Trend and Derivative Estimations using GAMMs: Estimating daily temperature trends in Montana

Rory G. Peck

Department of Mathematical Sciences

Montana State University

May 7, 2010

A writing project submitted in partial fulfillment

of the requirements for the degree

Master of Science in Statistics

APPROVAL

of a writing project submitted by

Rory G. Peck

This writing project has been read by the writing project advisor and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

5/7/2010   Mark Greenwood

Date

Mark C. Greenwood

Writing Project Coordinator and Advisor

Contents

Abstract

Climate change research has gained much attention recently and has been used to implement policies anywhere from the local to international level. Much of the research that has currently been done has used averaging of temperatures across many sites to gain a univariate time series that is easier to analyze. Methods such as linear regression are used on these data sets to search for a linear trend. Because averaging over many sites removes variability from the measurements and the trend in temperature data may not be linear over time, we propose using a generalized additive mixed model (GAMM) to analyze daily temperatures. Using this we can detect a nonlinear trend if it is present while also accounting for spatial and temporal correlations within and among different weather stations. We assess the coverage properties of confidence bands around the trends that are estimated, along with its derivative and confidence bands to understand how fast any rate of change is occurring. We then apply our methods to over 100 years of daily average temperatures from 9 weather stations in Montana, finding evidence for a nonlinear trend in the daily midrange temperature that is increasingly positive between 1945 and 1980 but may be leveling off between 1980 and 2008.

1 Introduction

Climate change research is concerned with estimating the "average of weather over time and space" (Gutro, 2008). This immediately poses new questions: (1) Over what spatial and/or temporal scale should the weather be averaged to provide useful inference to climate change? (2) In that process of averaging, is the variability in observations being propagated accurately? And (3), has the rate of change been constant over time? We propose methods that use a nonparametric mixed model to estimate the trend, its precision, the derivative of that trend and its precision. We illustrate the techniques using 100 year daily temperature records from a network of weather stations in Montana.

Many studies on temperature trends are based on a weighted average either over space or time, or both. Others rely on dimension reduction techniques such as Principal Component Analysis. These methods are done to produce a single series that can be analyzed and explored. Global or hemispheric average temperature time series are an example of these methods that have been used in previous studies (e.g., Jones & Moberg, 2003). Often smoothing lines are added to descriptions of these trends along with linear trends. These averages over space and/or time often provide strong visual evidence of change, but it is unclear how much variability has been lost by averaging from daily local values to the monthly or yearly hemispheric or global scales. This also assumes that the change is occurring similarly at the global or hemispheric scale, which is interesting for global scale policy questions but does not allow for or detect spatial variability in the impacts on the local climate. At the other extreme are analyses that use site specific records, looking for significant trends at each site, and then try to aggregate local results to provide more general conclusions (Stewart et al., 2005).

Climate change researchers are concerned with two aspects of the changes, both detecting that it has occurred and then exploring the drivers of that change. We only focus on the detection and estimation of the change in this work, leaving the attribution of that change for future research.

Combining information spatially and/or temporally can be performed to get an overall estimated average by the linear combination of the observations, possibly also including weights based on the locations and correlations between the observations. It is always true that the average of observations is less variable than the original observations, which makes a direct analysis of the aggregated means a potentially dangerous proposition. To be more specific, two averaged series could have the same mean trajectory and very different underlying variances. If the means are analyzed without incorporating the variability of the estimates, both series could be judged to contain similar evidence for a trend. But in reality, maybe only one series should provide strong evidence of a trend relative to the variability and the other is not statistically distinguishable from being a spurious result when the variability is incorporated.

To accurately propagate the variability from the initial observations that are taken on a daily scale to higher level multi-decadal estimates, we develop a generalized additive mixed model (Wood, 2006) to estimate the long term trend, seasonal effects, and account for both spatial and temporal correlations using multiple crossed random effects. It is only possible to estimate these models using sparse matrix implementations as the full data set contains over 335,923 daily observations. Using the estimated model, the first derivative of the trend is estimated to explore the rate of change in the trend.

Before we consider the application to the analysis of daily temperature data, we describe conventional generalized additive models (GAMs) and compare the use of Bayesian or Frequentist-based confidence intervals for the nonparametric model components. Then derivative estimation, as well as the associated confidence interval estimation, from GAMs is discussed. GAMMs are an extension of the conventional models that incorporate random effects into the models and also use a different method to determine the smoothness of the nonparametric effects. Simulation studies are used to understand these methods before considering our application to climate change modeling. The final sections describe a mixed-model for modeling daily temperature data and apply it to temperature records at a selection of sites in Montana, providing estimates of the trend, its derivative, and their respective variability.

2 Generalized Additive Models

2.1 Definition of Generalized Additive Models

Generalized additive models (GAMs) (Wood, 2006; Hastie & Tibshirani, 1990) are extensions of linear models with additional predictors that involve the sum of smooth functions of the covariates. The typical linear model is defined as

$$y_i = \mathbf{X}\boldsymbol{\beta}_* + \varepsilon_i \tag{1}$$

where $\mathbf{X}$ is a design matrix of observed covariates, $\boldsymbol{\beta}_*$ is its vector of coefficients, and ε_i is a vector of errors that follow a $N(0, \sigma^2)$ distribution. GAMs are similar to typical linear models, but where at least one smooth function of the covariates added. A smooth function is represented as $s(x)$ in the model $y_i = \mathbf{X}\boldsymbol{\beta}_* + s(x) + \varepsilon_i$. The smoothing term $s(x)$ is constructed as

$$s(x) = \sum_{i=1}^{k} \mathrm{B}_i(x)\beta_i, \tag{2}$$

where B is a matrix of B-spline basis functions and $\mathrm{B}_i(x)$ is the value of the i^{th} basis function evaluated across the values of x, providing the i^{th} column of B. $\boldsymbol{\beta}$ is the vector of spline coefficients

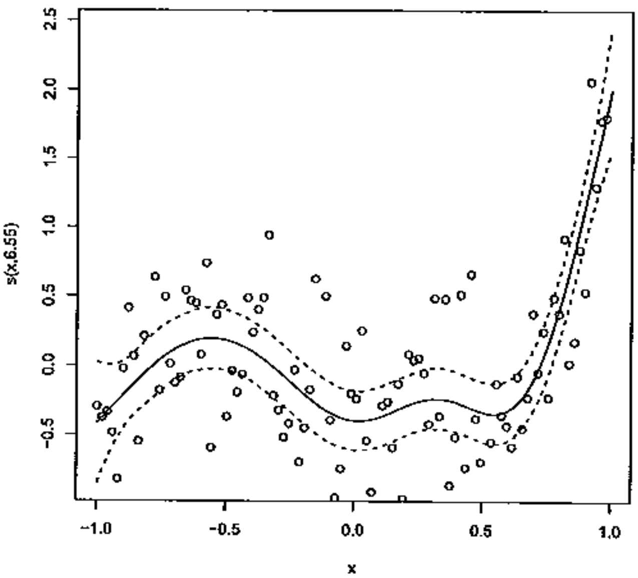

Figure 1: Example of a GAM of the form $y_i = s(x) + \varepsilon_i$ fit to 100 randomly generated points. The solid line is $s(x)$ and the dashed lines are 95% Bayesian confidence bands that are discussed in sections 2.3 and 3.2.

and is estimated as

$$\hat{\beta} = (\mathbf{B}^{\mathsf{T}}\mathbf{W}\mathbf{B} + \mathbf{S})^{-1}\mathbf{B}^{\mathsf{T}}\mathbf{W}\mathbf{y} \tag{3}$$

where $\mathbf{W}^{-1}\phi$ is the covariance matrix of the data, ϕ is the scale parameter, $\mathbf{y}$ is a vector of the responses, and $\mathbf{S}$ is chosen so that

$$\int_{x_i}^{x_k} [f''(x)]^2 dx = \beta^{\mathsf{T}}\mathbf{S}\beta. \tag{4}$$

This choice of $\mathbf{S}$ corresponds to working with penalized cubic regression splines. A smoothing parameter λ is then selected to minimize

$$||y - \hat{y}|| + \lambda \mathbf{B}^{\mathsf{T}}\mathbf{S}\mathbf{B} \tag{5}$$

by cross-validation or generalized cross validation. The choice of λ controls how wiggly the $s(x)$ can be.

2.2 Generalized Additive Mixed Models

Generalized additive mixed models (GAMMs) are GAMs with additional random effects included in the model. In GAMMs, the β vector, and thus the amount of smoothness in $s(x)$, is estimated based on a specific parameterization using random effects (Wood, 2006). The specific implementation used to estimate these models is found in gamm4 (Wood, 2009) which is based on the methods in amer (Scheipl, 2009).

An advantage to gamm4 is that it uses lme4 to estimate the mixed models, which has two related advantages for use here: first, it is implemented using sparse matrices, making large mixed models estimable, and second, it is optimized for crossed random effects. Even with these advanced, optimized techniques for estimating large mixed models, it is not possible to estimate these models in a typical Windows version of R due to the large matrices that must be built and constraints in memory use in 32-bit versions of R. To be able to fit these models on a Windows machine, we use REvolution Enterprise 3.1 (REvolution Computing, 2010) with a 64-bit version of Windows to be able to use all the physical memory on the computer.

Random effects in these models can be interpreted just as random effects in linear mixed models. Crossed or nested effects can also be included. The GAM part of the model can also be interpreted just as it would be in a GAM model from section 2.1. This will be helpful in our study when accounting for both spatial correlations among the 9 sites and temporal correlations over the 100+ years of observations, which are described in more detail in section 4.1.

2.3 Calculation of Confidence Bands

Confidence bands can be obtained for the smooth functions in GAMs to estimate the uncertainty of the estimates for each value of x in $s(x)$. A pointwise $C\%$ confidence band is obtained by calculating $C\%$ confidence intervals of the response for each value of $x \in [\min x, \max x]$. The standard errors for each point, SE_x, are the square roots of the diagonal values on the matrix $\mathbf{B}^\mathsf{T}\mathbf{V}\mathbf{B}$ where two different definitions of $\mathbf{V}$ are given below. The approximate $C\%$ confidence interval for $s(x)$ for any x is $\widehat{s(x)} \pm z_{\alpha/2}\mathrm{SE}_x$.

To obtain the intervals, a covariance matrix, $\mathbf{V}$, for β is required. The frequentist covariance matrix $\mathbf{V}_e$ can be obtained from a normal approximation. Recalling how $\hat{\beta}$ was estimated in equation 3, we get

$$\mathbf{V}_e = (\mathbf{B}^\mathsf{T}\mathbf{W}\mathbf{B} + \mathbf{S})^{-1}\mathbf{B}^\mathsf{T}\mathbf{W}\mathbf{B}(\mathbf{B}^\mathsf{T}\mathbf{W}\mathbf{B} + \mathbf{S})^{-1}\phi. \tag{6}$$

Wood (2006) shows that from the normality of y or the large sample multivarite normality of $\mathbf{B}^\mathsf{T}\mathbf{W}\mathbf{y}$ that

$$\hat{\beta} \sim N(\mathrm{E}(\hat{\beta}), \mathbf{V}_e), \tag{7}$$

approximately. However, $\mathrm{E}(\hat{\beta})$ usually will not equal β, resulting in biased estimates.

Nychka (1988) explains that the standard error bands will have less coverage than reported using this frequentist method. He also notes that the coverage is even worse at sharp curves in the function. He proposes a Bayesian approach to produce more accurate intervals. A prior distribution of the form

$$f_\beta(\beta) \propto e^{-\frac{1}{2}\beta^\mathsf{T}\Sigma[\mathbf{S}_i/\tau_i]\beta} \tag{8}$$

is placed on the parameter vector where the τ_i's are chosen to control the dispersion of the prior. Wood (2006) gives more detail on how these are chosen. This results in a posterior distribution of the form

$$\beta \sim N(\hat{\beta}, \mathbf{V}_\beta) \tag{9}$$

where $\mathbf{V}_\beta$ is the posterior covariance matrix of the parameters and is computed as

$$\mathbf{V}_\beta = (\mathbf{B}^\mathsf{T}\mathbf{W}\mathbf{B} + \mathbf{S})^{-1}\phi. \tag{10}$$

Replacing $\mathbf{V}_e$ with $\mathbf{V}_\beta$ and calculating new SE_x's provides approximate $C\%$ Bayesian confidence intervals.

Results from simulations are summarized in section 3.2 and show that the Bayesian approach may be giving more conservative results than reported and the frequentist approach does provide slight undercoverage of the true values, with these results applying to both GAM and GAMM models.

3 GAM Derivatives

3.1 Estimation of Derivatives

The derivatives of the smooth functions fit by GAMs can be understood by using a simple polynomial as an example. The smooth functions are built by multiplying the basis functions by $\boldsymbol{\beta}$. Just as the derivative of a constant times a polynomial is the constant times the derivative of the polynomial, the derivative of the smooth function is the sum of the derivatives of the basis functions times $\boldsymbol{\beta}$. This is discussed in further detail in Ruppert et al. (2003). Derivatives can be calculated one of two ways: analytically based on the underlying polynomial in spline bases or numerically using the definition of the derivative. The first method requires modifications based on the specific bases used while the second method can be applied more generally without modifications.

Derivatives of GAM's can be estimated by building a matrix of approximate derivatives of the basis functions, using

$$\mathbf{B}_i'(x) = \frac{\mathbf{B}_i(x+\epsilon) - \mathbf{B}_i(x)}{\epsilon}, \tag{11}$$

for some small ϵ. The matrix $\mathbf{B}'$ can then be multiplied by $\hat{\beta}$ to obtain an estimated first derivative of the estimated smooth function. Frequentist and Bayesian covariance matrices can be calculated for the derivative just like $\mathbf{V}_e$ and $\mathbf{V}_\beta$ were in 2.3 by replacing $\mathbf{B}$ with $\mathbf{B}'$. With these covariance matrices, confidence bands for the first derivative can then be estimated. Simulations show that the Bayesian estimates provide better coverage of the true derivative for most x values as seen in figure 3.

3.2 Simulation Study

We performed a simulation to verify the results of Nychka (1988) and Wahba (1983) and check the coverage of the confidence bands for both the function and its derivative. 100 equally spaced x's were generated on $[-1, 1]$ and evaluated in the function $f(x) = x^4 + 2x^3 - x$.

Two studies were performed, the first with a GAM of the form $y_i = s(x) + \varepsilon_i$ where $\varepsilon_i \sim N(0, 0.25)$. The second study was run with a GAMM of the form $y_{ij} = s(x) + \zeta_i + \varepsilon_{ij}$ where $i = 1, \ldots, 10$, $\zeta_i \sim N(0, 1)$, and $\varepsilon_{ij} \sim N(0, 0.25)$.

Both simulations generated 10,000 different data sets and the associated GAM or GAMM model was fit each time. The proportion of times the true function value fell within the estimated 95% confidence band was recorded for each simulation for both the Bayesian and frequentist variance estimates.

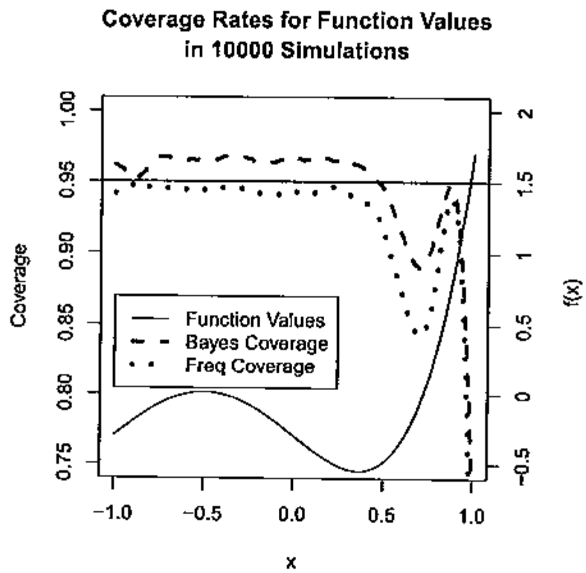

Figure 2: Plot of the true function $f(x)$ along with the coverage rates for the GAM model in 10,000 simulations.

The results from the GAM simulation are summarized in figure 2 and show that the Bayesian coverage is generally more conservative than reported. The frequentist bands only slightly undercover the true curve for most x values (93%-94% coverage). Both methods begin to fail as the function begins to rise faster on the right side of the function.

Coverage of the derivative for the GAM has similar results to the function coverage, with The results shown in figure 3. The frequentist coverage is slightly more liberal and the Bayesian coverage is slightly more conservative than reported. Again, both methods fail on the steepest part of the function. The frequentist coverage of the derivative began to improve on a section of the steepest part of the curve after initially failing and performed better than the Bayesian coverage in a small interval of the function ($x \sim [.8, .9]$). Similar results were found for the GAMM simulation but are not reported here.

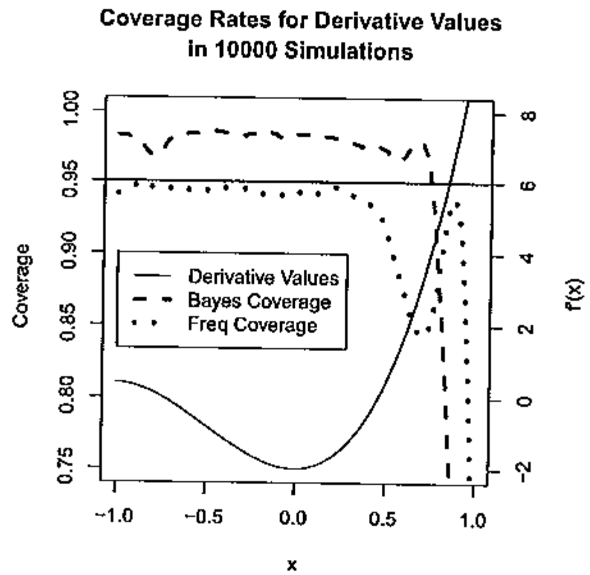

Figure 3: Plot of the true derivative $f'(x)$ along with the coverage rates for the GAM model in 10,000 simulations.

4 Application

4.1 Data

The data used for our study came from 9 weather stations in Montana, each with over 100 years of temperatures recorded daily in degrees Fahrenheit. These stations were chosen out of a potential 21 that were located along the Rocky Mountains in Montana that are available in the USHCN (Menne et al., 2008). The subset of stations was chosen to maximize the length and the completeness of the records to avoid as much bias as possible in the response. The length and completeness of each station's records is summarized in table 1. Due to types of bias that are unaccounted for, such as station moves and localized weather fluctuations, it may not be appropriate to interpret an individual site record as evidence for overall climate change. A model that will average over all the sites will help account for any bias that may be introduced by the station being moved during the

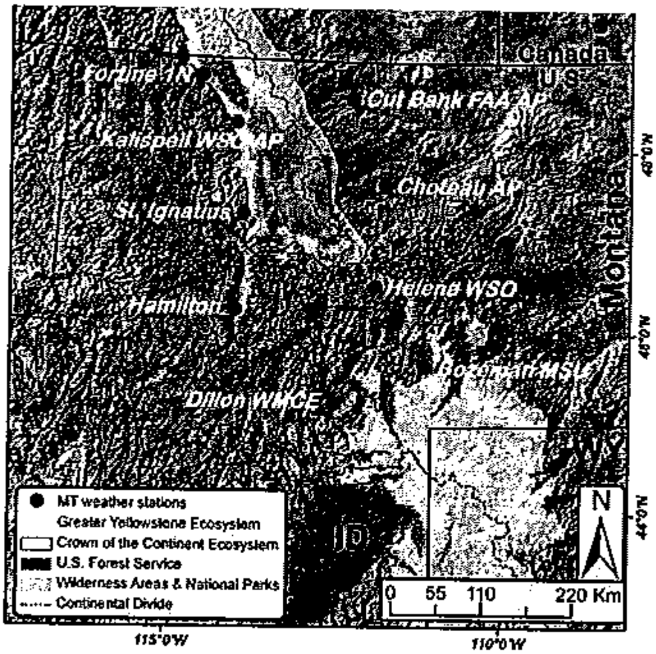

Figure 4: Locations of the 9 sites in Montana.

record, urban heat island effects, or other transient impacts on each record.

Using the daily average temperature from these locations, we develop a model in the next section to estimate the trend and account for other important sources of variation in the records. Since these results are based on the same observations as considered in Pederson et al. (2010) with 4 additional years included, we can directly compare our results to those based on averaging first and analyzing second.

4.2 Model

Our model for the daily average temperature, y_{ijk}, at location $i = 1, \ldots, 9$, month $j = 1, \ldots, 1308$, and day $k = 1, \ldots, 39812$ has the form

$$y_{ijk} = \alpha + s(time) + s_c(time) + \zeta_i + \gamma_j + \omega_k + \epsilon_{ijk}, \tag{12}$$

Table 1: Summary information from the 9 weather stations used in the study along with length and completeness of their records.

Station	Elevation (meters)	Latitude	Longitude	Completeness of Record in % from 1900-2008
Bozeman MSU	1497.5	45°40′N	111°03′W	99.5
Choteau AP	1172.0	47°49′N	112°12′W	83.6
Cut Bank FAA AP	1169.8	48°36′N	112°23′W	90.6
Dillon WMCE	1593.5	45°13′N	112°39′W	99.2
Fortine 1N	914.4	48°47′N	114°54′W	90.4
Hamilton	1080.5	46°15′N	114°10′W	90.9
Helena WSO	1166.8	46°36′N	111°58′W	99.9
Kalispell WSO AP	901.3	48°18′N	114°16′W	99.4
Saint Ignatius	883.9	47°19′N	114°06′W	90.2

where ζ_i is a random day effect, γ_j is a random month effect, ω_k is a random site effect, and ϵ_{ijk} is a random error. All four are are i.i.d. normal with their own variance estimates. All random effects are modeled as crossed effects. Crossing of site and month or day random effects induces a crude spatial-temporal correlation as locations within a site are correlated and observations for the same month or day are also correlated. This idea was developed in VanLeeuwen et al. (1996) as a way of accounting for the two directions of correlations in situations with multiple sites being modeled over time. The day random effect could be nested within each month but this is actually more computationally expensive to define a unique level of day within each month as opposed to a collection of 39,812 days. Treating these effects as crossed induces correlations between all the observations across the sites within a given month based on the day random effect and adds additional correlation to the observations that were observed on the same day (which is in the same month, of course). More details on the induced correlations for different "distances" between observations in time and space are found below, but this set of random effects induces two levels of temporal correlation (month and day) and one level of spatial correlation (at the same site or not). The connections to more refined spatial-temporal correlation structures that dampen out as a function of space and time can be easily seen. For large data problems, the luxury of these refined correlation structures is not available, especially while attempting to perform nonparametric trend estimation that also involves the use of random effects.

4.3 Results

There is a significant increasing trend in these data over the time period studied (edf = 6.96, p-value $< 10^{-16}$). The edf are the effective degrees of freedom and refer to the complexity of $s(x)$ and are equivalent to the number of independent basis functions required to describe the smooth function (although it is based on the specific $\hat{\beta}_i$). We allowed the estimated trend to use up to 60

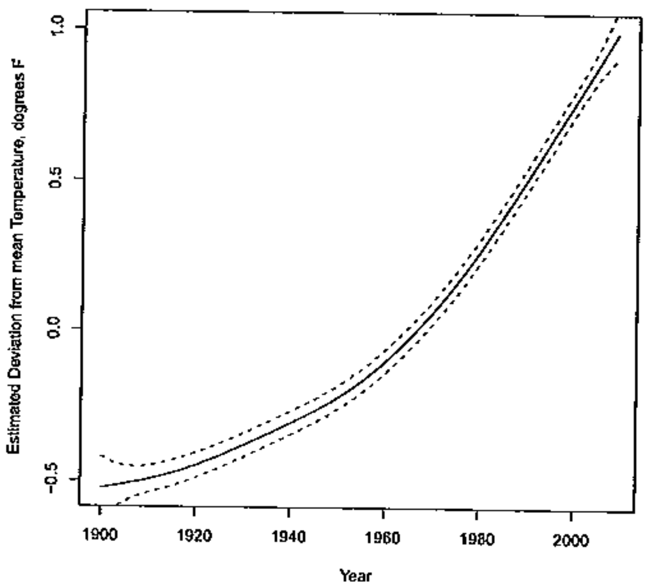

Figure 5: Estimated trend from the model using all 9 sites.

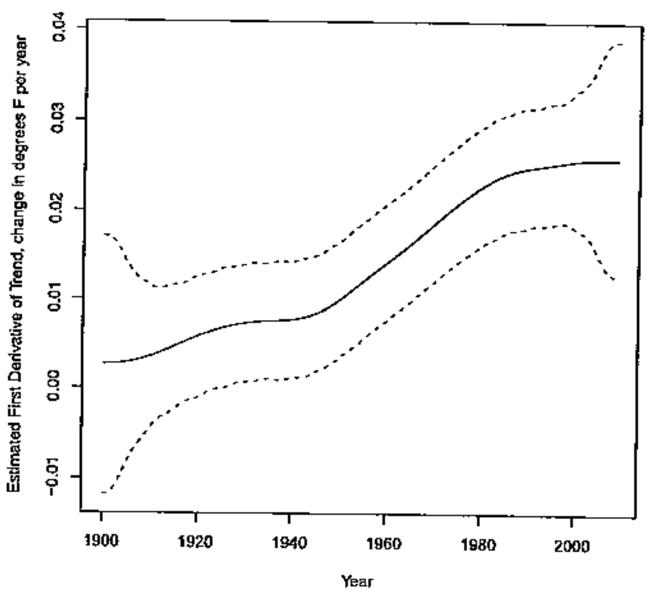

Figure 6: Estimated derivative of the trend from the model using all 9 sites.

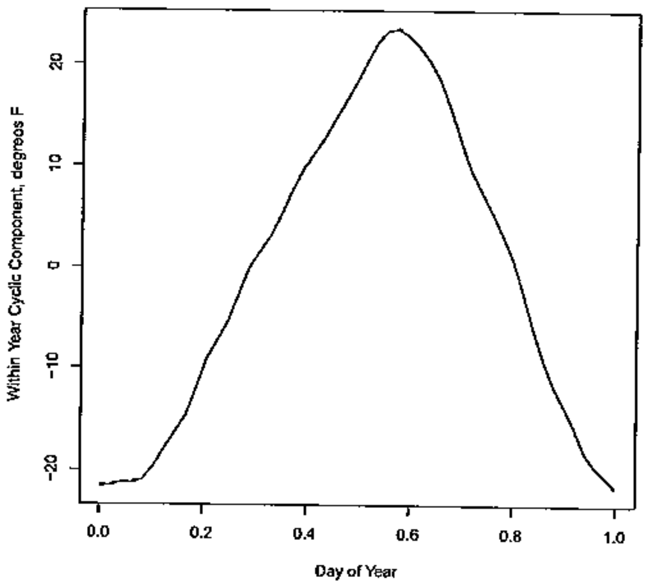

Figure 7: Estimated within year trend from the model using all 9 sites.

edf and it estimated the edf as 6.96. The maximum number of basis functions that can be used is the number of observations in the data. An initial edf of 60 provides a knot more than every other year (59 knots for 108 years) and enough flexibility to capture any slow variation in the trend.

The estimated derivative of the trend in figure 6 began fairly level, then saw a sharp increase, and now appears to be leveling off a bit again. The steepest increase occurred between 1945 and 1980. The derivative has significantly risen above zero over time as well.

The trend was significant (p-value < 0.0001) and shows an increase of about 1.5°F over the 109 year span of our data. The derivative along with the trend shows that this increase may also be occurring at a faster rate than previously. The results should be used with caution however. As seen in section 3.2, the confidence bands for values at the edge and values on a steeper part of the curve are unreliable. This model cannot and should not be used for any forecasting methods.

By including random effects for day, month, and site, all crossed with one another, a crude spatial temporal correlation is induced. There are 2 levels of temporal correlation that come from this (within day and within month) as well as the spatial correlation that comes from the site effect. The estimated variances for each effect are shown in table 2. The random effects induce a correlation of 0.0263 for any two observations from the same site that are in different months. For observations in the same site and same month, the correlation is 0.1768. For observations in different sites and different months, the induced correlation is 0, but for observations in the same month but different sites, the correlation is 0.1505. For observations on the same day but different sites, the correlation is 0.7321.

Two unusual residuals were observed after the model was initially fit. One site recorded a maximum temperature of 8°F in September. Temperatures this low do occasionally occur in Montana at this time of year, however the other 8 sites had maximum temperatures around 60° for the same day. That particular site had missing values before and after the 8°F was observed. Another site recorded a daily average in July as 114° with a maximum temperature that day of 182°F. These observations were not removed by the quality control procedures used in USHCN but are clearly problematic records identified by the residuals from our model. Both observations were removed and the model was refit. Doing so had practically no effect on the results due to the large sample size. This gives evidence that models such as the one we fit can be used as quality control for data sets of similar sizes in detecting data entry errors which are especially common in large data sets.

Table 2: Estimates of variances of random components from the model.

Variance	Estimate
$\hat{\sigma}^2_{Day}$	53.769
$\hat{\sigma}^2_{Month}$	13.914
$\hat{\sigma}^2_{Site}$	2.427
$\hat{\sigma}^2_{e}$	22.337

5 Discussion

5.1 Comparison to Other Methods

The data used in this study was the average of the daily minimum and maximum temperatures that were used in Pederson et al. (2010). Using linear regression they found a significant trend in both the daily minimum and maximum temperatures separately (both p-values were 0.000). The overall change was 1.13°C for the maximums and 1.55°C for the minimums in the 20^{th} century (about 2-2.75°F). While we found similar results for the overall change, using the GAMM model we found significant evidence for a non-linear trend for average temperatures in the 20^{th} century (p-value $< 2 * 10^{-16}$).

Our method of not averaging across sites gave us a larger data set in which we could more precisely estimate a trend with GAMMs. Accounting for the correlation between measurements both spatially and temporally allowed us to find the trend that showed a change of 1.5°F in a data set with over 300,000 observations that had a range of over 100°F.

5.2 Use of GAMMs

The model we fit pushed the boundaries of what gamm can do in R. Even using REvolution on a 64-bit version of Windows, the model took an hour to fit and used nearly 8 gigabytes of memory. Looking at longer time periods and more sites in a model such as the one we fit may not be possible with the current limitations of computing. The model still could be expanded to a larger spatial scale by grouping similar sites together and running separate analyses.

6 Future Research

6.1 Simulation of Confidence Bands

Our simulation study on the confidence bands for GAM smoothers only focused on one function that was similar to the trend we observed. It may be of interest to see if the results are similar for other functions.

6.2 Choice of Response Variable

The modeling of different responses may yield more interesting results. We used the midrange of the daily temperatures due to data collection limitations in early years of the data set. While this is an easy value to obtain and work with, it is highly sensitive to outliers because it is the average of only 2 values, the daily minimum and maximum. It is only useful when summarizing symmetric, platykurtic distributions. If the daily temperatures at these locations follow a skewed distribution, modeling other response variables would be required to obtain a better understanding of the temperature trends. Modeling the minimum, maximum, and range will be pursued in future research to avoid this issue.

6.3 Estimation of Derivatives

It is unclear how precise the numerical estimation of the derivative of the smoother is. Different choices of ϵ in equation 11 could give different results for the same data set. We used $\epsilon = 10^{-7}$ from Wood (2006). There no discussion on how this value was chosen or how much of an effect different values would have.

References

[1] Gutro, R. (2008), http://www.nasa.gov/mission_pages/noaa-n/climate/climate_weather.html

[2] Hastie T.J., Tibshirani R.J. (1990), Generalized Additive Models, Chapman and Hall/CRC, Boca Raton, FL

[3] Jones, P.D., Moberg, A. (2003), Hemispheric and Large-Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001, Journal of Climate, Vol 16, 206-223.

[4] Menne, M.J., Williams, Jr., C.N., and Vose R.S. (2008), United States Historical Climatology Network, Long-Term Daily and Monthly Climate Records from Stations Across the Contiguous United States. http://cdiac.ornl.gov/epubs/ndp/ushcn/access.html

[5] Nychka, D. (1988), Bayesian confidence intervals for smoothing splines, Journal of the American Statistical Association, 83.404, 1134+

[6] Pederson, G.T., et al. (2010), What do recent temperature trends portend? A view from western Montana, USA. Climatic Change 96: DOI 10.1007/s10584-009-9642-y, 22pp.

[7] Ruppert, D., et al. (2003) Semiparametric Regression, Cambridge University Press, New York, NY.

[8] Scheipl, F. (2009), amer: Additive mixed models with lme4. `http://CRAN.R-project.org/package=amer`

[9] Stewart, I.T., et al. (2005), Changes toward earlier streamflow timing across western North America, *Journal of Climate*, 18(8), 1136-1155.

[10] VanLeeuwen, D.M., Leigh, W.M., Urquhart, N.S. (1996), A Mixed Model with Both Fixed and Random Trend Components across Time, Journal of Agricultural, Biological, and Environmental Statistics, Vol. 1, No. 4, pp. 435-453.

[11] Wahba, G. (1983), Bayesian confidence intervals for the cross validated smoothing spline, Journal of the Royal Statistical Society, Series B 45, 133-150.

[12] Wood, S. (2006), Generalized Additive Models: An Introduction with R, Chapman and Hall/CRC, Boca Raton, FL.

[13] Wood, S. (2009), gamm4: Generalized additive mixed models using mgcv and lme4. R package version 0.0-2. `http://CRAN.R-project.org/package=gamm4`

[14] Zuur A.F., et al. (2009) *Mixed Effects Models and Extensions in Ecology with R.* Springer, New York.