ENVIRONMETRICS
Environmetrics 2009; **20**: 283–297
Published online 2 June 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.929

Nonparametric trend detection in river monitoring network data: a spatio-temporal approach

Lieven Clement*,† and Olivier Thas

Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Ghent, Belgium

SUMMARY

Since the introduction of the Water Framework Directive (WFD), regional authorities have to take actions to improve the aquatic environment. In Flanders (Belgium), for instance a Manure Decree (MD) has been introduced to reduce the impact of eutrophication. It is of course important to assess the evolution of the water quality. In this context, we develop a spatio-temporal method for the assessment of trends in river monitoring network data. The aim is to detect trends on a more regional scale. In contrast with most existing methods used for analysing river data, our model incorporates the spatio-temporal dependence structure explicitly. An AR(1) process is assumed for the temporal dependence. The spatial dependence is implied by the river topology and the flow direction. The dependence structure is modelled by a state variable which is embedded into an observation model that makes the correlation structure less rigid. An additive model incorporates a nonparametric long-term trend and it corrects for seasonal effects. To detect when the local trend is beneficial, tests on the first derivative of the nonparametric long-term trend are performed at each time step. This, however, results in a large number of simultaneous tests. To correct for multiplicity, we propose to use a variant of the free step-down method of Westfall and Young (1993) to take these dependences into account. The methodology is applied to a case study on the river Yzer (Belgium). A significant decrease in the nitrate concentration is detected after the implementation of the MD.

KEY WORDS: spatio-temporal model; water quality; GAM; multiplicity correction; short-term trends

1. INTRODUCTION

The Water Framework Directive (WFD) (EC, 2000) aims to trigger local authorities to improve the aquatic environment. To reach that goal, the Flemish Environmental Agency (VMM) develops basin management plans to improve the water quality of the rivers in Flanders (Belgium). A dominant problem in Flemish water bodies is the eutrophication due to nutrient pollution. One of the main nutrient pollution sources originates from agricultural activities. In Flanders there is an intensive pig farming activity and in the past the produced manure was mainly disposed on agricultural lands. A major action to reduce this nutrient load was the introduction of two Manure Decrees (MDs) (Vlaams Parlement, 1995, 1999).

*Correspondence to: L. Clement, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure links 653, 9000 Ghent, Belgium.
†E-mail: lieven.clement@ugent.be

Received 18 April 2008
Accepted 24 April 2008

The MDs restrict the amount of fertilisers that may be used by farmers in areas which are susceptible to eutrophication. The first MD was introduced in 1996 (Vlaams Parlement, 1995) and after an evaluation a new and more restrictive MD was implemented in 2000 (Vlaams Parlement, 1999). When such actions are taken, it is important to assess whether they indeed have an effect on the water quality. Therefore, water quality monitoring networks are needed to assess the evolution of the water quality. In Flanders, the VMM has developed several monitoring networks along the rivers. In their physico-chemical monitoring network, a basic spectrum of physico-chemical variables is evaluated monthly at each sampling location. In this paper we assess the evolution of the nitrate concentration in a small region of the Yzer basin. This river is located in the western part of Flanders. It is a rural area with a large agricultural activity.

The focus of this manuscript lays on the development of a methodology to detect and to locate trends in the water quality data. Instead of assessing trends at the level of individual sampling locations, our aim is to develop a methodology for trend detection on a more regional scale. Standard techniques cannot be used for this purpose because river monitoring networks typically generate data with a strong spatial and temporal dependence structure. In order for the statistical inference procedure to be formally valid, this dependence has to be taken into account. Many contributions exist on the assessment of trends in individual time series (e.g. Brillinger, 1989; Cleveland *et al*., 1990; El-Sharaawi and Niculescu, 1993; Esterby, 1993; El-Sharaawi, 1995; Esterby, 1996; Canjels and Watson, 1997; Thas *et al*., 1998; Sun and Pantula, 1999; Zheng and Basher, 1999, 2006; Wu and Zhao, 2007). Several approaches can be distinguished: (1) descriptive methods based on smoothing, (2) nonparametric methods based on the Mann–Kendall test and (3) time series methods that model the serial correlation explicitly. In environmental systems, however, the trend is often nonlinear and changes over time. The descriptive methods, such as the seasonal-trend decomposition procedure (STL) based on loess of Cleveland *et al*. (1990), can handle trend changes but they do not provide statistical tests to assess the significance of the observed trend/trend changes. Nonparametric methods based on the Mann–Kendall test are designed to detect monotonic trends. But, they are not useful when there are sign changes in the first derivative of the trend. The time series approach is useful in this setting if it allows the trend component to change over time. In contrast with the existing literature, we will not focus on single sampling locations, but we simultaneously study time series at multiple sampling locations of a river monitoring network instead. Thus, besides the temporal dependence, the spatial dependence structure due to the monitoring network topology also has to be taken into account. Contributions that focus on assessing trends in multiple time series observed at different sampling locations of a river monitoring network are limited. The few existing methods, however, have avoided the estimation of the spatio-temporal dependence in river monitoring network data by simply ignoring it or by using ad hoc methods. Burn and Hag Elnur (2002), for instance, adopted an approach to determine the *field significance* that is involved in the calculation of a regional value for the Mann–Kendall statistic. The field significance is obtained by the use of a permutation method. To correct for serial correlation, they proposed to first perform a pre-whitening step which retains the trend. They then used a permutation approach to obtain the distribution of the p-value under the null hypothesis of no trend. To correct for the spatial dependence, their method was designed as such to preserve the spatial dependence in each permutated dataset. Beside the stagewise removal of the dependences, another disadvantage of their approach is the assumption of monotonic trends.

Similar to the time series approach, we suggest to model the spatio-temporal dependence in river monitoring network data to control the type I error rate of the trend tests. In contrast with ad hoc methods, this modelling approach provides a very natural way to introduce the spatio-temporal dependence structure into the testing procedure. We will use the spatio-temporal correlation structure

DOI: 10.1002/env

for river networks that was proposed by Clement and Thas (2007). They pointed out that an important distinction has to be made with respect to the spatial dependence structures that are commonly used in geostatistics: the direction of the flow implies a causal interpretation of the correlations. Since rivers can also join or split, the river topology implies a branched unidirectional structure. In reality the environmental conditions may obscure this quite stringent unidirectional spatial dependence structure. Clement and Thas (2007) therefore only impose this restrictive topology-implied dependence structure on an unobservable state variable **S**. The latent variable **S** is embedded into an observation model **y** that allows cross-correlation between sampling locations that are located at different branches of the river. Besides the dependence structure, we also have to specify the marginal mean model that is used for modelling the trends. Trends in water quality are often nonlinear. We therefore propose a trend detection based on local polynomial regression smoothers. To enable testing for the trend on a regional scale, a common nonparametric trend is estimated at all sampling locations. The local trend can be assessed by performing tests on the first derivative of the nonlinear trend component. These tests have to be performed at each time instant, which leads to a large number of simultaneous tests. Therefore, a multiplicity correction procedure is required. In general, observations which are close in time are likely to have similar trends so that the trend tests are not independent. This reduces the actual dimension of the multiplicity problem. In this paper we present a procedure that corrects for multiplicity and takes the dependence between the tests explicitly into account. This leads to a correct statistical inference procedure that is not too conservative.

The organisation of the paper is as follows. First, the spatio-temporal model is briefly presented in Section 2. The parameter estimation procedure is introduced in Section 3. Section 4 deals with the trend detection method, and, in Section 5, the methodology is illustrated on a case study. Finally, we formulate some conclusions in Section 6.

2. SPATIO-TEMPORAL MODEL

Let $\mathbf{S} = (S_1, \ldots, S_p)^T$ represent the $p \times 1$ vector of response variables S_i at sampling locations $i = 1, \ldots, p$. The spatial correlation structure of **S** is completely defined by the river topology. This is illustrated in Figure 1, which shows the river topology of five sampling locations. The same figure can also be interpreted as a directed acyclic graph (DAG) (see e.g. Whittaker, 1990) in which the circles represent the S_i's and missing edges or arrows define conditional independences. The river monitoring network, however, generates data over time. Hence, the spatial pattern of the DAG is repeated over time. Let $\mathbf{S}_t = (S_{1t}, \ldots, S_{pt})^T$ and let us assume a Markovian temporal dependence structure which models $\mathbf{S}_t$ by conditioning on $\mathbf{S}_{t-1}$. In this case, we can use the equations in Section 2.3 of Clement and Thas

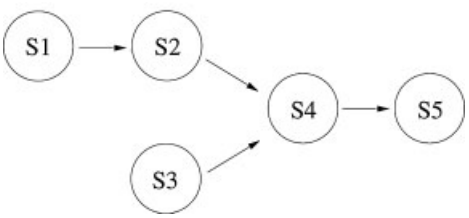

Figure 1. Directed acyclic graph (DAG) of five sampling locations along two joining river reaches

DOI: 10.1002/env

(2008) to model the covariance structure of $\mathbf{S}_t$. The resulting model is represented as

$$\mathbf{S}_t = (\mathbf{I}_p - \mathbf{A})^{-1}\mathbf{B}\mathbf{S}_{t-1} + (\mathbf{I}_p - \mathbf{A})^{-1}\boldsymbol{\eta}_t \tag{1}$$

where $t = 1, \ldots, n$, $\mathbf{I}_p$ is the $p \times p$ identity matrix, $\boldsymbol{\eta}_t \sim \text{MVN}(0, \boldsymbol{\Sigma}_\eta)$ with a $p \times p$ diagonal variance–covariance matrix $\boldsymbol{\Sigma}_\eta$, $\mathbf{A}$ is a $p \times p$ matrix that models the spatial correlation according to the DAG and $\mathbf{B}$ is a $p \times p$ matrix containing the temporal autocorrelation (diagonal elements) and the spatio-temporal cross-correlation coefficients (off-diagonal elements). More details on the structure of the matrices $\mathbf{A}$ and $\mathbf{B}$ can be found in Clement and Thas (2008). For completeness the initial conditions have to be defined at time instant 0. We assume $\mathbf{S}_0$ to be $\text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_{S_0})$.

In reality, however, the dependence structure might be obscured by common environmental confounders, such as rainfall. Therefore, the model is embedded into an observation model:

$$\mathbf{y}_t = \mathbf{S}_t + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, n \tag{2}$$

where $\mathbf{y}_t$ is the observation vector corresponding to $\mathbf{S}_t$, and $\boldsymbol{\epsilon}_t \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$. No restrictions are imposed on $\boldsymbol{\Sigma}_\epsilon$. This enables cross-correlations between sampling locations that are not connected according to the river topology.

Model (2) only defines the spatio-temporal dependence structure. It can be easily seen that $E[\mathbf{y_t}] = \mathbf{0}$ at all time instants. To model the trend, Equation (2) is extended with an additive model for the mean. Besides a trend, seasonal variation is typically present in water quality data. A common approach to deal with this variation is to include sinusoidal functions of fixed periods to describe the seasonal cycle within a year (e.g. Hirst, 1998; Cai and Tiwari, 2000; McMullan *et al*., 2003; McMullan, 2004). A function which is widely used for this purpose is $\alpha \cos(2\pi(t/P) + \theta)$, where P is the period, which is here taken to be 1 year, α is the amplitude of the seasonal trend and θ is a parameter to allow for a phase shift. The cosine function can also be expressed in a linear form by using a standard trigonometric expansion of the cosine term. This is the parametrisation of our choice and therefore we use Fourier basis functions to model the seasonal effect: $\gamma_1 \sin(2\pi t/365) + \gamma_2 \cos(2\pi t/365)$. Hence, the following mean model is proposed: $\text{E}[y_{it}] = \mathbf{X}_{it}\boldsymbol{\beta} + f(t)$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^T$ is the $q \times 1$ parameter vector and $\mathbf{X}_{it}$ is the $1 \times q$ design vector that includes the proper Fourier basis functions and some other linear predictors, and $f(t)$ is the nonlinear trend which will be estimated by the use of a local linear regression smoother. Note that $f(t)$ does not depend on the sampling location because we want to assess the trend on a regional scale. After embedding the mean model into Model (2), we obtain

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \mathbf{f}(t) + \mathbf{S}_t + \boldsymbol{\epsilon}_t \tag{3}$$

which specifies together with Model (1) the complete spatio-temporal state–space model.

An equivalent formulation of the spatio-temporal model is accomplished by recognising that the Models (1) and (3) can be written as a Structural Equation Model (SEM) (see e.g. Maruyama, 1997),

$$\mathbf{C}\mathbf{S}_N = \boldsymbol{\zeta} \tag{4}$$

$$\mathbf{Y}_N = \mathbf{X}_N\boldsymbol{\beta} + \mathbf{f}_N + \mathbf{S}_N + \boldsymbol{\psi} \tag{5}$$

where $\mathbf{C}$ is a $pn \times pn$ square matrix constructed from the elements of the matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{S}_N = (\mathbf{S}_1^T, \ldots, \mathbf{S}_n^T)^T$, $\boldsymbol{\zeta} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\zeta)$ where $\boldsymbol{\Sigma}_\zeta$ is a diagonal matrix built from the corresponding elements

DOI: 10.1002/env

of $\boldsymbol{\Sigma}_\eta$, $\mathbf{Y}_N = (\mathbf{y}_1^T, \ldots, \mathbf{y}_n^T)^T$, $\mathbf{X}_N = (\mathbf{X}_1^T, \ldots, \mathbf{X}_n^T)^T$, $\mathbf{f}_N = (\mathbf{f}^T(1), \ldots, \mathbf{f}^T(n))^T$, and $\boldsymbol{\psi} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\psi)$ where $\boldsymbol{\Sigma}_\psi$ is block-diagonal with blocks $\boldsymbol{\Sigma}_\epsilon$. From this SEM formulation the covariance structure of the observation vector $\mathbf{Y}_N$ is easily found

$$\boldsymbol{\Sigma}_{Y_N}(\boldsymbol{\Psi}_\alpha) = \text{var}\,(\mathbf{Y}_N) = \mathbf{C}^{-1}\boldsymbol{\Sigma}_\zeta\mathbf{C}^{-T} + \boldsymbol{\Sigma}_\psi \tag{6}$$

with $\boldsymbol{\Psi}_\alpha$ the vector that contains all parameters in $\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\Sigma}_{\mathbf{S_0}}$, $\boldsymbol{\Sigma}_\eta$ and $\boldsymbol{\Sigma}_\epsilon$.

3. PARAMETER ESTIMATION AND STATISTICAL INFERENCE PROCEDURE

It is possible to perform the parameter estimation and the inference procedure completely in the likelihood framework. However, to control the computational burden we will consider a slightly different approach where the mean model is estimated by ordinary least squares (OLS). OLS also provides an unbiased estimator, but under general conditions it is asymptotically less efficient than generalised least squares (GLS) (e.g. Shin and Oh, 2002). In practical applications, the covariance matrix is unknown and has to be estimated. When an estimator of the covariance matrix is used to perform GLS, this estimator is known as the feasible GLS (FGLS) estimator. If the roots of the transition matrix lie within the unit circle, both the OLS and FGLS are known to be asymptotically efficient (e.g. Canjels and Watson, 1997). The parameter estimation procedure of the mean model is given in Section 3.1. In Section 3.2 the estimation procedure for the parameters of the dependence structure is briefly discussed.

3.1. Mean model

In this paper the nonlinear trend is estimated by the use of a polynomial smoother (an overview of local polynomial smoothers can be found in Fan and Gijbels (1996)). Because a smoother is involved in the mean model, we will have to obtain the smoother matrix to fit the semiparametric mean model. In a GLS context, the dependence structure is involved in the calculation of the smoother matrix (see e.g. Giannitrapani *et al*., 2005). But, as we will see later on, the dependence structure will be estimated by the use of an iterative algorithm. This would imply the recalculation of the projection matrix of the smoother at each iteration and would result in a drastic increase of the computational power that is needed to estimate the model parameters. Therefore we use OLS to fit the mean model. From a computational point of view, this method has a considerable advantage because the parameters of the mean model only have to be estimated once and the parameters of the dependence structure can subsequently be estimated using the residuals of the OLS estimation procedure.

When the OLS procedure is applied to our particular additive model, the estimating equations have an analytical solution (Hastie and Tibshirani, 1990). The following results are obtained for our model:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_N^T\left(\mathbf{I}_N - \mathbf{S}_f\right)\mathbf{X}_N\right)^{-1}\mathbf{X}_N^T\left(\mathbf{I}_N - \mathbf{S}_f\right)\mathbf{Y}_N = \mathbf{H}_\beta\mathbf{Y}_N \tag{7}$$

$$\hat{\mathbf{f}} = \mathbf{S}_f\left(\mathbf{Y}_N - \mathbf{X}_N\hat{\boldsymbol{\beta}}\right) \tag{8}$$

DOI: 10.1002/env

where $\mathbf{S}_f$ is the smoother matrix and $\mathbf{I}_N$ is the $N \times N$ identity matrix. Hence, a projection matrix $\mathbf{H}_f$ can be constructed for the smoother term:

$$\mathbf{H}_f = \mathbf{S}_f \left(\mathbf{I}_N - \mathbf{X}_N \left(\mathbf{X}_N^T \left(\mathbf{I}_N - \mathbf{S}_f \right) \mathbf{X}_N \right)^{-1} \mathbf{X}_N^T \left(\mathbf{I}_N - \mathbf{S}_f \right) \right) \tag{9}$$

For inference procedures, this is advantageous. Once the covariance matrix of the observations $\mathbf{Y}_N$ is available, the covariance matrix of the smoother estimators can be obtained. To assess whether a beneficial trend occurs after a certain time, we have to infer on the first derivative of the trend. For local polynomial regression smoothers, a smoother matrix $\mathbf{S}_{f^{(1)}}$ for the first derivative $\mathbf{f}^{(1)}$ is available (Fan and Gijbels, 1996). The smoother, however, is embedded in an additive model, thus a projection matrix $\mathbf{H}_{f^{(1)}}$ for the first derivative has to be calculated. For local polynomial smoothers, this becomes

$$\begin{aligned} \hat{\mathbf{f}}^{(1)} &= \mathbf{S}_{f^{(1)}} \left(\mathbf{Y}_N - \mathbf{X}_N \hat{\boldsymbol{\beta}} \right) \\ &= \mathbf{S}_{f^{(1)}} \left(\mathbf{I}_N - \mathbf{X}_N \left(\mathbf{X}_N^T \left(\mathbf{I}_N - \mathbf{S}_f \right) \mathbf{X}_N \right)^{-1} \mathbf{X}_N^T \left(\mathbf{I}_N - \mathbf{S}_f \right) \right) \mathbf{Y}_N \\ &= \mathbf{H}_{f^{(1)}} \mathbf{Y}_N \end{aligned} \tag{10}$$

with

$$\mathbf{H}_{f^{(1)}} = \mathbf{S}_{f^{(1)}} \left(\mathbf{I}_N - \mathbf{X}_N \left(\mathbf{X}_N^T \left(\mathbf{I}_N - \mathbf{S}_f \right) \mathbf{X}_N \right)^{-1} \mathbf{X}_N^T \left(\mathbf{I}_N - \mathbf{S}_f \right) \right) \tag{11}$$

3.2. Dependence structure

The model is written as a state space model. Hence, an expectation-maximisation (EM) algorithm can be used for parameter estimation (Shumway and Stoffer, 1982). It is an iterative algorithm which ensures the likelihood to increase at each iteration. For our particular application, the EM algorithm of Shumway and Stoffer (1982) has to be adjusted to take the restrictions on the model matrices into account. We propose to apply an adjusted version of the ECM algorithm of Clement and Thas (2007). Their algorithm consists of an E-step where the Kalman filter and smoother are used to calculate the sufficient statistics for the latent process $\mathbf{S_N}$ conditionally on the current parameter values. In the first CM step the parameter values of the dependence structure are updated and the second CM step consists of the update of the parameters of the mean model. In this contribution, the parameters of the mean model are estimated by OLS. Hence, the second CM step is redundant. The different steps that are needed to estimate the dependence structure in our application are briefly given in the Appendix. The reader can find a detailed derivation of the E-step and the first CM step of the algorithm in Clement and Thas (2007), and Clement (2007).

4. STATISTICAL INFERENCE PROCEDURE

The parameters of the seasonal component and the nonlinear trend are a linear combination of the responses, $\hat{\boldsymbol{\beta}} = \mathbf{H}_\beta \mathbf{Y}_N$ and $\hat{\mathbf{f}} = \mathbf{H}_f \mathbf{Y}_N$. Thus inference on the mean parameters and the nonlinear trend require an estimator of the complete variance–covariance matrix of the observation vector $\mathbf{Y}_N$. A plug-in

DOI: 10.1002/env

estimator of $\mathbf{\Sigma}_{Y_N}$ can be calculated by using Equation (6). Let $\hat{\mathbf{\Sigma}}_{Y_N}$ denote this estimator. The variance–covariance matrix of the parameter estimators of the seasonal effect is thus consistently estimated by

$$\hat{\mathbf{\Sigma}}_{\beta} = \mathbf{H}_{\beta} \hat{\mathbf{\Sigma}}_{Y_N} \mathbf{H}_{\beta}^{T} \tag{12}$$

To know whether the nonlinear trend is present at a certain time t, an analysis of its derivative, $\mathbf{f}^{(1)}(t)$, is proposed. For local linear regression smoothers the derivative can be calculated and is linear in the response. Since projection matrices exist for the local polynomial regression smoother and its first derivative, the calculation of the estimators of the variance–covariance matrix of the nonlinear trend ($\mathbf{\Sigma}_f$) and of the derivative ($\mathbf{\Sigma}_{f^{(1)}}$) is straightforward. They are given by

$$\hat{\Sigma}_f = \mathbf{H}_f \hat{\Sigma}_Y \mathbf{H}_f^{T} \tag{13}$$

$$\hat{\Sigma}_{f^{(1)}} = \mathbf{H}_{f^{(1)}} \hat{\Sigma}_Y \mathbf{H}_{f^{(1)}}^{T} \tag{14}$$

Simple test statistics can thus be used for asymptotic pointwise inference on the derivatives, for example $t = f^{(1)}(t)/s_{f^{(1)}}(t)$ is asymptotically standard normally distributed under the null hypothesis of no trend. These tests are performed at each individual time instant and we therefore have to correct for multiplicity. A widely used method to take multiplicity into account is to use adjusted p-values. Well-known examples of this approach are classical methods such as the Bonferroni or Holm procedures. They assume all the tests to be independent and they are known to be too conservative when this is not the case (e.g. Shaffer, 1995). In our application, tests at time instants which are close to one another are likely to be correlated. Thus, the effective dimension of the multiplicity problem is reduced. We therefore propose to use a procedure which can take these dependences explicitly into account. In particular we have chosen to adapt the free step-down resampling method (algorithm 2.8 of Westfall and Young, 1993) to our application. The procedure proceeds as follows:

1. Rank the original p-values: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(n)}$, where (j) denotes the rank number and store the ranked p-values in the vector $(p_{(1)}, \ldots, p_{(n)})$.
2. Initialise the count variables: $\text{COUNT}_i = 0, i = 1, \ldots, n$.
3. Generate a vector $(p^*_{(1)}, \ldots, p^*_{(n)})$ from the same (or at least, approximately the same) distribution of the *original* p-values $(p_{(1)}, \ldots, p_{(n)})$ under the complete null hypothesis. Note that the sequence $\{(j)\}$ is fixed throughout the simulation. Thus the $p^*_{(j)}$ will not have the same monotonicity as the original p-values $p_{(j)}$.
4. Define the successive minima to enforce the same monotonicity:

$$\begin{aligned} q^*_n &= p^*_{(n)} \\ q^*_{n-1} &= \min(q^*_n, p^*_{(n-1)}) \\ q^*_{n-2} &= \min(q^*_{n-1}, p^*_{(n-2)}) \\ &\vdots \\ q^*_1 &= \min(q^*_2, p^*_{(1)}) \end{aligned}$$

5. If $q^*_i \leq p_{(i)}$, then $\text{COUNT}_i = \text{COUNT}_i + 1$.
6. Repeat steps (c)–(e) B times, compute the adjusted p-values $\tilde{p}^{(B)}_{(i)}$ as $\tilde{p}^{(B)}_{(i)} = \frac{\text{COUNT}_i}{B}$.

DOI: 10.1002/env

7. Enforce monotonicity using successive maximisation:

$$\begin{aligned}
\tilde{p}_{(1)}^{(B)} &= \tilde{p}_{(1)}^{B} \\
\tilde{p}_{(2)}^{(B)} &= \max(\tilde{p}_{(1)}^{(B)}, \tilde{p}_{(2)}^{(B)}) \\
&\vdots \\
\tilde{p}_{(n)}^{(B)} &= \max(\tilde{p}_{(n-1)}^{(B)}, \tilde{p}_{(n)}^{(B)})
\end{aligned}$$

Westfall and Young (1993) argue that once the monotonicity is enforced, and if B is sufficiently large, that the $\tilde{p}_{(j)}^{(B)}$ are reasonable approximations of the actual $\tilde{p}_{(j)}$. They also recommend to take $B \geq 10\,000$. One of the implementations Westfall and Young (1993) proposed for step (c), is to sample from a parametric estimate of the null distribution $\hat{F}_0$. When F_0 is a known function that depends on a vector of unknown parameters $\boldsymbol{\Theta}$, $F_0 = F_0(\boldsymbol{\Theta})$, one can sample from $\hat{F}_0 = F_0(\hat{\boldsymbol{\Theta}})$, where $\hat{\boldsymbol{\Theta}}$ is a consistent estimator of $\boldsymbol{\Theta}$. In our application, a simulated sample from $\hat{F}_0$ can be obtained by

1. sampling a new set of derivatives $\mathbf{f}^{(1)*}$ under the null hypothesis of no trend from MVN($\mathbf{0}$, $\hat{\boldsymbol{\Sigma}}_{f^{(1)}}$),
2. calculating the p-values p_k^* that correspond to each of the simulated derivatives $f_k^{(1)*}$, and
3. ranking these p-values according to the *original ranked* p-values $(p_{(1)}, \ldots, p_{(n)})$ to obtain $(p_{(1)}^*, \ldots, p_{(n)}^*)$.

In the above, inference is provided for the components of the mean model. To obtain standard errors of the parameter estimators of the dependence structure, we will estimate the observed Fisher information matrix. In this manuscript, this is done by numerical perturbation of the likelihood function (e.g. Harvey, 1989; Shumway and Stoffer, 2006).

5. CASE STUDY

The data used in this case study are part of a public database of the Flemish Environmental Agency (http://www.vmm.be). Five sampling locations along two joining river reaches located in the Yzer basin (Belgium) are used to assess whether there exists a trend in the nitrate concentrations between January 1990 and December 2003. Their DAG is represented in Figure 1. Sampling locations S1, S2, S4 and S5 are located on the Yzer while sampling location S3 is located on a joining creek. For each sampling location, monthly nitrate measurements are available between January 1990 and December 2003. Hence, the five sampling locations are sampled on 168 different time instants resulting in a total sample size of 840 observations. Since the observations are taken at time intervals which are much larger than the time scale of the water flow, the matrix $\mathbf{B}$ describing the temporal correlation, can be assumed to be diagonal, that is an AR(1) structure. Instead of looking for trends at the level of individual sampling locations, we aim to detect the trend on a more regional scale and impose the restriction that all locations have the same trend in common. The nonlinear trend is estimated by means of a local polynomial regression smoother for which we use the Epanechnikov kernel. Although the choice of the kernel is not that important from a practical point of view according to Fan and Gijbels (1996), they showed that the Epanechnikov kernel has some nice asymptotical properties. The bandwidth was selected by a grid search using the AIC criterion.

DOI: 10.1002/env

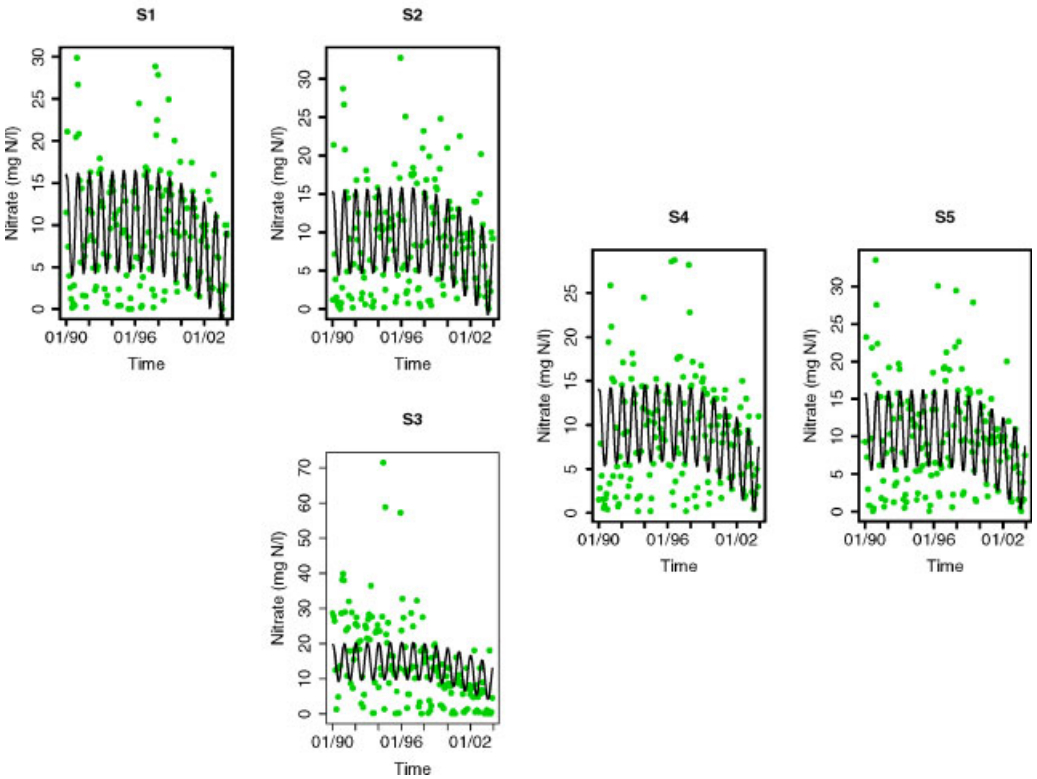

Figure 2. Model fit at five sampling locations of the river Yzer according to model II. Sampling locations S1, S2, S4, S5 are located on the main river and sampling location S3 is located on a tributary which drains into the Yzer between S2 and S4

We used the following mean model:

$$\begin{aligned} \mathrm{E}\left[y_{it}\right] &= \mu + \alpha_i + f(t) + \gamma_1 \sin(2\pi t/12) + \gamma_2 \cos(2\pi t/12) \\ &\quad + (\alpha\gamma)_{i1} \sin(2\pi t/12) + (\alpha\gamma)_{i2} \cos(2\pi t/12) \end{aligned} \tag{15}$$

where μ denotes the intercept at sampling location S5, and α_i the effect of the ith sampling location relative to sampling location S5 (hence $\alpha_5 = 0$). The value of the regional trend at time t is denoted by $f(t)$, the γ_k are the parameters of the seasonal component modelled by Fourier terms, and the $(\alpha\gamma)_{ik}$ are parameters of the sampling location–season interactions. The resulting OLS fit of the mean model is shown in Figure 2. The plot clearly shows the presence of seasonal variation and a decreasing trend

DOI: 10.1002/env

Table 1. Parameter estimates, standard errors and p-values for the linear part in the mean model (Model II)

Parameter	Estimate	Standard error	p-Value
μ	9.85	0.20	0.00
α_1	−0.66	0.07	0.00
α_2	−0.79	0.14	0.00
α_3	3.84	10.20	0.71
α_4	−0.90	0.14	0.00
γ_2	3.70	0.38	0.00
γ_1	3.71	0.39	0.00
$(\alpha\gamma)_{1,1}$	−0.18	0.15	0.23
$(\alpha\gamma)_{2,1}$	−0.04	0.25	0.87
$(\alpha\gamma)_{3,1}$	−0.23	1.20	0.85
$(\alpha\gamma)_{4,1}$	−0.92	0.26	0.00
$(\alpha\gamma)_{1,2}$	1.23	0.15	0.00
$(\alpha\gamma)_{2,2}$	0.50	0.25	0.04
$(\alpha\gamma)_{3,2}$	0.36	1.18	0.76
$(\alpha\gamma)_{4,2}$	−0.28	0.26	0.29

from 1998 until the end of the time series. The parameter estimates of the mean model are given in Table 1.

The parameters of the dependence structure consist of the elements of the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{\Sigma}_\eta$ and $\mathbf{\Sigma}_\epsilon$. Their estimates are listed below (standard errors are shown between brackets):

$$\hat{\mathbf{A}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.77\ (0.52) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1.05\ (0.42) & 0.04\ (0.07) & 0 & 0 \\ 0 & 0 & 0 & 0.39\ (0.29) & 0 \end{bmatrix}$$

$$\hat{\mathbf{B}} = \begin{bmatrix} 0.14\ (0.7) & 0 & 0 & 0 & 0 \\ 0 & 0.35\ (0.17) & 0 & 0 & 0 \\ 0 & 0 & 0.98\ (0.02) & 0 & 0 \\ 0 & 0 & 0 & -0.14\ (0.16) & 0 \\ 0 & 0 & 0 & 0 & 0.12\ (0.21) \end{bmatrix}$$

$$\hat{\mathbf{\Sigma}}_\eta = \begin{bmatrix} 12\ (13) & & & & \\ 0 & 0.01\ (5.9) & & & \\ 0 & 0 & 1.11\ (1.0) & & \\ 0 & 0 & 0 & 14.2\ (14.9) & \\ 0 & 0 & 0 & 0 & 0.02\ (7.40) \end{bmatrix}$$

DOI: 10.1002/env

and

$$\hat{\boldsymbol{\Sigma}}_{\epsilon} = \begin{bmatrix} 12\ (13.7) & & & & \\ 8.0\ (6.9) & 21.6\ (6.2) & & & \\ 10.4\ (3.7) & 13.5\ (4.1) & 90.4\ (10.4) & & \\ 4.9\ (8.2) & 6.7\ (6.4) & 4.6\ (3.8) & 3.2\ (18.8) & \\ 17.0\ (4.7) & 18.8\ (4.3) & 13.6\ (4.3) & 8.3\ (10.4) & 26.9\ (11.6) \end{bmatrix}$$

Thanks to the additive model structure, the contribution of each predictor can be studied individually. This enables us to decompose the model into components that can be represented graphically. The trend and its derivative are shown in Figure 3, along with 95% asymptotic pointwise confidence intervals. A naive approach to assess the trend is to perform a t-test at each individual time instant. An equivalent result is obtained by assessing at which time instants zero is not contained in the 95% confidence intervals of the derivatives. In the figure it can be seen that this is the case between January 1999 until July 2003.

For the test procedure to be formally valid, a multiplicity correction is needed to control the familywise Type I error at the α-level instead of controlling the Type I error of the individual tests. When the Holm procedure was applied to correct for multiplicity, no significant results were observed (results not shown). As explained previously, the Holm procedure is too conservative in our setting because it acts as if all tests are independent. The free step-down resampling method of Westfall and Young (1993), that was adapted to our problem, takes the dependences between the tests explicitly into account. The results of this approach are illustrated in Figure 3. Familywise significant first derivatives ($\alpha = 5\%$) are indicated with dots superimposed on the point estimates. The derivatives of the nonlinear trend are significantly different from 0 between September 1999 and January 2002. Since the estimates of the first derivatives of the nonlinear trend are negative, a significant decrease in nitrate concentration is concluded for this period. The fact that we exploited the dependences between the t-tests clearly leads to a less conservative test procedure than classical corrections such as the Holms procedure. Compared to the naive approach, the nonlinear trend is not significant in 2002 and 2003. This is not surprising as the variance of the predictions based on smoothers is typically inflated in the boundary regions.

An autocorrelation plot of the standardised innovations revealed no residual autocorrelation of the innovations of Y (results not shown), indicating that the AR(1) structure is good. The analysis of the innovations showed that the assumption of the regional trend was acceptable. No systematic pattern remained in the innovations at the sampling locations of the main river (S1–S4). Only in the sampling location of the joining creek, S3, a slightly decreasing trend was observed in the innovations during the first 3 years. Although it was not considered in this case study, our method easily allows the user to specify interactions between time and sampling location since it only involves the specification of the mean model.

From the case study it can be concluded that a decreasing trend in the nitrate concentration is established between the introduction of the first MD and the second MD in the study region. The trend remains significant until January 2002. Although this data analysis methodology has no causal interpretation, it suggests that the introduction of the MDs had a beneficial effect on the nitrate status in the study region.

DOI: 10.1002/env

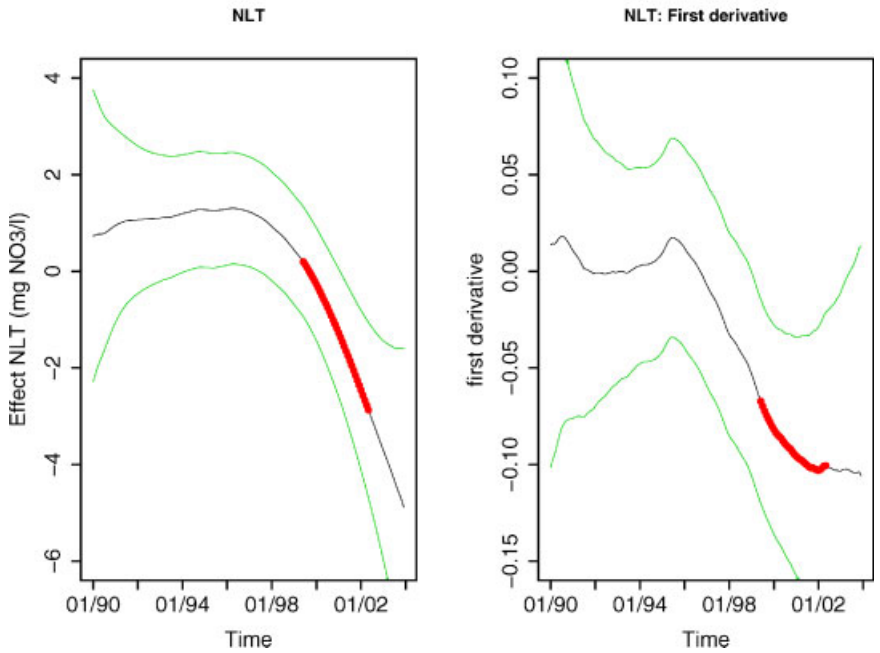

Figure 3. Evaluation of the common nonlinear trend along the river Yzer. The estimated trend is presented in the left panel, and its first derivative is shown in the right panel. In both graphs, 95% pointwise confidence bands are depicted. Familywise significant decreases are indicated with a dot superimposed on the point estimate

6. DISCUSSION AND CONCLUSIONS

In this paper a statistical methodology has been developed for the detection of nonlinear trends in river monitoring network data. A spatio-temporal model has been constructed to model the marginal mean and the dependence structure. According the specification of the marginal mean model, the trend can be studied at the level of individual sampling locations, or on a more regional scale.

In contrast with existing methodologies for trend detection, our procedure takes the spatio-temporal dependence explicitly into account. As compared to ad hoc methods, such as for example the methods based on the field significance (e.g. Burn and Hag Elnur, 2002), our method provides statistical inference which is formally valid. For the detection of trends in water quality, the use of a nonparametric regression method is more flexible. Classical nonparametric tests such as Mann–Kendall tests for trend detection are not appropriate when sign changes occur in first derivative of the trend. Our method also enables the detection of trends on a more local time scale. To verify at which time instants the nonlinear trend is beneficial, t-tests are performed at each time instant. Due to the specific dependence between these tests, classical multiplicity corrections are too conservative. We have adopted the free step-down resampling method of Westfall and Young (1993) and sampled from an appropriate null distribution to account for the dependences between the statistical tests.

The methodology has been illustrated in a case study where a significant decrease in the nitrate concentration was detected in the study region between September 1999 and January 2002 ($\alpha = 0.05$), suggesting a beneficial effect of the introduction of the MDs in Flanders.

DOI: 10.1002/env

ACKNOWLEDGEMENTS

Part of this research was supported by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy). The authors would also like to thank the Flemish Environmental Agency (VMM) for providing the data.

REFERENCES

Brillinger D. 1989. Consistent detection of a monotonic trend superposed on a stationary time series. *Biometrika* **76**(1): 23–30.
Burn DH, Hag Elnur MA. 2002. Detection of hydrologic trends and variability. *Journal of Hydrology* **255**: 107–122.
Cai Z, Tiwari RC. 2000. Application of a local linear autoregressive model to BOD time series. *Environmetrics* **11**: 341–350.
Canjels E, Watson M. 1997. Estimating deterministic trends in the presence of serially correlated errors. *The Review of Economics and Statistics* **79**(2): 184–200.
Clement L. 2007. Statistical validation and spatio-temporal modelling of river monitoring networks. *Ph.D. Thesis*, Ghent University, Ghent, Belgium.
Clement L, Thas O. 2007. Estimating and modelling spatio-temporal correlation structures for river monitoring networks. *Journal of Agricultural, Biological, and Environmental Statistics* **12**(2): 161–176.
Clement L, Thas O. 2008. Testing for trends in the violation frequency of an environmental threshold in rivers. *Environmetrics* (early view), DOI: 10.1002/env.911.
Clement L, Thas O, Vanrolleghem PA, Ottoy JP. 2006. Spatio-temporal statistical models for river monitoring networks. *Water Science and Technology* **53**(1): 9–15.
Cleveland RB, Cleveland WS, McRae J, Terpenning I. 1990. STL: a seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics* **6**: 3–73.
EC. 2000. Directive 2000/60/EC of the European Parliament and of the Council of October 23 2000 establishing a framework for community action in the field of water policy. *Official Journal of the European Communities* **43**(22.12.2000): L327/1–L327/72.
El-Sharaawi A. 1995. Trend detection and estimation with environmental applications. *Mathematics and Computurs in Simulation* **39**: 441–447.
El-Sharaawi A, Niculescu S. 1993. A simple test for detecting non-linear trend. *Environmetrics* **4**: 233–242.
Esterby S. 1993. Trend analysis methods for environmental data. *Environmetrics* **4**: 459–481.
Esterby S. 1996. Review of methods for the detection and estimation of trends with emphasis on water quality applications. *Hydrological Processes* **10**: 127–149.
Fan J, Gijbels I. 1996. *Local Polynomial Modelling and Its Applications* (1st edn). Monographs on Statistics and Applied Probability. Chapman & Hall: New York.
Giannitrapani M, Bowman AW, Scott EM. 2005. Additive models for correlated data with applications to air pollution monitoring. *Technical Report*, Department of Statistics, University of Glasgow.
Harvey AC. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter* (1st edn). Cambridge University Press: Cambridge.
Hastie TJ, Tibshirani RJ. 1990. *Generalized Additive Models* (1st edn). Monographs on Statistics and Applied Probability. Chapman & Hall: New York.
Hirst D. 1998. Estimating trends in stream water quality with a time-varying flow relationship. *Austrian Journal of Statistics* **27**: 39–48.
Maruyama GM. 1997. *Basics of Structural Equation Modeling*. Sage Publications, Inc.: Thousand Oaks.
McMullan A. 2004. Non-linear and nonparametric modelling of seasonal environmental data. *Ph.D. Thesis*, University of Glasgow, Glasgow, UK.
McMullan A, Bowman AW, Scott E. 2003. Non-linear and nonparametric modelling of seasonal environmental data. *Computational Statistics* **18**: 167–183.
Shaffer JP. 1995. Multiple hypothesis testing. *Annual Review of Psychology* **46**: 561–584.
Shin DW, Oh MS. 2002. Asymptotic efficiency of the ordinary least squares estimator for regressions with unstable regressors. *Economic Theory* **18**: 1121–1138.
Shumway RH, Stoffer DS. 1982. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis* **3**: 253–264.
Shumway RH, Stoffer DS. 2006. *Time Series Analysis and Its Applications* (2nd edn). Springer Texts in Statistics. Springer-Verlag: New York.
Sun H, Pantula S. 1999. Testing for trends in correlated data. *Statistics & Probability Letters* **41**: 87–95.
Thas O, Vooren LV, Ottoy JP. 1998. Nonparametric test performance for trends in water quality with sampling design applications. *Journal of the American Water Resources Association* **34**(2): 347–357.

Vlaams Parlement. 1995. Decreet van 20 december 1995 tot wijziging van het decreet van 23 januari 1991 inzake de bescherming van het leefmilieu tegen de verontreiniging door meststoffen. In *Belgisch Staatsblad* (5th edn), Vol. 249. 36069. Belgian Government: Brussels.

Vlaams Parlement. 1999. Decreet van 11 mei 1999 tot wijziging van het decreet van 23 januari 1991 inzake de bescherming van het leefmilieu tegen de verontreiniging door meststoffen en tot wijziging van het decreet van 28 juni 1985 betreffende de milieuvergunning. In *Belgisch Staatsblad* (1st edn). 30967. Belgian Government: Brussels.

Westfall PH, Young SS. 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. John Wiley and Sons: New York.

Whittaker J. 1990. *Graphical Models in Applied Multivariate Statistics* (1st edn). John Wiley & Sons Ltd: Chichester.

Wu WB, Zhao Z. 2007. Inference of trends in time series. *Journal of the Royal Statistical Society: Series B* **69**: 391–410.

Zheng X, Basher R. 1999. Structural time series models and trend detection in global and regional temperature series. *Journal of Climate* **12**: 2347–2358.

Zheng X, Basher R. 2006. Semiparametric estimation and testing of the trend of temperature series. *Econometrics Journal* **9**: 332–355.

APPENDIX

The dependence structure $\boldsymbol{\Psi} = (\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_{\mathbf{S_0}}, \boldsymbol{\Sigma}_\eta, \boldsymbol{\Sigma}_\epsilon)$ can be estimated by the following algorithm (the complete derivation can be found in Clement and Thas (2007) and Clement (2007)):

1. Choose initial estimates, $\boldsymbol{\Psi}^0$.
2. **E-step**
   (a) Apply the Kalman filter recursions given the current values of the parameters $\boldsymbol{\Psi}^{\boldsymbol{k}}$.
   For $t = 1, \ldots, N$ the following forward recursions are used:

$$\mathbf{a}_{t|t-1} = \boldsymbol{\Phi}\mathbf{a}_{t-1|t-1} \tag{16}$$

$$\mathbf{P}_{t|t-1} = \boldsymbol{\Phi}\mathbf{P}_{t-1|t-1}\boldsymbol{\Phi}^T + \mathbf{Q} \tag{17}$$

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{F}_t^{-1}\mathbf{v_t} \tag{18}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{F}_t^{-1}\mathbf{P}_{t|t-1} \tag{19}$$

$$\mathbf{F}_t = \mathbf{P}_{t|t-1} + \boldsymbol{\Sigma}_\epsilon^k, \tag{20}$$

$$\mathbf{P}_{t,t-1|t} = (\mathbf{I} - \mathbf{P}_{t|t-1}\mathbf{F}_t^{-1})\boldsymbol{\Phi}\mathbf{P}_{t-1|t-1}, \tag{21}$$

   where $\mathbf{a}_{t|s} = E[\mathbf{S_t}|\mathbf{Y_1}, \ldots, \mathbf{Y_s}]$, $\boldsymbol{\Phi} = (\mathbf{I} - \mathbf{A^k})^{-1}\mathbf{B^k}$, $\mathbf{P}_{t|s} = \mathrm{var}[\mathbf{S_t}|\mathbf{Y_1}, \ldots, \mathbf{Y_s}]$, $\mathbf{Q} = (\mathbf{I} - \mathbf{A^k})^{-1}\boldsymbol{\Sigma}_\eta^k(\mathbf{I} - \mathbf{A^k})^{-T}$, the innovations $\mathbf{v}_t = \mathbf{Y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}} - \mathbf{a}_{t|t-1}$, $\mathbf{P}_{t,t-1|s} = \mathrm{cov}[\mathbf{S_t}, \mathbf{S_{t-1}}|\mathbf{Y_1}, \ldots, \mathbf{Y_s}]$, $\mathbf{a_{0|0}} = E[\mathbf{S_0}] = \mathbf{0}$ and $\mathbf{P_{0|0}} = \mathrm{var}(\mathbf{S_0}) = \boldsymbol{\Sigma}_{\mathbf{S_0}}$.
   (b) Apply the Kalman smoother recursions given the current parameter values $\boldsymbol{\Psi}^{\boldsymbol{k}}$.
   For $t = N - 1, \ldots, 0$:

$$\mathbf{a}_{t|N} = \mathbf{a}_{t|t} + \mathbf{P}_t^*(\mathbf{a}_{t+1|N} - \mathbf{a}_{t+1|t}) \tag{22}$$

$$\mathbf{P}_{t|N} = \mathbf{P}_{t|t} + \mathbf{P}_t^*(\mathbf{P}_{t+1|N} - \mathbf{P}_{t+1|t})\mathbf{P}_t^{*T} \tag{23}$$

$$\mathbf{P}_t^* = \mathbf{P}_{t|t}\Phi^T\mathbf{P}_{t+1|t}^{-1} \tag{24}$$

$$\mathbf{P}_{t,t-1|N} = \mathbf{P}_{t,t-1|t} + (\mathbf{P}_{t|N} - \mathbf{P}_{t|t})\mathbf{P}_{t|t}^{-1}\mathbf{P}_{t,t-1|t} \tag{25}$$

DOI: 10.1002/env

(c) Calculate expected sufficient statistics:

$$E[\mathbf{S}_t|\mathbf{Y}] = \mathbf{a}_{t|N} \tag{26}$$

$$E[\mathbf{S}_t\mathbf{S}_t|\mathbf{Y}] = \mathbf{P}_{t|N} + \mathbf{a}_{t|N}\mathbf{a}_{t|N}^T \tag{27}$$

$$E[\mathbf{S}_t\mathbf{S}_{t-1}|\mathbf{Y}] = \mathbf{P}_{t,t-1|N} + \mathbf{a}_{t|N}\mathbf{a}_{t-1|N}^T \tag{28}$$

3. **M-step**

(a) Calculate maximum likelihood estimates:

$$\mathbf{\Sigma}_{S_0}^{k+1} = E[\mathbf{S_0}\mathbf{S_0^T}] \tag{29}$$

$$\mathbf{B}_{\mathbf{i}.}^{[\mathbf{b_i}]k+1} = \left[\sum_{t=1}^{N} S_t^i \mathbf{S}_{t-1}^{[b_i]\,T} - \left(\sum_{t=1}^{N} S_t^i \mathbf{S}_t^{[a_i]T}\right)\left(\sum_{t=1}^{N} \mathbf{S}_t^{[a_i]}\mathbf{S}_t^{[a_i]T}\right)^{-1}\left(\sum_{t=1}^{N} \mathbf{S}_t^{[a_i]}\mathbf{S}_{t-1}^{[b_i]\,T}\right)\right] \times \left[\sum_{t=1}^{N} \mathbf{S}_{t-1}^{[b_i]}\mathbf{S}_{t-1}^{[b_i]\,T} - \left(\sum_{t=1}^{N} \mathbf{S}_{t-1}^{[b_i]}\mathbf{S}_t^{[a_i]T}\right)\left(\sum_{t=1}^{N} \mathbf{S}_t^{[a_i]}\mathbf{S}_t^{[a_i]T}\right)^{-1}\left(\sum_{t=1}^{N} \mathbf{S}_t^{[a_i]}\mathbf{S}_{t-1}^{[b_i]\,T}\right)\right]^{-1} \tag{30}$$

$$\mathbf{A}_{\mathbf{i}.}^{[\mathbf{a_i}]k+1} = \left(\sum_{t=1}^{N} S_t^i \mathbf{S}_t^{[a_i]T} - \mathbf{B}^{[b_i]k+1}\sum_{t=1}^{N} \mathbf{S}_{t-1}^{[b_i]}\mathbf{S}_t^{[a_i]T}\right)\left(\sum_{t=1}^{N} \mathbf{S}_t^{[a_i]}\mathbf{S}_t^{[a_i]T}\right)^{-1} \tag{31}$$

$$\sigma_{\eta_i}^{2\,k+1} = \frac{\mathrm{RSS}_i^{k+1}}{N} \tag{32}$$

$$\mathbf{\Sigma}_{\epsilon}^{k+1} = \frac{\sum_{t=1}^{N}(\mathbf{Y}_t'\mathbf{Y}_t'^T) - \sum_{t=1}^{N}(\mathbf{Y}_t'\mathbf{S}_t^T) - \sum_{t=1}^{N}(\mathbf{S}_t\mathbf{Y}_t'^T) + \sum_{t=1}^{N}(\mathbf{S}_t\mathbf{S}_t^T)}{N} \tag{33}$$

where $\mathbf{A}_{i.}^{[a_i]}$ are the non-zero elements of the ith row of matrix $\mathbf{A}$ and $\mathbf{S}_t^{[a_i]}$ are the corresponding parents of $\mathbf{S}^i$ in the DAG at time step t, $\mathbf{B}_{i.}^{[b_i]}$ are the non-zero elements of the ith row of matrix $\mathbf{B}$ and $\mathbf{S}_{t-1}^{[b_i]}$ are the corresponding sampling locations at the previous time step, $\mathrm{RSS}_i^{k+1} = \sum_{t=1}^{N}(S_t^i - \mathbf{A}_{\mathbf{i}.}^{[\mathbf{a_i}]k+1}\mathbf{S}_t^{[a_i]} - \mathbf{B}_{\mathbf{i}.}^{[\mathbf{b_i}]k+1}\mathbf{S}_{t-1}^{[b_i]})^2$ and $\mathbf{Y}_t' = \mathbf{Y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}$.

(b) Replace the sufficient statistics by their conditional expectations calculated in step 2.

4. Repeat steps 2 and 3 until convergence.

DOI: 10.1002/env