A comparison of data-driven and model-driven approaches to brightness temperature diurnal cycle interpolation

F. van den Bergh

Meraka Institute
CSIR, Pretoria, South Africa
fvdbergh@csir.co.za

M.A. van Wyk, B.J. van Wyk

F'SATIE
TUT, Pretoria, South Africa
vanwykma1@tut.ac.za,
vanwykb@tut.ac.za

Abstract

This paper presents two new schemes for interpolating missing samples in satellite diurnal temperature cycles (DTCs). The first scheme, referred to here as the cosine model, is an improvement of the model proposed in [2] and combines a cosine and exponential function for modelling the DTC. The second scheme uses the notion of a Reproducing Kernel Hilbert Space (RKHS) interpolator [1] for interpolating the missing samples. The application of RKHS interpolators to the DTC interpolation problem is novel. Results obtained by means of computer experiments are presented.

1. Introduction

Remote sensing data obtained from earth observing satellites is frequently affected by cloud contamination, with about 50% of the globe covered by cloud at any given moment. The presence of clouds over a given pixel, even if the cloud cover is only partial, usually renders the data associated with that pixel useless to most applications outside of cloud studies. This paper presents various techniques of interpolating values that have been lost due to cloud cover.

Geostationary satellites are in a unique type of orbit that enable them to observe the same location on the planet continually. For example, the Meteosat series of satellites produce an image of the entire earth-disc every 30 minutes. Meteosat's successor, called Meteosat Second Generation (MSG), doubles this observation frequency to produce one image every 15 minutes. At such high update frequencies it becomes possible to build detailed representations of the diurnal cycle at the various wavelengths that the satellites monitor. Figure 1 is a plot of a typical diurnal temperature cycle over a period of 24 hours.

These Diurnal Temperature Cycles (DTC) are useful because there is a strong correlation between the observed brightness temperature[1] of a pixel in the 3.9μ wavelength, and the more familiar Land Surface Temperature (LST). These DTC models are also useful in their own right, and have been used in applications such as land surface type classification [2, 3] and fire detection [4].

Given that the diurnal temperature cycle (DTC) is a continuous, smooth curve with a well-defined shape, it is possible to interpolate the missing values caused by brief periods of cloud cover (less than 4 hours) with a high degree of accuracy [5].

[1]The brightness temperature of a pixel is a measure of its observed intensity at a specified wavelength, but expressed as the temperature at which a black body would have to be to emit an equivalent intensity of radiation.

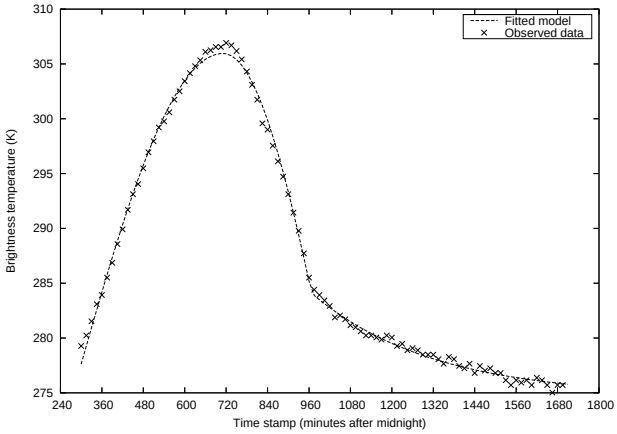

Figure 1: Sample MSG diurnal temperature data and fitted model.

This paper presents two new methods for interpolating the missing or cloud-contaminated samples in Meteosat diurnal temperature cycles. The first method is based on a pseudo-empirical analytical model, and the other is based on Reproducing Kernel Hilbert Space (RKHS) models.

Section 2 describes the cosine interpolation model, followed by an overview of Hilbert space kernels in Section 3.

2. Cosine model

An analytical approximation to the diurnal temperature cycle has been described by Schädlich *et al.* [3]. This model is reproduced in Equation (1), with Table 1 providing a definition of the symbols. Schädlich *et al.*'s model consists of a single cosine function fitted to the temperature peak of the day, followed by an exponential term describing the decay in temperature. Figure 1 illustrates the general shape of a typical diurnal temperature cycle.

$$T(t) = \begin{cases} T_0 + T_a \cos\left(\frac{\pi}{\omega}(t - t_m)\right), & t < t_s \\ T_0 + T_a \cos\left(\frac{\pi}{\omega}(t_s - t_m)\right) e^{-\frac{t-t_s}{k}}, & t \geq t_s \end{cases} \quad (1)$$

Schädlich *et al.* used the Levenberg-Marquardt algorithm (see [6] for a good introduction) to fit the parameterised cosine model to the observed temperature values [3]. Robust fitting techniques were employed to counter the effect of cloud contamination, but the specific algorithm that was used in [3] is

Table 1: Definition of symbols for Cosine diurnal model

Symbol	Meaning
T_0 (K)	residual temperature (at approx. sunrise)
T_a (K)	temperature amplitude
ω (minutes)	half-period of cosine term
t_m (solar time)	time of maximum temperature
t_s (solar time)	start of decay function
k (minutes)	attenuation constant

Table 2: Definition of symbols used in fitting functions

Symbol	Meaning
$\mathbf{a}$	model parameter vector, e.g. $(T_0, T_a, \ldots, k)$
t_i	timestamp of sample i
O_i	observed brightness temperature of sample i

not mentioned by name — given that the Levenberg-Marquardt algorithm was used to perform the fitting, it seems probable that the Iteratively Reweighted Least Squares [7] algorithm was used.

Various empirical trials with fitting Equation (1) to MSG brightness temperature data has shown that the width of the best-fitting cosine term on the rising slope of the DTC differs from the width of the best-fitting cosine term on the falling slope. This observation lead to the introduction of an additional term in the piecewise analytical model, resulting in the new model presented in Equation (2). Note that the width parameter ω has now been split into an ω_1 width parameter for the rising slope, and an ω_2 width parameter for the falling slope.

$$T(t) = \begin{cases} T_0 + T_a \cos\left(\frac{\pi}{\omega_1}(t - t_m)\right), & t < t_m \\ T_0 + T_a \cos\left(\frac{\pi}{\omega_2}(t - t_m)\right), & t_m \leq t < t_s \\ T_0 + T_a \cos\left(\frac{\pi}{\omega_2}(t_s - t_m)\right) e^{-\frac{t-t_s}{k}}, & t \geq t_s \end{cases} \tag{2}$$

In order to fit the model presented in Equation (2) to the observed data in a robust manner, an M-estimate [6] of the parameters was obtained.

Thus, instead of computing the usual sum-of-squared-errors as defined in Equation (3) (see Table 2 for a definition of the symbols), the robust error function presented in Equation (4) was used.

$$E_{ls}(\mathbf{a}) = \sum_i \left((O_i - T(t_i; \mathbf{a})\right)^2 \tag{3}$$

$$E_r(\mathbf{a}) = \sum_i \log\left(1 + \frac{1}{2}\left((O_i - T(t_i; \mathbf{a})\right)^2\right) \tag{4}$$

Press *et al.* recommend [6] that the Nelder-Mead simplex optimisation algorithm [8] should be used when fitting a model using Equation (4). In practice, it was found that the Nelder-Mead algorithm was not only more stable than gradient-based algorithms, but also faster overall on this particular problem.

3. Reproducing Kernel Hilbert Space model

The idea of a function space reproduced by a single kernel function as well as the question of whether or not there exists a kernel which will reproduce a specific function space has received attention since the beginning of the 20th century and even before. Aronszajn [1], however, was the first to formally define the notion of a Reproducing Kernel Hilbert space during the decade 1940 to 1950.

Today the applications of the theory of reproducing kernels are widely spread in mathematical statistics and engineering applications. In the 1960's (refer to [9],[10] and [11]) Parzen applied the theory of Reproducing Kernel Hilbert spaces to time series analysis. In the early 1970's Kailath ([12],[13],[14]) and his coworkers applied this theory to problems encountered in detection and estimation. More recently, the theory of reproducing kernel Hilbert spaces has found applications in generalised sampling theory, in wavelets and in graph matching (see [15],[16] and [17] as well as references therein).

In its simplest form a *Reproducing Kernel Hilbert Space* (RKHS) is a Hilbert space H equipped with an inner product $(\cdot, \cdot)$ and a *kernel* $K(\cdot, \cdot) : \mathrm{R} \times \mathrm{R} \rightarrow \mathrm{R}$ such that $K(t, \cdot) \in H$ for all $t \in \mathrm{R}$ and which has the *reproducing property*, i.e.

$$(F(\cdot), K(t, \cdot)) = F(t)$$

for all $t \in \mathrm{R}$. A consequence of the reproducing property is that $(K(s, \cdot), K(t, \cdot)) = K(s, t)$.

Suppose now we are given the input-output training data set $\mathcal{T} = \{t_i, f_i\}_{i=1}^N$ where $f_i = F(t_i) + \epsilon_i$ are noisy measurements of some unknown function $F(\cdot) : \mathrm{R} \rightarrow \mathrm{R}$. The following approximation problem is of interest: given $\mathcal{T}$ find the *minimum norm* approximation $\tilde{F}(\cdot)$ of $F(\cdot)$ in the RKHS H subject to the constraints $(\tilde{F}(\cdot), K(t_i, \cdot)) = f_i$. It can be shown that $\tilde{F}(\cdot)$ is of the form [18]

$$\tilde{F}(\cdot) = \sum_{i=1}^{N_a} a_i \, K(\tilde{t}_i, \cdot),$$

where usually $N_a \leq N$ due to the presence of noise and the *kernel centres* $\tilde{t}_i$ are inferred from $\mathcal{T}$ by means of some data reduction scheme [17]. The solution of this approximation problem is then obtained as

$$\mathrm{a} = \mathrm{G}^\dagger \, \mathrm{f}$$

where $\mathrm{a} = (a_i)$, $\mathrm{f} = (f_i)$ and $\mathrm{G} = (K(\tilde{t}_i, t_j))$. Here $\mathrm{G}^\dagger$ denotes the pseudo inverse of the matrix G.

For the application discussed here we have chosen the *Dirichlet kernel* [17] namely

$$K(s, t) = \frac{\sin\{(n + \frac{1}{2})\, u\, (s - t)\}}{\sin(\frac{1}{2}\, u\, (s - t))},$$

where u is simply a dilation parameter and n the harmonic number. The motivation for this choice is, firstly, the Shannon sampling theorem which says that an unknown bandlimited function can be reconstructed from its samples provided that the samples are sufficiently closely spaced and, secondly, the periodicity of the data.

4. Experimental results

In this section, the interpolation accuracy of the two techniques described in Sections 2 and 3 will be studied. Each method is expected to interpolate the DTC through the missing samples. The accuracy of the interpolation is measured using the Mean-Squared-Error metric, defined by

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (O_i - R_i)^2, \tag{5}$$

where O_i represents the observed brightness temperature value, and R_i represents the reconstructed/interpolated value.

4.1. Data set

The interpolation methods discussed in this paper have been developed to interpolate samples in a DTC that have been corrupted by cloud contamination. It would be infeasible to test these methods on DTC data that was really contaminated by clouds, since the expected values for the cloud-contaminated samples are unknown. Instead, cloud-free DTC data is used to simulate missing samples by deleting segments of the test data; this allows for direct comparison of the interpolated samples with the real samples, as expressed in Equation (5).

Some variation in the shape of the DTCs can be observed when comparing samples taken from different regions of South Africa, since the shape of the DTC is influenced by — amongst other things — the land cover type. The data set used in the experiments below was constructed to compensate for this effect to some degree by choosing 30 MSG pixels from six regions selected randomly over South Africa.

For each of these pixels, sequences of between four and six complete cycles (days) were selected from the complete data set spanning dates from 2004-08-01 to 2004-08-29. Each of these datasets were then subjected to a "cloud-simulation" filter that removed segments to simulate periods of cloud cover lasting 4 hours. This duration was chosen to correspond to the gap length used by Schädlich *et al.* [3], and corresponds to a segment of 16 consecutive samples. Through the application of the cloud-simulation filter, a data set of 90 complete DTC cycles was generated. This data set was thus composed of 30 unique source sequences, each of which was processed with three different cloud-simulation filters that removed different segments from each DTC cycle.

The first three complete cycles of each sequence was not affected by the cloud-simulation filter, thus providing uncorrupted data for the interpolating algorithm to train on.

4.2. Cosine model results

The results in this section were obtained by fitting Equations (1) and (2) to diurnal temperature cycle time-series data. The models were fitted individually to each cycle, thus requiring all the available samples of that cycle (excluding the 'missing' samples) in a batch. This makes the cosine model interpolation technique unsuitable for real-time applications like fire detection [4], however, it is still useful as a benchmark interpolation technique.

Table 3: Mean squared error values obtained by fitting both the original, and the newly proposed cosine model to DTC data (standard deviation in parentheses).

Model	All samples	Missing samples only
Eq. (1)	1.93221 (0.53349)	2.25000 (2.00443)
Eq. (2)	0.73041 (0.19548)	0.59428 (0.43810)

Table 3 presents the Mean-Squared-Error (MSE) of the fit produced by the two cosine models, computed over all 90 sequences in the data set. Two errors are reported: the MSE over the entire data set, and the MSE over the missing samples only. From these results it is clear that the new cosine model — represented by Equation (2) — is able to fit and interpolate the DTC data with greater accuracy. The difference in MSE between the two cosine models was confirmed to be significant at a 1% level of confidence, using the Wilcoxon signed-rank test.

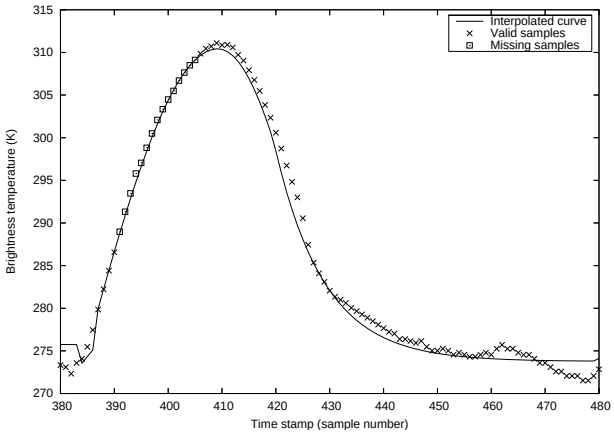

Figure 2: Sample DTC with missing values, reconstructed with the original cosine model, Equation (1). MSE = (1.4532, 0.62432).

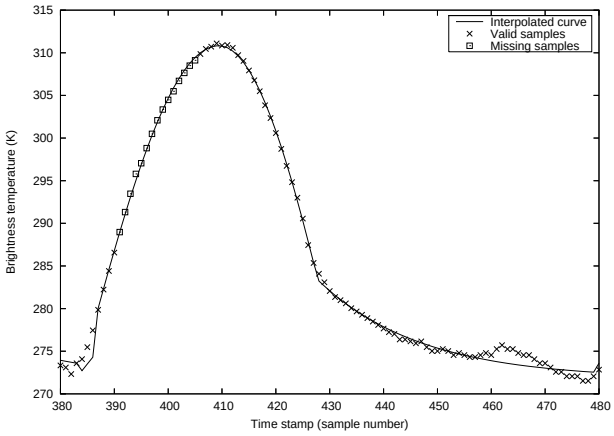

Figure 3: Sample DTC with missing values, reconstructed with the new cosine model, Equation (2). MSE = (0.4490, 0.248).

Specific examples, chosen to illustrate the nature of the cosine interpolation methods, are shown in Figures 2 and 3, representing Equation (1) and Equation (2), respectively. Note how the newly proposed model visibly fits better to the falling slope of the cosine part of the cycle. Note that the two MSE values in the caption correspond to the MSE over all samples, and the MSE over only the missing samples, respectively.

4.3. RKHS model results

Considering the test data, the kernel centres were positioned at intervals of approximately every 105 minutes yielding $N_a = 14$. The value of n was taken to be 7. These values were used for all the experiments performed.

In the first experiment the RKHS model was fitted to the first cycle in the given sequence of cycles thereby obtaining a reference model fit which may be considered as the RKHS equivalent to the cosine model's Equation (2). During the first cycle of the sequence no clouds were present, i.e. the cycle was complete without missing data. For each subsequent cycle this reference RKHS model was kept fixed and then just scaled and translated vertically to obtain the best least squares overlay. The

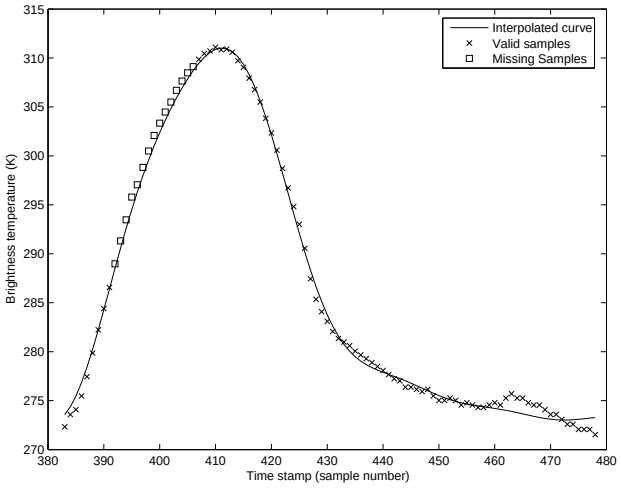

Figure 4: Sample MSG diurnal temperature data and fitted RKHS model for the first experiment. MSE = (0.5363, 0.7331).

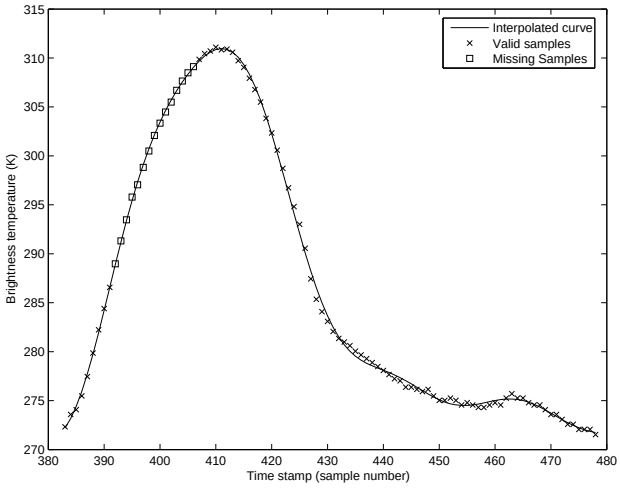

Figure 5: Sample MSG diurnal temperature data and fitted RKHS model for the second experiment. MSE = (0.1258, 0.2227).

motivation for this approach was that the amount of computation per cycle would be reduced significantly. The specific example in Figure 4 shows the RKHS model—initially fitted to cycle 1—overlaid onto cycle 5 of the same data sequence. The measured mean square errors over all samples and over the missing samples only are listed in the caption.

The same data sequence was used in the second experiment, but here the RKHS model was fitted to cycle 5 after which the mean square error was calculated over the complete cycle, as well as separately over the missing samples only. The results obtained are shown in Figure 5. The measured mean square errors over all samples and over the missing samples only are listed in the caption. Notice that the RKHS model was able to accommodate the hump on the right hand side which cannot be achieved with the cosine model. This hump appears to deviate from the expected temperature (that predicted by the pseudo-physical cosine model) by about 3K, so it is possible that it could have been caused by a fire. The difference between Figure 5 and Figure 4 can thus be used to detect anomalies such as this hotspot.

5. Conclusion

The results presented in Section 4.2 support the hypothesis that the cosine model is effective in reconstructing missing samples caused by brief periods of cloud cover. The evidence further supports the hypothesis that the new cosine model presented here is a significant improvement over the previous cosine model. Techniques for extending the cosine model to support real-time operation are currently being investigated. This would require building a 'typical' model over the most recent n cycles, and then adapting this model to fit partial cycles on-line as the data is received from the satellite, every 15 minutes.

The results presented in Section 4.3 clearly demonstrate the ability of RKHS interpolators to interpolate data by first deriving a reference fit from a complete cycle which produces the RKHS interpolator's equivalent of the cosine model's Equation (2). This reference model was fitted to remaining cycles by merely scaling and translating the reference model optimally. This approach produced good results while requiring significantly fewer computations than required by the cosine model. It is expected that multi-scale and hybrid RKHS-based approximators [15] would produce still better results and will be investigated in the near future.

Efforts to extend the analysis of the performance of these interpolation techniques to a an even larger data set are also under way.

6. References

[1] N. Aronszajn, "Theory of Reproducing Kernels," *Trans. Am. Math. Soc.*, vol. 68, pp. 337–404, 1950.

[2] F. M. Göttsche and F. S. Olesen, "Characterization of landsurfaces by modelling diurnal temperature waves extracted from Meteosat-IR data," in *Proceedings of the 9th conference on Satellite Meteorology and Oceanography*, (Boston, MA), pp. 426–427, American Meteorological Society, 1998.

[3] S. Schädlich, F. M. Göttsche, and F.-S. Olesen, "Influence of land surface parameters and atmosphere on meteosat brightness temperatures and generation of land surface temperature maps by temporally and spatially interpolating atmospheric correction," *Remote Sensing of Environment*, vol. 75, pp. 39–46, Jan. 2001.

[4] F. van den Bergh and P. E. Frost, "A multi-temporal approach to fire detection using msg data," in *Proceedings of Multitemp05*, (Biloxi, Mississippi, USA), May 2005.

[5] F.-M. Göttsche and F. S. Olesen, "Modelling of diurnal cycles of brightness temperature extracted from meteosat data," *Remote Sensing of Environment*, vol. 76, pp. 337–348, June 2001.

[6] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical recipes in C: the art of scientific computing*. Cambridge University Press New York, NY, USA, 1992.

[7] P. Green, "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 46, no. 2, pp. 149–192, 1984.

[8] J. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[9] E. Parzen, "Statistical Inference on Time Series by Hilbert Space Methods," tech. rep., Department of Statistics, Stanford University, Technical Report No. 24, 1959.

[10] E. Parzen, "Regression Analysis of Continuous Parameter Time Series," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability Theory*, (Berkeley, CA), pp. 469–489, University of California Press, 1961.

[11] E. Parzen, "An Approach to Time Series Analysis," *The Annals of Mathematical Statistics*, vol. 32, no. 4, pp. 337–404, 1961.

[12] T. Kailath, "RKHS Approach to Detection and Estimation Problems—Part I: Deterministic Signals in Gaussian Noise," *IEEE Trans. Information Theory*, vol. IT-17, no. 5, pp. 530–549, 1971.

[13] T. Kailath and D. Duttweiler, "An RKHS Approach to Detection and Estimation Problems—Part II: Gaussian Signal Detection," *IEEE Trans. Information Theory*, vol. IT-21, no. 1, pp. 15–23, 1975.

[14] T. Kailath and D. Duttweiler, "An RKHS Approach to Detection and Estimation Problems—Part III: Generalized Innovations Representations and a Likelihood-Ratio Formula," *IEEE Trans. Information Theory*, vol. IT-18, no. 6, pp. 730–745, 1972.

[15] M. A. van Wyk and T. S. Durrani, "A Framework for Multi-Scale and Hybrid RKHS-Based Approximators," *IEEE Trans. Signal Processing*, vol. 48, no. 12, pp. 3559–3568, 2000.

[16] M. A. van Wyk, T. S. Durrani, and B. J. van Wyk, "A RKHS Interpolator-Based Graph Matching Algorithm," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 24, no. 17, pp. 988–995, 2002.

[17] M. A. van Wyk, "Hilbert Space Methods for Non-linear Function Approximation and Filtering," tech. rep., CSIR / LEDGER, 2006.

[18] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley and Sons, New York, NY, 1969.