Note

Modeling and assessing climatic trends

The authors

Peter F. Craigmile is Professor of Statistics in the Department of Statistics at The Ohio State University in Colombus, Ohio, USA. Peter Guttorp is Professor Emeritus of Statistics in the Department of Statistics at University of Washington in Seattle, Washington, USA and Professor II at the Norwegian Computing Center in Oslo, Norway.

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in information and communication technology and applied statistical-mathematical modelling. The clients include a broad range of industrial, commercial and public service organisations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have in us is given by the fact that most of our new contracts are signed with previous customers.

Title	**Modeling and assessing climatic trends**
Authors	**Peter F. Craigmile** `<pfc@stat.osu.edu>`
	Peter Guttorp `<peter.guttorp@nr.no>`
Date	6th February 2017
Publication number	SAMBA/06/17

Abstract

Climate studies often fit linear trends to data. In many cases simplifying assumptions such as independent errors and constant variance are used. We review a variety of approaches to estimating linear trends, and illustrate with US temperature data how oversimplified assumptions may lead to false significance. We outline a variety of methods to fit nonlinear trend models. Using the Berkeley Earth global data set we show that a bent cable fit is better than a linear fit for this series. We also review spatial and spatiotemporal trend models in mean, variance and extremes, as well as models with long term memory structure.

Introduction

Studies of climate change often involve the statistical analysis of climate variables (or their meteorological counterparts) indexed in space and time (von Storch and Zwiers, 2002). Usually these variables are observed directly from instrumental measurements (e.g., from thermometers for temperature; from rain gauges for precipitation), but it is becoming more common to observe data indirectly from satellite measurements or as the output of global or regional climate models. As the datasets become more complex, the statistical analyses can become more involved.

A common statistical problem is to model and assess trend in climate data (e.g., Cohn and Lins (2005); Craigmile et al. (2004); Shen et al. (2012); Smith (1993); Weatherhead et al. (1998)). Trends are often described by smooth changes in the certain features of a stochastic process over longer scales. Traditionally trend refers to smooth changes in the mean of a process over time but, as we will demonstrate, this definition can be extended to allow us to consider the smooth changes of other characteristics of a stochastic process, or can be extended to other dimensions of change, such as over space.

In the latest IPCC report (Hartmann et al., 2013) the issue of trends is explained in Box 2.2. While they discuss both linear and nonlinear trend models, they chose to emphasize linear trends, arguing that "The linear trend fit is used in this chapter because it can be applied consistently to all the data sets, is relatively simple, transparent and easily comprehended, and is frequently used in the published research assessed here." (p.180)

In this chapter we mainly focus on the estimation and assessment of trend in the presence of dependence in time. Statistical methods that assume that the errors (after removal of the trend component) are independent and identically distributed are usually a poor assumption for climate data. Indeed assuming that the errors can be approximated by a simple time series model such as the autoregressive process of order one, AR(1), is also unlikely to represent the residual variation in the climate process. The choice of the temporal (and spatial) scale for the trend is difficult and not well defined. We present methods to estimate both linear and nonlinear trends, with associated uncertainty quantification. Trend estimation can be different from changepoint analysis (the estimation of breaks in a time series). (For a discussion of changepoints and trends see Gallagher et al. (2013).) We discuss the estimation of trend in non-traditional settings, such as in the analysis of climate extremes, and point to future directions in the statistical assessment of climatic trends.
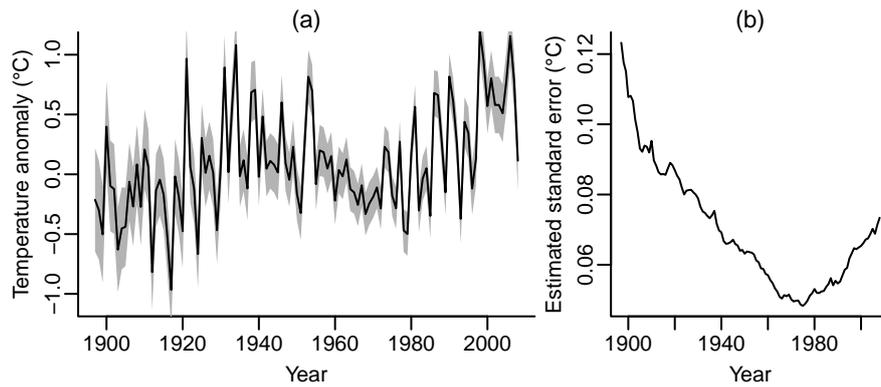
Figure 1. (a) US annual mean temperature anomalies with respect to the 1961–1990 climatology. The gray shaded region denotes simultaneous 95% confidence intervals for the mean, calculated using estimated standard errors from Shen et al. (2012); (b) A plot of the estimated standard errors for the mean by year.

Two motivating examples

We motivate the estimation of trend in climate time series using two datasets of average annual temperature. Each series has a well-defined estimate of uncertainty. The first series we consider is the average temperature over the contiguous United States (US), while the other is a global temperature series.

US average temperature anomaly

Shen et al. (2012) produced statistical estimates of US temperature anomalies from 1897–2008, using the US Historical Climatology Network data set version 2 (Menne et al., 2009), corrected for the fact that the time of day that measurements are made at can differ by site. Since US land temperature is a statistical estimate and not a direct measurement, it has a quantifiable standard error, with components coming from measurement error at individual stations, spatial dependence, and orographic effects, etc. Shen et al. (2012) was mainly concerned with the statistical estimation of this standard error. Figure 1(a) shows a time series plot of the anomalies with respect to the 1961–1990 climatology (i.e., residuals from the average temperature over this time period). The shaded gray regions denotes simultaneous 95% confidence intervals for this mean, calculated using the estimated standard errors with a Bonferroni correction (Abdi, 2007). Figure 1(b) shows that the standard errors decrease until 1975, and then increase again to the end of the record. The reason for this increase is the removal of stations mainly due to reduced funding.

Global temperature series

The Berkeley Earth project (Rohde et al., 2013) uses isotropic geostatistical tools (kriging) to estimate the global mean temperature. The land data have been collected from 14 databases and
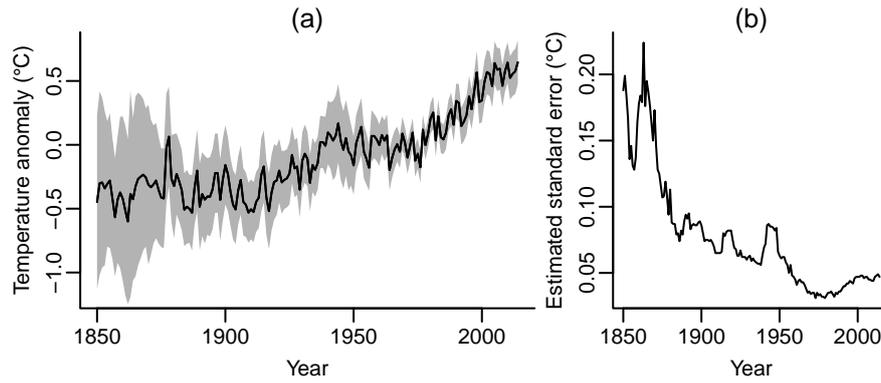
Figure 2. (a) Global annual mean temperature anomalies with respect to the 1961–1980 climatology. The gray shaded region denotes simultaneous 95% confidence intervals for the mean, using estimated standard errors from Rohde et al. (2013); (b) A plot of the estimated standard errors for the mean by year.

almost 45,000 stations. One of the main differences between the Berkeley Earth approach and most other global approaches is that the former group does not attempt to "homogenize" stations (Trewin, 2010). If a measurement device is moved or replaced, it is considered a different station, rather than being "corrected." The ocean data used in the Berkeley Earth series (Rohde and Muller, 2015) come from the Hadley Center sea surface temperature data set HadSST3 (Kennedy et al., 2011a,b), modified by kriging of missing grid squares. This allows for a statistically justifiable estimate of global mean estimation uncertainty. Figure 2, formatted identically as Figure 1, demonstrates that the standard errors tend to be higher in the past, although this over-simplifies the temporal patterns in the standard errors.

That temperatures tend to increase with the years is obvious in the global temperature series, but possible trend effects are more nuanced for the US average temperatures. Both series exhibit dependence over time that need to be accounted for before we can assess the significance of possible trends. We also need to account for the uncertainty in each mean series, as measured by the time-varying standard errors.

Time series approaches

In this chapter we mainly focus on methods of estimation and assessment of trend for time series processes observed discretely in time. Let $T \subset \mathbb{Z}$ denote a set of possible time points. Suppose we observe a climate series $\{y_t : t \in T\}$ regularly sampled in time. Let $\{Y_t : t \in T\}$ denote the associated discrete-time time series process, the stochastic process that generated $\{y_t\}$. An *additive decomposition* for trend then assumes that

$$Y_t = \mu_t + \eta_t, \quad t \in T, \tag{1}$$

where we refer $\{\mu_t : t \in T\}$ to be the *trend component* and $\{\eta_t : t \in T\}$ to be the *(irregular) noise component*, that captures everything in Y_t that is not captured by the trend. We could also extend the definition to include a seasonal component, $\{s_t : t \in T\}$ that repeats over time. (For a further discussion of statistical methods for estimating seasonal components in time series, see, e.g., Bloomfield (2000); Brockwell and Davis (2002); Percival and Walden (1993).)

A multiplicative decomposition is often used to analyze data that are positively skewed or exhibit a mean-variance relationship. The *multiplicative decomposition* for trend posits that

$$Y_t = \mu_t \times \eta_t, \quad t \in T.$$

When μ_t and η_t are positively valued for all t then we take logarithms to obtain the following the additive decomposition:

$$\log Y_t = \log \mu_t + \log \eta_t, \quad t \in T.$$

While the log transformation may not be variance-stabilizing for the data in hand, the additivity of the transformed response is an attractive feature that can simplify modeling (e.g., Section 2.1 of McCullagh and Nelder (1999)).

Clearly an additive decomposition (1) for a time series is not unique, as it may not be obvious for a given application what constitutes the trend $\{\mu_t\}$ and what is the noise $\{\eta_t\}$. While Kendall (1973) encapsulates that "the essential idea of trend is that it shall be smooth", he does not indicate "how smooth". An estimate of trend is defined purely in terms of the statistical model or method that we use to estimate it. As with any additive statistical estimation procedure, anything that is not captured by the estimated trend, $\widehat{\mu}_t$ say, will be appear in the estimate of the errors $\widehat{\eta}_t = y_t - \widehat{\mu}_t$. This has led to defining the trend in terms of how smooth the process is. For example: (i) letting $\mu_t = f(t)$ for some deterministic function f of time t, we assume a certain number of derivatives for f or by defining the trend in terms of linear combinations of known smooth functions (e.g., Wahba (1990); Wood (2006)); (ii) using certain functions such as wavelets we can define the trend in terms of averages over certain temporal scales (e.g., Brillinger (1994a, 1996); Craigmile et al. (2004)).

Candidate models for the noise

Different statistical models for the noise will influence our ability to estimate climatic trends. Thus, before we outline statistical methods for estimating and assessing trend, we discuss commonly used classes of time series models for the noise. A typical assumption is that the noise process $\{\eta_t\}$ is a mean zero stationary process. For a mean zero time series process, stationarity requires that the covariance of η_t and η_{t+h} depends only on the time lag h, and not on the time index t. This assumption simplifies estimation of the parameters driving the noise, and can be a reasonable assumption if the trend is able to capture the time-varying features of climate.

As we argued at the start of the chapter, it is typically unreasonable to assume that the noise process is uncorrelated in time (also known as a *white noise process*) or independent and identic-

ally distributed in time (also known as an *IID process*). In climate studies an *autoregressive process of order one, AR(1)*, also called a red-noise process, is commonly used to model the noise $\{\eta_t\}$. This model is defined by letting

$$\eta_t = \phi\eta_{t-1} + Z_t, \quad t \in T,$$

where $\{Z_t\}$ is a white noise or IID process. The noise process is stationary (and causal) when the autoregressive parameter ϕ satisfies $|\phi| < 1$. More general dependence structures can be obtained by increasing the order of autoregression to yield an autoregressive process of order p, AR(p):

$$\eta_t = \sum_{j=1}^{p} \phi_j\eta_{t-j} + Z_t, \quad t \in T.$$

We can also filter the $\{Z_t\}$ process to obtain the ARMA(p, q) process, the autoregressive moving average process of orders p and q:

$$\eta_t = \sum_{j=1}^{p} \phi_j\eta_{t-j} + Z_t + \sum_{k=1}^{q} \theta_k Z_{t-k}, \quad t \in T.$$

In practice since stationary ARMA model can be written as an infinite AR model (that in practice is truncated to finite order), some statisticians working in climate prefer to use lower order AR processes rather than ARMA processes (the parameters of AR processes can be easier to estimate than those of ARMA processes, for example).

In more recent years, long memory processes have been used as a model for noise in studies of climate (e.g., Barboza et al. (2014); Cohn and Lins (2005); Craigmile et al. (2004); Hussain and Elbergali (1999); Koutsoyiannis (2003)). The choice of this class of processes could be due to the fact these models capture the self-similar behavior of climate processes over long time scales (e.g., Beran (1994)), but also because stationary long memory processes exhibit slowly decaying autocorrelations. These slowly decaying correlations give the appearance of local deviations that we commonly see in climate series, but lead to greater uncertainty in the trend estimates.

For our two motivating datasets, we need to extend the class of models for the noise to include modulated stationary noise, with time dependent variance. A modulated noise process $\{\eta_t\}$ is defined by

$$\eta_t = \sigma_t\epsilon_t, \quad t \in T,$$

where $\{\sigma_t\}$ are the time-varying standard deviations and $\{\epsilon_t\}$ is a stationary process.

Linear trends

For some climate series it may be reasonable to assume that the mean level μ_t of the process at time t, is linearly associated with some covariate x_t of interest. Commonly this covariate x_t

may be a linear function of time, but it could be another series such as carbon dioxide levels at time t (or some smoothed version of this series). A *simple linear regression/trend model* is then given by

$$\mu_t = \beta_0 + \beta_1 x_t, \quad t \in T. \tag{2}$$

A naive estimator of the intercept parameter β_0 and slope parameter β_1 is given by *ordinary least squares (OLS)*. For series of length N, The OLS estimates,

$$\arg\min_{\beta_0, \beta_1} \sum_{t=1}^{N} (y_t - \beta_0 - \beta_1 x_t)^2, \tag{3}$$

are

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x} \quad \text{with} \quad \widehat{\beta}_1 = r \frac{s_y}{s_x},$$

where $\overline{y}$ is the sample mean of $\{y_t\}$, $\overline{x}$ is the sample mean of $\{x_t\}$, s_y is the sample standard dervation of $\{y_t\}$, s_x is the sample standard dervation of $\{x_t\}$, and r is the (Pearson) sample correlation between $\{x_t\}$ and $\{y_t\}$. When the noise $\{\eta_t\}$ is IID we have that the OLS estimates are the best linear unbiased estimates (BLUE) of β_0 and β_1 (they are unbiased estimates of the true parameters, and the variance of any linear combinations of the parameter estimates is smallest amongst all linear estimators). Since $\{\eta_t\}$ are usually not independent, a key question is how the OLS estimator behaves when $\{\eta_t\}$ is a dependent time series. When $\{\eta_t\}$ is a mean zero stationary process the OLS estimators are unbiased but are no longer the best when it comes to minimizing their variance among linear competitors. The same is true for modulated stationary processes.

When the variances of $\{\eta_t\}$ are changing with time, but the series are still independent, the best linear unbiased estimators come from *weighted least squares (WLS)*. The WLS estimates are

$$\arg\min_{\beta_0, \beta_1} \sum_{t=1}^{N} w_t (y_t - \beta_0 - \beta_1 x_t)^2, \tag{4}$$

where $w_t = 1/\text{var}(\eta_t)$ is the reciprocal of the variance of the noise at time t.

When the error terms are dependent, let $\mathbf{\Sigma} = \text{cov}(\boldsymbol{\eta})$ the covariance matrix for the noise, where $\boldsymbol{\eta}$ is the column vector of the η_t, meaning that $\mathbf{\Sigma}_{t,t'} = \text{cov}(\eta_t, \eta_{t'})$ for each t and t'. Letting $\boldsymbol{y}$ and $\boldsymbol{x}$ be the column vectors of the y_t and x_t, respectively, the *general least squares (GLS)* estimates,

$$\arg\min_{\beta_0, \beta_1} (\boldsymbol{y} - \beta_0 \mathbf{1} - \beta_1 \boldsymbol{x})^T \mathbf{\Sigma}^{-1} (\boldsymbol{y} - \beta_0 \mathbf{1} - \beta_1 \boldsymbol{x}) \tag{5}$$

(where $\mathbf{1}$ is a vector of ones), are the BLUEs of β_0 and β_1. OLS and WLS is a special case of GLS when $\mathbf{\Sigma}$ is a scaling of the identity matrix, or a diagonal matrix with diagonal entries $\text{var}(\eta_t)$, respectively.

Usually $\boldsymbol{\Sigma}$ is unknown. In that case, OLS is commonly used for exploratory purposes, even when we have correlated noise. Using the residuals calculated using via the OLS estimates,

$$\widehat{\eta}_t = y_t - \widehat{\beta}_0 - \widehat{\beta}_1 x_t,$$

we find candidate time series models. Using a given candidate model, we plug in the estimated $\boldsymbol{\Sigma}$ from the model in (5). Alternatively the time series parameters can be part of the minimization process (Goldstein, 1986), using OLS, followed by a number of steps of GLS, re-estimating the time series parameters after each OLS or GLS step.

If we are willing to assume a distribution for the noise process $\{\eta_t\}$, then we can estimate the parameters of the trend and time series process jointly using maximum likelihood (ML). For example, suppose that $\{\eta_t\}$ is a Gaussian process (i.e., the joint distribution of the noise at any collection of time indexes is normal). Let $\boldsymbol{\theta}$ denote the parameters characterizing the noise model, which drives $\boldsymbol{\Sigma}$. The likelihood function that needs to be maximized is

$$\begin{aligned} &L(\beta_0, \beta_1, \boldsymbol{\theta}) \\ &= (2\pi)^{-N/2} \left[\det(\boldsymbol{\Sigma})\right]^{-1/2} \exp\left\{(\boldsymbol{y} - \beta_0 \mathbf{1} - \beta_1 \boldsymbol{x})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \beta_0 \mathbf{1} - \beta_1 \boldsymbol{x})\right\}. \end{aligned}$$

It can be shown that the ML estimates of β_0 and β_1 are the the GLS estimates substituting in the ML estimate of $\boldsymbol{\theta}$ into $\boldsymbol{\Sigma}$. Typically the ML estimate of $\boldsymbol{\theta}$ is not available in closed form for most time series models; it is obtained via numerical optimization.

Assuming the noise vector to be Gaussian with mean $\mathbf{0}$ (a vector of zeros) and covariance $\boldsymbol{\Sigma}$, the vector of the GLS estimators $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1)^T$ of $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ have a bivariate normal sampling distribution:

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_2\left(\boldsymbol{\beta}, \left[\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X}\right]^{-1}\right), \tag{6}$$

where here the design matrix $\boldsymbol{X}$ is a $N \times 2$ matrix with first column all ones and the second column being $\boldsymbol{x}$. This distributional assumption assumes that we have the correct model for both the trend and the noise. In practice for statistical inference of $\boldsymbol{\beta}$, we plug-in an estimate of $\boldsymbol{\Sigma}$. The sampling distribution of the trend estimate, $\widehat{\boldsymbol{\mu}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ follows naturally (e.g., Ravishanker and Dey (2001)):

$$\widehat{\boldsymbol{\mu}} \sim N_n(\boldsymbol{\mu}, \boldsymbol{X}\text{cov}(\widehat{\boldsymbol{\beta}})\boldsymbol{X}^T). \tag{7}$$

The standard error for the fitted curve is then

$$\text{se}(\widehat{\mu}_t) = \sqrt{\boldsymbol{x}_t^T \text{cov}(\widehat{\boldsymbol{\beta}})\boldsymbol{x}_t} \quad t \in T, \tag{8}$$

where $\boldsymbol{x}_t$ is the tth row of $\boldsymbol{X}$. We can also derive a marginal sampling distribution for $\widehat{\mu}_t$ based on the t distribution when $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R}$. In this case $\boldsymbol{R}$ is a positive definite matrix and σ^2 is some positive variance parameter that is estimated using OLS; multivariate inference follows from F-tests (e.g., Ravishanker and Dey (2001)). (The t and F distribution assumes that the parameters characterzing $\boldsymbol{R}$ are known.)

Nonlinear and nonparametric trends

Clearly there is no *a priori* reason why the trend component $\{\mu_t\}$ of the climate series needs to be linear in time or in some covariate. Again, introducing the covariate $\{x_t\}$ and considering it now to be defined as a linear function of time, a simple model that allows for the trend to be nonlinear is the *polynomial trend/regression model*. A polynomial trend model of degree p has functional form

$$\mu_t = \beta_0 + \sum_{j=1}^{p} \beta_j\, x_t^j, \quad t \in T.$$

As with the simple linear trend model the coefficients in the polynomial model $\{\beta_j\}$ can be estimated in a number of different ways (e.g., OLS, WLS, GLS, and ML). As this is a linear model (e.g., Ravishanker and Dey (2001)), the properties of each estimator will again be driven by the statistical properties of the noise process, and these properties are essentially the same as for the simple linear trend model. Compared to (6), the GLS estimator now has a $(p+1)$-variate normal sampling distribution for a Gaussian noise process, where the design matrix $\boldsymbol{X}$ has ith row

$$\boldsymbol{x}_i = (1, x_i, x_i^2, \ldots, x_i^p)^T, \quad i = 1, \ldots, N.$$

(7) and (8) also hold for this model. In choosing a polynomial trend model for data we need to select the degree of the polynomial, p. This is a common problem in both regression and time series analysis,where there is a need to balance the the smoothness of the trend that we observe in the climate series, with writing down a model that adequately captures the noise. We should be wary of overfitting (writing down a statistical model for the trend for which the estimated coefficient are near zero, but are highly uncertain). Overfitting leads to statistical models with poor predictive performance. Penalized regression and penalized likelihood methods can be applied to ensure that we do not fit models that are overly complicated and poor descriptors of the data. Demonstrating the idea of penalized ML methods, we will use the Akaike Information Criterion (AIC) in our case studies in Section 4 below.

There are others examples of parametric models for the trend beyond polynomial models. To see this, introducing the $p+1$ functions $\{f_j(x) : j = 0, \ldots, p\}$, defined as

$$f_j(x) = x^j, \quad j = 0, \ldots, p,$$

we see that the polynomial trend can be written in more a general notation:

$$\mu_t = \sum_{j=0}^{p} \beta_j f_j(x_t), \quad t \in T.$$

Replacing the polynomial functions by other smooth functions, lead to more generalized models for trend. Commonly used examples of so-called *basis functions* include sinusoids (e.g., Bloomfield (2000); Percival and Walden (1993)), wavelets (e.g., Brillinger (1994a, 1996); Craigmile

et al. (2004); Percival and Walden (2000)), splines (e.g., Wahba (1990); Wang (1998); Wood (2006)), and local regression models (e.g., Cleveland et al. (1992); Loader (1999)). Given the multitude of possible functions that can be used to model trend, model selection and comparison becomes more problematic. We again need to be wary of overfitting; penalized regression and penalized regression models are popularly used for fitting trends constructed from a linear combination of basis functions. In the Bayesian paradigm there is extensive research into building priors for the regression parameters to enforce smoothness, while guarding against overfitting by enforcing sparseness.

Uncertainty quantification for the estimated trends are typically more involved and more computationally intensive to calculate, especially when Bayesian penalized methods are used. Depending on the paradigm, Markov chain Monte Carlo (MCMC) (e.g., Brooks et al. (2011); Gelman et al. (2013)), or Monte Carlo and other resampling methods are commonly used for assessing the uncertainty or significance of trend estimates (e.g., Cabilio et al. (2013); Craigmile et al. (2004); Efron and Tibshirani (1994); Lahiri (2013)).

A particularly simple and effective nonlinear model for trend is the *broken stick model*, which is a first order spline model with one knot. The broken stick model is useful when there is a qualitative change in the system at some time point. Sometimes this time point is known, sometimes only suspected. As with linear models, generalized least squares or maximum likelihood, for example, can be used to estimate the parameters in the model, while estimating the dependence in the errors. Given a change point time (or knot) τ we fit a straight line to $y_1, \ldots, y_\tau$ as a function of $x_1, \ldots, x_\tau$, and a connecting line to $y_{\tau+1}, \ldots, y_N$ as a function of $x_{\tau+1}, \ldots, x_N$. If desired, the parameter τ can also be estimated in the model. If a smoother version is desirable, a *bent cable model* (Chiu and Lockhart, 2010) connects the two linear parts with a quadratic connector. The model is then

$$\mu_t = \beta_0 + \beta_1 t + \beta_2 \left[\frac{(t - \tau + \gamma)^2}{4\gamma} I\{|t - \tau| \leq \gamma\} + (t - \tau)I\{t > \tau + \gamma\} \right], \quad t \in T. \qquad (9)$$

In other words, a straight line is fitted between indexes 1 and $\tau - \gamma$, a quadratic between $\tau - \gamma$ and $\tau + \gamma$, and another line between $\tau + \gamma$ and n, with the three segments connecting continuously. The model can be fit using least squares or via ML.

Smoothing and filtering to estimate the trend

The *moving average (MA) filter* can be used to estimate the trend of a climate series $\{Y_t : t = 1, \ldots, n\}$ nonparametrically. The trend estimate from an MA(q) filter is defined by

$$\widehat{\mu}_t = \sum_{j=-q}^{q} \left[\frac{1}{2q+1} \right] Y_{t-j}, \quad t = q+1, \ldots, n-q$$

(care needs to be taken to define the trend estimate for $t < (q + 1)$ and $t > (n - q)$). The value of q (a positive integer) controls the level of smoothing of the filter, while controlling the influence of the noise process upon the trend estimate (i.e., it manages the bias–variance

tradeoff in estimating the trend). Any MA(q) filter provides an unbiased estimate of a linear or locally-linear trend (Chapter 1 of Brockwell and Davis (2002)).

Again for some positive integer q, a more general smoothing filter $\{a_j : j = -q, \ldots, q\}$ can be designed so that the trend estimate,

$$\widehat{\mu}_t \;\; = \;\; \sum_{j=-q}^{q} a_j Y_{t-j},$$

can unbiasedly estimate trends of a given degree of polynomial (e.g., Exercise 1.12 of Brockwell and Davis (2002)). With the additive decomposition for trend given by (1), uncertainty quantification for these trend estimates follow by assuming a model for the noise, and calculating the covariance of the filtered noise using the stationarity preserves a stationary result.

We need to be cautious with some choices of smoothing filter, because it is possible to introduce features in the estimated trend that do not exist in the original time series. Often it can help to think of smoothing in terms of its effect upon the low and high frequency components of the climate series $\{Y_t\}$; namely, smoothing emphasizes the low frequency components of the series ("the trend"), and reduces the effect of the high frequency component ("the noise"). Looking at the so-called spectral properties of filters can indicate issues that may occur with smoothing and filtering; see, e.g., Percival and Walden (1993) for more details of the spectral approach.

Removing or simplifying trend by differencing

Rather than estimating the trend component for a climate series, we may choose to remove or simplify it. For example the differencing operator, ∇, has the ability to remove trend by taking a polynomial of degree p and yielding a polynomial of degree $p - 1$ (e.g., Exercise 1.10, p.42 of Brockwell and Davis (2002)). Formally for the series $\{Y_t\}$, the differencing operator ∇ is defined by

$$\nabla Y_t = Y_t - Y_{t-1}, \quad t \in T.$$

Now suppose that $\mu_t = \beta_0 + \beta_1 t$ is a linear trend in t. Then $\nabla \mu_t = \beta_1$ for all t. Thus differencing a linear trend yields a constant trend, with a trend equal to the slope of the original series. This demonstrates another way to estimate the slope for a simple linear trend model as defined by (2): we difference the climate series, and estimate the slope using the sample mean for the differencing time series. The statistical properties of this estimate of the slope may be involved, and are driven by the statistical properties of the differenced noise process, $\{\nabla \eta_t\}$.

Hierarchical and dynamic linear model decompositions for trend

Writing down statistical models defined in terms of differencing, allows for stochastic representations of trend. These stochastic representations are more commonly used in financial applications (e.g., Tsay (2010)), but can also be used in climatic analyses. A simple example of a model defined using differencing is the autoregressive integrated moving average ARIMA(p, d, q) process.

By noting that an ARIMA model can be written as a form of random walk model, we are drawn to define stochastic models for trend in terms of a hierarchical statistical model (Durbin and Koopman, 2012; Hyndman et al., 2008). Here follows a simple example. Again suppose $\{Y_t : t \in T\}$ is our time series process of interest and assume the additive decomposition of

$$Y_t = \mu_t + \eta_t, \quad t \in T,$$

where $\{\eta_t\}$ is the noise process. Now, rather than assuming that the trend μ_t is deterministic we assume a random walk model for μ_t:

$$\mu_t = \mu_{t-1} + Z_t, \quad t \in T,$$

where $\{Z_t\}$ is a white noise or IID process. For this model we can estimate the latent trend component process using the computationally efficient Kalman filtering algorithm (e.g., Durbin and Koopman (2012); Shumway and Stoffer (1982, 2010)). Parameter estimation follows in the Gaussian process setting using ML. In the Bayesian context this is an example of a *Dynamic Linear Model (DLM)* (Harrison and West, 1999). This class of models can be related to the smoothing and spline methods discussed above (e.g., Hyndman et al. (2008); Lindgren and Rue (2008); Wecker and Ansley (1983)).

Two case studies

We now return to the two data sets presented in Section 2. We will demonstrate that failure to investigate residual structure, or to make assumptions appropriate for the data, can yield incorrect conclusions.

US annual temperatures

The standard approach in the climate literature is to fit a linear trend to a time series using ordinary least squares (3). Doing that for the US annual temperature anomalies yields an estimate of 0.55°C per century (see Table 1 for details of all the fits in this section). The slope is highly significantly different from zero, but that assumes the noise is IID and does not take into account the variability of the measurements, leading to an overfit of the more uncertain early measurements. In order to deal with this, we do a weighted least squares fit (4). The resulting estimate is somewhat smaller (0.48°C per century) and somewhat more uncertain, although still highly significant (again assuming the noise is IID).

Looking at the sample autocorrelation function (ACF) of the residuals from the two least squares fits in Figure 3, we see clear evidence of dependence for both fits. Indeed, compared to the OLS fit, there is evidence that the dependence is stronger once we account for the change in variability over time using the WLS fit. The last column of Table 1, showing the Ljung-Box (Ljung and Box, 1978) P-value when testing that the residuals are IID (based on 10 lags

Table 1. Fits of linear trend slopes to US annual average temperature anomalies

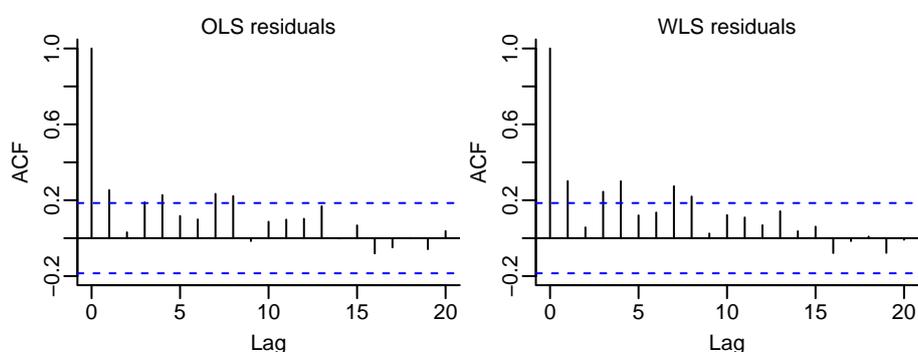| Model	Slope per century	Standard error	Slope P-value	AIC	Ljung-Box P-value
OLS	0.55	0.12	<0.0001	116.0	0.000
WLS	0.48	0.14	0.0006	124.8	0.000
Shen et al.	0.57	0.17	0.0008	N/A	N/A
AR(1)	0.55	0.15	0.0004	110.6	0.066
Weighted AR(1)	0.47	0.18	0.0107	116.1	0.013
AR(4)	0.58	0.20	0.0038	109.3	0.593
Weighted AR(4)	0.53	0.26	0.0490	110.1	0.741
ARMA(3,1)	0.64	0.24	0.0092	107.4	0.814
Weighted ARMA(3,1)	0.60	0.32	0.0655	109.1	0.849

Figure 3. Sample autocorrelation functions for the residuals from the ordinary and weighted least squares linear trend fit to the US annual mean temperature anomalies.

of the ACF), demonstrates that there is significant dependence in the OLS and WLS residuals that needs to be accounted for.

The most common time series model in the climate literature is AR(1). To fit the two AR(1) models (unweighted: assuming a constant variance over time; weighted: assuming that the variance is time-varying) we use the generalized least squares estimator, given by (5). In essence, Shen et al. (2012) use this approach to estimate slopes. The resulting slopes are almost the same as for the corresponding fits using the independence assumption, but the standard errors become 20-30% larger. The fitted slopes are still significantly different from zero under the AR(1)-assumption for the noise structure. At the $\alpha = 0.05$ level, the Ljung-Box test rejects the null hypothesis of IID noise for either set of AR(1) residuals.

Sticking to the autoregressive error structure, the best such fit is AR(4). The resulting slopes as well as standard errors become larger, and the weighted AR(4) test for the significance of the slope parameter has a P-value of just lower than 0.05. But the best ARMA-type error structure
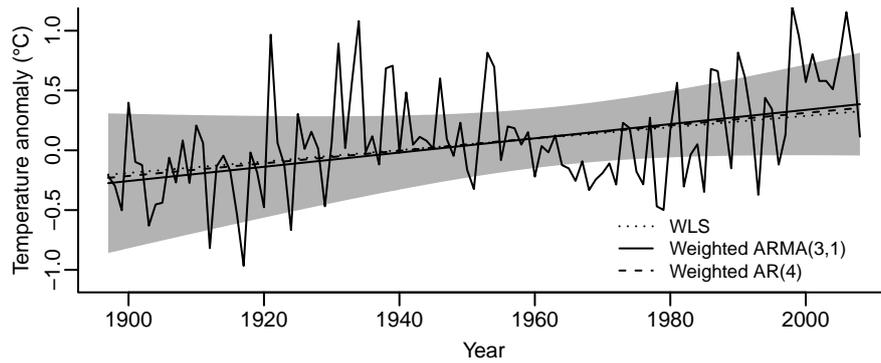
Figure 4. US annual mean temperature anomalies with trend lines overlaid from the WLS fit (dotted line), generalized least squares with weighted ARMA(3,1) errors (solid line), and weighted AR(4) errors (dashed line). Simultaneous 95% confidence bands for the ARMA(3,1) trend are shown in gray.

is ARMA(3,1), which also yield the largest slopes. The weighted ARMA(3,1)-slope is not significantly different from zero. Figure 4 shows three of the fitted lines. All fits from these models pass the Ljung-Box IID noise test.

Table 1, in addition to showing that the statistical significance of the slope strongly depends on the assumptions made, indicates that unweighted ARMA(3,1) model has the best fit according to the AIC criterion. Choosing the model with the lowest AIC assumes that the errors are Gaussian and that the trend is linear. Since, the weighted residuals are slightly less skewed than the unweighted residuals, we choose the weighted ARMA(3,1) or weighted AR(4) general least square fits to explain the dependence, which leads us to conclude that the linear trend is only weakly significant, once we account for the time-varying variance and dependence in the series.

Global annual mean temperature

When looking at the global temperature time series, we start by fitting a straight line using least squares. The linear fit is not particularly good, and the residuals (Figure 5) indicate that a quadratic fit might be better.

Using a similar approach to that in Section 4.1, we use a generalized least squares fit to a trend that is a quadratic function of time, with time series errors following an ARMA(4,1)-model and weights corresponding to the estimated observation variance. Figure 6 shows the fitted model. Again, the residuals pass various tests for IID noise. We calculate the standard error of the fitted curve from (8). A Bonferroni calculation shows that a simultaneous 95% confidence band is given by going up and down 3.6 standard errors from the fitted curve.

A broken stick model is suggested by the idea that at some point in the twentieth century the human-generated greenhouse gases may have started to dominate the natural forcings of the climate system, thereby changing the rate of increase in global temperatures. The estimated
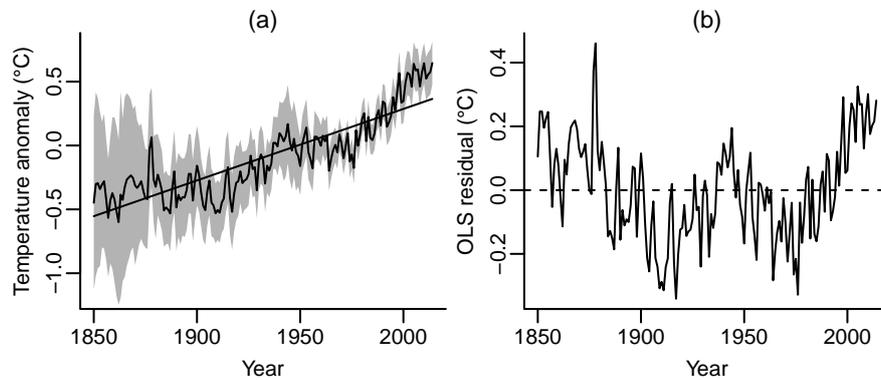
Figure 5. (a) Global annual mean temperature anomalies with respect to the 1951-1970 climatology. The gray shaded region denotes simultaneous 95% confidence intervals for the mean, using estimated standard errors from Rohde et al. (2013). The trend line correspond to ordinary least squares. (b) Residuals from the ordinary least squares fit.
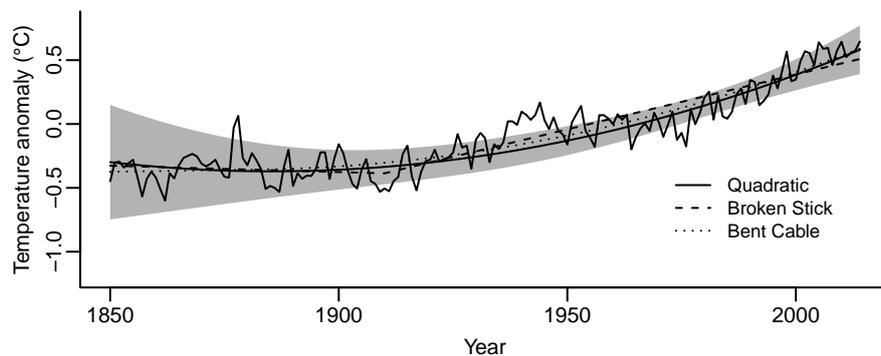

Figure 6. Global annual mean temperature anomalies with respect to the 1951-1970 climatology. Different line types denote different estimated trends. The gray shaded regions denotes simultaneous 95% confidence bands for the quadratic trend estimate, assuming ARMA(4,1) errors and time-varying variances.

broken stick model changes slope at 1909 (confidence interval 1894 – 1924). When fitting a bent cable model using AR(4) error structure to the global temperature series, the resulting curve is very similar to the quadratic fit, as seen in Figure 6. The quadratic part of the fit goes between 1863 and 2014; i.e. almost the entire series. Both the broken stick and bent cable models fall inside the quadratic simultaneous confidence band.

Since the global series is poorly fit by a straight line, the decision in the IPCC Fifth Assessment Report to only report straight lines is not supported by this data set. It should be noted that the report uses the Hadley global temperature series (Morice et al., 2012), but an analysis of that series yields nearly identical results to our analysis in this section.

Spatial and spatio-temporal trends

Statistical methods of trend estimation and assessment extend naturally to the spatial and spatio-temporal settings. (For a review of spatial and spatio-temporal models, see e.g., Cressie and Wikle (2011); Gelfand et al. (2010); Le and Zidek (2006); see Chorley and Haggett (1965) for an early discussion of spatial trend estimation.) Not all of the papers that appear in this section are climate examples, but the methodology can be applied to the analysis of climate variables observed over space and time.

An an example, suppose we wish to estimate trend for a point-referenced (also known as a geostatistical) spatial process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^p\}$, based on n observations from the process, $\boldsymbol{z} = (z(\boldsymbol{s}_1), \ldots z(\boldsymbol{s}_n))^T$. Assuming an additive decomposition for this process $\{Z(\boldsymbol{s})\}$ we have

$$Z(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \eta(\boldsymbol{s}), \quad \boldsymbol{s} \in D,$$

where $\{\mu(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ is the spatial trend component and $\{\eta(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ is the spatial noise component. As in the time series case, we typically assume that the noise is a mean zero stationary process. Again, linear model representations are commonly used for the trend component:

$$\mu(\boldsymbol{s}) = \sum_{j=0}^{p} \beta_j \, \boldsymbol{x}_j(\boldsymbol{s}),$$

where $\{\boldsymbol{x}_j(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ are $p + 1$ spatial basis functions. Traditionally, simple polynomials (e.g., Cressie and Wikle (2011); Gelfand et al. (2010)), splines (e.g., Cressie and Johannesson (2008); Nychka (2000); Sangalli et al. (2013); Wood (2003)), and other functions such as wavelets are used to model the trend in spatial settings (e.g., Álvarez and Sansó (2008); Cressie and Johannesson (2008); Nychka et al. (2002)). Inference follows in the same way as for time series (e.g., via least squares, ML, penalized methods, and Bayesian methods).

Also, given that we can represent spline models (and indeed other basis models) using random spatial processes (e.g., Kleiber and Nychka (2015); Nychka (2000)), we could also assume that the trend component $\{\mu(\boldsymbol{s}) : \boldsymbol{s} \in D\}$ is a random process that smoothly varies over space.

In such a case we need to specify the model for the climate variables in such a way that the trend component captures the smooth (long-range) variation of the process over space, and the errors captures the noise, the short range variation of the process over space (cf Sang and Huang (2011)).

These ideas extend to modeling spatial trends for areal processes (e.g., Besag et al. (1991); Gelfand et al. (2010)), point processes (see e.g., Baddeley and Turner (2006); Brillinger (1994b); Diggle (2013); Lawless and Thiagarajah (1996); Lawson (1988) for temporal and spatial examples), and to river networks (e.g., O'Donnell et al. (2013)). Spatio-temporal methods follow naturally (e.g., Craigmile and Guttorp (2011); Cressie and Wikle (2011); Lemos and Sansó (2009); Paciorek and McLachlan (2009); Sampson et al. (2011); Waller et al. (1997); Wikle et al. (2001)).

Assessing climatic trends in other contexts

Up to this point we have considered the trend to be smooth changes over a mean over time and/or space. (Indeed the trend for a point process, as measured by the intensity function (e.g., Diggle (2013)), can also be considered to be a mean number of events occurring in a specific time interval or spatial region.) More recently there has been an interest in modeling trends in other features of the distribution of climate; for example, trends in the variance (e.g., Craigmile and Guttorp (2011)), trends in the quantiles (e.g., Chandler and Scott (2011); Lee et al. (2013); Reich (2012)), and trends in extremes.

In the latter case, there are countless examples of using spatio-temporal models to assess trends in the extreme-value distribution of climate variables, usually in the location parameter of the generalized extreme value distribution for block maxima. Some examples of extreme value studies using different climate variables include: temperature (Brown and Katz, 1995; Craigmile and Guttorp, 2013; Fuentes et al., 2013; Kottas et al., 2012), paleoclimate proxies of temperature (Mannshardt et al., 2013), and precipitation (Cooley et al., 2007; Davison et al., 2012; Schliep et al., 2009). The use of hierarchical Bayesian modeling has revolutionized our ability to examine trends in extremes. For example, we can introduce a spatial process for the spatially-varying location parameter of the extreme value distribution to introduce some (albeit limited) smoothness over space. See Davison et al. (2012) for a comprehensive review of the issues underlying the modeling of spatial extremes.

Discussion

In this chapter we have discussed the many issues underlying the modeling and assessment of climate trends. Defining trend to be smooth changes in time or space, we reviewed the idea that

trend is ill-posed because we have not specified how smooth a trend we have. This requires us to introduce some prior belief about the trend (relative to the noise), via a suitable model or modeling framework for the climate variable of interest. For example, a Gaussian additive model with a specified component capturing the trend as the mean of the process may be reasonable for modeling long-term average temperatures, whereas for modeling precipitation block monthly maxima, the location parameter of a generalized extreme value distribution is more appropriate. We demonstrated that trends do not need to be linear, but once we allow for nonlinear trends, we need to be wary of overfitting to the climate data at hand.

Some authors have argued that the increasing temperature trend observed in many climate series is due to a long memory error structure, caused by the slow reaction to forcing in oceans as compared to atmosphere. This has led to a rich literature in researchers trying to purely describe the variation (and indeed the trends themselves) using only long memory processes (for a review, see, e.g., Koutsoyiannis (2003)). Smith (1993) demonstrated that a linear trend, after accounting for long memory errors, was statistically significant when estimated from the Hansen and Lebedeff (1987) global temperature series. (For other examples see Craigmile and Guttorp (2011); Craigmile et al. (2004); Lemos and Sansó (2009); see Cohn and Lins (2005) for further discussion of the long memory versus trend question.)

It is not possible to highlight all methods for assessing and modeling trend in this chapter. For example, tests for trend based on Kendall's tau statistics have some popularity in environmental and climate science (e.g., Darken et al. (2000); El-Shaarawi and Niculescu (1992)). We also note the popularity of principal components (also known as empirical orthogonal components; see, e.g., Berliner et al. (2000); Thiebaux and Pedder (1987); von Storch and Zwiers (2002)) as trend components in atmospheric and oceanic sciences.

There are numerous future directions for statistical research into trend estimation. Using statistical learning methodologies, there is interest in applying different loss functions to the efficient, while robust, estimation of trend (e.g., Kim et al. (2009)). Further development of spatio-temporal methods for trend estimation is also needed in non-Gaussian and multivariate settings; a thought-provoking example in climate is the joint estimation of trends in the occurrence and intensity of precipitation amounts (e.g., Cowpertwait (1995); Guttorp (1988)).

Acknowledgments

References

Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. In Salkind, N. J., editor, *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, CA. 5

Álvarez, G. and Sansó, B. (2008). Bayesian wavelet regression for spatial estimation. *Journal of Data Science*, 6:219–229. 18

Baddeley, A. and Turner, R. (2006). Modelling spatial point patterns in R. In *Case studies in spatial point process modeling*, pages 23–74. Springer, New York, NY. 19

Barboza, L., Li, B., Tingley, M. P., Viens, F. G., et al. (2014). Reconstructing past temperatures from natural proxies and estimated climate forcings using short-and long-memory models. *The Annals of Applied Statistics*, 8:1966–2001. 8

Beran, J. (1994). *Statistics for Long Memory Processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, NY. 8

Berliner, L. M., Wikle, C. K., and Cressie, N. A. (2000). Long-lead prediction of pacific ssts via bayesian dynamic modeling. *Journal of Climate*, 13:3953–3968. 20

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–20. 19

Bloomfield, P. (2000). *Fourier analysis of time series (Second Edition)*. Wiley-Interscience, New York, NY. 7, 11

Brillinger, D. R. (1994a). Some river wavelets. *Enivronmetrics*, 5:211–220. 7, 11

Brillinger, D. R. (1994b). Trend analysis: time series and point process problems. *Environmetrics*, 5:1–19. 19

Brillinger, D. R. (1996). Some uses of cumulants in wavelet analysis. *Nonparametric Statistics*, 6:93–114. 7, 11

Brockwell, P. J. and Davis, R. (2002). *Introduction to Time Series and Forecasting (Second Edition)*. Springer, New York, NY. 7, 13

Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L. (2011). *Handbook of MCMC*. Chapman and Hall/CRC, Boca Raton, FL. 12

Brown, B. and Katz, R. (1995). Regional analysis of temperature extremes: Spatial analog for climate change? *Journal of Climate*, 8:108–119. 19

Cabilio, P., Zhang, Y., and Chen, X. (2013). Bootstrap rank tests for trend in time series. *Envronmetrics*, 24:537–549. 12

Chandler, R. and Scott, M. (2011). *Statistical methods for trend detection and analysis in the environmental sciences*. John Wiley & Sons. 19

Chiu, G. S. and Lockhart, R. A. (2010). Bent-cable regression with autoregressive noise. *Canadian Journal of Statistics*, 38:386–407. 12

Chorley, R. J. and Haggett, P. (1965). Trend-surface mapping in geographical research. *Transactions of the Institute of British Geographers*, 37:47–67. 18

Cleveland, W. S., Grosse, E., and Shyu, W. M. (1992). Local regression models. In Chambers, J. and Hastie, T., editors, *Statistical Models in S*, chapter 8. Wadsworth and Brooks/Cole. 12

Cohn, T. A. and Lins, H. F. (2005). Nature's style: Naturally trendy. *Geophysical Research Letters*, 32:5. 4, 8, 20

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102:824–840. 19

Cowpertwait, P. S. (1995). A generalized spatial-temporal model of rainfall based on a clustered point process. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 450, pages 163–175. The Royal Society. 20

Craigmile, P. F. and Guttorp, P. (2011). Space-time modelling of trends in temperature series. *Journal of Time Series Analysis*, 32:378–395. 19, 20

Craigmile, P. F. and Guttorp, P. (2013). Can a regional climate model reproduce observed extreme temperatures? *Statistica*, 73:103. 19

Craigmile, P. F., Guttorp, P., and Percival, D. B. (2004). Trend assessment in a long memory dependence model using the discrete wavelet transform. *Environmetrics*, 15:313–335. 4, 7, 8, 11, 12, 20

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 70:209–226. 18

Cressie, N. A. C. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, New York, NY. 18, 19

Darken, P., Holtzman, G., Smith, E., and Zipper, C. E. (2000). Detecting changes in trends in water quality using modified Kendall's tau. *Environmetrics*, 11:423–434. 20

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27:161–186. 19

Diggle, P. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*. CRC Press, Boca Raton, FL. 19

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press, Oxford, United Kingdom. 14

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press, Boca Raton, FL. 12

El-Shaarawi, A. and Niculescu, S. P. (1992). On Kendall's tau as a test of trend in time series data. *Environmetrics*, 3:385–411. 20

Fuentes, M., Henry, J., and Reich, B. (2013). Nonparametric spatial models for extremes: Application to extreme temperature data. *Extremes*, 16:75–101. 19

Gallagher, C., Lund, R., and Robbins, M. (2013). Changepoint detection in climate time series with long-term trends. *Journal of Climate*, 26:4994–5006. 4

Gelfand, A., Diggle, P., Fuentes, M., and Guttorp, P. (2010). *Handbook in Spatial Statistics*. Chapman and Hall, Boca Raton, FL. 18, 19

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis, 3rd edition*. Chapman and Hall, Boca Raton, FL. 12

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 78:43–56. 10

Guttorp, P. (1988). Analysis of event based precipitation data with a view towards modeling. *Water Resources Research*, 24:35–44. 20

Hansen, J. and Lebedeff, S. (1987). Global trends of measured surface air temperature. *Journal of Geophysical Research*, 92:13345–13372. 20

Harrison, J. and West, M. (1999). *Bayesian Forecasting and Dynamic Models*. Springer, New York, NY. 14

Hartmann, D., Klein-Tank, A., Rusticucci, M., Alexander, L., Broìnnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A., Soden, B., Thorne, P., Wild, M., and Zhai, P. (2013). Observations: Atmosphere and surface. In Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter 2, pages 159–254. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 4

Hussain, S. and Elbergali, A. (1999). Fractional order estimation and testing, application to Swedish temperature data. *Environmetrics*, 10:339–349. 8

Hyndman, R., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer Science and Business Media, Berlin, Germany. 14

Kendall, M. (1973). *Time Series*. Charles Griffin, London. 7

Kennedy, J., Rayner, N., Smith, R., Saunby, M., and Parker, D. (2011a). Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 1: measurement and sampling errors. *Journal of Geophysical Research*, 116(D14103). 6

Kennedy, J., Rayner, N., Smith, R., Saunby, M., and Parker, D. (2011b). Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 2: biases and homogenisation. *Journal of Geophysical Research*, 116(D14104). 6

Kim, S.-J., Koh, K., Boyd, S., and Gorinevsky, D. (2009). l_1 trend filtering. *SIAM review*, 51:339–360. 20

Kleiber, W. and Nychka, D. W. (2015). Equivalent kriging. *Spatial Statistics*, 12:31–49. 18

Kottas, A., Wang, Z., and Rodriguez, A. (2012). Spatial modeling for risk assessment of extreme values from environmental time series: a Bayesian nonparametric approach. *Environmetrics*, 23:649–662. 19

Koutsoyiannis, D. (2003). Climate change, the Hurst phenomenon, and hydrologic statistics. *Hydrological Sciences Journal*, 48:3–24. 8, 20

Lahiri, S. N. (2013). *Resampling methods for dependent data*. Springer, New York, NY. 12

Lawless, J. and Thiagarajah, K. (1996). A point-process model incorporating renewals and time trends, with application to repairable systems. *Technometrics*, 38:131–138. 19

Lawson, A. (1988). On tests for spatial trend in a non-homogeneous Poisson process. *Journal of Applied Statistics*, 15:225–234. 19

Le, N. D. and Zidek, J. V. (2006). Statistical analysis of environmental space-time processes. *Springer, New York*. 18

Lee, K., Baek, H.-J., and Cho, C. (2013). Analysis of changes in extreme temperatures using quantile regression. *Asia-Pacific Journal of Atmospheric Sciences*, 49:313–323. 19

Lemos, R. and Sansó, B. (2009). A spatio-temporal model for mean, anomaly, and trend fields of north atlantic sea surface temperature. *Journal of the American Statistical Association*, 104:5–18. 19, 20

Lindgren, F. and Rue, H. (2008). On the second-order random walk model for irregular locations. *Scandinavian Journal of Statistics*, 35:691–700. 14

Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65:297–303. 14

Loader, C. (1999). *Local Regression and Likelihood*. Springer, New York, NY. 12

Mannshardt, E., Craigmile, P. F., and Tingley, M. P. (2013). Statistical modeling of extreme value behavior in North American tree-ring density series. *Climatic Change*, 117:843–858. 19

McCullagh, P. and Nelder, J. A. (1999). *Generalized Linear Models*. Chapman and Hall Ltd. 7

Menne, M. J., Williams Jr., C. N., and Vose, R. S. (2009). The U.S. Historical Climatology Network monthly temperature data, version 2. *Bull. Amer. Meteor. Soc.*, 90:993–1007. 5

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 dataset. *Journal of Geophysical Research*, 117. 18

Nychka, D., Wikle, C. K., and Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, 2:315–331. 18

Nychka, D. W. (2000). Spatial-process estimates as smoothers. In Schimek, M. G., editor, *Smoothing and Regression. Approaches, Computation and Application*, pages 393–424. Wiley, New York, NY. 18

O'Donnell, D., Rushworth, A., Bowman, A. W., Marian Scott, E., and Hallard, M. (2013). Flexible regression models over river networks. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 63:47–63. 19

Paciorek, C. J. and McLachlan, J. S. (2009). Mapping ancient forests: Bayesian inference for spatio-temporal trends in forest composition using the fossil pollen proxy record. *Journal of the American Statistical Association*, 104:608–622. 19

Percival, D. and Walden, A. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press, Cambridge, United Kingdom. 7, 11, 13

Percival, D. and Walden, A. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge, United Kingdom. 12

Ravishanker, N. and Dey, D. (2001). *A First Course in Linear Model Theory*. Chapman and Hall. 10, 11

Reich, B. J. (2012). Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61:535–553. 19

Rohde, R. and Muller, R. (2015). The average temperature of 2014. Results from Berkeley Earth. Technical report, Berkeley Earth. Available from: http://static.berkeleyearth.org/memos/Global-Warming-2014-Berkeley-Earth-Newsletter.pdf. 6

Rohde, R., Muller, R. A., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D., and Wickham | , C. (2013). A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinfor Geostat: An Overview*, 1. 5, 6, 17

Sampson, P. D., Szpiro, A. A., Sheppard, L., Lindström, J., and Kaufman, J. D. (2011). Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data. *Atmospheric Environment*, 45:6593–6606. 19

Sang, H. and Huang, J. Z. (2011). A full scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 74:111–132. 19

Sangalli, L. M., Ramsay, J. O., and Ramsay, T. O. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 75:681–703. 18

Schliep, E. M., Cooley, D., Sain, S. R., and Hoeting, J. A. (2009). A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, 13:219–239. 19

Shen, S. S. P., Lee, S. K., and Lawrimore, J. (2012). Uncertainties, trends, and hottest and coldest years of U.S. surface air temperature since 1895: An update based on the USHCN V2 TOB data. *J. Climate*, 25:4185–4203. 4, 5, 15

Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3:253–264. 14

Shumway, R. H. and Stoffer, D. S. (2010). *Time series analysis and its applications: with R examples*. Springer, New York, NY. 14

Smith, R. (1993). Long-range dependence and global warming. In Barnett, V. and Turkman, F., editors, *Statistics for the Environment*, pages 141–161. John Wiley, Hoboken, NJ. 4, 20

Thiebaux, H. and Pedder, M. (1987). *Spatial Objective Analysis*. Academic Press, Cambridge, MA. 20

Trewin, B. (2010). Exposure, instrumentation, and observing practice effects on land temperature measurements. *WIREs Climate Change*, 1:490–506. 6

Tsay, R. S. (2010). *Analysis of Financial Time Series (Third Edition)*. Wiley, New York, NY. 13

von Storch, H. and Zwiers, F. W. (2002). *Statistical analysis in climate research*. Cambridge University Press, Cambridge, United Kingdom. 4, 20

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM (Society for Industrial and Applied Mathematics), Philadelphia. 7, 12

Waller, L. A., Carlin, B. P., Xia, H., and E, G. A. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association*, 92:607–617. 19

Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, 93:341–348. 12

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X.-L., Choi, D., Cheang, W.-K., Keller, T., DeLuisi, J., Wuebbles, D. J., and Kerr, J. (1998). Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *Journal of Geophysical Research*, 103(D14):17–149. 4

Wecker, W. E. and Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, 78:81–89. 14

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001). Spatiotemporal hierarchical Bayesian modeling tropical ocean surface winds. *Journal of the American Statistical Association*, 96:382–397. 19

Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press, Boca Raton, FL. 7, 12

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 65:95–114. 18