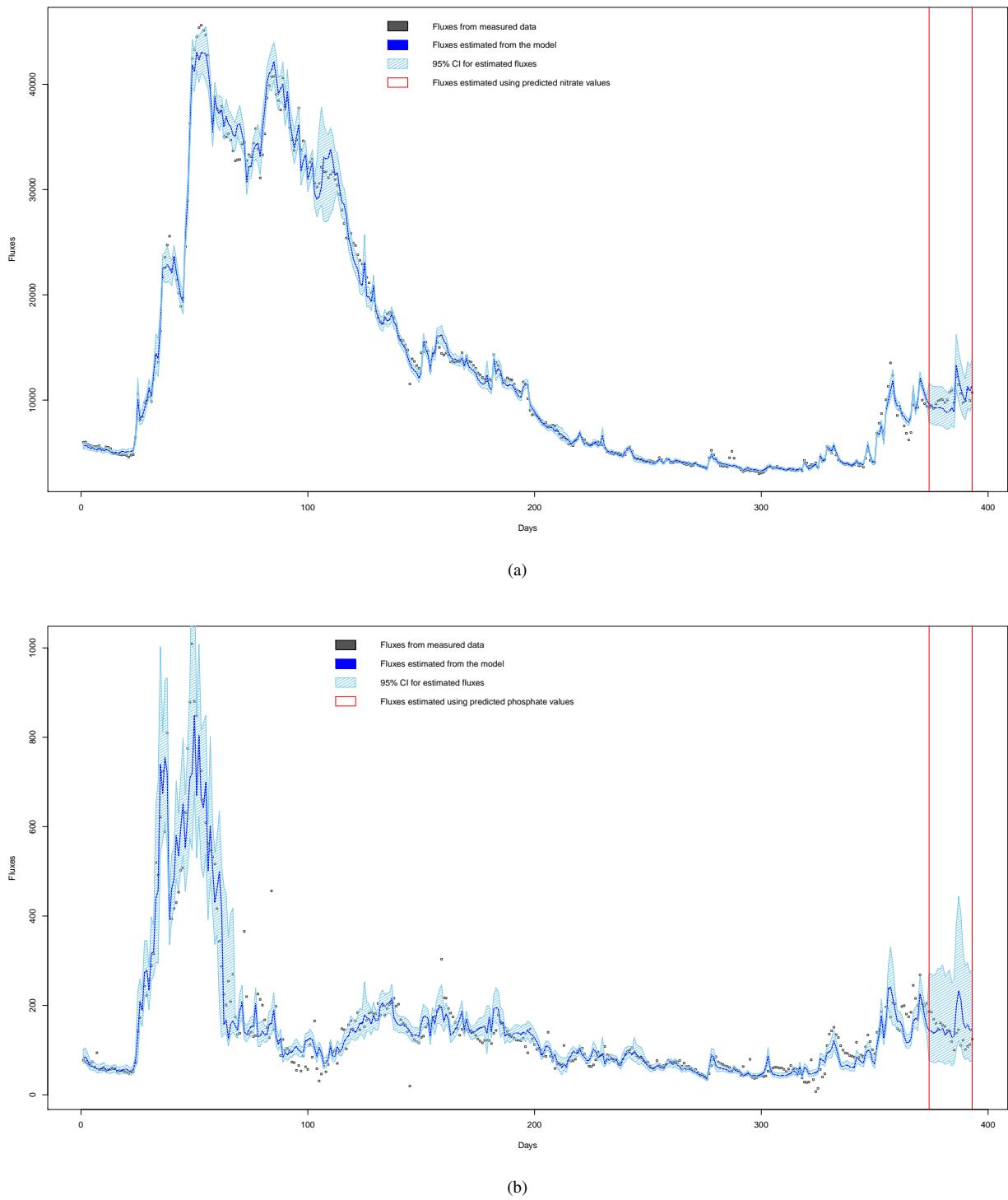


Figure 10: Nitrate fluxes (Kg/Day) using the estimated parameters for the change point structures. Black dots are the observed fluxes, black solid line represents the fluxes estimated from the model M1 and shadow area represents the 95% CI for the estimated fluxes.

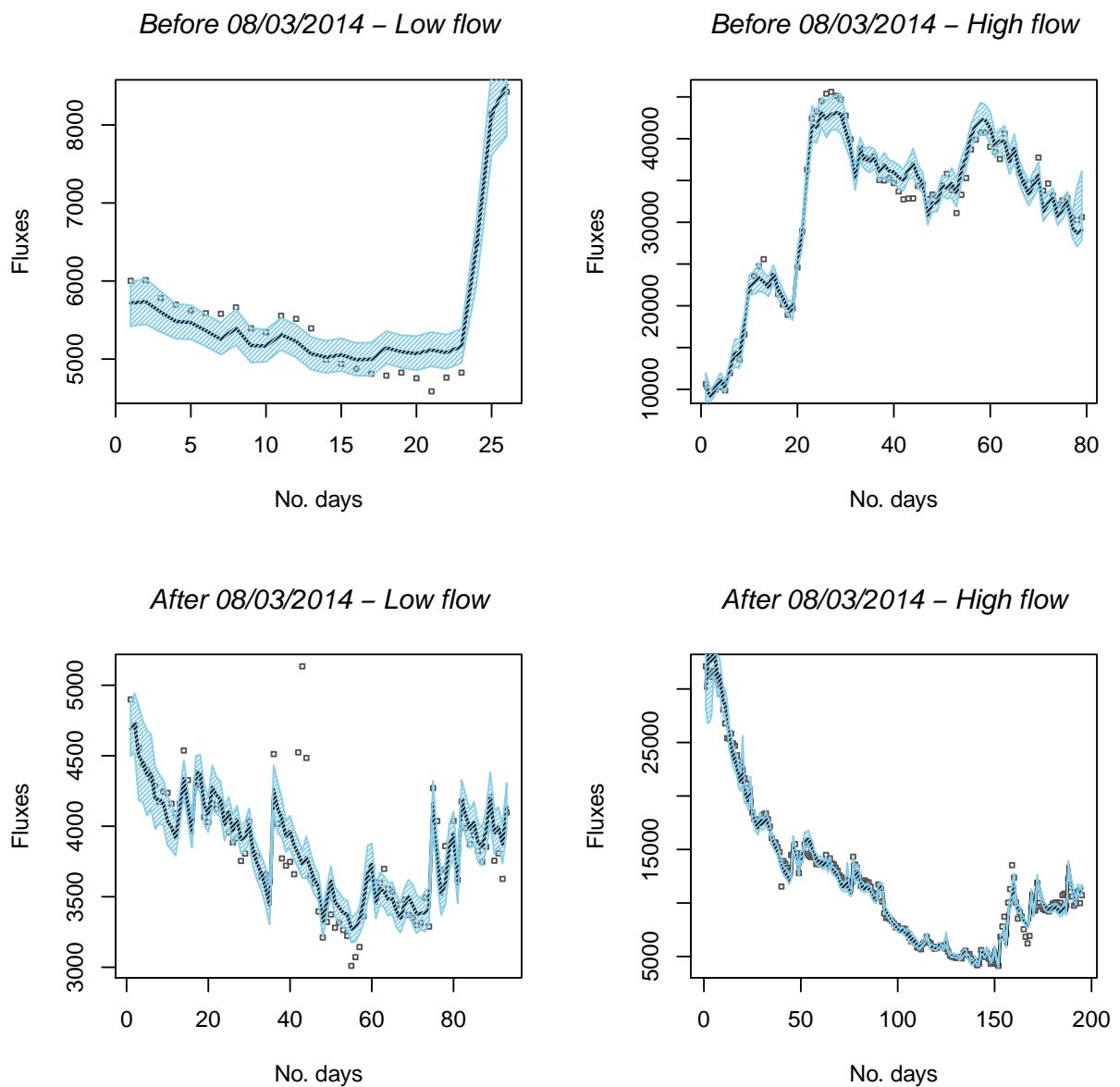


Figure 11: Phosphate fluxes (Kg/Day) using the estimated parameters for the change point structures. Black dots are the observed fluxes, black solid line represents the fluxes estimated from the model M1 and shadow area represents the 95% CI for the estimated fluxes.

