Handbook of Environmental and Ecological Statistics

Alan Gelfand, Montse Fuentes, Jennifer A. Hoeting, Richard L. Smith

Modeling and assessing climatic trends

Publication details
Peter F. Craigmile, Peter Guttorp
Published online on: 15 Sep 2017

How to cite :- Peter F. Craigmile, Peter Guttorp. 15 Sep 2017, Modeling and assessing climatic trends from: Handbook of Environmental and Ecological Statistics CRC Press
Accessed on: 07 Dec 2023

PLEASE SCROLL DOWN FOR DOCUMENT
Modeling and assessing climatic trends

Peter F. Craigmile
Department of Statistics, The Ohio State University, Columbus, Ohio, USA.

Peter Guttorp
Department of Statistics, University of Washington, Seattle, Washington, USA and Norwegian Computing Center, Oslo, Norway

CONTENTS

27.1 Introduction ................................................................. 619
27.2 Two motivating examples ............................................. 620
  27.2.1 US average temperature anomaly ............................... 620
  27.2.2 Global temperature series ....................................... 621
27.3 Time series approaches .............................................. 621
  27.3.1 Candidate models for the noise .................................. 622
  27.3.2 Linear trends ......................................................... 623
  27.3.3 Nonlinear and nonparametric trends ........................... 625
  27.3.4 Smoothing and filtering to estimate the trend ............... 627
  27.3.5 Removing or simplifying trend by differencing .............. 627
  27.3.6 Hierarchical and dynamic linear model decompositions for trend ... 628
27.4 Two case studies ......................................................... 628
  27.4.1 US annual temperatures .......................................... 628
  27.4.2 Global annual mean temperature ............................... 630
27.5 Spatial and spatio-temporal trends ................................. 632
27.6 Assessing climatic trends in other contexts ..................... 633
27.7 Discussion ................................................................. 633
Bibliography ................................................................. 634

27.1 Introduction

Studies of climate change often involve the statistical analysis of climate variables (or their meteorological counterparts) indexed in space and time [98]. 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 trends in climate data [19, 21, 84, 90, 102]. Trends are often described by smooth changes in 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 illustrate, 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. 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”).

In the latest IPCC report [46] 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 [36].) We discuss the estimation of trend in nontraditional settings, such as in the analysis of climate extremes, and point to future directions in the statistical assessment of climatic trends.

27.2 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.

27.2.1 US average temperature anomaly

[90] produced statistical estimates of US temperature anomalies from 1897–2008, using the US Historical Climatology Network data set version 2 [70], 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, natural variability, orographic effects, etc. [90] was mainly concerned with the statistical estimation of this standard error. Figure 27.1(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 region denotes simultaneous 95% confidence intervals for this mean, calculated using the estimated standard errors with a Bonferroni correction [1]. Figure 27.1(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.
Modeling and assessing climatic trends

27.2.2 Global temperature series

The Berkeley Earth project [84] uses isotropic geostatistical tools (kriging; see Chapter 5) to estimate the global mean temperature. The land data have been collected from 14 databases and 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 [96]. If a measurement device is moved or replaced, it is considered a different station, rather than being “corrected.” For a statistician, this seems to be a natural approach. The ocean data used in the Berkeley Earth series [83] come from the Hadley Center sea surface temperature data set HadSST3 [51, 52], modified by kriging of missing grid squares. This allows for a statistically justifiable estimate of global mean estimation uncertainty. Figure 27.2, formatted identically as Figure 27.1, demonstrates that the standard errors tend to be higher in the past. In particular, temperature reconstructions were very unreliable before about 1880.

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.

27.3 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

\[
Y_t = \mu_t + \eta_t, \quad t \in T,
\]

FIGURE 27.1
(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 [90]; (b) A plot of the estimated standard errors for the mean by year.

27.2.2 Global temperature series

The Berkeley Earth project [84] uses isotropic geostatistical tools (kriging; see Chapter 5) to estimate the global mean temperature. The land data have been collected from 14 databases and 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 [96]. If a measurement device is moved or replaced, it is considered a different station, rather than being “corrected.” For a statistician, this seems to be a natural approach. The ocean data used in the Berkeley Earth series [83] come from the Hadley Center sea surface temperature data set HadSST3 [51, 52], modified by kriging of missing grid squares. This allows for a statistically justifiable estimate of global mean estimation uncertainty. Figure 27.2, formatted identically as Figure 27.1, demonstrates that the standard errors tend to be higher in the past. In particular, temperature reconstructions were very unreliable before about 1880.

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.

27.3 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

\[
Y_t = \mu_t + \eta_t, \quad t \in T,
\]
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., [8, 12, 77].)

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

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

Clearly an additive decomposition (27.1) 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 [50] 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, $\hat{\mu}_t$ say, will appear in the estimate of the errors $\hat{\eta}_t = y_t - \hat{\mu}_t$. 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 [99, 106]; (ii) using certain functions such as wavelets we can define the trend in terms of averages over certain temporal scales [9, 11, 19].

There are cases for which the trend tends to be less smooth. For example, letting $\mu_t$ be related to known covariates, plus a possibly smooth function of time $t$, may lead to a rougher trends. Introducing an abrupt change in $\mu_t$ (e.g., using the broken stick model presented later in this chapter) or using stochastic models for $\mu_t$ often also produces rougher trends.

### 27.3.1 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 \( \eta_t \) and \( \eta_{t+h} \) 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. For further details of stationary processes see Chapter 3.

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 identically 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. This process is stationary when the autoregressive parameter \( \phi \) satisfies \( \mid \phi \mid < 1 \). 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 \( q \):

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

Since a stationary ARMA model can be written as an infinite order AR model (that in practice would be 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 [4, 19, 21, 48, 56]. 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 [6], 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.

### 27.3.2 Linear trends

For some climate series it may be reasonable to assume that the mean level \( \mu_t \) 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

\[
\mu_t = \beta_0 + \beta_1 x_t, \quad t \in T.
\]
A naive estimator of the intercept parameter $\beta_0$ and slope parameter $\beta_1$ is given by ordinary least squares (OLS). For series of length $N$, the OLS estimates,

$$\arg\min_{\beta_0,\beta_1} \sum_{t=1}^{N} (y_t - \beta_0 - \beta_1 x_t)^2,$$  
(27.3)

are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{with} \quad \hat{\beta}_1 = \frac{s_y}{s_x},$$

where $\bar{y}$ is the sample mean of $\{y_t\}$, $\bar{x}$ is the sample mean of $\{x_t\}$, $s_y$ is the sample standard deviation of $\{y_t\}$, $s_x$ is the sample standard deviation 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 $\beta_0$ and $\beta_1$ (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

$$\arg\min_{\beta_0,\beta_1} \sum_{t=1}^{N} w_t (y_t - \beta_0 - \beta_1 x_t)^2,$$  
(27.4)

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 $\Sigma = \text{cov}(\eta)$ the covariance matrix for the noise, where $\eta$ is the column vector of the $\eta_t$, meaning that $\Sigma_{t,t'} = \text{cov}(\eta_t, \eta_{t'})$ for each $t$ and $t'$. Letting $y$ and $x$ be the column vectors of the $y_t$ and $x_t$, respectively, the general least squares (GLS) estimates,

$$\arg\min_{\beta_0,\beta_1} \langle y - \beta_0 1 - \beta_1 x \rangle^T \Sigma^{-1} \langle y - \beta_0 1 - \beta_1 x \rangle$$  
(27.5)

(where $1$ is a vector of ones), are the BLUEs of $\beta_0$ and $\beta_1$. OLS and WLS is a special case of GLS when $\Sigma$ is a scaling of the identity matrix, or a diagonal matrix with diagonal entries $\text{var}(\eta_t)$, respectively.

Usually $\Sigma$ 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,

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

we find candidate time series models. Using a given candidate model, we plug in the estimated $\Sigma$ from the model in (27.5). Alternatively the time series parameters can be part of the minimization process [39], 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 $\theta$ denote the parameters characterizing the
Modeling and assessing climatic trends

noise model, which drives \( \Sigma \). The likelihood function that needs to be maximized is

\[
L(\beta_0, \beta_1, \theta) = (2\pi)^{-N/2} [\det(\Psi)]^{-1/2} \exp\left\{ (y - \beta_0 1 - \beta_1 x)^T \Sigma^{-1} (y - \beta_0 1 - \beta_1 x) \right\}.
\]

It can be shown that the ML estimates of \( \beta_0 \) and \( \beta_1 \) are the GLS estimates substituting in the ML estimate of \( \theta \) into \( \Sigma \). Typically the ML estimate of \( \theta \) 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 \( \Sigma \), the vector of the GLS estimators \( \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T \) of \( \beta = (\beta_0, \beta_1)^T \) have a bivariate normal sampling distribution:

\[
\hat{\beta} \sim N_2\left( \beta, \left[ X^T \Sigma^{-1} X \right]^{-1} \right),
\]

where here the design matrix \( X \) is a \( N \times 2 \) matrix with first column all ones and the second column being \( x \). This distributional assumption assumes that we have the correct model for both the trend and the noise. In practice when doing statistical inference for \( \beta \), we plug in an estimate of \( \Sigma \). The sampling distribution of the trend estimate, \( \hat{\mu} = X \hat{\beta} \) of \( \mu = (\mu_1, \ldots, \mu_n)^T \) follows naturally [81]:

\[
\hat{\mu} \sim N_n(\mu, X \text{cov}(\hat{\beta}) X^T).
\]

The standard error for the fitted curve is then

\[
\text{se}(\hat{\mu}_t) = \sqrt{x_t^T \text{cov}(\hat{\beta}) x_t}, \quad t \in T,
\]

where \( x_t \) is the \( t \)th row of \( X \). We can also derive a marginal sampling distribution for \( \hat{\mu}_t \) based on the \( t \) distribution when \( \Sigma = \sigma^2 R \). In this case \( R \) is a positive definite matrix and \( \sigma^2 \) is some positive variance parameter that is estimated using OLS; multivariate inference follows from F-tests [81]. (The \( t \) and \( F \) distribution assumes that the parameters characterizing \( R \) are known.)

### 27.3.3 Nonlinear and nonparametric trends

Clearly there is no \textit{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 \textit{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 [81], 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 (27.6), the GLS estimator now has a \( (p+1) \)-variate normal sampling distribution for a Gaussian noise process, where the design matrix \( X \) has \( i \)th row

\[
x_i = (1, x_i, x_i^2, \ldots, x_i^p)^T, \quad i = 1, \ldots, N.
\]
Handbook of Environmental and Ecological Statistics

(27.7) and (27.8) 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 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 coefficients 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 27.4 below. There are several other criteria, such as BIC, AICc, etc. [61].

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 [8, 77], wavelets [9, 11, 19, 78], and splines [99, 101, 106]. Local regression methods based on kernels can also be used [20, 67]. Given the multitude of possible functions and methods that can be used to model trend, model selection and comparison becomes more problematic.

We again need to be wary of overfitting; 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 [16, 42, 43, 76].

Compared to the case of linear trends, uncertainty quantification for nonlinear and non-parametric trend estimates are typically more involved and more computationally intensive to calculate. Depending on the paradigm (Bayesian or frequentist), Markov chain Monte Carlo (MCMC) [13, 38], or Monte Carlo and other resampling methods are commonly used for assessing the uncertainty or significance of trend estimates [15, 19, 33, 57].

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) \( \tau \) 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 \( \tau \) can also be estimated in the model. If a smoother version is desirable, a bent cable model [18] connects the two linear parts with a quadratic connector. The model is then

\[
\mu_t = \beta_0 + \beta_1 t + \beta_2 \left[ \frac{(t - \tau + \gamma)^2}{4\gamma} I\{|t - \tau| \leq \gamma\} + (t - \tau)I\{|t > \tau + \gamma\} \right], \quad t \in T. \quad (27.9)
\]

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...
Modeling and assessing climatic trends

continuously. The model can be fit using least squares or via ML using the bentcableAR R package (https://cran.r-project.org/web/packages/bentcableAR/) – this package is able to estimate the changepoint and can calculate confidence intervals for the trend estimate. For a climate application, see [107].

### 27.3.4 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

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

(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: if \( q \) is too large the trend will be biased, while if \( q \) is too small the variability is too high). Any MA\((q)\) filter provides an unbiased estimate of a linear or locally-linear trend (Chapter 1 of [12]).

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,

\[
\hat{\mu}_t = \sum_{j=-q}^{q} a_j Y_{t-j},
\]

can unbiasedly estimate trends of a given degree of polynomial (e.g., Exercise 1.12 of [12]). For applications to climate series, see the review in [68].

Another common trend estimate is a cubic spline [89], which is a piecewise cubic function with differentiable connectors at a series of knots. A common choice of knot positions is to place them at equal quantiles of the x-variable, which is usually time for climate applications. The number of quantiles is called the degrees of freedom. Roughly speaking if there are \( N \) observations in the data set, there are \( N/\text{df} \) observations between each knot. Other common approaches in climate analyses use generalized additive models (GAM) ([47]; 2.SM.3) or singular spectrum analysis (SSA) [80].

With the additive decomposition for trend given by (27.1), 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 (See Chapter 3).

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. Looking at the so-called spectral properties of filters can indicate issues that may occur with smoothing and filtering; see, e.g., [77] for more details of the spectral approach.

### 27.3.5 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, \( \nabla \), 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 [12]). Formally for the series \( \{Y_t\} \), the differencing operator \( \nabla \) 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 (27.2): 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, \( \{\nabla \eta_t\} \).

### 27.3.6 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 [97], 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 that was defined in Chapter 3.

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 [32, 49]. 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 \( \mu_t \) is deterministic we assume a random walk model for \( \mu_t \):

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

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 [32, 64, 91]. Parameter estimation follows in the Gaussian process setting using ML. In the Bayesian context this is an example of a Dynamic Linear Model (DLM) ([45]; also see Chapter 4). This class of models can be related to the smoothing and spline methods discussed above [49, 65, 103].

### 27.4 Two case studies

We now return to the two data sets presented in Section 27.2. We will demonstrate that failure to investigate residual structure, or to make assumptions appropriate for the data, can yield incorrect conclusions. All our analyses were carried out using the R software package [76] – the R code is available from [http://www.stat.osu.edu/~pfc/](http://www.stat.osu.edu/~pfc/).

#### 27.4.1 US annual temperatures

The standard approach in the climate literature is to fit a linear trend to a time series using ordinary least squares (27.3). Doing that for the US annual temperature anomalies yields an estimate of 0.55°C per century (see Table 27.1 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
TABLE 27.1

Fits of linear trend slopes to US annual average temperature anomalies

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

FIGURE 27.3

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

least squares fit (27.4), with weights inversely proportional to the variance estimates. The resulting estimate is somewhat smaller (0.48°C per century) and somewhat more uncertain, although still highly significant (again assuming the noise is independent). We also remark that any estimate can be sensitive to choice of time range for the data (in this case, the years 1897–2008).

Looking at the sample autocorrelation function (ACF) of the residuals from the two least squares fits in Figure 27.3, 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 27.1, showing the Ljung-Box [66] P-value when testing that the residuals are IID (based on 10 lags 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 (27.5). In essence, [90] 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. Restricting to ARMA($p$, $q$) models with $p \leq 10$ and $q \leq 10$, the best ARMA-type error structure is ARMA(3,1), which also yield the largest slopes. The weighted ARMA(3,1)-slope is not significantly different from zero. Figure 27.4 shows three of the fitted lines. All fits from these models pass the Ljung-Box IID noise test.

Table 27.1, 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.

### 27.4.2 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 27.5) indicate that a quadratic fit might be better.

Using a similar approach to that in Section 27.4.1, 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 27.6 shows the fitted model. Again, the residuals pass various tests for IID noise. We calculate the standard error of the fitted curve from (27.8). 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...
**FIGURE 27.5**
(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 [84]. The trend line correspond to ordinary least squares.
(b) Residuals from the ordinary least squares fit.

**FIGURE 27.6**
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.
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 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. 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 [71], but an analysis of that series yields nearly identical results to our analysis in this section. In the IPCC report there is a suggestion to use a GAM model with AR(1) error. We have tried that fit, as well as an SSA fit and a cubic spline with 8 degrees of freedom (corresponding to about 20 observations between each knot). They are all very similar. In Figure 27.6 we show the spline fit, together with a spline with 5 degrees of freedom (about 32 observations between each knot). The reason for the latter choice of degrees of freedom is that 30 years is the standard amount of time used to estimate a changing climate [41]. We see that the smoother of the spline fits is quite close to the quadratic fit.

27.5 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., [28, 37, 60]; see [19] 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 \( \{Z(s) : s \in D \subset \mathbb{R}^p \} \), based on \( n \) observations from the process, \( Z = (z(s_1), \ldots, z(s_n))^T \). Assuming an additive decomposition for this process \( \{Z(s)\} \) we have

\[
Z(s) = \mu(s) + \eta(s), \quad s \in D,
\]

where \( \{\mu(s) : s \in D\} \) is the spatial trend component and \( \{\eta(s) : s \in D\} \) 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:

\[
\mu(s) = \sum_{j=0}^{p} \beta_j x_j(s),
\]

where \( \{x_j(s) : s \in D\} \) are \( p + 1 \) spatial basis functions. The simplest case is often to include latitude and longitude as covariates. But traditionally, simple polynomials [28, 37], splines [27, 73, 87, 105], and other functions such as wavelets are used to model the trend in spatial settings [2, 27, 72]. 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 [54, 73], we could also assume that the trend component \( \{\mu(s) :
Modeling and assessing climatic trends

$s \in D$ 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 [86]).

These ideas extend to modeling spatial trends for areal processes [7, 37], point processes (see e.g., [3, 10, 31, 58, 59] for temporal and spatial examples), and to river networks [74]. Spatio-temporal methods follow naturally [24, 28, 63, 75, 85, 100, 104].

27.6 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 [31], 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 [24], trends in the quantiles [17, 62, 82], 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 [14, 25, 35, 55], paleoclimate proxies of temperature [64], and precipitation [22, 30, 88]. 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 [30] for a comprehensive review of the issues underlying the modeling of spatial extremes.

27.7 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.

Our modeling setup indicates that the measurement errors modulate a stationary, zero mean error process, i.e., that the natural variability does not change over time. This may be a strong assumption, but there is little evidence of changing climate variability in the literature.

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., [56]). [84] demonstrated that a linear trend, after accounting for long memory errors, was statistically significant when estimated from the [44] global temperature series. (For other examples see [19, 24, 63, 94]; see [21] 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 [29, 34]. We also note the popularity of principal components (also known as empirical orthogonal components; see, e.g., [6, 95, 98]) as trend components in atmospheric and oceanic sciences. In addition, the dynamic linear models approach of [45] has been applied to power projection in Brazil by [64].

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 [53]. 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 [23, 40].

Bibliography


Modeling and assessing climatic trends


Modeling and assessing climatic trends


Modeling and assessing climatic trends


