Preprint ANL/MCS-P5141-0514

Data-Driven Model for Solar Irradiation Based on Satellite Observations

Ilias Bilionis[a], Emil M. Constantinescu[a], Mihai Anitescu[a]

[a]*Mathematics and Computer Science Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL 60439. Emails: {ebilionis,anitescu,emconsta}@mcs.anl.gov*

Abstract

We construct a data-driven model for solar irradiation based on satellite observations. The model yields probabilistic estimates of the irradiation field every thirty minutes starting from two consecutive satellite measurements. The probabilistic nature of the model captures prediction uncertainties and can therefore be used by solar energy producers to quantify the operation risks. The model is simple to implement and can make predictions in realtime with minimal computational resources. To deal with the high-dimensionality of the satellite data, we construct a reduced representation using factor analysis. Then, we model the dynamics of the reduced representation as a discrete (30-minute interval) dynamical system. In order to convey information about the movement of the irradiation field, the dynamical system has a two-step delay. The dynamics are represented in a nonlinear, nonparameteric way by a recursive Gaussian process. The predictions of the model are compared with observed satellite data as well as with a similar model that uses only ground observations at the prediction site. We conclude that using satellite data in an area including the prediction site signficantly improves the prediction compared with models using only ground observation site data.

Keywords:
insolation, irradiance, Bayesian, recursive Gaussian process, factor analysis, dimensionality reduction, dynamical system

1. Introduction

Solar irradiation is the amount of power per square meter that reaches the Earth from the Sun. In solar energy applications, part of the solar irradiation can be converted to electricity. In contrast to conventional power

sources such as coal or gas, solar irradiation is volatile and uncontrollable by the user. The two most important factors of the solar irradiation variability are the movement of the Sun and weather fluctuations. The former can be captured mathematically to great accuracy, because it is a deterministic effect. The latter is a chaotic effect, and hence the main cause for the difficulties associated with forecasting solar irradiation.

All the stages of a solar-power conversion project need to take into account the risks associated with solar irradiation. For the feasibility and design phases of the project, historical data can be employed to quantify these risks. The risks associated with the operation phase, however, require the ability to make short-term predictions (from 1 to 8 hours ahead) of the solar irradiation.

The most widely used solar irradiation forecasting methodologies are those that rely only on pointwise ground measurements of the solar irradiation. The reason is that ground measurements are readily available at any solar energy production plant. Mathematically, these techniques fall into the category of single-value time series analysis. Time series analysis methods include the autoregressive integrated moving average (ARIMA) processes [1] (see [2] for a first-order autoregressive model (AR(1)) and [3, Ch. 15.2.2.] for ARIMA examples), and artificial neural networks (ANNs) [4]. The accuracy of the predictions of these models degrades rapidly as the forecasting window is increased. This result is expected because the weather fluctuations exhibit a nonlocal behavior; see [3, Ch. 15] for a comprehensive review.

More accurate forecasts can be achieved only if nonlocal data are taken into account. For short-term (0-30 minutes) forecasts, a promising approach is to use total sky imager technologies [5]. One takes pictures of the sky from a particular site, extracts information about the clouds, constructs the cloud motion vectors (CMVs) and moves the image forward in time. Using geometrical arguments and semi-empirical models, one recovers the solar irradiation from the cloud information at that later time. The time frame for which this approach is useful depends on the velocity of the clouds.

Longer forecasts (hours to days) are feasible if satellite data are used. For forecasts ranging from 30 minutes to 6 hours ahead, the data-driven technique introduced in [6] may be used. As a first step, the semi-empirical heliostat method of [7] is used to extract cloud structure information from the satellite images. Then, as in the sky imager-based techniques, two consecutive images are compared in order to construct the CMVs. The cloud information is moved forward in time using the CMVs and goes through a

final smoothing phase. Solar irradiation is recovered by again employing the heliostat method. Numerical Weather Prediction (NWP) may be used for forecasts of up to 6 days or longer. For example, in [3, Ch. 10] the sky-cover fraction of the U.S. National Digital Forecast Database is coupled with semi-empirical models to produce long-term forecasts of solar irradiation.

The main disadvantage of most these aforementioned techniques is that they are difficult to use, and, sometimes, unable to quantify the uncertainty in their predictions. Given the current evolution of decision systems for energy toward incorporating stochastic representations [8], this may be a serious shortcoming. It is hard to see how to consistently add an uncertainty model to the heliostat approach. NWP models can in principle be modified to support an ensemble-based approach to uncertainty, but at a significant computational cost that requires a dedicated supercomputer [9]. Among the methods described, only the ARIMA-based methods can provide error bars for the predictions with a small or moderate effort. Yet, it is exactly these error bars that help quantify the potential risks and allow the stakeholders to properly price them.

These considerations have motivated us to develop a fully stochastic model that can quantify forecast uncertainties. In addition, aiming for a model that is convenient even for lean operations, we propose a model that is considerably easier to implement than any satellite-based model and can produce predictions in realtime with minimal computational resources.

Our model building philosophy and paper can be sumarized as follows. As a first step, the Sun's movement effect on the satellite observed solar irradiation field (Sec. 2.1) is removed by dividing it with a clear sky model (Sec. 2.2) to get the clear sky index field. Our goal is to use consecutive observations of the clear sky index field in order to learn its dynamics. Because of its high-dimensional nature, we construct a reduced-dimensionality representation of it (Sec. 2.3). To learn the dynamics of this low-dimensional representation, we use a nonlinear, nonparametric technique known as recursive Gaussian process (Sec. 2.4). Having constructed the dynamics of the reduced space, forecasts can be performed for an arbitrary number of time steps ahead (Sec. 2.5). Our recursive Gaussian process is similar in concept to the ANN used in [4]. However, our model is Bayesian, a key feature that enables us to make not only best estimates but also probabilistic forecasts. We then present (Sec. 3) our numerical results for an 8-hours-ahead forecast, and we compare them pointwise with those obtained by a recursive Gaussian process model based only on ground observations. We observe that using

satellite data significantly reduces the forecast uncertainties and improves the forecast itself. We attribute this improvement to the nonlocal information carried by the satellite images and to the space-time correlation between the solar irradiation at the prediction site and at neighboring sites.

2. Methodology

Throughout this work, we denote the solar irradiation field by $I(\phi, \lambda, t)$, where ϕ, λ and t are the latitude, longitude, and time, respectively. The units of $I(\phi, \lambda, t)$ are power per square meter.

2.1. Observations of Solar Irradiation

The solar irradiation field can be observed almost instantaneously by processing satellite images. For simplicity, we restrict our attention to data coming from the continental United States (CONUS) scan of the GOES-13 satellite. The CONUS scan takes place almost every 30 minutes and has a resolution of 1 km. The solar irradiation field is constructed by processing the raw data collected by the Advanced Very High Resolution Radiometer (AVHRR) via the algorithms described in [10]. This product is reported as part of the Clouds from AVHRR Extended (CLAVR-x, `https://cimss.ssec.wisc.edu/clavr/`) data and is available in realtime from the University of Wisconsin (`ftp://ftp.ssec.wisc.edu/clavr/goes_east/processed/`. The CLAVR-x data are available on a grid of latitudes and longitudes described by ϕ_{ij} and λ_{ij} for $i = 1, \ldots, P_\phi, j = 1, \ldots, P_\lambda$, where P_ϕ and P_λ denote the number of pixels on each dimension. That is, at time t we observe a matrix $\mathbf{I}(t) = (I_{ij}(t))$ of size $P_\phi \times P_\lambda$:

$$I_{ij}(t) = I(\phi_{ij}, \lambda_{ij}, t). \tag{1}$$

The dimensions of the observed matrix for the CONUS scan are $P_\phi = 1,900, P_\lambda = 3,100$. For computational reasons, we work with a submatrix of the CONUS scan of dimensions $P_\phi = P_\lambda = 400$ that is centered on a site of interest. We refer to this submatrix as the *patch*.

The time resolution of the data is irregular. In particular, we have observations every thirty minutes, with the exception of 03:00, 06:00, 09:00, 12:00, 15:00, 15:30, 18:00, 21:00, and 00:00 UTC daily as well as 17:00 UTC every Wednesday. In addition, early morning and late afternoon measurements are not useful, because of the very low irradiation values that are observed.

Therefore, we drop from our dataset observations that take place during nighttime (i.e., the period between two hours before sunset up to two hours after sunrise the next day, when sunset and sunrise times are calculated for an observer standing at the center of the patch). We denote the times for which we actually have observations by $t_k, k = 1, \ldots K_t$. Then, our observed dataset consists of

$$\mathcal{D}_I = \{\mathbf{I}(t_k) : k = 1, \ldots, K_t\}\,. \tag{2}$$

2.2. Clear Sky Model and the Clearness Index

The presence of daily and seasonal trends makes the direct modeling of $I(\phi, \lambda, t)$ problematic. On the one hand, the daily part is characterized by (1) no variation over night (the solar irradiation field is equal to zero), (2) increasing values and variance until noon, and (3) decreasing values and variance from noon till sunset. On the other hand, the seasonal trend is characterized by (1) low values during winter and high values during summer, and (2) high variance during winter and low variance during summer. The change in the mean values of the field is directly chained to the solar zenith angle, which varies daily as well as seasonally. The higher variance during winter days is attributed to the rapidly changing weather conditions. To partially address these trends, we model the *clearness index* instead of the solar irradiation field. In order to properly define the clearness index, the concept of a *clear sky model* is required.

A clear sky model approximates the expected solar irradiation when no clouds are present. Typically, it depends on the extraterrestrial irradiation ($(\mathrm{W/m^2})$ at the top of Earth's surface), the solar zenith angle at a particular Earth site, the elevation of the site above sea level, the composition of atmospheric gases such as water vapor and ozone content, and the atmospheric aerosol content (see [3, 2.3]). Many clear sky models may be used (e.g., [11, 12]). However, given the variability in the total solar irradiation due to the cloud variability, very high accuracy of the clear sky model is not extremely important here. We thus opt for Ineichen's model [12], which is a surrogate of the more accurate Solis model [13]. The Ineichen model depends on the extraterrestrial irradiation, the solar zenith, the elevation of the site above sea level, the atmospheric water vapor content, and the aerosol optical depth at 700 nm. We denote the global clear sky irradiation of Ineichen's model by $I_{\mathrm{cls}}(\phi, \lambda, t)$. Details on the various model parameters are given in Appendix A.

We can now define the clearness index field as the ratio of the solar irradiation field to the clear sky model:

$$C(\phi, \lambda, t) = \frac{I(\phi, \lambda, t)}{I_{\text{cls}}(\phi, \lambda, t)}. \tag{3}$$

The observed data of Eq. (1) may be expressed in terms of the clearness index by

$$\mathcal{D}_C = \{\mathbf{C}(t_k) : k = 1, \ldots, K_t\}, \tag{4}$$

where $\mathbf{C}(t)$ is the $P_\phi \times P_\lambda$-matrix defined as

$$C_{ij}(t) = \frac{I_{ij}(t)}{I_{\text{cls,ij}}(t)} = \frac{I(\phi_{ij}, \lambda_{ij}, t)}{I_{\text{cls}}(\phi_{ij}, \lambda_{ij}, t)},$$

for $i = 1, \ldots, P_\phi, j = 1, \ldots, P_\lambda$.

Our goal is to model the evolution of the clearness index field $C(\phi, \lambda, t)$ based on the observations included in Eq. (4). In reality, we will be focusing on a discrete version of the problem, namely, modeling the evolution of the matrix $\mathbf{C}(t)$. Since the clear sky model may be evaluated anywhere at will, the solar irradiation field can be recovered in a trivial way from Eq. (3).

We note that some trends do persist even after switching to the clearness index. In particular, the clearness index field exhibits a seasonal trend in its variance similar to that of the solar irradiation field. However, this remaining trend is inconsequential, because our goal is to make short-term forecasts.

2.3. Dealing with the Curse of Dimensionality

The mathematical problem we are facing is the data-driven determination of the dynamics of an extremely high-dimensional system. The very small patch of data we are considering has $160,000$ ((400×400) dimensions. Attempting to learn the dynamics of the clearness index directly is not tractable. Clearly, we must resort to some kind of *dimensionality reduction technique*.

The observed data are in the high-dimensional space $\mathbb{R}^{P_\phi \times P_\lambda}$. However, we expect them to be constrained on a low-dimensional manifold embedded in this high-dimensional space. This expectation is based on the spatial correlations of the field that are induced by physical laws involving cloud transportation, nucleation, and annihilation. In the hypothetical scenario of a truly $P_\phi \times P_\lambda$-dimensional manifold, there would be no spatial correlations. But such a scenario is not supported by the data: displaying succesive maps

of the solar irradiation field shows evolution of persistent moderate scale patterns, which is consistent with nontrivial spatial correlation.

To be precise, let $\mathcal{M}_h \subset \mathbb{R}^{P_\phi \times P_\lambda}$ denote the manifold in which the clearness index lives. Mathematically, it is defined as

$$\mathcal{M}_h = \{\mathbf{C}(t) : \text{ for all } t\}. \tag{5}$$

We expect the true dimensionality of this manifold to be less than $P_\phi \times P_\lambda$. That is, we expect that it can be parametrized with fewer than $P_\phi \times P_\lambda$ variables.

A dimensionality reduction technique should provide two maps: a *reduction* map, and a *reconstruction* map from and to $\mathcal{M}_h$, respectively. A reduction map $\mathcal{R} : \mathcal{M}_h \to \mathbb{R}^R$ projects the original data to the low-dimensional space $\mathbb{R}^R$ with $R \ll P_\phi \times P_\lambda$. A *reconstruction map* $\mathcal{C} : \mathbb{R}^R \to \mathcal{M}_h$ maps the low-dimensional space back to the original one. A *good* pair of reduction-reconstruction maps $(\mathcal{R}, \mathcal{C})$ should have an R as small as possible and also satisfy the following property,

$$\mathcal{C}\left(\mathcal{R}\left(\mathbf{C}(t)\right)\right) \approx \mathbf{C}(t), \tag{6}$$

with "$\approx$" being interpreted as "close" in some appropriate norm. In other words, the reconstruction map should approximately be the inverse of the reduction map. In reality, we always expect some information loss along the way. After finding a good $(\mathcal{R}, \mathcal{C})$ pair, our goal will be to model the dynamics of the R-dimensional projections of the original data. This is a much easier task than dealing directly with $\mathbf{C}(t)$. To clarify this point, let us denote by $\mathbf{x}(t) \in \mathbb{R}^R$ the *reduced* variables:

$$\mathbf{x}(t) = \mathcal{R}\left(\mathbf{C}(t)\right). \tag{7}$$

Our goal is to capture the dynamics of $\mathbf{x}(t)$. The original dynamics can be recovered by passing $\mathbf{x}(t)$ through the reconstruction map. The observations we have at hand for this are induced by Eq. (4) through the reduction map

$$\mathcal{D}_x = \{\mathbf{x}(t_k) = \mathcal{R}\left(\mathbf{C}(t_k)\right) : k = 1, \ldots, K_t\}. \tag{8}$$

This is the topic of Sec. 2.4. For the moment, we focus on finding $(\mathcal{R}, \mathcal{C})$.

The only information available for determining the pair $(\mathcal{R}, \mathcal{C})$ is the observed clearness index dataset $\mathcal{D}_C$ given in Eq. (4). The dimensionality reduction technique we will use is the factor analysis (FA)[14, 15, 16]. FA can

be thought of as a generalization of probabilistic principal component analysis (pPCA) [17, 18], a probabilistic interpretation of the celebrated principal component analysis (PCA) (see [19, 20] and for a more recent reference [21, Ch. 12.1]). PCA works under the assumption that the data manifold $\mathcal{M}_h$ is linear. In other words, one may think of it as an attempt to approximate $\mathcal{M}_h$ as being embedded in a low-dimensional affine space. In addition, pPCA and FA attempt to characterize the uncertainty of the reduction/reconstruction operations. Both fall into the category of latent linear Gaussian models; but pPCA uses the same variance to characterize the reconstruction uncertainty, whereas FA uses a different variance for each distinct feature.

FA works with vectors. Therefore, all quantities must be vectorized before proceeding further. For notational convenience, let $\text{vec}_{m,n} : \mathbb{R}^{m\times n} \to \mathbb{R}^{mn}$ be the *vectorization* operator. The action of the vectorization operator on a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ transforms it into a vector $\mathbf{a} \in \mathbb{R}^{mn}$, as follows:

$$\mathbf{a} = \text{vec}_{m,n}\left(\mathbf{A}\right) := \left(A_{11}, \ldots, A_{1n}, \ldots, A_{m1}, \ldots, A_{mn}\right). \tag{9}$$

In addition, we need the inverse transform $\text{vec}_{m,n}^{-1}$, which transforms a vector $\mathbf{a} \in \mathbb{R}^{mn}$ into a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$:

$$\mathbf{A} = \text{vec}_{m,n}^{-1}\left(\mathbf{a}\right) := \begin{pmatrix} a_1 & \ldots & a_n \\ \vdots & \ddots & \vdots \\ a_{m(n-1)+1} & \ldots & a_{mn} \end{pmatrix}. \tag{10}$$

Since we are interested in vectorizing $\mathbf{C}(t) \in \mathbb{R}^{P_\phi \times P_\lambda}$, we need to use $\text{vec}_{P_\phi \times P_\lambda}$. For brevity, we define the following:

$$\text{vec} := \text{vec}_{P_\phi \times P_\lambda} \text{ and } \text{vec}^{-1} := \text{vec}_{P_\phi \times P_\lambda}^{-1}. \tag{11}$$

Our problem now is to find a reduced representation of the K_t $(P_\phi P_\lambda)$-dimensional observations:

$$\mathcal{D}_c := \left\{\mathbf{c}(t_k) = \text{vec}\left(\mathbf{C}(t_k)\right) : k = 1, \ldots, K_t\right\}. \tag{12}$$

In our discussion of FA we follow closely the work of Bishop [21, Ch. 12.2.4]. Let $\mathbf{x} \in \mathbb{R}^R$ be the latent reduced representation of an observed clearness index vector $\mathbf{c} \in \mathbb{R}^{P_\phi P_\lambda}$. In FA, we assume that the observed clearness vector $\mathbf{c}$ is generated from $\mathbf{x}$ as follows,

$$p(\mathbf{c}|\mathbf{x}, \mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Psi}) = \mathcal{N}\left(\mathbf{c}|\mathbf{W}\mathbf{x} + \boldsymbol{\mu}, \boldsymbol{\Psi}\right), \tag{13}$$

where $\mathbf{W} \in \mathbb{R}^{(P_\phi P_\lambda)\times R}$ and $\boldsymbol{\mu} \in \mathbb{R}^{P_\phi P_\lambda}$ are the *reconstruction matrix*, and the *mean vector*, respectively, and $\boldsymbol{\Psi} \in \mathbb{R}^{(P_\phi P_\lambda)\times(P_\phi P_\lambda)}$ is a diagonal matrix containing the *reconstruction variance* of each component of $\mathbf{c}$ (each component is known as a *feature*). Equation (13) defines a probabilistic version of the reconstruction map $\mathcal{C}$.

To close the model, one must say how $\mathbf{x}$ would be generated if $\mathbf{c}$ was not observed. In FA, one assumes that $\mathbf{x}$ is generated from a unit variance Gaussian,

$$p(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}|\mathbf{0}, \mathbf{I}\right), \tag{14}$$

where $\mathbf{I}$ is the unit matrix.

The reduction map $\mathcal{R}$ is also defined probabilistically. One has to compute the posterior of $\mathbf{x}$ conditioned on $\mathbf{c}$ and all the parameters $(\mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Psi})$. This is again a Gaussian of the form

$$p(\mathbf{x}|\mathbf{c}, \mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Psi}) = \mathcal{N}\left(\mathbf{x}|\mathbf{G}\mathbf{W}^T\boldsymbol{\Psi}^{-1}(\mathbf{c}-\boldsymbol{\mu}), \mathbf{G}\right), \tag{15}$$

where

$$\mathbf{G} = \left(\mathbf{I} + \mathbf{W}^T\boldsymbol{\Psi}^{-1}\mathbf{W}\right)^{-1}. \tag{16}$$

FA can be trained efficiently by maximizing the likelihood of $\mathcal{D}_c$ of Eq. (12) using the expectation-maximization (EM) algorithm [22]. In particular, combining Eqs. (13) and (14) and assuming independence of the vector $\mathbf{c}$ conditional on the projection on principal components $\mathbf{x}$, we need to solve the following maximization problem:

$$\mathbf{W}^*, \boldsymbol{\mu}^*, \boldsymbol{\Psi}^* = \arg\max_{\mathbf{W},\boldsymbol{\mu},\boldsymbol{\Psi}} p(\mathcal{D}_c|\mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Psi}), \tag{17}$$

where the likelihood term is given by

$$p(\mathcal{D}_c|\mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Psi}) = \prod_{k=1}^{K_t} p\left(\mathbf{c}(t_k)|\mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Psi}\right), \tag{18}$$

with

$$p\left(\mathbf{c}(t_k)|\mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Psi}\right) = \int p\left(\mathbf{c}(t_k)|\mathbf{x}(t_k), \mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Psi}\right) p\left(\mathbf{x}(t_k)\right) d\mathbf{x}(t_k). \tag{19}$$

From this point on, for notational convenience we write

$$\boldsymbol{\phi}^* = (\mathbf{W}^*, \boldsymbol{\mu}^*, \boldsymbol{\Psi}^*) \tag{20}$$

to denote the solution of the maximization problem of Eq. (17). The projection operator is then stated as

$$\mathcal{R}(\mathbf{C}) := \mathbf{G}^* \mathbf{W}^{*T} (\mathbf{\Psi}^*)^{-1} \left(\mathrm{vec}\,(\mathbf{C}) - \boldsymbol{\mu}^* \right), \tag{21}$$

with $\mathbf{G}^*$ as in Eq. (16) evaluated for $\mathbf{W} = \mathbf{W}^*$ and $\mathbf{\Psi} = \mathbf{\Psi}^*$.

Certainly the assumption of independence of $\mathbf{c}$ conditional on $\mathbf{x}$ is arguable, although the problem would become much harder if we did not use it. On the other hand, should $\mathbf{c}$ be entirely contained in the space spanned by $\mathbf{x}$ for all times t, the assumption would be satisfied. The assumption is thus less restrictive the more accurate the PCA projection is. Therefore, we anticipate the conditional independence being a good approximation for forecasting large-scale behavior, which tends to be contained in the principal components.

We now come to the problem of determining the appropriate value for R (the dimension of the reduced space). This could be attacked in a fully Bayesian manner by following the ideas of [23]. In this work, however, we simply fix R to the maximum value we can afford, given the constraints of our model. In particular, as R increases so does the number of parameters of the reduced dynamics model (see Sec. 2.4). Because of the limited number of observations we have, it does not make sense to select an R that is greater than 8.

2.4. Learning the Reduced Dynamics

The task of this section is to learn the dynamics of the reduced variables $\mathbf{x}(t)$ (see Eq. (7)) based on $\mathcal{D}_x$ (see Eq. (8)). We use a model for nonlinear time-series known as a *recursive* or *dynamic* Gaussian process [24, 25]. Such models have been used extensively over the past decade in diverse tasks including human motion modeling/tracking [26, 27] and nonlinear signal processing [28]. For notational clarity, we assume that the time t counts units of 30 minutes; that is, $t-1$ means 30 minutes before t, $t-2$ means 60 minutes before t, and so forth. We assume that the evolution of $\mathbf{x}(t)$ obeys the following discrete dynamics:

$$\mathbf{x}(t) = \mathbf{f}\left(\mathbf{x}(t-1), \mathbf{x}(t-2) \right), \tag{22}$$

where $\mathbf{f} : \mathbb{R}^R \times \mathbb{R}^R \to \mathbb{R}^R$ is an unknown function to be determined from the observations $\mathcal{D}_x$ of Eq. (8). The reason we have assumed a dependence on $\mathbf{x}(t-2)$ in addition to $\mathbf{x}(t-1)$ is to include some information about the

velocity field that transports the clearness index. Simply including $\mathbf{x}(t-1)$ would yield a Markovian model that would not be able to capture any transport properties.

The idea of a recursive Gaussian process is to represent the function $\mathbf{f}$ as a Gaussian process to be learned from the data. In turn, a Gaussian process may be thought as a nonparametric way of performing regression tasks [29]. We denote the input of $\mathbf{f}$ collectively by $\mathbf{z}(t) \in \mathbb{R}^{2R}$, where

$$\mathbf{z}(t) = (\mathbf{x}(t-1), \mathbf{x}(t-2)) \,. \tag{23}$$

Its output is simply $\mathbf{x}(t)$. The data we have available for learning $\mathbf{f}$ are

$$\mathcal{D}_R = \{(\mathbf{z}(t_k), \mathbf{x}(t_k)) : t_k - t_{k-1} = 1, t_{k-1} - t_{k-2} = 1, k = 3, \ldots, K_t\} \,.$$

That is, each of the $(\mathbf{z}(t), \mathbf{x}(t))$ pairs requires exactly three consecutive measurements of the clearness index to be available. For future reference, we assume that there are N observations in $\mathcal{D}_R$ as follows:

$$\mathcal{D}_R = \left\{ \left(\mathbf{z}^{(i)}, \mathbf{x}^{(i)}\right) : i = 1, \ldots, N \right\} . \tag{24}$$

The core methodology for Gaussian processes aims at learning real functions, that is, functions with one output. In contrast, here we are dealing with the problem of learning a multioutput function. Despite the fact that there is a wealth of methods for learning multioutput functions [30, 31, 32, 33]; we chose a simple approach that treats each output dimension of $\mathbf{f}$, independently. In particular, we assume that each of the components $f_r, r = 1, \ldots, R$ of $\mathbf{f}$ is a Gaussian process representing the evolution of the rth principal component

$$x_r(t) = f_r\left(\mathbf{z}(t)\right) . \tag{25}$$

Each of these functions can be learned from a fraction of the data contained in $\mathcal{D}_R$ of Eq. (24). In particular, f_r can be learned from

$$\mathcal{D}_{R,r} = \left\{ \left(\mathbf{z}^{(i)}, x_r^{(i)}\right) : i = 1, \ldots, N \right\} , \tag{26}$$

where x_r is the rth component of $\mathbf{x}$.

At this point, we elaborate on how each of the functions f_r is constructed based on $\mathcal{D}_{R,r}$ of Eq. (26) using Gaussian process regression. For notational convenience, we drop the index r from any equation in this paragraph. For

more details, the reader may consult [34]. Prior to seeing the data, we assume that the f_r is a draw from a zero mean Gaussian process,

$$f_r \sim \mathrm{GP}\left(f_r|0, k(\cdot,\cdot;\boldsymbol{\theta}_r)\right), \tag{27}$$

where $k : \mathbb{R}^{2R} \times \mathbb{R}^{2R} \to \mathbb{R}$ is a covariance function and $\boldsymbol{\theta}_r$ all its parameters. Notice that the same covariance function will be used for each r but its parameters $\boldsymbol{\theta}_r$ will be different. The particular form of the covariance functions used in the numerical examples is discussed in Sec. 3. The likelihood of the observed data is

$$p(\mathcal{D}_{R,r}|\boldsymbol{\theta}) = \mathcal{N}\left(\mathbf{X}_r|\mathbf{0}, \mathbf{A}_r\right), \tag{28}$$

where $\mathbf{X}_r$ is the vector of all the x_r's observed in $\mathcal{D}_{R,r}$ of Eq. (26),

$$\mathbf{X}_r = \left(x_r^{(1)}, \ldots, x_r^{(N)}\right), \tag{29}$$

and $\mathbf{A}_r \in \mathbb{R}^{N\times N}$ is the covariance matrix evaluated at $\boldsymbol{\theta}_r$:

$$\mathbf{A}_{r,nm} = k\left(\mathbf{z}^{(n)}, \mathbf{z}^{(m)}; \boldsymbol{\theta}_r\right). \tag{30}$$

To train the model, we maximize the logarithm of the likelihood, Eq. (28), with respect to $\boldsymbol{\theta}_r$,

$$\boldsymbol{\theta}_r^* = \arg\max_{\boldsymbol{\theta}_r} \log p(\mathcal{D}_{R,r}|\boldsymbol{\theta}_r), \tag{31}$$

subject to any constraints we might have. Having found a point estimate $\boldsymbol{\theta}_r^*$ of $\boldsymbol{\theta}_r$, we have as the *predictive distribution*

$$p(x_r(t)|\mathbf{z}(t), \mathcal{D}_{R,r}, \boldsymbol{\theta}_r^*) = \mathcal{N}\left(x_r(t)|m^*\left(\mathbf{z}(t)\right), \sigma^{*2}\left(\mathbf{z}(t)\right)\right), \tag{32}$$

where m^* is the *predictive mean*,

$$m^*\left(\mathbf{z}(t)\right) = \mathbf{a}_r^*\left(\mathbf{z}(t)\right)\mathbf{A}_r^{*-1}\mathbf{X}_r, \tag{33}$$

and σ^{*2} is the *predictive variance*,

$$\sigma^{*2}\left(\mathbf{z}(t)\right) = k\left(\mathbf{z}(t), \mathbf{z}(t); \boldsymbol{\theta}_r^*\right) - \mathbf{a}_r^*\left(\mathbf{z}(t)\right)\mathbf{A}_r^{*-1}\mathbf{a}_r^*\left(\mathbf{z}(t)\right)^T, \tag{34}$$

where $\mathbf{A}_r^*$ is the covariance matrix of Eq. (30) evaluated at $\boldsymbol{\theta}_r^*$ and $\mathbf{a}_r^*\left(\mathbf{z}(t)\right)$ is the *cross-covariance*:

$$\mathbf{a}_r^*\left(\mathbf{z}(t)\right) = \left(k\left(\mathbf{z}(t), \mathbf{z}^{(1)}; \boldsymbol{\theta}_r^*\right), \ldots, k\left(\mathbf{z}(t), \mathbf{z}^{(N)}; \boldsymbol{\theta}_r^*\right)\right). \tag{35}$$

Combining the results for every r, we get the *multioutput predictive distribution*:

$$p\left(\mathbf{x}(t)|\mathbf{z}(t), \mathcal{D}_R, \boldsymbol{\theta}^*\right) = \prod_{r=1}^{R} p\left(x_r(t)|\mathbf{z}(t), \mathcal{D}_{R,r}, \boldsymbol{\theta}_r^*\right), \tag{36}$$

where $\boldsymbol{\theta}^* = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_R\}$. Equation (36) characterizes in a probabilistic manner all the information about the dynamics that we are able to get from the observations. As shown in Sec. 2.5, iteratively taking samples from Eq. (36) allows us to make probabilistic forecasts about the evolution of the reduced dynamics.

2.5. Forecasting the Quantities of Interest

Now we come to the problem of forecasting the clearness index. Our predictions are going to be probabilistic, and they will be based solely on Eq. (13), Eq. (15), and Eq. (36).

Assume that we have observed the clearness index at times t_0 and t_0-1; in other words, we have observed the matrices $\mathbf{C}(t_0)$ and $\mathbf{C}(t_0-1)$, respectively. We would like to predict the clearness index at future times $t = t_0+1, t_0+2, \ldots$. First we need to project the observations to the reduced components using Eq. (15),

$$\begin{array}{rcl} \mathbf{x}(t_0) & \sim & p(\mathbf{x}(t_0)|\operatorname{vec}\left(\mathbf{C}(t_0)\right), \boldsymbol{\phi}^*), \\ \mathbf{x}(t_0-1) & \sim & p(\mathbf{x}(t_0-1)|\operatorname{vec}\left(\mathbf{C}(t_0-1)\right), \boldsymbol{\phi}^*), \end{array} \tag{37}$$

or, in short notation,

$$\mathbf{z}(t_0) = \left(\mathbf{x}(t_0-1), \mathbf{x}(t_0-2)\right). \tag{38}$$

Note that this kind of probabilistic projection accounts for the uncertainty of the reduction map.

We can take sample paths from the second-order Markov chain defined by the predictive distribution of Eq. (36) to produce a sample path n steps long (see Alg. 1). The initial points $\mathbf{x}(t_0-1)$ and $\mathbf{x}(t_0)$ are assumed to be sampled as in Eq. (37). That is, we can produce sample paths that go arbitrarily far into the future, albeit in multiples of 30 minutes. Suppose that we have taken $S > 0$ such samples. Let us denote each of them by $\left\{\mathbf{x}^{(s)}(t_0+i)\right\}_{i=1}^{n}, s = 1, \ldots, S$. These can be transformed to sample paths of the clearness index matrix $\left\{\mathbf{C}^{(s)}(t_0+i)\right\}_{i=1}^{n}, s = 1, \ldots, S$ using Eq. (13):

$$\begin{array}{rcl} \mathbf{c}^{(s)}(t_0+i) & \sim & p\left(\mathbf{c}^{(s)}|\mathbf{x}^{(s)}(t_0+i), \boldsymbol{\phi}^*\right), \\ \mathbf{C}^{(s)}(t_0+i) & = & \operatorname{vec}^{-1}\left(\mathbf{c}^{(s)}(t_0+i)\right). \end{array} \tag{39}$$

Equation 13 accounts for the uncertainty of the reconstruction map. If one wants sample paths $\left\{\mathbf{I}^{(s)}(t_0+i)\right\}_{i=1}^{n}, s=1,\ldots,S$, of the solar irradiation, these can be trivially acquired by using Eq. (3),

$$\mathbf{I}^{(s)}(t_0+i) = \mathbf{C}^{(s)}(t_0+i) * \mathbf{I}_{\text{cls}}(t_0+i), \tag{40}$$

where $\mathbf{I}_{\text{cls}}(t_0+i)$ is the $P_\phi \times P_\lambda$ matrix containing the evaluation of the clear sky model at time t_0+i over the whole patch and "$*$" stands for *elementwise* multiplication of two matrices of the same dimensions.

Algorithm 1 Algorithm taking a single sample path from the second-order Markov chain defined by Eq. (36)

Require: Observed data: $\mathbf{x}(t_0), \mathbf{x}(t_0-1)$, number of steps ahead to predict: $n \geq 1$.

Ensure: $\{\mathbf{x}(t_0+i)\}_{i=1}^{n}$ is a sample path of the second order Markov chain defined by Eq. (36).

for $i = 1$ to n **do**

Sample $\mathbf{x}(t_0+i)$ from $p\left(\mathbf{x}(t_0+i)|\mathbf{x}(t_0+i-1), \mathbf{x}(t_0+i-2), \mathcal{D}_R, \boldsymbol{\theta}^*\right)$ as defined in Eq. (36).

end for

3. Numerical Examples

As discussed in Sec. 2.1, the satellite observations of solar irradiation are obtained from the CLAVR-x project. We have been systematically storing the CLAVR-x data since October 10, 2013. We will use all valid observations until February 25, 2014, to train our model. There are 822 such observations. The number is reduced because of two facts: (1) only observations that take place 2 hours after sunrise and 2 hours before sunset are retained, (2) there is a gap in observations during January 2014 because of technical problems faced by the CLAVR-x group. This constitutes our observed solar irradiation dataset $\mathcal{D}_I$ of Eq. (2). From $\mathcal{D}_I$, the observed clearness index dataset $\mathcal{D}_C$ of Eq. (4) is constructed as outlined in Sec. 2.2. The available data from February 26, 2014, to March 4, 2014, are used to test our predictions.

Because we would like to compare the performance of our model with one based on ground observations, we center our 400×400 pixel patch on Lamont, OK (see Fig. 3 for a map of the region and Fig. 2(a) for a sample

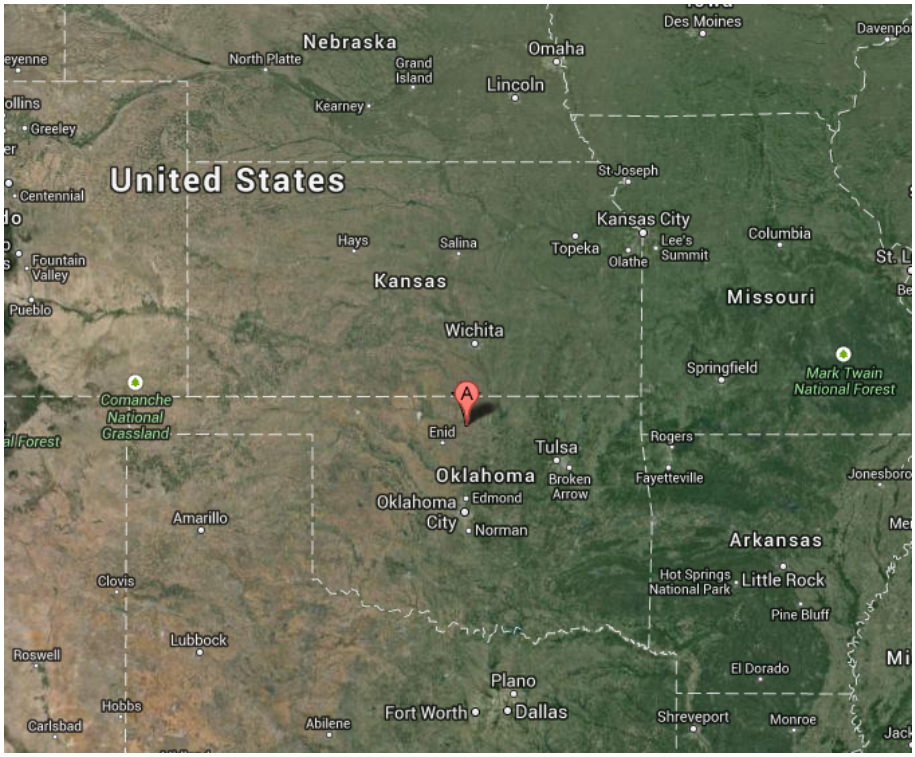
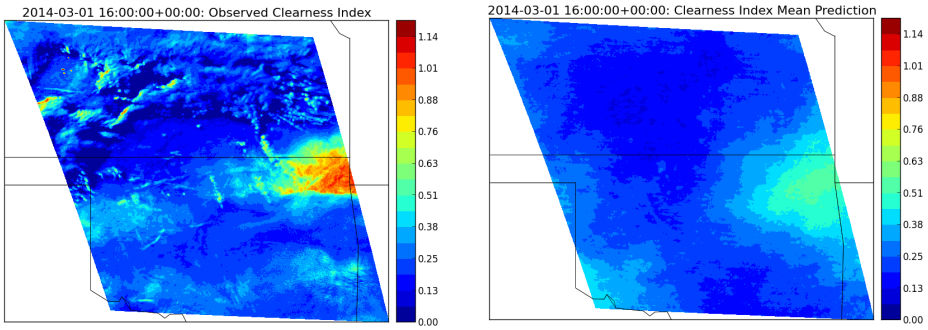


Figure 1: Map of the patch around Lamont, OK, that we consider in our numerical examples. Source: Google Maps.



(a) Observed $\mathbf{C}(t)$ (b) Reconstruction $\mathcal{C}(\mathcal{R}(\mathbf{C}(t)))$

Figure 2: Sample of the clearness index on a 400×400 patch around Lamont, OK, on March 3, 2014, at 16:00 UTC (a) and its reconstruction with $R = 8$ components (b).

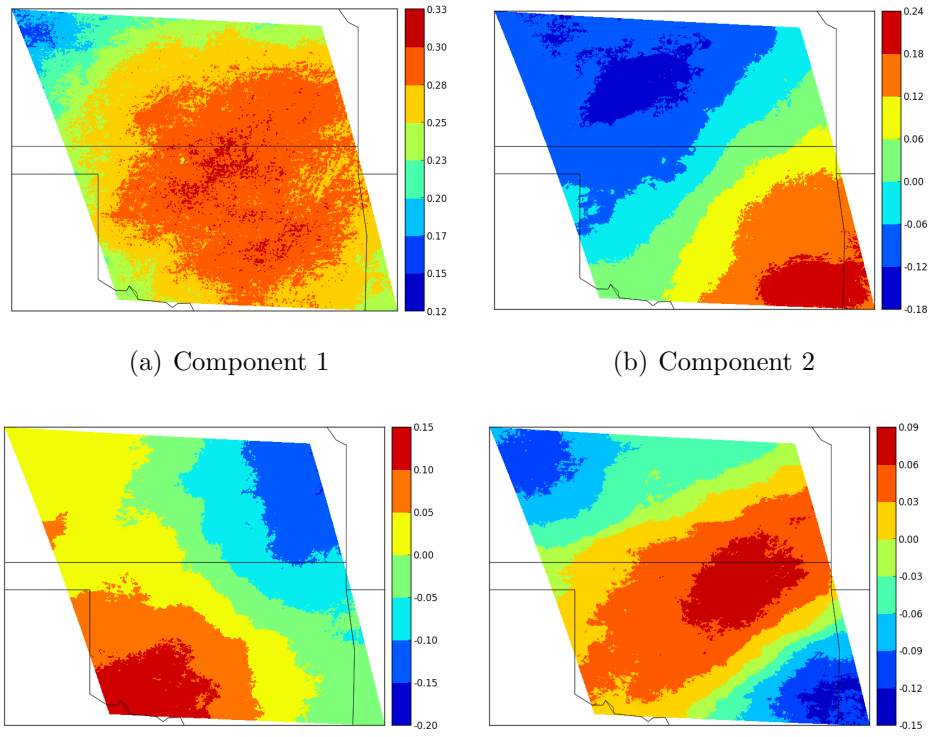
clearness index field). For this particular location there exist excellent ground observations of solar irradiation that we can use for validation [35]. The ground model to which we compare our results is again a recursive GP on the clearness index observed at this site. The ground data used for training span the period from October 1, 2013, to February 25, 2014.

The observed clearness index dataset $\mathcal{D}_C$ is used to perform the dimensionality reduction task (see Sec. 2.3). In our numerical examples, we have experimented with $R = 4$ and $R = 8$ components without observing any measurable performance gains by adding more complexity. For $R = 16$ components, the likelihood maximization encounters many local maxima because the number of parameters is considerably larger. In this case, either a global approach to maximizing the likelihood or some form of regularization, for example a fully Bayesian treatment with priors on the covariance parameters in (27), is required. We will present results only for the $R = 8$ case. The first four principal components $\mathbf{u}_r$ are visualized in Fig. 3. In Fig. 2(b) we plot the reconstruction of the observed clearness index of Fig. 2(a) using just $R = 8$ components. There certainly is a loss of information when we project the high-dimensional description to the low-dimensional one. The compression ratio is about 10^5, however, and we are able to estimate the error in prediction by the methods described in the previous section.

Having constructed the reduced representation of the data, we come to the problem of learning the evolution of the reduced dynamics described in Sec. 2.4. In Fig. 4 we illustrate the evolution of $\mathbf{x}(t)$ of Eq. (7). The stars correspond to observations. Notice that since the collection procedure has both weekly and daily irregularities in its protocol, as described in Sec. 2.1, the data in Fig. 4 show occasional temporal gaps. As a result, of the 822 observations of $\mathbf{x}(t)$, only 252 remain in $\mathcal{D}_R$ of Eq. (24) and can, therefore, be used to train the recursive GP of Eq. (22). The reason is that we can only use triplets of three consecutive measurements with a 30-minute lag, and the gaps in data imply that not all observations are part of such triplets. Such a limited number of observations in $\mathcal{D}_R$ constrains the number of degrees of freedom in the statistical model and consequently limits the size of the reduced dimension R.

The covariance function we are using in Eq. (27) has the form

$$k(\mathbf{z}, \mathbf{z}'; \boldsymbol{\theta}_r) = v_{r,0} + \sum_{i=1}^{2R} v_{r,i} z_i z_i' + k_0(\mathbf{z}, \mathbf{z}'; \boldsymbol{\theta}_{r,0}) + \sigma_r^2 \delta(\mathbf{z} - \mathbf{z}'), \qquad (41)$$



(a) Component 1

(b) Component 2

(c) Component 3

(d) Component 4

Figure 3: Principal components ($\text{vec}^{-1}(\mathbf{u}_r)$) 1 to 4.

where $\boldsymbol{\theta}_r = (v_{r,0}, \ldots, v_{r,2R}, \boldsymbol{\theta}_{r,0}), \sigma_r,\ v_{r,i} \geq 0,\ k_0$ and is a classic covariance function (see [34, Ch. 4]). Here, $r = 1, 2, \ldots, R$ denotes the index of the reduced components whose evolution is modeled. The first term in Eq. (41) captures a constant bias. The second, nonstationary, term captures a structured linear dependence between $\mathbf{z}$ and $\mathbf{x}$. The third term captures any nonlinear dependence. The second term can be interpreted as the effect of a mean 0 prior on a linear mean term [34, Ch. 2.7]. The last term models the possible effects of measurement noise (though we anticipate its effect to be small, since the vector $\mathbf{x}$) is spatially filtered by the projection on the principal component). The covariance function forms $k_0(\cdot, \cdot; \cdot)$ we have experimented with include the squared exponential, the exponential and several covariances belonging to the matern class. None of these choices yielded a significant improvement in the final forecasting capabilities of the model.

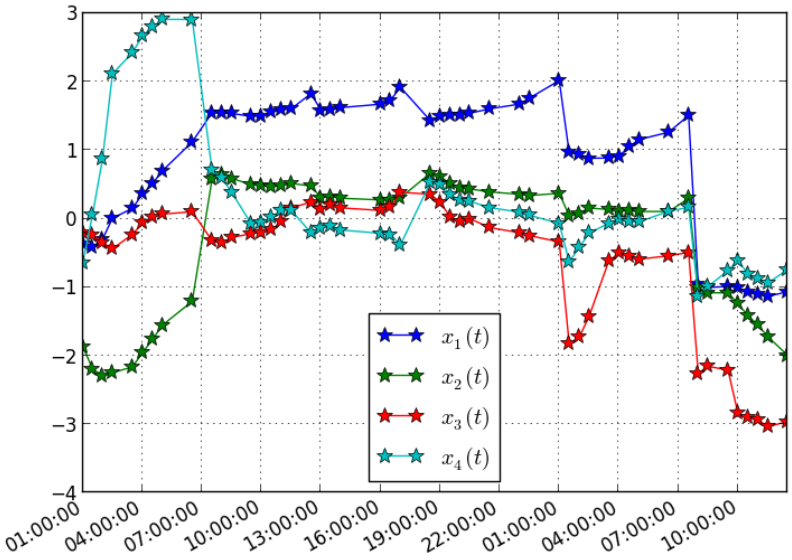


Figure 4: The first four reduced variables $x_1(t), \ldots, x_4(t)$. The time is in UTC, and the stars indicate observations.

Therefore, we present only those results obtained by the exponential covariance function whose form is

$$k_0\left(\mathbf{z}, \mathbf{z}', \boldsymbol{\theta}_{r,0}\right) = s_r^2 \exp\left\{-\sum_{i=1}^{2R} \frac{|z_i - z_i'|}{\ell_{r,i}}\right\}, \tag{42}$$

where $\boldsymbol{\theta}_{r,0} = (s_r, \ell_{r,1}, \ldots, \ell_{r,2R})$. The parameters s_r and $\ell_{r,i}$ may be interpreted as the nonlinear part of the signal strength and its length scale with respect to the input variable z_i, respectively. The number of parameters per r is $|\boldsymbol{\theta}_r| = 4R + 3$, and the total number of parameters is $R(4R + 3)$. For $R = 8$ this translates to 67 parameters per r that need to inferred from $\mathcal{D}_{R,r}$ of Eq. (26). It now becomes apparent that the low number of observations (about 250) we have at our disposal does not allow for choosing a large R.

The optimization problem of each r (see Eq. (31)) is solved by using the BFGS algorithm [36]. To accommodate the nonnegative constraints, we use an exponential change of variable. The results are shown in Table 1. The '*' symbols for the $\ell_{r,i}$ scale parameters indicate that their corresponding values

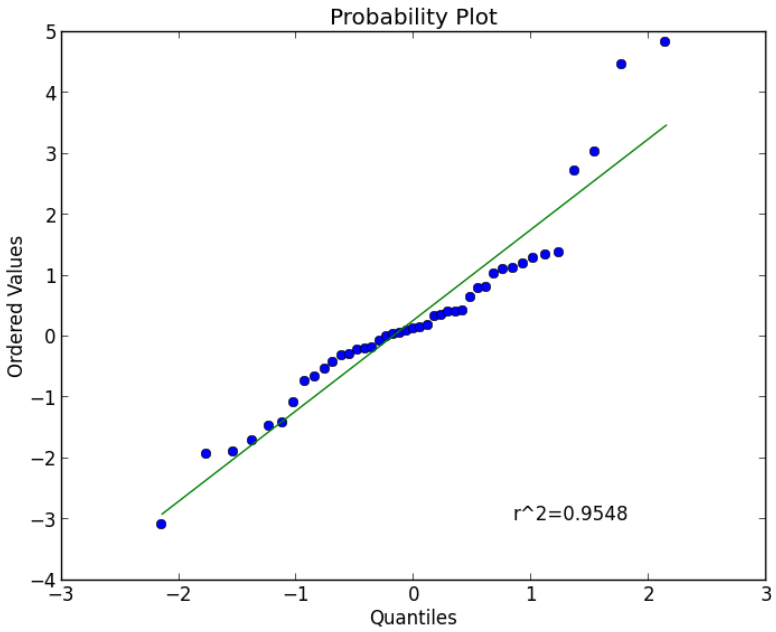
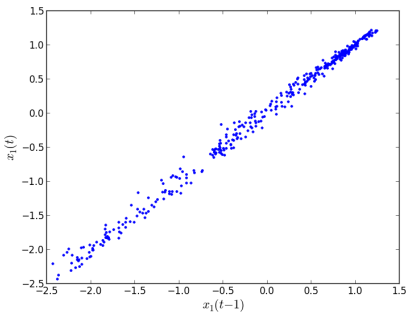
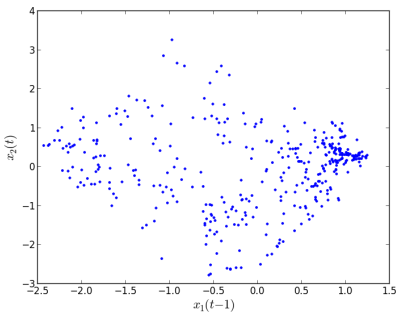


Figure 5: Quantile plot for one-step-ahead probabilistic predictions.

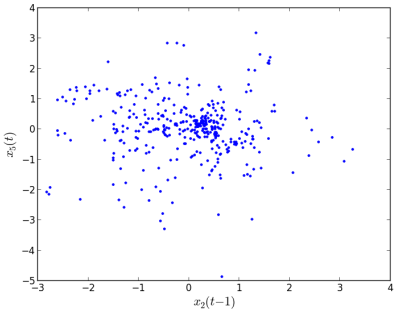
were greater than 10 and, hence, tend to affect the model signficantly less than those with smaller values, and displaying them would make the dependence patterns less visible. Parameters displayed as 0.00 are not necessarily 0; they are simply displayed in fixed point to emphasize their lack of significance. We notice that the bias terms $v_{r,0}$ are effectively zero. This result was expected since $\mathbf{x}(t)$ are the projections of quantities with mean removed (21). We also notice notice that given the size of the parameters s_r^2, the most important effect observed originates from the nonstationary component of our covariance function of Eq. (41). This is the part that captures the linear part of the correlations between $\mathbf{z}(t)$ and $\mathbf{x}(t)$. However, the nonlinear part, coming from k_0 of Eq. (42), is also considerable and accounts for about 3% to 20% of total signal strength as measured by $s_r/\sqrt{\sum_{i=0}^{2R} v_{r,i} + s_r^2}$ (though that is not a ratio of the variances of the components, which has a far more complicated expression). This effect becomes more important as we go to higher FA coefficients. Another important point to notice is that there is no significant exchange of energy between the linear parts of the signal. That



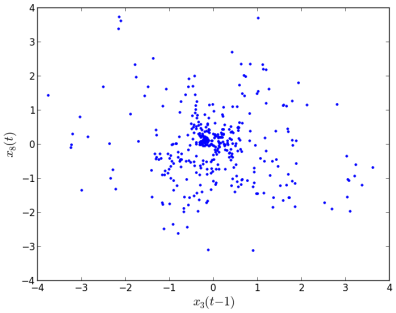
(a) $x_1(t)$ versus $x_1(t-1)$ scatterplot



(b) $x_2(t)$ versus $x_1(t-1)$ scatterplot



(c) $x_5(t)$ versus $x_2(t-1)$ scatterplot

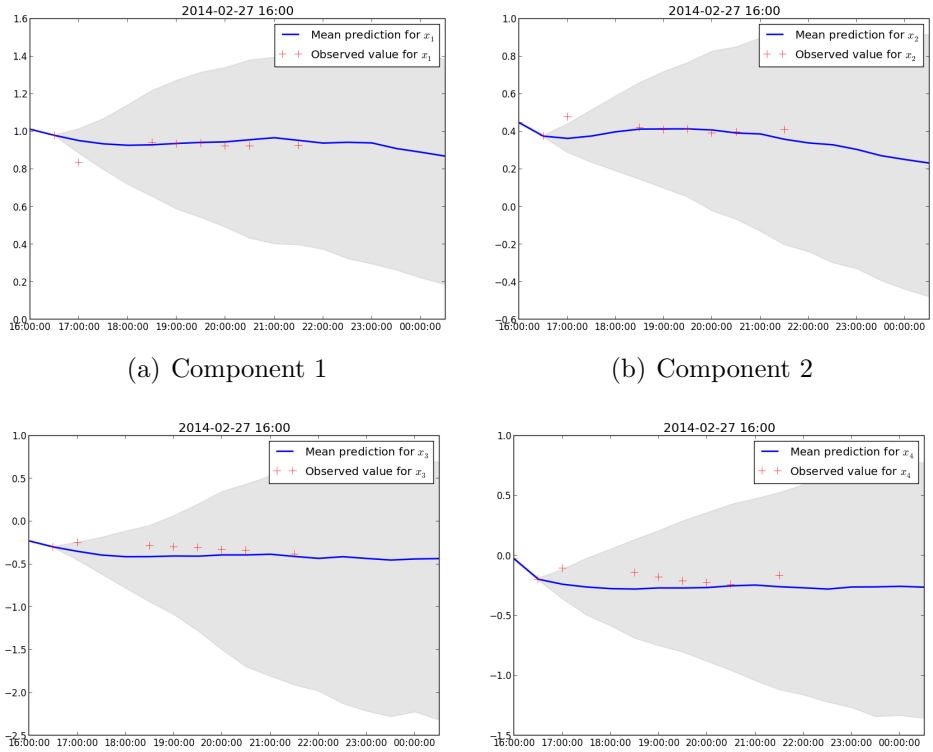


(d) $x_8(t)$ versus $x_3(t-1)$ scatterplot

Figure 6: Scatterplots of $x_i(t)$ by $x_j(t-1)$, for $(i,j) = (1,1), (1,2), (2,5), (3,8)$

is, $x_r(t)$ depends linearly only on $z_r(t) = x_r(t-1)$ and $z_{r+R}(t) = x_r(t-2)$. As expected, the dependence on $x_r(t-1)$ is larger than on $x_r(t-2)$. An explanation of this phenomenon appears in Fig. 6. We see that the scatterplots of $x_i(t)$ and $x_j(t-1)$ for various values of $j \neq i$ appear to be null plots: essentially indistinguishable from the plot of independent random variables. This phenomenon is essentially the same for other $i \neq j$ cases not displayed and this explains the insignificant $v_{r,i}$ coefficients for $r \neq i$. The length scale parameters $\ell_{r,i}$ of the nonlinear part are interpreted in a reciprocal manner to $v_{r,i}$. In particular, the larger the length scale, the less important the interaction. As for the linear part, again the main effect is that $x_r(t)$ depends mostly on $x_r(t-1)$ and $x_r(t-2)$, though here there is a significant number of cross-component dependence coefficients that are not zero. The parameter σ_r^2 was found to be close to zero in all cases. From the first plot in Fig. 6,

which again is consistent with $x_i(t)$ by $x_i(t-1)$ plots other than the displayed $i = 1$, we see that the noise levels are low and are likely to be masked by the other variance components. The small value of σ_r does not mean that the uncertainty in our predictions will be low (see Eq. (36) for the way that σ_r^2 affects the predictive variance). Even if these parameters were perfectly known and σ_r were 0, the uncertainty could go to 0 only in the limit of infinite data.



(a) Component 1

(b) Component 2

(c) Component 3

(d) Component 4

Figure 7: Median and 95% confidence interval for $\mathbf{x}(t)$ on February 26, 2014.

We can now make forecasts for an arbitrary number of steps ahead. To investigate the suitability of normal densities usage, we start by ploting in Fig. 5 the quantiles for one-step-ahead normalized prediction (shifted by its mean and divided by the standard deviation). This plot is generated by using all the observed values of the irradiance up to March 4, 2014. The plot demonstrates that the model is able to properly quantify the forecast

uncertainty to a satisfactory degree, in the sense that the predictive density and the observed density generally agree. Problems remain, however, near the tails of the predictive distribution. Since the prediction is Gaussian, we can investigate the issue further with the Shapiro-Wilk [37] test. The value of this normality test is $W = 0.9077$, and the corresponding p-value is 0.0022. While the p value is small, we note that when applying the test to a completely misspecified distribution, such as for having a uniform or Cauchy instead of a normal, the p value would be on the order of $1e-9$ for the same amount of data. Hence, we deem the normal distribution assumption to be reasonable for our purposes, particularly since with this amount of data a meaningful alternative seems difficult to posit.

We show results for two representative days: February 26, 2014, which was a sunny day, and March 1, 2014, which was a cloudy day. To initiate our forecast, we observe the clearness index at 16:00 and 16:30 UTC. Then, we make predictions for 8 hours ahead by sampling Eq. (36) recursively as explained in Alg. 1. For the computation of the quantiles we present here, we use 1,000 such samples.

Forecast for February 27, 2014 (sunny day). Figure 7 depicts the evolution of the median and 95% confidence intervals of our prediction for the first four components of $\mathbf{x}(t)$. We notice that the observed values always fall within the error bars. Figures 8 and 9 show the evolution of the mean and the standard deviation of $\mathbf{C}(t)$, respectively. Since the day is sunny, the mean remains relatively constant. Notice that the standard deviation of $\mathbf{C}(t)$ does increase with time but not very fast. In Fig. 10 we compare the observed $\mathbf{C}(t)$ with our mean predictions at three different times. Notice that high-frequency features are not captured properly. This result is expected, however, since the model is built on the projected $\mathbf{C}(t)$. In Fig. 11, we compare the error between projection of $\mathbf{C}(t)$ and our mean predictions, with the forecasted standard deviation. We notice that the error is indeed well bounded by the standard deviation, which shows that our model is at least conservative. In Fig. 12 we compare our point predictions for the solar irradiation at Lamont, OK, with that of a recursive GP model based only on the ground observations, which can be seen as our model when using only one component but which is obtained directly from the data. In this case the model has 7 parameters and 350 observations. We see that the mean prediction is considerably better for our model compared with the one that uses only data from the prediction site, particularly a few hours after the start of the forecast window. Moreover,

the standard deviation of the global data approach is smaller. We note here that a discrepancy between the satellite and the ground observations. This is expected, since the satellite solar irradiation is averaged over a square kilometer while the ground measurement is almost pointwise. Overall we find that the uncertainty margins are improved and that the predictions are vastly improved when using satellite data compared with only site data.

Forecast for March 1, 2014 (cloudy day). Figure 13 depicts the evolution of the median and 95% confidence intervals of our prediction for $\mathbf{x}(t)$. We notice that $\mathbf{x}(t)$ takes more extreme values compared with Fig. 7. As a result, some of the observations and, in particular, those of $x_2(t)$ do not always fall within the error bars, though their errors are certainly on the order of the standard deviations. Figures 14 and 15 show the evolution of the mean and the standard deviation of $\mathbf{C}(t)$, respectively. Here we notice a much larger variation of the mean prediction compared with Figures 8 and 9. In addition, the standard deviation increases much more rapidly. In Fig. 16 we compare the observed $\mathbf{C}(t)$ with our mean predictions at three different times. Here, the error is much larger, and the standard deviation captures the order of the larger errors but not their pattern, though it does so for the large feature persistent error. A look at the projected version of the error (see Fig. 17) validates this observation. Another interesting property is that the standard deviation increases faster near the boundaries of the patch than at the center. The cause is that effects near the boundary depend also on the clearness index field that is outside our small patch. In Fig. 18 we compare our point predictions for the solar irradiation at Lamont, OK, with that of a recursive GP model based only on ground observations. The mean prediction has smaller error for the satellite data compared with the site data, whereas the standard deviation is significantly smaller after a couple of hours. The performance of our model is reasonable and certainly better than the model based on ground observations. However, sharp variations such as the ramps occurring at 20:00 UTC are not captured. We believe that in order to capture ramp events one has to model the finer frequencies, which is not done in the version of the model. We note here that ramps could not be captured even with $R = 8$, which was the largest model that our limited observations allowed us to train. The discrepancy between the measurements at 16:00 and 16:30 UTC and the starting points of our forecasts is due to the fact that with $R = 8$ the clearness index cannot be reconstructed perfectly from the reduced space. Such a discrepancy is present also at 16:00 UTC in

Fig. 12 but much less pronounced. However, we also note that our pPCA approach reasonably captures that uncertainty.

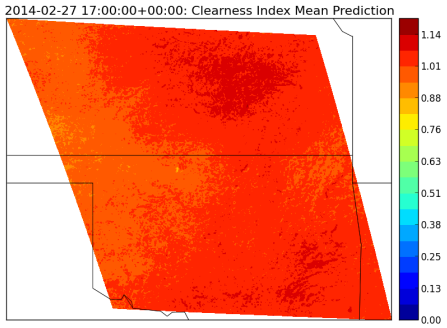
In any case, we again observe a significant improvement in forecast quality in comparison to the model that uses only site data, though in this case the improvement is more pronounced in the variance estimation.

4. Conclusions

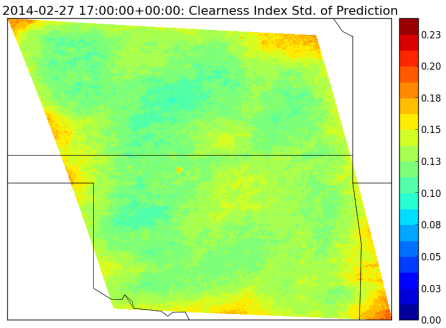
We presented a probabilistic forecast model for solar irradiation based on satellite observations. The approach is based on separate modeling of the clear sky model and of the clearness index, the ratio of irradiation and the clear sky model. Since good clear sky models exist, only the clearness index needs to be modeled statistically. The resulting model yields probabilistic estimates of the irradiation field every thirty minutes starting from consecutive satellite measurements. The model is based on a probabilistic PCA approach to reduce the forecast dimension, followed by a Gaussian process approach to learn its dynamics. Using ground data from Lamont, OK, we demonstrate that the model results in better predictions and uncertainty estimates for the solar radiation compared with models that use only ground data at the prediction site, which is the prevailing forecast method for solar energy operations. Moreover, since the data are available over the entire United States, forecasts can be produced at any site in the United States with (what we anticipate) to be comparable accuracy, without needing ground measurements. This may prove important for very distributed solar energy systems.

This initial investigation can be improved in multiple ways, not the least of which is access to more data as the AVHRR instrument acquires more measurements. This extra data will allow us to consider more components and richer covariance functions, and, in particular, to model the larger space frequencies which we currently do not do as well as the satellite resolution errors. As a result, we may be able to represent higher-fidelity spatial features such as the ramps that we cannot capture well for the March 1, 2014, forecast.

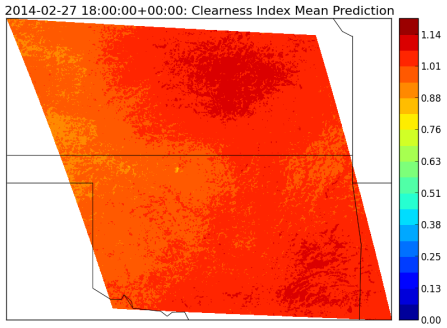
We note, however, that in some applications solar energy sources are spread over large (and perhaps well-chosen) areas. In such cases, the fact that we capture well the low-frequency components of the error (interpreted here as the projection on the space of the principal components, as is seen in Figs. 11 and 17 indicates that spatially aggregated statistics of available solar power are likely to be well captured. However, we would have no way of obtaining that much ground data to validate the predictions. Nevertheless,



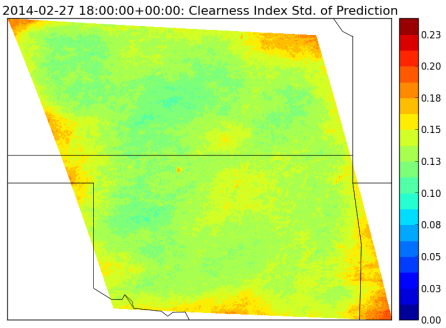
(a) 2014-02-27 17:00 UTC, mean



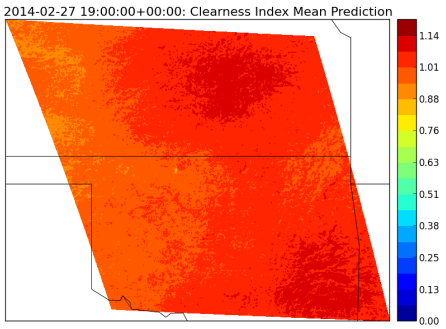
(b) 2014-02-27 17:00 UTC, std.



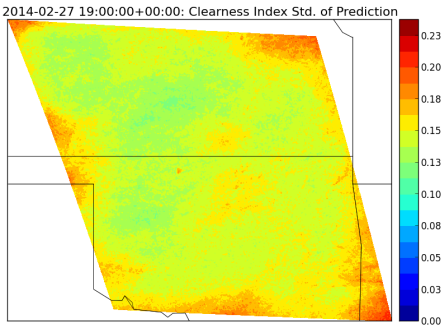
(c) 2014-02-27 18:00 UTC, mean



(d) 2014-02-27 18:00 UTC, std.

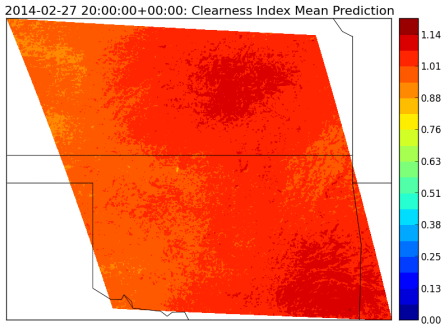


(e) 2014-02-27 19:00 UTC, mean

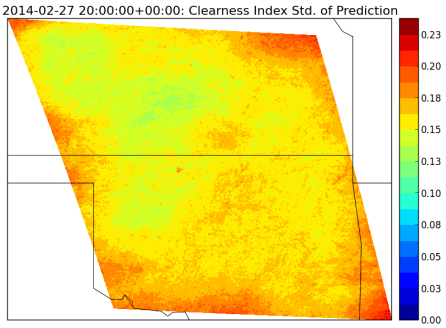


(f) 2014-02-27 19:00 UTC, std.

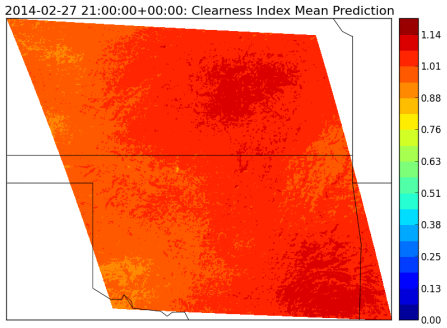
Figure 8: Starting with two observations on February 27, 2014, at 16:00 and 16:30 UTC we take 1,000 sample paths of the clearness index $\mathbf{C}(t)$ following Alg. 1. The left column shows the mean forecast and the right column the standard deviation. The forecast is shown for every hour from 17:00 to 19:00 UTC.



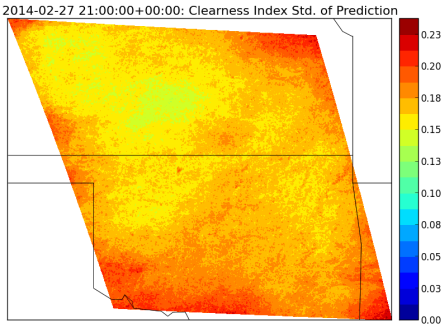
(a) 2014-02-27 20:00 UTC, mean



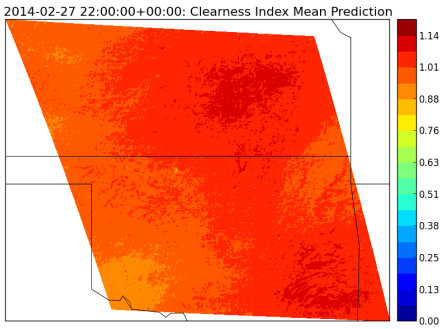
(b) 2014-02-27 20:00 UTC, std.



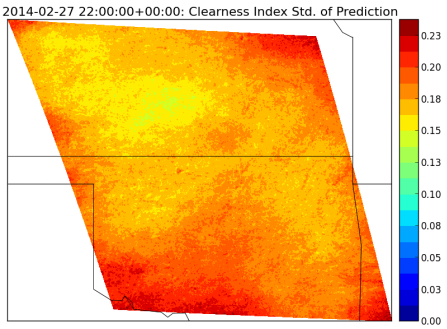
(c) 2014-02-27 21:00 UTC, mean



(d) 2014-02-27 21:00 UTC, std.



(e) 2014-02-27 22:00 UTC, mean



(f) 2014-02-27 22:00 UTC, std.

Figure 9: Starting with two observations on February 27, 2014, at 16:00 and 16:30 UTC we take 1,000 sample paths of the clearness index $\mathbf{C}(t)$ following Alg. 1. The left column shows the mean forecast and the right column the standard deviation. The forecast is shown for every hour from 20:00 to 22:00 UTC.

looking at the plots of the ground irradiation and satellite radiation data, we find a good agreement, which indicates that this is likely to be true. Moreover, one would need to choose a pattern for the solar plants, which requires obtaining some realistic configurations; this will be the subject of future analyses.

Another interesting direction is to combine these measurements with weather forecast information, which –by itself– has difficulties in representing accurately irradiation information because of poor high-resolution cloud forecasting. However, since weather forecast may capture certain large-scale trends (e.g. wind), such information when combined with this approach conceivably may result in better forecasts.

Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Science, under Contract No. DE-AC02-06CH11357. We gratefully acknowledge the use of the Blues cluster in the Laboratory Computing Resource Center at Argonne National Laboratory. We thank Christine Molling from the Cooperative Institute for Meteorological Satellite Studies, Space Science and Engineering Center, at the University of Wisconsin-Madison for providing access to CLAVRx Satellite data. We also acknowlegde the Atmospheric Radiation Measurement (ARM) Program for providing access to the SGP ground irradiance measurements.

Appendix A. Details of the clear sky model

As discussed in Sec. 2.2, we use Ineichen's model [12] as our clear sky model. Here we briefly discuss the details of its implementation that pertain to this work. For the actual model, the reader may consult the original reference. Ineichen's model requires the solar zenith, the extraterrestrial irradiation, the surface elevation, the water vapor content, and the aerosol optical depth at 700 nm.

The solar zenith is computed by using the Solar Position Algorithm (SPA) [38]. The extraterrestrial irradiation is modeled as [39]

$$I_{\text{ext}}(t) = I_0 \left(\frac{R(t)}{R_{\text{av}}} \right)^2 , \tag{A.1}$$

where $R(t)$ is the Earth-Sun distance at time t, $I_0 = 1361$ W/m^2 and $R_{\text{av}} = 1.0000010178$ AU is the average Earth-Sun distance. The Earth-Sun distance is also computed by using SPA. Since the sensitivity of the model to the surface elevation is small and the patch of the CONUS scan we are concentrating on is relatively flat, we take the surface elevation to be constant and equal to 320 m. The water vapor content and the aerosol optical depth at 700 nm are also taken to be constant at 0.3 cm and 0.15, respectively.

References

[1] G. Box, G. Jenkins, G. Reinsel, Time Series Analysis: Forecasting and Control, Wiley Series in Probability and Statistics, Wiley, 2008.
URL http://books.google.com/books?id=lJnnPQAACAAJ

[2] J. Boland, Time series modelling of solar radiation, in: V. Badescu (Ed.), Modeling Solar Radiation at the Earths Surface, Springer Berlin Heidelberg, 2008, pp. 283–312. doi:10.1007/978-3-540-77455-6_11.
URL http://dx.doi.org/10.1007/978-3-540-77455-6_11

[3] J. Kleissl, Solar Energy Forecasting and Resource Assessment, Academic Press, 2013.

[4] A. Mellit, Artificial intelligence technique for modelling and forecasting of solar radiation data: a review, Int. J. Artif. Intell. Soft Comput. 1 (1) (2008) 52–76. doi:10.1504/IJAISC.2008.021264.
URL http://dx.doi.org/10.1504/IJAISC.2008.021264

[5] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, B. Washom, Intra-hour forecasting with a total sky imager at the {UC} san diego solar energy testbed, Solar Energy 85 (11) (2011) 2881 – 2893. doi:http://dx.doi.org/10.1016/j.solener.2011.08.025.
URL http://www.sciencedirect.com/science/article/pii/S0038092X11002982

[6] E. Lorenz, D. Heinemann, A. Hammer, Short-term forecasting of solar radiation based on satellite data, in: EuroSun Conference Freiburg, Germany.

[7] A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorenz, R. Mller, H. G. Beyer, Solar energy assessment using remote sensing technologies, Remote Sensing of Environment

86 (3) (2003) 423 – 432, urban Remote Sensing. doi:http://dx.doi.org/10.1016/S0034-4257(03)00083-X.
URL http://www.sciencedirect.com/science/article/pii/S003442570300083X

[8] J. H. Eto, R. J. Thomas, Computational needs for the next generation electric grid, department of Energy, (2011).
URL http://energy.gov/sites/prod/files/FINAL_CompNeeds_Proceedings2011.pdf

[9] E. M. Constantinescu, V. M. Zavala, M. Rocklin, S. Lee, M. Anitescu, A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation, Power Systems, IEEE Transactions on 26 (1) (2011) 431–441.

[10] I. Laszlo, P. Ciren, H. Liu, S. Kondragunta, J. D. Tarpley, M. D. Goldberg, Remote sensing of aerosol and radiation from geostationary satellites, Advances in Space Research 41 (11) (2008) 1882 – 1893. doi:http://dx.doi.org/10.1016/j.asr.2007.06.047.
URL http://www.sciencedirect.com/science/article/pii/S0273117707007156

[11] R. E. Bird, R. L. Hulstrom, A simplified clear sky model for direct and diffuse insolation on horizontal surfaces, Tech. Rep. SERI/TR-335-76181, Solar Energy Research Institute (1981).

[12] P. Ineichen, A broadband simplified version of the solis clear sky model, Solar Energy 82 (8) (2008) 758 – 762. doi:http://dx.doi.org/10.1016/j.solener.2008.02.009.
URL http://www.sciencedirect.com/science/article/pii/S0038092X08000406

[13] R. Mueller, K. Dagestad, P. Ineichen, M. Schroedter-Homscheidt, S. Cros, D. Dumortier, R. Kuhlemann, J. Olseth, G. Piernavieja, C. Reise, L. Wald, D. Heinemann, Rethinking satellite-based solar irradiance modelling: The {SOLIS} clear-sky module, Remote Sensing of Environment 91 (2) (2004) 160 – 174. doi:http://dx.doi.org/10.1016/j.rse.2004.02.009.
URL http://www.sciencedirect.com/science/article/pii/S0034425704000690

[14] B. Everitt, An Introduction to Latent Variable Models, Monographs on statistics and applied probability, Chapman and Hall, 1984.
URL http://books.google.com/books?id=2UglOgAACAAJ

[15] A. Basilevsky, Statistical Factor Analysis and Related Methods: Theory and Applications, Wiley Series in Probability and Statistics, Wiley, 2009.
URL http://books.google.com/books?id=uvZkWmRTZhUC

[16] D. Bartholomew, M. Knott, I. Moustaki, Latent Variable Models and Factor Analysis: A Unified Approach, Wiley Series in Probability and Statistics, Wiley, 2011.
URL http://books.google.com/books?id=5YxvE1S56DwC

[17] S. Roweis, Em algorithms for pca and spca, in: in Advances in Neural Information Processing Systems, MIT Press, 1998, pp. 626–632.

[18] M. E. Tipping, C. M. Bishop, Probabilistic principal component analysis, Journal of the Royal Statistical Society, Series B 61 (1999) 611–622.

[19] K. Pearson, On lines and planes of closest fit to systems of points in space, Philosophical Magazine 2 (1901) 559–572.

[20] H. Hotelling, Analysis of a complex of statistical variables into principal components, Journal of Educational Psychology 24.

[21] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[22] D. Rubin, D. Thayer, EM algorithms for ML factor analysis, Psychometrika 47 (1) (1982) 69–76. doi:10.1007/bf02293851.
URL http://dx.doi.org/10.1007/bf02293851

[23] T. P. Minka, Automatic choice of dimensionality for PCA, Tech. Rep. 514 (2000).

[24] A. Girard, J. Q. Candela, R. Murray-smith, C. E. Rasmussen, Gaussian process priors with uncertain inputs - application to multiple-step ahead time series forecasting (2003).

[25] J. M. Wang, D. J. Fleet, A. Hertzmann, Gaussian process dynamical models, in: In NIPS, MIT Press, 2006, pp. 1441–1448.

[26] C. K. Liu, A. Hertzmann, Z. Popovic, Learning physics-based motion style with nonlinear inverse optimization, ACM Trans. Graph 24 (2005) 1071–1081.

[27] L. Raskin, E. Rivlin, M. Rudzsky, Using gaussian process annealing particle filter for 3d human tracking, EURASIP Journal on Advances in Signal Processing.

[28] F. Perez-Cruz, S. Van Vaerenbergh, J. J. Murillo-Fuentes, M. Lazaro-Gredilla, I. Santamaria, Gaussian Processes for Nonlinear Signal Processing: An Overview of Recent Advances, IEEE Signal Processing Magazine 30 (4) (2013) 40–50. doi:10.1109/msp.2013.2250352.
URL http://dx.doi.org/10.1109/msp.2013.2250352

[29] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, USA, 2006.

[30] S. Conti, A. O'Hagan, U. of Sheffield. Dept. of Probability, Statistics, Bayesian Emulation of Complex Multi-output and Dynamic Computer Models, Research report (University of Sheffield. Dept. of Probability and Statistics), Department of Probability & Statistics, University of Sheffield, 2007.
URL http://books.google.com/books?id=0bQOMwEACAAJ

[31] M. Alvarez, N. D. Lawrence, Sparse Convolved Gaussian Processes for Multi-output Regression, in: NIPS, 2008, pp. 57–64.
URL http://books.nips.cc/papers/files/nips21/NIPS2008_0170.pdf

[32] I. Bilionis, N. Zabaras, Multi-output local gaussian process regression: Applications to uncertainty quantification, Journal of Computational Physics 231 (17) (2012) 5718 – 5746. doi:http://dx.doi.org/10.1016/j.jcp.2012.04.047.
URL http://www.sciencedirect.com/science/article/pii/S0021999112002513

[33] I. Bilionis, N. Zabaras, B. A. Konomi, G. Lin, Multi-output separable gaussian process: Towards an efficient, fully bayesian paradigm for uncertainty quantification, Journal of Computational Physics 241 (0) (2013) 212 – 239. doi:http://dx.doi.org/10.1016/j.jcp.2013.01.011.
URL http://www.sciencedirect.com/science/article/pii/S0021999113000417

[34] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning), The MIT Press, 2005.

[35] Atmospheric Radiation Measurement (ARM) Climate Research Facility. 1994, updated hourly. Multifilter Radiometer (MFR10M). 2013-10-01 to 2014-03-05, 36.605 N 97.485 W: Southern Great Plains (SGP) Central Facility, Lamont, OK (C1). Compiled by G. Hodges. Atmospheric Radiation Measurement (ARM) Climate Research Facility Data Archive: Oak Ridge, Tennessee, USA. Data set accessed 2014-03-07 at http://dx.doi.org/10.5439/1025224.

[36] M. Avriel, Nonlinear Programming: Analysis and Methods, Prentice-Hall, Englewood Cliffs, NJ, 1976.
URL http://www.ams.org/mathscinet-getitem?mr=58:9264

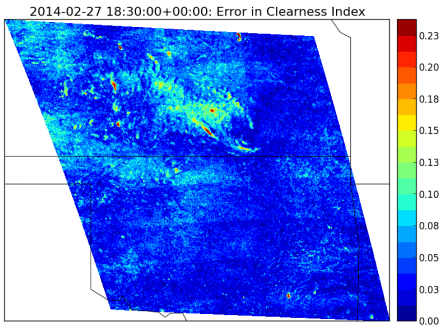
[37] S. S. Shapiro, M. B. Wilk, An analysis of variance test for normality (complete samples), Biometrika 52 (3/4) (1965) 591–611.
URL http://links.jstor.org/sici?sici=0006-3444%28196512%2952%3A3%2F4%3C591%3AAAOVTF%3E2.0.CO%3B2-B

[38] I. Reda, A. Andreas, Solar position algorithm for solar radiation applications, Solar Energy 76 (5) (2004) 577 – 589. doi:http://dx.doi.org/10.1016/j.solener.2003.12.003.
URL http://www.sciencedirect.com/science/article/pii/S0038092X0300450X

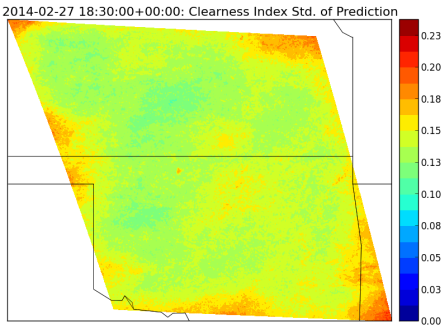
[39] G. W. Paltridge, C. M. R. Platt, Radiative processes in meteorology and climatology / by G. W. Paltridge and C. M. R. Platt, Elsevier Scientific Pub. Co Amsterdam ; New York, 1976.

Parameter	1	2	3	4	5	6	7	8
$v_{r,0}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$v_{r,1}$	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$v_{r,2}$	0.00	0.24	0.00	0.00	0.00	0.00	0.00	0.00
$v_{r,3}$	0.00	0.00	0.35	0.00	0.01	0.00	0.00	0.00
$v_{r,4}$	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00
$v_{r,5}$	0.00	0.00	0.00	0.00	0.32	0.00	0.00	0.00
$v_{r,6}$	0.00	0.00	0.00	0.00	0.00	0.24	0.00	0.00
$v_{r,7}$	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.00
$v_{r,8}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30
$v_{r,9}$	1.72	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$v_{r,10}$	0.00	0.89	0.00	0.00	0.00	0.00	0.00	0.00
$v_{r,11}$	0.00	0.00	2.39	0.00	0.01	0.00	0.00	0.00
$v_{r,12}$	0.00	0.00	0.00	2.01	0.00	0.00	0.00	0.00
$v_{r,13}$	0.00	0.00	0.00	0.00	2.25	0.00	0.00	0.00
$v_{r,14}$	0.00	0.00	0.00	0.00	0.00	2.02	0.00	0.00
$v_{r,15}$	0.00	0.00	0.00	0.00	0.00	0.00	4.27	0.00
$v_{r,16}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.08
s_r^2	0.01	0.01	0.02	0.02	0.04	0.04	0.07	0.06
$\ell_{r,1}$	0.17	*	*	*	*	*	*	1.20
$\ell_{r,2}$	2.24	2.19	*	*	1.22	0.44	*	1.40
$\ell_{r,3}$	*	*	*	*	0.87	*	*	2.02
$\ell_{r,4}$	*	*	2.69	*	5.99	*	*	3.91
$\ell_{r,5}$	*	*	*	3.79	*	1.37	*	*
$\ell_{r,6}$	*	*	*	0.21	*	*	*	*
$\ell_{r,7}$	*	*	*	*	*	*	1.68	*
$\ell_{r,8}$	*	*	*	3.08	0.96	*	0.98	0.46
$\ell_{r,9}$	0.11	0.26	1.16	1.26	*	0.78	*	*
$\ell_{r,10}$	*	1.09	*	*	*	0.89	0.91	*
$\ell_{r,11}$	*	1.65	0.31	3.24	0.80	*	*	*
$\ell_{r,12}$	1.52	0.63	*	*	*	*	2.40	0.96
$\ell_{r,13}$	*	*	*	*	0.29	*	8.32	*
$\ell_{r,14}$	*	*	5.70	0.39	*	0.32	4.07	4.16
$\ell_{r,15}$	*	*	1.50	4.14	*	*	0.75	*
$\ell_{r,16}$	*	*	0.59	6.07	*	*	*	0.37
σ_r^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

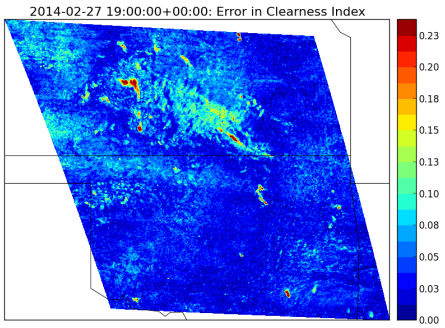
Table 1: The infered parameters of the dynamic model. The '*' stand for length scales that were estimated to be larger than 10 and, therefore, unimportant.



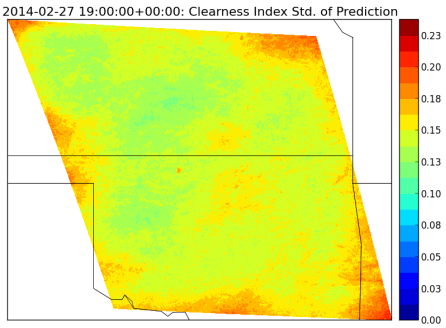
(a) 2014-02-27 18:30 UTC, error in prediction



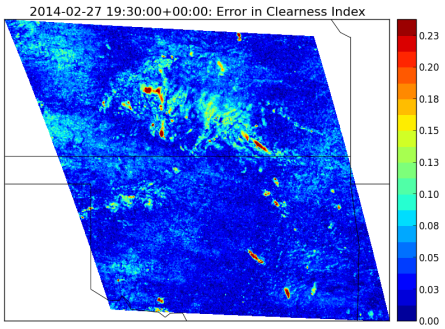
(b) 2014-02-27 18:30 UTC, std.



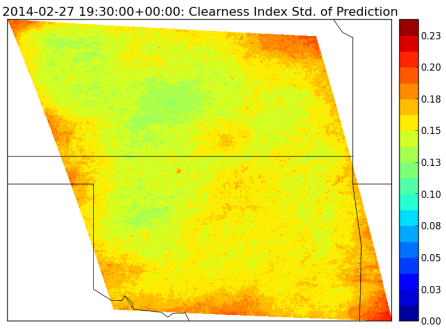
(c) 2014-02-27 19:00 UTC, error in prediction



(d) 2014-02-27 19:00 UTC, std.

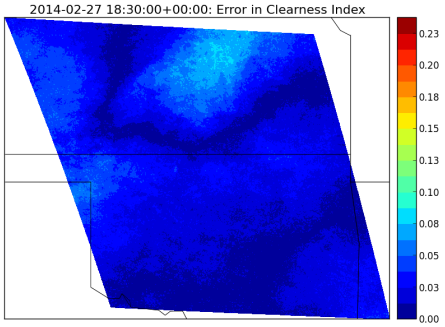


(e) 2014-02-27 19:30 UTC, error in prediction

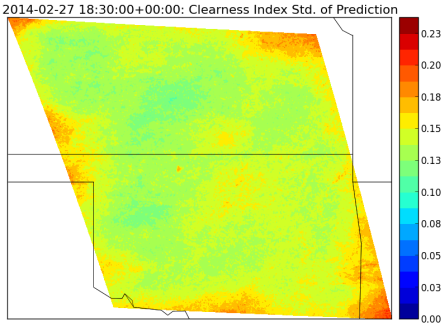


(f) 2014-02-27 19:30 UTC, std.

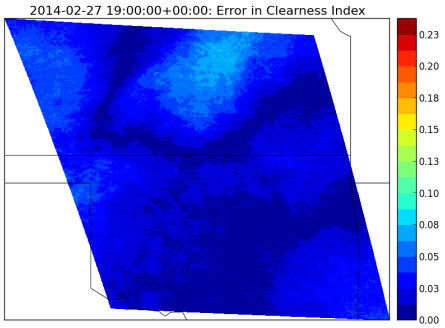
Figure 10: Comparison of the real absolute error in the prediction (left) with the standard deviation of our forecast (right) on February 27, 2014.



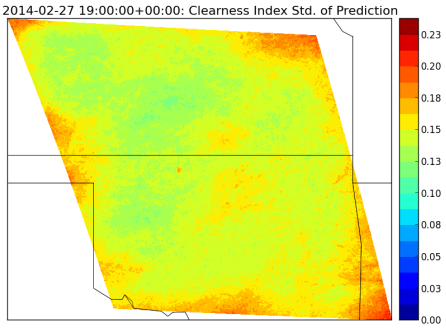
(a) 2014-02-27 18:30 UTC, projected error in prediction



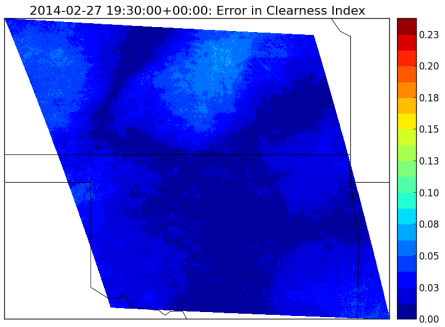
(b) 2014-02-27 18:30 UTC, std.



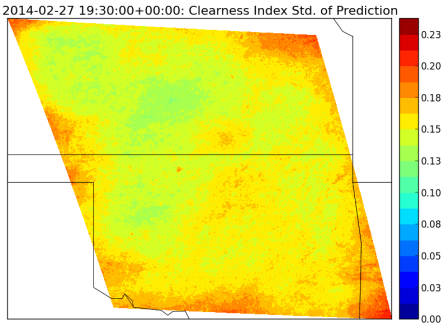
(c) 2014-02-27 19:00 UTC, projected error in prediction



(d) 2014-02-27 19:00 UTC, std.



(e) 2014-02-27 19:30 UTC, projected error in prediction



(f) 2014-02-27 19:30 UTC, std.

Figure 11: Comparison of the absolute projected error in the prediction (left) with the standard deviation of our forecast (right) on February 27, 2014.

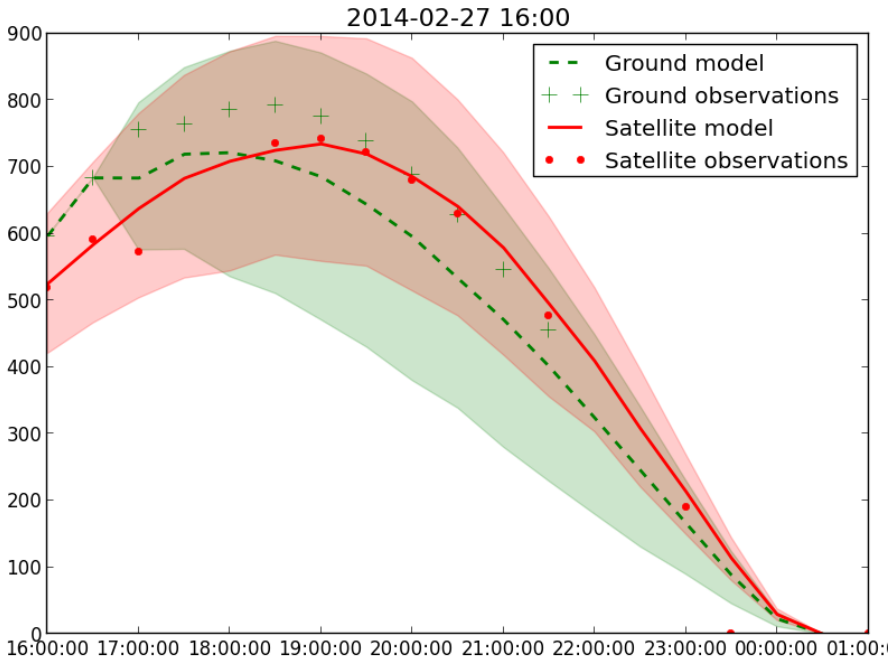
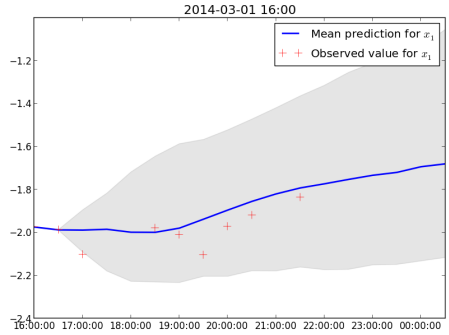
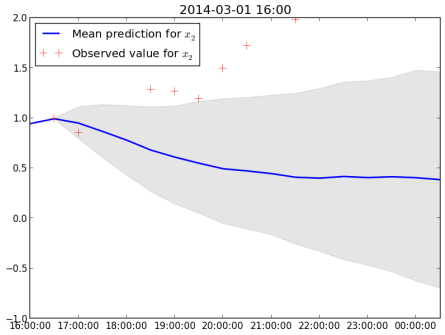


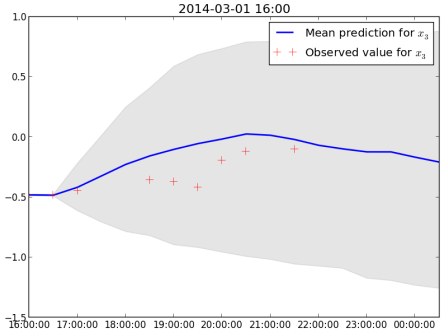
Figure 12: Comparison of our point forecast at Lamont, OK, with a recursive GP model based solely on ground observations.



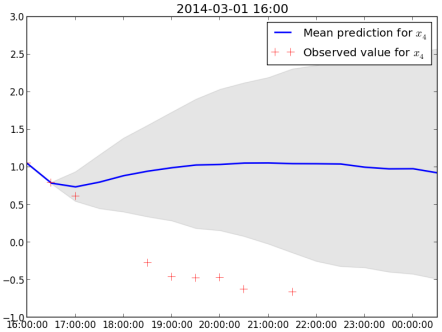
(a) Component 1



(b) Component 2

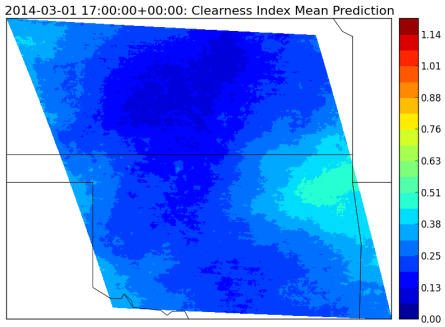


(c) Component 3

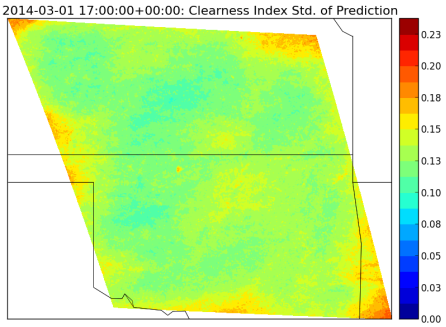


(d) Component 4

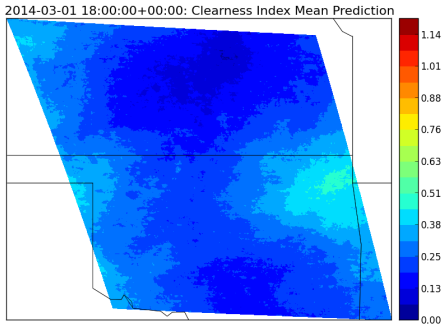
Figure 13: Median and 95% confidence interval for $\mathbf{x}(t)$ on March 1, 2014.



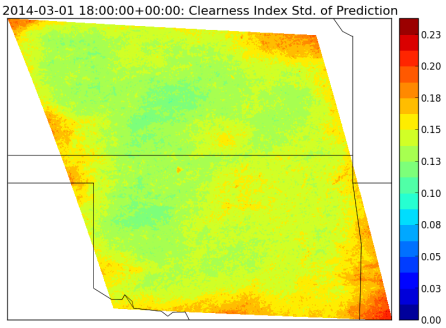
(a) 2014-03-01 17:00 UTC, mean



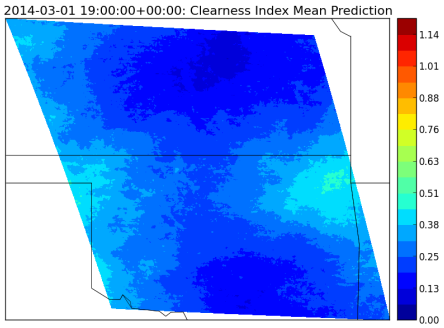
(b) 2014-03-01 17:00 UTC, std.



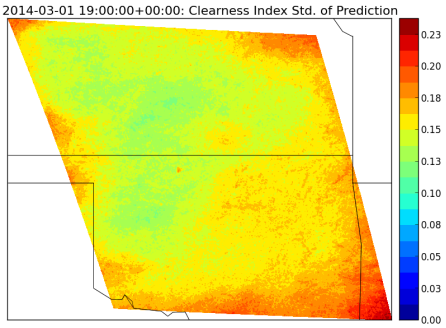
(c) 2014-03-01 18:00 UTC, mean



(d) 2014-03-01 18:00 UTC, std.

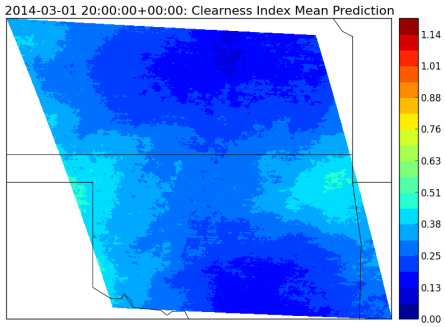


(e) 2014-03-01 19:00 UTC, mean

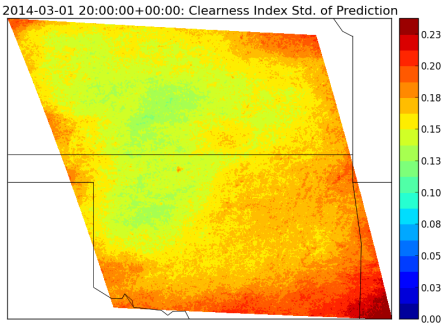


(f) 2014-03-01 19:00 UTC, std.

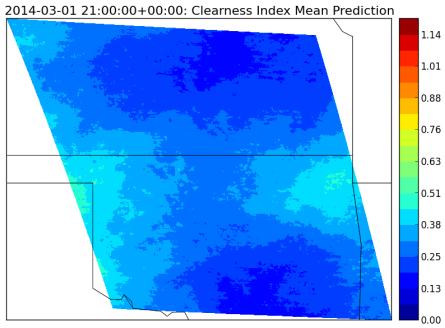
Figure 14: Starting with two observations on March 1, 2014, at 16:00 and 16:30 UTC we take 1,000 sample paths of the clearness index $\mathbf{C}(t)$ following Alg. 1. The left column shows the mean forecast and the right column the standard deviation. The forecast is shown for every hour from 17:00 to 19:00 UTC.



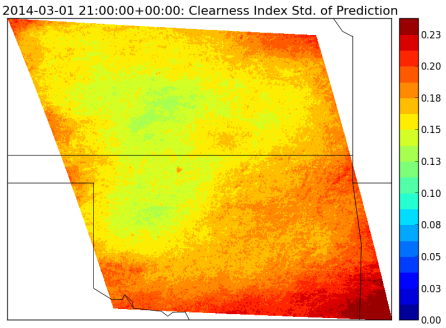
(a) 2014-03-01 20:00 UTC, mean



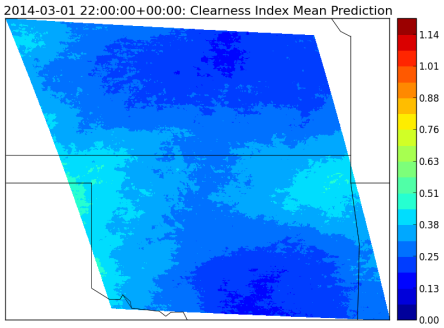
(b) 2014-03-01 20:00 UTC, std.



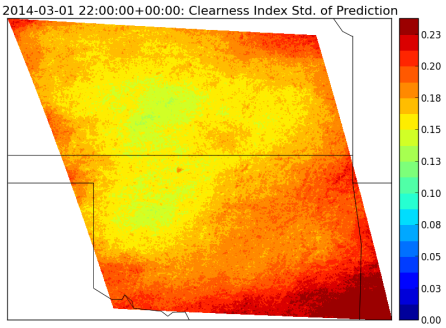
(c) 2014-03-01 21:00 UTC, mean



(d) 2014-03-01 21:00 UTC, std.

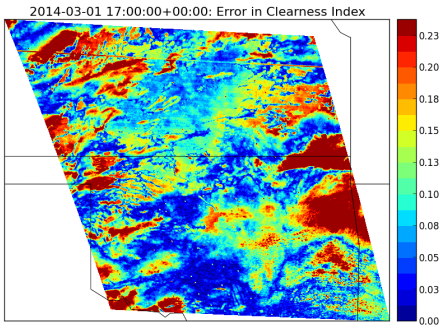


(e) 2014-03-01 22:00 UTC, mean

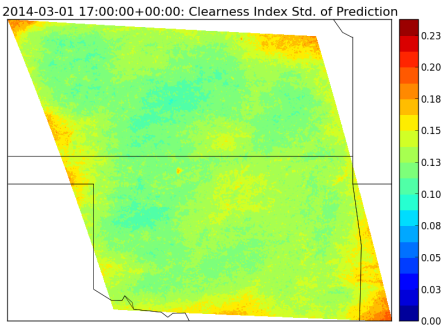


(f) 2014-03-01 22:00 UTC, std.

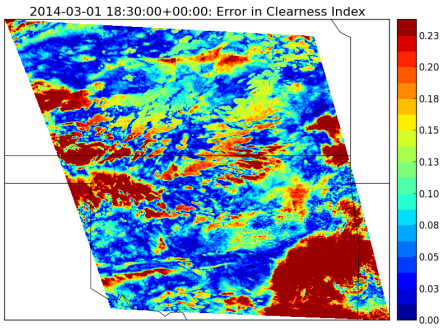
Figure 15: Starting with two observations on March 1, 2014, at 16:00 and 16:30 UTC we take 1,000 sample paths of the clearness index $\mathbf{C}(t)$ following Alg. 1. The left column shows the mean forecast and the right column the standard deviation. The forecast is shown for every hour from 20:00 to 22:00 UTC.



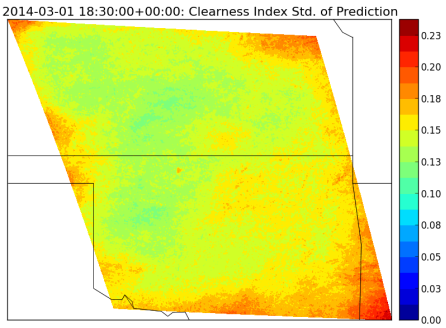
(a) 2014-03-01 17:00 UTC, error in prediction



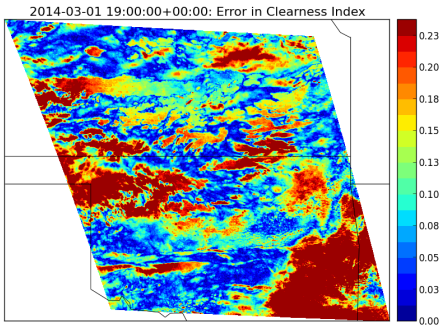
(b) 2014-03-01 17:00 UTC, std.



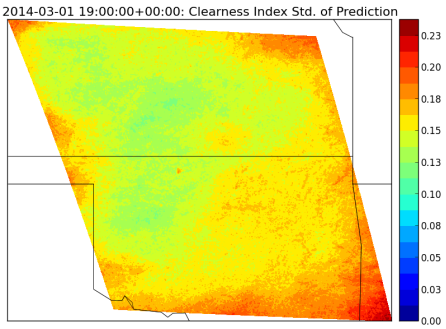
(c) 2014-03-01 18:30 UTC, error in prediction



(d) 2014-03-01 18:30 UTC, std.

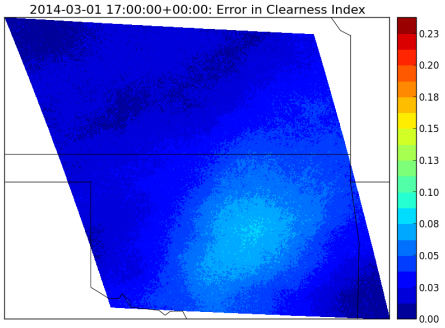


(e) 2014-03-01 19:00 UTC, error in prediction

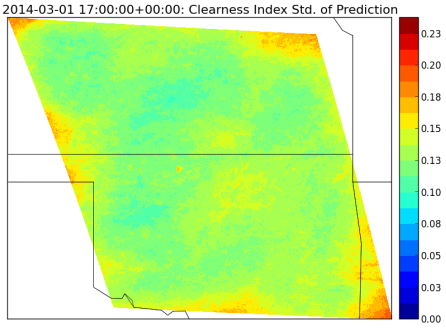


(f) 2014-03-01 19:00 UTC, std.

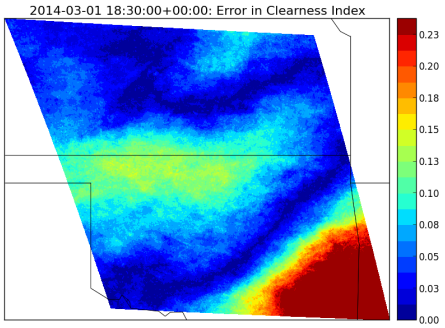
Figure 16: Comparison of real absolute error in the prediction (left) with the standard deviation of our forecast (right) on March 1, 2014.



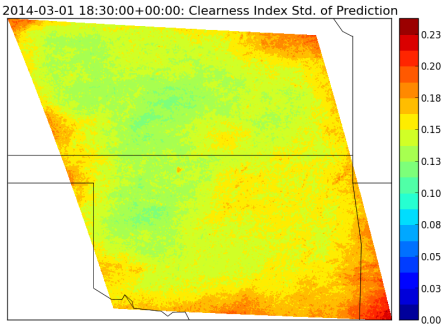
(a) 2014-03-01 17:00 UTC, projected error in prediction



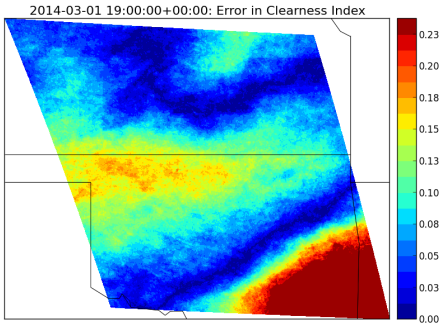
(b) 2014-03-01 17:00 UTC, std.



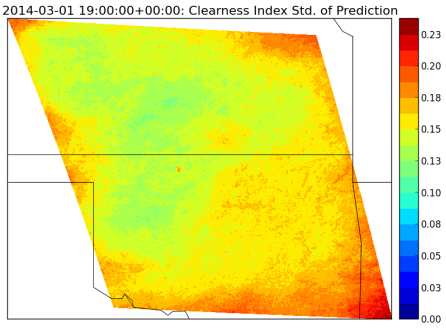
(c) 2014-03-01 18:30 UTC, projected error in prediction



(d) 2014-03-01 18:30 UTC, std.



(e) 2014-03-01 19:00 UTC, projected error in prediction



(f) 2014-03-01 19:00 UTC, std.

Figure 17: Comparison of absolute projected error in the prediction (left) with the standard deviation of our forecast (right) on March 1, 2014.

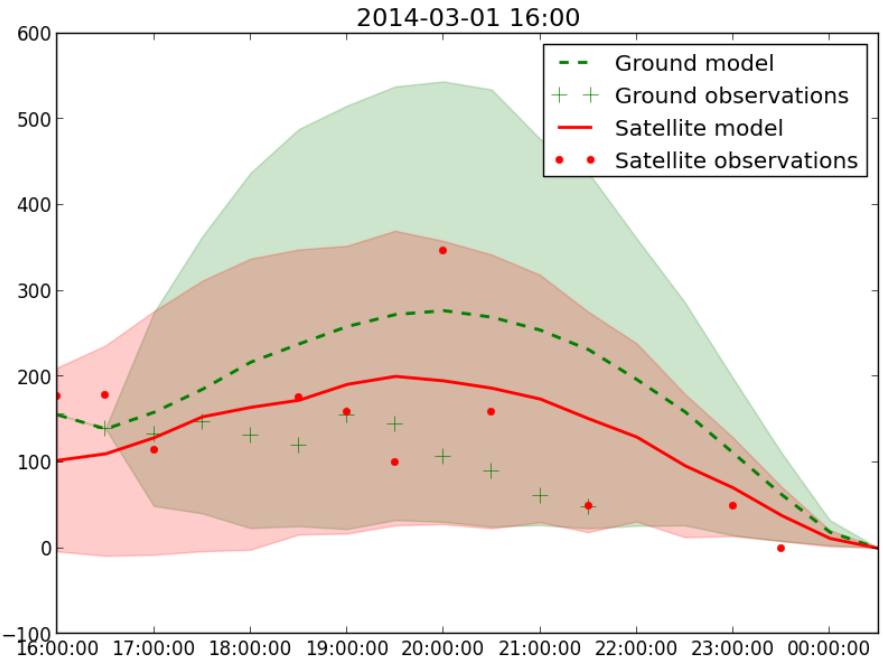


Figure 18: Comparison of our point forecast at Lamont, OK, with a recursive GP model based solely on ground observations.