Temporal interpolation of global surface skin temperature diurnal cycle over land under clear and cloudy conditions

F. Aires

Laboratoire de Météorologie Dynamique, CNRS, France

Department of Applied Physics and Applied Mathematics, NASA Goddard Institute for Space Studies, Columbia University, New York, USA

C. Prigent

Laboratoire de l'Etude du Rayonnement et de la Matiére en Astrophysique, CNRS, Observatoire de Paris, France

W. B. Rossow

NASA Goddard Institute for Space Studies, New York, USA

Received 22 February 2003; revised 30 July 2003; accepted 9 December 2003; published 26 February 2004.

[1] The surface skin temperature is a key parameter at the land-atmosphere interface. An accurate description of its diurnal cycle would not only help estimate the energy exchanges at the interface, it would also enable an analysis of the global surface skin diurnal cycle and its variability within the last 20 years. This study is based on the 3-hourly surface skin temperature estimated by the International Satellite Cloud Climatology Project (ISCCP) from the infrared measurements collected by the polar and geostationary meteorological satellites. The diurnal cycle of surface skin temperature is analyzed almost globally (60N–60S snow-free areas), using a Principal Component Analysis. The first three components are identifyed as the amplitude, the phase, and the width (i.e., daytime duration) of the diurnal cycle and represent 97% of the variability. PCA is used to regularize estimates of the diurnal cycle at a higher time resolution. A new temporal interpolation algorithm, designed to work when only a few measurements of surface temperature are available, is developed based on the PCA representation and an iterative optimization algorithm. This method is very flexible: only temperature measurements are used (no ancillary data), no surface model constraints are used, and the time and number of measurements are not fixed. The performance of this interpolation algorithm is tested for various diurnal sampling configurations. In particular, the potential to use the satellite microwave observations to provide a full diurnal surface temperature cycle in cloudy conditions is investigated. *INDEX TERMS:* 3322 Meteorology and Atmospheric Dynamics: Land/atmosphere interactions; 3360 Meteorology and Atmospheric Dynamics: Remote sensing; *KEYWORDS:* surface skin temperature, diurnal cycle, temporal interpolation

Citation: Aires, F., C. Prigent, and W. Rossow (2004), Temporal interpolation of global surface skin temperature diurnal cycle over land under clear and cloudy conditions, *J. Geophys. Res.*, *109*, D04313, doi:10.1029/2003JD003527.

1. Introduction

[2] A global data set of surface skin temperature providing information on the whole diurnal cycle would constitute a valuable source of information at the land-atmosphere interface. Two main applications of such a data set can be identified.

[3] First, the surface skin temperature could be used to evaluate surface models. Land surface temperature and its difference with the temperature of the overlying air mass govern the exchange of energy and water fluxes at the land-atmosphere interface. Coupled land-atmosphere climate models (Soil/Vegetation/Atmosphere Transfer (SVAT)) are usually forced by in situ surface air temperature measurements, and then they predict land surface skin temperatures. Adequate measurements of the land surface skin temperature over the whole diurnal cycle would not only enable model evaluation, it could also provide an additional constraint on latent and sensible heat flux calculations.

[4] The second main application of a surface skin temperature data set is related to climate change studies. Routine in situ observations of the surface air temperature at meteorological stations are currently the primary source of information to estimate global climate change. They show an increase in the mean global temperature over the last century with an increased rate over the last 20 years [e.g., *Hansen et al.*, 1999; *Jones et al.*, 1999]. Closer analysis shows that the temperature change is different for the daily maxima and minima, resulting in a damping of the

0148-0227/04/2003JD003527$09.00

diurnal temperature amplitude [e.g., *Karl et al.*, 1993; *Easterling et al.*, 1997]. However, incomplete space and time coverage complicates the analysis of global trends from the meteorological network.

[5] The surface skin temperature T_s is not conventionally observed by the meteorological weather station network, although it could be determined from observations with an infrared radiometer, provided that the land surface emissivity is known. T_s can be estimated from satellite infrared radiance observations. Instruments like AVHRR on board polar orbiters provide good spatial resolution with a limited time sampling (two overpasses per day per location), whereas radiometers on board geostationary weather satellites offer adequate sampling of the diurnal cycle but with a poorer spatial resolution. The main problem of satellite infrared measurements of surface skin temperature is their inability to penetrate clouds, limiting them to clear conditions. Clouds not only cover more than half of the globe at a given time, they also alter the radiative energy exchanges, reducing surface insolation and increasing the downward longwave radiation, both of which reduce the diurnal temperature amplitude.

[6] The most extensive data set of land skin temperature available is produced since 1983 at 3-hour intervals over the globe, every 30 km, by the International Satellite Cloud Climatology Project (ISCCP) [*Rossow and Schiffer*, 1999]. It combines all the infrared measurements from polar and geostationary operational weather satellites, yielding full coverage of the globe. Based on infrared measurements, these results are limited to clear-sky conditions [*Rossow and Garder*, 1993], although a statistical estimation is provided for the skin temperature for cloudy scenes. Although the 3-hourly sampling of the ISCCP data set resolves the diurnal cycle, this sampling is performed at constant universal times, not at constant local times. Hence the representation of the diurnal T_s cycle varies in quality with longitude.

[7] Recently, a neural network inversion scheme, including first-guess information, has been developed to retrieve the surface skin temperature over land from passive microwave and infrared radiance observations [*Aires et al.*, 2001; *Prigent et al.*, 2003a; 2003b] using predetermined monthly mean microwave emissivities [*Prigent et al.*, 1997, 1998], ISCCP cloud and surface parameters [*Rossow and Schiffer*, 1999], and the meteorological analysis from National Center for Environmental Prediction (NCEP) [*Kalnay et al.*, 1996] as first guess information. This technique was developed specifically to extend the clear-sky-limited ISCCP results to cloudy situations. However, the microwave observations extracted from the Special Sensor Microwave Imager (SSM/I) on board the Defense Meteorological Satellite Program (DMSP) polar orbiters do not provide the adequate time sampling for a full description of the surface skin temperature cycle under cloudy conditions.

[8] The objective of this study is to develop a method to reconstruct the complete daily surface skin temperature diurnal cycle over the globe from a limited number of observations for both clear and cloudy conditions, based on a statistical analysis of the observations data sets rather than model calculations. *Ignatov and Gutman* [1999] already suggested a statistical method to derive monthly mean T_s diurnal cycle over land based on a lower spatial and time resolution ISCCP product (the monthly mean data set with a spatial resolution of 280 km). *Jin and Dickinson* [1999] supplement the limited time sampling of polar orbiter infrared measurements for clear pixels with an analysis of the surface skin diurnal cycle from model simulations and experiment campaigns. To estimate surface skin temperatures from infrared measurements in cloudy conditions, *Jin* [2000] and *Jin and Dickinson* [2000] propose a technique that uses both neighboring (in time or space) clear pixels and the surface energy balance, combined with an adjustment derived from surface air temperature estimates for pixels under extended or quasi-permanent cloud cover. This method relies on parameterized diurnal cycle shapes dependent on ancillary information such as soil moisture, vegetation, cloud properties, season, and latitude. In a recent study, *Vinnikov et al.* [2002] use a harmonic representation of the diurnal cycles of the surface air temperatures at selected meteorological stations to developed a technique to study seasonal cycles of climatic trends.

[9] This paper focuses on the technical aspects of the temporal interpolation of surface skin temperature diurnal cycles under clear and cloudy conditions. A new methodology is presented and tested in various experiments, some for real conditions, some with synthetic cases to be able to check the results. No surface model constraint or ancillary data are used; only surface skin temperature measurements themselves are required by the technique. This is particularly important for studying the impact on surface temperature of surface parameters like vegetation or soil moisture. Furthermore, the number of satellite surface skin temperature measurements will increase with new instruments (AQUA, TERRA) and our approach, able to deal with variable time and number of measurements, is flexible enough to process such new observations. The interpolation works for daily or monthly diurnal cycles. The method is flexible enough to allow the use of any kind of a priori information. The interpolation technique that is developed is general and can be used in many other contexts like spatial interpolation instead of temporal interpolation and for different variables. (In the spatial interpolation problem, the PCA/iterative approach developed in this paper would be identical, except that the PCA is performed on fields instead of time series, i.e., the diurnal cycles. In this case, the base function are called EOFs (Empirical Orthogonal Functions.) The ISCCP data set is used (section 2) to infer a statistical representation of the diurnal cycle (see *Cairns* [1995] for a similar application for an analysis of the diurnal variation of clouds). The statistical analysis serves several purposes: relating diurnal cycle features to surface and other characteristics, suppressing part of the noise in the measurements (from instrumental noise, retrieval errors, or sampling problems), interpolating the diurnal cycle when data are missing especially for measurements derived from polar orbiting instead of geostationary satellites. Interpolation algorithms are presented in section 3. Various experiments are performed for clear sky conditions to evaluate these algorithms (section 4). Synthetic experiments are also presented to check the extension to cloudy cases. The potential of past and future microwave instruments for the reconstruction of the diurnal cycle of surface skin temperature is then investigated. The

conclusions and perspectives of this work are given in section 5.

2. Statistical Analysis of the Surface Skin Temperature Diurnal Cycle

2.1. ISCCP Database of Surface Skin Temperature

[10] The ISCCP data set is used as a sampling database to estimate the main statistical features of the surface skin temperature diurnal cycles. The ISCCP algorithm discriminates between clear and cloudy scenes and cloud parameters and related quantities are retrieved from visible (VIS $\sim$0.6 μm wavelength) and infrared (IR $\sim$11 μm wavelength) radiances provided by the set of polar and geostationary meteorological satellites [*Rossow and Schiffer*, 1999]. Calibration issues are carefully reviewed (orbital drift, instrument intercalibration): a consistent analysis is provided of the satellite measurements for up to seven satellite simultaneously in orbit, over more than 20 years [*Brest and Rossow*, 1992; *Desormeaux et al.*, 1993]. Surface skin temperature T_s is retrieved from the IR radiance under clear conditions. The original ISCCP product contains values of T_s retrieved assuming that the surface emissivity is unity everywhere. These values are corrected using an emissivity-type specification from the Goddard Institute for Space Studies general circulation model. In any case this correction does not affect the diurnal cycle. If the ISCCP DX scenes are cloudy, a clear-sky compositing procedure is conducted within the ISCCP analysis to derive an estimate of the surface temperatures under cloudy conditions (see *Rossow and Garder* [1993] for more details). The pixel level data set (the DX data set) is selected for its spatial sampling of about 30 km and its sampling interval of 3 hours [*Rossow et al.*, 1996]. The 1999 year is chosen because the presence of Meteosat 5 instrument over India provides more complete coverage. This study will focus on temperature diurnal cycles of land surfaces. T_s diurnal cycles over ocean have a much smaller amplitude compared to the uncertainty in any T_s measurements and different approaches would be required. (For example, one could try to determine first the minimum and maximum temperatures over the day to estimate the amplitude of the cycle and then fit a simple shape to the mean-level and amplitude characteristics.)

2.2. Spline Interpolation

[11] The first step in the processing of the DX diurnal cycle of surface skin temperature converts the GMT times of ISCCP values to local times to enable direct comparison of diurnal cycles at different locations. Second, the 3-hourly ISCCP values are interpolated at a higher temporal resolution for a more accurate description of the diurnal cycle. Cubic Spline Interpolation [*Press et al.*, 1992] is used for interpolation to 1-hour intervals, similar to the method used by *Ignatov and Gutman* [1999]. Piecewise linear interpolation has zero second derivative in the interior of each interval and an undefined second derivative at the measurements points. In contrast, the goal of cubic spline interpolation is to obtain an interpolated curve that has smoothed first derivatives and continuous second derivatives (within the intervals and at the measurement points). Writing down these constraints, a system of linear equations is defined that gives, when solved, the interpolation values.

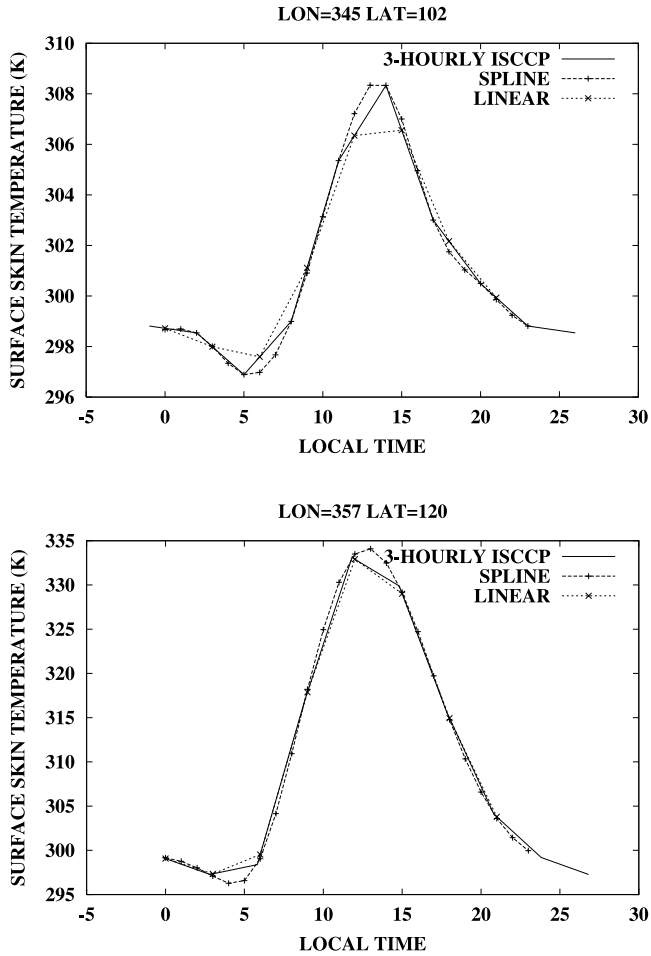

Figure 1. Two samples of the interpolation of a 3-hourly ISCCP surface skin temperature, y, into a 1-hourly cycle, using cubic spline and linear interpolations.

[12] Figure 1 presents two examples of such spline interpolations, along with the corresponding linear interpolation. The linear interpolation tends to damp the diurnal cycle amplitude, whereas the spline interpolation partially avoids this problem by allowing cubic curvatures between the original points. The diurnal amplitude of the spline-interpolated estimates can thus be larger than the amplitude in the original ISCCP T_s, and the time of maximum and minimum values are more reliably located on the cycle interpolated by cubic spline.

2.3. Principal Component Analysis

[13] The statistical characteristics of the surface skin temperature diurnal cycle are described so as to account for the variability of the diurnal cycle as completely as possible. An Independent Component Analysis approach [*Aires et al.*, 2000, 2002a] could have been used to analyze the statistical regularities of the T_s diurnal cycle regularities, but a more traditional Principal Component Analysis (PCA) is adequate for this application, as will be shown.

[14] Let $\mathcal{D} = \{y^e; e = 1, \ldots, E\}$ be a database of E diurnal cycles, y, of dimension $M = 24$. We note y to be a generic diurnal cycle; it can represent a daily observation, y^d, or a monthly-mean diurnal cycle, y^m (see Notation section). In

the following, we will denote diurnal cycles, i.e., a vector, with lower case letter, y, and the "mean-level" of the diurnal cycle, i.e., a scalar, with capital letter, Y (see Notation section). Let $\Sigma = < (y - \langle y \rangle) \cdot (y - \langle y \rangle)^t >$ be the $M \times M$ covariance matrix of $\mathcal{D}$, where $\langle \cdot \rangle$ represents the mathematical expectation. Let V be the $M \times M$ matrix with columns equal to the eigenvectors of Σ and let L be the diagonal $M \times M$ matrix with the M associated eigenvalues in decreasing order (by definition $\Sigma \cdot V = V \cdot L$).

[15] We define the $M \times M$ filter matrix $F = L^{-1/2} \cdot V^t$. For convenience, we also denote $G = F^{-1}$. The matrix F is used to project diurnal cycles, y, onto a new orthonormal base composed of the columns of G, $\{F_{\star i};\ i = 1, \ldots, M\}$:

$$\begin{cases} h = F \cdot y = F_{1\star} \cdot y_1 + \ldots + F_{M\star} \cdot y_M \\ y = F^{-1} \cdot h = G \cdot h = G_{\star 1} \cdot h_1 + \ldots + G_{\star M} \cdot h_M, \end{cases} \quad (1)$$

where ^{t} is the transpose operator. The vectors $\{F_{i\star};\ i = 1, \ldots, M\}$, i.e., rows of F, are called the filters, and the normalized eigenvectors $\{G_{i\star};\ i = 1, \ldots, M\}$ are called the PCA basis functions. These eigenvectors are an orthogonal basis set for representing the diurnal cycles, y. By definition, the elements of the new data h are uncorrelated, since

$$\langle h \cdot h^t \rangle = \langle F \cdot y \cdot y^t \cdot F^t \rangle = \langle F \cdot \sum \cdot F^t \rangle = I_{M \times M}. \quad (2)$$

[16] Practically, the first step in a PCA analysis is to compute the 24×24 covariance matrix Σ of the database of the diurnal cycles y composed of the 24 local hours. The eigenvalue matrix L and the corresponding eigenvectors V of this covariance matrix Σ are then computed using a Cholesky or a Singular Value Decomposition (SVD).

2.3.1. PCA of the T_s Diurnal Cycle

[17] The PCA is performed globally to determine generic (global) base functions that can be used for analysis or for interpolation of the diurnal cycles. The PCA is applied on the global data set, over land and ocean, for a whole year of spline-interpolated monthly mean ISCCP T_s diurnal cycles (year 1999). To avoid any confusion, we define the monthly mean diurnal cycle as the vector whose elements are the monthly mean temperatures at the local times. Only locations with complete diurnal cycles (eight daily measurements) are used for the PCA (Table 1). The latitudinal range is also limited to areas between $-60°$ and $+60°$ to avoid, as much as possible, snow-covered pixels that would require a study by itself [*Prigent et al.*, 2003b]. Previous works have performed separate statistical analyses of diurnal cycles for different latitudes, vegetation types, and seasons. Global and all-season data are used here to avoid any discontinuities. Such a global analysis makes it possible to study the link between the PCA components and the surface characteristics. The statistical sampling provided by a year of global data generates an effective time resolution that is higher than the original 3-hour ISCCP time sampling.

[18] The monthly mean diurnal cycle data, y^m, are first centered: the mean level of the monthly mean cycle $Y^m = \langle y^m \rangle$ is subtracted from each monthly mean diurnal cycles y^m. For each location and each month, the mean level, Y^m, of the monthly mean surface skin temperature diurnal cycle is stored for later use with the PCA base functions for the interpolation algorithms described in section 3. Without this preprocessing step (results not shown here), the first PCA components are determined by the regional and seasonal variations of the mean level Y^m but contaminated by other information such as the amplitude, phase, or width (i.e., daytime duration) of the diurnal cycle. This makes the physical interpolation of the PCA components most difficult. Subtracting the mean level of the diurnal cycle before performing the PCA is equivalent to constraining the first PCA base function to represent the variations of Y^m. This allows a "cleaner" interpretation of the PCA components and insures the generality of their meaning globally.

Table 1. Statistics of Missing Values for Monthly Mean DX Diurnal Cycles Under Clear Sky Conditions

Number of Missing Data in Monthly Mean Diurnal Cycle	Percentage of Pixels for Ocean (1 Month)	Percentage of Pixels for Land (1 Month)
0	60.55	58.05
1	5.12	9.48
2	15.25	20.90
3	10.45	8.60
4	4.53	2.47
5	1.97	0.38
6	1.30	0.07
7	0.72	0.03

[19] The PCA is then applied to the centered data set. We do not use any normalization (analyzing the correlation matrix instead of the covariance matrix Σ) to emphasize local times that have more variability. The first three components represent 92.83, 2.36, and 1.90% of the total variability, respectively, a total of 97%. In order to obtain a satisfactory level of precision, the second and third PCA component that represent together more than 4% of the variability cannot be neglected. Furthermore, we have shown in Figure 4 that they have a physical meaning and are useful to modify the first PCA component shape in terms of time shifting or width adjustment. Each of the higher-order components represents less than 1% of the variability. Given that there is some noise in the ISCCP cycles interpolated by cubic spline, 97% of variability explained with the first three components is adequate for an accurate description of the cycles. Discarding the other components effectively suppresses measurement and fitting noise [*Aires et al.*, 2002b]. The resulting PCA base functions are presented in Figure 2.

[20] The RMS difference between the spline-interpolated ISCCP data and its PCA representation (called "PCA-reconstructed" in the following) is shown in Figure 3. By projecting the observed cycle onto the first three PCA components, we deliberately suppress the variability of the cycle not associated to these three components. Less than 1 K RMS difference is observed, confirming that three PCA components, i.e., three degrees of freedom (four if we count the mean-level of the diurnal cycle), are enough to describe the diurnal cycle. Note that the RMS error represents the statistics for a very large data set where each diurnal cycle sample has different time measurements.

2.3.2. Interpretation of the T_s Diurnal Cycle PCA

[21] The effect of a variation of each of the three components on the diurnal cycle is illustrated in Figure 4.

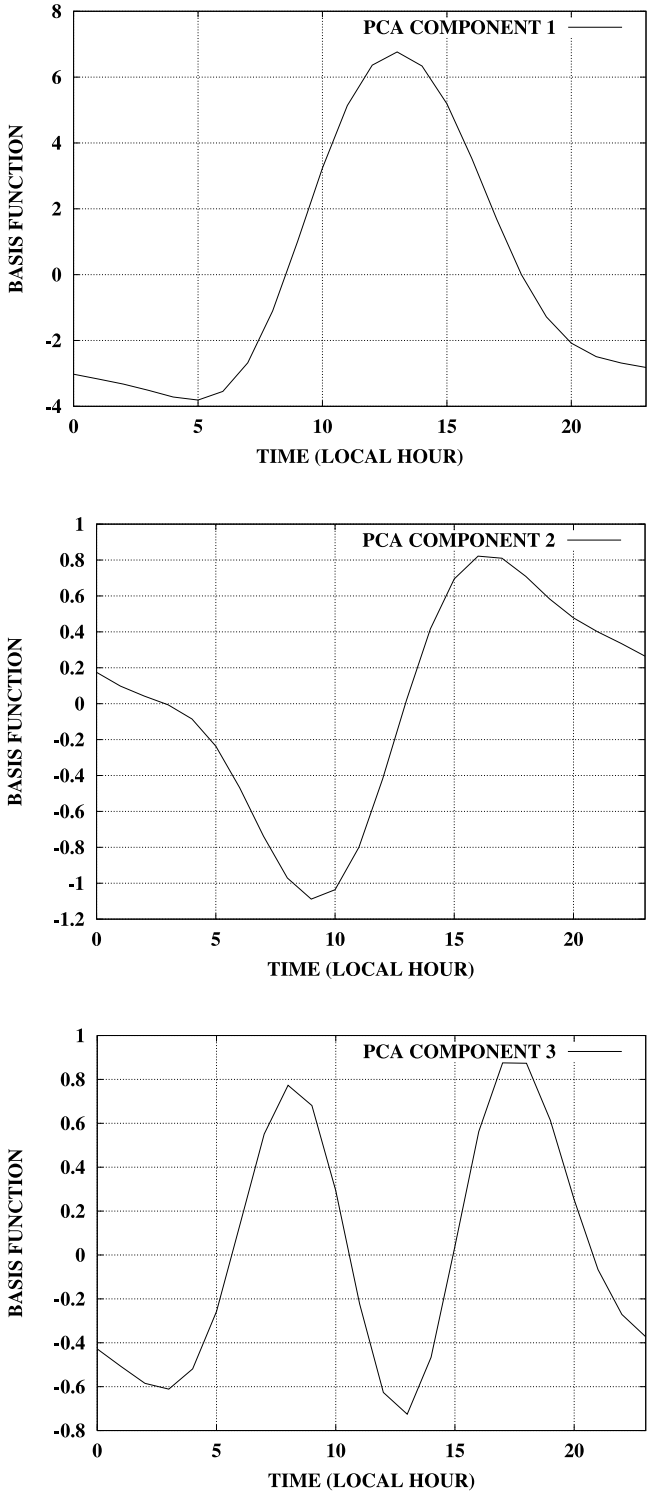

Figure 2. From top to bottom, the first three PCA base functions, g_1, g_2 and g_3, of surface skin temperature diurnal cycle.

The first component describes the amplitude of the diurnal cycle. The second represents an information on the local time of maximum temperature (phase shifting). The third component relates to the duration of daytime part of the diurnal cycle. Note that the effect of one PCA component interacts with that of the other components. For example, the phase component h_2 will not have the same effect when the amplitude component h_1 is higher or lower. Although the PCA components are calculated to be decorrelated, it does not mean that they do not interact with each other. As a consequence, the interpretation of one of the components is dependent on the value of the other components (in particular the first one).

[22] The maps of the PCA components (not shown) have realistic spatial structures coherent with our interpretation of the three first PCA components. Linear correlation coefficients larger than $\sim$0.96 are found for the map of the first PCA component and the diurnal T_s amplitude for all months in 1999. The map is well correlated to surface characteristics, essentially vegetation density. The second PCA component is less structured. The third component associated with the duration of the daytime portion of the diurnal cycle shows strong latitudinal and seasonal variations as expected.

[23] The PCA base functions are able to describe a large variety of diurnal cycle shapes. Six examples of the PCA representation of diurnal cycles are given in Figure 5, using the first three components. Although most diurnal cycles can be correctly described by the PCA projection, cycles that have more irregular shapes and that are not frequently represented in the data base cannot be accurately well reconstructed from the PCA. The PCA representation acts as a filter that keeps only the ''regular'' part of the diurnal variability in the ISCCP data.

[24] In order to interpolate a time series with irregularly and incomplete time sampling and with other sources of variability, the chosen method needs to use two sources of information: the measurements and a priori information about the shape of the diurnal cycle. Any method will use a compromise of these two sources of information. The compromise means that the PCA representation of the diurnal cycle at the measurement times can be slightly different than the actual measurement. This is the consequence of the shape that is imposed: the regularities in the curvature might force the diurnal cycle to pass through the neighborhood of actual measurements only. If the user of

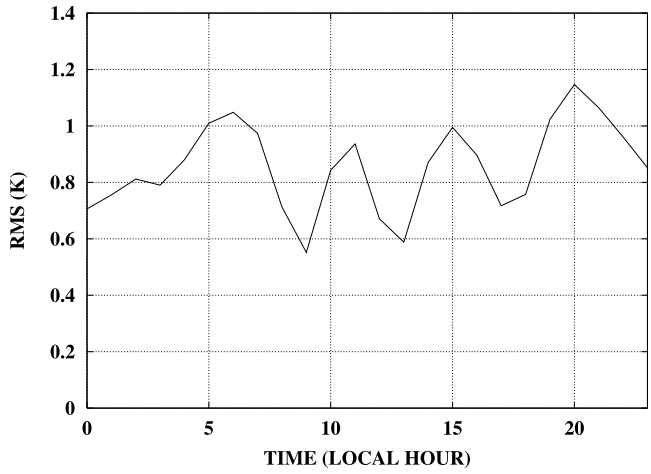

Figure 3. Root mean square differences between the 1-hourly spline interpolated diurnal cycles, y^s, and the PCA reconstructed cycles.

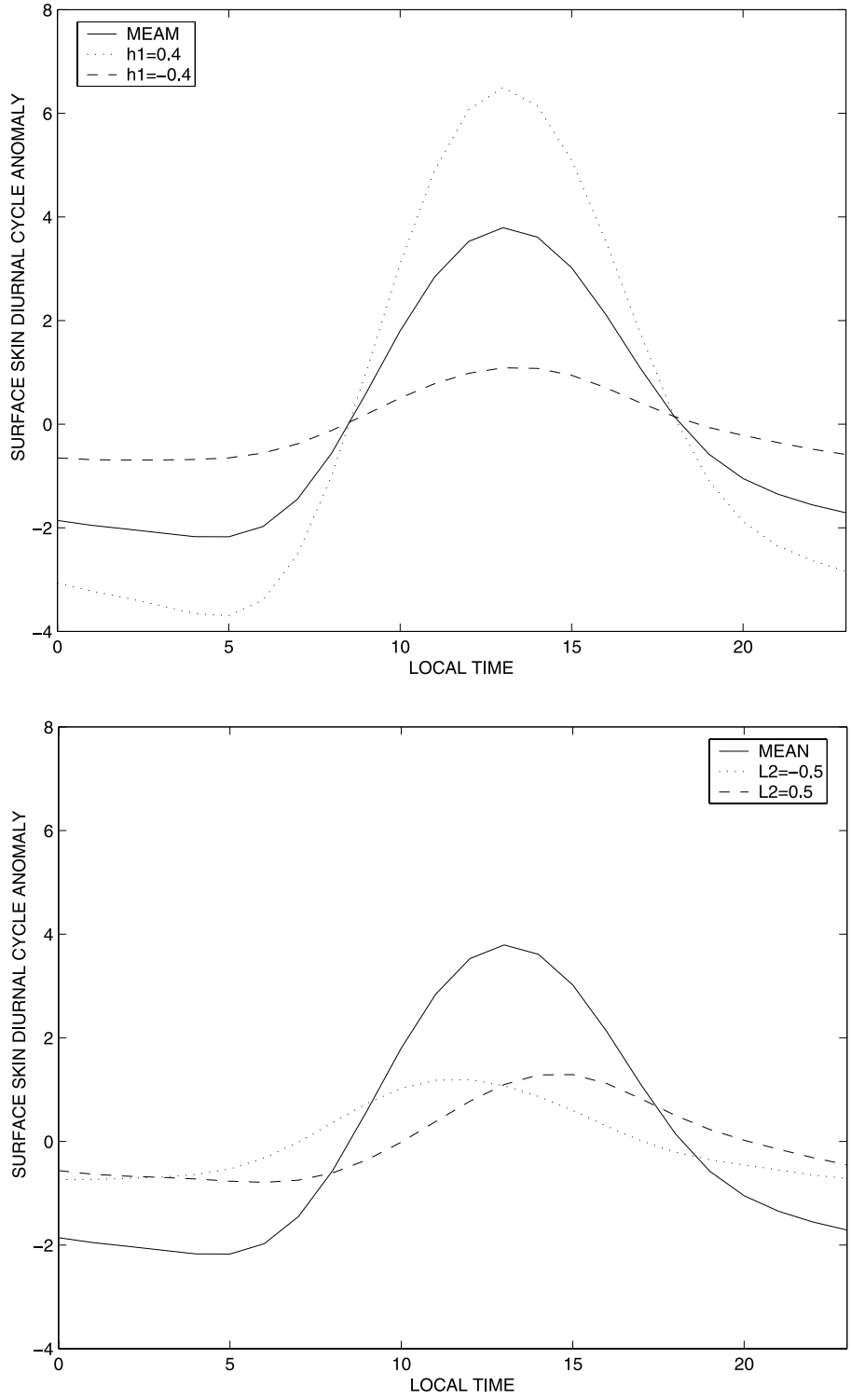

Figure 4. Effect of the first three PCA components: for component h_1 representing diurnal amplitude information, component h_2 representing lag of maximum of diurnal cycle, and component h_3 representing a diurnal cycle width information.

the interpolation algorithm does not want this behavior, he can simply use the interpolated values at the times that have no measurements and keep those measurements unmodified with the risk of generating jumps in the Ts cycles. This would give a zero level for the measurement times. Note however that this approach confuses synoptic variations (and measurement errors) with the diurnal variation, whereas our approach separates out the diurnal cycle from other variations using the shape derived from the statistics. It should also be remembered that we interpolate on remotely sensed measurements. The noise level associated with these measurements is close to 2 K

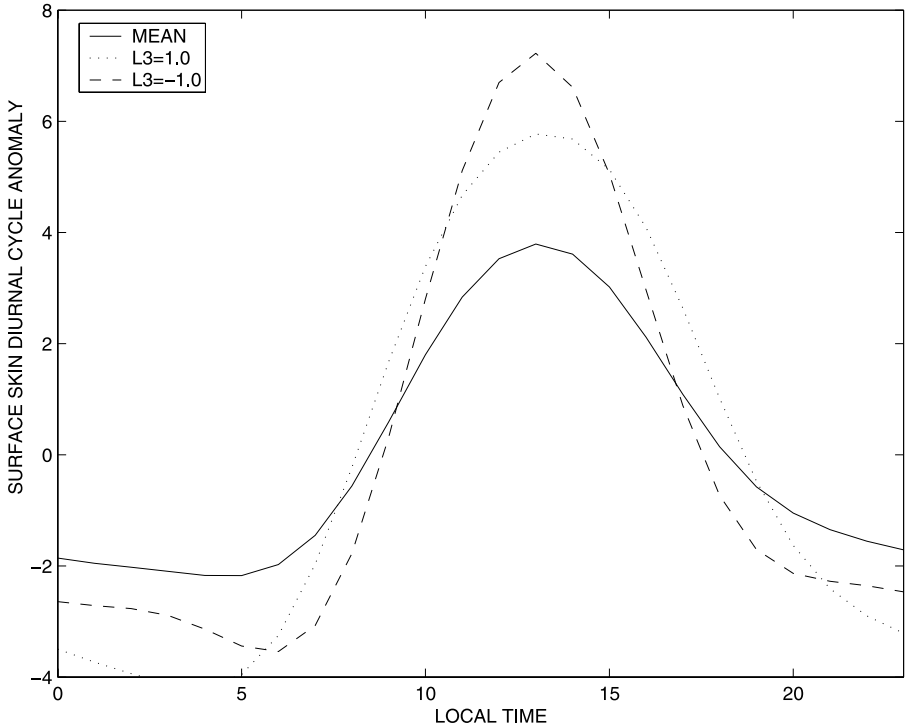

Figure 4. (continued)

RMS, if not more. That means that the actual measurements can be quite distant from the real cycle. Using a PCA representation can be considered, in this context, as a denoising technique (see *Aires et al.* [2002b] for another application of PCA to suppress noise in measurements). The PCA representation can correct the measurement and a smaller RMS difference would consequently be present. Examples of this situation are provided in Figure 5 which still exhibit quite good results.

[25] The following analysis is based on 1 year (1999) of the DX diurnal cycles projected into the first three PCA components providing diurnal cycles of T_s with 1-hour time resolution. Figure 6 shows the T_s diurnal cycles for different vegetation types in January and July 1999 between latitudes 40°N and 45°N. For each location, the mean daily T_s is subtracted. Whatever the season, the amplitude of the T_s cycle decreases with increasing vegetation density. For a given vegetation type, changes from winter to summer are primarily driven by the solar flux with the amplitude of the T_s cycle increasing with solar flux and the width of the cycle increasing with the length of the day. The averaged T_s cycles for 4 different months are presented in Figure 7 for shrublands between 40°N and 45°N: The normalized shape of the diurnal cycle clearly shows that the width of the cycle increases from winter to summer. For a given latitude range, the width of the T_s diurnal cycle is related to the length of the day but is also modulated by the vegetation/soil moisture characteristics (see Figure 6). The width of the diurnal cycle is calculated as the time in hours that separates the two times during the day when T_s is equal to the mean T_s. Figure 8 shows the latitudinal mean width of the cycle along with the calculated length of the day and the mean PCA third component for January and July 1999. The standard deviations for each latitude zone are also indicated (except for the calculated length of the day).

3. Algorithms for the Temporal Interpolation of T_s Diurnal Cycle

[26] In practice, analysis of the diurnal cycle of a specific meteorological variable is often limited by the time sampling of the available measurements. Most studies on the T_s diurnal cycle are based on observations derived from instruments on board polar orbiting satellite with a typical sampling rate of twice a day [e.g., *Lakshmi et al.*, 1998; *Jin and Dickinson*, 1999].

[27] Interpolation problems are a particular application of the inverse problem theory [*Tarantola*, 1987] from which many useful concepts can be derived, such as the concept of a priori information that we will use in the following. Various methods have been developed to interpolate the T_s diurnal cycles from a limited number of T_s estimates a day. For example, in the work of *Jin and Dickinson* [1999], a linear regression on the minimum and maximum T_s is used to estimate the temperature at other times. This assumes that the minimum and maximum temperatures are known. The following sections will investigate more general algorithms able to work in more diverse conditions.

3.1. Estimation of the Mean Level of the Diurnal Cycle

[28] In the following, $\bar{y}$ is the incomplete temperature diurnal cycle observation. For example, the ISCCP diurnal cycle averaged each 3 hours on a monthly basis still has missing data for some locations (Table 1) that need to be interpolated. $\bar{y}$ can also be the observations of a microwave instrument such as SSM/I.

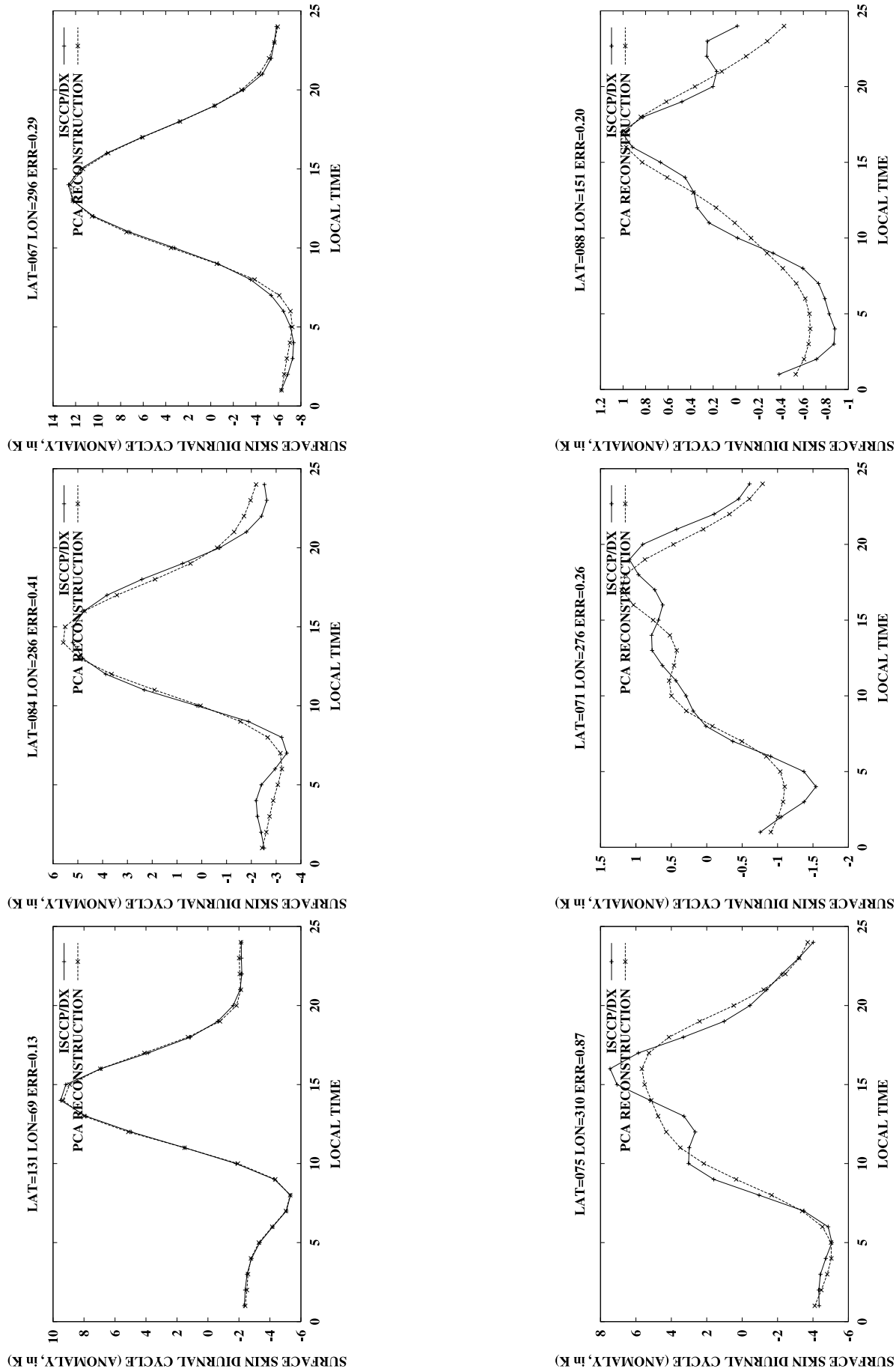

Figure 5. Six examples of PCA reconstruction using the first three PCA base functions.

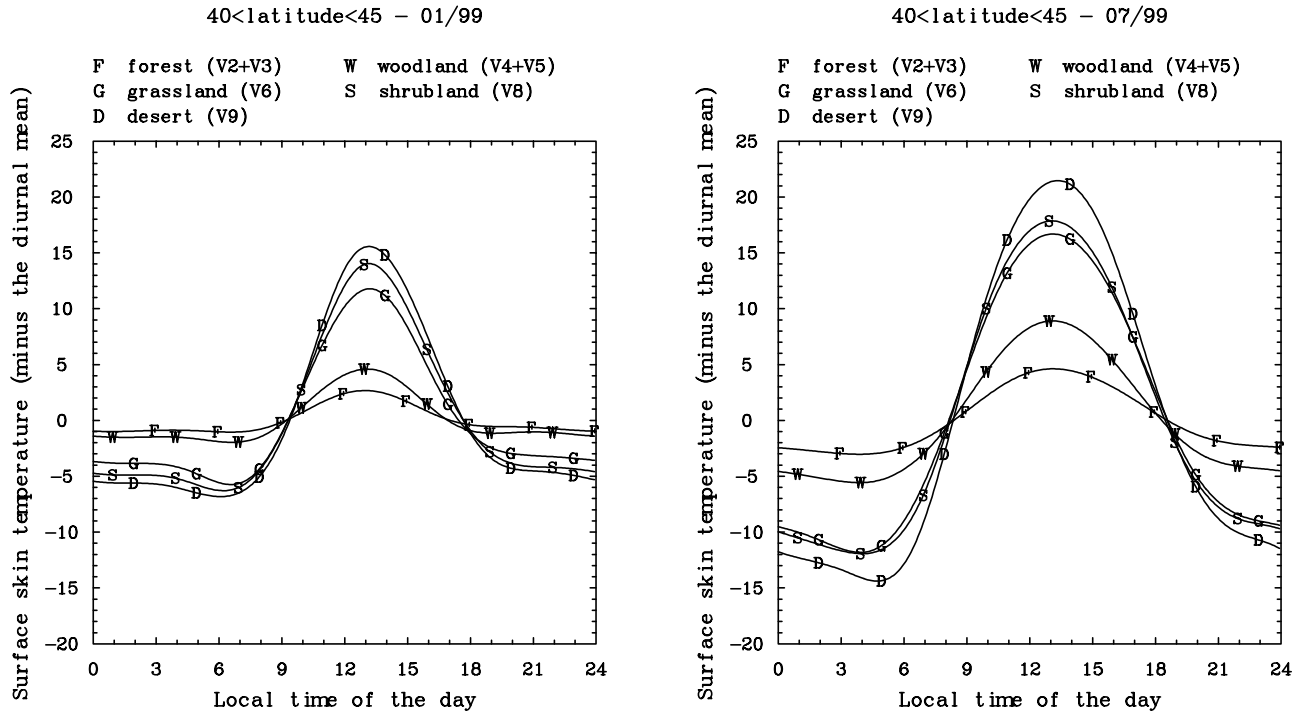

Figure 6. Anomaly of the surface skin temperature diurnal cycle for five surface types for January 1999 (left) and July 1999 (right).

[29] The mean level has been subtracted from the diurnal cycle, y, before using the PCA (section 2.3.1), and in order to use the PCA decomposition for the interpolation of the diurnal cycle, estimation of this mean level is the first step. Since $\bar{y}$ is incomplete, estimating Y^d by $\bar{Y} = \langle \bar{y} \rangle$ is not optimal. A simple solution would be to estimate Y^d by using Y^m, the mean level of the monthly mean diurnal cycle at the same location, but global statistics show that the RMS of $Y^d - Y^m$ is 1.92 K, i.e., far from negligible (Table 2).

[30] A bias, b, between the monthly mean cycle, y^m, and the incomplete daily cycle, $\bar{y}$, can be estimated using only the local times available in $\bar{y}$ (1, 2, or more measurements). The mean level Y^d of the daily diurnal cycle is then estimated by $\tilde{Y}^d = Y^m + b$. The quality of this estimation has been tested as a function of the number of missing values in $\bar{y}$. The second column of Table 2 reports the RMS error of $Y^d - \tilde{Y}^d$ for increasing number of missing values (first column). For no missing values, Y^d is perfectly estimated, being just the average of the eight 3-hourly DX temperatures. The RMS estimation error increases from 0.26 K, for one missing value, to 1.92 K, when all eight values are missing (i.e., when $\tilde{Y}^d$ is just equal to Y^m). The third column of Table 2 reports the RMS of $b = \tilde{Y}^d - Y^m$, the RMS difference between the estimate $\tilde{Y}^d$ and Y^m. The RMS difference between the estimate $\tilde{Y}^d$ and the a priori information Y^m increases from 1.92 K, when there is no missing values, to 2.66 K, when seven values are missing. This corresponds to the RMS from the estimation error $\tilde{Y}^d - Y^d$ and from the difference between Yd and Y^m. When eight values are missing, Y^d is estimated by Y^m so the RMS is zero.

[31] These statistics show that it is possible to estimate the mean level of the daily diurnal cycle with a satisfactory level of quality when values are missing using this "bias estimate under incomplete observation" method. However, the quality deteriorates with an increasing number of missing values as expected. This approach can be used in clear or in cloudy situations.

3.2. Splines/PCA Projection Interpolation Method

[32] The first step in this method is to use cubic spline interpolation (section 2.2) in order to get a 24-hour data, $\bar{y}^s$.

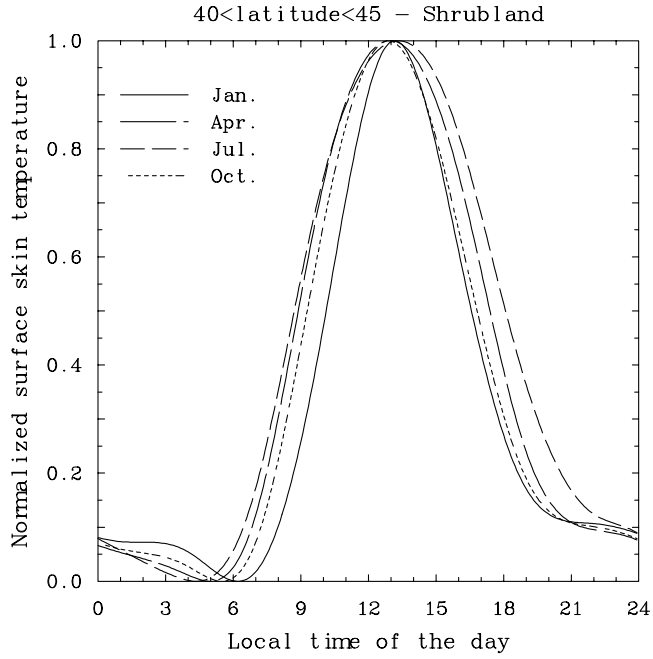

Figure 7. Normalized surface skin temperature diurnal cycle on shrubland surface type for January, April, July, and October 1999.

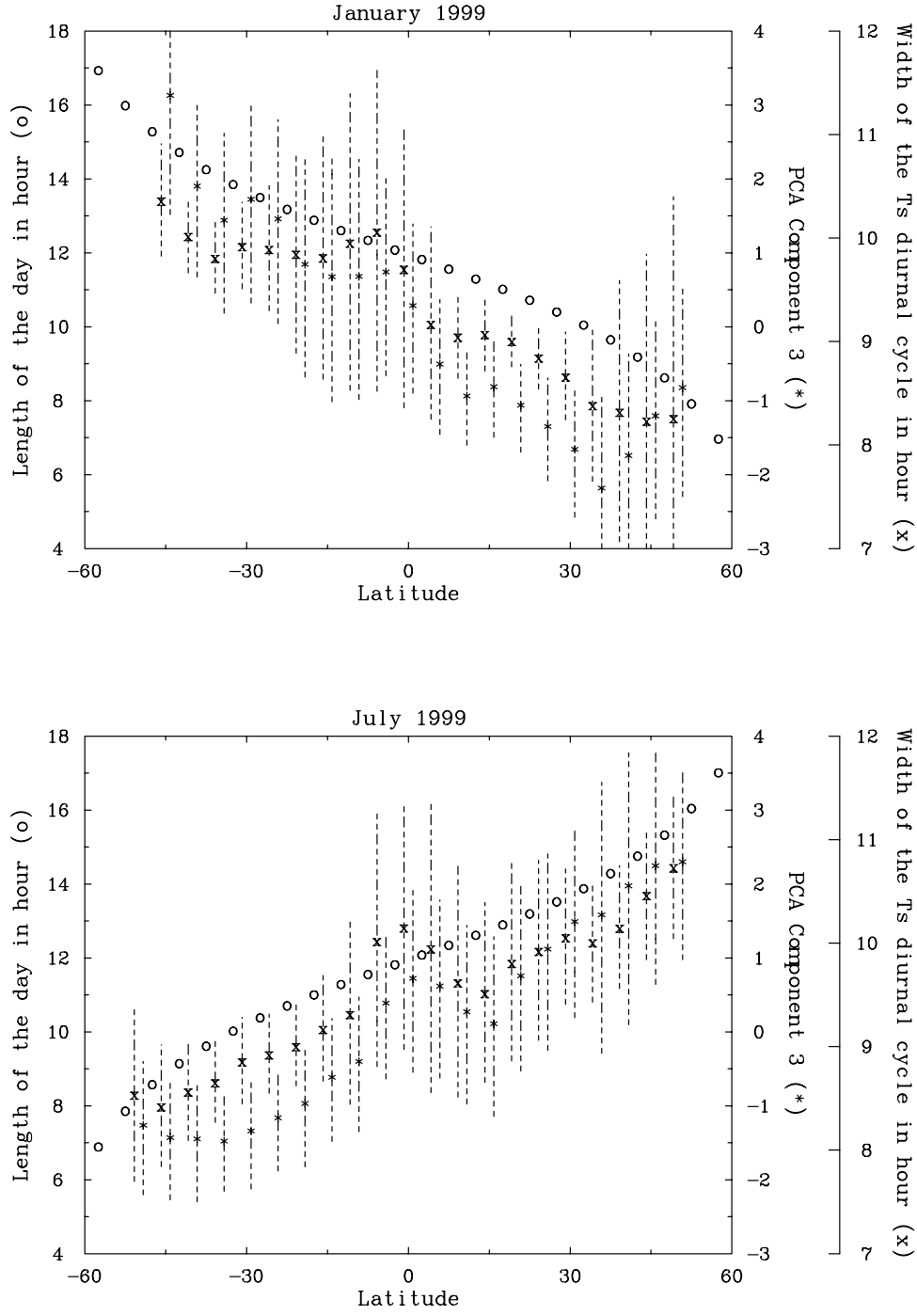

Figure 8. Length of the day, PCA component 3, and width of the surface skin temperature diurnal cycle, with respect to latitude, for January 1999 (top) and July 1999 (bottom).

If there is only a few missing estimates in the cycle and if the missing values are located at strategic times (i.e., close to the maximum or the minimum of the diurnal cycle), this procedure can provide a satisfactory reconstruction of the complete diurnal cycle $\bar{y}^s$ from the incomplete one $\bar{y}$. However, this spline interpolation cannot reduce the noise in the observations, see section 2.3.2; in fact, it can be sensitive to the noise, becoming very irregular and exhibiting outliers. After this spline interpolation to obtain a regular grid local time sampling, a second step is generally recommended. It consists in projecting $\bar{y}^s$ into the PCA base functions:

$$h = G \cdot \bar{y}^s. \tag{3}$$

in order to suppress small-scale variability like synoptic variations. The reconstruction $\tilde{y} = G \cdot h$ (see equation (1)) is the resulting regularized 24-hours interpolated diurnal cycle.

[33] It is expected that this approach will be satisfactory when only a few missing values are present. Figure 9 shows

Table 2. RMS Error for the Estimation of the Mean Daily Skin Temperature Y^d and RMS Difference Between Daily and Monthly Mean Skin Temperature, With Respect to the Number of Missing Data in the Daily Diurnal Cycle Measurements

Number of Missing Times in Daily Diurnal Cycle	RMS Error for the Estimation of Y^d	RMS Difference Between Estimate $\tilde{y}^d$ and Y^m
0	0.00	1.92
1	0.26	1.94
2	0.40	1.96
3	0.54	1.99
4	0.69	2.04
5	0.90	2.12
6	1.20	2.26
7	1.85	2.66
8	1.92	0.00

the difference between a spline interpolation using all measurements and using incomplete data. The error in the incomplete spline interpolation increases dramatically with the number of missing values, especially when these are located near the maximum or the minimum of the diurnal cycle.

3.3. Algebraic/Statistical Methods

[34] The first step in this approach is to construct a set of new PCA base functions, $\bar{G}$, interpolated at the local hours of the measurements in $\bar{y}$. That way, the incomplete diurnal cycle $\bar{y}$ does not need to be spline interpolated as in section 3.2, which can cause problems when multiple observations are missing. Cubic splines are then used to interpolate the PCA base functions G into the transformed ones $\bar{G}$. PCA components h are then calculated so that $\bar{y} = \bar{G} \cdot h$. One possibility to find such an h is to invert the matrix $\bar{G}$ with a size of $M' \times 8$, 8 being the number of DX measurements over the day, and M' the number of PCA components used (determined by the user, 2 or 3 in our case). Since the matrix $\bar{G}$ is not square and possibly ill-conditioned, its inversion might be difficult. However, many algebraic methods can be used, e.g., the pseudoinverse, the regularized pseudoinverse, or the SVD inverse [*Aires*, 1999]. If some statistical a priori information is available, such as the a priori probability distribution of the diurnal cycle, methods such as Bayesian techniques can also be used to help regularize the matrix $\bar{G}$ inversion. Once $\bar{G}^{-1}$ is determined, the PCA components are just the projection:

$$h = \bar{G}^{-1} \cdot \bar{y}, \tag{4}$$

and the reconstruction $\tilde{y} = G \cdot h$, is the resulting hourly interpolated diurnal cycle using the M' original PCA base functions.

3.4. Iterative Methods

[35] Based on the same approach as described in section 3.3 (algebraic/statistical method), it is possible to use optimization algorithms to estimate the PCA components h such that $\bar{y} = \bar{G} \cdot h$.

3.4.1. Quality Criterion

[36] The first step is to define a cost function to minimize:

$$C(h) = D_E(\bar{y}, \bar{G} \cdot h)^2 + \lambda \cdot D(h, h_{fg}), \tag{5}$$

where D_E is the Euclidean distance (other distances can be used as well, see [*Aires*, 1999]. For example, if it is expected that some of the measurements are for cloudy conditions and some are for clear-sky, a distance that puts less emphasis on outliers than the Euclidean distance is recommended.), D is a generic distance (it is important here that this distance takes into account the fact that the first PCA component h_1 is much more important than the following ones. The Mahalanobis distance is a potential candidate.), h_{fg} is a priori information for the PCA components, and λ is the parameter controlling the relative weight given to observations, $\bar{y}$, and the a priori information, h_{fg}. The term $D(h, h_{fg})$ is a stabilizer constraining the solution, h, to be close to the first-guess a priori information, h_{fg} [*Tikhonov and Arsenin*, 1977; *Vapnik*, 1998]. In order to estimate the optimal parameter λ we recommend performing some tests using using various values of λ for artificial data (so that the answer is known).

[37] The stabilizer, $D(h, h_{fg})$, can also be used as an a posteriori test for the quality of the solution, h. If $D(h, h_{fg})$ is too large, it can mean that the optimization has failed due to numerical instabilities.

[38] One of the advantages of this method is that it starts from a first-guess solution for h. This is very important when the temperature measurement $\bar{y}$ does not provide enough information to complete the entire diurnal cycle. This first-guess can be provided for example by a climatological dataset. We will see in section 4.2 how important this a priori information is for cloudy cases.

3.4.2. Optimization Algorithm

[39] Various algorithms can be used to minimize the criterion, $C(h)$, in equation (5): Metropolis algorithm, simulated annealing, stochastic gradient descent. A "deterministic gradient descent with momentum" is used here for its speed and quality of results. The steps of this algorithm are (1) estimate the gradient by finite differences: $grad = (C(h + \Delta h) - C(h))/\Delta h$, (2) use the gradient descent $h^{n+1} = h^n - \rho \cdot grad + \mu \cdot (h^n - h^{n-1})$ to modify the current estimate h^n, and (3) repeat these first two steps until $\|C(h^{n+1}) - C(h^n)\|$ is lower than a threshold, τ.

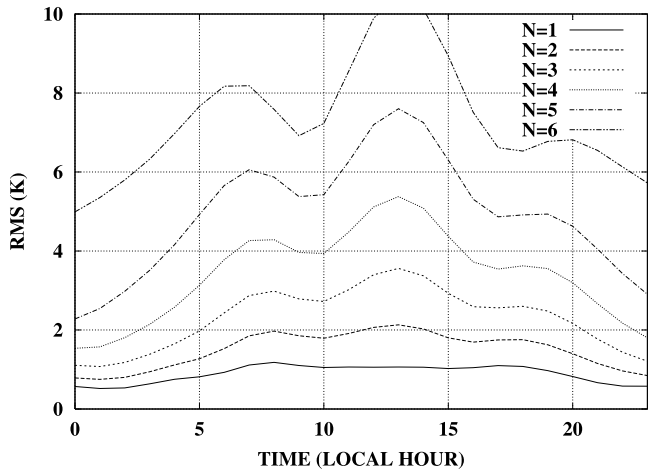

Figure 9. Root mean square errors between a complete diurnal cycle and its spline interpolation (from bottom to top, $N = 1$ to 6 missing data).

[40] First, the increments Δh have to be defined. In this study, this vector is nonzero in one element only, meaning that *grad* is calculated by applying a perturbation on only one element of h, i.e., on only one PCA component h_i. We start with the first component, h_1, then the second, h_2, etc., then repeat the cycle starting again on the first component.

[41] Second, the learning rate ρ has to be determined. If it is too small, the optimization will be too slow (possibly the optimum will not be reached) and the optimization search will have difficulties escaping from local minima. If it is too large, the optimization can oscillate from one side of the "valley" of the minima to the other side.

[42] The first element of the vector Δh corresponds to the learning step of the first PCA component that represents amplitude information and a much larger percentage of variability than the following components. Consequently, this component deserves a larger learning rate ρ_1 than the other ones associated with higher order PCA components. Here $\rho_i = \frac{l(i)}{5\sum_j l(j)}$ is selected (one–fifth of the explained variance of component i), where $l(i)$ is the diagonal elements of the eigen-values matrix L (see section 2.3). An adaptative ρ_i can also be used [*Hertz et al.*, 1992].

[43] The so-called momentum term $\mu \cdot (h^n - h^{n-1})$ in the gradient descent adds inertia during the optimization. This is similar, in practice, to increasing the size of the effective learning rate ρ when the current solution h^n is in a flat region of the cost function during the minimization. The momentum term also smoothes out potential oscillations occurring when the optimization goes from one side of the minimum "valley" of the cost function to the other one. The inclusion of a momentum generally improves significantly the performance of the optimization. The momentum rate needs to be between 0 and 1; $\mu = 0.25$ is selected here.

[44] A stochastic gradient descent algorithm can be adopted instead of the deterministic version. It is generally preferable to get to a global minimum, but in this application, the first guess is good enough so that the local minima provided by the determinist gradient is in the right region. In addition, deterministic gradient descent is faster.

[45] Here $\tau = 0.00001$ is a good compromise between a good fit and a reasonable number of iterations. This stopping criterion is used once the modifications have been done for each of the components; in this way, only one τ needs to be used. Otherwise, each component would require a different τ. The number of iterations is constrained to be more than 1000 but less than 5000. During the gradient descent, the best solution encountered is stored: sometimes, it can differ from the last one.

4. Application to the Interpolation of the Temperature Surface Skin Diurnal Cycle

[46] A database of examples is constructed to represent different experimental situations in order to test which interpolation algorithm works better, under which conditions, and for what reasons.

4.1. Clear Sky

4.1.1. Methodology

[47] Only the spline/PCA method and the iterative approaches are investigated. The algebraic/statistical approach is not robust enough because of numerical instabilities. To implement this method, it would be necessary to regularize it using statistical information such as a priori probability distribution functions, for example, in the Bayesian context [*Gelman et al.*, 1995]. In practice, the statistical information is very difficult to estimate correctly and exhaustively.

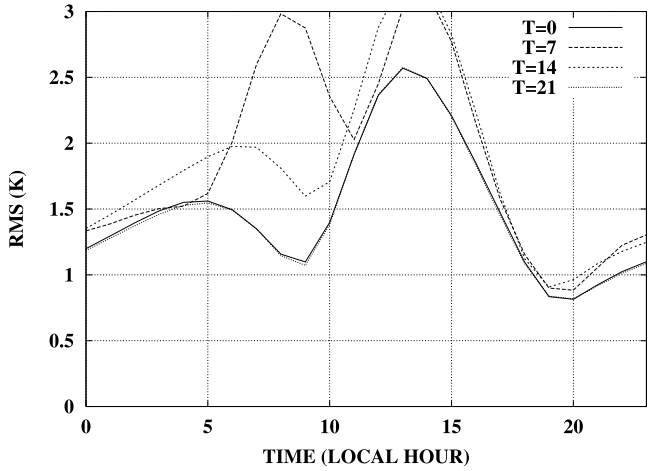

Figure 10. Root mean square errors between PCA reconstructed and spline/PCA interpolation for one-missing data at local time T = 0,7, 14 and 21.

[48] For the iterative method, the stabilizer that constrains the solution to be close to the a priori information is set to zero ($\lambda = 0$ in equation (5)) because the a priori information coming from the use of h_{fg} as a first-guess solution is already a strong enough constraint. This depends upon the experimental conditions and a stabilizer might be necessary for another application.

[49] Only the first three PCA components are used: they represent more than 97% of the variability in the ISCCP climatological data set, they describe all the variability that is of interest, and suppressing higher-order PCA components regularizes the cycles and reduces the noise.

4.1.2. Data Set

[50] A data set of complete clear-sky daily diurnal cycles y^d is extracted from the DX data set described in section 2.1. Each daily cycle is associated to its corresponding monthly mean diurnal cycle y^m. This data set represents 957,860 diurnal cycles for January 1999 over land. To test the interpolation algorithms, one or more temperatures from y^d are artificially removed and then we try to retrieve the correct diurnal cycle y^d from the incomplete one $\bar{y}^d$.

4.1.3. Results

[51] In a first experiment, the DX monthly diurnal cycles of January 1999 over land are collected and one of the eight 3-hourly temperatures is suppressed. Figures 10 and 11 represent the RMS error of the spline/PCA and the iterative interpolation methods for a missing observation at local time T = 0, 7, 14, and 21, respectively. The error is generally lower than 3 K for both methods, but the iterative algorithm has better results for all local times. The spline/PCA approach is sensitive to the time of the missing data, whereas the iterative algorithm shows more robust statistics. When T, the missing data time, is close to the maximum or minimum of the cycle, the spline interpolation can miss the

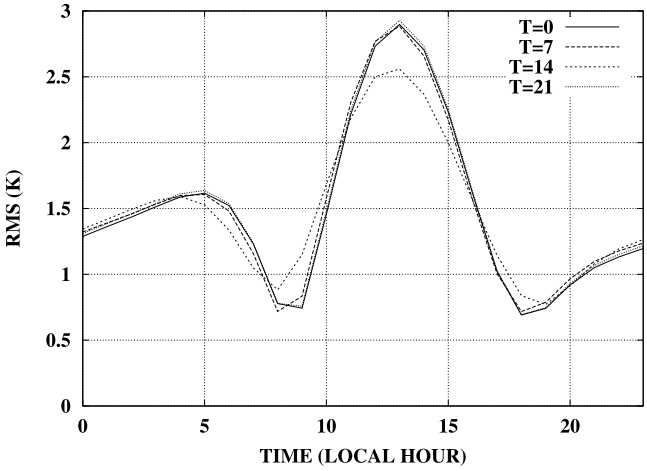

Figure 11. Root mean square errors between PCA-reconstructed and iterative interpolation for one-missing data at local time T = 0,7, 14 and 21.

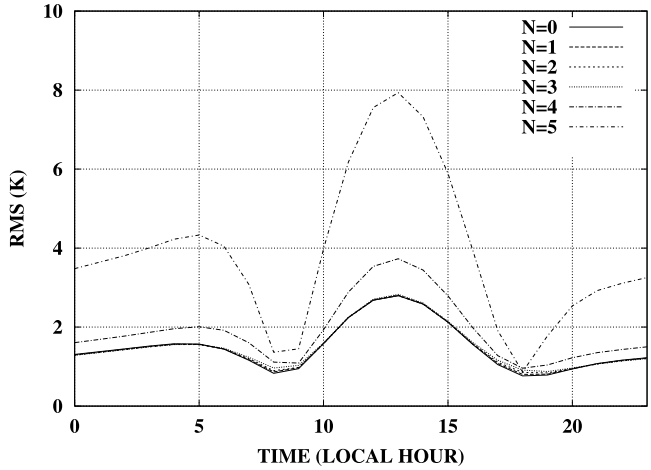

Figure 13. Root mean square errors between PCA-reconstructed and iterative interpolation for N-missing data.

amplitude of the cycle. In the iterative approach, the use of a priori information of the PCA shapes from the climatology helps the extrapolation of the extremum temperatures.

[52] The previous experiment is then repeated with multiple missing measurements instead of one. The times of the missing values are randomly selected in the eight 3-hourly measurements of DX data.

[53] Figures 12 and 13 present the RMS errors of the spline/PCA and the iterative interpolation methods for N-missing data, respectively. With the spline/PCA approach, the error increases rapidly with the number of missing data from a mean RMS error of less than 2 K for $N = 0$ to more than 6 K for $N = 4$. In contrast, the mean RMS error for the iterative approach is quite stable between 1.5 and 2.5 K for missing values $N \leq 4$: the iterative method benefits from the a priori information (the first-guess solution from monthly diurnal cycle at the same location) to compensate for the missing observations.

[54] In Figure 14, an example of interpolation when only three measurements are available is presented (i.e., $N = 5$).

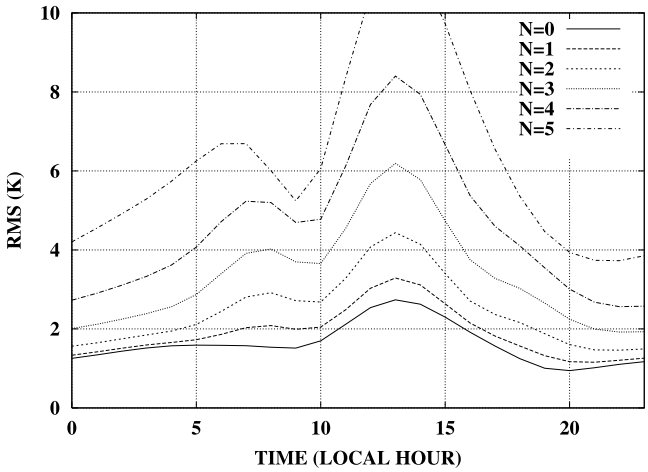

Figure 12. Root mean square errors between PCA-reconstructed and spline/PCA interpolation for N-missing data.

The times of the three measurements are close to the best case (with one almost at the maximum and the two others constraining the width of the cycle). Despite this choice, the spline/PCA method fails because the three measurements are not enough for the cubic spline algorithm to correctly interpolate the cycle (first stage of the spline/PCA approach). In contrast, the iterative approach converges to the correct diurnal cycle.

[55] An example of the curves representing the evolution of the estimates of the estimates of the first three PCA components h during the optimization of the iterative approach are shown in Figure 15. The elements of h start from the initial solution (i.e., the first-guess h_{fg}), and then evolve in order to minimize the cost function (equation (5)). The associated RMS error $R(h)$ between $\bar{y}$, the incomplete measurements of skin temperature, and $\tilde{y} = G \cdot h$, the estimation of the diurnal cycle at the local times of the measurements, is also shown. The optimization process ends when the rate of decrease of $R(h)$ becomes small.

[56] It is clear from these experiments that the most promising technique for the interpolation of surface temperature in clear sky is the iterative approach. In an operational context, some quality control tests are recommended such as the use of the stabilizer $D(h, h_{fg})$ presented in section 3.4.1 or a check based on the neighboring pixels: some instabilities can occur when the data are of bad quality (noise in observations or missing values at very bad times such as the maximum or minimum surface temperature).

4.2. Cloudy Sky

[57] The iterative method is tested here for days that are totally cloudy. Under cloudy conditions, the importance of the a priori information precludes the use of the other methods. In this study, we assume that microwave observations provide T_s estimate under cloudy sky [*Aires et al.*, 2001], with specific time sampling characteristics.

4.2.1. Creation of a Synthetic Data Set

[58] The efficiency of the interpolation technique in cloudy cases will be evaluated on synthetic data because in order to perform error statistics, the "correct" answer has to be known and a dataset with an entire diurnal cycle of surface skin temperature under cloudy conditions is not yet available.

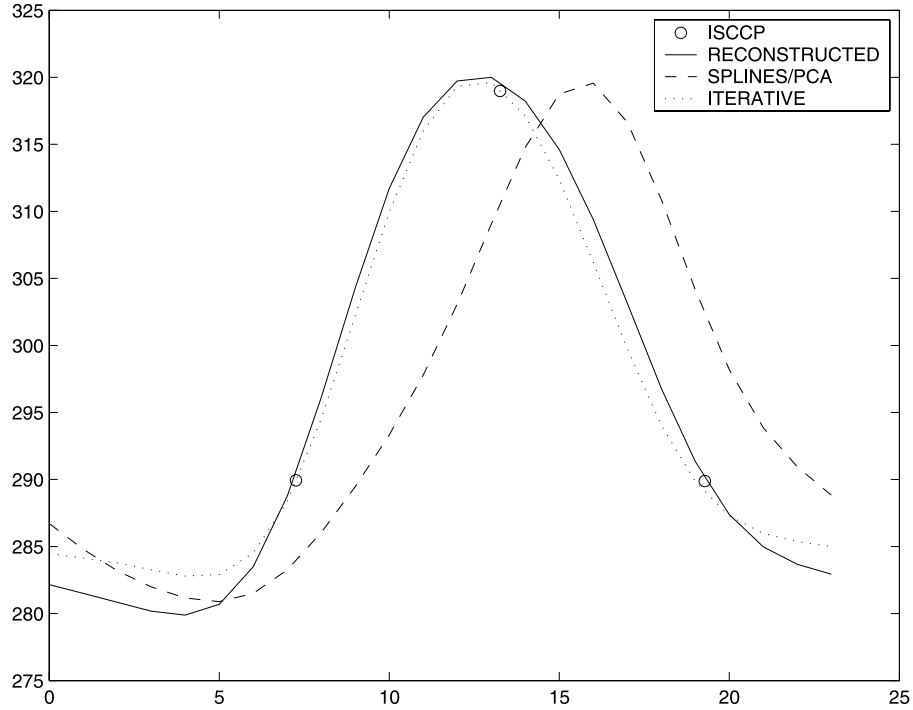

Figure 14. Example of interpolation using spline/PCA and iterative methods in the *N*-missing data experiment. In this example, only three measurements were kept from the original 8 DX reported temperatures (i.e., N = 5).

[59] To test our algorithm, we first need to generate the synthetic cloudy dataset. The shape of the diurnal cycle in cloudy conditions is chosen to be the same as for clear conditions. To generate this synthetic data set, we take the mean levels Y^d_{cloudy} and Y^d_{clear} to be identical (changing this parameter would not change the results of this experiment). The presence of clouds decreases the diurnal cycle amplitude of the surface skin temperature in clear sky by a factor, γ:

$$y^d_{cloudy}(i) - Y^d_{cloudy} = \gamma \cdot \left(y^d_{clear}(i) - Y^d_{clear}\right), \quad \text{for } i = 1, \ldots, 24. \tag{6}$$

A damping coefficient $\gamma = 0.75$ is selected to perform these synthetic tests. It corresponds to a reduction of 25% of the amplitude of the diurnal cycle in clear conditions.

4.2.2. *N*-Missing Data Experiment

[60] The clear-sky monthly mean diurnal cycle at the same location is used as a priori information for the interpolation experiment. Other authors also use clear-sky information to derive the cloudy diurnal cycle using, for example, time and/or space interpolation [*Jin*, 2000]. First, the clear-sky monthly mean diurnal cycle is used to estimate the mean level, Y^d_{cloudy}, as in section 3.1. Then, the first guess used in the iterative algorithm is the PCA projection:

$$h_{fg} = F \cdot y^m_{cloudy}, \tag{7}$$

of the centered, damped, clear-sky monthly mean diurnal cycle:

$$y^m_{cloudy}(i) = \gamma \cdot \left(y^m_{clear}(i) - Y^m_{clear}\right), \quad \text{for } i = 1, \ldots, 24. \tag{8}$$

[61] Two tests are performed to assess the impact of a good first-guess on the interpolation: one with the amplitude of the cycle decreased by a factor $\gamma = 0.75$ (supposed to be known) and the other without it ($\gamma = 1.0$) to assess the sensitivity of the iterative algorithm to this a priori information. Figures 16 and 17 represents the RMS error for the iterative interpolation methods for *N*-missing data with and without modification of the first guess for cloudy conditions. Without correction, the amplitude of the diurnal cycle is overestimated at the beginning of the iterative process, potentially causing distortions since the convergence can be trapped in local minima. When the first guess has been damped to account for the cloudy conditions, the RMS error is quite stable between 1.5 and 3 K for missing values $N \leq 4$, as for the clear-sky case, and depending on local time. These levels seem reasonable considering that

[62] 1. The missing values are chosen randomly over the eight 3-hourly sample times. It is well understood that if the temperature measurements around the minimum and maximum of the cycle are missing, the interpolation will not be of good quality. However, this is legitimate; one cannot invent information where none exists. With the maximum and minimum temperature given, the error is much smaller. The positive point is that we know the times of the measurement and can derive an estimate of the quality of the temporal interpolations. If a user of the interpolation technique prefers to be conservative and to use the interpolation only with smaller RMS scores, then he can use only the observations that have good time locations, but this will limit the number of situations that can be processed.

[63] 2. The remote sensing surface skin temperature measurements have a RMS error close to 2K for clear sky and are larger for cloudy situations. These levels of preci-

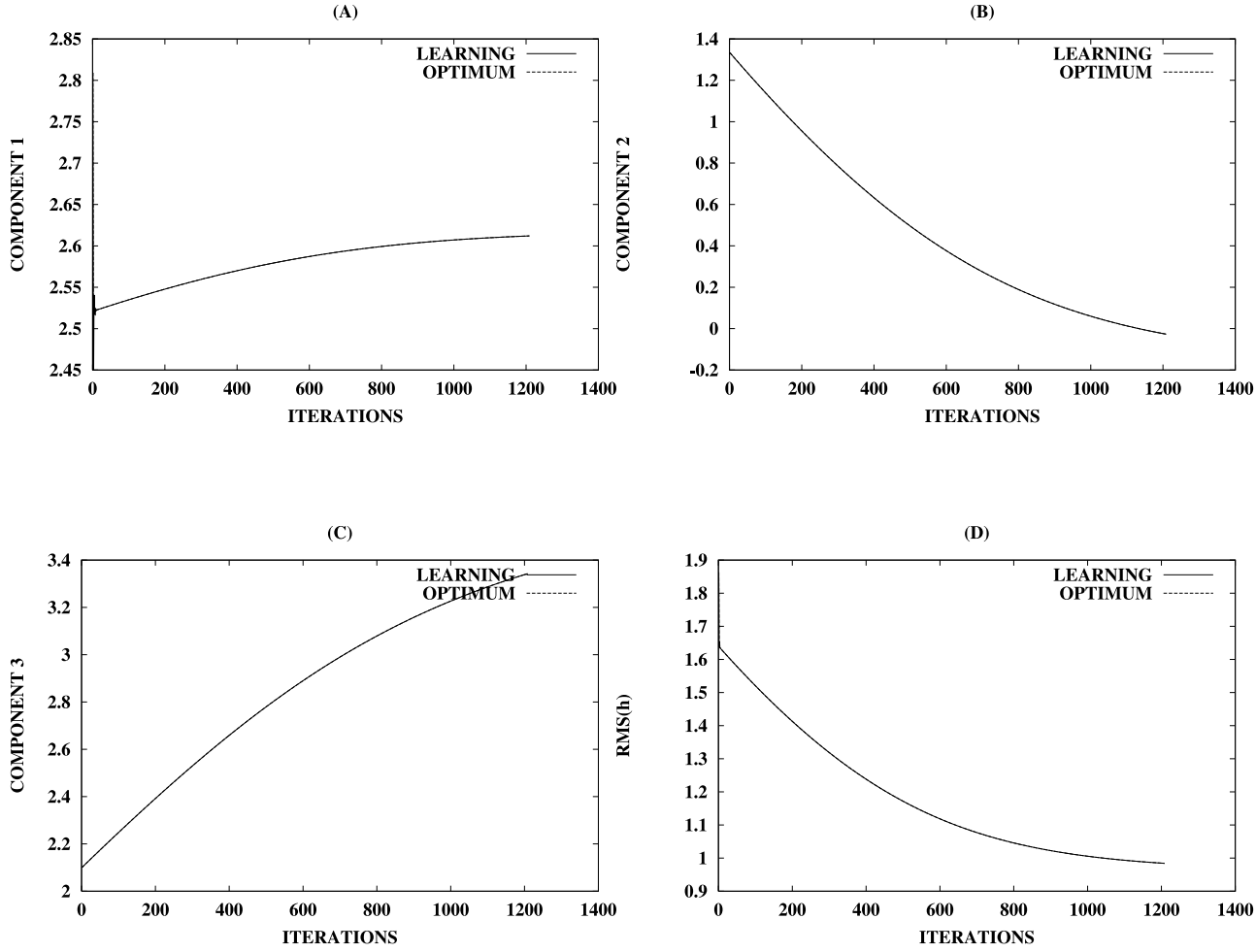

Figure 15. Learning curves for the first three PCA components (a–c) associated to the corresponding RMS error function $R(h)$ (d).

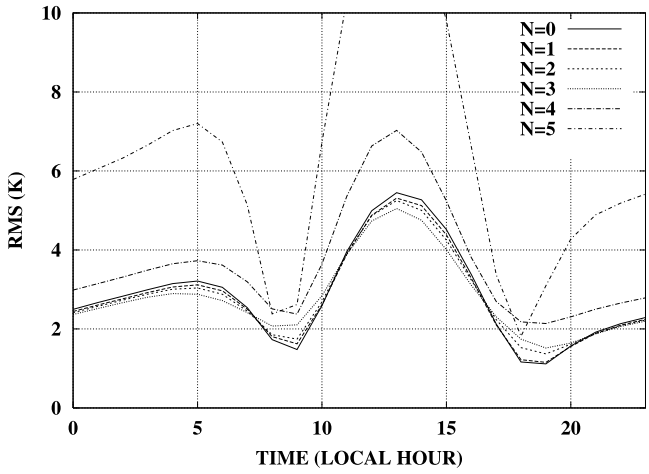

Figure 16. Root mean square errors between PCA reconstructed and Iterative interpolation in the synthetic cloudy conditions, without correction of the first guess.

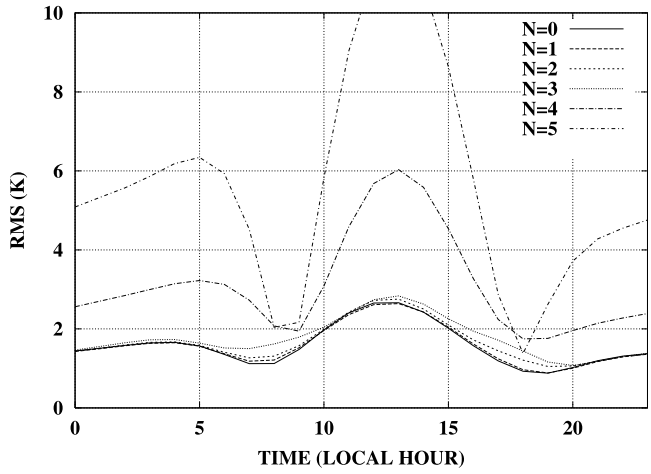

Figure 17. Root mean square errors between PCA reconstructed and iterative interpolation in the synthetic cloudy conditions, with correction of the first guess.

sion seem to be in the order of what we can obtain with our interpolation scheme.

[64] This experiment shows that a priori information for the damping parameter γ is important for the quality of the diurnal cycle reconstruction under cloudy situations. We will discuss, in section 4.2.4, some ways of obtaining this a priori information.

4.2.3. Synthetic Experiment With Current and Future Microwave Instruments

[65] At present and in the near future, passive microwave satellite measurements are only performed from polar orbits, thus limiting the number of overpasses per day at a given location. All SSM/I instruments for instance are on board polar satellites with Equator Crossing Times (ECT) close to 0630/1830 local times, even when there is more than one satellite in orbit, except for the F-10 satellite that missed its orbit and had its ECT increasing each year at a rate of 47 min (corresponding to 0913/2130 local times on 1 January 1993). The AMSU (Advanced Microwave Sounding Unit) channels are specifically selected for atmospheric sounding not for surface retrievals. Two AMSU instruments are in space, in quadrature with ECT at 0130/1330 and 0730/1930 local times. The reconstruction of the diurnal cycle of surface skin temperature with AMSU is shown for illustrative purposes: development of specific retrieval methods would be required to use AMSU observations for T_s retrieval. The Advanced Microwave Scanning Radiometer (AMSR) on board the AQUA mission since May 2002 with an ECT at 0130/1330 local times and the AMSR, launched in December 2002 on board the ADEOS-II mission with an ECT at 1030/2230 local times, could be used to retrieve surface skin temperature.

[66] Figure 18 illustrates the RMS error of the iterative approach for the reconstruction of the surface skin temperature diurnal cycle based on microwave measurements from various combinations of instruments. The RMS values only account for the time sampling errors (number of measurements and time-location). Individual T_s retrieval errors are not considered. The daily cycles have their amplitude damped using equations (7) and (8) to simulate cloudy conditions. When we use the time sampling of F11 SSM/I with an ECT at 0630 local time, the reconstruction of the diurnal cycle has quite large RMS errors, between 1 and 6 K, due to a lack of information around the time of maximum T_s. When additional SSM/I measurements are available at a different time from the F10 satellite, the errors decrease between 1 and 3 K, due to the additional constraint that the extra measurements provide. In terms of time-sampling considerations, the AMSU configuration is slightly worse than the 2 SSM/I configuration. With only AMSR on board ADEOS II, reconstruction of the diurnal cycle is only a little bit better than with one SSM/I; combination with the AMSR measurements on board AQUA reduces the error to between 0.5 and 3 K.

[67] The local times of the microwave measurements are the key for a correct reconstruction of the T_s diurnal cycle under cloudy conditions. With the new generation of instrument such as AMSR, possibly coupled to the previous generation of microwave instruments, the monitoring of T_s diurnal cycle for all-weather conditions is possible.

4.2.4. A Priori Information Under Cloudy Situations

[68] A priori information (mean-level and first-guess h_{fg}) being particularly important for our iterative algorithm, we comment on how to obtain this information when only partial cloudy observations are available.

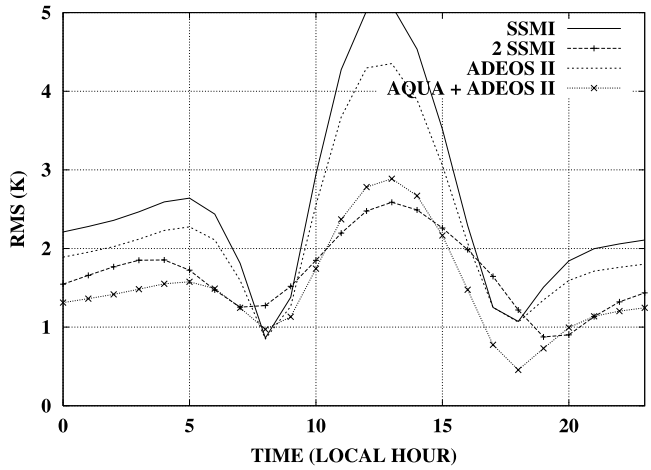

Figure 18. Root mean square errors between PCA reconstructed and iterative interpolation for SSMI only, 2 SSMI, AMSR from ADEOS II, and 2 AMSR from ADEOS II + AQUA).

[69] The mean level of the daily diurnal cycle under cloudy conditions, Y^d_{cloudy}, can be estimated using the procedure of "bias estimate under incomplete observation" method described in section 3.1. This procedure uses the a priori information from the clear-sky monthly mean diurnal cycle, y^m_{clear}; this is satisfactory if the cloud presence does not change the shape of the diurnal cycle. If this is not the case, a data set providing a cloudy-sky monthly mean diurnal cycle, y^m_{cloudy}, is necessary. The creation of such data set will be possible once a thorough microwave measurement of the surface skin temperature diurnal cycle becomes available.

[70] In order to have a priori information on the amplitude damping coefficient γ for each location, month, and surface type, a possible approach would be to use in situ air temperatures available from station measurements along with cloud observations. Such a study has been performed by Y. Zhang (personal communication, 2003), using the reports from surface weather stations. The method uses cloud fraction information to determine the diurnal cycle change relative to clear-sky conditions This first attempt to obtain a diurnal cycle of surface skin temperature for cloudy pixel is a good first candidate for a good first guess when interpolating a diurnal cycle in cloudy conditions. However, using T_{air} instead of T_s (i.e., skin temperature) introduces a bias since the amplitude of T_{air} is lower than the amplitude of T_s.

[71] Another approach would be to use, as a priori information, a climatology derived from microwave instruments. Such climatology is not possible yet since passive microwave measurements are not enough yet to provide a full description of the diurnal cycle of surface skin temperature. New instruments such as in the GPM (Global Precipitation Mission) experiment will include TMI-like microwave instruments with a 3-hourly sampling. These measurements combined with an inversion scheme such as the one described by *Aires et al.* [2001] will provide a full

climatology of mean-level and diurnal amplitude decrease for a variety of cloudy conditions.

5. Conclusion and Perspectives

[72] The diurnal cycle of the surface skin temperature has been analyzed, based on the global ISCCP 3-hourly T_s estimates for clear scenes over a year. The statistical characteristics of the diurnal cycle are extracted, using a PCA decomposition. The first three PCA components represent, respectively, the amplitude, the phase, and the width (i.e., duration of the daytime portion) of the diurnal cycle and explain 97% of the variability. Reprojection on these three PCA components filters the data, suppressing additional variability particularly measurement noise.

[73] The PCA decomposition coupled to an iterative algorithm is developed to interpolate T_s diurnal cycles in case of a limited number of measurements per day. This temporal interpolation method is general and can be applied to other problems, for time or spatial interpolations and for other variables. The algorithm is based on statistics, and no surface model is used.

[74] Our goal in this paper is to design an algorithm for the temporal interpolation of diurnal cycles without any other sources of information than the surface skin temperature measurements themselves. We do not use surface model constraints and we do not use ancillary data such as vegetation, wind, or soil moisture. There are two very important reasons for this: (1) we want to study whether surface parameters like vegetation and soil moisture could be inferred from the diurnal cycle of temperature so we cannot use such ancillary data or surface models in the analysis itself, (2) with future instruments (AQUA, TERRA) added to the current ones, and with surface temperature retrievals under cloudy conditions, we have or will soon have available more than twice-daily measurements of the diurnal cycle. Our technique aims at processing such new measurements and to keep as much flexibility as possible: the time of the measurements and their number are variable and the surface conditions are unknown. The interpolation works for monthly or daily data. It is possible to introduce first guess information if available. This approach is very useful for many reasons. For example, some practical questions can be answered, like what kind of time sampling, in terms of time-location or in terms of number of measurements, is needed in the future to obtain a satisfactory diurnal cycle of surface temperature.

[75] The effects of the number of measurements per day and their time are analyzed. The method can be used for instance to interpolate polar satellite IR T_s estimates to a full diurnal cycle. The impact of the quality of the a priori information is tested for cloudy days, a decrease of the amplitude of the diurnal cycle in the first-guess improves the results. It has been shown that reasonable T_s estimates can be retrieved from passive microwave measurements under cloudy conditions [*Aires et al.*, 2001]. Combinations of IR clear-sky T_s estimates and microwave T_s retrievals under cloud cover then determine the full T_s diurnal cycle globally, whatever the cloud cover. In clear-sky conditions, the RMS error of the temporal interpolation algorithm is quite stable between 1.5 and 2 K for $N < 4$ missing values compared to a maximum of eight samples. For cloudy situations, when the first guess is of good quality, for example by damping the amplitude of a priori initial condition taken from clear sky, the RMS error is quite stable between 1.5 and 2 K for missing values $N < 4$, as for the clear-sky case.

[76] A statistical analysis would be necessary to study the impact of the presence of clouds on the mean-level and on the amplitude of the diurnal cycle for cloudy scenes. Given the lack of in situ or microwave T_s measurements, a complete T_s diurnal cycle cannot yet be measured and directly validated. In the near future, new microwave instruments such as in GPM project will allow a complete monitoring of the surface skin temperature diurnal cycle. These data could be used to perform statistics on the mean level, shape, and diurnal amplitude decrease. That information could then be used as the a priori information needed for an operation use of our iterative algorithm. This could allow the use of previous measurements (like SMM/I observations) for the construction of a much longer data set of T_s diurnal cycle. One SSM/I instrument being insufficient to reconstruct the entire diurnal cycle but two being satisfactory, we think that our approach could provide a data set from 1993 to present time.

[77] Two main studies can follow, using a data set of surface skin temperature. First, we can analyze the potential of the diurnal cycle to force SVAT schemes and eventually to infer soil moisture. Since no model is used in the interpolation algorithm, the diurnal cycle can be also used either to evaluate SVAT schemes or to constrain surface models. SVAT schemes now include detailed parameterizations of soil and canopy processes, along with their interaction with the atmosphere. Mass and energy fluxes at the land/atmosphere interface are strongly constrained by the T_s diurnal cycle. The T_s diurnal cycle estimates can be used in the models to help validate, update, or constrain the T_s produced by the SVAT scheme at hourly scale for climate or mesoscale applications. It would also help constrain the soil moisture, given the strong relationship between the soil moisture and the amplitudes of the T_s cycles. The Global Soil Wetness Project (GSWP) initiative of the Global Land Surface Study (GLSS) of GEWEX (Global Energy and Water Experiment) for instance could benefit from this independent T_s diurnal cycle data set.

[78] The second application of this work is climate change studies. Once the full diurnal cycle of surface skin temperature is known, including accurate determination of the minimum and maximum values, it will be possible to analyze at least a decadal time record of such values and to compare them to the surface air temperature trends that are today the main source of information to diagnose global climate changes.

Notation

$\mathcal{D}$ Database of diurnal samples.
E Number of samples in $\mathcal{D}$.
M Dimension of a "complete" diurnal cycle y.
N Number of missing data in the diurnal cycle.
y Generic temperature diurnal cycle (i.e., vector with dimension M).
y^d Daily temperature diurnal cycle (i.e., vector with dimension M).

y^m Monthly mean temperature diurnal cycle (i.e., vector with dimension M).
$\bar{y}$ Incomplete temperature diurnal cycle (i.e., vector with dimension M-N).
y^s Twenty-four hour spline-interpolated diurnal cycle.
$\tilde{y}$ Final interpolated/reconstructed diurnal cycle.
Y^d Mean level of the daily temperature diurnal cycle.
Y^m Mean level of the monthly mean temperature diurnal cycle.
h PCA components.
h_{fg} The a priori first-guess PCA components.
M' Number of PCA components used for the interpolation.
Σ $M \times M$ covariance matrix.
V $M \times M$ eigen-vector matrix.
L $M \times M$ eigen-value diagonal matrix.
F PCA projection $M' \times M$ matrix.
$G = F^{-1}$ $M' \times M$ matrix of the M PCA base funtions.
λ Parameter controlling observation and a priori information in iterative method.
D_E, D Euclidean distance, generic distance.
$C(h)$ Cost function used for the optimization.
τ Threshold for the gradient descent.
ρ Learning rate for the gradient descent.
μ Momentum rate for the gradient descent.
$\langle \cdot \rangle$ Mathematical expectation.
$_t$ Transpose.

[79] **Acknowledgments.** The authors would like to thank Violeta Golea, Ting Chen, and Yuanchong Zhang for providing valuable information and for fruitful discussions. This work was partly supported by special NASA Climate and Radiation Program funding provided by Donald Anderson.

References

Aires, F. (1999), Problèmes inverses et réseaux de neurones: Application à l'interféromètre haute résolution IASI et à l'analyse de séries temporelles, Ph.D. thesis, 220 pp., Univ. Paris IX/Dauphine, Paris.

Aires, F., A. Chédin, and J.-P. Nadal (2000), Independent component analysis of multivariate time series: Application to the tropical SST variability, *J. Geophys. Res.*, *105*, 17,437.

Aires, F., C. Prigent, W. B. Rossow, and M. Rothstein (2001), A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature and emissivities over land from satellite microwave observations, *J. Geophys. Res.*, *106*, 14,887–14,907.

Aires, F., W. B. Rossow, and A. Chédin (2002a), Rotation of EOFs by the independent component analysis: Towards a solution of the mixing problem in the decomposition of geophysical time series, *J. Atmos. Sci.*, *59*, 111–123.

Aires, F., W. B. Rossow, N. Scott, and A. Chédin (2002b), Remote sensing from the IASI instrument: 1. Compression, denoising, and first-guess retrieval algorithms, *J. Geophys. Res.*, *107*(D22), 4619, doi:10.1029/2001JD000955.

Brest, C. L., and W. B. Rossow (1992), Radiometric calibration and monitoring of NOAA AVHRR ISCCP, *Int. J. Remote Sens.*, *13*, 235–273.

Cairns, B. (1995), Diurnal variations of cloud from ISCCP data, *Atmos. Res.*, *37*, 133–146.

Desormeaux, Y., W. B. Rossow, C. L. Brest, and C. G. Campbell (1993), Normalization and calibration of geostationary satellite radiances for ISCCP, *J. Atmos. Ocean. Technol.*, *10*, 304–325.

Easterling, D. R., et al. (1997), Maximum and minimum temperature trends for the globe, *Science*, *177*, 364–367.

Gelman, A. B., J. S. Carlin, H. S. Stern, D. B. Rubin (1995), *Bayesian Data Analysis*, Chapman and Hall, New York.

Hansen, J., R. Ruedy, J. Glascoe, and M. Sato (1999), GISS analysis of surface temperature change, *J. Geophys. Res.*, *104*, 20,997–21,022.

Hertz, J., A. Kroght, and R. Palmer (1992), *Introduction to the Theory of Neural Computation*, Cambridge Univ. Press, New York.

Ignatov, A., and G. Gutman (1999), Monthly mean diurnal cycles in surface temperatures over land for global studies, *J. Clim.*, *12*, 1900–1910.

Jin, M. (2000), Interpolation of surface radiative temperature measured from polar orbiting satellites to a diurnal cycle: 2. Cloudy-pixel treatment, *J. Geophys. Res.*, *105*, 4061–4076.

Jin, M., and R. E. Dickinson (1999), Interpolation of surface radiative temperature measured from polar orbiting satellites to a diurnal cycle: 1. Without clouds, *J. Geophys. Res.*, *104*, 2105–2116.

Jin, M., and R. E. Dickinson (2000), A generalized algorithm for retrieving cloudy sky skin temperature from satellite thermal infrared radiances, *J. Geophys. Res.*, *105*, 27,037–27,047.

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Igor (1999), Surface air temperature and its changes over the 150 years, *Rev. Geophys.*, *37*, 173–199.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*, 437–470.

Karl, T. R., P. D. Jones, R. W. Knight, G. Kukla, N. Plummer, V. Razuvayev, K. P. Gallo, J. Lindseay, R. Charlson, and T. C. Peterson (1993), Asymetric trends of daily maximum and minimum temperature, *Bull. Am. Meteorol. Soc.*, *74*, 1007–1023.

Lakshmi, V., J. Susskind, and B. Choudhury (1998), Determination of land surface skin temperature and surface air temperature and humidity from TOVS HIRS2/MSU data, *Adv. Space Res.*, *22*, 629–636.

Press, W. H. P., S. A. Teukolsky, W. T. Vetterling, B. P. Flannery (1992), *Numerical Recipes in Fortran*, Cambridge Univ. Press, New York.

Prigent, C., W. B. Rossow, and E. Matthews (1997), Microwave land surface emissivities estimated from SSM/I observations, *Bull. Am. Meteorol. Soc.*, *102*, 21,867–21,890.

Prigent, C., W. B. Rossow, and E. Matthews (1998), Global maps of microwave land surface emissivities: Potential for land surface characterization, *Radio Sci.*, *33*, 745–751.

Prigent, C., F. Aires, and W.B. Rossow (2003a), Land surface skin temperatures from a combined analysis of microwave and infrared satellite observations for an all-weather evaluation of the differences between air and skin temperatures, *J. Appl. Meteorol.*, *108*(D10), 4310, doi:10.1029/2002JD002301.

Prigent, C., F. Aires, and W. B. Rossow (2003b), Retrieval of surface and atmospheric geophysical variables over snow-covered land from combined microwave and infrared satellite observations, *J. Appl. Meteorol.*, *42*, 368–380.

Rossow, W. B., and L. C. Garder (1993), Validation of ISCCP cloud detections, *J. Clim.*, *6*, 2370–2393.

Rossow, W. B., and R. A. Schiffer (1999), Advances in understanding clouds from ISCCP, *Bull. Am. Meteorol. Soc.*, *80*, 2261–2287.

Rossow, W. B., A. W. Walker, D. E. Beuschel, M. D. Roiter (1996), International Satellite Cloud Climatology Project (ISCCP): Document on new cloud datasets, Natl. Aeronaut. and Space Admin. Goddard Inst. for Space Stud., New York.

Tarantola, A. (1987), *Inverse Problem Theory*, Elsevier Sci., New York.

Tikhonov, A., and V. Arsenin (1977), *Solutions of Ill-Posed Problems*, V. H. Vinsten, Washington, D.C.

Vapnik, V. (1998), *Statistical Learning Theory*, John Wiley, Hoboken, N. J.

Vinnikov, K., A. Robok, and A. Basist (2002), Diurnal and seasonal cycles of trends of surface air temperature, *J. Geophys. Res.*, *107*(D22), 4641, doi:10.1029/2001JD002007.

F. Aires, Laboratoire de Météorologie Dynamique, École Polytechnique, 91128, Palaiseau Cedex, France. (filipe.aires@lmd.polytechnique.fr)

C. Prigent, Laboratoire de l'Etude du Rayonnement et de la Matière en Astrophysique, CNRS, Observatoire de Paris, 61, avenue de l'Observatoire, 75014, Paris, France. (catherine.prigent@obspm.fr)

W. B. Rossow, NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA. (wrossow@giss.nasa.gov)