

## Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: a statistical approach

E. N. FLORIO\*<sup>†</sup>, S. R. LELE<sup>‡</sup>, Y. CHI CHANG<sup>§</sup>, R. STERNER<sup>†</sup>  
and G. E. GLASS<sup>¶</sup>

<sup>†</sup>The Johns Hopkins University, Applied Physics Laboratory,  
Johns Hopkins Road, Laurel, MD 20723-6099, USA

<sup>‡</sup>University of Alberta, Department of Mathematical Sciences

<sup>§</sup>IBM Watson Research Center

<sup>¶</sup>The Johns Hopkins University, Bloomberg School of Hygiene Public Health

(Received 19 July 2001; in final form 16 September 2003)

**Abstract.** Ground station temperature data are not commonly used simultaneously with the Advanced Very High Resolution Radiometer (AVHRR) to model and predict air temperature or land surface temperature. Technology was developed to acquire near-synchronous datasets over a 1 000 000 km<sup>2</sup> region with the goal of improving the measurement of air temperature at the surface. This study compares several statistical approaches that combine a simple AVHRR split window algorithm with ground meteorological station observations in the prediction of air temperature. Three spatially dependent (kriging) models were examined, along with their non-spatial counterparts (multiple linear regressions). Cross-validation showed that the kriging models predicted temperature better (an average of 0.9°C error) than the multiple regression models (an average of 1.4°C error). The three different kriging strategies performed similarly when compared to each other. Errors from kriging models were unbiased while regression models tended to give biased predicted values. Modest improvements seen after combining the data sources suggest that, in addition to air temperature modelling, the approach may be useful in land surface temperature modelling.

### 1. Introduction

Accurate surface air temperature measurement is useful in climatological, environmental, ecological and public health applications. Many public health studies benefit from timely and accurate measurements of temperature. In vector-borne disease studies, for example, Bertrand and Wilson (1996) demonstrated the effect of ground-level temperature on the survival of disease vectors (*Ixodes scapularis* ticks), and Mount and Haile (1989) simulated the effect of temperature on the population dynamics of the dog tick. Similarly, the transmission of urban dengue fever (Focks *et al.* 1995) is temperature-dependent, affected by the rate of development of virus in *Aedes aegypti* mosquitoes. Other public health applications, such as the effect of extreme temperatures on mortality, are also better studied with

---

\*Corresponding author; e-mail: elaine.florio@jhuapl.edu

accurate temperature predictions over large areas because the effects are individually small and are best demonstrated in aggregate (Ramlow and Kuller 1990, Whitman *et al.* 1997).

Satellite data, particularly that from the thermal infrared bands of Landsat and Advanced Very High Resolution Radiometer (AVHRR), have been used extensively in land surface temperature (LST) modelling (Qin and Karnieli 1999), however few procedures exist that use satellite observations for surface air temperature estimation (Prince *et al.* 1998). Generally, satellite-based land surface temperature algorithms apply physical principles, such as atmospheric phenomena and surface properties, to model the relationship between the channel data and the true land surface temperature. Qin and Karnieli (1999) summarize the class of split-window algorithms that use AVHRR thermal channels 4 and 5. Less common, models are also developed for surface air temperature using AVHRR thermal infrared channels, such as Prihodko and Goward (1997), and Prince *et al.* (1998), who use a land surface temperature algorithm with the Normalized Difference Vegetation Index (NDVI) and assumptions about vegetation canopy temperature.

The proliferation of modelling approaches has necessitated methods to evaluate and validate the various algorithms, often with surface observations. For land surface temperature modelling, Vasquez *et al.* (1997) noted that the absence of coincident field and satellite measurements of land temperature, particularly over a wide range of atmospheric conditions, land surface temperatures, and emissivities, rendered large-scale field assessments daunting. They note that the level of precision generally accepted as 'accurate' for land surface temperature modelling from space is 1–2°C, and most models reported a performance level in this range. Prihodko and Goward (1997) used five local ground monitors to assess the performance of the LST/NDVI method for air temperature, in a study area that covered approximately 1° latitude by 1° longitude. They report mean errors of 2.92°C for the LST/NDVI method. Rather than limiting the use of surface observations to evaluating satellite algorithms, our method takes advantage of both data sources in the modelling itself. An example of combining satellite and surface measurements is Prata and Platt (1991), who demonstrated the collection of simultaneous AVHRR and ground observations and their immediate use in a regression model for LST.

The goal of the current study is to improve performance in surface air temperature modelling by integrating satellite and surface information in the model, incorporating covariates, and assuming patterns of spatial dependence. Hypothetically speaking, in the absence of satellite measurements, one can use information such as ground station measurements and covariates such as land cover, elevation and others to model land surface temperature via regression or kriging models. Given this, our study addresses whether combining such models and information with satellite data improves the performance in surface air temperature prediction. The methods presented in this study take advantage of (1) technology that allows the acquisition of near-simultaneous AVHRR and ground observations over a large area, and (2) statistical models that incorporate covariates and spatial dependence.

The study was conducted in two steps:

1. evaluate the predictive performance of the statistical approaches that use only the ground level data and no satellite data, and,
2. introduce methods that utilize the satellite data along with ground level data and evaluate their predictive performance.

Satellite data may be used in two ways. One can use the raw channel measurements as covariates in the statistical models. These, however, do not correct for the errors that can be addressed by a split window algorithm. Hence, it seems more logical to use the Celsius-scale LST measurements predicted by one of the AVHRR-based LST algorithms. For this study, we selected the earliest form of the AVHRR-based Price (1984) split-window algorithm for its relatively simple form, its previous use in comparative studies (Qin and Karnieli 1999, Vasquez *et al.* 1997), its use in other air temperature measurement approaches (Prihodko and Goward 1997), and its past application in public health scenarios (Hay *et al.* 2000). Price's algorithm is based on the principle that the difference in top-of-atmosphere brightness temperature is due to differential atmospheric absorption only. Hence, he concluded that the coefficients obtained could be useful for predicting most ground surfaces.

## 2. Description of the statistical models

A review of scatter plots between the Price algorithm and the ground observations (table 1, column 2) suggested a strong enough association that there was information to be gained about surface air temperature from the Price algorithm in a combined statistical model. We further conducted exploratory simple regressions (table 2) that indicated strong statistical association between the two, particularly in the absence of heavy cloud cover.

To combine the data sources, we consider two sets of statistical models. The first set consists of multiple linear regression models where spatial dependence in temperature is not assumed. The second set of models, popularly known as kriging models, explicitly utilizes spatial dependence.

### 2.1. Multiple linear regression models

Two multiple regression models are included in the analysis; they are the simplest extension from previous work referenced in the use of simultaneous datasets, by Prata and Platt (1991).

*Model 1a:* This multiple regression model assumes that no satellite data are available, and attempts to predict surface air temperature using only the measurements at weather stations as the response variable and latitude, longitude and elevation as the covariates. This model can be written as:

$$\text{Temp} = \beta_0 + \beta_1 \text{lat} + \beta_2 \text{long} + \beta_3 \text{elev} + e \quad (1)$$

*Model 1b:* This multiple regression model contains all of the elements of the first, but also assumes that satellite data, in the form of the Price algorithm values, are available. As a result, it adds the satellite dataset to the covariates.

$$\text{Temp} = \beta_0 + \beta_1 \text{lat} + \beta_2 \text{long} + \beta_3 \text{elev} + \beta_4 \text{Price} + e \quad (2)$$

In both regression models, the temperatures are assumed to be spatially independent of each other. The parameters in these models were estimated using the least squares method and were based on up to 82 weather station locations. Details of the data used for estimation are provided in the appendix.

### 2.2. Kriging models

The three remaining models assumed spatial continuity in surface air temperature, i.e. temperatures at locations on the ground closer to each other

were assessed to be more similar based on variogram analysis of the ground monitor data. Prihodko and Goward (1997) also report the decaying effect of temperature dependence at increasing distance. This spatial dependence was captured in a covariance function  $\Sigma$  that modelled spatial dependence at an exponential rate of decay.

*Model 2a:* The first kriging model used the same datasets as model 1a; however, the weather station dataset, as a response variable, used this spatial dependence assumption. The ground-based covariates latitude, longitude and elevation were the same as in model 1a.

$$\text{Temp} = \beta_0 + \beta_1 \text{lat} + \beta_2 \text{long} + \beta_3 \text{elev} + e, \quad e \sim N(0, \Sigma) \quad (3)$$

*Model 2b:* This model contains all of the elements of model 2a, but with the Price algorithm temperature added as a covariate.

$$\text{Temp} = \beta_0 + \beta_1 \text{lat} + \beta_2 \text{long} + \beta_3 \text{elev} + \beta_4 \text{Price} + e, \quad e \sim N(0, \Sigma) \quad (4)$$

The above two models are known in spatial statistics (Cressie 1991) as Universal Kriging models.

*Model 3:* The third kriging model differed from the preceding models in that it used an alternative approach to combine the ground and satellite-based datasets. It was a two-stage kriging model with the satellite and ground measurements treated as simultaneous responses. The satellite measurement was placed in the second stage of the hierarchy as a response to the surface air temperature, which was built as model 2a.

### 3. Description of the data and processing

#### 3.1. Image and ground data acquisition

AVHRR High Resolution Picture Transmission (HRPT) images from the National Oceanic and Atmospheric Administration (NOAA)-14 Polar Orbiter were collected at the Applied Physics Laboratory (APL) of Johns Hopkins University during six days from January – August 2000. Thermal infrared channels 4 and 5 were calibrated and output in terms of brightness temperature. For each day, 1 000 000 pixel images were extracted over a large portion of the south-central United States (figure 1).

Channel 4 and 5 brightness temperatures, and ancillary data were processed and imported into Idrisi for Windows, Version 2, image processing software (Clark Laboratories, Clark University, Worcester, MA, USA). Ground brightness temperature was then calculated using the split-window approach in the Price (1984) formulation:

$$\text{LST} = \text{BT}(\text{Ch } 4) + 3.33(\text{BT}(\text{Ch } 4) - \text{BT}(\text{Ch } 5)) \quad (5)$$

Where LST is land surface temperature, BT(Ch4) is channel 4 brightness temperature, and BT(Ch5) is channel 5 brightness temperature.

Based on the overpass time of the AVHRR data, the NOAA on-line database of hourly Local Unedited Climatological Data from the National Climate Data Center (NCDC) was queried by robot. The dry bulb temperature was selected for each point station in the study area. Ground data within one hour of satellite overpass were collected. For the purposes of this model, data within two hours of AVHRR scan were also assumed simultaneous, if closer data values were missing.

The latitude and longitude datasets are retrieved directly with the AVHRR

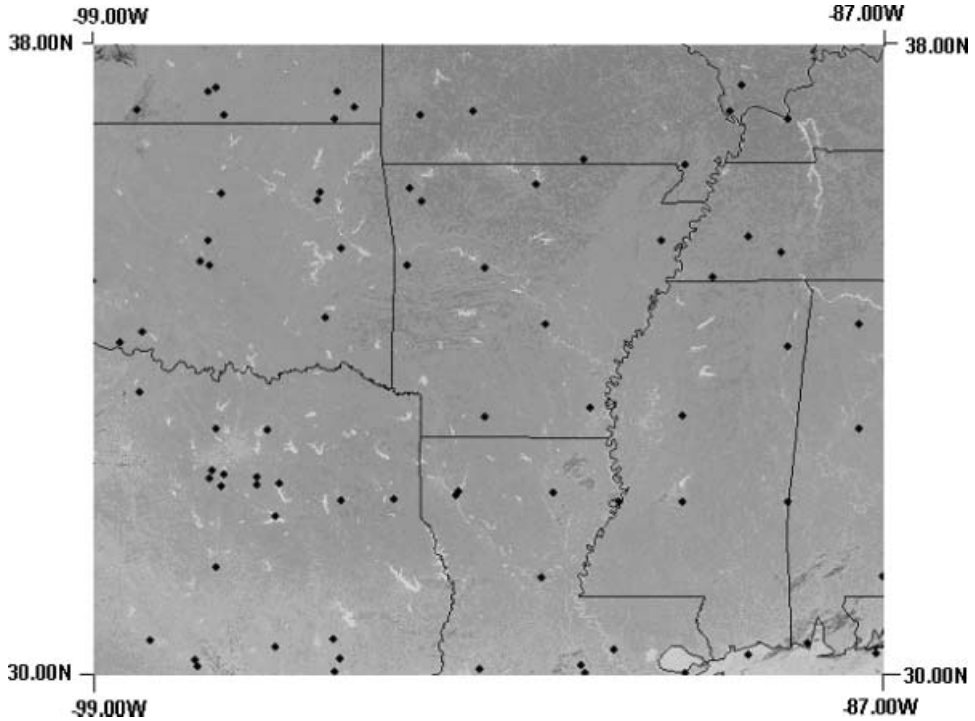


Figure 1. Study area. Typical day of Price algorithm data and ground station points.

dataset, and the elevation dataset, in metres, was based on the US Geological Survey Digital Elevation Model (USGS DEM) data, pan-american tile #w100n40.

The point locations were then passed to Idrisi image processing, which extracted covariates (latitude, longitude and elevation), land surface temperature algorithm predictions, and coordinates of the pixels that corresponded to each day's ground observations. Splus statistical software (Insightful Corporation, Seattle, WA, USA) automatically collected the results and was used to compute that particular day's maximum likelihood estimates of model parameters. Depending on the model and extent of cloud cover, between 33 and 82 locations were available for a particular day.

### 3.2. Cloud cover and data editing

Typically, LST modelling such as Price's algorithm used in this paper relies on cloud-free areas to use satellite information. However, the presence of at least some clouds over large areas presented a challenge, because complete temperature surfaces were required. Because cloud-covered pixels degrade, rather than improve, a model, we discarded them from the estimation procedure for models that relied on satellite data. Thus, pixels determined to be cloud-covered were removed from models 1b, 2b, and 3, by a crude cut-off criterion. Given the bimodal distribution in a temperature dataset for a partially clouded image, a  $k$ -means procedure ( $k=2$ ) was used. For each day's Price algorithm dataset, a random sample of 500 pixels was taken and divided into two clusters via  $k$ -means. The 5th percentile of the upper (presumed clear) cluster was selected as the cut-off point, below which any

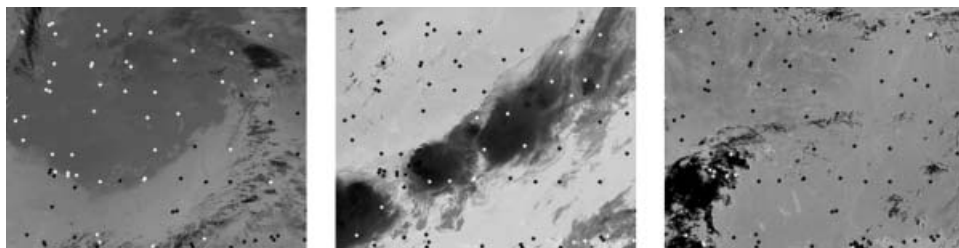


Figure 2. Output of  $k$ -means cloud editing. The geographic coverage is the same as in figure 1. White points represent locations discarded from AVHRR-based models, black points are those retained.

pixels and their corresponding ground readings and covariates were discarded from the statistical models that used AVHRR (figure 2).

Three of the daily AVHRR images (20 January, 30 May and 24 July) were predominantly cloud free, while the remaining dates had moderate (9 August) or extensive (3 March, 15 June) cloud cover (figure 2; table 1). Consequently, 12, 49 and 32 ground stations were excluded from models 1b, 2b and 3 during the respective cloudy days (table 2). Additionally, one (30 May) and five ground stations (24 July) were excluded from two of the cloud free days because of the absence of synchronous data from selected locations (table 2).

### 3.3. Data synopsis


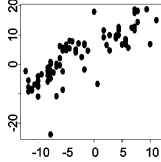
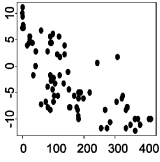
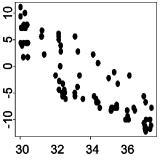

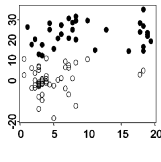
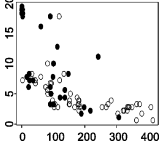
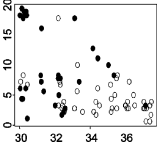
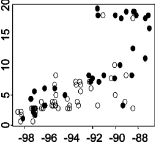

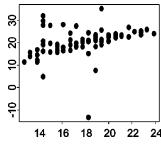
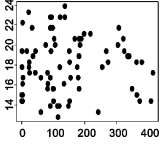
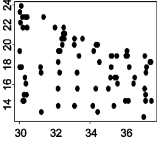
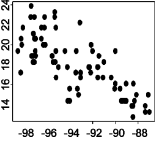
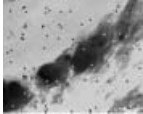
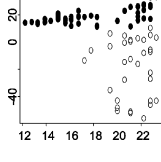
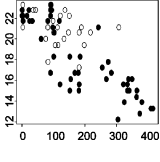
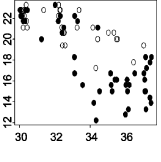
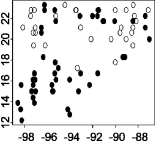

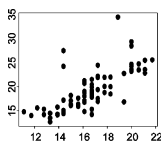
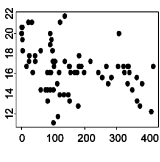
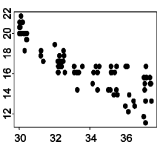
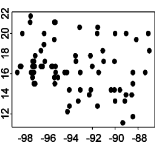
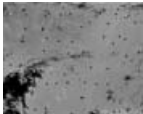
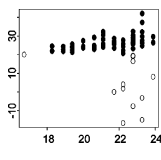
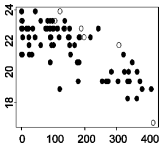
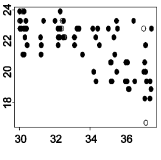
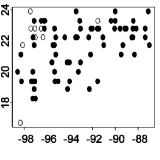
Thumbnail images of the AVHRR datasets summarized the data ranges and were used for some basic exploratory plots of the covariates (table 1). Generally, there was good correspondence between the Price algorithm alone and the temperatures measured at ground stations (tables 1 and 2). Cloud-covered stations were notable by the marked underestimation of surface temperature associated with the algorithm. Bivariate associations between measured temperatures and covariates appeared strongest during the winter and weakened during the later portion of the spring/early summer (table 1).

## 4. Cross-validation study and results

Cross-validations exploited the availability of ground truth as a means to evaluate model performance. For each of the five models, the model parameters were estimated with one 'known' (ground station-measured) point removed from the dataset, and then the estimated parameters were used to predict the value of the holdout point. The prediction error for each model was obtained by comparing the predicted value with the ground station observation that had been removed. Results are first shown for each model type and for each day, and then for model types aggregated over all dates in the study. The total number of data points contributing to the prediction error averages are also included in each table cell.

The unaggregated models showed substantial day-to-day variation in their performance, measured as average squared prediction error (table 3). During two of the six days the regression models were not sufficiently accurate, having average errors at or above  $2^{\circ}\text{C}$ . In contrast, the average daily errors of all the kriging models indicated that they obtained generally accurate (less than  $1.5^{\circ}\text{C}$ ) predictions during almost all days studied. Errors tended to be highest during cloudy days, especially using standard regression models (1a and 1b). The distinctions among the


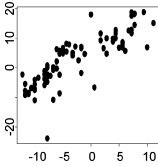
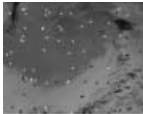
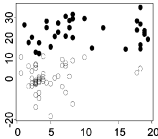

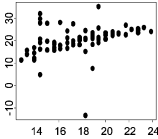
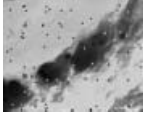
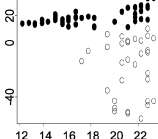

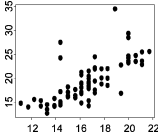

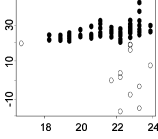
Table 1. Data summaries for six image dates. Solid circles represent clear data points, hollow circles represent data points removed by cloud editor.

Date	$x = \text{monitor}$ $y = \text{satellite}$ solid = $k$ clear hollow = $k$ cloud	$x = \text{elevation}$ $y = \text{monitor}$ solid = $k$ clear hollow = $k$ cloud	$x = \text{latitude}$ $y = \text{monitor}$ solid = $k$ clear hollow = $k$ cloud	$x = \text{longitude}$ $y = \text{monitor}$ solid = $k$ clear hollow = $k$ cloud
	20 January 2000 			
3 March 2000 				
30 May 2000 				
15 June 2000 				
24 July 2000 				
9 August 2000 				

kriging models over clear and cloudy days were more subtle and are discussed in further detail below. Daily comparisons of the five models (table 3) indicated that at least one of the kriging models outperformed both regression models on five of the six days examined.

When prediction errors for the models were aggregated (table 4), the kriging models uniformly outperformed the regression models in each of three measures of prediction error: average squared error, average absolute error, and average error.

Table 2. Exploratory analysis of association between surface monitors and AVHRR Price algorithm.

Date	$x = \text{monitor}$ $y = \text{satellite}$ solid = $k$ clear hollow = $k$ cloud	Simple linear regression results, Price $\sim$ monitor	
20 January 2000 		coefficient	1.05
		$p$ -value	< 0.0001
		$r^2$	0.68
		No. points	82
3 March 2000 		coefficient	0.09
		$p$ -value	0.6088
		$r^2$	0.01
		No. points	33
30 May 2000 		coefficient	0.77
		$p$ -value	0.0015
		$r^2$	0.12
		No. points	81
15 June 2000 		coefficient	0.69
		$p$ -value	0.0001
		$r^2$	0.30
		No. points	47
24 July 2000 		coefficient	1.27
		$p$ -value	< 0.0001
		$r^2$	0.52
		No. points	77
9 August 2000 		coefficient	1.10
		$p$ -value	< 0.0001
		$r^2$	0.24
		No. points	70

The regression models 1a and 1b both tended to underestimate ground surface temperature. In addition, the distributions of errors for the regression models were substantially broader than for any of the kriging models (figure 3).

The three kriging formulas performed similarly—with an average prediction error within  $0.034^\circ\text{C}$  of each other. In addition, the prediction errors were symmetric around zero with approximately 90% of stations having errors less than  $\pm 2^\circ\text{C}$ , and 84% of stations having errors less than  $\pm 1.5^\circ\text{C}$ .



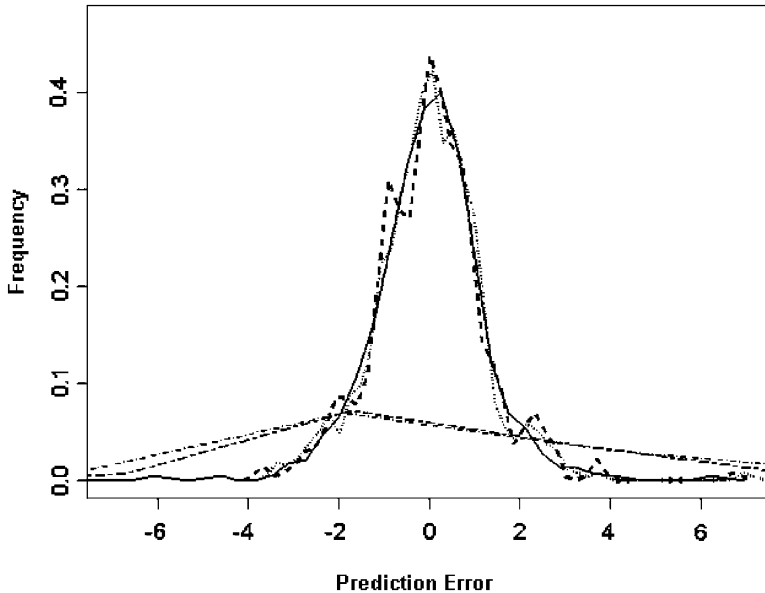


Figure 3. Distribution of prediction errors. ----- regression 1a, -.-.-.-.- regression 1b, ——— Kriging 2a, ..... Kriging 2b, - - - - - Kriging 3.

After it was observed that kriging models were superior in performance to the regression models, further comparisons were made among the kriging models only. Overall, model 2a, which did not use satellite data, predicted more accurately than model 2b, which incorporated satellite data into the mean function. This was somewhat surprising because we expected the addition of satellite information to improve predictions in any version of the model. We proposed an explanation that model 2a used all ground stations to estimate ground surface temperature regardless of the presence of clouds, whereas when the satellite data were used in model 2b, only the ground points matching unclouded pixels were used. Therefore, there was a possible advantage of increased sample size when kriging was conducted ‘under the clouds’.

To study this question, we compared the performance of models 2a and 2b (as well as model 2a and model 3) by analysing the clear day datasets separately from the clouded day datasets (figure 4). As expected, using model 2b (incorporating satellite data) showed an improvement of  $0.06^{\circ}\text{C}$  in prediction errors from model 2a on clear days (figure 4(a)). On the clouded days, however, incorporating the satellite data in model 2b reduced the sample size and was associated with a worse ( $0.03^{\circ}\text{C}$ ) performance in prediction error. Similar behaviour was seen during clear and cloudy days comparing model 2a to model 3 (figure 4(b)).

## 5. Discussion

The results provide an assessment of surface air temperature predictive performance of various statistical models, with and without satellite data, over different imaging conditions. It is clear that models assuming spatial continuity (i.e. the kriging models), regardless of whether or not satellite data are included, are

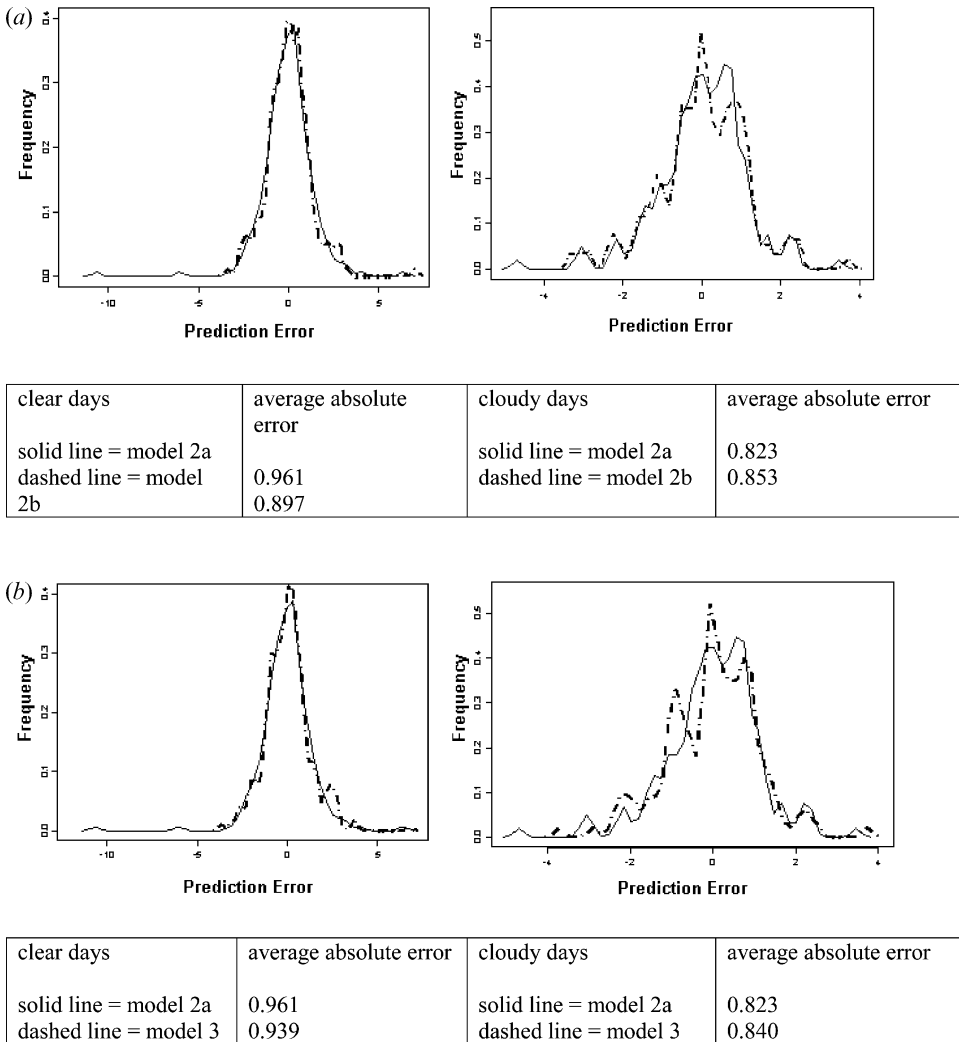


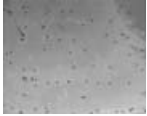
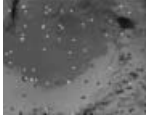
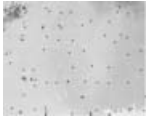
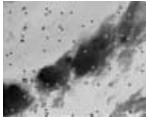
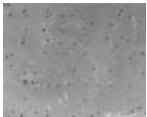

Figure 4. Comparison of clear vs cloudy days from (a) model 2a to model 2b (b) model 2a to model 3.

consistently superior to non-spatial models. Among the kriging models, it was further determined that models incorporating satellite data can be expected to perform slightly better under optimal (i.e. clear day) conditions. Somewhat unexpectedly, kriging models without satellite data performed similarly to those with satellite data under more general conditions.

The implication is that assumptions about spatial continuity are useful in surface air temperature modelling regardless of whether an algorithm uses satellite data.

Additionally, it appears that, depending on weather conditions, there may be day-to-day differences in which kriging models can be expected to be most accurate, suggesting that satellite data may not be needed on a daily basis or possibly not at all for sub-optimal conditions.

Table 3. Cross-validation results; average squared prediction errors and number of ground station observations used in each analysis.

Date	Multiple regression 1a	Kriging model 2a	Multiple regression 1b	Kriging model 2b	Kriging model 3
20 January 2000 	3.978 82 points	2.367 82 points	3.988 82 points	2.504 82 points	2.945 82 points
3 March 2000 	5.890 82 points	1.709 82 points	6.363 33 points	2.298 33 points	2.223 33 points
30 May 2000 	1.264 81 points	2.985 81 points	1.120 81 points	1.077 81 points	1.109 81 points
15 June 2000 	2.690 79 points	0.767 79 points	2.142 47 points	1.140 47 points	0.851 47 points
24 July 2000 	1.399 77 points	0.806 77 points	1.342 77 points	0.722 77 points	0.828 77 points
9 August 2000 	1.129 82 points	0.843 82 points	0.677 70 points	0.914 70 points	0.993 70 points

This interpretation is based on using a simple split-window algorithm. The use of more advanced AVHRR algorithms in concert with ground station data may produce marked improvements in performance, if the improved satellite algorithms are eliminating additional sources of error. If so, this improvement can be expected to propagate through the combined models such as kriging model 2b and model 3. For example, the Price algorithm does not include information on solar or viewing geometry. These adjustments might upgrade the performance of the combined

Table 4. Cross-validation results by model; average prediction errors and number of ground station observations involved in analysis.

	Multiple regression 1a 483 points	Kriging model 2a 483 points	Multiple regression 1b 390 points	Kriging model 2b 390 points	Kriging model 3 390 points
Average squared prediction error	2.770	1.593	2.422	1.410	1.476
Average absolute prediction error	1.199	0.867	1.556	0.898	0.901
Average prediction error	-0.035	-0.307	-0.066	0.018	0.020

kriging models. In addition, there are additional covariate datasets relevant to emissivity classification that are important to explore in future model comparisons. These data could improve prediction and help draw further distinctions regarding model performance. For example, Normalized Difference Vegetation Indices (NDVI), used in the Prihodko and Howard (1997) model, and up-to-date Land Use/Land Cover classifications, both of which are used in compiling emissivity atlases, would be relevant in models which include the satellite measurements, but may be less relevant in models which only kriged surface points (such as model 2a).

The modest improvement that we observed in the kriging model predictions that included satellite-derived land surface temperature suggest that a similar approach, i.e. of including surface air temperature measurements, might be useful in improving satellite algorithms for land surface temperature measurement. Despite the less-commonly practised use of surface air monitors, we found that they are more plentiful spatially and temporally than in-ground probes, and did show an association with land surface temperature.

In conclusion, the integration of satellite and ground-based datasets into one spatial statistical model shows promise as an approach for increasing the precision of air temperature measurement at the surface. The advantage of such modelling is that information from the more accurate direct measurements (the relatively sparse ground data points) with the more spatially plentiful but less direct dataset (i.e. the satellite algorithm's land surface temperature) both have information that can be applied to the overall predicted surface.

Such capability, particularly in real-time or repeated experiments, relies heavily on the ability to acquire both datasets efficiently. For this particular study, an automated search engine such as the IBM-developed *servelet* that recognizes the satellite overpass time and instantly queries the NCDC database, was critical.

### Acknowledgments

The authors wish to thank Rich Gasparovic at the Johns Hopkins University Applied Physics Laboratory, Ocean Remote Sensing Group, for indispensable assistance with AVHRR data acquisition. We would also like to thank Chung Sheng Li of International Business Machines for the significant contribution of automated methodologies for the simultaneous acquisition of AVHRR and NOAA weather station data. This work was supported by NCC5-305 NASA grant.

**Data sources**

1. Ground station temperature NCDC

Local Unedited Climatological Data

<http://www.ncdc.noaa.gov/doc/lcdudocumentation.txt>

2. AVHRR raw data

Johns Hopkins University Applied Physics Laboratory, via ftp

3. Grid latitude and longitude data

Johns Hopkins University Applied Physics Laboratory, via ftp

4. Elevation

USGS Digital Elevation Model

<http://edcwww.cr.usgs.gov/landdaac/gtop30/w100n40.html>

**Appendix. Kriging model formulae**

*Model 2a and Model 2b: Universal Kriging*

The assumptions for the Universal Kriging model are as follows.  $Z$  is the surface air temperature (of which the ground station observations are realizations), assumed to be a spatially continuous phenomenon, arising from a Multivariate Normal distribution with mean  $X\beta$ , and variance  $\Sigma$ .  $X$  is a matrix of covariates,  $\beta$  is a vector of regression coefficients,  $\Sigma$  is the variance-covariance matrix of true temperature, where the covariance of temperatures at any two points is an exponential function of distance between the points  $d(i,j)$ , that is

$$\Sigma = \sigma^2 * \rho^{d(i,j)}, \sigma > 0, 0 < \rho < 1 \tag{A1}$$

Thus, given  $n$  surface observations and  $(N-n)$  ground points without any surface measurements, the predicted temperature  $Z_i, i=1(N-n)$  at any unmeasured point is given by:

$$E[Z_i|Z_n, X, d(i,j)] \sim N\left(X_i\beta + \Sigma_{i,n}\Sigma_{n,n}^{-1}[Z_n - X_n\beta], \Sigma_{i,i} - \Sigma_{i,n}\Sigma_{n,n}^{-1}\Sigma_{n,i}\right) \tag{A2}$$

The distinction between the two versions of Universal Kriging is:

Model 2a: The full ground observation dataset was used and the mean vector function included elevation, latitude and longitude (figure A1).

Model 2b: A smaller, cloud-edited dataset was used (because of the incorporation of the Price algorithm information) and the mean vector function included elevation, latitude, longitude and satellite data.

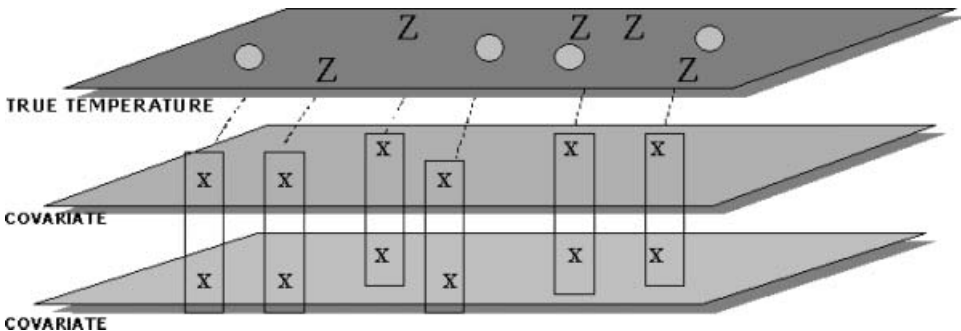


Figure A1. Illustration of kriging model (model 2a). The  $Z$ s represent known weather station measurements,  $X$ s represent covariates, and open circles represent unknown locations whose temperatures are to be predicted.

*Model 3: Two-stage Universal Kriging*

The second kriging model built upon model 2a, but placed the satellite observation as a response variable to the temperature surface rather than a covariate in the mean function as in model 2b. The statistical framework for model 3 was modelled on the method proposed by Lele and Das (2000) of using a hierarchy to address the problem of combining simultaneous measurements on a phenomenon. The hierarchy was built to mirror the satellite thermal channels responding to the surface air temperature (figure A2).

Thus, in model 3, the following were assumed. The surface observations and the AVHRR land surface temperature data values were treated as arising from a joint distribution.  $Z$  is the surface air temperature (of which the ground station observations are realizations), assumed to be a spatially continuous phenomenon, arising from a Multivariate Normal distribution with mean  $X\beta$ , and variance  $\Sigma$ , where:  $X$  is a matrix of covariates,  $\beta$  is a vector of regression coefficients.  $\Sigma$  is the variance-covariance matrix of true temperature, where the covariance of temperatures at any two points is an exponential function of distance between the points  $d(i,j)$ , that is

$$\Sigma = \sigma^2 * \rho^{d(i,j)}, \sigma > 0, 0 < \rho < 1 \tag{A3}$$

$E/Z$  are the land surface temperature algorithm readings, given the surface air temperatures, assumed independent and normally distributed, with mean  $\eta Z$  and variance  $\tau^2$ , where:  $Z$  is as above, and  $\eta$  is a vector of regression coefficients.

We may predict the surface temperature using the expected value of the

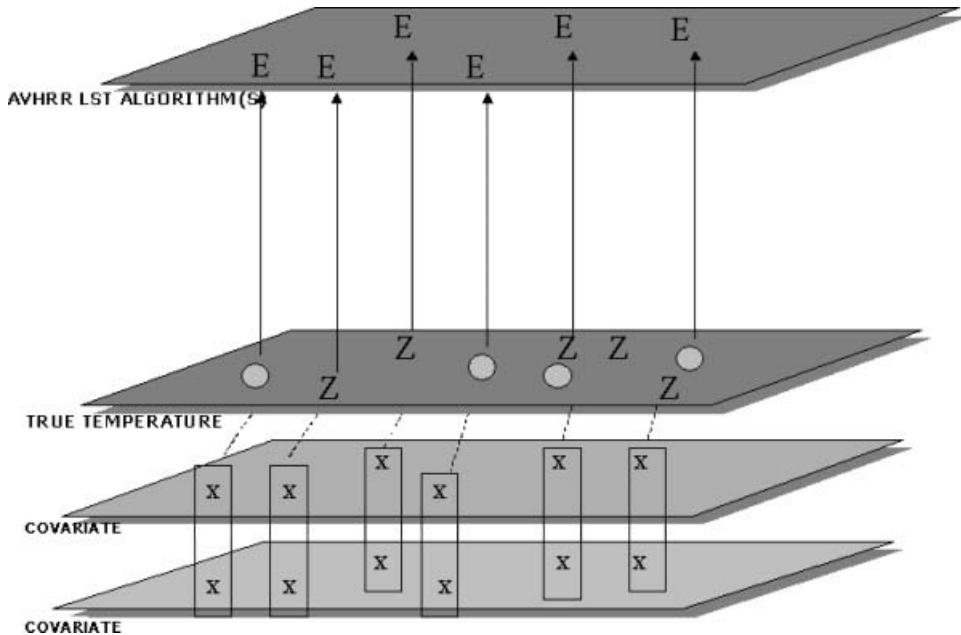


Figure A2. Illustration of two-stage kriging setup (model 3). The  $Z$ s represent known weather station measurements,  $X$ s represent covariates, and open circles represent unknown locations whose temperatures are to be predicted. The land surface temperature algorithm ( $E$ ), was assumed dependent upon the ground temperature ( $Z$ ).

marginal distribution, which is Multivariate Normal:

$$\begin{aligned}
 E[E] &= E(E(E|Z)) \\
 &= E(\eta_0 + \eta_1 Z) \\
 &= \eta_0 + \eta_1 E[Z] \\
 &= \eta_0 + \eta_1 X\beta
 \end{aligned}
 \tag{A4}$$

and variance:

$$\begin{aligned}
 \text{Var}[E] &= E(\text{Var}(E|Z)) + \text{Var}(E(E|Z)) \\
 &= E(\tau^2) + \text{Var}(\eta_0 + \eta_1 Z) \\
 &= \tau^2 + (0 + \eta_1^2 \text{Var}[Z]) \\
 &= \tau^2 + \eta_1^2 \Sigma
 \end{aligned}
 \tag{A5}$$

The joint distribution of the surface and satellite is thus:

$$(Z, E) \sim N \left( \begin{bmatrix} z_1 & 1 & & x_1 \\ \dots & \dots & & \dots \\ z_n & 1 & & x_n \\ e_1 & 1 & 1 & x_1 \\ \dots & \dots & \dots & \dots \\ e_n & 1 & 1 & x_n \end{bmatrix}, \begin{bmatrix} \Sigma & \eta_1 \Sigma \\ \eta_1 \Sigma & \tau^2 + \eta_1^2 \Sigma \end{bmatrix} \right)
 \tag{A6}$$

**References**

BERTRAND, M. R., and WILSON, M. L., 1996, Microclimate-dependent survival of unfed adult *Ixodes scapularis* (Acari: Ixodidae) in nature: life cycle and study design implications. *Journal of Medical Entomology*, **33**, 619–627.

CRESSIE, N. A., 1991, *Statistics for Spatial Data*, Wiley Series in Probability and Mathematical Statistics (New York: John Wiley & Sons).

FOCKS, D. A., DANIELS, E., HEILE, D. G., and KEESLING, J. E., 1995, A simulation-model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *American Journal of Tropical Medicine and Hygiene*, **53**, 489–506.

HAY, S. I., RANDOLPH, S. E., and ROGERS, D. J., 2000, *Remote Sensing and Geographical Information Systems in Ecdemiology*, 1st edn (Oxford: Academic Press).

LELE, S. R., and DAS, A., 2000, Elicited data and incorporation of expert opinion for statistical inference in spatial studies. *Mathematical Geology*, **32**, 465–486.

MOUNT, G. A., and HAILE, D. G., 1989, Computer simulation of population dynamics of the American dog tick (Acari: Ixodidae). *Journal of Medical Entomology*, **26**, 60–76.

PRATA, A. J., and PLATT, C. M. R., 1991, Land surface temperature measurements from the AVHRR. *Proceeding of the 5th AVHRR Data Users' Meeting' Tromsø, Norway, 24–28 June 1991* (Darmstadt: EUMETSAT), pp. 433–438.

PRICE, J. C., 1984, Land surface temperature measurements from the split window channels of the NOAA 7 Advanced Very High Resolution Radiometer. *Journal of Geophysical Research*, **89**, 7231–7237.

PRIHODKO, L., and GOWARD, S. N., 1997, Estimation of air temperatures from remotely sensed surface observations. *Remote Sensing of Environment*, **60**, 335–346.

PRINCE, S. D., GOETZ, S. J., DUBAYAY, R. O., CZAJKOWSKI, K. P., and THAWLEY, M., 1998, Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using Advanced Very High-Resolution Radiometer satellite observations: comparison with field observations. *Journal of Hydrology*, **212–213**, 230–249.

- QIN, Z., and KARNIELI, A., 1999, Progress in the remote sensing of land surface temperature and ground emissivity using NOAA-AVHRR data. *International Journal of Remote Sensing*, **20**, 2367–2393.
- RAMLOW, J. M., and KULLER, L. H., 1990, Effects of the summer heat wave of 1988 on daily mortality in Allegheny County, PA. *Public Health Reports*, **105**, 283–289.
- VASQUEZ, D. P., REYES, F. J. O., and ARBOLEDAS, L. A., 1997, A comparative study of algorithms for estimating land surface temperature from AVHRR data. *Remote Sensing of Environment*, **62**, 215–222.
- WHITMAN, S., GOOD, G., DONOGHUE, E. R., BENBOW, N., SHOU, W., and MOU, S., 1997, Mortality in Chicago attributed to the July 1995 heat wave. *American Journal of Public Health*, **87**, 1515–1518.