Spatio-Temporal Reconstruction of Missing Forest Microclimate Measurements

Francesco Tonini[a, *], Whalen W. Dillon[a,b], Eric S. Money[a], and Ross K. Meentemeyer[a,b]

[a]Center for Geospatial Analytics, North Carolina State University, Raleigh, NC 27695-7106 USA.

[b]Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695-7106 USA.

*Corresponding author. Tel.: +1 (919) 513 0411

E-mail addresses: ftonini@ncsu.edu (Francesco Tonini), wwdillon@ncsu.edu (Whalen W. Dillon), esmoney@ncsu.edu (Eric Money), rkmeente@ncsu.edu (Ross K Meentemeyer)

Abstract

Scientists and land managers are increasingly monitoring forest microclimate environments to better understand ecosystem processes, such as carbon sequestration and the population dynamics of species. Obtaining reliable time-series measurements of microclimate conditions is often hindered by missing and erroneous values. In this study, we compare spatio-temporal techniques, space-time kriging (probabilistic) and empirical orthogonal functions (deterministic), for reconstructing hourly time series of near-surface air temperature recorded by a dense network of 200 forest understory sensors across a heterogeneous 349 km² region in northern California. The reconstructed data were also aggregated to daily mean, minimum, and maximum in order to understand the sensitivity of model predictions to temporal scale of measurement. Empirical orthogonal functions performed best at both the hourly and daily time scale. We analyzed several scenarios to understand the effects that spatial coverage and patterns of missing data may have on model accuracy: (a) random reduction of the sample size/density by 25%, 50%, and 75% (*spatial coverage*); and (b) random removal of either 50% of the data, or three consecutive months of observations at randomly chosen stations (random and seasonal *temporal missingness*, respectively). Here, space-time kriging was less sensitive to scenarios of spatial coverage, but more sensitive to temporal missingness, with less marked differences between the two approaches when data were aggregated on a daily time scale. This research contextualizes trade-offs between techniques and provides practical guidelines, with free source code, for filling data gaps depending on the spatial density and coverage of measurements.

Keywords (max 6): missing data, spatio-temporal prediction, microclimate sensors, empirical orthogonal functions, near-surface air temperature, California.

1. Introduction

The availability of complete long-term, high-resolution datasets of climatic conditions measured near the earth's surface is crucial for understanding ecological processes affected by environmental heterogeneity (McDonald and Urban, 2004; Meentemeyer, 1978; Turner and Chapin, 2005; Waring and Running, 1998). Fluctuations in these microclimate (Geiger, 1965) conditions may influence processes such as the dynamic evolution of an infectious disease in natural ecosystems (Meentemeyer et al., 2012), species distributions (Gehlhausen et al., 2000), and the structure of landscape patterns of carbon and nutrient cycling (Band et al., 1991). In forest ecosystems, weather conditions have historically been measured either at insufficient spatial densities to capture landscape-level variation in microclimate (Lookingbill and Urban, 2003) or at a "representative" site often cleared of vegetation (Bolstad et al., 1998). Technological advancements have led to high resolution data sets becoming more affordable and common, but this has produced a fresh challenge. Even when high-resolution meteorological and hydrologic observational datasets are obtained they are frequently beset with periods of missing data caused by the failure of data loggers (e.g. due to power outages), climatological events (e.g. snow, ice, or precipitation) (Henn et al., 2013), and even cattle interference (personal observation). In other cases, erroneous or unreasonable values are recorded and must be manually removed. Establishing effective and efficient methods for predicting microclimate conditions to fill in these gaps will increase the utility of these datasets for examining thresholds governing ecological dynamics at fine spatial and temporal scales.

Accurately filling in the missing data at fine spatial and temporal resolutions provide the necessary detail for modeling organisms that interact, but may respond differently to similar environmental conditions (e.g. a plant pathogen versus its host). A complete data set also allows for examination and testing of environmental thresholds under natural conditions. For example, the number of hours or days at or above (or below) a threshold temperature for disease development may be used to help predict whether a pest or pathogen is likely to occur, such as the models developed for controlling powdery mildew on grapes (Thomas et al., 1994). Seasonal averages of maximum or minimum daily temperatures at specific locations have also been used to assess the sensitivity of *Ixodes pacificus* tick densities (a Lyme disease carrier) in California forests, requiring finely resolved spatial and temporal data (Swei et al., 2011). These detailed data enable empirically based predictions of how pest-pathogen-host interactions may respond to future climate conditions (e.g. Caffarra et al. 2012).

Several statistical methods are available for reconstructing time series of missing air temperature data. In general, techniques for estimating missing data are similar to spatial interpolation, extrapolation, and forecasting in that available observations are used to reconstruct missing observations (Henn et al., 2013). Common spatial interpolation techniques include thin-plate splines (Pape et al., 2009), inverse-distance weighting (Daly et al., 2000), radial basis functions (Myers, 1992), and kriging (Cressie, 1993; Garen et al., 1994; Tobin et al., 2011). Temporally correlated processes are typically modeled using autoregressive time series (Raible et al., 1999). The above techniques are concerned either with estimating unknown values for single temporal realizations (spatial domain), or separately for each station regardless of their spatial proximity (temporal domain). Recent advancements in the theory of spatio-temporal processes have extended the above techniques for application to processes correlated in time and space (Cressie and Wikle, 2011).

In this paper, we present a framework, with comparison of two statistical techniques, for estimating missing time series data collected across heterogeneous landscapes. We specifically examined (i) probabilistic space-time kriging (Cressie and Wikle, 2011; Heuvelink and Griffith, 2010) on residuals from temporally-smoothed data and (ii) deterministic spatio-temporal correlations in the form of empirical orthogonal functions (EOF) (Beckers and Rixen, 2003; Lorenz, 1956). We applied these methods to an unusually dense network of hourly temperature measurements collected in forest understory environments across a spatially heterogeneous landscape. Although temperature measurements are often needed at different temporal scales (e.g. hourly, daily) depending on the analysis, for example, how species and communities respond to climate (Bolstad et al., 1998), little attention has been given to studying how temporal aggregation impacts the performance of methods for reconstructing missing data. A previous study inspected the effect of temporal aggregation on spatial prediction of understory temperature using physiographic and ecological factors over the same area used herein (Vanwalleghem and Meentemeyer, 2009). However, the presence of temporal dependence in the temperature time series was ignored in favor of focusing on single temporal aggregations, e.g. average maximum temperature for the month.

In this study, we address several key questions: how does a geostatistical technique (space-time kriging) compare to a deterministic one (EOF) when trying to accurately estimate data gaps in microclimate measurements? How does spatial coverage (i.e. sampling size and density of stations) and temporal completeness (i.e. amount of missing data and gap length) affect the accuracy of model estimates? Do larger temporal aggregations of observations influence model accuracy?

Finally, we offer guidelines and free source code for practitioners to fill gaps in understory temperature data, depending on the spatial density of available microclimate stations, duration of the missing data, and scale of analysis.

Section 2 describes the study area, the temperature dataset, and the methodologies used to reconstruct the missing data across scenarios of spatial coverage and temporal gaps. Results are presented in section 3. Section 4 presents discussions and conclusions based on the results of the study.

2. Materials and Methods

The steps required to pre-process the available dataset and all subsequent statistical analyses were implemented in the R environment for statistical computing (R Core Team, 2013), and we provide free source code[1].

2.1. Study area and data

During 2003, we established 200 ecological monitoring sites in forested stands across a 349-km^2 heterogeneous area in Sonoma County, California (Fig. 1). Sonoma County experiences a Mediterranean climate, with distinct wet and dry seasons (Barbour and Billings, 2000). Precipitation typically falls as rain from October through April, and a dry season occurs from May through September. The landscape is a mix of public and private property near the cities of Santa Rosa, Petaluma, and Sonoma with varying levels of agricultural and urban development

[1] **https://github.com/f-tonini/Microclimate-Sonoma**

(Fig. 1A). The forested areas are characterized by open oak woodlands, denser stands of mixed evergreen trees, and a few locations dominated by Coast redwood or Douglas-fir. Sites were randomly located across the forested landscape, resulting in an elevation gradient of the geo-referenced site centers from 55 m to 800 m (mean = 378 m). Each site was equipped with a microclimate data logger, and has been revisited annually through 2014 to monitor conditions and download the data. Initially, each location had a temperature/relative humidity data logger (model H08-032-08; Onset Computer Corporation, Bourne, MA, USA) installed inside a solar radiation shield (model RS1, Onset Computer Corporation, Bourne, MA, USA) 1-m off the ground (Fig. 1C). These loggers began to fail in 2008, and so they were replaced with new temperature-only data loggers (model UA-001-64, Onset Computer Corporation, Bourne, MA, USA) inside the same solar radiation shields (Fig. 1B). During this ongoing project different personnel have managed data collection, leading to inconsistencies in the setting of the temporal resolution of the loggers across the sites and between years. In the remainder of this paper we address this challenge, as well as methods for handling varying levels of missing data that may often plague datasets collected over many years.

Figure 1-caption at the end of file

2.2. Data processing

Several pre-processing steps were necessary in order to prepare the data for the analysis: (1) temperature values from each data logger were pre-screened to check for outliers exceeding plausible maximum and minimum values; (2) The full database was aligned to match a common hourly resolution. Data recorded at intervals shorter than an hour were averaged. This step was required because some data loggers recorded temperature values at different time intervals (e.g. 30, 45 minutes). (3) The full database was sliced to begin on May 1st, 2003 and end on April 30th, 2014 in order to keep a large number of available stations every year ($\geq$ 100). (4) Time series points in which the rate of change between hours was considered excessive were replaced by a missing-value flag. In this case, a rate of change >4°C h^{-1} was chosen as reasonable threshold based on other studies (Henn et al., 2013). (5) Lastly, temperature time series were visually inspected for each station to spot the presence of erroneous patterns (e.g. if a data logger kept recording data after being removed from the field). If present, these data points were replaced with a missing-value flag.

The analyses presented herein were conducted using data from the year 2004, which together with 2005 had the largest number of actively recording stations (n = 200). Overall, 10% of the observations were missing for 2004, providing a good trade-off between missing data and number of useable stations for testing the performance of our models. A map showing the percentages of missing values at each station for the year 2004 can be found in Appendix A (online version).

2.3. Statistical methods

We incorporated spatio-temporal correlations between observations and estimated missing values in the recorded time series of understory temperature by using (i) local space-time kriging (STK, hereafter) (Cressie and Wikle, 2011; Heuvelink and Griffith, 2010) and (ii) empirical orthogonal functions (EOF, hereafter) (Beckers and Rixen, 2003; Lorenz, 1956). These two statistical approaches were chosen to compare a geostatistical (STK) to a deterministic (EOF) technique, and because of their applicability to different fields of study, such as oceanography and meteorology (Alkuwari et al., 2013; Beckers and Rixen, 2003; Hengl et al., 2012; Lorenz, 1956; Youzhuan et al., 2008; Yu and Chu, 2010). Deterministic techniques, compared to geostatistical ones, do not make use of any *a priori* hypothesis based on some probability distributions and, hence, no statistical test (Cressie, 1993). An outline of the theory, main assumptions, and modeling settings used in each technique follows.

2.3.1. Local space-time kriging

Consider a continuous variable Z, e.g. temperature, varying over a spatial domain S and a time interval T. Let $z(\boldsymbol{s}_i, t_i), i = 1,2, \dots, n$ be a set of n data observed at a finite set of locations in space and points in time, where $\boldsymbol{s}$ is a vector of spatial coordinates, $\{\boldsymbol{s} = (x, y)\}$, and t represents a series of points in time. In a space-time geostatistical framework, unobserved values $z(\boldsymbol{s}_0, t_0)$ are typically predicted at a number of nodes $(\boldsymbol{s}_0, t_0)$ of a spatio-temporal grid. Predictions are made by exploiting the spatio-temporal correlation between observed locations $z(\boldsymbol{s}_i, t_i)$ using techniques such as kriging (Cressie, 1993). Commonly used kriging models include space-time ordinary kriging and universal kriging, also known as kriging with an external drift (Hengl et al., 2012). Kriging requires directly estimating spatio-temporal covariances or, more commonly, the semivariances between observed values using spatio-temporal variograms as follows:

$$\gamma(h,u) = \frac{1}{2n(h,u)} \sum_{i=1}^{n(h,u)} [z(\boldsymbol{s}_i, t_i) - z(\boldsymbol{s}_i + h, t_i + u)]^2,$$

where γ measures the average dissimilarity between data separated by a given spatial and temporal lag (h, u), where h is the Euclidean spatial distance $|\mathbf{h}|$ and u is the time interval.

A diverse range of models have been proposed to capture the structure of spatio-temporal autocorrelation, including the product model (Rodriguez-Iturbe and Mejia, 1974), the metric model (Dimitrakopoulos and Luo, 1994), the product-sum model (De Cesare et al., 2001), and the sum-metric model (Heuvelink et al., 2012). The sum-metric model was adopted for this study because it handles the space-time interaction in a flexible way, without imposing symmetry constraints between the spatial and temporal correlation components. The sum-metric variogram structure is defined as:

$$\gamma(h, u) = \gamma_S(h) + \gamma_T(u) + \gamma_{ST}\left(\sqrt{h^2 + (\alpha \cdot u)^2}\right), \qquad (1)$$

where $\gamma(h, u)$ represents the semivariance for h and u units of spatial and temporal distance, respectively. $\gamma_S(h)$ describes the purely spatial components, while $\gamma_T(u)$ describes the purely temporal component. The space-time interaction component is described by $\gamma_{ST}(h, u)$, where the geometric anisotropy between space and time, i.e. the range variation in different dimensions, is handled by the parameter α, which converts units of temporal distances into units of spatial distance (Kilibarda et al., 2014). In local space-time kriging, the spatio-temporal covariance function is evaluated only for the "strongest" neighbors of a prediction point, i.e. only the first n number of observations with the strongest correlation are used, with n assigned by the user.

It is necessary to remove large-scale spatial trends and seasonality prior to investigating the spatio-temporal covariance structure of the data because space-time kriging assumes stationarity and spatial isotropy (Kilibarda et al., 2014). For this purpose, a loess ("locally-weighted scatterplot smoothing") smoothing curve (Cleveland and Devlin, 1988) was applied separately for each station and residuals of each temperature time series were used for local space-time kriging (e.g. Fig. 2).

Figure 2-caption at the end of file

All spatio-temporal kriging models were implemented using the gstat (Pebesma, 2004), spacetime (Pebesma, 2012), and stats (R Core Team, 2013) R packages.

2.3.2. Empirical orthogonal functions

The spatio-temporal correlation structure of a dataset may also be described by a set of orthogonal functions, called empirical orthogonal functions. Let $T_n(t)$ represent the temperature values recorded at N stations as a function of time. Assuming these values are observed at M times, $t_1, t_2, \ldots, t_M$, it is possible to expand $T_n(t)$ as follows:

$$T_n(t_i) = \sum_{k=1}^{N} Y_{kn} Q_k(t_i)$$

where Y_{kn} represent the time-independent basis functions, i.e. EOF, and $Q_k(t_i)$ represent time-dependent coefficients or weights. Standard singular value decomposition (SVD) techniques (Klema and Laub, 1980) can be applied to the spatio-temporal dataset matrix to generate a set of EOF, where the first orthogonal components contain the bulk of the variance and explain the dominant patterns of spatio-temporal variation (Beckers and Rixen, 2003). The leading components are most likely to describe large-scale spatio-temporal patterns, while the latter ones might contain a mix of local-scale patterns and instrument noise (Henn et al., 2013). EOF can be considered as a set of optimally defined functions of space with associated temporal weights at each time. However, SVD cannot be used when a dataset matrix contains missing data. To overcome this issue, Beckers and Rixen (2003) developed a parameter-free iterative estimation technique to reconstruct both missing data and the complete EOF. Missing data are first replaced by an unbiased guess, i.e. the overall dataset mean, and then iteratively estimated by using a truncated series of EOF until reaching convergence. A detailed description of the estimation algorithm can be found in Appendix B (online version). EOF have been widely applied in both oceanography and meteorology (Beckers and Rixen, 2003; Lorenz, 1956; Youzhuan et al., 2008; Yu and Chu, 2010), as well as used in statistical downscaling methods in geophysics (Alkuwari et al., 2013). A recent study applied the EOF reconstruction technique to estimate missing values in air temperature datasets (Henn et al., 2013). The truncated EOF algorithm was implemented using a set of R functions (R Core Team, 2013), following the method proposed by Beckers and Rixen (2003).

2.4. Missing data scenarios

We developed three scenarios of missing data to assess the influence of spatial coverage and temporal completeness on the efficacy of each statistical technique for filling data gaps in space and time. Specifically, we artificially altered the number of sampling locations (sampling size/density), the number of randomly missing observations, and serial missingness of observations (consecutively missing observations at different times and at different locations). Each scenario is described in further detail in the following sections.

2.4.1. Sampling size/density scenario

We examined the case in which the available network of microclimate stations has a lower spatial density. The complete set of available microclimate stations (n = 200, 0.57 stations/km^2) was artificially altered by removing an increasing number of locations. Specifically, we examined reductions in sampling size of 25% (n = 150, 0.43 stations/km^2), 50% (n = 100, 0.29 stations/km^2), and 75% (n = 50, 0.14 stations/km^2). In each case, stations were removed in a randomized fashion (see Fig. 3 A-D). We reconstructed microclimate measurements only within these spatially reduced datasets, without any attempt to spatially interpolate values at locations that were removed from the original dataset.

Figure 3-caption at the end of file

2.4.2. Gap length and amount of missing values scenario

We also examined cases where the missing data may not be stations but temporal gaps instead. These temporal gaps may be in the form of either random occurrence (random missingness) or a long section of consecutively missing values (seasonal missingness). To examine these two types of missing data the available spatio-temporal dataset matrix was either artificially altered by randomly removing 50% of the observations (random missingness, Fig. 4C), or randomly removing three consecutive months of observations (seasonal missingness, Fig. 4B).

Figure 4-caption at the end of file

2.5. Performance metrics

We evaluated each statistical technique in terms of prediction accuracy, ignoring missing values in the dataset during model evaluation. A 10-fold cross validation was carried out, where a single subsample was retained as the validation data for testing the model, while using the remaining portion for model training. The following prediction metrics were quantified in order to compare the original data to model predictions:

Root-mean-square error (RMSE):

The root-mean-square error is defined as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}[\hat{T}(\boldsymbol{s}_{i,}t_i) - T(\boldsymbol{s}_{i,}t_i)]^2},$$

where $\hat{T}(\boldsymbol{s}_{i,}t_i) - T(\boldsymbol{s}_{i,}t_i)$ represents the difference between the predicted and observed temperature at space-time points $(\boldsymbol{s}_{i,}t_i)$ and m is the length of the time series of observations for each station. The root-mean-square error was also used in the iterative EOF reconstruction algorithm as a criterion to determine the optimal number of EOF for minimizing the error (see Appendix B, online version, for more details).

Mean absolute error (MAE):

The mean absolute error is a simple arithmetic average of the absolute errors and is defined as follows:

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{T}(\boldsymbol{s}_{i,}t_i) - T(\boldsymbol{s}_{i,}t_i)\right|$$

Mean-square-error skill score:

A skill score measures the forecast accuracy with respect to the accuracy of a reference forecast. Positive values correspond to a skill, while negative ones correspond to no skill. The mean-square-error skill score (SS_{MSE}) is defined as follows:

$$SS_{MSE} = 1 - \frac{MSE}{MSE_{ref}},$$

where MSE is defined as the quantity within the square root in the RMSE formula above. SS_{MSE} was calculated by using the observed average as baseline reference (Murphy, 1988).

Correlation coefficient (COR):

Perhaps the simplest overall measure of performance, the correlation coefficient is defined as:

$$COR = \frac{Cov(\hat{T}(\boldsymbol{s}_{i,}t_i), T(\boldsymbol{s}_{i,}t_i))}{\boldsymbol{s}_{\hat{T}}\boldsymbol{s}_T},$$

where $\boldsymbol{s}_{\hat{T}}$ and $\boldsymbol{s}_T$ indicate the standard deviations of predicted and observed temperature values, respectively. The correlation coefficient measures the linear association between forecast and observation. However, it only performs well when data are normally distributed and it is extremely sensitive to large values and outliers (Taylor, 2001).

Both the RMSE and MAE disregard the direction of over- or under- prediction. All four metrics were averaged over the total number of available stations to come up with an overall measure of accuracy. However, these can also be calculated and mapped separately for each station to assess where the statistical techniques had higher/lower accuracy (Fig. C.1-C.2, Appendix C, online version). In order to analyze the impact of temporal aggregation on prediction accuracy the original dataset and modeled predictions were aggregated from an hourly to daily resolution (daily mean, maximum, and minimum). Days for which this aggregation process did not remove missing values were ignored in the calculation of both performance metrics.

3. Results

3.1. Exploratory Data Analysis

In order to inspect the degree to which pairs of time series are correlated, we selected a group of microclimate stations in close proximity (Fig. 5A) and looked at the cross-correlation function (CCF) (Berezin et al., 2012).

Figure 5-caption at the end of file

Similar patterns of cross-correlations (Fig. 5B) suggested the presence of a strong spatial dependence among stations in close proximity. The extended correlation over time can perhaps be explained by the buffering effect that forest canopies have on the hourly temperature measurements. This preliminary analysis was replicated on other groups of stations and revealed approximately the same patterns of spatio-temporal dependency.

3.2. Local space-time kriging for hourly temperature

Residuals from loess-smoothed hourly temperature data show a clear spatio-temporal correlation pattern (Fig. 6A), following the main hypothesis of space-time variograms, i.e. the spatial structure becomes weaker as the time differences increase (Fig. 6B). Therefore, spatio-temporal kriging of residuals is applicable.

Figure 6-caption at the end of file

The three components of the sum-metric variogram model and their parameters (Table 1) were chosen based on the combination that gave the best accuracy in predicting the observed hourly temperature.

Table 1-caption at the end of file

The parameters show that all components are needed to capture the residual spatio-temporal pattern in the loess-smoothed time series of hourly temperature. The range parameters in both the purely spatial and joint space-time components are very large, indicating that the residuals are correlated over distances up to ~240 km. The highest prediction accuracy for the local space-time kriging technique was reached when setting the maximum number of neighbors equal to 10 (i.e. using the 10 strongest correlated). The spatio-temporal anisotropy (α = 2.96 m/hour) shows that stations with a temporal lag of 1 day exhibit a similar correlation as stations that are about 70 meters (2.96 * 24) apart.

3.3. Empirical orthogonal functions for hourly temperature

The iterative EOF estimation routine (Appendix B, online version) suggested that the optimal number of orthogonal functions to use with the hourly temperature measurements is equal to nine (Fig. 7). The RMSE starts leveling out after the first six components until reaching a minimum for nine EOF. This number is considered as the optimal number of EOF needed to explain most spatio-temporal variability found in the hourly temperature dataset. The first EOF alone explained 55% of the dataset variance, and the first nine together explained about 70%.

Figure 7-caption at the end of file

3.4. Comparison of model predictions

Both STK and EOF predicted hourly temperatures accurately (Fig. 8A-B), with a correlation of about 0.98 between predicted and observed values for each method. A slightly higher heteroscedasticity can be observed for STK (Fig. 8A).

Figure 8-caption at the end of file

We selected two temporal windows from the hourly time series of a representative microclimate station to show the resulting predictions for both statistical techniques. The reconstructed time series of hourly temperatures shows some differences between the two models (Fig. 9A-B). The extended gap length likely affected model predictions over the missing portion of the time series (Fig. 9A). The reconstructed values show a similar variability between the two methods, with STK predicting lower hourly temperature values compared to EOF. The high prediction accuracy of both techniques is confirmed by looking at the reconstructed values over a portion of available data (Fig. 9B). Although similar in their accuracy, the EOF technique predicted the observed time series better compared to STK.

Figure 9-caption at the end of file

Performance differences between STK and EOF were more pronounced at the hourly temporal resolution compared to the chosen daily aggregations (mean, max, min), with magnitudes depending on the missing data scenario used (Fig. 10). The model performance measured in terms of MAE, MSE skill score (SS_{MSE}), and COR confirmed a similar trend to that observed for the RMSE (Appendix D, online version). Overall, the EOF technique predicted the observed time series of data more accurately (lower RMSE and MAE, higher COR and SS_{MSE}) than STK regardless of the temporal aggregation. The accuracy of both modeling methods converged as the sampling size/density was reduced, indicating that the performance of EOF degrades rapidly as spatial coverage decreases. Conversely, STK was less sensitive to the reduction of sampling density with only slight decreases in model accuracy. At the daily resolution, prediction accuracy was almost identical when 75% of the stations were removed, however, at an hourly resolution EOF provided greater accuracy. EOF was affected by the random and seasonal temporal missingness scenarios to a lesser extent than STK.

Figure 10-caption at the end of file

4. Discussion and Conclusions

Our focus was to accurately reconstruct incomplete forest microclimate measurements rather than inspecting the relative importance of ecological and physiographic variables on microclimate dynamics (e.g. Vanwalleghem and Meentemeyer, 2009).

Results indicate that a reduction in sampling size/density has a greater effect on EOF model predictions than temporal missingness. In contrast, STK was more affected by temporal missingness compared to EOF, with seasonal gaps (ss_noise) reducing STK prediction accuracy more than the presence of a larger number of randomly missing data (rnd_noise). The lower performance of the space-time kriging technique in these settings may be explained by a lack of stationarity and spatial isotropy of the spatio-temporal covariance structure (Kilibarda et al., 2014). We speculate that additional temporal gaps contributed to degrading stationarity and spatial isotropy. Stationarity assumptions are likely inaccurate when evaluating the raw (hourly) temperature data, leading to less accurate predictions. We tried to address this issue by using residuals of loess-smoothed time series, separately for each station, and by using a flexible space-time variogram model. In contrast, EOF offers a convenient method for characterizing dominant spatial patterns of variability by exploiting the spatio-temporal covariance structure without making any assumptions on the probabilistic distribution of data. The effect of the number of stations for the automatic EOF reconstruction routine has been demonstrated in a previous study, where the method performed best with more than 16 stations (Henn et al., 2013). Temporal aggregation reduced the error in each modeling technique, as well as the differences in the accuracy between them, with the lowest RMSE values for both EOF and STK observed when modeling the daily mean. The smaller differences between model accuracies at the daily resolution can be explained by fewer missing values left in the dataset after aggregation.

Depending on the characteristics of their dataset, Fig. 10 can instruct practitioners on the different trade-offs that need to be considered when choosing a method for filling in missing temperature data recorded by understory microclimate stations. On the basis of the above results, it was clear that the station density, the amount of missing values, and the length of data gaps affected the performance of the chosen statistical methods. Although people have successfully used kriging with as few as seven data points (Jernigan, 1986), successful applications also depend on the extent of the study area over which they are distributed. On the one hand, a general rule of thumb in the literature appears to be around a minimum of 30 stations (ASTM Standard D5922, 2010). On the other hand, although a large amount of data typically improves the predictive power of space-time kriging, it can also pose computational challenges due to the *big n* problem (Banerjee et al., 2004). Whenever the primary goal is to predict temperature data at unobserved locations, space-time kriging represents the most common and immediate solution to spatially interpolate observed data and have an associated measure of prediction uncertainty. The use of space-time kriging requires a fair amount of time to calibrate all the parameters and tune the model (see section 3.2), thus representing a limitation compared to the automatic EOF estimation routine.

4.1. Guidelines for reconstructing missing forest microclimate measurements

We herein summarize general “rules-of-thumb” and trade-offs between the two statistical approaches in order to guide method selection:

- EOF should be preferred over STK for highly correlated hourly observations. The increase of temporal aggregation levels (e.g. daily) resulting in a smaller dependence among observations reduces discrepancies between predictive performance of EOF and STK.
- Use EOF when dealing with either random or consecutive seasonal patterns of temporal gaps in the observations. Discrepancies between the predictive performance of both modeling approaches decrease when increasing temporal aggregation.

- Use STK when interpolating values at unobserved locations. Although this methodology has been successfully applied to small numbers of ground stations over small spatial extents, a minimum of 30 stations should be used as a rule-of-thumb. EOF would simply apply the mean value of all the observations to the unobserved locations, thus not capturing physical influences.
- STK should be preferred over EOF for sparse networks of ground stations (< 50 or 0.14 stations/ km2) with few temporal gaps (preferably random) in the available observations.

We demonstrated methods to reconstruct hourly time series of microclimate data by exploiting the spatio-temporal correlation between microclimate sensors placed under forest canopy and compared the predictive accuracies of two spatio-temporal statistical techniques, a geostatistical (STK) and a deterministic one (EOF), in reconstructing hourly time series of data. To the best of our knowledge, this is the first study to quantify in a comprehensive way the performance of both methods at a landscape-level based on several missing data scenarios as well as the impact of temporal aggregation. A dense network of 200 microclimate stations allowed us to analyze the impact that sampling size/density, the overall amount of missing data, and the length of data gaps had on model predictions. The framework presented herein could be used to assimilate multiple data sources measured at different temporal resolutions providing an avenue for integrating key aspects of fine-scale spatial heterogeneity into ecosystem studies.

Acknowledgements

The authors would like to thank everybody in the Meentemeyer Landscape Dynamics Lab at the Center for Geospatial Analytics for their feedback and valuable suggestions on the present work; the Sonoma State University's Fairfield Osborn Preserve, the Sonoma Mountain Ranch Preservation Foundation, the Sonoma Agriculture and Open Space District, the Sonoma Land Trust, California State Parks, Sonoma County Regional Parks, as well as multiple private land owners for granting us access to their lands to collect the data; Brian Anacker and Deanne DiPietro who were instrumental in installing microclimate stations. This research was supported by grants from the National Science Foundation (DEB-EF-0622677 and EF-0622770) as part of the joint NSF-NIH Ecology of Infectious Disease program. We also gratefully acknowledge financial support from the United States Department of Agriculture Forest Service-Pacific Southwest Research Station.

References

Alkuwari, F.A., Guillas, S., Wang, J., 2013. Statistical downscaling of an air quality model using Fitted Empirical Orthogonal Functions. Atmos. Environ. 81, 1-10.

ASTM Standard D5922, 2010. Analysis of spatial variation in geostatistical site investigations. ASTM International, West Conshohocken, PA, USA.

Band, L.E., Peterson, D.L., Running, S.W., Coughlan, J., Lammers, R., Dungan, J., et al., 1991. Forest ecosystem processes at the watershed scale: basis for distributed simulation. Ecol. Model. 56, 171-196.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical modeling and analysis for spatial data. Chapman & Hall/CRC, Boca Raton, FL, USA.

Barbour, M.G., Billings, W.D., 2000. North American terrestrial vegetation. Cambridge University Press, Cambridge, U.K., 695 pp.

Beckers, J.M., Rixen, M., 2003. EOF calculations and data filling from incomplete oceanographic datasets. J. Atmos. Oceanic Technol. 20, 1839-1856.

Berezin, Y., Gozolchiani, A., Guez, O., Havlin, S., 2012. Stability of climate networks with time. Sci. Rep. 2, 666.

Bolstad, P.V., Swift, L., Collins, F., Regniere, J., 1998. Measured and predicted air temperatures at basin to regional scales in the southern Appalachian mountains. Agric. For. Meteorol. 91, 161-176.

Caffarra, A., Rinaldi, M., Eccel, E., Rossi, V., Pertot, I. 2012. Modelling the impact of climate change on the interaction between grapevine and its pests and pathogens: European grapevine moth and powdery mildew. Agr. Ecosyt. Environ, 148, 89-101.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: An approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83, 596-610.

Cressie, N., 1993. Statistics for spatial data. John Wiley & Sons, Inc., New York, NY, U.S.A.

Cressie, N., Wikle, C.K., 2011. Statistics for spatio-temporal data. John Wiley & Sons, Inc., Hoboken, New Jersey, 571 pp.

Daly, S.F., Davis, R., Ochs, E., Pangburn, T., 2000. An approach to spatially distributed snow modeling of the Sacramento and San Joaquin basins, California. Hydrol. Processes 14, 3257-3271.

De Cesare, L., Myers, D.E., Posa, D., 2001. Estimating and modeling space-time correlation structures. Stat. Probabil. Lett. 51, 9-14.

Dimitrakopoulos, R., Luo, X., 1994. Spatiotemporal modeling: covariances and ordinary kriging systems. In: Dimitrakopoulos, R. (ed.), Geostatistics for the next century. Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 88-93.

Garen, D.C., Johnson, G.L., Hanson, C.L., 1994. Mean areal precipitation for daily hydrologic modeling in mountainous regions. J. Amer. Water Resour. Assoc. 30, 481-491.

Gehlhausen, S.M., Schwartz, M.W., Augspurger, C.K., 2000. Vegetation and microclimatic edge effects in two mixed-mesophytic forest fragments. Plant Ecol. 147, 21-35.

Geiger, R., 1965. The climate near the ground. Harvard University Press, Cambridge, MA, U.S.A.

Hengl, T., Heuvelink, G.B.M., Tadic, M.P., Pebesma, E., 2012. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. Theor. Appl. Climatol. 107, 265-277.
Henn, B., Raleigh, M.S., Fisher, A., Lundquist, J.D., 2013. A comparison of methods for filling gaps in hourly near-surface air temperature data. J. Hydrometeor. 14, 929-945.
Heuvelink, G.B.M., Griffith, D.A., 2010. Space-time geostatistics for geography: A case study of radiation monitoring across parts of Germany. Geographical Analysis 42, 161-179.
Heuvelink, G.B.M., Griffith, D.A., Hengl, T., Melles, S.J., 2012. Sampling design optimization for space-time kriging. In: Mateu, J., Müller, W.G. (eds.), Spatio-temporal design: advances in efficient data acquisition. John Wiley & Sons, Ltd, Oxford, U.K., pp. 207-230.
Jernigan, R.W., 1986. A primer on kriging. U.S. Environmental Protection Agency, Washington, D.C., U.S.A.
Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E., Tadic, M.P., et al., 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J. Geophys. Res. Atmos. 119, 2294-2313.
Klema, V., Laub, A.J., 1980. The singular value decomposition: Its computation and some applications. IEEE Trans. Autom. Control 25, 164-176.
Lookingbill, T., Urban, D.L., 2003. Spatial estimation of air temperature differences for landscape-scale studies in montane environments. Agric. For. Meteorol. 114, 141-151.
Lorenz, E.N. 1956. Empirical orthogonal functions and statistical weather prediction. Scientific Report No. 1, Cambridge, MA, U.S.A.
McDonald, R.I., Urban, D.L., 2004. Forest edges and tree growth rates in the North Carolina Piedmont. Ecol. 85, 2258-2266.
Meentemeyer, R.K., 1978. Macroclimate and lignin control of litter decomposition rates. Ecol. 59, 465-472.
Meentemeyer, R.K., Haas, S.E., Vaclavik, T., 2012. Landscape epidemiology of emerging infectious diseases in natural and human-altered ecosystems. Ann. Rev. Phytopathol. 50, 379-402.
Murphy, A.H. 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Mon. Wea. Rev. 116, 2417-2424.
Myers, D.E., 1992. Kriging, cokriging, radial basis functions and the role of positive definiteness. Comput. Math. Appl. 24, 139-148.
Pape, R., Wundram, D., Löffler, J., 2009. Modelling near-surface temperature conditions in high mountain environments: an appraisal. Clim. Res. 39, 99-109.
Pebesma, E., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683-691.
Pebesma, E., 2012. Spacetime: Spatio-temporal data in R. J. Stat. Software 51, 1-30.
R Core Team, 2013. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Raible, C.C., Bischof, G., Fraedrich, K., Kirk, E., 1999. Statistical single-station short-term forecasting of temperature and probability of precipitation: Area interpolation and NWP combination. Wea. Forecasting 14, 203-214.
Rodriguez-Iturbe, I., Mejia, J.M., 1974. Design of rainfall networks in time and space. Water Resour. Res. 10, 713-728.

Swei, A., Meentemeyer, R.K., Briggs, C. J. 2011. Influence of abiotic and environmental factors on the density and infection prevalence of *Ixodes pacificus* (Acari: Ixodidae) with Borrelia burgdorferi. J. Med. Entomol. 48, 20-28.
Taylor, K.E. 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. 106, 7183-7192.
Thomas, C. S., Gubler, W. D., and Leavitt, G. 1994. Field testing of a powdery mildew disease forecast model on grapes in California. Phytopathology. 84, 1070.
Tobin, C., Nicotina, L., Parlange, M.B., Berne, A., Rinaldo, A., 2011. Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region. J. Hydrol. 401, 77-89.
Turner, M.G., Chapin, F.S., 2005. Causes and consequences of spatial heterogeneity in ecosystem function. In: Lovett, G.M., Jones, C.G., Turner, M.G., Weathers, K.C. (eds.), Ecosystem function in heterogeneous landscapes. Springer, New York, pp. 9-30.
Vanwalleghem, T., Meentemeyer, R.K., 2009. Predicting forest microclimate in heterogeneous landscapes. Ecosystems 12, 1158-1172.
Waring, R.H., Running, S.W., 1998. Forest ecosystems: analysis at multiple scales, 2nd ed. Academic Press, San Diego, CA, 370 pp.
Youzhuan, D., Dongyang, F., Zhihui, W., Zhihua, M., Juhong, Z., 2008. Reconstruction of incomplete satellite oceanographic data sets based on eof and kriging methods. Proc. SPIE Int. Soc. Opt. Eng. Image and Signal Processing for Remote Sensing XIV
Yu, H.-L., Chu, H.-J., 2010. Understanding space-time patterns of groundwater system by empirical orthogonal functions: A case study in the Choshui River alluvial fan, Taiwan. J. Hydrol. 381, 239-247.

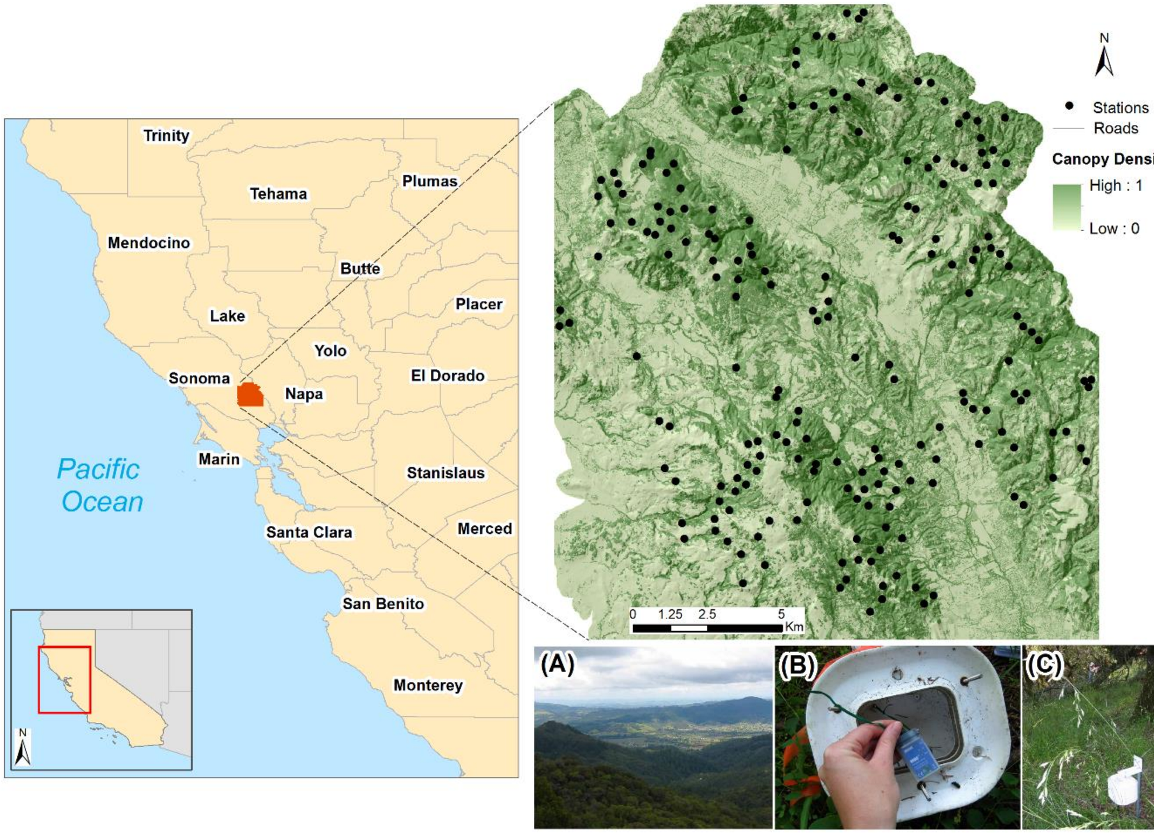


Figure 1 Overview of the Sonoma study system, California, with indication of canopy density and microclimate sensor locations. (A) Photo of the landscape characteristic of the study area. (B) The temperature-only logger installed beginning in 2008, and (C) A solar radiation shield housing the temperature logger installed in the forest understory (HOBO H8 Pro, Onset Corp., Bourne, MA, USA). Available in color online.

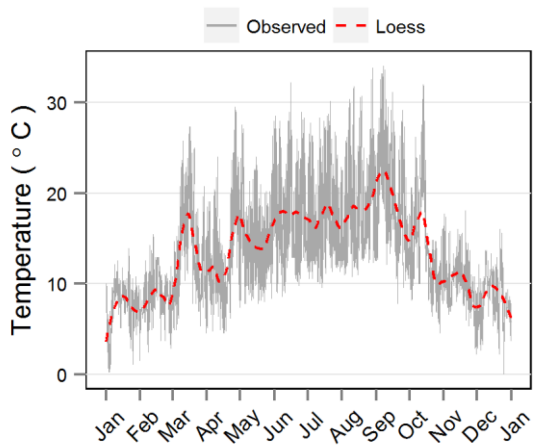


Figure 2 Hourly temperature (gray solid line) and loess smoother (span = 0.1) of hourly temperature (red dashed line) for station ANN01. Available in color online.

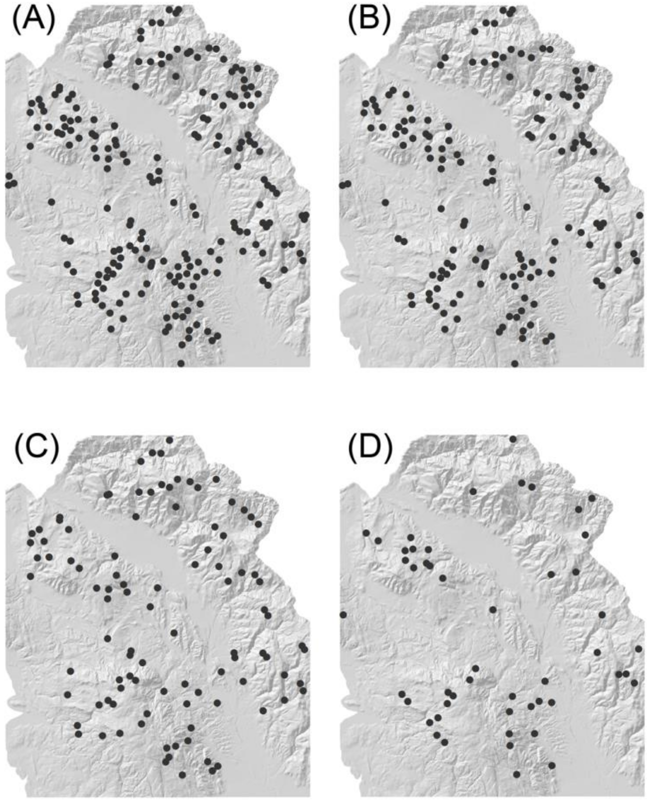


Figure 3 **Sampling size/density reduction scenarios. (A) All locations (n = 200, 0.57 stations/km^2). (B) Random removal of 25% of the stations from the original network (n = 150, 0.43 stations/km^2). (C) Random removal of 50% of the stations from the original network (n = 100, 0.29 stations/km^2). (D) Random removal of 75% of the stations from the original network (n = 50, 0.14 stations/km^2).**

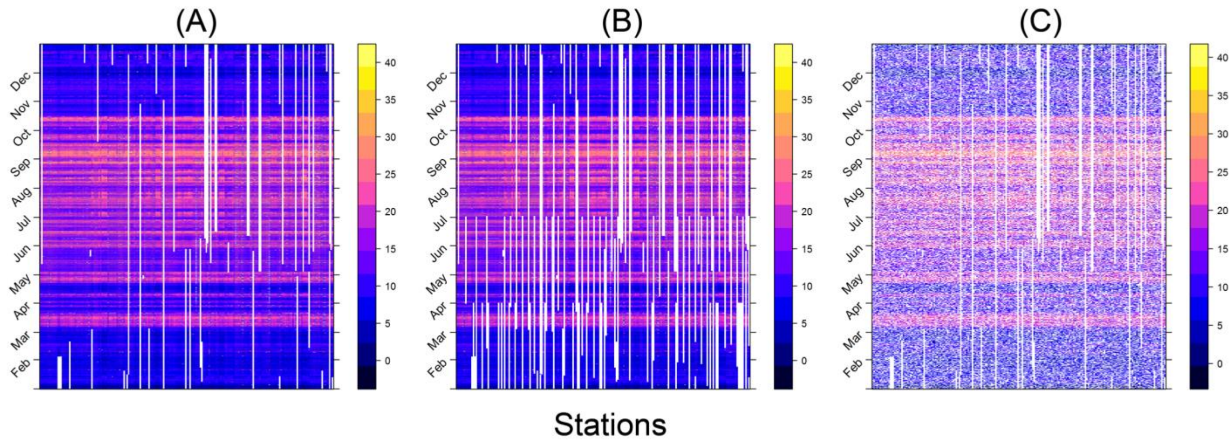


Figure 4 Data removal scenarios. (A) Original hourly temperature measurements in 2004. (B) Randomized removal of consecutive 3-month blocks of hourly temperature measurements for randomly selected stations (C) Randomized removal of 50% of hourly temperature measurements. Lighter areas correspond to higher temperature values. Completely white sections are missing data. Available in colors online.

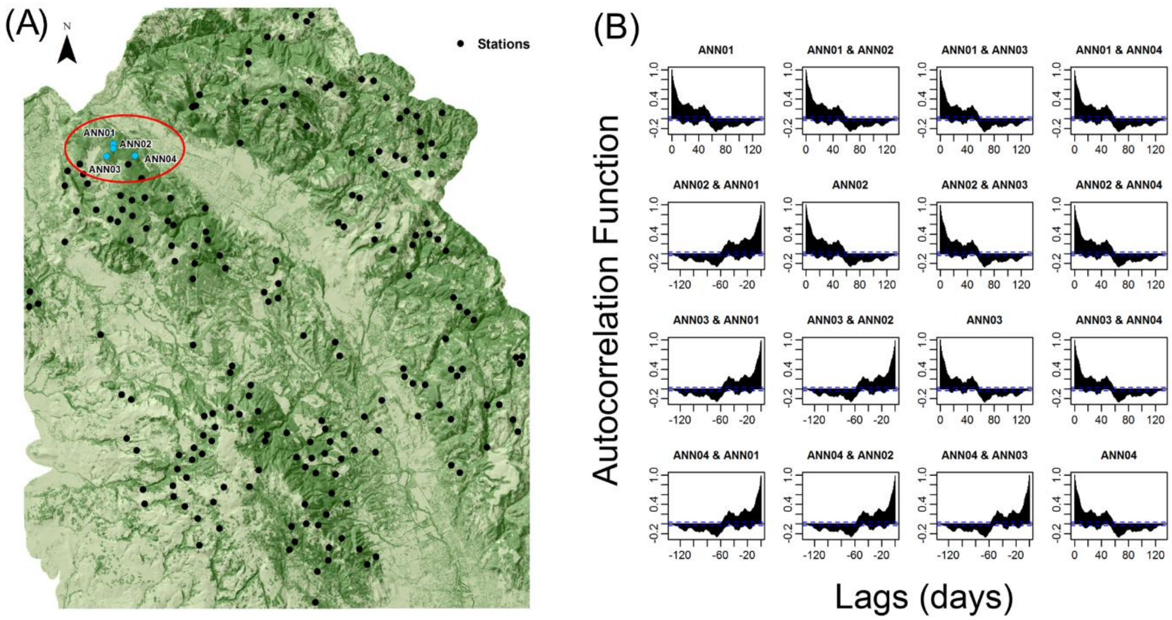


Figure 5 Temporal cross-correlation between four closely located stations within the study area. (A) Selected stations: ANN01, ANN02, ANN03, ANN04. (B) Cross-correlation plot. Available in colors online.

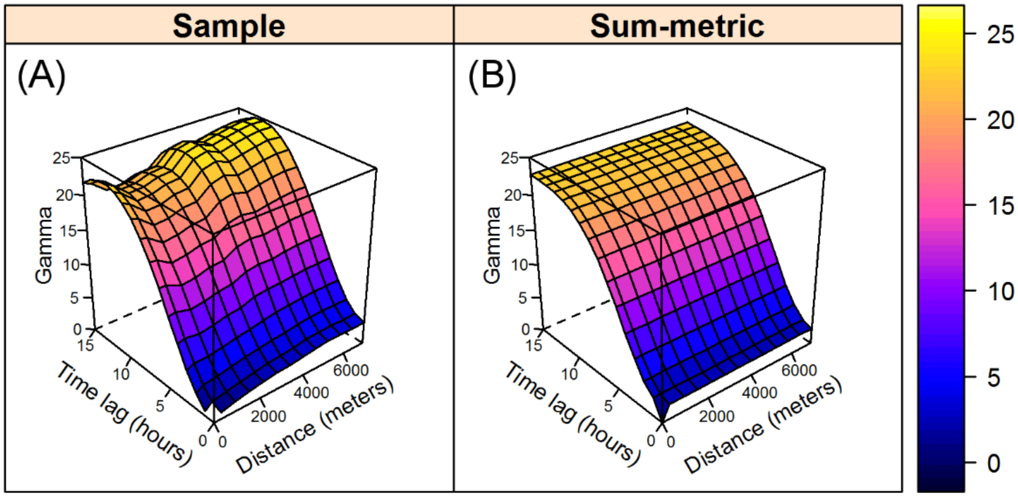


Figure 6 Sample (left) space-time variogram of residuals from loess smoothing of hourly temperature and fitted (right) sum-metric model. The variogram surface is presented in 3-D. Lighter areas correspond to higher values. Available in colors online.

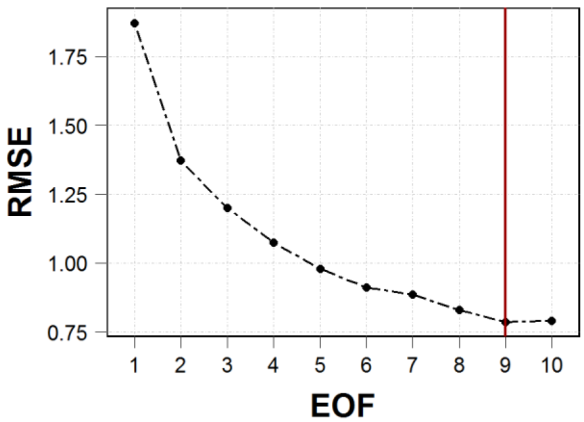


Figure 7 Root-mean-square error calculated for hourly temperature by the iterative estimation routine for an increasing number of EOF. The optimal number of EOF (n = 9) is chosen based on the convergence criteria set up in the algorithm. Available in colors online.

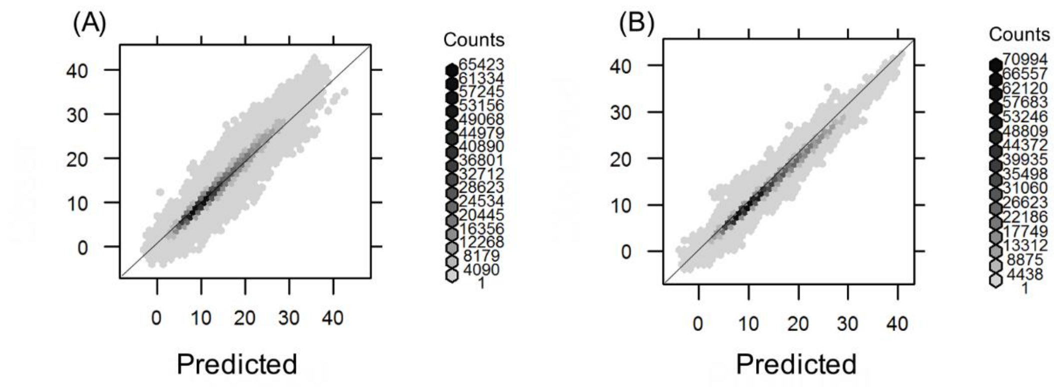


Figure 8 General relationship between observed and predicted hourly temperature. (A) Space-time kriging. (B) Empirical orthogonal functions. Cells along the diagonal are in a 1:1 relationship. Hexagonal bins are used to group points.

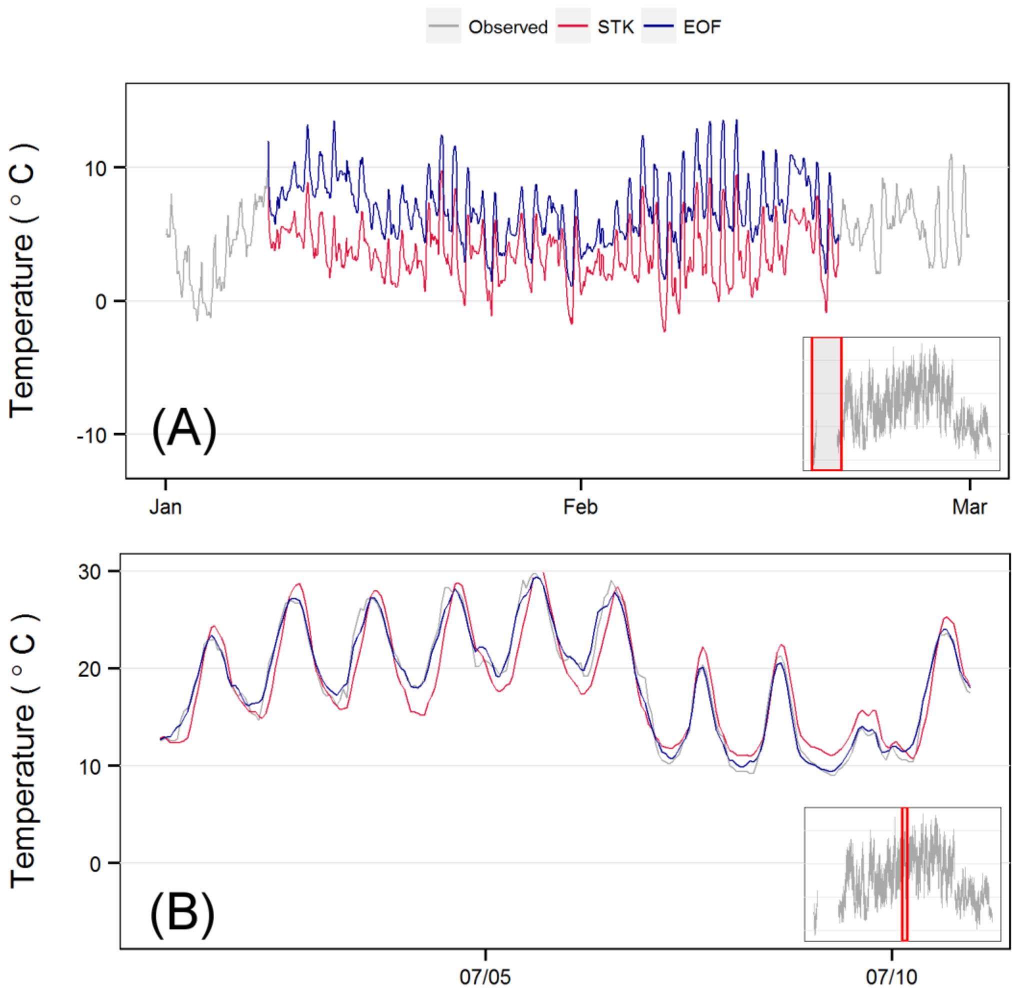


Figure 9 Hourly temperature (gray line) with overlaid predicted hourly temperature by space-time kriging (STK, red line) and empirical orthogonal functions (EOF, blue line) for station LARS01. (A) Predictions over a missing section of the dataset. (B) Predictions over a complete section of the dataset. Available in colors online.

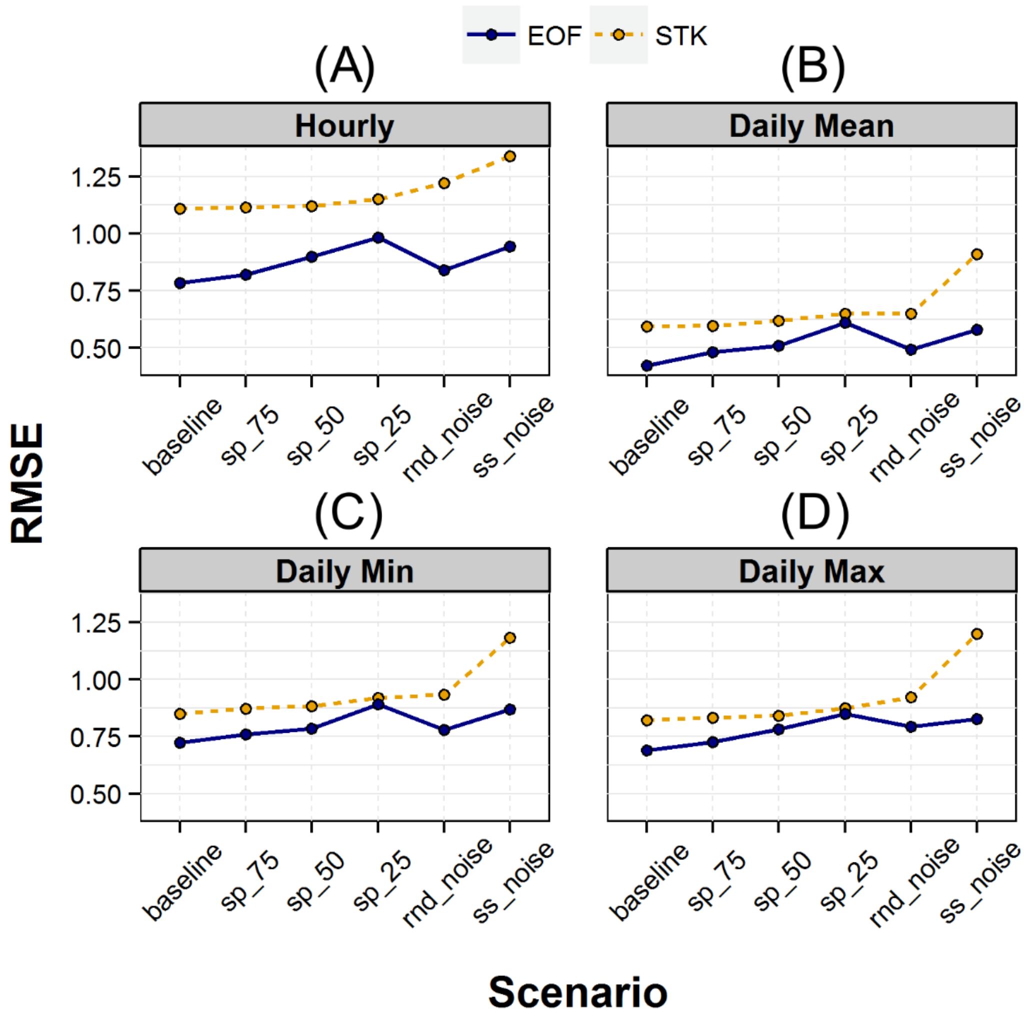


Figure 10 Root-mean-square error (RMSE) for space-time kriging (STK, golden dashed line) and empirical orthogonal functions (EOF, blue solid line) with different scenarios of missing data (see section 2.4). (A) Hourly temperature (B) Daily mean temperature. (C) Daily minimum temperature. (D) Daily maximum temperature. Available in colors online.

Table 1 Parameters of the fitted sum-metric variogram model for hourly temperature loess residuals. The model used for each component (see eq.(1)) is also specified.

	Model	Nugget (semivariance)	Partial Sill (semivariance)	Range	Anisotropy Ratio
Space	Exponential	0	7.21	243.28 km	
Time	Gaussian	0	20.74	6.28 hours	
Joint	Spherical	2.09	7.15	243.47 km	2.96 m/hour

Appendix A.

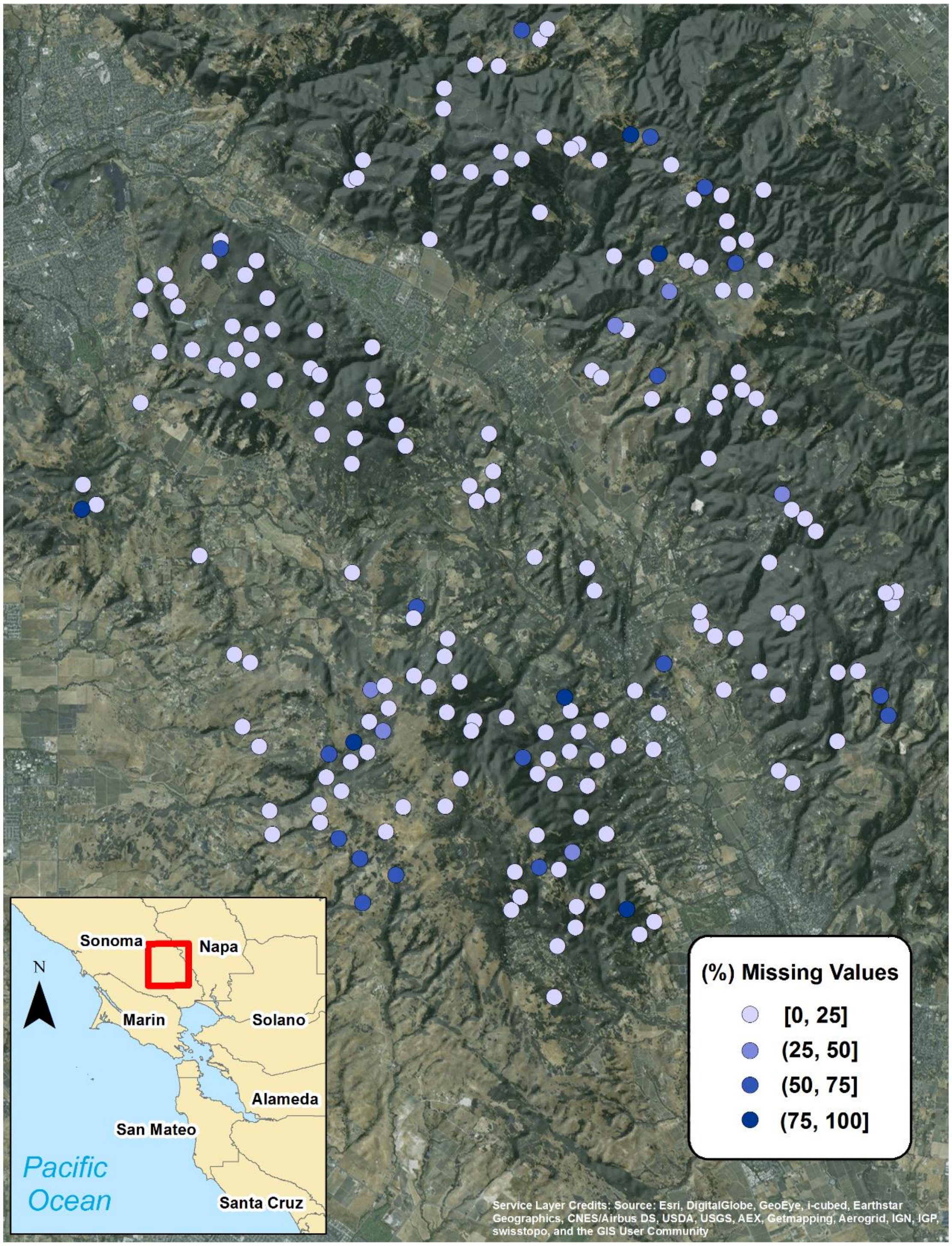


Figure A.1 Percentages of missing values in the observed understory temperature at each microclimate station for the year 2004.

Appendix B. Empirical Orthogonal Function Estimation Algorithm

The following pseudo-code algorithm is illustrated to match the R code we implemented (github link). The structure has been re-adapted from the study by Beckers and Rixen (2003) and the appendix presented in Henn et al. (2013).

$\mathbf{X} = m \times n$ (m = hours, n = stations)

1. Calculate X_mean (overall dataset mean) and X_sd (overall dataset standard deviation)

2. $\mathbf{X_0} = \mathbf{X}$ – X_mean / X_sd (standardize variable)

3. $\mathbf{X_0}$ [sub] = $\mathbf{X_0}$**_val**. Subset portion of data from $\mathbf{X_0}$ for validation. Replaced with missing values in $\mathbf{X_0}$.

4. $\mathbf{X_0}$ [sub] = missing. Values set aside for validation are replaced with missing values in $\mathbf{X_0}$.

5. $\mathbf{X_0}$ [missing] = 0; replace all missing values with unbiased guess.

6. *Outer FOR LOOP*:

    FOR *Ne* (number of EOF) = min (n, 30); Minimum between number of stations n and 30.

        $\mathbf{X_1} = \mathbf{X_0}$; Make a copy of $\mathbf{X_0}$. $\mathbf{X_1}$ will be iteratively improved within the inner loop.

7. *Inner FOR LOOP*:

        FOR k *(iteration)* = 2 to *Nit* (max number of iterations)

8.	[$\mathbf{U}$, $\mathbf{D}$, $\mathbf{V}$] = SVD($\mathbf{X_1}$);	Singular value decomposition (SVD).
	$\mathbf{U_t} = \mathbf{U}$ [,1:*Ne*];	Columns are left singular vectors of $\mathbf{X_1}$. Truncate components using first *Ne* EOFs.
	$\mathbf{D_t} = \mathbf{D}$ [1:*Ne*,1:*Ne*];	Diagonal matrix with singular values. Truncate components using first *Ne* EOFs.
	$\mathbf{V_t} = \mathbf{V}$ [,1:*Ne*];	Columns are right singular vectors of $\mathbf{X_1}$. Truncate components using first *Ne* EOFs.
9.	$\mathbf{X_a} = \mathbf{U_t}\,\mathbf{D_t}\,\mathbf{V_t}^{\mathbf{T}}$;	$\mathbf{X_a}$ is the reconstructed matrix.
10.	$\mathbf{X_a}$ [!missing] = $\mathbf{X_0}$ [!missing];	Restore original data in the estimated matrix except where missing in $\mathbf{X_0}$.
11.	dx = sum[($\mathbf{X_a}$ - $\mathbf{X_1}$)^2];	Calculate deviance of estimated matrix from $\mathbf{X_1}$.
12.	mx = sum[($\mathbf{X_a}$)^2];	Calculate deviance of estimated matrix.

13. IF dx/mx < *tol* BREAK (go to outer loop); Test for convergence.

ELSE $\mathbf{X_1} = \mathbf{X_a}$; Make a copy of $\mathbf{X_a}$ and NEXT ($k = k + 1$)

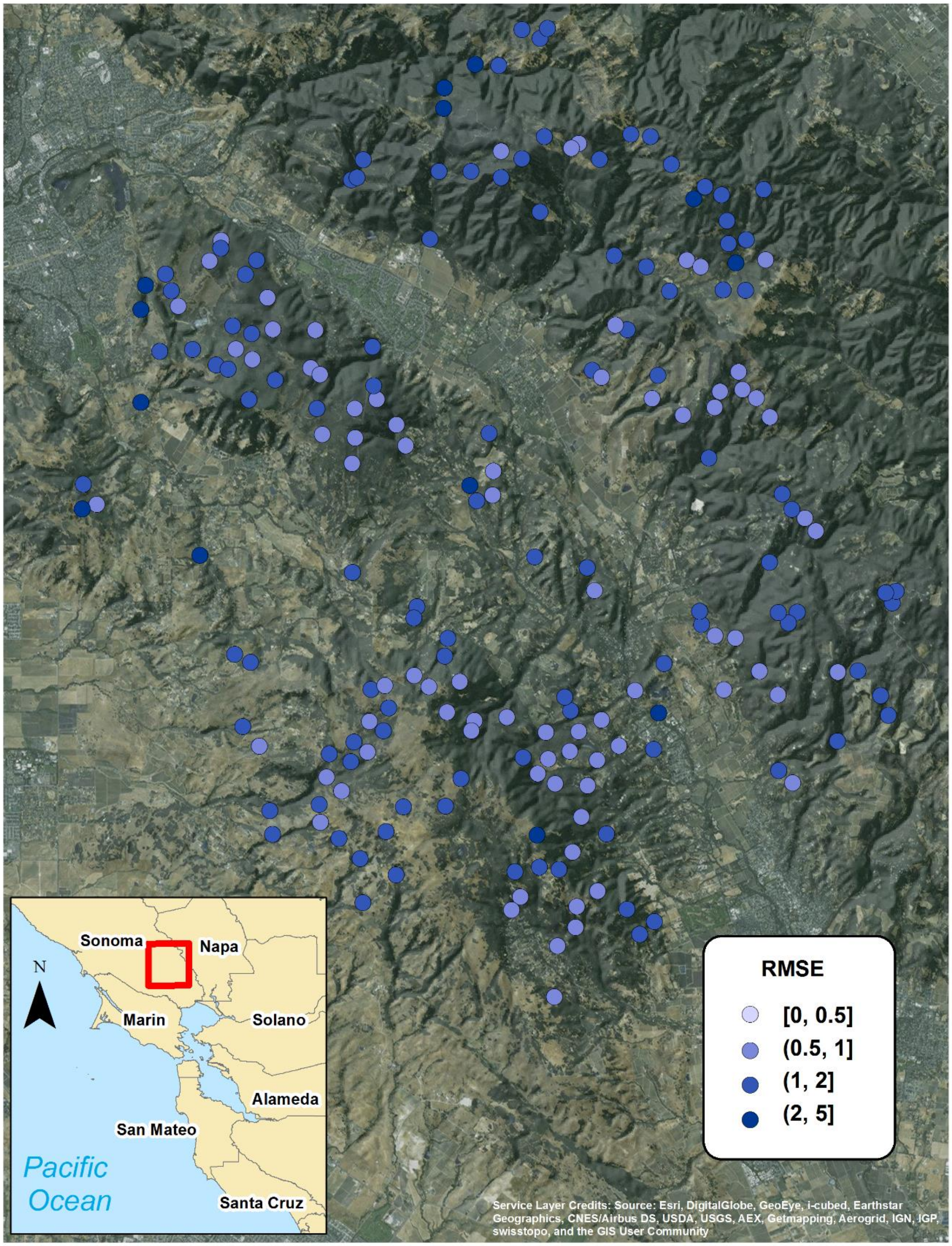
14. $RMSE[Ne] = \sqrt{(X_a[sub] - X_{0_{val}})^2}$; Calculate RMSE using *Ne* EOFs.

IF $Ne > 1$ & $(RMSE[Ne-1] - RMSE[Ne]) < 0.01$ BREAK (exit outer loop)

ELSE NEXT ($Ne = Ne + 1$); If the decrease in RMSE is almost zero, exit the outer look. Otherwise increase number of EOFs by one and restart.

END

913 **Appendix C.**



914

915 **Figure C.1 RMSE of the predicted hourly temperature at each microclimate station for the**

916 **year 2004 using STK.**

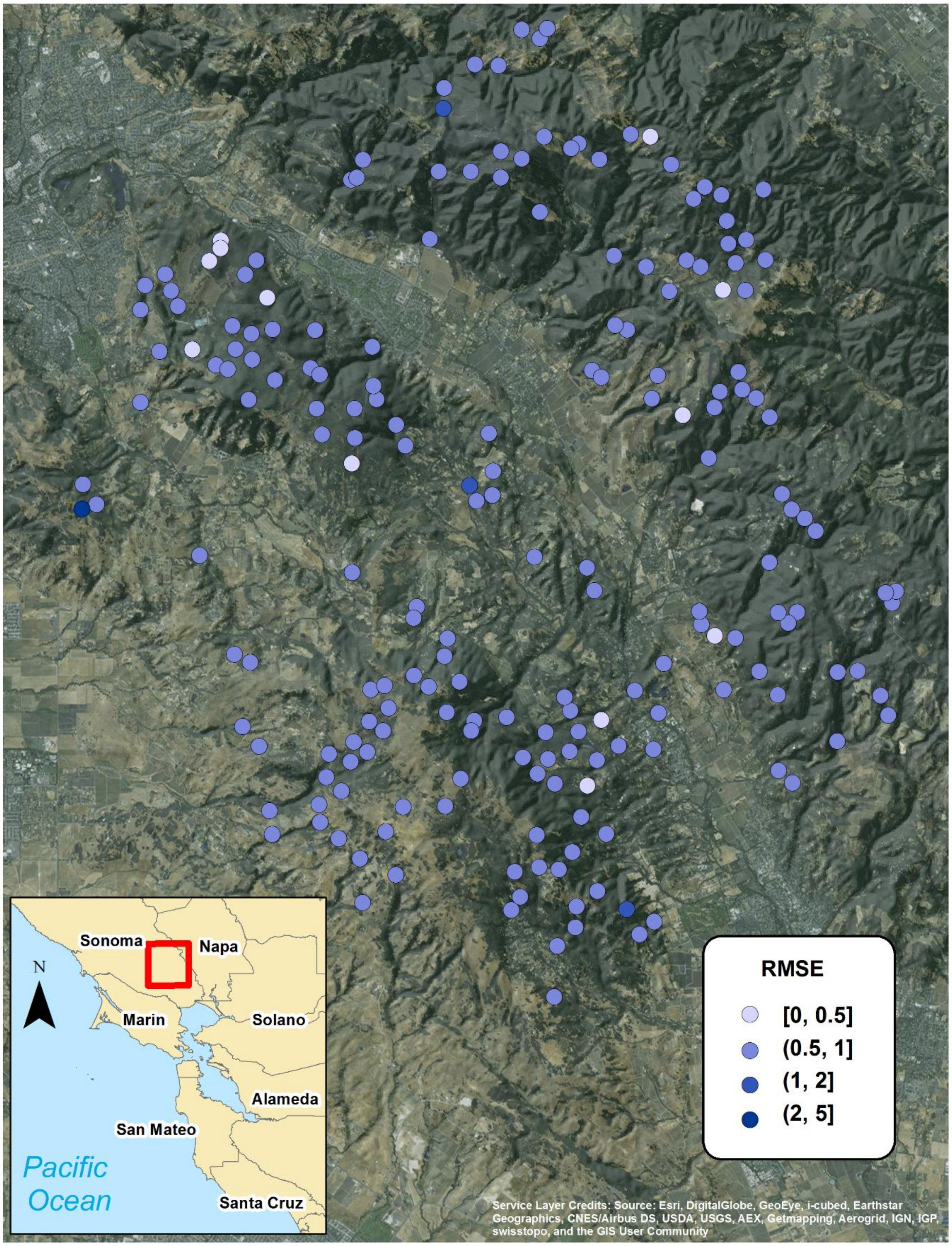


Figure C.2 RMSE of the predicted hourly temperature at each microclimate station for the year 2004 using EOF.

Table D.1 Root-mean-square error (RMSE) for space-time kriging (STK) and empirical orthogonal functions (EOF) with different scenarios of missing data at different temporal aggregations.

	HOURLY					
	baseline	sp_75	sp_50	sp_25	rnd_noise	ss_noise
EOF	0.78	0.82	0.9	0.98	0.84	0.94
STK	1.11	1.11	1.12	1.15	1.22	1.34
	DAILY MEAN					
EOF	0.42	0.48	0.51	0.61	0.49	0.58
STK	0.59	0.59	0.62	0.65	0.65	0.91
	DAILY MINIMUM					
EOF	0.72	0.76	0.78	0.89	0.78	0.87
STK	0.85	0.87	0.88	0.92	0.93	1.18
	DAILY MAXIMUM					
EOF	0.69	0.72	0.78	0.85	0.79	0.83
STK	0.82	0.83	0.84	0.87	0.92	1.20

Table D.2 Mean Absolute Error (MAE) for space-time kriging (STK) and empirical orthogonal functions (EOF) with different scenarios of missing data at different temporal aggregations.

	HOURLY					
	baseline	sp_75	sp_50	sp_25	rnd_noise	ss_noise
EOF	0.56	0.58	0.61	0.68	0.62	0.64
STK	0.93	0.92	0.93	0.96	0.97	1.02
	DAILY MEAN					
EOF	0.37	0.41	0.45	0.51	0.44	0.52
STK	0.48	0.49	0.49	0.50	0.53	0.60
	DAILY MINIMUM					
EOF	0.52	0.55	0.59	0.66	0.58	0.63
STK	0.75	0.77	0.77	0.79	0.81	0.95
	DAILY MAXIMUM					
EOF	0.48	0.52	0.58	0.64	0.61	0.65
STK	0.73	0.74	0.77	0.77	0.80	0.91

Table D.3 Correlation (COR) for space-time kriging (STK) and empirical orthogonal functions (EOF) with different scenarios of missing data at different temporal aggregations.

	baseline	sp_75	sp_50	sp_25	rnd_noise	ss_noise
			HOURLY			
EOF	0.95	0.92	0.89	0.87	0.93	0.88
STK	0.94	0.93	0.91	0.89	0.90	0.87
			DAILY MEAN			
EOF	0.97	0.94	0.92	0.91	0.92	0.90
STK	0.96	0.96	0.94	0.93	0.92	0.89
			DAILY MINIMUM			
EOF	0.93	0.91	0.89	0.85	0.91	0.88
STK	0.89	0.89	0.88	0.86	0.89	0.84
			DAILY MAXIMUM			
EOF	0.94	0.92	0.91	0.88	0.92	0.89
STK	0.92	0.92	0.89	0.89	0.88	0.85

Table D.4 Mean-square-error skill Score (SS_{MSE}) for space-time kriging (STK) and empirical orthogonal functions (EOF) with different scenarios of missing data at different temporal aggregations.

	baseline	sp_75	sp_50	sp_25	rnd_noise	ss_noise
			HOURLY			
EOF	0.98	0.97	0.95	0.92	0.93	0.90
STK	0.93	0.93	0.90	0.90	0.91	0.86
			DAILY MEAN			
EOF	0.96	0.95	0.94	0.92	0.92	0.91
STK	0.96	0.95	0.95	0.93	0.93	0.87
			DAILY MINIMUM			
EOF	0.94	0.94	0.92	0.89	0.90	0.88
STK	0.91	0.91	0.91	0.90	0.89	0.85
			DAILY MAXIMUM			
EOF	0.96	0.95	0.92	0.88	0.92	0.89
STK	0.94	0.92	0.91	0.91	0.89	0.87