Modelling macronutrient dynamics in the Hampshire Avon river: A Bayesian approach to estimate seasonal variability and total flux

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 affected by weather pattern and extreme rainfall events. As a result, the nutrient budget in the receiving estuaries and coasts can change suddenly and seasonally, causing ecological damage to resident wildlife and fish populations. In this paper, we propose a statistical change-point model with interactions between time and river flow, to capture the macronutrient dynamics and their responses to river flow threshold behaviour. It also accounts for the nonlinear effect of water quality properties via nonparametric penalised splines. This model enables us to estimate the daily levels of riverine macronutrient fluxes and their seasonal and annual totals. 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 meteorological conditions leading to localised flooding. Adopting a Bayesian inference framework, we have quantified riverine macronutrient fluxes based on input river flow values. Out of sample empirical validation methods justify our approach, which captures also the dependencies of macronutrient concentrations with water body characteristics.

Keywords: Change-point analysis, Bayesian inference, Macronutrients, Fluxes, River flows, Water quality properties.

1. Introduction

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

In addition to these disturbances, natural features of the environment and climate changes can compromise macronutrient cycles in fresh waters (Woodward et al., 2010; Whitehead and Crossman, 2012). Climate changes are likely to impact on weather pattern and bring an increasing number of extreme events, including increased frequency and intensity of storms, leading to high winds and heavy rainfall. These events might cause changes in the process that governs macronutrient behaviour.

In this paper, we are concerned with the dynamics of riverine nitrate and phosphate concentrations and their response to rapid changes in river flow and water quality properties, such as temperature, conductivity, dissolved oxygen and turbidity. River flow alteration is indeed an environmental factor to which riverine ecosystems respond considerably (Poff and Zimmerman, 2010; Rolls et al., 2012). Rapid changes in river flow are greatly driven by extreme weather pattern and events such as storms, which can impact on the macronutrient runoff and budget, but which ecological importance is not well documented (Leigh et al., 2014). Inclusion of the effect of variations in river flow on riverine macronutrient dynamics provides new insights into the dynamics of the macronutrient fluxes from the river to the estuary.

*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)

Our study is based on measurements collected at the Knapp Mill gauging station on the River Hampshire Avon in the UK under the *Macronutrient Cycles Programme* funded by Natural Environmental Research Council (NERC). In particular, we use macronutrient and river flow data, as well as water quality properties, from the Hampshire Avon 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.

We adopt a novel change-point approach, within a Bayesian hierarchical structure, which results in a generalized additive model that is able to: (i) differentiate the effect of changes in river flow on nitrate and phosphate according to the time of year in which they occur, and (ii) capture the complex nonlinear relationships among macronutrients with the water quality properties through unspecified smooth functions of these properties. The resulting model allows us to estimate the annual total flux of the modelled micronutrients from the river to the downstream estuary, often called the annual budget, with quantified uncertainties.

The hierarchical model detects changes in micronutrient dynamics by simultaneously estimating change-points in river flow and also a change-point in time. Thus the changes in river flow rate, that we call change thresholds following the terminology adopted by Alameddine et al. (2011), can be different according to the period of the year in which they occur. Hence, the model accommodates a temporal window introducing a possible shift in time, that we simply call switch-point, to distinguish it from the terminology used to refer to changes in river flow. Henceforth, we call the full statistical model identifying the switch-point in time and the change thresholds in river flow simply as change-point structures.

Change-point analysis has become a popular tool in ecological studies and in the simplest form it detects distributional changes within observations that are ordered in time. Its use in water quality models resulted in very interesting contributions. For example, Fortin et al. (2004) reformulated the shifting-level model to show that it belongs to the class of hidden Markov models, and developed Bayesian methods for evaluating its parameter uncertainty and demonstrated its utility for forecasting of streamflows. Alameddine et al. (2011) used a change-point and threshold

model to analyse the relationship between nitrogen concentrations and flow regimes during a long period of 29 years, quantifying changes in this relationship across time.

The modelling approach proposed in this paper, is similar in spirit to the Bayesian model of Alameddine et al. (2011), as we describe the macronutrient dynamics through change-point structures, modelling their locations as unknown. However our model formulation is different from the Alameddine et al. (2011) contribution, as (i) we focus on a short study period, (ii) we use interaction terms between time and river flow to capture a potential seasonal behavior in freshwater, that is known to be an important determinant when considering macronutrient loadings (Sigleo and Frick, 2003; Laud and Ibrahim, 2008), and (iii) we include different physico-chemical water properties without imposing any parametric form (e.g. linear) in their relationship with macronutrients.

2. Methods

2.1. Study area

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

The sampling site for this study is located at the lowest water flow gauging station on Hampshire Avon at Knapp Mill (latitude: 50.74807, longitude: –1.77968), encompassing a catchment area of 1706 km^2. Fig. 1 provides a map of the study area.

2.2. Macronutrient and water quality samples

Sampling at Knapp Mill was carried out between 22 November 2013 and 19 December 2014. Water quality properties, including temperature, conductivity, dissolved oxygen, turbidity and chlorophyll concentration, were measured *in situ* every 10 minutes using an EXO2 multiparameter sonde (Xylem, UK). Samples for macronutrient analysis were collected every 8 to 15 hours with an ISCO automated water sampler (RS Hydro, UK). Water samples were fixed immediately with 0.015M mercuric chloride (750 μL in 150 mL) and later filtered through a glass fibre filter

upon return to the laboratory. Concentrations of inorganic macronutrients were determined at the University of Portsmouth using a QuAAtro segmented flow nutrient analyser (SEAL Analytical, UK). River flow data were obtained from the UK Environment Agency. To regularise the sampling intervals between measurements, the 24-hour (daily) means were calculated and used for further analyses.

2.3. Exploratory analysis

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

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

2.3.1. Data pre-processing

The various measured water quality properties have a range of different units, therefore for modelling purposes these are standardised to have zero mean and unit variance. This procedure

makes the magnitude of the coefficients comparable. Macronutrient concentrations and river flow data are modelled on logarithmic scale to stabilise their variance. Moreover, logarithmic transformation of the data is convenient for macronutrients, as they are nonnegative and their distributions are often skewed to the right.

2.3.2. Selection of water quality properties

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

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

3. Model set-up

The discussion in the previous section leads us to consider a regime switching model for macronutrients according to both temporal window and river flow that is able to adjust for non-linear effects of the chosen water quality properties. Here the two macronutrients, nitrate and phosphate, are modelled separately, although it is possible to model them jointly. Joint modelling of the two macronutrients nitrate and phosphate, is not of interest here since our objective is not to study their inter-relationships, which is seen to be rather weak (correlation –0.16 in Table 1), but to predict their individual daily and annual fluxes into the estuary.

The model is developed for data y_t, which denotes the natural logarithm of the observed macronutrient concentration at day t, for $t = 1, \ldots, T = 393$. We construct a Bayesian hierarchical model, which encompasses the model for the observed data, the dynamics of the process and the specification of parameters and hyperparameters (Berliner, 1996). At the first stage of the modelling hierarchy, we assume an independent Gaussian measurement error model:

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

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

The second stage of the hierarchy defines the model for μ_t. To incorporate nonlinear effects of each of the p water quality properties, we incorporate a nonparametric smoothing function $g_j(x_{tj})$ of x_{tj} at each time point t, where x_{tj} denotes the value of the jth water quality property at the tth time point. The choice of the $g_j(\cdot)$ functions ranges from linear to nonparametric penalised splines (Eilers and Marx, 1996; Ruppert et al., 2003) which are well-known to be very flexible. In our implementation, following Crainiceanu et al. (2005), we construct the splines using radial basis functions, which provides a more stable fit than traditional truncated linear basis. By denoting x to be a generic covariate, we define a set of K knots, $k_1 < k_2 < \cdots < k_K$ taken to be equally spaced

over the range of x. We consider a low-rank thin-plate spline representation given by:

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

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

Model (2), assumed for the jth covariate at tth time point, x_{tj}, is given by:

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

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:

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

However, the p separate intercept terms will not be identified and hence we only take one global 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 $j = 1, \ldots, p$. Now, each b_{kj} for $k = 1, \ldots, K$ and $j = 1, \ldots, p$ is given an independent normal prior distribution with mean zero and unknown variance, σ_b^2 as mentioned above.

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

Now we turn to modelling the step changes in nitrate and phosphate concentrations due to temporal changes and discontinuities river flow. The exploratory analysis in Section 2.3 has made it clear that the nutrient concentration, especially nitrate, is severely impacted upon by not only river flow but also seasonality. However, it is likely that variations in river flow will have different

effects on concentration in different temporal windows. Moreover, natural rain fall, and hence river flow, does not strictly adhere to the calendar dates. That is why, we let the parameter τ denote the switch-point in time that serves as the unknown boundary between the end of the winter high flow season and the start of the low flow season spanning the rest of the year. Since τ is unknown we estimate it along with all other parameters. To allow for interactions between seasonal windows and river flow levels we imagine that there are two change thresholds in river flow which occur once during the winter and the other during the rest of the year. Let φ_1 and φ_2, denote these flow threshold parameters. Hence, we introduce the following four terms in the model:

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 switch-point in time,

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 the switch-point in time,

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 switch-point in time,

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 the switch-point in time,

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

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

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

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

- M2. A linear regression model for the water quality properties, but with no change-point structures, that allows us to compare the proposed modelling innovations with a straw method:

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

- M3. A linear regression model for the water quality properties, with only a switch-point in time:

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

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

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

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

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

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

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

To account for temporal dependence that is expected to occur between measurements collected on consecutive days, we also evaluated the additional inclusion in (4) of a random intercept, modelled as a linear stationary first-order autoregressive process, η_t, which is a very popular choice in 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 is white noise, that is normally distributed with mean 0 and variance σ_η^2, and the parameter ρ is assumed be in the interval $[-1, 1]$. However, we were not able to fit this model to the data due to lack of identifiability.

The Bayesian model is completed by assuming prior distributions for all the unknown parameters. We assume that the switch-point in time, τ, is uniformly distributed on $[1, 2, \ldots, T]$. Note

that $\tau = 1$ and $\tau = T$ does not imply any change. We also assess a discrete uniform prior for the switch-point in time, that showed lead to a better convergence for the phosphate model, though requiring a higher computational effort. Similarly, we adopt uniform prior distributions for the two change thresholds on river flow φ_1 and φ_2, in the interval $[1.995, 4.631]$, which are the minimum and maximum values of the river flow on the logarithmic scales. The precision parameters (i.e. inverse of the variance parameters) specific for each macronutrient, σ^{-2}, are assumed to follow a Gamma distribution $\text{Ga}(c, d)$ independently, with shape parameter, c, and expectation, c/d. In particular, we assume a proper prior specification by taking $c = 2$ and $d = 1$ for these parameters. We assume normal prior distributions for β_0 and the fixed effect parameters $\boldsymbol{\beta}$ specified as $\text{Normal}(0, 10^4)$. Moreover, as previously mentioned, an independent normal prior distribution, centered at zero, is chosen for the random effects parameters $\mathbf{b}$ associated with the splines for the water quality properties. For σ_b^2, which controls the amount of smoothness of the water quality properties, we consider two different prior distributions: (i) a widely accepted Gamma 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 half-Cauchy distribution for the standard deviation parameter, $\sigma_b \sim \text{half-Cauchy}(A)$, with $A = 25$ as suggested by Marley and Wand (2010). Using a half-Cauchy, in fact, we can restrict the standard deviation parameter, σ_b, to be away from very large values (Gelman, 2006), that could bias the distribution against zero. By comparing model fits under both of these prior distributions, we adopt the first parameterization for the nitrate model and the latter for the phosphate model. Finally, we assume a normal prior distribution for the $\boldsymbol{\delta}$ parameters associated with the change-point structures.

Fig. 5 presents the Directed Acyclic Graph (DAG) of our more complex model M1, that is a simplified graphical representation of the hierarchical modelling structure. In this graph each quantity is represented by a node and links between nodes show direct dependence. The ellipses represent stochastic nodes (that is, variables that have a probability distribution) or logical nodes (that is, deterministic relationships or functions). The small squares identify nodes that are constants. Stochastic dependence and functional dependence are denoted by solid and dashed arrows, respectively. Finally, the large square plates represent repetitive structures (i.e. the 'loop' from $t = 1$ to T).

To compare the quality of the model fit of the proposed modelling approach in comparison

to the above described simpler statistical models, we adopt the predictive model choice criterion (PMCC; Laud and Ibrahim (1995); Gelfand and Ghosh (1995)) defined by:

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

where y_t^{rep} denote the future replicate of the observed macronutrient concentrations y_t^{obs}. The PMCC essentially quantifies the fit of the model by comparing the posterior predictive distribution 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 goodness of fit measure (G) which will decrease with increasing model complexity and the second term of (10) is a penalty term (P) which tends to be larger for complex models. The model with the smallest value of PMCC is the preferred model.

To facilitate model comparisons using the traditional coefficient of determination (often termed as adjusted R^2), we consider the analogus Bayesian statistic $R_B^2 = 1 - \frac{\sigma^2}{S_Y^2}$, where S_Y^2 is the sample variance of Y (i.e. the macronutrient concentrations) and σ^2 is the model variance (Ntzoufras, 2009). The R_B^2 quantity can be interpreted as the proportional reduction of uncertainty concerning the macronutrient concentrations, Y, achieved by incorporating in the model the water quality properties and the change-point structures. Alternatives to the R_B^2 are the estimating model skill methods proposed by Jolliff et al. (2009), and the traditional Nash-Sutcliffe calculation (see e.g. Krause et al., 2005). However, these are not considered any further in the paper. Instead, we use simple to interpret and use out-of-sample validation tests as noted below in Section 3.2.

3.1. Computation

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

The implementation of the models has been performed using the freely available software package WinBUGS (version 1.4.3; Lunn et al. (2000)), that was executed in batch mode using the R library R2WinBUGS (version 2.1-19; Sturtz et al. (2005)). WinBUGS code for the model

M1 is available in the Supplementary material. We have run two parallel MCMC chains independently starting at two very different initial values for 50,000 iterations with 20,000 burn-in, and we thinned the Markov chains by a factor of 10, resulting in samples of size 6,000 to estimate the posterior distributions for the parameters of interest. Convergence was assessed by checking the trace plots of the samples of the posterior distribution and the estimated autocorrelation functions and the Monte Carlo standard errors.

3.2. Prediction and estimation of macronutrient fluxes

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

The Bayesian methods allow us to estimate the daily total deposit (mass flux) of each macronutrient as follows. Note that macronutrient flux is defined as the product of concentration times river flow rate (Sigleo and Frick, 2003; Quilbé et al., 2006), measured in Kg/day, i.e. flux at day t, denoted by ξ_t is $\mu_t \times f_t$ where μ_t is converted to be measured in Kg/m^3 and river flow is converted in m^3/day. We estimate ξ_t and its uncertainty by using $\xi_t^{(\ell)} = \mu_t^{(\ell)} f_t$ where $\ell = 1, \ldots, 6000$ indexes the thinned MCMC iterates.

We predict macronutrient fluxes for the 20 days used in the out-of-sample validation test to assess the predictive accuracy of the model. We also estimate daily and total fluxes for the entire study period using the whole data set available. Finally, to allow a comparison with similar literature contributions, we quantify the annual macronutrient fluxes, from 22/11/2013 to 21/11/2014, computing catchment-normalised estimates (that is, our estimated annual macronutrient fluxes are divided by the total area of the catchment).

4. Results

Table 3 presents the values of the PMCC and the Bayesian statistic R_B^2 that inform us about the quality of the fit and predictive abilities of each model. From this analyses, we are able to judge the worth of each of the modelling strategies: change-point structures, penalised splines and linear regression model for the water quality properties. Model M5 based on the linear regression model for the water quality properties provides almost equal performance but shows a worse goodness-of-fit as expected, since the spline based models are more flexible. Interestingly, the straw method based on simple linear model, M2, without any modelling innovation does not perform well as expected. We also note that both the PMCC and R_B^2 choose the same model M1 as the best model, which is adopted henceforth in this paper.

To assess the adequacy of the chosen model M1 for the macronutrients data, we have checked the residuals plots. Fig. 6 illustrates the median of the posterior distributions of the standardised residuals plotted against the time period for nitrate and phosphate. No discernible pattern is present for nitrate, with a random scattering of points. For phosphate, the residuals scatter around zero randomly with a few large values. This result supports an overall adequacy of the model for the data.

4.1. Parameter and flux estimates

Parameter estimates for the chosen model M1 are presented in Table 4. The switch-point in time for nitrate, estimated to occur on 08/03/2014 (95%CI: 05/03/2014, 13/03/2014), identifies essentially two seasonal periods that are, clearly, winter and summer times. The change thresholds before and after this switch-point captures two regimens in river flow, occurring at 27.87 m^3/s (95%CI: 16.26, 43.64) and 10.64 m^3/s (95%CI: 7.64, 13.41) in winter and summer times respectively. Taking low flow conditions in summer as reference category, the results suggest that a higher level of river flow in winter, as well as in summer, is associated with increased concentrations of nitrate, such that a difference of 1 in river flow corresponds, on original scale, to an increase in nitrate of about 1.17 mg/L in winter, and about 1.22 mg/L in summer.

Phosphate shows a considerable different change-point structure, with a no clearly identifiable seasonal variation. The switch-point in time for phosphate is estimated to occur on 24/01/2014

(95%CI: 22/01/2014, 28/01/2014). Because of this early identification of the switch-point in time, during which the Hampshire Avon is still experiencing extremely high flow levels, the associated estimation of the change thresholds in river flow lacks of precision. This is clearly showed in a larger uncertainty in the estimation of the change threshold parameters, occurring at 66.35 m^3/s (95%CI: 7.89, 100.28) before the switch-point in time and at 34.29 m^3/s (95%CI: 29.19, 40.32) after the switch-point in time. The increase in phosphate before the switch-point in time, associated with high river flow is not significant, however after the switch-point in time, a higher level of river flow seems associated with a dilution of phosphate of about 0.29 mg/L.

Figs. 7 and 8 show the different change-point regimes in the macronutrient dynamics and river flow as estimated by the model. Between regime variations in macro-nutrients can be clearly seen from these two graphs, although the variations are more pronounced in the case of nitrate than phosphate as expected.

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

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

We also estimate the total macronutrient fluxes from the complete model, according to the estimated parameters for the change-point structures. We find strong seasonal effects in the riverine nitrate fluxes as shown in Fig. 10. For example, in winter time low-flow conditions (that is, before the 08/03/2014) the mean of the daily observed nitrate fluxes is 5,552 Kg/Day (that correspond to an estimated daily posterior mean of 5,531 Kg/Day from our model), while in winter time

high-flow conditions, the mean increases to 31,696 Kg/Day (that correspond to an estimated daily posterior mean of 31,668 Kg/Day from our model). The seasonal structure is not so clear in the model for phosphate. From Fig. 11 we can see that most of the days (no. 260), occurring after the 24/01/2014, are classified as low-flow conditions. However, we can still estimate the effect of high flow caused by extreme rainfall events in the model for phosphate. For example, before 24/01/2014, the mean of the daily observed nitrate fluxes in low-flow conditions is 197.18 Kg/Day (that corresponds to an estimated daily posterior mean of 208.68 Kg/Day from our model), and the mean in high-flow conditions is definitively higher, being equal to 605.7 Kg/Day (that corresponds to an estimated daily posterior mean of 573.0 Kg/Day from our model).

Finally, Table 5 presents the posterior median estimates and 95%CI for the catchment area normalised annual total nitrate and phosphate fluxes in Kg/Km^2 for the year from 22/11/2013 to 21/11/2014, according to our best model M1 and the linear regression model, M2. We can observe that the flux estimates under model M2 are lower than the corresponding estimates under model M1. However, the estimates under M2 have higher uncertainties as seen by comparing the lengths of the 95% intervals. Hence the estimates under model M1 are seen to be more accurate than those under M2 and, hence, the former model continues to be our preferred model.

The last two rows of Table 5, respectively, provide estimates of mean annual fluxes for the Hampshire Avon at Knapp Mill reported by Jarvie et al. (2005) and the UK wide average reported by Nedwell et al. (2002). Our estimates are of broadly similar magnitude to both of these sets of estimates. However, for nitrate we note that our estimate is for a very unusual year with exceptionally high rainfall leading to higher nitrate fluxes. Regarding phosphate, both the other estimates are for dates which are more than two decades in the past and during these last two decades sewage treatment works have improved significantly reducing phosphate levels. In addition to these points, there are substantial differences in the methodologies used to calculate the fluxes. Our estimates are based on a detailed model based calculation of daily concentrations and river flow that takes advantage of sudden changes in flow levels. On the contrary, the estimates reported by Jarvie et al. (2005) and Nedwell et al. (2002) are based on simple calculations of monthly average concentration levels and monthly average flow levels which are likely to miss peaks and troughs, and seasonality in the deposition levels.

5. Discussion and conclusion

The principal aim of this paper consists in understanding how different macronutrient species respond to changes in river flow, which can be largely driven by weather pattern and severe weather conditions such as storm events. Therefore, we develop a model for riverine data collected during a relatively short study period characterised by unusual frequency of storms and heavy rainfall. We propose to describe the inter-annual variability of these macronutrient species using a Bayesian modelling approach featured by an interaction between temporal window and river flow via change-point structures. It is also complemented by a flexible representation of the effect of the water quality properties, that are modelled free from parametric constrains. In the application considered in the paper, we observe that the change-point structures better depict the temporal behaviour of riverine nitrate and phosphate, and that nonparametric spline based model outperforms the standard multiple linear regression model. This is coherent with Walther (2010), who also noted that the relationships among the components of ecological systems are complex and that interactions and feed-back mechanisms can lead to nonlinear and abrupt changes.

The identification of switch-points or threshold behaviour in hydrological processes is indeed an active area of research, which is, in a growing number of examples, accomplished within a Bayesian modelling framework (e.g. Fortin et al., 2004; Alameddine et al., 2011; Jo et al., 2016). Recently, the importance of encapsulating environmental thresholds behaviour has been also raised in context of hydrological process-based models. Wellen et al. (2014) proposed, for example, a Bayesian approach for accommodating environmental thresholds and extreme events within SWAT models, assuming that the watershed response to precipitation occurs in distinct state, thus allowing the parameters values to vary between states. In our statistical approach applied to the Hampshire Avon river's waters, we did not assess directly the macronutrient response to extreme events, but instead we assess their response to the threshold behaviour of river flow, which is, however, largely controlled by weather pattern. Therefore, taking advantage of the interaction with distributional changes in time, we found that the threshold changes in river flow causes very different dynamics in nitrate and phosphate time-series.

An important feature of our model is that it allows the predictions of macronutrient concen-

trations and also the quantification of riverine input fluxes to the estuary. We illustrate this by providing estimates of daily fluxes and annual totals for nitrate and phosphate along with their uncertainties. These daily fluxes can be aggregated to coarser temporal levels, e.g., monthly, quarterly or annually, as demonstrated in our application, where we find that the amount of macronutrients delivered to the estuary can change dramatically according to the period of the year in which river flow experiences larger changes. This is particularly evident for nitrate which shows a clear seasonal pattern, while flux estimates for phosphate present a weaker seasonal structure, that leads to a higher uncertainty in our modelling approach.

The Bayesian modelling framework adopted here can be extended in various ways by including more relevant covariates, such as wind field that may have a short-term mixing effect on water quality, increasing the sediment re-suspension and be a driving force in exporting nutrients in the estuary. This can lead to a better estimate of changes in macronutrient concentrations and fluxes. Multivariate modelling for both the aquatic pollutants and for data from multiple sites may also lead to fruitful research.

In conclusion, the Bayesian approach introduced here is able to facilitate the description of complex and nonlinear environmental processes and is able to assess the associated uncertainties of the reported estimates. We present an application for modelling macronutrients dynamics in relationship to water quality properties and changes in river flow. Our method can be easily adapted to similar data modelling and analysis problems for estuarine pollution using the accompanying computer code.

6. Acknowledgement

This study is supported by the Natural Environmental Research Council (NERC) through *The Christchurch Harbour Macronutrients Project* (grant number: NE/J012238/1), which is one of the four consortium projects of the *Macronutrients Cycles Programme*. We wish to thank Dr Fay Couceiro for her help with the macronutrient analyses.

7. Conflict of interest

The authors declare that there are no conflicts of interest.

References

Alameddine I, Qian SS, Reckhow KH. A Bayesian changepoint-threshold model to examine the effect of TMDL implementation on the flow-nitrogen concentration relationship in the Neuse River basin. Water Research 2011;45:51–62.

Berliner LM. Hierarchical Bayesian time series models. In: Hanson K, Silver R, editors. Maximum Entropy and Bayesian Methods. Kluwer Academic Publishers; 1996. p. 15–22.

Conley DJ, Paerl HW, Howarth RW, Boesch DF, Seitzinger SP, Havens KE, Lancelot C, Likens GE. Controlling eutrophication: nitrogen and phosphorus. Science 2009;323:1014–5.

Crainiceanu C, Ruppert D, Wand MP. Bayesian analysis for penalized spline regression using WinBUGS. Journal of Statistical Software 2005;14:1–24.

Eilers PHC, Marx BD. Flexible smoothing with b-splines and penalties. Statistical Science 1996;11:89–121.

Environment Agency . Hampshire Avon WFD Management Area Abstraction Licensing Strategy. Environment Agency, 2012.

Fortin V, Perreault L, Salas J. Retrospective analysis and forecasting of streamflows using a shifting level model. Journal of Hydrology 2004;296:135 –63.

Gelfand A, Ghosh S. Model choice: A minimum posterior predictive loss approach. Biometrika 1995;85:1–11.

Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 2006;1:515–34.

Gilks WR, Richardson S, Spiegelhalter DJ. Markov Chain Monte Carlo in practice. Chapman & Hall/CRC, 1996.

Hans C. Bayesian lasso regression. Biometrika 2009;69:835–45.

Jarvie HP, Neal C, Withers PJ, Wescott C, Acornley RM. Nutrient hydrochemistry for a groundwater-dominated catchment: The Hampshire Avon, UK. Science of The Total Environment 2005;344:143–58.

Jo S, Kim G, Jeon JJ. Bayesian analysis to detect abrupt changes in extreme hydrological processes. Journal of Hydrology 2016;In press.

Jolliff JK, Kindle JC, Shulman I, Penta B, Friedrichs MA, Helber R, Arnone RA. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. Journal of Marine Systems 2009;76:64–82.

Krause P, Boyle DP, Bäse F. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences 2005;5:89–97.

Laud P, Ibrahim J. Predictive model selection. Journal of the Royal Statistical Society, Series B 1995;57:247–62.

Laud P, Ibrahim J. Seasonal variation of riverine nutrient inputs in the northern Bay of Biscay (France), and patterns of marine phytoplankton response. Journal of Marine Systems 2008;72:309–19.

Leigh C, Bush A, Harrison ET, Ho SS, Luke L, Rolls RJ, Ledger ME. Ecological effects of extreme climatic events on riverine ecosystems: insights from Australia. Freshwater Biology 2014;60:2620–38.

Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 2000;10:325–37.

Lykou A, Ntzoufras I. On Bayesian lasso variable selection and the specification of the shrinkage parameter. Statistics and Computing 2013;23:361–90.

Marley JK, Wand MP. Non-standard semiparametric regression via BRugs. Journal of Statistical Software 2010;37:1–30.

Muchan K, Lewis M, Hannaford J, Parry S. The winter storm of 2013/2014 in the UK: hydrological responses and impacts. Weather 2015;70:55–61.

Nedwell D, Dong L, Sage A, Underwood G. Variations of the nutrients loads to the Mainland U.K. estuaries: Correlation with catchment areas, urbanization and coastal eutrophication. Estuarine, Coastal and Shelf Science 2002;54:951–70.

Ntzoufras I. Bayesian Modeling Using WinBUGS. Wiley, 2009.

O'Hara RB, Sillanpää MJ. A review of Bayesian variable selection methods: what, how and which. Bayesian Analysis 2009;4:85–117.

Paerl HW. Controlling eutrophication along the freshwater-marine continuum: dual nutrient (N and P) reductions are essential. Estuaries and Coasts 2009;32:593–601.

Park T, Casella G. The Bayesian Lasso. Journal of the American Statistical Association 2008;103:681–6.

Poff NL, Zimmerman JKH. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. Freshwater Biology 2010;55:194–205.

Quilbé R, Rousseau AN, Duchemin M, Poulin A, Gangbazo G, Villeneuve JP. Selecting a calculation method to estimate sediment and nutrient loads in streams: Application to the Beaurivage river (Québec, Canada). Journal of Hydrology 2006;326:295–310.

Rolls RJ, Leigh C, Sheldon F. Mechanistic effects of low-flow hydrology on riverine ecosystems: ecological principles and consequences of alteration. Freshwater Science 2012;31:1163–86.

Ruppert D. Selecting the number of knots for penalized splines. Journal of Computational and Graphical Statistics 2002;11:735–57.

Ruppert D, Wand MP, Carroll RJ. Semiparametric Regression. Cambridge University Press, 2003.

Sigleo AC, Frick WE. Seasonal variations in river flow and nutrient concentrations in a northwestern USA watershed. In: Proceedings of the first interagency on research in the watersheds. USDA, Benson, Arizona; 2003. p. 370–6.

Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. Journal of Statistical Software 2005;12:1–16.

Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society B 1996;58:267–88.

Walther GR. Community and ecosystem responses to recent climate change. Philosophical Transactions of the Royal Society B 2010;365:2019–24.

Wellen C, Arhonditis GB, Long T, Boyd D. Accommodating environmental thresholds and extreme events in hydro-

logical models: A Bayesian approach. Journal of Great Lakes Research 2014;40:102–16.

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.

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.

Withers PJA, Neal C, Jarvie HP, Doody DG. Agriculture and eutrophication: where do we go from here? Sustainability 2014;6:5853–75.

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
Macronutrients						
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
Water properties						
Temperature ($^{\circ}C$)	4.71	8.11	12.26	12.21	16.24	21.96
Conductivity ($\mu S/cm$)	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 ($\mu g/L$)	0.89	1.58	2.71	2.40	3.41	8.73
River Flow						
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: Quality of the model fit and the predictive abilities of the competing models: goodness of fit term (G), penalty term (P) and overall predictive model choice criterion (PMCC: G+P), along with the Bayesian statistic R_B^2.

		Nitrate				Phosphate			
	Models	G	P	PMCC (G+P)	R_B^2	G	P	PMCC (G+P)	R_B^2
M1.	Penalised spline for water quality data + change-point structures	1.31	3.76	5.07	0.69	33.84	42.61	76.45	0.74
M2.	Linear model for water quality data, no change-point structures	3.87	5.97	9.84	0.46	66.79	71.42	138.21	0.47
M3.	Linear model for water quality data + a switch-point in time	3.02	5.14	8.16	0.53	73.50	78.58	152.08	0.49
M4.	Linear model for water quality data + a change threshold in river flow	2.56	4.70	7.26	0.57	67.72	72.42	140.14	0.53
M5.	Linear model for water quality data + change-point structures	1.56	4.04	5.60	0.65	44.24	50.67	94.91	0.67
M6.	Penalised splines for water quality data, no change-point structures	2.37	4.77	7.14	0.58	50.24	57.91	108.15	0.63

Table 4: Parameter estimations.

	Nitrate		Phosphate	
Parameters	Median	95%CI	Median	95%CI
Change-point structures				
τ (Switch-point 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 switch-point in time)	27.87	(16.26, 43.64)	66.35	(7.89, 100.28)
φ_2 (Change threshold in flow after switch-point in time)	10.64	(7.65, 13.41)	34.29	(29.19, 40.32)
δ_1 (Slope for low flow before switch-point in time)	−0.80	(−0.63, −0.88)	0.07	(−0.63, 1.26)
δ_2 (Slope for high flow before switch-point in time)	1.17	(1.09, 1.32)	0.06	(−0.74, 1.94)
δ_4 (Slope for high flow after switch-point in time)	1.21	(1.16, 1.27)	−1.25	(−1.59, −0.97)
Penalised splines				
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)
Other				
σ^2 (Measurement error variance)	0.01	(0.00, 0.01)	0.10	(0.08, 0.12)

Table 5: Posterior median and 95% credible interval (CI) for the catchment area normalised total annual macronutrient fluxes from 22/11/2013 to 21/11/2014 according to models M1 and M2. Values are catchment area standardised with Kg/Km^2 units. The last two rows present comparable estimates from the literature.

	Nitrate (N)		Phosphate (P)	
Models/Methods	Annual budget	95%CI	Annual budget	95% CI
M1. Penalised spline for water quality data + change-point structures	2978.9	(2937.9, 3016.4)	31.6	(30.2, 33.1)
M2. Linear model for water quality data, no change-point structures	2936.7	(2890.4, 2981.8)	29.8	(28.3, 31.3)
Jarvie et al. (2005): Mean annual load during 1993-2000	2050	–	71	–
Nedwell et al. (2002): UK average during 1995-96	1400	–	152	–

Figure 1: Map of the study area.

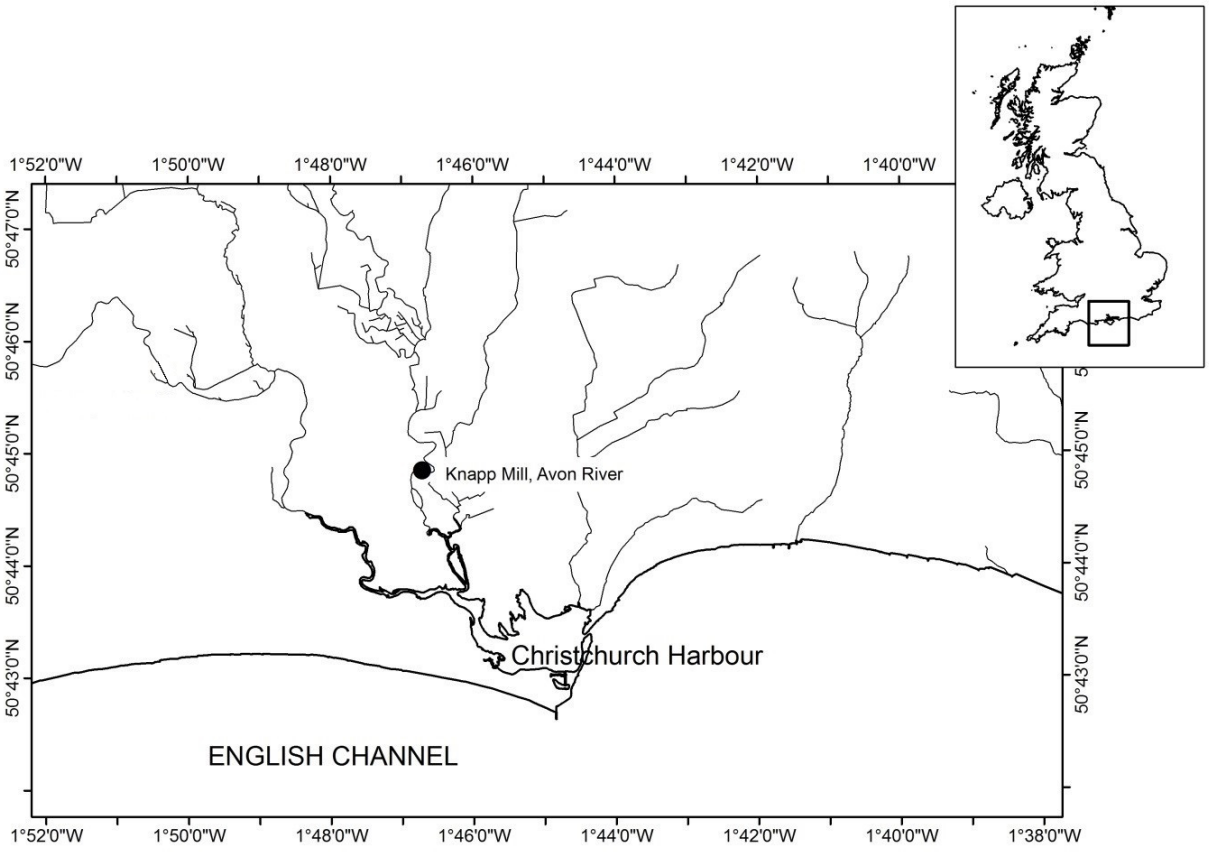


Figure 2: Daily macronutrient and river flow data (22/11/2013 to 19/12/2014). Data are plotted on original scale: nitrate (solid line) in mg/L, phosphate (dashed line) in mg/L and river flow (dotted line) in m^3/s.

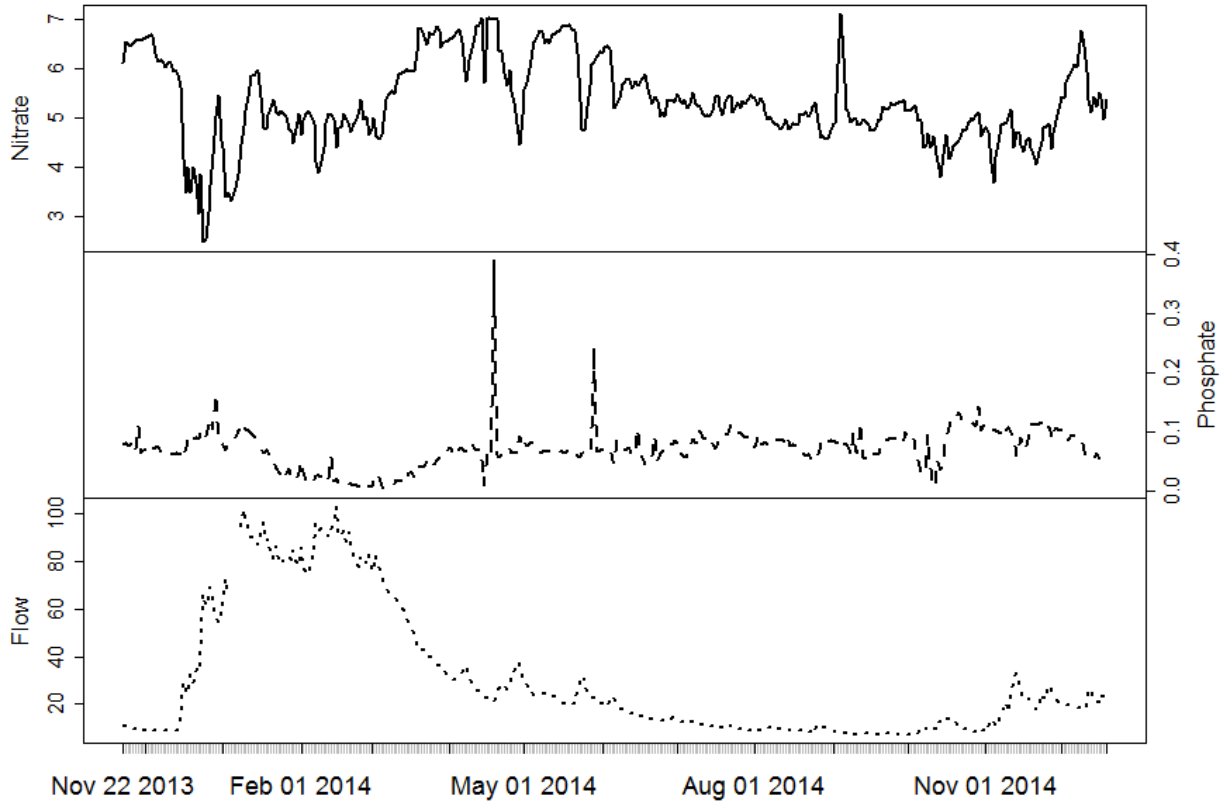


Figure 3: Daily water quality data values (22/11/2013 to 19/12/2014). Data are plotted on original scale: temperature (solid line) in $°C$, conductivity (dashed line) in $\mu S/cm$, dissolved oxygen saturation (dotted line) in %, turbidity (dotdash line) in NTU and chlorophyll (longdash line) in $\mu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

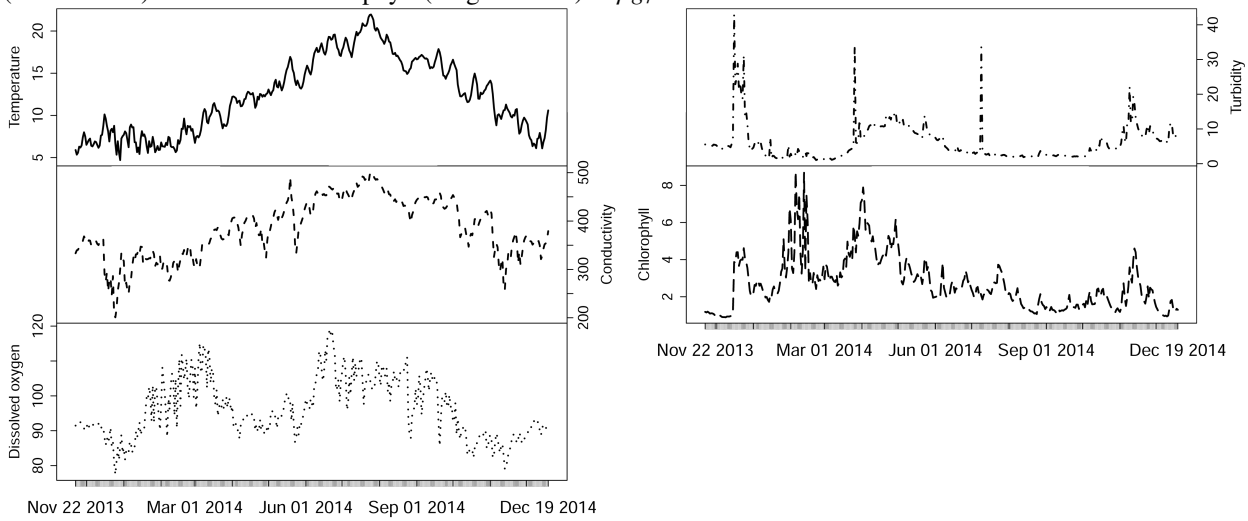


Figure 4: Scatterplot of macronutrient data *versus* water quality properties. Data are plotted on original scale: nitrate in mg/L, phosphate in mg/L, temperature in $°C$, conductivity in $\mu S/cm$, dissolved oxygen (DO) in %, turbidity in NTU and chlorophyll in $\mu g/L$.

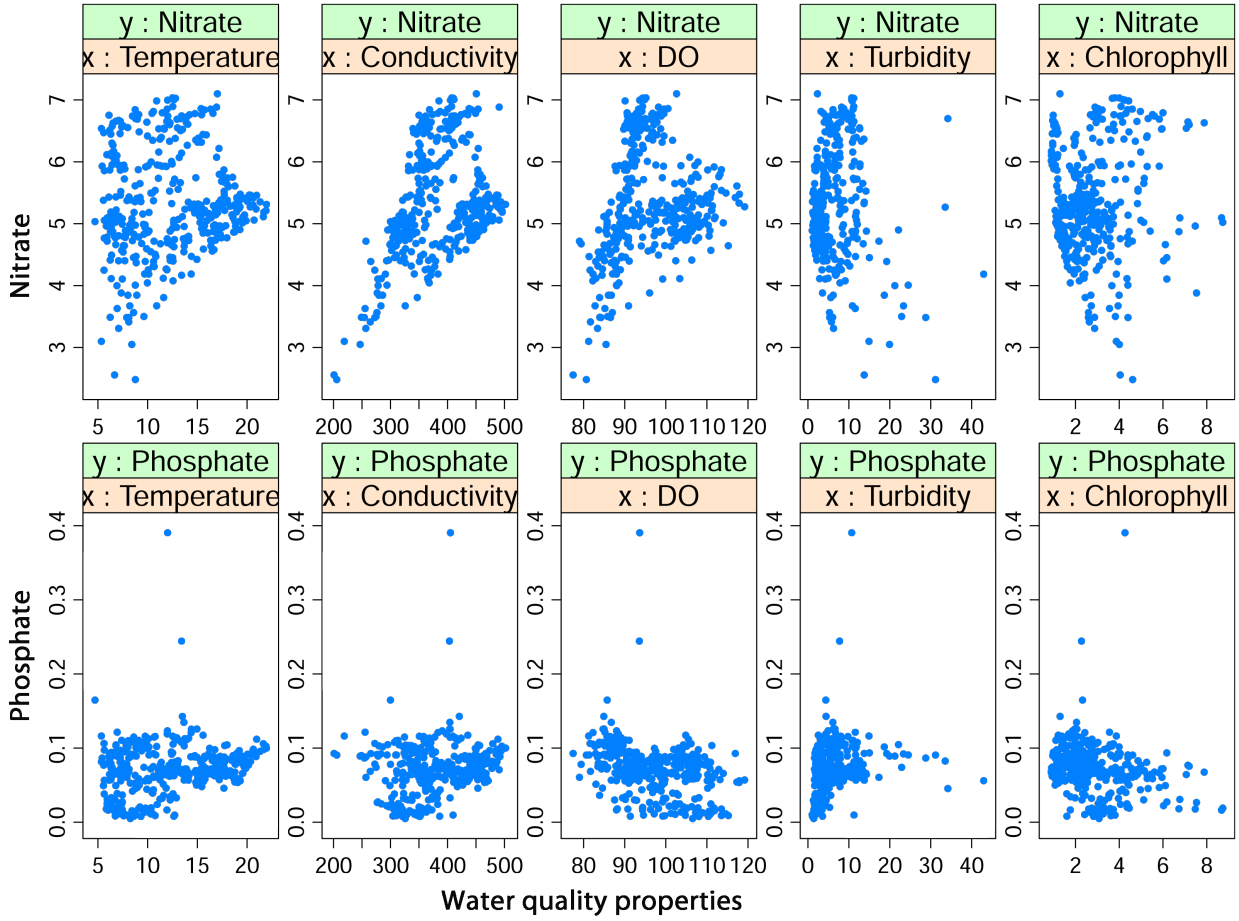
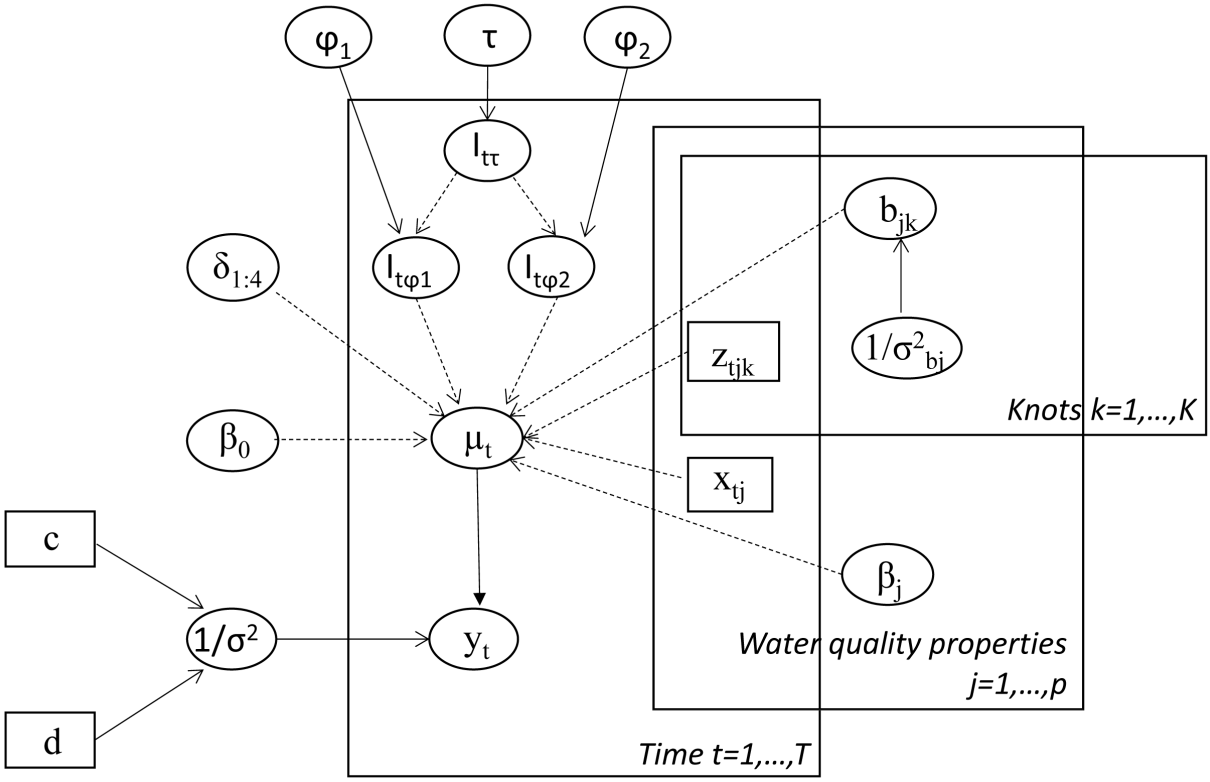
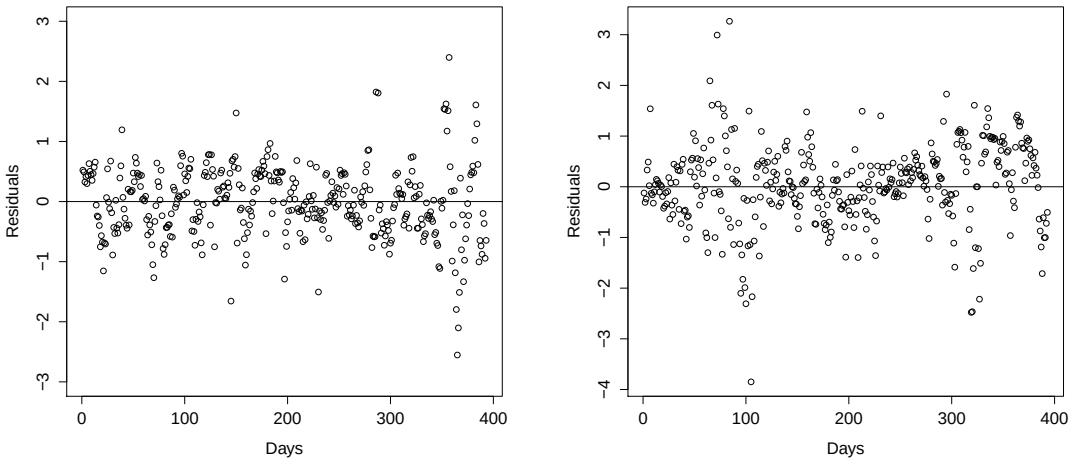


Figure 5: Directed Acyclic Graph (DAG) for the model M1, depicting the relationship among macronutrient, river flow and water quality properties.



In the graph, y_t is the observed concentration for a macronutrient in day t. The macronutrient concentrations are assumed to be distributed normally around the mean μ_t. The parameter $1/\sigma^2$ represents the precision (i.e. 1/variance) of the normal distribution. μ_t is modelled as a function of: (i) a global intercept, β_0; (ii) the water quality properties, which are parameterised via penalised splines, where β_j is the fixed coefficient for each water quality property, x_{tj}, and b_{jk} are the random coefficients associated with the design matrix with elements $z_{tjk} = (x_{tj} - k_{kj})_+^d$; (iii) the change-point structures constructed with interaction terms described by indicator functions: $I_{t\tau}$ is the indicator for the switch-point in time, τ, and $I_{t\varphi 2}$ and $I_{t\varphi 2}$ are the indicators for the two change thresholds in river flow, φ_1 and φ_2. Finally, $\delta_1, \ldots, \delta_4$ are the coefficients associated with the change-point structures.

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



(a) (b)

Figure 7: Relationship between nitrate and river flow using the estimated parameters for the change-point structures: red dotted vertical line is the switch-point in time and black solid horizontal lines are the change thresholds in river flow. Data are plotted on original scale: nitrate (black solid line) in mg/L, and river flow (gray dotted line) in m^3/s.

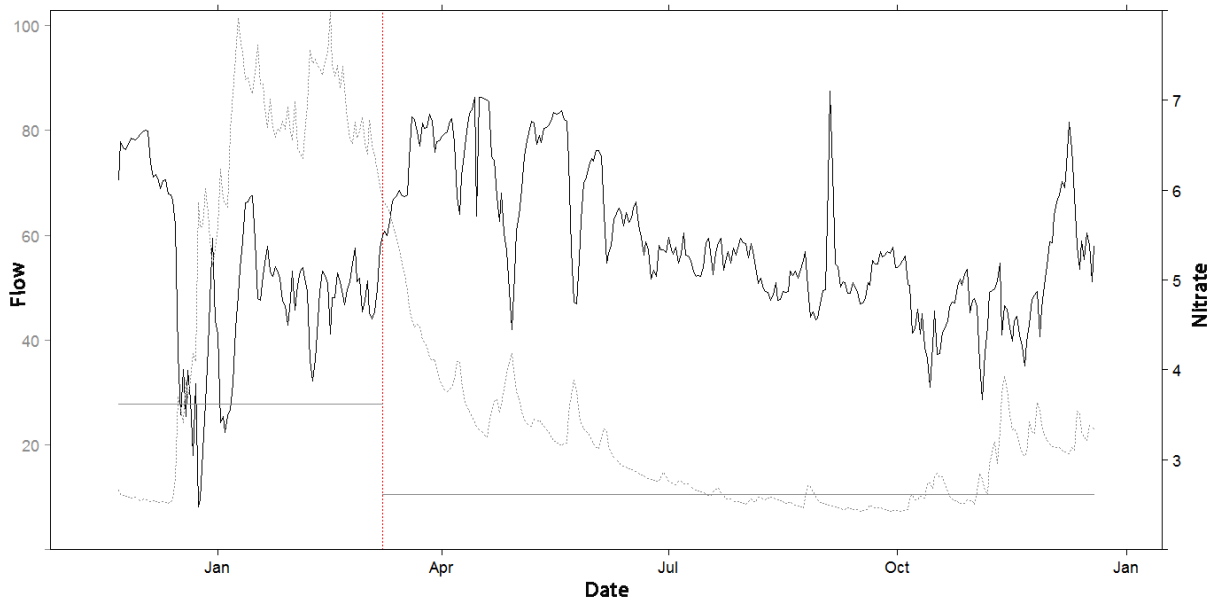


Figure 8: Relationship between phosphate and river flow using the estimated parameters for the change-point structures: red dotted vertical line is the switch-point in time and black solid horizontal lines are the change thresholds in river flow. Data are plotted on original scale: phosphate (black solid line) in mg/L, and river flow (gray dotted line) in m^3/s.

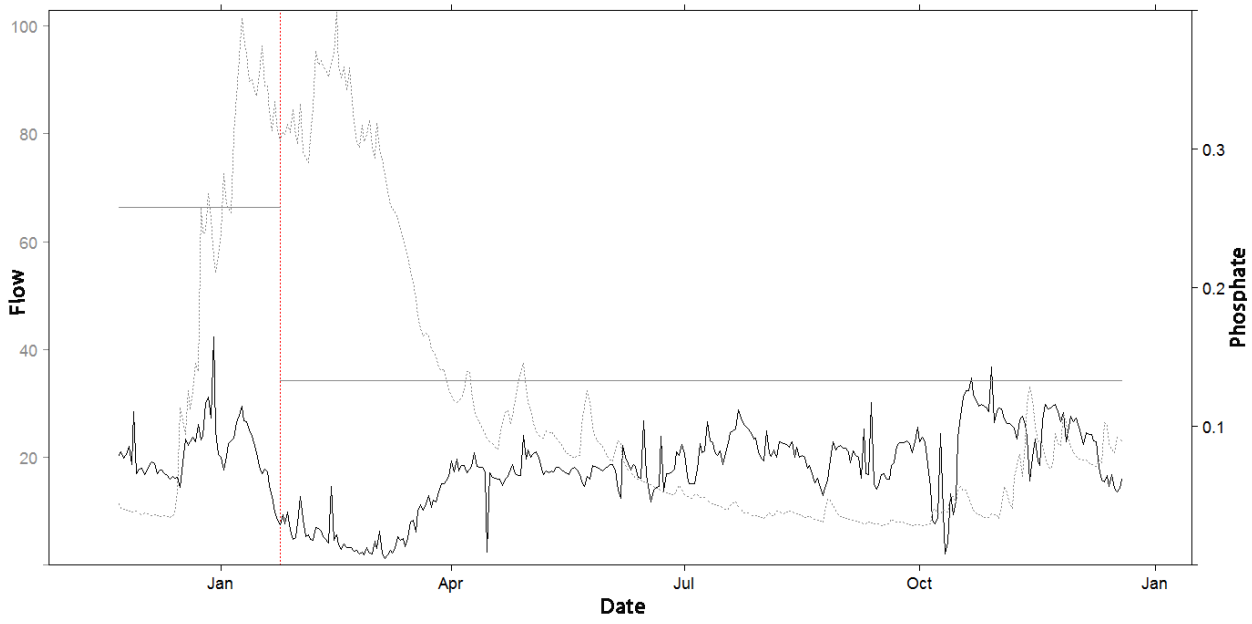


Figure 9: Plots of daily fluxes (Kg/Day) for nitrate (top) and phosphate (bottom). Red rectangle on the right-hand of the panel contains the predicted fluxes from model M1 for the period 30/11/2014 to 19/12/2014.

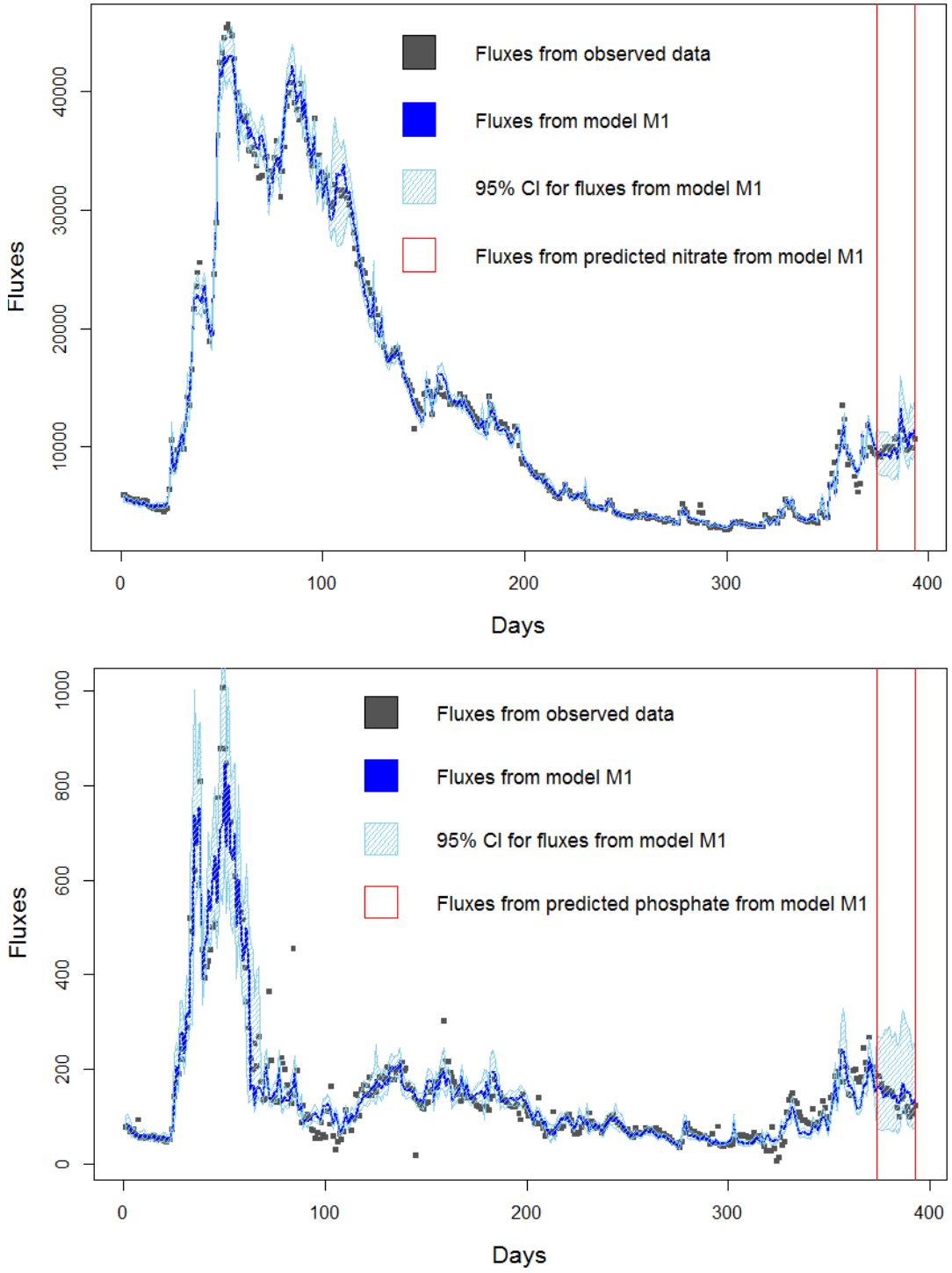


Figure 10: Nitrate fluxes (Kg/Day) for the entire study period using the estimated parameters for the change-point structures. Black dots are the observed fluxes, black solid line represents the fluxes estimated from model M1 and shadow area represents the 95% CI for the estimated fluxes.

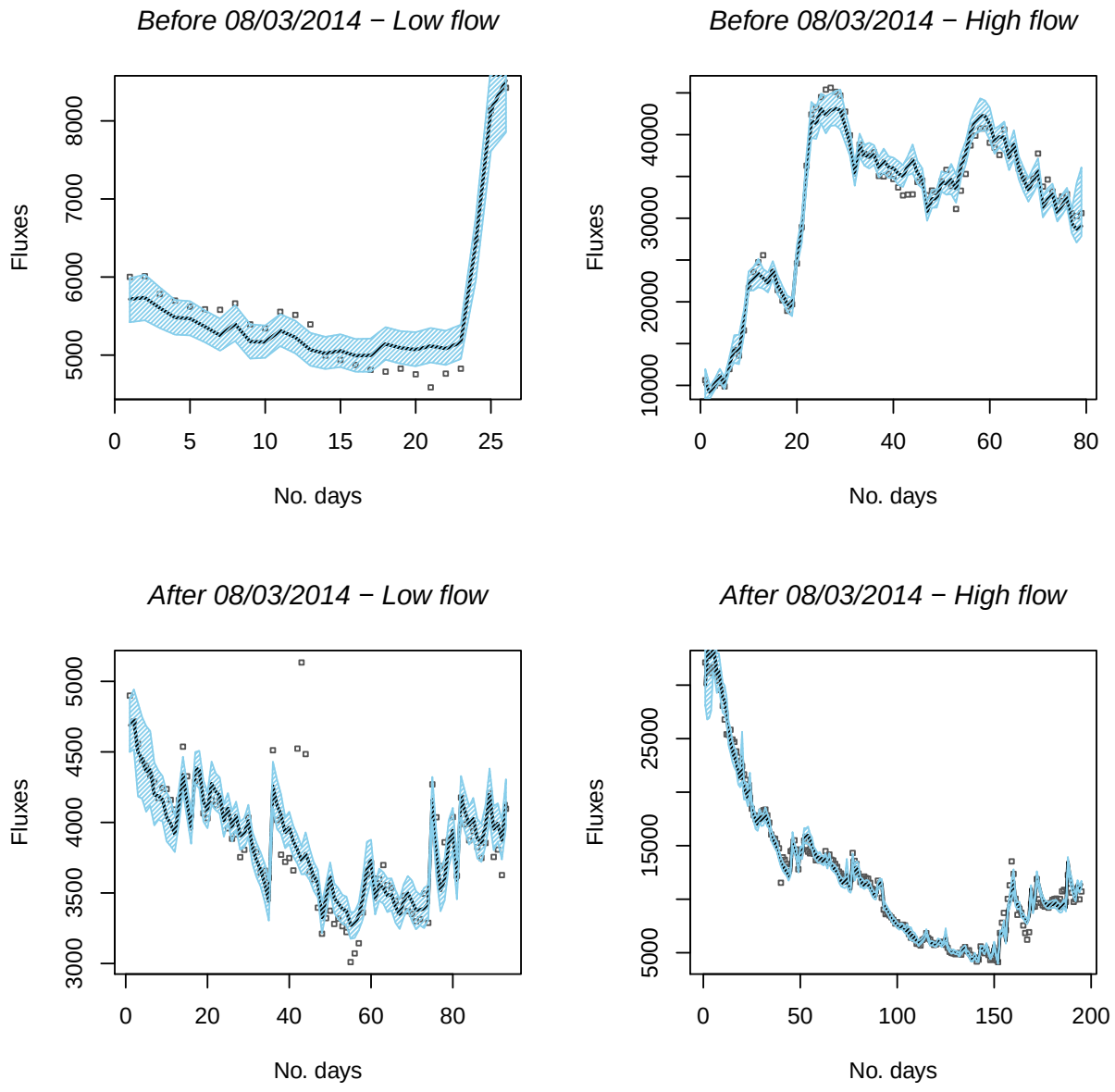


Figure 11: Phosphate fluxes (Kg/Day) for the entire study period using the estimated parameters for the change-point structures. Black dots are the observed fluxes, black solid line represents the fluxes estimated from model M1 and shadow area represents the 95% CI for the estimated fluxes.