Spatially Varying Temperature Trends in a Central California Estuary

Ricardo T. Lemos[*], Bruno Sansó[†] and Marc Los Huertos[‡]

Abstract

We consider monthly temperature data collected over a period of 16 years at 24 stations in the Elkhorn Slough National Estuary, located in the Monterey Bay area in Central California, USA. Our goal is to develop a statistical model in order to separate the annual cycle, short term fluctuations and long term trends, while accounting for the spatial variability of these features. In the model, each station has a specific, time-invariant mixture of two seasonal components, to encompass the spatial gradient of oceanic influence. Likewise, long term trends are modeled as local mixtures of two components. Finally, all stations share a common baseline, whose temporal variability is linearly dependent on a variable that summarizes several atmospheric measurements. We use a Bayesian approach with a purposely developed Markov chain Monte Carlo method to explore the posterior distribution of the parameters. We find that seasonal patterns have changed in time, that neighboring stations can have substantially different behaviors and that most stations show significant warming trends.

[*]Maretec–IST, Avenida Manuel da Maia, 36, 3°esq., 1000–201 Lisboa. rtl@net.sapo.pt

[†]Department of Applied Mathematics and Statistics, University of California, Santa Cruz, U.S.A. bruno@ams.ucsc.edu, www.ams.ucsc.edu/∼bruno

[‡]Science and Environmental Policy, California State University Monterey Bay, Seaside CA, U.S.A. Marc_Loshuertos@csumb.edu

Key words: Bayesian Modeling; Mixture models; Space-time models; Water quality data; Time-varying trends.

1 Introduction

Estuaries are the most highly anthropogenically impacted of all habitat types (Edgar et al., 2000), yet they host rich, distinctive biodiversity, including migratory shorebirds, nursery fishes, among others (Laprise and Dodson, 1994; Whitfield, 1994; Price, 2002). Even though Elkhorn Slough is a small, shallow estuary on a national scale, it is the largest estuary between San Francisco and Morro Bay, California USA. It has an extremely important ecological role with a variety of habitats, including extensive marshes and mudflats. Each estuary has unique, often complicated, spatial and temporal characteristics. Therefore, it is a challenge to develop and maintain monitoring programs that capture this variability while developing record lengths long enough to discern water quality trends in the context of seasonal variation.

One of the goals of the Elkhorn Slough National Estuarine Research Reserve (ESNERR) is to examine spatial and temporal variation water quality and nutrient concentrations, to assess changes in ecosystem status of the whole estuary and at the site specific level. Under that scope, we consider if there are significant long-term trends in temperature during the period 1988–2004 and if these changes occur similarly across the reserve, or some regions present stronger signals. We examine records from 24 sites in the ESNERR where monthly water quality data have been collected for as long as 16 years. The slough is located in the Monterey Bay area, between the cities of Watsonville and Salinas, along the central coast of California. Much of the surrounding land use is dominated by row crop agriculture. The locations of the stations can be seen in Figure 1. They have been renumbered based on their ranked distance from the mouth of the slough. More information about the Elkhorn Slough is available from `http://www.elkhornslough.org/`.

We expected that, of all the water quality measures, temperature should be the most tractable. For example, nutrient concentrations tend to be highly variable, and dissolved

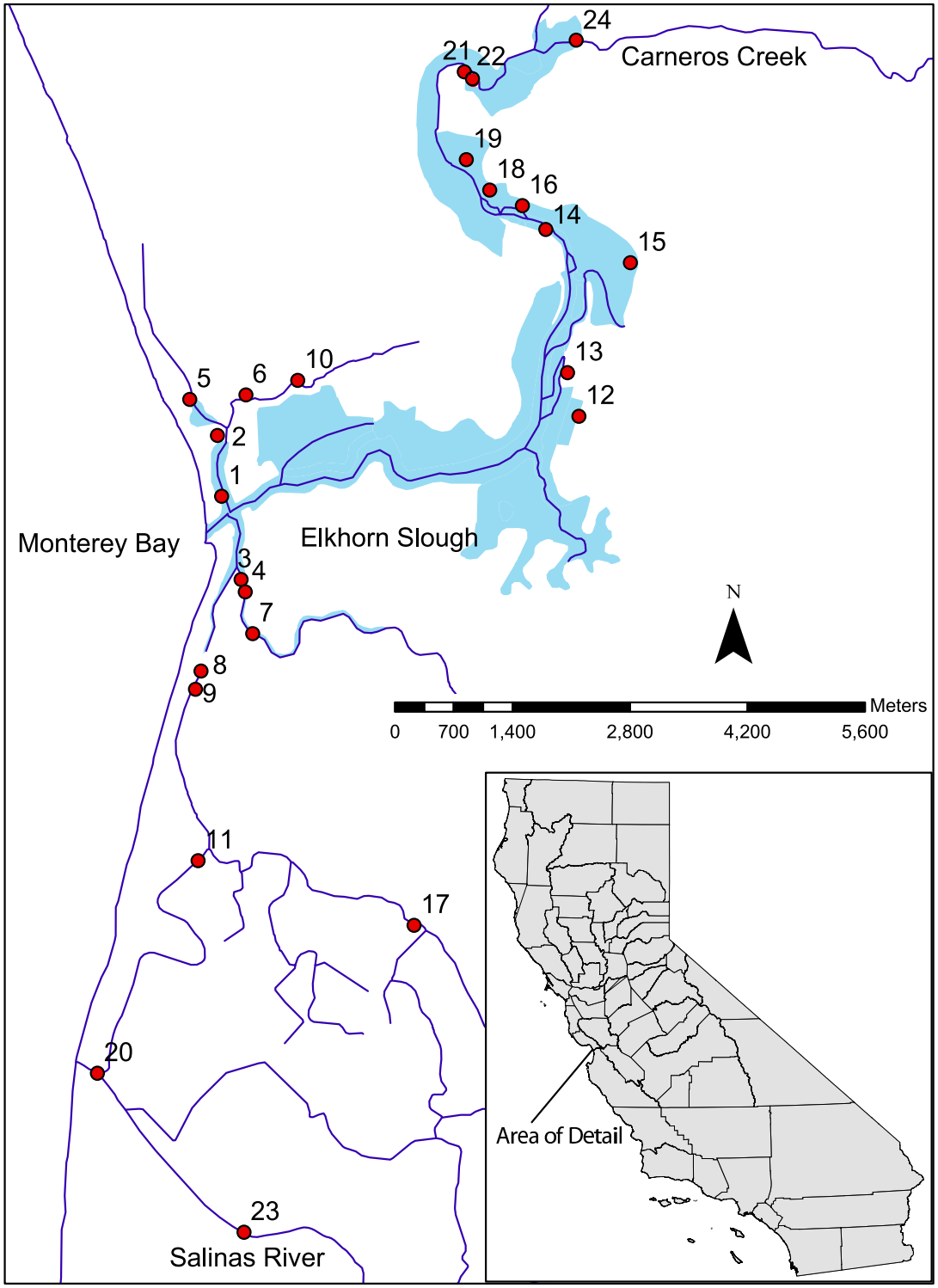

Figure 1: Locations of the monitoring stations in the Elkhorn Slough. Shaded areas show the extent of the tidal salt marsh and continuous lines represents riverine or tidal channels

oxygen or turbidity measures are controlled by several factors simultaneously (e.g., sediment concentration, planktonic and benthic algae, water mixing, nutrient concentrations, and day length). The results presented in this paper were obtained after using increasingly complex statistical approaches. At the onset, visual inspection of the data strongly suggests temperatures and their seasonal amplitudes are increasing in the slough in most of the 24 sampled sites. In this paper we carefully quantify such trends.

Water temperature is strongly influenced by human changes to the slough's hydrology, which date back to the 1880s. Because the main channel is artificially opened to create a harbor for fishing and research vessels, the slough experiences the semi-diurnal tidal action. In the lower portion of the slough, water has a residence time of less than one day. In the northern portion of the main channel, water has a longer residence time, estimated to be 3 weeks (per. comm. Steve Monosmith). Currently, the main channel extends inland for 11.4 km from Monterey Bay in Central California (Figure 1). The slough includes a variety of habitats that receive seawater exchange through the mouth and terrestrial freshwater from a few seasonal streams in upper Elkhorn main channel (Carneros and Corncob Canyon Creek), plus flow from Salinas River, via the Old Salinas Channel, and runoff from local terrestrial sources. Thus, some areas have limited tidal exchange and temperature is isolated from the main channel, while other areas have extremely high tidal exchange and temperatures similar to the ocean. Flow in many of the freshwater sources is augmented by agricultural run-off, often in the form of summer irrigation tailwater or winter storm driven events.

The main results from our analysis are summarized in the quantification of the temperature trends. We show that such trends vary substantially from month to month and from station to station. In most cases the summer temperature has increased by up to 5 °C in 16 years, while a comparable decrease has been observed in April in other stations. Establishing the cause of this increase was beyond the scope of this effort.

The paper is organized as follows: in the next section we discuss the salient features of the data, based on an exploratory data analysis; in Section 3 we present a statistical model and discuss the Monte Carlo method that we use to estimate the parameters and

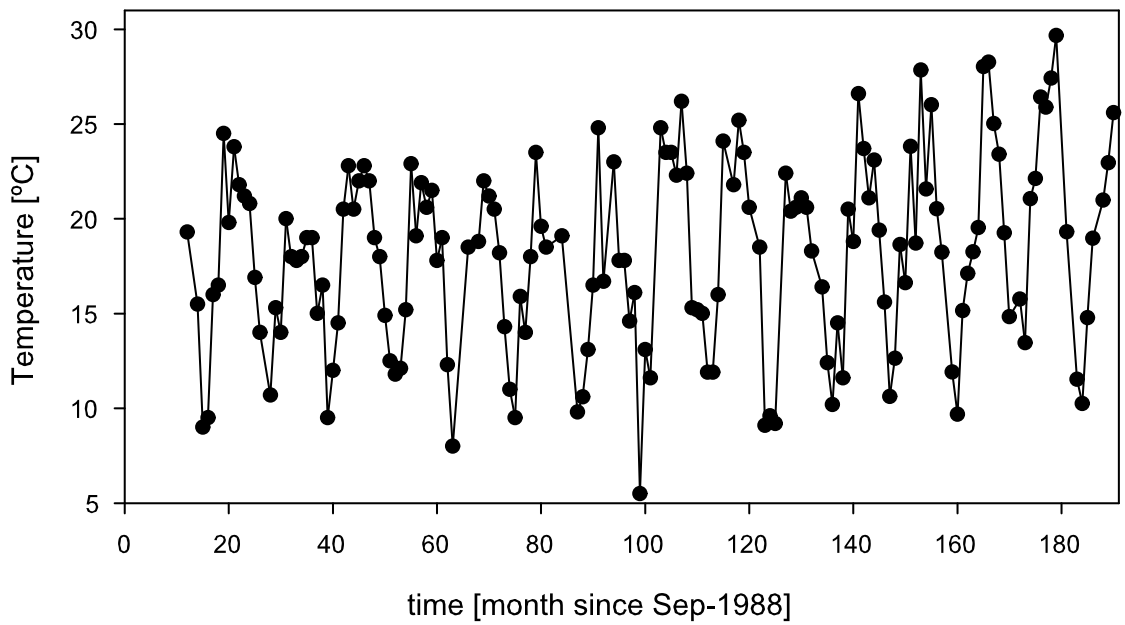

Figure 2: Time series of the temperature readings for Station 22. This is one of the stations with the most complete record.

check the goodness of fit; in Section 4 we present the results obtained from the model and in the last section we present a discussion.

2 Exploratory data analysis

The data considered in this paper are the result of a 16-year program carried out by a combination of volunteers and professional support. Starting in the fall of 1988, several stations in the Elkhorn Slough were sampled for temperature, salinity, pH, turbidity, dissolved oxygen, ammonium, nitrate, and soluble reactive phosphorous. The number of stations increased over time, from 6 to 24. As an illustration of the data under analysis we present the plot of the temperature measured at Station 22 in Figure 2.

Temperature measurements were made with YSI multiprobe sensors using thermistor technology. The accuracy of temperature measurements is $\pm 0.15\%$°C with a resolution

of 0.01°C. Temperature thermistors are very reliable and require no calibration or maintenance (YSI, Series 6 Owners Manual, 069300B). They exhibit less than 0.01°C drift that is usually associated with a change in the thermistor resistance. Thermistor drift is generally caused by exposure to high temperatures, i.e. well outside the range of values in estuarine environments. This information is important to exclude the possibility that the observed trends are due to equipment malfunction.

Although data were nominally collected on a monthly basis, some observations were missing. On the other hand, there were a few months where two samples were taken. When a month with no samples followed or was preceded by a month with two samples, these were often close to the extreme days of the month. In those cases, we moved the sample closest to the empty month. Otherwise, we removed the sample with fewest measured variables. Even after this 22% of the data were still completely missing.

Our exploratory data analysis began with time series plots of temperature at the 24 stations. Given the strong seasonal signal that was immediately apparent, we attempted to model each time series as the sum of a stationary seasonal cycle, with sinusoidal shape, a linear trend and random white noise. Preliminary results derived from these models revealed several important features in the data. First, the seasonal signal varied markedly from station to station in terms of amplitude, ranging between 4.5°C and 12.5°C. In contrast, the phase seemed to be locked, with the yearly minimum occurring in January. Secondly, 19 out the 24 stations presented long-term warming trends, up to 4°C in 16 years. Despite being smaller than yearly amplitudes, these trends seemed evident upon visual inspection of fitted values together with observations. The significance of such trends is supported by a seasonal Kendall test, which consists of a non-parametric test well suited to data with strong seasonal patterns (Helsel and Hirsch, 1992). In this case, we used month as the season (i.e. 12 seasons per year). Results suggested that 13 of the 24 sites had a significant increase in temperature over the sampling period. The increasing temperatures range from 1.1 to 5.4°C/(16 years). Stations where significant trends are present are scattered throughout the slough and do not reflect an obvious spatial pattern. In some stations, an amplification of the seasonal cycle, with warmer

summer months, seemed to be occurring, rather than a year-round temperature increase.

In contrast to our initial expectations, both trends and seasonal cycles were not always similar for nearby stations. Even when factors such as connectivity and tidal influence were accounted for, we could not find simple methods based on proximity to appropriately estimate the results for a given station given the neighboring ones. This reflects the complex circulation patterns present in estuarine systems, and hinders the interpolation of observations to other locations in the estuary, as we first intended.

More importantly, the simple statistical models mentioned above could not adequately describe temperature variability in the Elkhorn Slough, since residual analysis revealed the presence of substantial unexplained structure. For instance, observations made in April systematically produced positive residuals, indicating that one sinusoidal component was not enough to capture the annual cycle of temperature. As will be depicted below, this cycle was so complex that it required a form-free model where the effects of each month were not linked. On the other hand, significant spatial and temporal correlations were found in the residuals. This impaired any estimate of significance assigned to the trends, and implied the need for additional model parameters. Interestingly, the time series of residuals obtained with stationary form-free models with linear trends were similar among stations. This feature pointed to a common explanatory variable to the short-term variability of temperature in the slough. We considered solar radiation, rainfall and wind speed, measured in the weather station of Castroville, as well as sea surface temperature (SST) data from a National Oceanic and Atmospheric Administration (NOAA) buoy located off the Monterey Bay. From a visual inspection of time-series plots, we concluded that a known combination of some of these variables – potential evapotranspiration – might provide the best regressor.

As we moved into models with form-free seasonal components, we observed that coastal stations, clearly influenced by the entrance of seawater into the slough, displayed a damped annual cycle of temperature variation similar to that of offshore SST. In contrast, temperature in inland stations behaved more like air temperature, cooling slightly more in winter and warming much more during summer. Thus, we hypothesized that each

station's cycle could be a mixture of two seasonal patterns. Although the exact mixture could not be deducted from the station's location, the dimensionality of the problem was greatly reduced, as will be shown below.

3 Statistical model

Let $\theta_{m,y}(s)$ be the temperature of station s, month m and year y. We assume that such temperature can be expressed as the sum of the following: a seasonal component, a trend, an atmospheric factor dependent baseline and random noise. The seasonal component is location dependent and is the result of mixing two seasonal patterns. The trend is also location dependent and results from mixing two trend components. For both the seasonal component and the trend, one of the mixing terms corresponds to a coastal behavior and the other to an inland behavior. More explicitly, the model for $\theta_{m,y}(s)$ can be written as

$$\theta_{m,y}(s) = \alpha(s)\eta_m^{(1)} + (1-\alpha(s))\eta_m^{(2)} + \beta(s)\gamma_m^{(1)}(t-\bar{t}) + (1-\beta(s))\gamma_m^{(2)}(t-\bar{t}) + \lambda_t + \epsilon_{m,y}(s) \quad (1)$$

where $t = t(y,m) = 12(y - y_1) + m$, with $y_1 = 1988$, and $\bar{t} = 95$. For month m, $\eta_m^{(1)}$ and $\eta_m^{(2)}$ are the two temperature seasonal components, which we identify with the inland and coastal seasonal components, respectively. $\gamma_m^{(1)}$ and $\gamma_m^{(2)}$ are the two long-term linear trend components. For station s, $\alpha(s)$ and $\beta(s)$ correspond to weights (between zero and one) assigned to $\eta_m^{(1)}$ and $\gamma_m^{(1)}$, respectively. λ_t corresponds to the short-term temperature variability in the slough. The variability of λ_t has serial correlation and is partially explained using atmospheric factors, as will be seen below. Finally $\epsilon_{m,y}(s)$ is random noise. We will assume that the vector $\boldsymbol{\epsilon_{m,y}} = (\epsilon_{m,y}(s_1), \ldots, \epsilon_{m,y}(s_{24}))'$ follows a multivariate normal distribution with mean zero and covariance matrix V, for all m and y and that $\boldsymbol{\epsilon_{m,y}}$ and $\boldsymbol{\epsilon_{m',y'}}$ and independent if $m \neq m'$ or $y \neq y'$.

The salient features of the model in Equation (1) include two different seasonal patterns described by twelve parameters each. This provides the flexibility needed to capture the lack of symmetry observed in the data, in particular, the dip observed from April to May at some stations. The hydrology of the slough suggests that monthly trends vary with geographical location. A priori, we have no reason to believe that either coastal

or inland stations display stronger long-term trends, because we do not know the cause for such temperature change. Therefore, we use different sets of weights for the seasonal signal (α) and the trends (β). In both cases, the model does not impose any spatial regularity; in fact, all our earlier attempts at considering seasonal patterns or trends that were linked by proximity were unsuccessful. Nor does the model impose any relation between the parameters that change with the month, since this would smooth out the peculiar effect of months like April and December.

Denote the observations taken at station s, month m and year y as $x_{m,y}(s)$. We assume that

$$x_{m,y}(s) = \theta_{m,y}(s) + \chi_{m,y}(s), \quad \chi_{m,y}(s) \sim N(0, \tau_\chi^2). \tag{2}$$

In words, the temperature at a given time and location is subject to a measurement error $\chi_{m,y}(s)$. We assume that such errors are all independent across time and location.

Local weather stations (California Irrigation Management Information System – CIMIS) continuously measure temperature, solar radiation, rainfall, relative humidity, and wind speed and direction and calculate reference evapotranspiration using the Penman-Monteith equation. The resulting output provides daily assessments that might be associated to drivers of short term temperature trends in the slough. We complete the model specified by Equations (1) and (2) by incorporating reference evapotranspiration (ET_0), denoted as z_t, as an explanatory variable for λ_t. Daily reference evapotranspiration expresses the evaporation power of the atmosphere from a standardized vegetated surface. ET_0 has a very regular cyclical pattern with a yearly periodicity, as can be seen in Figure 3. Thus, we assume that z_t and λ_t follow the model

$$\begin{pmatrix} z_t \\ \lambda_t \end{pmatrix} = \begin{pmatrix} 1 & \cos(2\pi t/12) \\ \phi & 0 \end{pmatrix} \begin{pmatrix} \delta_t \\ \kappa_t \end{pmatrix} + \boldsymbol{\zeta}_t, \quad \boldsymbol{\zeta}_t \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_z^2 & 0 \\ 0 & \tau_\lambda^2 \end{pmatrix}\right] \tag{3}$$

$$\begin{pmatrix} \delta_t \\ \kappa_t \end{pmatrix} = \begin{pmatrix} \delta_{t-1} \\ \kappa_{t-1} \end{pmatrix} + \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_\delta^2 & 0 \\ 0 & \tau_\kappa^2 \end{pmatrix}\right] \tag{4}$$

We chose such a simple seasonal model for ET_0 after a preliminary regression analysis suggested that no additional sinusoids are needed.

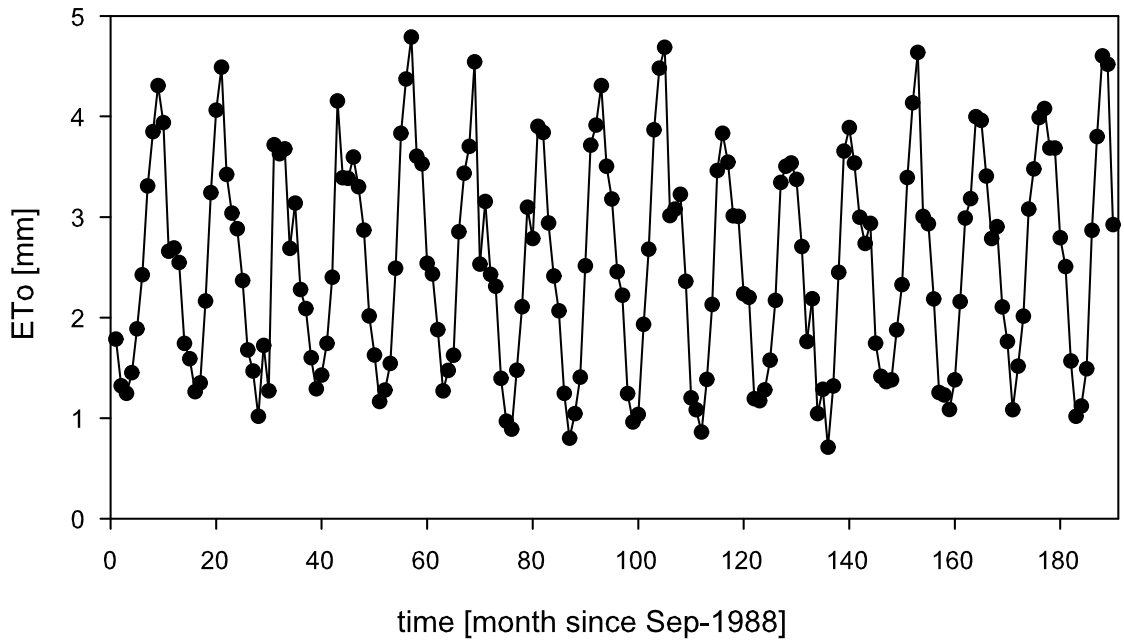

Figure 3: Time series of the reference evapotranspirations (ET_0).

3.1 Prior distributions

We assume that all parameters have independent priors. We use sea surface temperatures from the National Oceanic and Atmospheric Administration buoy 46042, and air temperature data from CIMIS station 19, Castroville, to provide proper normal priors for the coastal and inland average monthly temperatures, respectively. For the trends, δ_1, κ_1 and ϕ, we assign vague normal priors, with mean zero.

For $\alpha(s)$ and $\beta(s)$ we use uniform priors with support $(0, 1)$. The prior for the covariance matrix V is an inverse Wishart with 25 degrees of freedom and scale matrix equal to 25 times the identity. The priors for τ_λ^2, τ_δ^2 and τ_κ^2 are given by inverse gammas with parameters 0.05 and 5. Thus, we are allowing the monthly variability in ET_0 to be explained by changes in either the baseline or the seasonality, and we are not providing much information as to how closely δ_t and λ_t vary. In contrast, τ_χ^2 and τ_z^2 receive informative priors with inverse gamma distribution, with parameters 6.8 and 2000, and 0.05 and 500, respectively. These parameters are set so that twice the expectation of τ_χ and

τ_z are close to the accuracy of water temperature and ET_0 measurements, respectively.

3.2 Fitting the model

In order to fit the proposed model we use a Markov Chain Monte Carlo method (for example, see Gamerman and Lopes, 2006). For most parameters it is possible to obtain the full conditionals in closed form. The detailed distributions are presented in the appendix. For $m = 1, \ldots, 12$, $\eta_m^{(1)}$, $\eta_m^{(2)}$, $\gamma_m^{(1)}$ and $\gamma_m^{(2)}$ can be sampled from univariate normals. Denote $\boldsymbol{\theta}_{m,y} = (\theta_{m,y}(s_1), \ldots, \theta_{m,y}(s_{24}))'$, $\boldsymbol{\alpha} = (\alpha(s_1), \ldots, \alpha(s_{24}))'$, $\boldsymbol{\beta} = (\beta(s_1), \ldots, \beta(s_{24}))'$ and let $\mathbb{1}$ be a vector of ones. Then, according to Equation (1),

$$\boldsymbol{\theta}_{m,y} - \boldsymbol{\alpha}\eta_m^{(1)} - (\mathbb{1} - \boldsymbol{\alpha})\eta_m^{(2)} - \boldsymbol{\beta}\gamma_m^{(1)}t - (\mathbb{1} - \boldsymbol{\beta})\gamma_m^{(2)}t - \mathbb{1}\lambda_t \sim N_{24}(0, V), \quad \forall\, m, y. \tag{5}$$

Since the prior for V is an inverse Wishart, the full conditional will also be an inverse Wishart. Let $\boldsymbol{X}_{m,y}$ denote the vector of observations at time (m, y). We write $\boldsymbol{X}_{m,y} = (\boldsymbol{X}_{m,y}^O, \boldsymbol{X}_{m,y}^*)$, where $\boldsymbol{X}_{m,y}^O$ are the observed values and $\boldsymbol{X}_{m,y}^*$ the missing ones. As is customary in a Bayesian framework, we treat the missing values as unknown parameters and sample them within the MCMC, from the appropriate multivariate normal distribution.

Let $\boldsymbol{Y}_t = \boldsymbol{\theta}_{m,y} - \boldsymbol{\alpha}\eta_m^{(1)} - (\mathbb{1} - \boldsymbol{\alpha})\eta_m^{(2)} - \boldsymbol{\beta}\gamma_m^{(1)}t - (\mathbb{1} - \boldsymbol{\beta})\gamma_m^{(2)}t$, then, sampling λ_t can be done by noticing that, conditional on all the remaining parameters, $\boldsymbol{Y}_t = \mathbb{1}\lambda_t + \boldsymbol{\epsilon}_t$. From Equation (3) we have that $\lambda_t \sim N(\phi\delta_t, \tau_\lambda^2)$. Thus the full conditional for λ_t is given by a normal with mean $(\mathbb{1}'V^{-1}Y_t + \phi\delta_t/\tau_\lambda^2)/(\mathbb{1}'V^{-1}\mathbb{1} + 1/\tau_\lambda^2)$ and variance $(\mathbb{1}'V^{-1}\mathbb{1} + 1/\tau_\lambda^2)^{-1}$.

To sample from the joint distribution of $(\delta_1, \ldots, \delta_T, \kappa_1, \ldots, \kappa_T)$ we use the forward filtering, backward sampling algorithm (for example, see West and Harrison, 1997, Chp. 15) applied to the conditional multivariate dynamic linear model given by observation equation (3) and evolution equation (4).

We use the appropriate normal distribution to obtain samples of ϕ. Samples of $\tau_\chi^2, \tau_z^2, \tau_\delta^2, \tau_\kappa^2$ and τ_λ^2 correspond to inverse gamma distributions.

To obtain samples of $\alpha(s)$ and $\beta(s)$, for each s, we first considered an approach based on a Metropolis step. Unfortunately the resulting samples showed very slow mixing.

As an alternative, we follow the approach of Neal (2003) and use an overrelaxed slice sampler, switching to a regular slice sampler every ten iterations of the MCMC. Since the approach is the same for both sets of parameters, let us consider $\alpha(s)$ only and fix s. Also, let us denote $\alpha^n(s)$ as a newly sampled value of $\alpha(s)$and $\alpha^p(s)$ as a value sampled in the previous iteration of the MCMC. The following paragraph summarizes the rationale and algorithm of the slice sampling approach.

In both slice samplers, a slice of the distribution is defined, where the density is always greater than a threshold given by a random fraction (between 0 and 1) of the density at $\alpha^p(s)$. While in the regular sampler $\alpha^n(s)$ is sampled independently from $\alpha^p(s)$, in the overrelaxed sampler it is chosen to be on the opposite side of the mode, thereby avoiding random walks. An implicit assumption in the overrelaxed sampler is that the full conditional distribution is unimodal. We followed the scheme of Neal (2003) to define the initial slice, trim its edges and obtain $\alpha^n(s)$. The initial slice is set to have a width of 0.05 and is randomly placed in the interval $(0, 1)$, provided that it includes $\alpha^p(s)$. A random variate with standard uniform distribution is drawn and multiplied to the full conditional of $\alpha(s)$ to obtain the threshold. With an iterative procedure, the limits of the initial slice are extended or contracted so that the posterior at its edges remains bigger than the threshold. Once the lower (L) and the upper (U) edges of final slice are defined, the overrelaxed slice sampler chooses the new candidate according to $\alpha^n(s) = L + U - \alpha^p(s)$, while the regular slice sampler uses $\alpha^n(s) = L + z(U - L)$, where z is a random variate with standard uniform distribution. The candidate is accepted if its posterior exceeds the threshold. Otherwise, the overrelaxed slice sampler sets $\alpha^n(s) = \alpha^p(s)$, while the regular slice sampler redefines the slice using $\alpha^n(s)$ as one of the edges and samples a new candidate. As Neal (2003) points out, for unimodal distributions the candidate is almost never rejected, as long as the edges of the slice are accurately estimated. In our application we obtained rejection rates of about 0.1% for the various $\alpha(s)$ and 0.07% for the $\beta(s)$.

For convergence diagnostics, we use the methods developed by Heidelberger and Welch (1983); Gelman and Rubin (1992); Geweke (1992); Raftery and Lewis (1992b,a); Brooks

and Gelman (1998), which are available in the package Bayesian Output Analysis Program (`BOA`) (Smith, 2005) within `R` (R Development Core Team, 2005). We used the default values of `BOA` to define the length of the burn-in stage, thin the chain, check stationarity and define the adequate sample size to achieve the precision required, when sampling from the posterior distribution.

3.3 Model checking

To perform model checking we plot, for each station, the time-series of the observations together with the corresponding 95% posterior intervals, given by the model. In this way we assess if the model closely follows the observations, while providing narrow intervals. For stations with few missing values, we compare the ordinary least squares mean trend and standard error estimates with the corresponding values provided by the model, month by month. We perform this to see if our model is able to separate the long-term trends from the short-term variability. We perform a third informal analysis to check if λ_t successfully captures the short-term variability, or if some temporal structure is left in the residuals. This consists of randomly choosing 100 iterations from the stationary part of the Markov chain and, for each iteration and station, compute the temporal autocorrelations of the model's residuals. We plot each station's results and search for remaining temporal structure.

A more systematic analysis of the residuals is performed following the ideas in Kim et al. (1998). Let Θ denote the collection of all parameters. For each site s consider the random variable $X_t(s)$ that corresponds to the temperature at time t and location s. The one step ahead distribution of the temperature is

$$u_{t,s}(x) = P(X_t(s) \leq x | \Theta, x_i(s), i = 1, \ldots, t-1).$$

Following Rosenblatt (1952), $u_{1,s}(X_1(s)), \ldots, u_{T-1,s}(X_{T-1}(s))$ are independent and uniformly distributed provided the underlying distribution is continuous. Gneiting et al. (2005) denote this transformation as Probability Integral Transform and give an extensive list of references regarding its application.

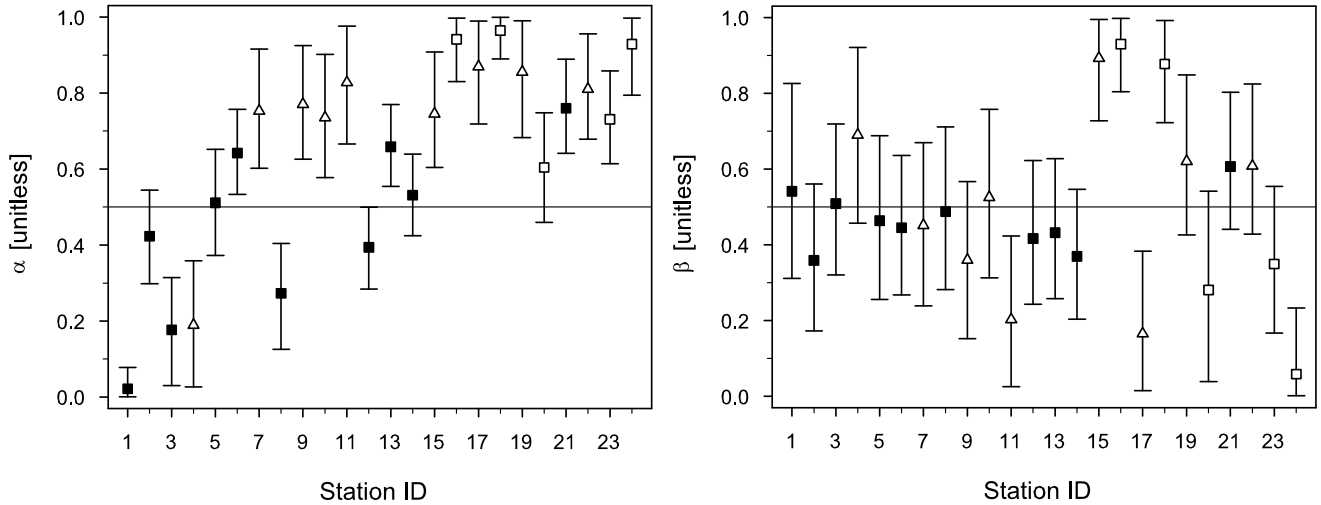

Figure 4: Posterior means and 95% probability intervals for the weights α (left panel) and β (right panel). Black squares correspond to tidal stations, white triangles to muted stations and white squares to non-tidal stations. The horizontal line provides the reference of even mixing of the coastal and inland components.

We notice that, conditional on Θ, the one step ahead predictive distribution of the temperature at any given site is normal. So, from each iteration of the MCMC after convergence, we can obtain a collection of random variables that should be independent and uniformly distributed. As in Kim et al. (1998) we consider a transformation to normality given by $\Phi^{-1}(u_t(s))$. We use quantile-quantile plots, correlograms and periodograms to check that these variables are, respectively, normally distributed and independent.

4 Results

We present results based on 30,000 iterations of the MCMC after a burn-in of 20,000 iterations. `BOA` convergence diagnostics and parameter trace plots can be examined from `http://www.ams.ucsc.edu/~bruno/`. In Figure 4 we present the 95% posterior intervals that correspond to the parameters $\alpha(s)$ and $\beta(s)$ for each one of the 24 stations. These parameters determine the proportion of each type (i.e. coastal or inland) of seasonality or

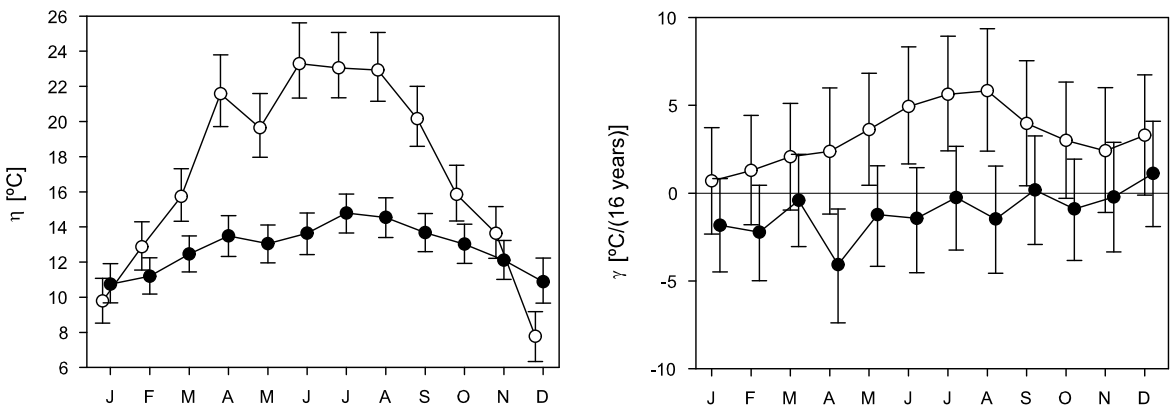

Figure 5: Posterior mean and 95% probability intervals for the coefficients that define the seasonality (left panel) and the long-term trend (right panel). The symbol $\circ$ corresponds to $\eta^{(1)}$ and $\gamma^{(1)}$; the symbol $\bullet$ corresponds to $\eta^{(2)}$ and $\gamma^{(2)}$.

seasonal trend that corresponds to a station. Since the stations are ordered with respect to their distance to the mouth, we expect some association between the x-axis in the plots and the weights. In fact $\alpha(s)$ generally increases with increasing distance inland. This is particularly evident for the first seven stations. However, there is a great deal of variability due to local conditions. The behavior of the weights for the monthly seasonal trends is more irregular. Tidal stations and stations close to the coast (rank ≤ 14) with restricted flow due to structures in the channel, like tidal gates, culverts, etc. (muted), admit an even mixing of the two seasonal trend components. In contrast, most northern stations display average weights above 0.5. Station 24 is a clear exception.

Figure 5 displays the 95% posterior intervals for the parameters that correspond to the two types of seasonality and trends. We observe both types of seasonality present a dip in May. This is consistent with the behavior observed in the data, where a strong decrease in temperature in May is present for most stations. $\eta^{(1)}$ is typical of inland stations, with higher temperatures, especially during the summer. For most months, the main effects of the two types of seasonality have little overlap. The long-term trends show substantial overlaps during the Fall and Winter months. Of particular interest is

the fact that the components of $\gamma^{(1)}$ are mostly positive, while those of $\gamma^{(2)}$ are mostly negative. The monthly differences indicate that, where present, the warming trend is not consistent through the year. In general we observe that the trend is stronger for summer months than for the rest of the year and that there may be a cooling trend in April. Recall that the trend for a station indexed as s, corresponding to month m is given by

$$\beta(s)\gamma_m^{(1)} + (1 - \beta(s))\gamma_m^{(2)}.$$

By using samples of $\beta(s)$, $\gamma_m^{(1)}$ and $\gamma_m^{(2)}$ for all months and stations we obtained a detailed illustration of the yearly dynamics of the trends and their differences from site to site. This is presented in Figures 6 and 7. We notice variations through the year for all of the stations. For the majority of the stations there is some evidence of warming, especially during the summer months. April is peculiar, since for some stations there is some evidence of cooling trend during that month. The stations for which this effect is strongest are 2, 7, 9, 11, 14, 17 and 24. A strong warming trend is present in December for almost all stations. In some cases the warming can be as high as high as 5 °C/(16 years) in median, but its value is highly variable.

An example of how the seasonal pattern changes with time is presented in Figure 8. We observe substantial changes along the year, but the strongest differences are present during the summer months when the mean of the seasonal component has had an increase of about 4°C.

Figure 9 shows the estimation of the common baseline for all stations, λ_t, the baseline for ET_0, δ_t, and the parameter that controls the seasonality of ET_0, κ_t. Despite the large monthly fluctuations, the link between λ_t and δ_t is clear. This is strengthened by the fact that a posteriori, the parameter ϕ, which regresses λ_t onto δ_t, is positive almost with probability one. Its 95% posterior probability interval is $(0.588, 1.293)$; the mean is 0.975. Another noticeable feature in Figure 9 is that most of the temporal variability in ET_0 is explained by the baseline, whilst the seasonality remains nearly identical. This can also be deducted from Table 1, where the posterior mean of the evolution variance for δ_t is 30 times greater than the evolution variance for κ_t.

As discussed in Section 2, the dependence among stations has a rather complicated

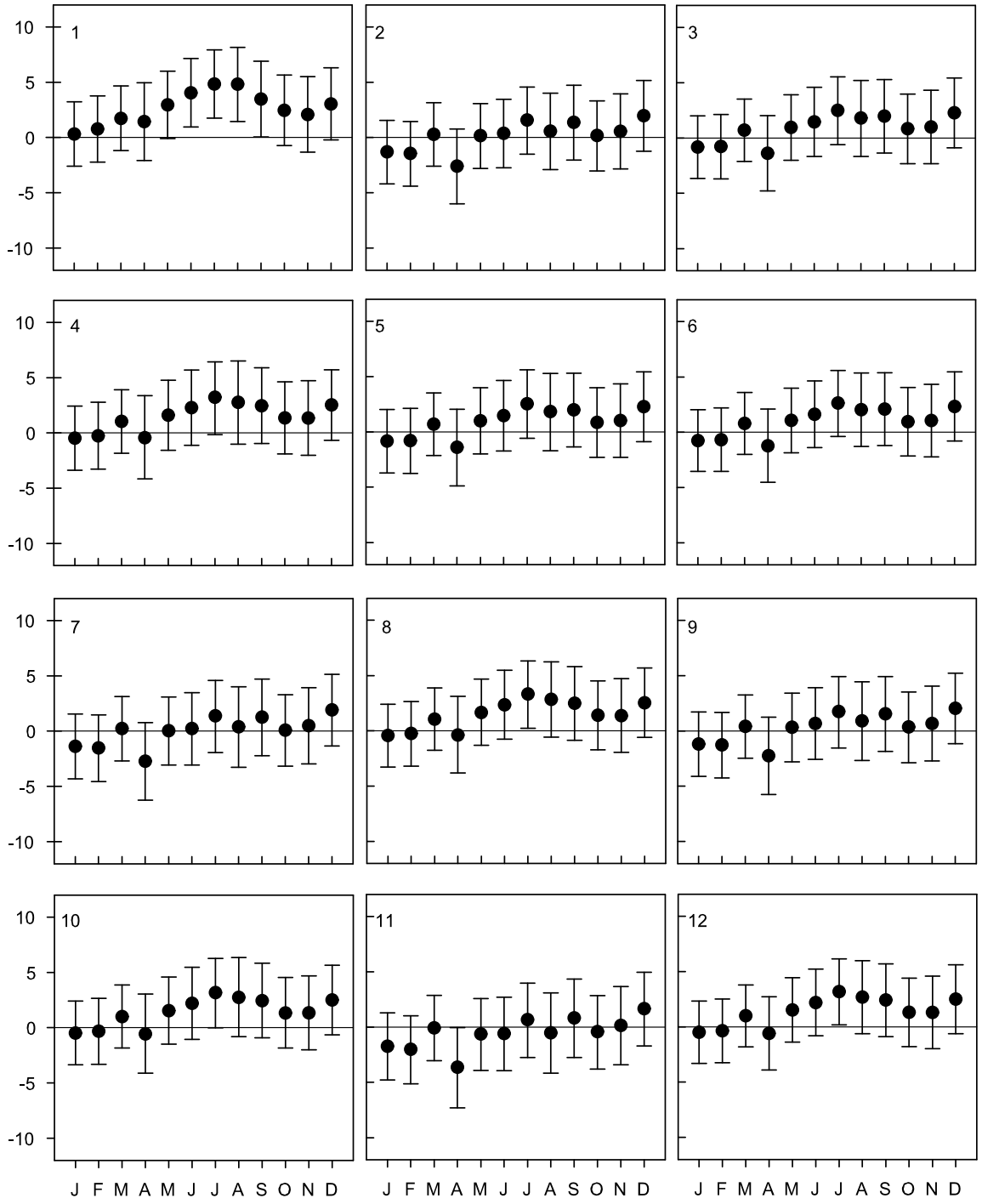

Figure 6: Posterior mean and 95% probability intervals for the temperature trends estimated for Stations 1 to 12. The y-axis scale corresponds to °C/(16 years), so that it represents the difference in temperature between 2004 and 1988.

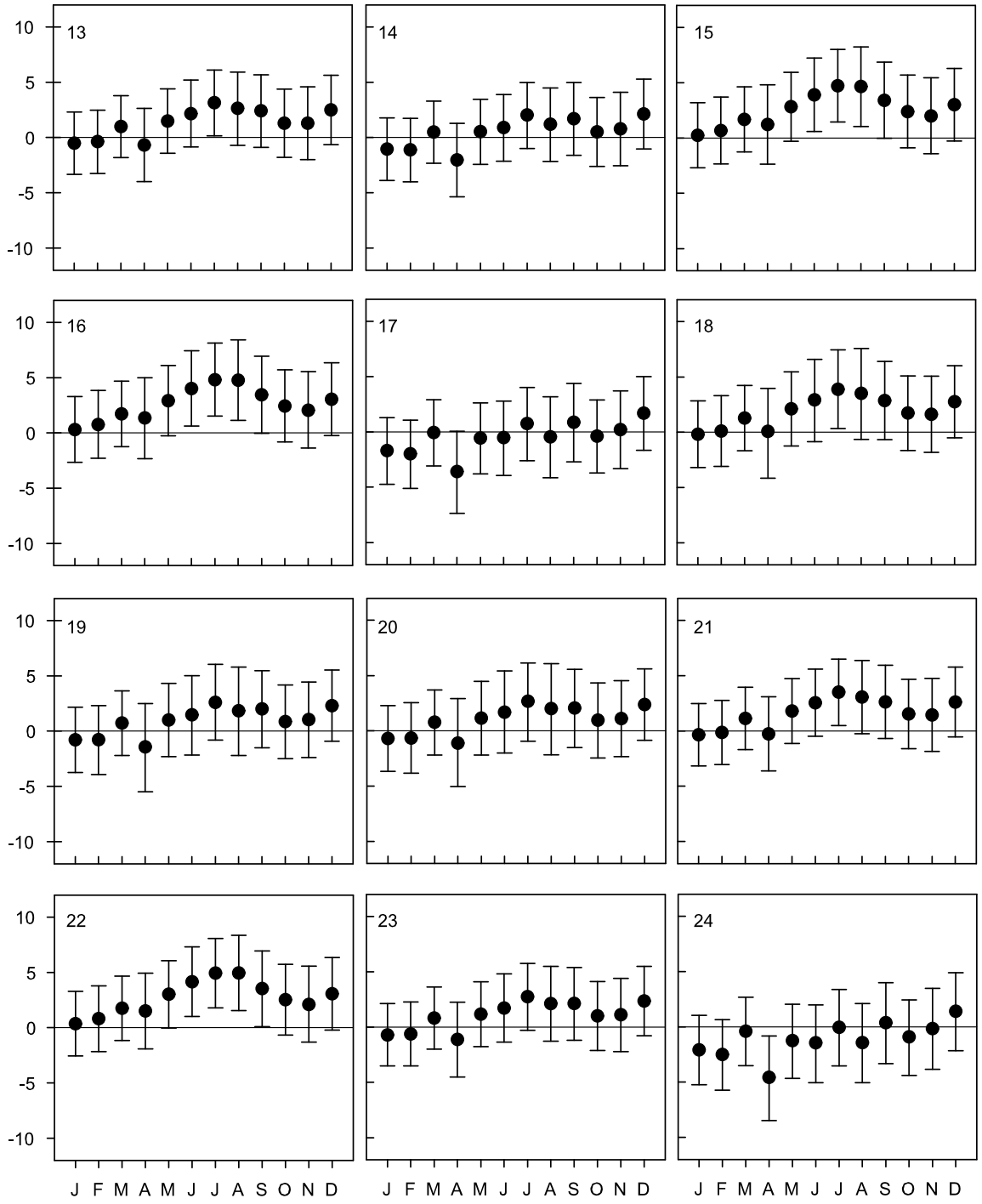

Figure 7: Posterior mean and 95% probability intervals for the temperature trends estimated for Stations 13 to 24. The y-axis scale corresponds to °C/(16 years), so that it represents the difference in temperature between 2004 and 1988.

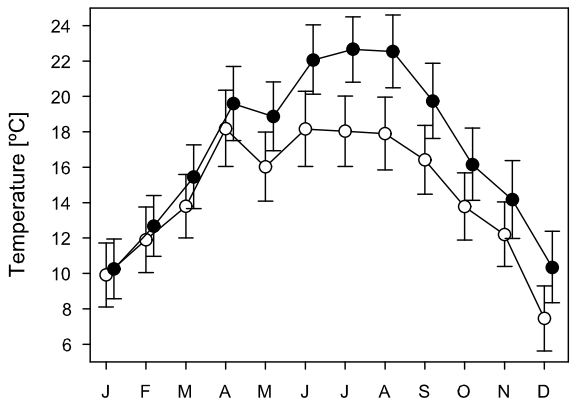

Figure 8: Posterior mean and 95% probability intervals for the seasonal cycle in the first year (○) and the last year (●) of the time-series for Station 22.

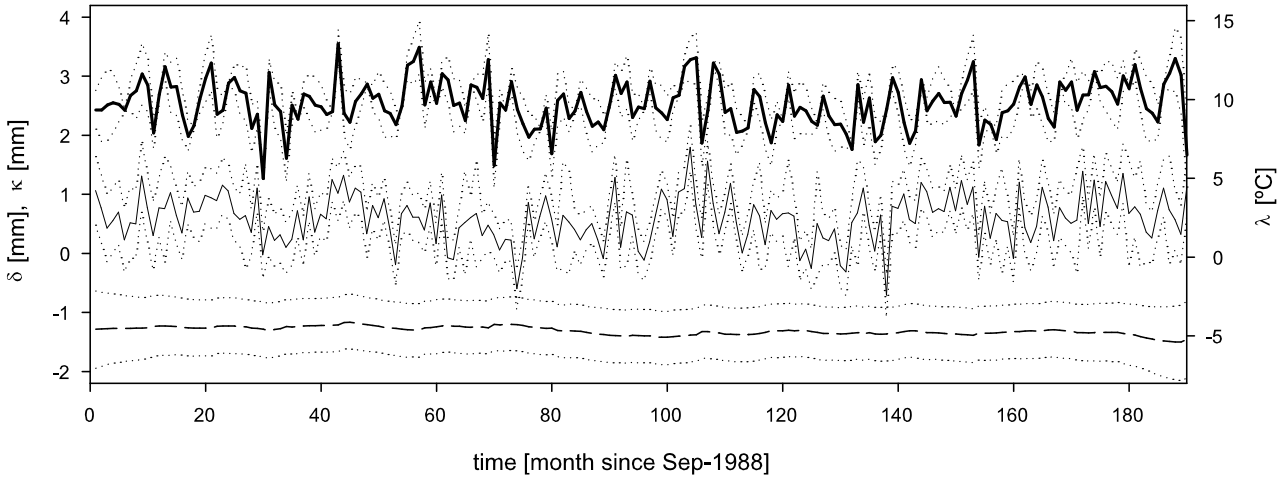

Figure 9: Posterior mean for δ_t (left vertical axis, thick line), κ_t (left vertical axis, stippled line) and λ_t (right vertical axis, thin line), and corresponding 95% intervals (dotted lines), at different time steps.

Table 1: Posterior means and 95% intervals for the variance parameters.

	τ_δ^2	τ_κ^2	τ_λ^2	τ_χ^2	τ_z^2
Upper Endpoint	0.290	0.0188	3.42	0.00362	0.000114
Mean	0.236	0.0075	2.66	0.00340	0.000100
Lower Endpoint	0.192	0.0029	2.05	0.00320	0.000089

structure. This is illustrated by the correlations estimated from the posterior mean of V in the right panel of Figure 10. We notice that Figure 10 shows two distinctive clusters. A block of gray areas corresponding to Stations 1 through 11 and another one corresponding to Stations 12 to 24, with the exception of Stations 17, 20 and 23. Stations 14, 15, 16, 18, 19, 21, 22 and 23 share agricultural land use influence, are all fairly distant from the mouth and are located in the northern part of the slough. We observe that Stations 17, 20, and 23 have peculiar behavior with respect to its neighbors. They are strongly influenced by Salinas River discharge which can be relatively high during the rainy season and low during the dry season. Additionally, Station 20 is a lagoonal system, periodically open to the ocean but dominated by upstream riverine processes from the Salinas River. Figure 10 demonstrates the relative strengths of the correlations but does not show the differences in variances between stations. These are presented in the left panel of Figure 10. We observe substantial differences even between stations that are close together. Posterior inference for the other variance components in the model are presented in Table 1.

We use the method proposed in Section 3.3 to check the validity of the model. We found that all of the 24 quantile-quantile plots conformed well to a normal distribution. The autocorrelation functions display weak residual correlations, the largest value among all stations being close to .20. The estimated periodograms present irregular patterns with random fluctuations consistent with white noise.

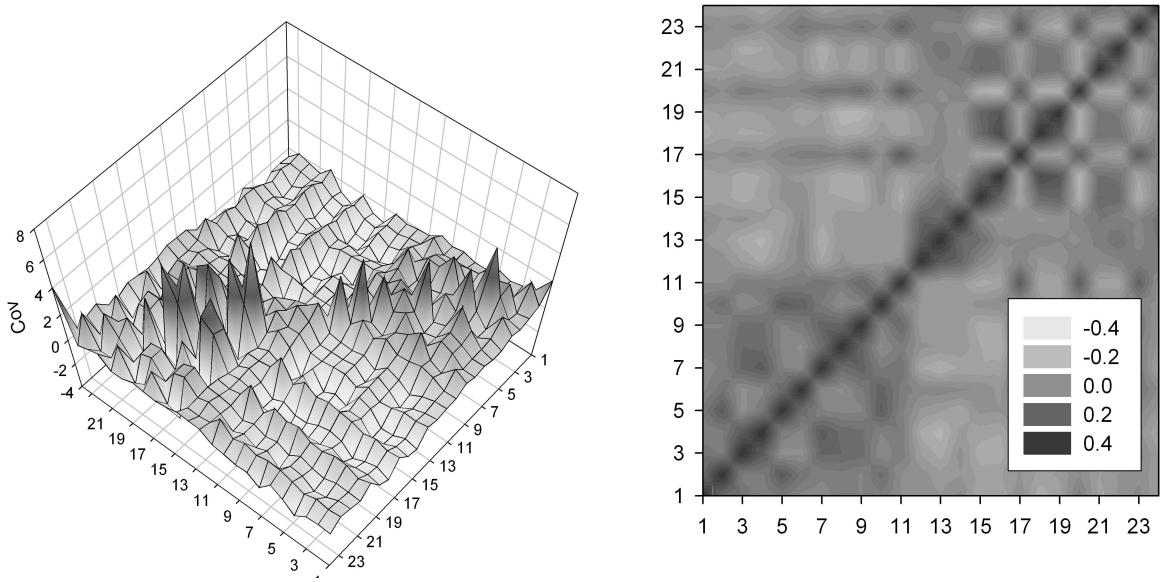

Figure 10: Covariance (left) and correlation (right) structure of the error.

5 Conclusions

We have presented a detailed analysis of the spatial and temporal variations of temperature in an estuary in central California, based on observations collected over the last 16 years. Based on our exploratory analysis, we designed our model based on the assumption that the within-year dynamics needed to be described as a mixture of two form-free components. The monthly variation of temperature is clear in Figure 5: only three months separate the coldest and the warmest months of the year, i.e. December and April respectively, and periods of strong temperature fluctuations contrast with others of relative constancy or steady variation. This behavior can also be captured using sinusoidal components. But the lack of symmetry would require many terms, involving nearly as many parameters as the form-free representation. Given the results, we can broadly identify the two form-free components as describing the coastal and the inland variation of temperature. The coastal component, when compared to its counterpart, has smaller annual amplitude and fluctuations. This is likely a reflex of the buffering action of the ocean, since in this region the annual SST amplitude is smaller than 5°C.

The maximum and minimum annual slough temperatures are not reflected in either SST or air temperature, and thus they may reflect regional extremes in the balance between radiation absorption and emission. Apart from tidal influence, site-specific effects seem to come into play, since nearby stations can display quite distinct behaviors. As mentioned in Section 2, this impairs the interpolation of the results to other locations in the estuary. On the other hand, it highlights the importance of maintaining a network of stations covering the Elkhorn Slough, for the longest time span possible.

Recent examples of models that consider time varying parameters appear in Shaddick and Wakefield (2002), Huerta et al. (2004) and Lemos and Sansó (2006). The approach taken in Lemos and Sansó (2006) considers temperature trends that vary smoothly in space. The model in Huerta et al. (2004) focuses in the spatial variation of the amplitudes of the series. Those approaches are not appropriate for the problem considered in this paper, since the complex hydrology of the slough produces very localized effects. In fact, we observed locations that are spatially adjacent with different warming trends. Also, some stations that are very far from the mouth of the estuary exhibit a seasonal pattern similar to the one observed along the coast, due to intense tidal flushing.

The above conclusion is also reached from the analysis of temperature trends. Here, the two form-free components place northern and southern inland stations on opposite poles, while stations close to the mouth have an even mix of the two behaviors. Southern stations (11, 17, 20 and 23) display a marked cooling in April and warming in December, which both reduce the annual amplitude over time. This may reflect an increasing tidal flushing over time, as a consequence of channel erosion (Van Dyke and Wasson, 2005). In contrast, northern stations reveal strong warming trends for several months, with summer months and December being the most noteworthy. This warming in the northern part of the Elkhorn Slough may have important biological implications, namely changes in species composition and rates of biochemical processes. The cause for this temperature change is unknown; it may be due to a natural or anthropogenic long-term change in the hydrology of the slough, or to a combination of both.

The relevance of natural forcing on temperature in the slough is confirmed by our

model, when it includes ET_0 as a regressor (Figure 9). ET_0 affects all stations' temperature identically and describes a large fraction of the short-term variability. Thus, it shows that location-independent phenomena contribute to temperature fluctuations as well. Another component that may contain some natural variability is the error, since its covariance structure presents interesting patterns that connect nearby stations. In all, our approach demonstrates how temperature trends can be determined in a hydrologically complex estuary.

Acknowledgments

Funding for this research was provided in part by the Community Foundation of Monterey County and the Elkhorn Slough Foundation, through a donor-advised fund with the Central Coast Regional Water Quality Control Board. R. Lemos acknowledges grant SFRH/BD/17929/2004 from Fundação para a Ciência e a Tecnologia. The authors are grateful to three referees and an Associate Editor for their useful comments.

A Full conditional distributions

The dots are shorthand for the data and all the remaining parameters. μ and σ^2 respectively denote the prior mean and variance of the parameter considered.

- $\left(\eta_m^{(1)}|...\right) \sim N\left(\frac{\boldsymbol{\alpha}'V^{-1}\boldsymbol{\theta}_{m,y}^{\dagger}+\mu/\sigma^2}{\boldsymbol{\alpha}'V^{-1}\boldsymbol{\alpha}+1/\sigma^2}, \frac{1}{\boldsymbol{\alpha}'V^{-1}\boldsymbol{\alpha}+1/\sigma^2}\right)$, where $\boldsymbol{\theta}_{m,y}^{\dagger} = \boldsymbol{\theta}_{m,y} - (\mathbb{1}-\boldsymbol{\alpha})\eta_m^{(2)} - \boldsymbol{\beta}\gamma_m^{(1)}t - (\mathbb{1}-\boldsymbol{\beta})\gamma_m^{(2)}t - \mathbb{1}\lambda_t$
- $\left(\eta_m^{(2)}|...\right) \sim N\left(\frac{(\mathbb{1}-\boldsymbol{\alpha})'V^{-1}\boldsymbol{\theta}_{m,y}^{\dagger}+\mu/\sigma^2}{(\mathbb{1}-\boldsymbol{\alpha})'V^{-1}(\mathbb{1}-\boldsymbol{\alpha})+1/\sigma^2}, \frac{1}{(\mathbb{1}-\boldsymbol{\alpha})'V^{-1}(\mathbb{1}-\boldsymbol{\alpha})+1/\sigma^2}\right)$, where $\boldsymbol{\theta}_{m,y}^{\dagger} = \boldsymbol{\theta}_{m,y} - \boldsymbol{\alpha}\eta_m^{(1)} - \boldsymbol{\beta}\gamma_m^{(1)}t - (\mathbb{1}-\boldsymbol{\beta})\gamma_m^{(2)}t - \mathbb{1}\lambda_t$
- $\left(\gamma_m^{(1)}|...\right) \sim N\left(\frac{(\boldsymbol{\beta}t)'V^{-1}\boldsymbol{\theta}_{m,y}^{\dagger}+\mu/\sigma^2}{(\boldsymbol{\beta}t)'V^{-1}(\boldsymbol{\beta}t)+1/\sigma^2}, \frac{1}{(\boldsymbol{\beta}t)'V^{-1}(\boldsymbol{\beta}t)+1/\sigma^2}\right)$, where $\boldsymbol{\theta}_{m,y}^{\dagger} = \boldsymbol{\theta}_{m,y} - \boldsymbol{\alpha}\eta^{(1)} - (\mathbb{1}-\boldsymbol{\alpha})\eta_m^{(2)} - (\mathbb{1}-\boldsymbol{\beta})\gamma_m^{(2)}t - \mathbb{1}\lambda_t$
- $\left(\gamma_m^{(2)}|...\right) \sim N\left(\frac{(\mathbb{1}-\boldsymbol{\beta}t)'V^{-1}\boldsymbol{\theta}_{m,y}^{\dagger}+\mu/\sigma^2}{(\mathbb{1}-\boldsymbol{\beta}t)'V^{-1}(\mathbb{1}-\boldsymbol{\beta}t)+1/\sigma^2}, \frac{1}{(\mathbb{1}-\boldsymbol{\beta}t)'V^{-1}(\mathbb{1}-\boldsymbol{\beta}t)+1/\sigma^2}\right)$, where $\boldsymbol{\theta}_{m,y}^{\dagger} = \boldsymbol{\theta}_{m,y} - \boldsymbol{\alpha}\eta^{(1)} - (\mathbb{1}-\boldsymbol{\alpha})\eta_m^{(2)} - \boldsymbol{\beta}\gamma_m^{(1)}t - \mathbb{1}\lambda_t$

- $(\phi|\ldots) \sim N\left(\frac{\sum \delta_t \lambda_t/\tau_\lambda^2+\mu/\sigma^2}{\sum \delta_t^2/\tau_\lambda^2+1/\sigma^2}, \frac{1}{\sum \delta_t^2/\tau_\lambda^2+1/\sigma^2}\right)$
- $(\tau_\chi^2|\ldots) \sim IG(3280, (6.8+\sum_{y,m}(\boldsymbol{X}_{m,y}-\boldsymbol{\theta}_{m,y})'(\boldsymbol{X}_{m,y}-\boldsymbol{\theta}_{m,y}))/2)$
- $(\tau_z^2|\ldots) \sim IG(345, (0.05+\sum_{t=1}^{190}(z_t-\delta_t-\kappa_t\cos(2\pi t/12))^2)/2)$
- $(\tau_\delta^2|\ldots) \sim IG(97, (0.05+\sum_{t=2}^{190}(\delta_t-\delta_{t-1})^2)/2)$
- $(\tau_\kappa^2|\ldots) \sim IG(97, (0.05+\sum_{t=2}^{190}(\kappa_t-\kappa_{t-1})^2)/2)$
- $(\tau_\lambda^2|...) \sim IG(97.5, (0.05+\sum_{t=1}^{190}(\lambda_t-\phi\delta_t)^2)/2)$
- $(\boldsymbol{\theta}_{m,y}|...) \sim N\left(\boldsymbol{X}_{m,y}-\tau_\chi^2(V+\tau_\chi^2 I)^{-1}(\boldsymbol{X}_{m,y}-\boldsymbol{A}_{m,y}), \tau_\chi^2(I-\tau_\chi^2(V+\tau_\chi^2 I)^{-1})\right)$, where $\boldsymbol{A}_{m,y}=\boldsymbol{\alpha}\eta_m^{(1)}+(\mathbb{1}-\boldsymbol{\alpha})\eta_m^{(2)}+\boldsymbol{\beta}\gamma_m^{(1)}t+(\mathbb{1}-\boldsymbol{\beta})\gamma_m^{(2)}t+\mathbb{1}\lambda_t$.
- $\left(\boldsymbol{X}_{m,y}^*|...\right) \sim N(\boldsymbol{\theta}_{m,y}^*, \tau_\chi^2 I)$.

References

Brooks, S. and Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**, 434–55.

Edgar, G. J., Barrett, N. S., Graddon, D. J. and Last, P. R. (2000) The conservation significance of estuaries: a classification of Tasmanian estuaries using ecological, physical and demographic attributes as a case study. *Biological Conservation*, **92**, 383–397.

Gamerman, D. and Lopes, H. F. (2006) *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. London, UK: Chapman and Hall, second edn.

Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–72.

Geweke, J. (1992) Evaluating the accuracy of sampling-based i approaches to calculating posterior moments. In *Bayesian Statistics 7* (eds. J. M. Bernardo, J. O. Berger, P. Dawid, A. F. M. Smith and M. West). Clarendon Press, Oxford, UK.

Gneiting, T., Balabdaoui, F. and Raftery, A. (2005) Probabilistic forecasts, calibration and sharpness. *Tech. Rep. 483*, Department of Statistics, University of Washington.

Heidelberger, P. and Welch, P. (1983) Simulation run length control in the presence of an initial transient. *Operations Research*, **31**, 1109–44.

Helsel, D. R. and Hirsch, R. M. (1992) *Statistical Methods in Water Resources.* New York: Elsevier.

Huerta, G., Sansó, B. and Stroud, J. R. (2004) A spatio-temporal model for Mexico city ozone levels. *Applied Statistics*, **53**, 231–248.

Kim, S., Shephard, N. and Chib, S. (1998) Stochastic volatility: likelihood inference and comparison with arch models. *Rev. Fin. Stud.*, **65**, 361–393.

Laprise, R. and Dodson, J. J. (1994) Environmental variability as a factor controlling spatial patterns in distribution and species-diversity of zooplankton in the St-Lawrence-Estuary. *Marine Ecology-Progress Series*, **107**, 67–81.

Lemos, R. and Sansó, B. (2006) Spatio-temporal variability of ocean temperature in the portugal current system. *Journal of Geophysical Research Oceans*, **111**.

Neal, R. (2003) Slice sampling. *Annals of Statistics*, **31**, 705–767.

Price, A. (2002) Simultaneous 'hotspots' and 'coldspots' of marine biodiversity and implications for global conservation. *Marine Ecology-Progress Series*, **241**, 23–27.

R Development Core Team (2005) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL: `http://www.R-project.org`. ISBN 3-900051-07-0.

Raftery, A. E. and Lewis, S. M. (1992a) Comment: One long run with diagnostics: Implementation strategies for markov chain monte carlo. *Statistical Science*, **7**, 493–7.

— (1992b) How many iterations in the Gibbs sampler? In *Bayesian Statistics 4* (eds. J. M. Bernardo, J. O. Berger, P. Dawid and A. F. M. Smith), 765–776. Oxford University Press.

Rosenblatt, M. (1952) Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, **23**, 470–472.

Shaddick, G. and Wakefield, J. (2002) Modelling daily multivariate pollutant data at multiple sites. *Applied Statistics*, **51**, 351–372.

Smith, B. J. (2005) *boa: Bayesian Output Analysis Program (BOA) for MCMC*. URL: `http://www.public-health.uiowa.edu/boa`. R package version 1.1.5-2.

Van Dyke, E. and Wasson, K. (2005) Historical ecology of a central california estuary: 150 years of habitat change. *Estuaries*, **22**, 173–189.

West, M. and Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models*. New York: Springer Verlag, second edn.

Whitfield, A. K. (1994) Fish species-diversity in Southern African estuarine systems - an evolutionary perspective. *Environmental Biology of Fishes*, **40**, 37–48.