INTERPOLATION OF NONSTATIONARY HIGH FREQUENCY SPATIAL–TEMPORAL TEMPERATURE DATA[1]

BY JOSEPH GUINNESS AND MICHAEL L. STEIN

University of Chicago

The Atmospheric Radiation Measurement program is a U.S. Department of Energy project that collects meteorological observations at several locations around the world in order to study how weather processes affect global climate change. As one of its initiatives, it operates a set of fixed but irregularly-spaced monitoring facilities in the Southern Great Plains region of the U.S. We describe methods for interpolating temperature records from these fixed facilities to locations at which no observations were made, which can be useful when values are required on a spatial grid. We interpolate by conditionally simulating from a fitted nonstationary Gaussian process model that accounts for the time-varying statistical characteristics of the temperatures, as well as the dependence on solar radiation. The model is fit by maximizing an approximate likelihood, and the conditional simulations result in well-calibrated confidence intervals for the predicted temperatures. We also describe methods for handling spatial–temporal jumps in the data to interpolate a slow-moving cold front.

1. Introduction. When analyzing surface meteorological data, we often encounter observations that are recorded at regular and frequent intervals in time but at irregular and sparse locations in space. The data usually come in the form of multiple time series, in which a small number of lengthy time series are associated with fixed locations in space. The methods developed in Stein (2005, 2009) exploit this structure in constructing computationally efficient likelihood approximations for data that may be modeled as a realization of a stationary spatial–temporal Gaussian process or a very constrained type of nonstationary Gaussian process. To model the temperature data that we consider in this paper, a more flexible nonstationary time series model is needed to capture accurately the statistical properties of the data, especially the dependence on solar radiation and time of day. The time series model that we employ relies on the idea of the evolutionary spectrum, as introduced by Priestley (1965) and advanced theoretically by Dahlhaus (1996, 2012). Through a simple modification of the nonseparable spatial–temporal covariance function in Stein (2005, 2009), we incorporate evolutionary spectra to introduce nonstationary behavior in the time domain of the spatial–temporal process. Using

the computational techniques described in Guinness and Stein (2013), we are able to efficiently compute and maximize approximate Gaussian likelihoods for these data, which comprise 648,000 temperature observations. We also develop specialized methods for modeling and interpolating spatial–temporal jumps in the data.

The goal of this work is to develop methods for producing accurate space–time interpolations of temperature data at sites for which no observations are available. This objective has a long history in the meteorological literature, and many researchers have made important contributions, including Daly, Neilson and Phillips (1994), who interpolate precipitation data in the U.S., and Hijmans et al. (2005), who interpolate several climate variables at a high resolution with global land coverage. While we have similar goals, our methods differ in that we characterize the entire spatial–temporal statistical distribution of the temperature process so that we may naturally incorporate uncertainty into the interpolations. Specifically, we use the fitted statistical model to produce a suite of simulated temperature data at unobserved locations conditional on the observed temperature data, accounting for the various uncertainties associated with the fitted model. Such simulations are similar in spirit to the idea of multiple imputations [Rubin (1987)] and have been referred to as data ensembles [Schneider (2006), Stein (2009)]. In the geostatistics literature, this approach is usually called geostatistical simulation [Lantuéjoul (2002), e.g.]. Because our statistical model tries to reflect the nonstationary nature of the temperature records, the resulting conditional simulations closely resemble the observed temperature data; not only are the simulations accurate, but the variance and the correlation structure are consistent with the observations. Such simulations may be useful in any application that requires a measure of uncertainty in addition to a point interpolation of temperature. For example, a high resolution regional climate model may require meteorological fields on a spatial grid as inputs, and the simulations provide a way to propagate the uncertainty of the interpolations through the climate model.

2. Atmospheric radiation measurement program data. The data are provided by the Atmospheric Radiation Measurement (ARM) Program, which was established by the U.S. Department of Energy in 1989 to study how weather processes, especially cloud formation, affect global climate. The ARM program includes several mobile measurement facilities, but the majority of the observations come from permanent facilities in three primary regions: the North Slope of Alaska, the Tropical Western Pacific, and the Southern Great Plains (SGP). The SGP field measurement site consists of a central facility in northern Oklahoma and a modest number of extended facilities spread over 55,000 square miles in northern Oklahoma and southern Kansas. The data analyzed in this paper were collected at the SGP site, and all of the data may be accessed on the web at http://www.archive.arm.gov. The extended facilities collect a host of meteorological observations, including surface measurements of air temperature, air pressure, relative humidity, and horizontal wind speed and direction. We analyze
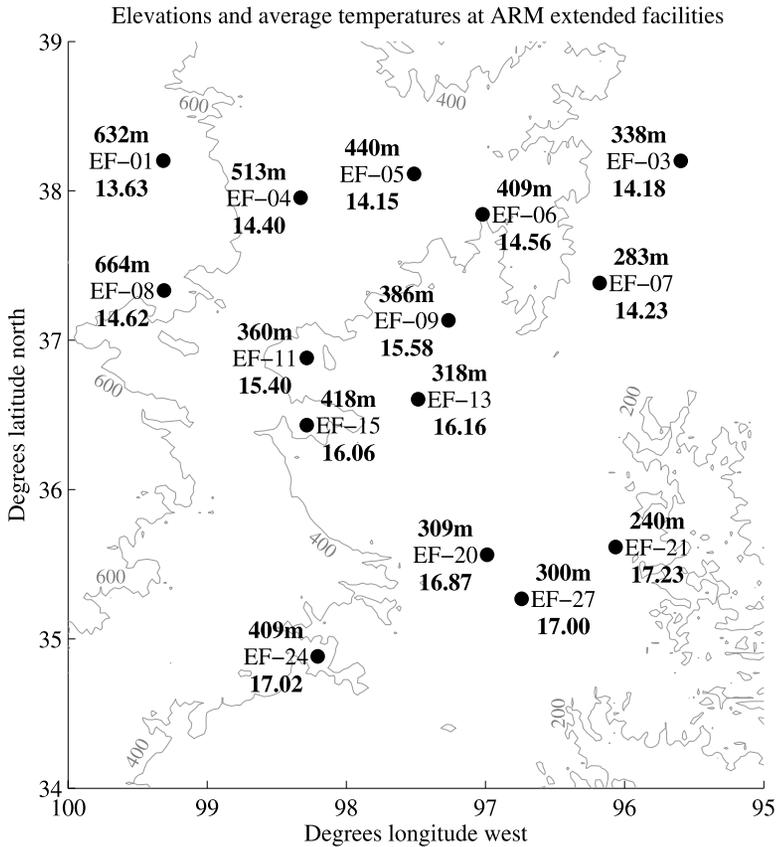
Elevations and average temperatures at ARM extended facilities

FIG. 1.   *Locations and elevations* (*meters*) *of the monitoring sites and their October* 2005 *average temperatures* (°C). *One degree latitude is approximately* 111 *km, and one degree longitude at* 37 *degrees latitude is approximately* 89 *km. The central facility is located at site EF-13.*

the temperature data from 15 extended facilities from the first 30 days of October, 2005, and we also make use of a solar radiation measurement collected at the central facility. The positions and the elevations of the 15 monitoring sites and the average temperatures from the first 30 days of October 2005 are plotted in Figure 1. The temperature data are collected every minute to the nearest 0.01°C. Modeling and analysis of meteorological data at such a high temporal frequency are of basic scientific interest, as evidenced by the emergence of the field of micrometeorology, which aims to describe the physical properties of atmospheric processes on very fine scales. In addition, the high frequency interpolations may be of use as inputs to high resolution regional climate models. When a regional climate model has a spatial resolution of 10 km or less, it is typically run at time steps of less than one minute. Such high resolution models include the SPoRT-WRF model, which includes some 1 km resolution modeling
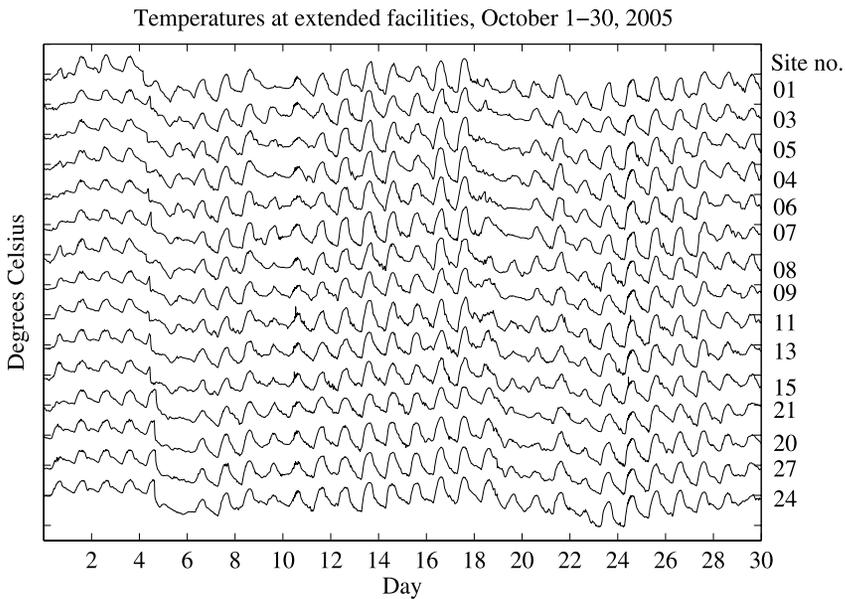
FIG. 2. *Temperature record for October* 1–30 *at* 15 *ARM SGP extended facilities. The tick marks on the vertical axis are separated by* 20°*C, and the sites are ordered from north* (*top series*) *to south* (*bottom series*) *and offset from each other by* 20°*C.*

(http://weather.msfc.nasa.gov/sport/modeling/), and HIRLAM-B, which has a target resolution of 2.5 km (http://hirlam.org). The ARM SGP data we analyze are of high quality in that only 50 are missing out of 648,000, and since the missing data are so few, we simply linearly interpolate them in time. Figure 2 plots the temperature time series from each site.

As expected, on most days the temperature data exhibit a clear diurnal cycle in which the temperature begins to increase shortly after sunrise, warms throughout the morning, and then begins to cool in the late afternoon. In addition, the temperatures in the middle of the day usually display more temporal variability than the nighttime temperatures do. As a result, it is clear that a model for the temperature (or differences of temperature) time series with a stationary variance function is not sufficient for these data. Several authors, including Benth and Šal-tytė Benth (2007) and Campbell and Diebold (2005), have addressed the issue of nonstationary variance in temperature observations in the context of pricing financial products on the weather derivatives markets. Their work mostly focuses on modeling seasonal variances for mean daily temperatures, whereas here we consider very high frequency data. For one-minute time resolution data, there may be some hope in explaining the changing variability of the temperature differences with other meteorological covariates. In fact, incoming solar radiation, which is measured at the central facility at one minute increments, has a strong connection to the temperature variability, as seen in Figure 3. We explore this connection more
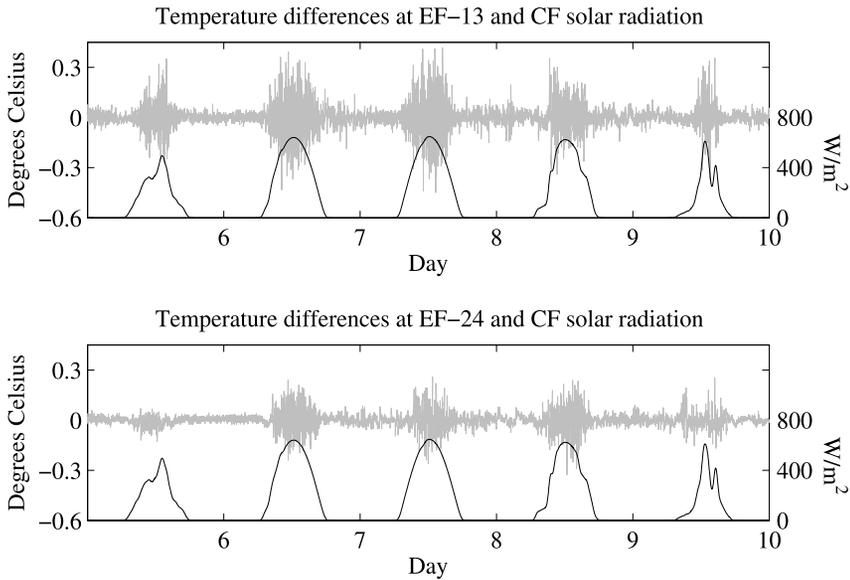
Temperature differences at EF−13 and CF solar radiation

Temperature differences at EF−24 and CF solar radiation

FIG. 3.   *Comparison between solar radiation measured at the central facility and temperature differences measured at two sites*, *EF*-13 *and EF*-24. *Site EF*-13 *is the site nearest the central facility*, *where the solar radiation is measured.*

closely in Section 6. In addition to modeling a nonstationary variance, we model a nonstationary correlation function, the details of which are discussed in Sections 5 and 6. Another conspicuous aspect of the data is the existence of extreme jumps in temperature, most notably during the fifth day, when the temperature at each site drops suddenly over a short period of time. In Section 4 we develop methods for interpolating the jumps to unobserved sites.

3. Statistical model.   We model the temperature process X at time t and location $\mathbf{u}$ as

$$(1) \qquad X(t, \mathbf{u}) = m(t) + s(\mathbf{u}) + J(t, \mathbf{u}) + Y(t, \mathbf{u}),$$

and we observe $X(t, \mathbf{u}_j)$ for $t = 1, \ldots, T = 43{,}200$ and for $j = 1, \ldots, n = 15$, the number of observation locations. The function m is a nonrandom temporal mean function, s is a spatial Gaussian process, J is a random spatial–temporal jump process, and Y is a nonstationary spatial–temporal Gaussian process. The temporal mean function, m, is estimated by smoothing the time series of spatially averaged temperatures. Specifically, we compute

$$(2) \qquad \overline{X}(t) = \frac{1}{n} \sum_{j=1}^{n} X(t, \mathbf{u}_j),$$

and we smooth $\overline{X}(t)$ in time with a 20-minute-bandwidth Daniell kernel applied three times to obtain $\hat{m}(t)$. The bandwidth was chosen manually to balance smoothness and fit. We treat $s(\mathbf{u})$ as known at each observation location and set its value to

$$(3) \qquad\qquad s(\mathbf{u}_j) = \frac{1}{T} \sum_{t=1}^{T} X(t, \mathbf{u}_j) - \overline{\overline{X}},$$

where $\overline{\overline{X}}$ is the grand mean of all observed temperatures. The spatial Gaussian process model for s allows us to provide predictions and conditional simulations of the spatial means at unobserved locations.

The random spatial–temporal jump process, J, is discussed in Section 4. The residual term, Y, is modeled as a nonstationary spatial–temporal Gaussian process using techniques from the spectral analysis of nonstationary time series. We discuss the details of the time series model in Sections 5 and 6, and we describe how we combine the time series models into a spatial–temporal process model in Section 7.

4. Spatial–temporal jump process. It is clear that incorporating nonstationarity into the temporal domain will provide a significant improvement to the fit. There is, however, one important aspect of the data that remains difficult to capture if we are to keep the amount of nonstationarity relatively constrained: the extreme drop in temperature occurring on October 5, plotted in Figure 4. On this day, each site experiences a rapid drop in temperature of roughly 4 to 10°C within a period of no more than 20 minutes, during which the first differences of temperature can be as high as 2°C. In contrast, the differences during the rest of the month have a sample standard deviation of 0.06°C. The issue is further complicated by the fact that the temperature drops are not simultaneous across the sites. Indeed, the site in the southeastern corner of the region records its drop more than 12 hours after the site in the northwestern corner records its drop. There seems to have been a weather system moving slowly across the region on that day.

We do not think it would be a good idea to model the jumps as part of the nonstationary spatial–temporal Gaussian process. For these data, such a Gaussian process model would need to be sufficiently pathological that it would violate the idea that the model is nearly stationary on short time scales. A more suitable option is to fit a random spatial–temporal jump process model in which the timing, sizes and steepness of the jumps are random spatial processes. This not only allows us to interpolate the jumps at unobserved sites but also provides a way to model the uncertainty of the interpolated jumps.

Consider the following parametric form for the jump process:

$$(4) \qquad J(t, \mathbf{u}) = \begin{cases} f(t; \tau(\mathbf{u}), D(\mathbf{u}), \lambda(\mathbf{u})), & \text{if } 1440 \times 4 < t \leq 1440 \times 5, \\ 0, & \text{if otherwise,} \end{cases}$$
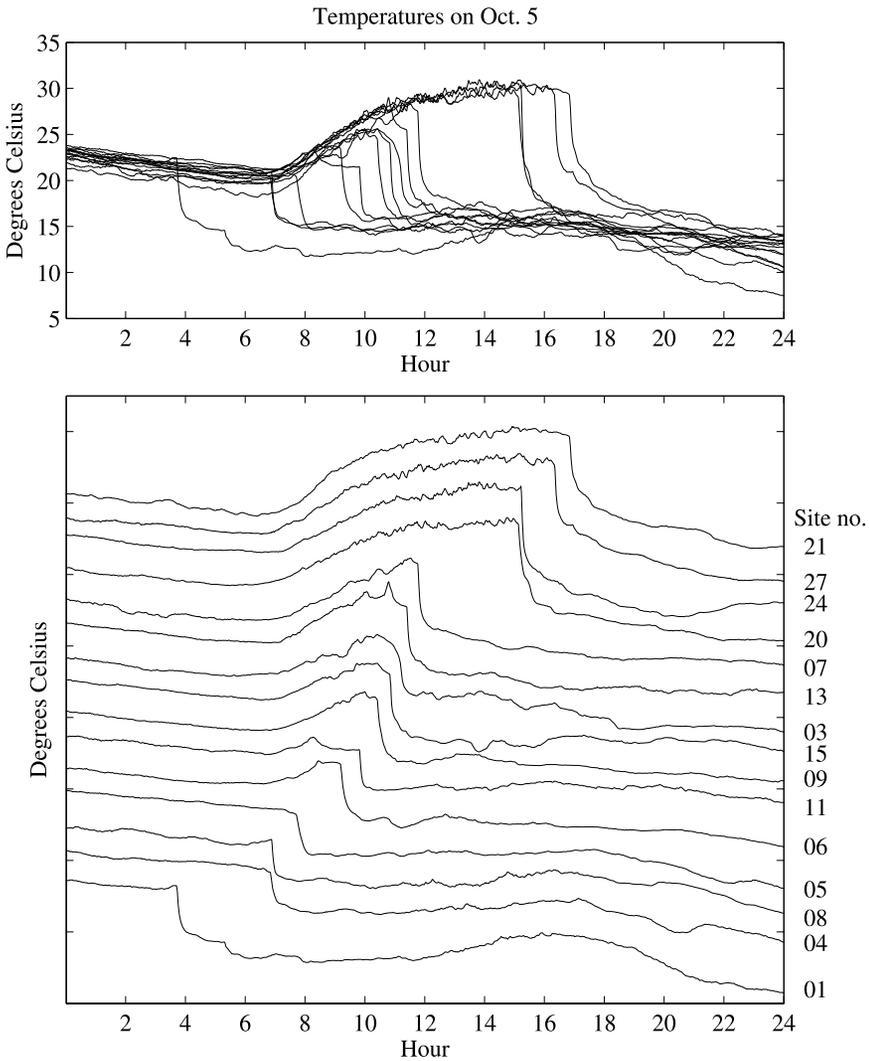
FIG. 4.   *Temperature record for October* 5 *at* 15 *ARM SGP extended facilities. On the lower plot, the vertical axis tick marks are separated by* 10°*C, and the temperatures are offset by* 4°*C per site, ordered by the time of the extreme temperature drop.*

so that the jump process is nonzero only on day 5, and where f is a nonrandom function of time given $\tau(\mathbf{u})$, the time the jump occurs, $D(\mathbf{u})$, the size of the jump, and $\lambda(\mathbf{u})$, the steepness of the jump. We introduce randomness into the jump process by modeling these parameters as random spatial processes, allowing us to model the uncertainty in the timing, size and steepness of the jumps. The idea is that if we can estimate these parameters using the data from locations $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$, we can conditionally simulate the parameters at unobserved sites $(\mathbf{u}_{01}, \ldots, \mathbf{u}_{0m})$
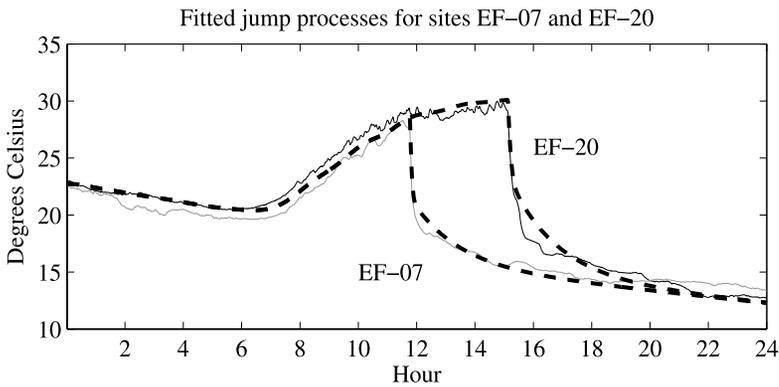
FIG. 5. *Example of fitted jump processes at sites EF*-07 *and EF*-20. *The fitted jump processes are the dashed lines*, *the temperatures from site EF*-07 *are the solid gray line*, *and the temperatures from site EF*-20 *are the solid black line*.

via the spatial models and plug the conditionally simulated parameters into f to construct conditionally simulated jump processes at the unobserved sites. Two examples of the fitted jump processes are included in Figure 5.

Choosing the specific form for f requires careful input from the modeler, and we leave the details to Appendix A. The particular formulation that we chose is highly specialized for these data, so we do not propose that our analysis may be blindly applied to other non-Gaussian features in the data or even other jumps. We do not see at present a way to automate the modeling of this kind of feature in high frequency meteorological data. However, we may in general be able to apply the idea of fitting parametric functions to features in the data at each site and then interpolating the feature to unobserved sites by spatially interpolating the fitted parameters.

In addition to the extreme drops in temperature, sites EF-11 and EF-15 record short upward bursts in which the temperatures are raised by several degrees Celsius and subsequently return to the pre-burst temperature a short time later, usually within 30 minutes. The bursts are recorded on two separate days, Oct. 11 and 25. Because the bursts are short-lived and do not cause any major shift in the process, we simply remove and replace them with linear functions connecting the pre- and post-burst temperatures. We do not know what is causing the bursts; however, it is interesting to note that the bursts occur at two sites separated by 50 km and the burst at the northern site follows the burst at the southern site by 75 minutes on day 11 and by 125 minutes on day 25. In any case, since they are short-lived, and the temperatures return to near the pre-burst temperatures, little is lost in ignoring them. In contrast, the drop on Oct. 5 is sustained, so it is not possible to ignore it.

Once the mean functions and the jump process have been estimated, we may subtract them from the data and fit a nonstationary spatial–temporal Gaussian pro-

cess model to

$$\hat{Y}(t, \mathbf{u}_j) = X(t, \mathbf{u}_j) - \hat{m}(t) - s(\mathbf{u}_j) - \hat{J}(t, \mathbf{u}_j), \tag{5}$$

where $\hat{m}(t)$ and $s(\mathbf{u}_j)$ are described in Section 3, and $\hat{J}(t, \mathbf{u}_j)$ is the estimate of the jump process, which is outlined in Appendix A. The details of the model for $\hat{Y}$ are to be described in the following sections.

5. Locally stationary time series model. The spatial–temporal model for the residuals, which will be discussed in Section 7, specifies nonstationary marginal time series models at fixed locations in space. Before describing the spatial–temporal model, it is instructive to first introduce some concepts from nonstationary time series. Dahlhaus (1996, 2012) introduced an asymptotic framework for nonstationary time series in which the process may be considered stationary over short time scales, thus allowing for consistent estimation of the model. Dahlhaus's work builds on the idea of the evolutionary spectrum, introduced by Priestley (1965). Specifically, let T be a positive integer and Z be a complex orthogonal increment process on $\mathbb{T}$, the unit circle, subject to the constraint that $Z(\omega) = Z(-\omega)^*$, where $*$ denotes the complex conjugate. Furthermore, let A be a complex function on $[0, 1] \times \mathbb{T}$ with $A(u, \omega) = A(u, -\omega)^*$ for every u and ω, and $\int_{\mathbb{T}} |A(u, \omega)|^2 \, d\omega < \infty$ for every u. Then

$$X_T(t) = \int_{\mathbb{T}} A(t/T, \omega) e^{i\omega t} \, dZ(\omega) \tag{6}$$

is a real-valued nonstationary process on $t = 1, \dots, T$.

Estimation of the model amounts to estimation of A, which is a time-varying transfer function, or, more frequently, of $|A|^2$, which is usually called an evolutionary spectrum. Dahlhaus (1997, 2000) proposes Gaussian likelihood approximations that may be maximized over some parametric family for $|A|^2$. Guinness and Stein (2013) provide an improved Gaussian likelihood approximation. Their likelihood approximates the integral in (6) with a sum over Fourier frequencies, $\omega_j = 2\pi j / T$, so that the nonstationary time series vector $\mathbf{X}_T$ may be written approximately as a linear transformation of a vector of uncorrelated random variables $\hat{\mathbf{Z}}$,

$$\mathbf{X}_T \approx C_T(A)\hat{\mathbf{Z}}, \tag{7}$$

where $C_T(A)$ is a $T \times T$ matrix with $(t, j + 1)$th entry $A(t/T, \omega_j) \exp(i\omega_j t)$. Therefore, the linear transformation $C_T(A)^{-1}\mathbf{X}_T$ is approximately decorrelating, and the negative loglikelihood approximation has the form

$$l_T(A) = \frac{T}{2} \log(2\pi) + \log|\det C_T(A)| + \frac{1}{2} \|C_T(A)^{-1}\mathbf{X}_T\|^2. \tag{8}$$

The likelihood approximation effectively makes the same approximation that the Whittle likelihood [Whittle (1953)] makes for stationary time series. In practice,

computation of the log determinant part of the likelihood is costly, so Guinness and Stein [(2013), Equation 10] provide an approximation to that term. At present, their log determinant approximation does not carry any known general theoretical guarantees, but it works well and outperforms other existing approximations in many examples. The quadratic form term may be computed efficiently using iterative methods and a fast Fourier transform (FFT), especially when A takes the form

$$(9) \qquad A(t/T, \omega) = \sum_{k=1}^{K} m_k(t/T)\mu_k(\omega)$$

with K being a small integer [Guinness and Stein (2013), Section 2]. We will always make the additional assumption that m_k and μ_k are nonnegative. The form in (9) includes the stationary model as a special case with $K = 1$ and m_1 constant. It also includes the uniformly modulated model as described by Priestley (1965), with $K = 1$ and m_1 not constant. When $K > 1$, we refer to the μ_k's as "regimes," so that at any time t/T, the transfer function is a positive linear combination of the various regimes. This representation is similar to the approach taken in hidden state modeling, where the process can be described at any time by a particular state or a superposition of several states. Fuentes (2002) proposed a similar formulation in which the nonstationary process is a spatially-varying superposition of several independent stationary processes.

For a spectral analysis of time series, it is often advantageous to prewhiten the data to smooth out peaks in the power spectrum [Priestley (1981)]. The spectra of the high frequency temperature time series tend to have peaks at the origin, and one way to smooth out a peak at the origin is to prewhiten the data with the difference filter, that is, perform a spectral analysis on the first differences, $\Delta X(t) = X(t) - X(t-1)$. Differencing is exactly a transformation to white noise when the time series is Brownian motion. For the temperature time series, we argue that differencing is too strong a filter because temperatures do exhibit some mean-reversion over monthly time scales, whereas Brownian motion does not over any time scale. Therefore, we propose to perform the analysis on partial differences, defined as

$$(10) \qquad \Delta_\alpha X(t) = X(t) - \alpha X(t-1).$$

The parameter, $\alpha \in [0, 1]$, which controls the amount of differencing, may be estimated with maximum likelihood. Partial differencing also stabilizes the simulated time series upon the operation of undifferencing, which has a huge impact on the usefulness of the conditional simulations (see Appendix C).

6. Solar radiation and choosing regimes. One striking feature of the data set is the relationship between the temporal variability in the first differences of temperature and the amount of incoming solar radiation: large amounts of sunshine usually result in more variable temperatures. Figure 3 shows an example of

the relationship. Meteorologists have hypothesized that convective forces in the atmosphere, which are most active during the middle of the day, may be driving the variability. One theory predicts that the standard deviation of surface air temperature changes should be linearly related to the heat flux at the earth-atmosphere boundary raised to the 2/3 power [Arya (2001), page 183]. We have found that, for these data, the standard deviation of the partial differences is approximately linearly related to incoming solar radiation. We use the radiation as a covariate to attempt to standardize the temporal variance in the differences. In order to quantify the ability of solar radiation to explain the time-varying variance, we fit three spatial–temporal models to the first differences of temperature at the 15 monitoring locations: (a) a stationary model, as described in Stein (2009), (b) a model that is stationary except for a separate variance for day and night, and (c) a model that is stationary except for a standard deviation that depends linearly on smoothed solar radiation measured at the central facility (the smoothing is described below). The day/night model increased the loglikelihood by 67,543 units over the stationary model, and the radiation-dependent model increased the loglikelihood by 104,286 units over the stationary model, so by any measure such as AIC or BIC, the additional variance parameters in both models are highly meaningful, and solar radiation explains the variance significantly better than does time of day alone.

To get a preliminary estimate for the relationship between the differences and solar radiation, we find $\hat{a}_0$ and $\hat{a}_1$ by fitting the model

$$(11) \qquad \Delta_\alpha \hat{Y}(t, \mathbf{u}) = [a_0 + a_1 r(t - \theta \mathbf{u}' \boldsymbol{\phi})] \varepsilon(t, \mathbf{u}),$$

where $\varepsilon(t, \mathbf{u})$ are i.i.d. $N(0, 1)$. The smoothed radiation, $r(t)$, is obtained by applying a left-sided, 10-minute-bandwidth Daniell kernel three times to the observed radiation. The parameters θ and $\boldsymbol{\phi}$ control the size and direction of the phase. For radiation from the sun, the phase should move west $[\boldsymbol{\phi} = (-1, 0)']$ at a speed of $\theta = 4$ minutes per degree longitude. Later, we will refine the estimates of a_0 and a_1 by maximizing the full approximate likelihood over them. Figure 3 shows first differences of temperature at site EF-13, which is located at the central facility where the radiation is measured, and at site EF-24, which is more than 200 km away from the central facility. It is not surprising that the relationship between variability and solar radiation is weaker when the site is far away from where the radiation measurement was taken. If one-minute resolution radiation data were available at every site, using that data would be preferable, but the extended facilities record radiation data only on the hourly resolution. There is also a matter of convenience: if central radiation measurements are used, there is no need to interpolate the radiation to unobserved sites. Despite these shortcomings, since a single covariate explains so much of the changing variability, it is worthwhile to include it in the model.

Another feature of the data, although less striking than the radiation phenomenon, is that the spectral properties of the time series differ depending on
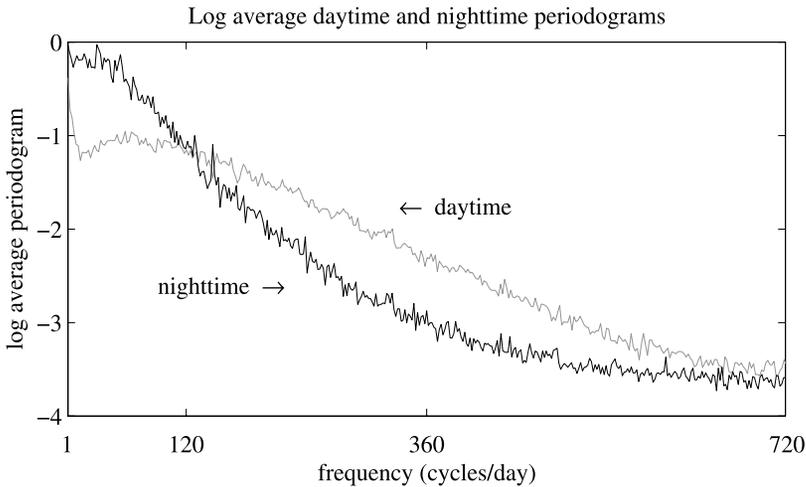
FIG. 6.   *Log average daytime (gray) and nighttime (black) periodograms computed with no sunrise or sunset offsets. Daytime (nightime) average is taken over all daytime (nighttime) blocks over all sites.*

the time of day. Furthermore, the difference cannot be explained with a uniform modulating function; the mean-zero daytime process is not simply a multiple of the mean-zero nighttime process. To illustrate this point, after dividing $\Delta_1 \hat{Y}(t, \mathbf{u})$ by $\hat{a}_0 + \hat{a}_1 r(t - \theta \mathbf{u}' \boldsymbol{\phi})$, we partition the data into daytime and nighttime blocks according to approximate local sunrise and sunset times,[2] compute periodograms for each block of data, and average all the periodograms from all the daytime blocks from all the sites to get an average daytime periodogram, and we average all the periodograms from all the nighttime blocks from all the sites to get an average nighttime periodogram. We plot the two averages in Figure 6. It is clear that the daytime spectrum cannot be modeled as a multiple of the nighttime spectrum. This means that not only does the variance of the process change over time, but the correlation structure changes as well.

Based on this evidence, we propose that the time series model requires at least two regimes [$K = 2$ in (9)], and a simple candidate model is one with a "daytime" regime and a "nighttime" regime. It is not clear how one would automatically estimate the two regimes, μ_1 and μ_2, via maximum likelihood. However, if we take the regimes to be fixed and known, it is possible to maximize the likelihood over

---

[2]The exact time of each day's local sunrise and sunset at the central facility was recorded, and then the approximate local sunrises and sunsets at the other sites were assumed to be offset by four minutes per degree longitude from the sunrise and sunset at the central facility on each day. At the beginning of the month, this approximation is nearly exact, as Oct. 1 is close to the equinox. At the end of the month, the approximation is off by roughly four minutes for the most northern site, which for our purposes is an acceptable error.

a family of modulating functions. Therefore, we estimate the regimes beforehand and treat them as fixed when we maximize the likelihood. In short, after defining the time at which the day starts and ends, which we do not assume is exactly at local sunrise and sunset, we estimate the regimes by fitting B-splines to the log of the average daytime and nighttime periodograms. One issue with the daytime and nighttime periodograms is that they do not provide information about the very low frequency behavior of the process because the periodograms are taken over short blocks of data. To estimate the very low frequency behavior, we borrow some information from the periodograms of the entire month of data. The details of the estimation of the regimes are left to Appendix B.

With the regimes fixed, we introduce a parametric form for the modulating functions, m_1 and m_2, in (9). We partition the time interval into B blocks, and we assume that, aside from a radiation-dependent uniform modulation, the evolutionary spectrum is constant in time on the blocks

(12)
$$M(t/T, \omega) = \sum_{k=1}^{2} \sum_{b=1}^{B} w_{kb} \mathbf{1}_b(t/T) \mu_k(\omega),$$

$$A(t/T, \omega) = (a_0 + a_1 r(t)) M(t/T, \omega),$$

where $w_{kb} > 0$, the μ_k's are the regimes, and $\mathbf{1}_b(x)$ is an indicator function that equals 1 if $x \in$ block b and 0 otherwise. Therefore, the parameters w_{kb} determine the weight assigned to regime k in block b. The model in (12) bears some resemblance to a cyclostationary model, in which the process is nonstationary within each day, but the nonstationary structure exactly repeats itself every day [Gardner, Napolitano and Paura (2006)]. Here, we do not require the spectrum to be cyclic, but we do enforce a relatively simple dependence on just two "regimes."

Because the regimes are taken to be fixed, a maximum approximate likelihood procedure involves choosing the linear radiation coefficients, the weights assigned to the regimes within each block, and the positions of the changepoints defining the blocks (in addition to fitting the spatial dependence structure, which is discussed in Section 7). Guinness and Stein (2013) implemented a genetic algorithm to find the changepoints, which works well when there are 1–10 changepoints. Here, however, where there are 60 changepoints per site, a more constrained search is necessary. We propose that it is reasonable to impose that the changepoints occur a fixed amount of time before or after local sunrise and sunset at each site and on each day. This constraint leaves us with just two changepoint parameters to estimate, so we may perform a grid search to find the best sunrise and sunset offsets according to the associated likelihoods. For each choice of sunrise and sunset offsets, we re-estimate the daytime and nighttime regimes as outlined in this section and maximize the likelihood over the radiation coefficients, the w_{kb}'s, and the parameters describing the spatial dependence to be described in Section 7.

7. Nonstationary spatial–temporal Gaussian process model. Recent work on incorporating nonstationarity into Gaussian process models has come in several forms. One of the approaches is to assume that data come from an isotropic random field on an unobserved domain E, and the domain on which we observed the data, D, is related to E via an unknown invertible mapping. Early work on this approach is due to Sampson and Guttorp (1992), and more recent advances include Anderes and Stein (2008) and Anderes and Chatterjee (2009). Several authors have constructed nonstationary processes via the convolution of a family of independent (of each other) stationary Gaussian processes. Nonstationarity enters when the convolution kernels are allowed to vary across the domain. Higdon, Swall and Kern (1999) use squared exponential covariance functions for the component processes, and Paciorek and Schervish (2006) provide an approach to include the Matérn class of covariance functions. Fuentes (2002) expresses the nonstationary process at each spatial location as a spatially-varying linear combination of stationary processes and estimates the model in the spectral domain. Others have modeled the nonstationary process in the spectral domain through the use of spatially- or spatial-temporally-varying spectral densities, analogous to the evolutionary spectrum approach studied in Section 5. Fuentes, Chen and Davis (2008) propose a class of nonseparable and nonstationary covariance functions through the use of a spectrum that varies across space and time.

The stationary, nonseparable covariance function in Stein (2009) is specially suited to facilitate fast approximate likelihood computations when the data are collected at a sparse set of fixed monitoring stations at regular intervals in time. In particular, he considered stationary spatial–temporal covariance functions of the form

$$(13) \qquad K(t, \mathbf{u}) = \int_{-\pi}^{\pi} S(\omega) R(|\mathbf{u}|/\gamma(\omega)) e^{i\omega(t-\theta\mathbf{u}'\boldsymbol{\phi})} \, d\omega,$$

where S is a spectral density, R is a one-dimensional correlation function invoking spatial coherence, γ is a positive function allowing the coherence to vary with temporal frequency, θ is a scalar, and $\boldsymbol{\phi}$ is a 2×1 vector. The parameters θ and $\boldsymbol{\phi}$ give the size and direction of the phase shift.

We construct a process that can be considered nonstationary in time by allowing the spectrum in (13) to depend on time. Rather than writing the covariance function, we give a spectral representation of the process,

$$\Delta_\alpha Y(t, \mathbf{u}) = \int_{-\pi}^{\pi} A\big((t - \theta\mathbf{u}'\boldsymbol{\phi})/T, \omega\big) e^{i\omega t} \, dZ_{\mathbf{u}}(\omega).$$

Because the temporal argument of A is shifted based on the location $\mathbf{u}$, we make the slight modification that A is a function defined on $\mathbb{R} \times \mathbb{T}$. In practice, the phase shift parameters have a very small effect on the likelihood, but rather than setting $\theta = 0$ and ignoring the phase, the default phase should correspond to a shift to local time because we know the temperature process is dependent on the time of day,

so we fix $\boldsymbol{\phi} = (-1, 0)'$, and $\theta = 4$ minutes per degree longitude. The orthogonal increment process, $Z_{\mathbf{u}}$, has spatial coherence

$$E\big(dZ_{\mathbf{u}}(\omega)\, dZ_{\mathbf{v}}(\omega)^*\big) = R\bigg(\frac{|\mathbf{u} - \mathbf{v}|}{\gamma(\omega)}\bigg)e^{-i\theta\omega(\mathbf{u}-\mathbf{v})'\boldsymbol{\phi}}\, d\omega,$$

where R is again a one-dimensional correlation function. We take $R(d) = \exp(-|d|)$, and use the convention that $R(0/0) = 1$. We parameterize $\gamma(\omega)$ as a B-spline whose derivative at $\omega = 0$ is 0, equals 0 for all $|\omega| > \omega_0$, and whose first and second derivatives at $\omega = \omega_0$ are 0. We choose ω_0 to be 48 cycles per day, and we place the interior knots in the B-spline at $(1, 3, 6, 12, 24)$ cycles per day, leaving us with 5 basis functions in the B-spline representation for γ. To ensure $\gamma(\omega) \geq 0$, we require that the B-spline coefficients be greater than 0. The purpose of γ is to allow the low (temporal) frequency fluctuations of the process to be more strongly correlated across space than the high frequency fluctuations of the process are, a phenomenon observed in the analysis of the Irish wind data [Haslett and Raftery (1989), Stein (2005)] and the ARM SGP air pressure data [Stein (2009)]. By setting $\gamma(\omega) = 0$ for all $|\omega| > \omega_0$, we are assuming that there is no spatial coherence above the 48 cycles per day frequency. This assumption is not unrealistic, at least at the available spacings between monitoring sites for these data, and it provides significant computational advantages in the likelihood approximation. Stein (2009) also considered a spatial nugget effect, but for these data and choice of coherence function, a nugget did not improve the fit, so we ignore it here.

In principle, one could allow the spectrum to depend on space as well. In fact, our spectrum does depend on space through the phase shift, but this is a very constrained dependence. For the ARM temperature data, which have just 15 spatial locations, we do not attempt to model any nonstationary aspects of the spatial covariance function. However, in the temporal domain, where we have 43,200 observations over the month per site, introducing nonstationarity has a substantial impact on the model fit.

8. Likelihood approximation and model fitting. We fit a spatial–temporal model to the data with two sites removed, site EF-08, which is on the western edge of the observation domain, and EF-09, which is near the center of the observation domain. We refer to their locations as $\mathbf{u}_{01}$ and $\mathbf{u}_{02}$. Given fixed values of the sunrise and sunset offsets and the differencing parameter, our method for fitting the spatial–temporal model proceeds as follows:

1. Estimate the temporal and spatial mean functions (Section 3) and the jump process parameters (Section 4).
2. Construct $\hat{Y}(t, \mathbf{u}_j)$ as in equation (5).
3. Preliminarily estimate a_0 and a_1 by fitting the model in (11) to obtain $\tilde{a}_0$ and $\tilde{a}_1$.
4. Divide each time series $\Delta_\alpha \hat{Y}(t, \mathbf{u}_j)$ by $\tilde{a}_0 + \tilde{a}_1 r(t)$ and estimate the regimes using the average daytime and nighttime periodograms, where the start of the

daytime and nighttime blocks are determined by local sunrise and sunset plus the offsets.

5. Maximize the approximate likelihood, which will be described at the end of this section, simultaneously over a_0, a_1, the 122 w_{bk} parameters and the 5 parameters describing the B-spline representation for $\gamma(\omega)$.

To speed the optimization, we compute analytic gradients of the likelihood with respect to a_0, a_1 and each w_{bk}, and we compute finite difference gradients with respect to the $\gamma(\omega)$ parameters. It should also be noted that the model in (12) is overspecified in that we may multiply the radiation coefficients by a constant and divide all the weights by the same constant and arrive at the same function A. In the optimization procedure, we fit the overspecified model, as the naturally specified model tends to get stuck in local minima. For fixed values of the offsets and differencing parameter, the time it takes to complete the optimization will vary but usually takes on the order of 2–3 hours on a single processor. In Figure 7 we plot the maximum likelihood weights, w_{bk}. As expected, the daytime blocks usually receive more weight from the "daytime" regime, and the nighttime blocks usually receive more weight from the "nighttime" regime, although there are some exceptions. The maximum likelihood coherence, $C(|d|/\gamma(\omega))$, as a function of frequency at various distances is plotted in Figure 8. The fitted model for the temperature data exhibits much weaker spatial coherence than is estimated for the pressure data in Stein (2009).

To explore the effect of the sunrise and sunset offsets, we repeat the optimization first over a coarse grid and then over a finer grid of sunrise and sunset offsets (with α fixed at 0.99). Once we have obtained results on the finer grid, which has 10 minute spacings, we fit a quadratic to the points near the maximum and then repeat the optimization procedure at several points near the maximum of the quadratic fit.
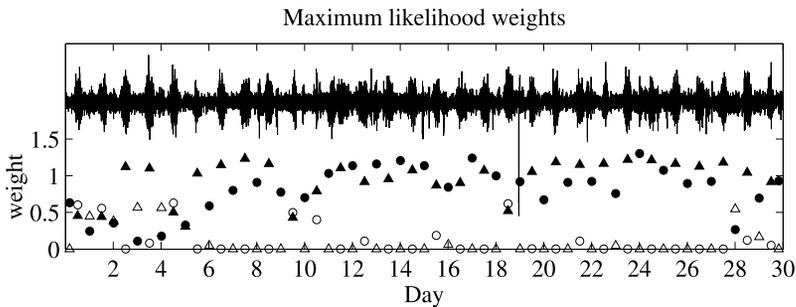

FIG. 7. *Maximum likelihood estimates of the weights assigned to the daytime (triangles) and nighttime (circles) regimes within each block. The filled markers indicate regimes that match the time of day, that is, if it is daytime, then the triangles are filled, and if it is nighttime, the circles are filled, so we expect the filled markers to receive more weight (e.g., we expect the daytime blocks to receive more weight from the daytime regime). We also plot the time series of partial differences for one of the sites, EF-27.*
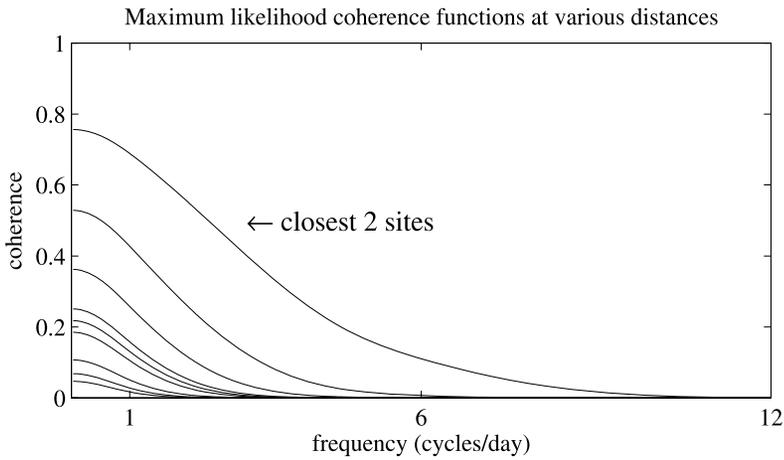
Maximum likelihood coherence functions at various distances

FIG. 8.   *Maximum likelihood estimate of the coherence as a function of frequency. Each line corre-*
sponds to the coherence at a fixed spatial distance. The coherence function was allowed to be nonzero
at up to 48 *cycles per day.*

The optimization over a grid is easily parallelizable because we can assign each
grid point to a processor, so the grid search is relatively fast when one has access to
multiple processors. The results of the grid search are given in Table 1. The offset
parameters have a big effect on the likelihood, and we find that sunrise offset 111
minutes and sunset offset -125 minutes give the highest maximum approximate
likelihood; the temperature process seems to undergo a change roughly two hours
after sunrise and two hours before sunset.

One may also attempt to obtain a higher likelihood by altering α, the amount of
partial differencing. Assuming that the differencing does not alter the locations of
the changepoints of the process, we may use the estimates of the sunrise and sun-
set offsets obtained previously, then repeat the likelihood optimization as before
over a grid of differencing parameters. In Figure 9 we plot the maximum approx-
imate likelihood as a function of α for sunrise offset 111 and sunset offset -125.
Choosing $\alpha = 0.997$ gives the highest maximum approximate likelihood.

The Gaussian likelihood approximation for the spatial–temporal data proceeds
in much the same way as it does for the time series data. After the mean func-
tions and the estimated jump process have been removed from the data, we con-
struct $\Delta_\alpha \hat{\mathbf{Y}}_j$, which is the vector of partial differences of the residuals at site $\mathbf{u}_j$,
for $j = 1, \ldots, n - 2 = 13$, the number of observed sites. For some choice of A,
and for each site $\mathbf{u}_j$, we compute the approximate transformation to independence
$\hat{\mathbf{Z}}_j = C_T(A)_j^{-1} \Delta_\alpha \hat{\mathbf{Y}}_j$, where the subscript j on $C_T(A)_j$ reminds us that the phase
shift in the temporal argument of A depends on location $\mathbf{u}_j$ (we omit from the
notation the dependence on θ and $\boldsymbol{\phi}$, which are taken to be fixed). The inverse
transformation is efficiently computed using an iterative algorithm and the FFT, as

TABLE 1
Results of the grid search over sunrise and sunset offsets (minutes) with differencing parameter
$\alpha = 0.99$. *Table entries refer to difference in loglikelihood from that found at sunrise offset* 111 *and*
sunset offset −125. *The units are thousands of loglikelihood units*

Sunrise offset	Sunset offset					
	−200	**−150**	**−100**	**−50**	**0**	**50**
−50	10.05	8.32	8.03	8.30	8.22	10.34
0	8.43	6.47	6.31	6.92	7.03	9.55
50	5.96	3.51	3.41	5.00	5.53	8.38
100	3.78	0.71	0.51	3.43	5.03	7.80
150	4.28	1.45	1.68	4.67	6.31	9.19
200	7.22	5.39	6.07	7.73	8.63	11.81
	−150	**−140**	**−130**	**−120**	**−110**	**−100**
80	1.426	1.047	0.940	0.948	1.062	1.243
90	1.050	0.648	0.533	0.537	0.655	0.848
100	0.712	0.302	0.177	0.176	0.303	0.506
110	0.603	0.187	0.069	0.068	0.205	0.425
120	0.682	0.278	0.162	0.174	0.326	0.566
130	0.737	0.345	0.241	0.264	0.438	0.694

described in Guinness and Stein (2013). Then the approximate covariance matrix
of $(\hat{\mathbf{Z}}_1, \ldots, \hat{\mathbf{Z}}_n)$ is block diagonal with block sizes equal to the number of spatial
locations because we allowed for spatial correlation in the orthogonal increment
process. Inverting this block diagonal covariance matrix is much easier than invert-
ing the full covariance matrix of the differences of the residuals, especially since
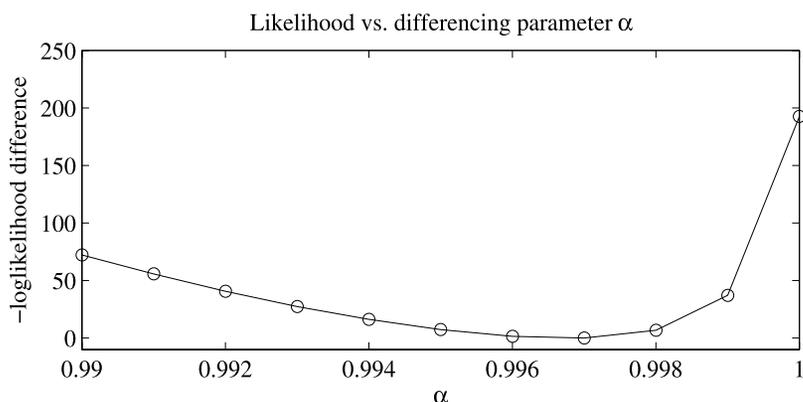

FIG. 9. *Results of a grid search over* α, *the differencing parameter. Here we plot the difference
from the loglikelihood maximized when* $\alpha = 0.997$. *We use sunrise offset* 111 *and sunset offset* −125.

we assume that the spatial correlation in $Z_{\mathbf{u}}$ is zero for all frequencies greater than 48 cycles per day. The log determinant term is computed as in Guinness and Stein (2013) with the modification that the spatial correlation of $(\hat{\mathbf{Z}}_1, \ldots, \hat{\mathbf{Z}}_n)$ must be accounted for, but this is also quite simple because the log determinant of a block diagonal matrix is easily obtained when the block sizes are small.

9. Conditional simulations.

Here we describe how we conditionally simulate the process at the unobserved sites, $\mathbf{u}_{01}$ and $\mathbf{u}_{02}$. To account for some of the uncertainty in the fitted model, we construct $\tilde{A}$ with weights $\tilde{w}_{kb}$ that are sampled from the asymptotic distribution of the maximum likelihood weights, $\hat{w}_{kb}$, whose covariance is given by the inverse Hessian of the loglikelihood. We simulate the partially differenced process in the spectral domain by jointly simulating the complex Fourier coefficients $\mathbf{Z}_{01}$ and $\mathbf{Z}_{02}$ from the multivariate complex normal distribution conditional on the coefficients from the other sites, $(\hat{\mathbf{Z}}_1, \ldots, \hat{\mathbf{Z}}_n)$, where the coefficients at the same frequency from different sites are related with the spatial coherence function, and coefficients at differing frequencies are uncorrelated. The partially differenced process is then constructed with the transformation $\Delta_{\hat{\alpha}} \mathbf{Y}_{0j} = C_T(\tilde{A})_{0j} \mathbf{Z}_{0j}$, with $j = 1, 2$.

As part of fitting the jump process to the observed sites, we estimated three parameters for each site: $\tau(\mathbf{u})$, the time of the jump at site $\mathbf{u}$; $D(\mathbf{u})$, the size of the jump; and $\lambda(\mathbf{u})$, the steepness of the jump. To interpolate the jump process at the unobserved sites, we model the three parameters, $\tau(\mathbf{u})$, $\log D(\mathbf{u})$ and $\log 1/\lambda(\mathbf{u})$, as independent intrinsic stationary spatial Gaussian processes with mean functions $d_k - \theta_k \mathbf{u}' \boldsymbol{\phi}_k$ (which can be written as a linear function of the coordinates) and generalized covariance functions $G_k(d) = -\eta_k d$ [$k = 1, 2, 3$ refers to the parameters $\tau(\mathbf{u})$, $D(\mathbf{u})$, $\lambda(\mathbf{u})$]. The best linear unbiased predictors of the jump process parameters depend only on contrasts of the observed jump process parameters, so we may use restricted maximum likelihood to estimate η_k and to predict the parameters at the unobserved sites [see, e.g., Stein (1999), page 171]. We conditionally simulate the jump process parameters at the unobserved sites using bivariate t distributions on 10 degrees of freedom (13 observed sites—3 mean parameters) to account for uncertainty in the estimates of η_k. We then reconstruct the jump process from the conditionally simulated jump process parameters. As an example, we plot in Figure 10 the data from October 5 at site EF-09, along with 20 conditional simulations of the jump process and 2 conditional simulations of the temperature process.

We also model the spatial mean function as a spatial Gaussian process (independent of the other processes). Its generalized covariance has the same form as that of the jump process parameters, and its mean function is linear in latitude and elevation. We use the fitted model to simulate $(s(\mathbf{u}_{01}), s(\mathbf{u}_{02}))$ jointly and conditionally on the means from the other sites. The conditionally simulated process at
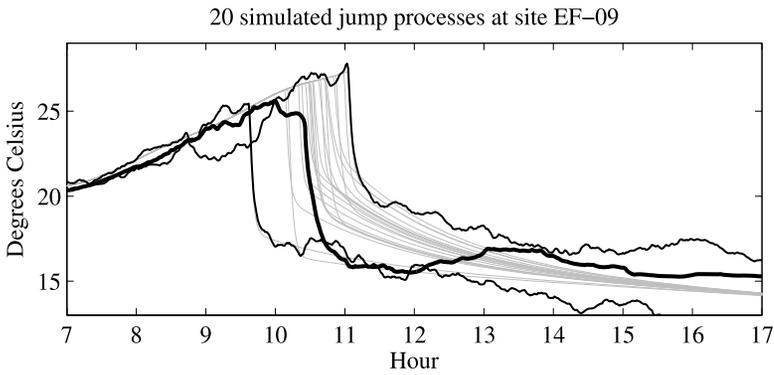
FIG. 10. 20 *conditionally simulated jump processes* (*gray*), *along with the data* (*thick black line*) *and* 2 *full conditional simulations* (*thin black lines*) *corresponding to the earliest and latest simulated jump times among the* 20.

site $\mathbf{u}_{0j}$ is then constructed with

$$\hat{Y}(1, \mathbf{u}_{0j}) = 0,$$

$$\hat{Y}(t, \mathbf{u}_{0j}) = \hat{\alpha}\hat{Y}(t-1, \mathbf{u}_{0j}) + \Delta_{\hat{\alpha}}\hat{Y}(t, \mathbf{u}_{0j}), \qquad t > 1,$$

$$\hat{X}(t, \mathbf{u}_{0j}) = \hat{Y}(t, \mathbf{u}_{0j}) + \hat{m}(t) + s(\mathbf{u}_{0j}) + \hat{J}(t, \mathbf{u}_{0j}).$$

We simulate 99 conditionally independent bivariate time series using the methods described above. The simulations are very fast; each bivariate simulation takes just a few seconds on a single processor. Confidence bands may be constructed by computing the quantiles at each time point among the conditional simulations. To evaluate our conditional simulations, we constructed 90% confidence bands and found that they had a coverage rate of 89.6% for the central site and a coverage rate of 93.3% for the peripheral site. The roughly accurate coverage rates are a promising result, although we are slightly conservative for the peripheral site. Table 2 shows that the predicted temperatures at the peripheral site and during the daytime have wider confidence bands. This is not surprising, as we expect it to be more

TABLE 2
Average width of 90% *confidence intervals in* °C *for predicted temperatures at the two unobserved sites over daytime blocks, nighttime blocks and overall. Sample standard deviations are given in parentheses*

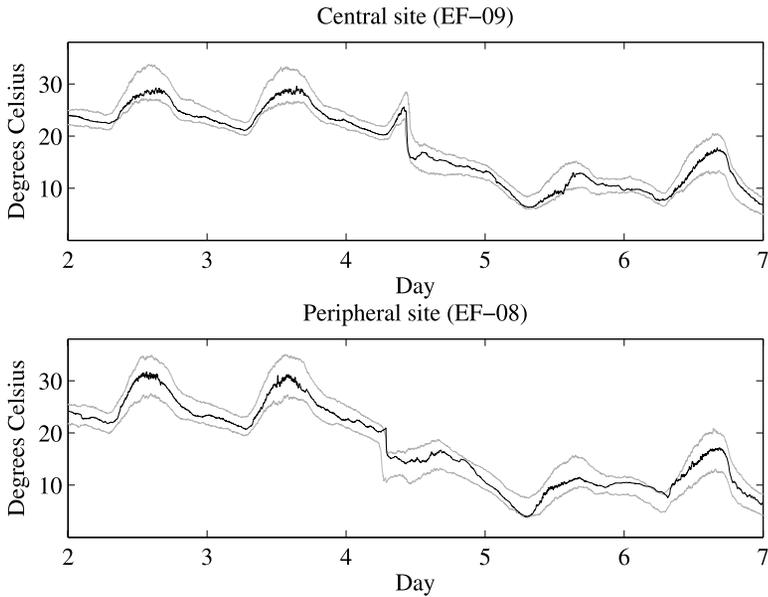|	Daytime	Nighttime	Overall
Peripheral	5.95 (1.23)	4.79 (1.08)	5.14 (1.24)
Central	5.03 (1.17)	3.90 (1.00)	4.24 (1.18)

FIG. 11.   90% *confidence bands for predicted temperature at the two held-out sites over days* 3
through 7. *This time period includes the extreme drop in temperature, which occurred on day* 5.

difficult to predict temperatures at the boundary of the observation region, and
the daytime temperatures tend to be more variable. The confidence bands around
the time of the jump become very wide; at the peripheral site, which undergoes
a smaller jump because the drop occurs earlier in the day, the confidence band is
wider than 8°C for more than 40 minutes, and at the central site, the confidence
band is wider than 9.5°C for 30 minutes near the time of its drop. In Figures 11
and 12 we plot a portion of the data from the unobserved sites along with the 90%
confidence bands for predicted temperature.

10. Conclusions.   We have proposed a spatial–temporal model that aims to
capture nonstationary variance and correlation in a set of high frequency tempera-
ture data. The model also allows for the variance to depend on another meteorolog-
ical covariate, solar radiation, and includes spatial–temporal jumps. We provided
computationally efficient methods for fitting the model to a large data set and for
generating spatial–temporal simulations from the model, conditional on the obser-
vations. Most of the computational effort is spent on fitting the model. Once the
model is fit, the conditional simulations can be computed very quickly. The condi-
tional simulations result in a suite of temperature values at unobserved locations,
and we have shown that the simulated data reflect some of the uncertainties that
we expect, namely, that the interpolations are more uncertain during the daytime,
that there is more uncertainty in predicting at a peripheral location than there is in
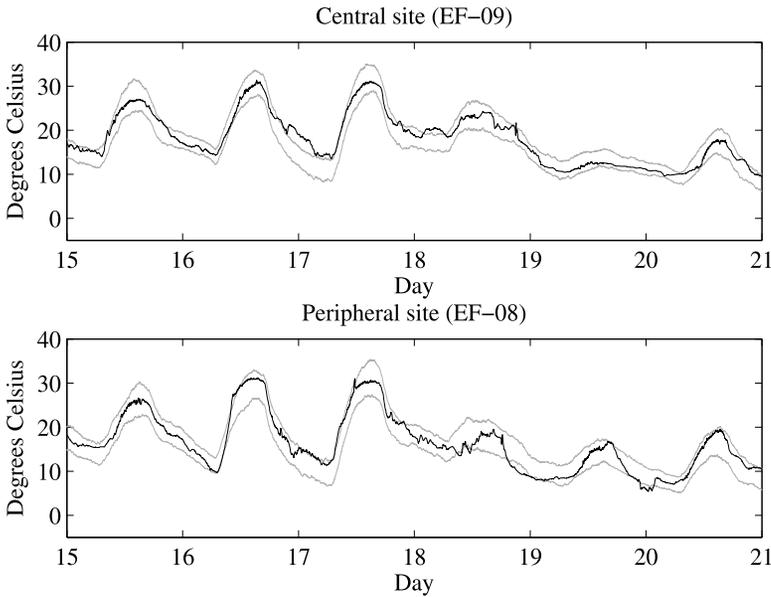
FIG. 12.   *90% confidence bands for predicted temperature at the two held-out sites over days* 16 *through* 21. *The confidence bands are usually wider during the daytime, and the peripheral site's confidence band is generally wider than that of the central site.*

predicting at an interior location, and the uncertainties are inflated near the time of the jump.

APPENDIX A: PARAMETRIC FORM AND FITTING OF JUMP PROCESS

As seen in Figure 4, the temperatures at each site follow the usual diurnal cycle until the jump occurs, when they undergo a very sharp drop followed by a slow decay toward a fixed temperature, roughly 15°C. This suggests that the following may be an appropriate formulation of the jump process:

$$
(14) \qquad J(t, \mathbf{u}) = \begin{cases} b_1(t) - m(t), & \text{if } 4 \times 1440 < t < \tau(\mathbf{u}), \\ b^*(t, \mathbf{u}), & \text{if } \tau(\mathbf{u}) \le t \le 5 \times 1440, \\ 0, & \text{otherwise}, \end{cases}
$$

where b_1 is a mean function that shadows the diurnal cycle, and

$$
\begin{aligned}
b^*(t, \mathbf{u}) = {} & \big[b_1(\tau(\mathbf{u})) - m(\tau(\mathbf{u}))\big] \exp(-\nu_1(t - \tau(\mathbf{u}))) \\
& - D(\mathbf{u}) \int_0^{t-\tau(\mathbf{u})} \frac{\lambda(\mathbf{u})^\beta}{\Gamma(\beta)} x^{\beta-1} \exp(-\lambda(\mathbf{u})x)\,dx \\
& + D(\mathbf{u})\big[1 - \exp(-\nu_2(t - \tau(\mathbf{u})))\big] \\
& + \big[b_2(t) - m(t)\big]\big[1 - \exp(-\nu_3(t - \tau(\mathbf{u})))\big].
\end{aligned}
$$

(15)

The integral term is simply the incomplete gamma function with rate $\lambda(\mathbf{u})$ and shape β. We use the incomplete gamma function to model the drop because the first differences of temperature immediately following the jump resemble gamma densities. Additionally, we could allow the shape parameter, β, to vary with location, but for simplicity, we assume that β is constant in space and fix it at $\beta = 1.3$. The function b_2 is a post-jump mean that we take to be linear. The rates ν_1, ν_2 and ν_3 are less critical, so they may be taken to be equal, but they should be much smaller than $\lambda(\mathbf{u})$. In our analysis, we assume $\nu_j = 0.01$ for each j.

To fit $b_1(t)$, we compute a moving average of the average temperature at time t, where the average is taken over the sites that have not undergone a jump up to time t. Thus, for each $\mathbf{u}$, we must find a preliminary estimate of $\tau(\mathbf{u})$, which we call $\overline{\tau}(\mathbf{u})$, found by taking the first time at each site for which the first differences are below $-0.35°C$. Later, the estimates of the jump times will be refined. To estimate $b_2(t)$, we fit a linear function to the average of the temperatures at time t from the sites for which $\overline{\tau}(\mathbf{u}) + 60 < t$. The line is constrained so that $b_2(1440 \times 5 + 1) = m(1440 \times 5 + 1)$, the value of the temporal mean function at the start of the sixth day. To fit $\tau(\mathbf{u})$, $D(\mathbf{u})$ and $\lambda(\mathbf{u})$, we minimize the sum of squared differences between the first differences of the fitted jump process with those parameters and the first differences of temperature, where the sum of squares is taken over all sites and for times $\overline{\tau}(\mathbf{u}) - 3, \ldots, \overline{\tau}(\mathbf{u}) + 20$.

APPENDIX B: CHOOSING THE REGIMES

One problem with the average daytime and average nighttime periodograms is that the periodograms are taken over blocks of data that are never longer than 18 hours, so they do not give any information about the very low frequency (less than 1 or 2 cycles per day) behavior of the process. Therefore, we are forced to try to learn something about the low frequency behavior from periodograms taken over longer blocks of data. We estimate the shape of A at low frequencies $(-\omega_S, \omega_S)$ by computing the periodogram using the entire month of data at each site and averaging the periodograms over the sites, which we call the average monthly periodogram. Here, we take $\omega_S = 2$ cycles per day.

Based on the nonmonotonic shape of the daytime and nighttime periodograms and the requirement that the regimes be positive, it seems reasonable to fit cubic B-splines to one half of the log of the periodograms to model the regimes. The knots are chosen manually, and we take them to be more dense at lower frequencies. The positions of the knots are $(1/3, 2/3, 1, 4/3, 5/3, 2, 4, 8, 12, 24, 60, 120, 360)$ in units of cycles per day. Under the constraints that the first derivative of $\mu_k(\omega)$ at $\omega = 0$ and π be equal to zero, and $\mu_k(\omega) = \mu_k(-\omega)$, we are left with 15 basis functions per regime in the cubic B-spline representation, of which 8 are nonzero in the interval $(-\omega_S, \omega_S)$. We fit the coefficients for those 8 basis functions by minimizing the sum of squares between the associated B-spline and one half the log of the average monthly periodogram at the frequencies in the interval $(-\omega_S, \omega_S)$.

Holding those 8 coefficients constant at their least squares estimates, we choose the remaining 7 coefficients for each regime by minimizing the sum of squares between the associated B-spline and one half the log of the average daytime (and nighttime) periodogram. This procedure ensures that μ_1 and μ_2 are equal in the interval $(-\omega_S, \omega_S)$ but free to vary at higher frequencies.

APPENDIX C: MOTIVATION FOR PARTIAL DIFFERENCING

Until now, we have given only a brief explanation for the need for partial differencing. If we set $\alpha = 1$ and perform the analysis on the first differences of temperature, the conditional simulations become almost useless. As the month progresses, the simulated temperatures tend to drift away from the data, resulting in very wide confidence bands for most of the month. This is a problem that was not observed in the analysis in Stein (2009), in which he fit and simulated from a smooth, uniformly modulated model. Here, the undifferencing operation interacts with the discontinuous jumps in A, causing the undifferenced simulations to drift. More specifically, in the stationary case the basis functions for the process are trigonometric functions, which integrate to zero over every cycle. In the locally stationary case, the basis functions are amplitude-modulated trigonometric functions. When A is smooth as a function of time, as in Stein (2009), all but the highest frequency functions may approximately integrate to zero, but when A has jumps, this is no longer the case, so when the simulations are undifferenced, the results can be quite unpredictable. Partial differencing appears to solve the problem and has some theoretical merit given the nature of the temperature process, as discussed in Section 5.

REFERENCES

ANDERES, E. and CHATTERJEE, S. (2009). Consistent estimates of deformed isotropic Gaussian random fields on the plane. *Ann. Statist.* **37** 2324–2350. MR2543694

ANDERES, E. B. and STEIN, M. L. (2008). Estimating deformations of isotropic Gaussian random fields on the plane. *Ann. Statist.* **36** 719–741. MR2396813

ARYA, S. P. (2001). *Introduction to Micrometeorology*, 2nd ed. Academic Press, San Diego, CA.

BENTH, F. E. and ŠALTYTĖ BENTH, J. (2007). The volatility of temperature and pricing of weather derivatives. *Quant. Finance* **7** 553–561. MR2358919

CAMPBELL, S. D. and DIEBOLD, F. X. (2005). Weather forecasting for weather derivatives. *J. Amer. Statist. Assoc.* **100** 6–16. MR2166065

DAHLHAUS, R. (1996). On the Kullback–Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.* **62** 139–168. MR1388767

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916

DAHLHAUS, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794. MR1835040

DAHLHAUS, R. (2012). Locally stationary processes. In *Handbook of Statistics* 30, *Time Series Analysis*: *Methods and Applications* **30** 351–408. Elsevier, Amsterdam.

DALY, C., NEILSON, R. P. and PHILLIPS, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* **33** 140–158.

FUENTES, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika* **89** 197–210. MR1888368

FUENTES, M., CHEN, L. and DAVIS, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics* **19** 487–507. MR2523910

GARDNER, W. A., NAPOLITANO, A. and PAURA, L. (2006). Cyclostationarity: Half a century of research. *Signal Processing* **86** 639–697.

GUINNESS, J. and STEIN, M. L. (2013). Transformation to approximate independence for locally stationary Gaussian processes. *J. Time Series Anal.* **34** 574–590.

HASLETT, J. and RAFTERY, A. E. (1989). Space–time modelling with long-memory dependence: Assessing Ireland's wind power resource. *J. Appl. Stat.* **38** 1–50.

HIGDON, D., SWALL, J. and KERN, J. (1999). Non-stationary spatial modeling. In *Bayesian Statistics* 6: *Proceedings of the Sixth Valencia International Meeting* 761–768. Oxford Univ. Press, New York.

HIJMANS, R. J., CAMERON, S. E., PARRA, J. L., JONES, P. G. and JARVIS, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25** 1965–1978.

LANTUÉJOUL, C. (2002). *Geostatistical Simulation*: *Models and Algorithms*. Springer, New York.

PACIOREK, C. J. and SCHERVISH, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17** 483–506. MR2240939

PRIESTLEY, M. B. (1965). Evolutionary spectra and non-stationary processes (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **27** 204–237. MR0199886

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, San Diego. Seventh printing, 1992.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York. MR0899519

SAMPSON, P. D. and GUTTORP, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *J. Amer. Statist. Assoc.* **87** 108–119.

SCHNEIDER, T. (2006). Analysis of incomplete data: Readings from the statistics literature. *Bulletin of the American Meteorological Society* **87** 1410–1411.

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. Springer, New York. MR1697409

STEIN, M. L. (2005). Statistical methods for regular monitoring data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 667–687. MR2210686

STEIN, M. L. (2009). Spatial interpolation of high-frequency monitoring data. *Ann. Appl. Stat.* **3** 272–291. MR2668708

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423–434. MR0060797

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27695
USA
E-MAIL: jsguinne@ncsu.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637
USA
E-MAIL: stein@galton.uchicago.edu