The use of temporal datasets of remotely sensed images in verifying environmental models

Keith R. McCloy  
Danish Institute of Agricultural Sciences, P.O. Box 50, 8830 - Tjele, Denmark  
Keith.mccloy@agrsci.dk

Wolfgang Lucht  
Potsdam Institute for Climate Impact Research, D-14412 Potsdam, Germany  
Wolfgang.lucht@pic-potsdam.de

Keywords: temporal analysis, remote sensing, model verification

ABSTRACT: Environmental modelling is a critical tool in understanding complex environmental processes, and in predicting the effects of proposed actions on those processes. Many models have been or are being developed for this purpose. This paper will review the development of techniques to monitor the effectiveness of the Lund-Potsdam-Jena (LPJ) model of global vegetation and discuss the role of these techniques in showing the spatial and temporal effectiveness of that model. For this work, global monthly NDVI data, derived from the AVHRR data, were converted into a monthly estimate of Fraction, Absorbed Photosynthetically Active Radiation (FAPAR). This estimated FAPAR was then compared with the monthly FAPAR values derived by the model over the 18 years of the AVHRR record.

1 INTRODUCTION

Science has historically used observational data to formulate hypotheses, and then used other data to evaluate the validity of those hypotheses. Such work not only develops and tests a hypotheses within a range of data values, but often provides a perception of the range of values that can be expected in the main parameters that drive or are affected by the processes involved in the hypotheses. This form of scientific research has been very successful in developing and testing many scientific hypotheses. It has also shown the complexity of many environmental processes due to the complex interdependencies that exist in the environment, and the existence of feedback and feed forward loops of varying duration, intensity and spatial extent. In the identification of these complexities, some limitations of this method of scientific investigation have been revealed.

The main limitation in relation to the physical environment is that it is not powerful enough, on its own, when dealing with very complex situations that cannot be mimicked in the experimental methodology. This complexity arises either because of scaling issues or because interactions between processes severely limits the value of experimental work that focuses on sections of the complex set of processes. In these situations, the experimental methodology is of help in understanding the processes within subsets of the complex interaction between processes, but it is not sufficient.

At the same time, some experimental observations are showing that some environmental parameters are moving outside of the range expected for them, and indeed that some exhibit dynamic
traits not assumed in the original experimental work. When this happens then these assumptions that are made in the original work are no longer valid. Under these conditions, the conclusions drawn from the original experimental work may no longer be valid either. Most of the experimental work conducted over the last 100 years has assumed stable atmospheric conditions. Experimental data from Hawaii and other places is showing that the concentration of atmospheric CO₂ is increasing and there is significant scientific opinion that this will affect atmospheric dynamics in complex, and as yet not understood ways. Thus the assumption of stable atmospheric conditions is no longer valid, and this may affect the validity of some of the scientific work conducted over the last century or more.

The complexity of many of these processes means that descriptions of these processes cannot be based on single, or few interdependent processes, but rather need to account for a complex suite of interdependent processes. Numerical modelling by computer is seen as the most important tool available to us to analyse these complex interdependencies. The first purpose of this numerical modelling is to improve our scientific understanding of this complex set of interactive processes. In this context, numerical modelling has been implanted in the scientific methodology as another step that needs to be implemented in the development and testing of hypotheses. Such a tool is potentially very powerful. Theoretically it can overcome the problem of scale, just as it can, theoretically, be made to mimic complex situations. The second purpose is to use the models to predict the possible outcomes of anticipated or proposed actions or processes. Such a goal is also potentially very powerful. We have never had reliable predictive tools available to us before, so that the value of such tools must be very imperfectly understood. Despite this situation, many people value the potential of predictive tools for many situations. Consequently it is little wonder that much effort is being expended on numerical modelling in many different ways.

Modeling is beset, however, with a number of significant problems. Three of the more significant problems are the recognition that our ability to model many complex sets of processes are limited by our scientific knowledge of those processes, by the occurrence of quasi-random events in the processes, and by non-numerically model able response characteristics that can occur in the processes. The complexity of many environmental processes is not only beyond our current scientific knowledge, but they are very often beyond our current computing power. The most common response to this situation is to attempt to simplify the process, either by simplifying the datasets, by using coarser data, and/or by simplifying the models used in the process. The existence of quasi-random processes defies current numerical techniques. Not only can such events not be handled, but also they have often been found to lead to an exponential divergence from reality. Small errors soon become monumental ones. Non-numerically model able situations arise when either the forward numerical model cannot be inverted, if that is necessary, or when the event itself cannot be modelled.

The most common response to these difficulties in model development is to attempt to understand either the temporal or spatial dynamics of divergence that occurs in the model, but rarely to attempt to understand both. Yet it is precisely this interaction between these two that is most important in the assessment of a models capacity to estimate actual conditions due to a specific complex set of processes. Thus models may be found to work very well under some conditions, and very poorly under others, so that both the spatial and temporal divergences need to be considered. Related to this issue is the one of data used to both train the model and to evaluate the model. Models use and create huge amounts of data. A Global Climate Model at 1 degree resolution will create 64800 values for each date that it derives an estimate. If it is used to derive daily estimates, then it will derive 23, 652,000 values in a year. If it is assumed that a 1% sample is sufficient to assess the quality of this result, then 2, 365,200 independent field data values are required each year. Since model outputs are often of parameters that are difficult to measure, then this represents a huge field
data collection load. It would seem that there is a significant need to develop alternative ways to calibrate and evaluate many models.

Remote sensing has often been sited as a method of collecting data for calibrating and evaluating models. For both of these purposes, the 1% sample that is quoted above may be adequate. But a 1% sample is totally inadequate if the purpose of the remotely sensed data is to understand how, where, when and why models diverge from reality. For this task the sampling needs to be much closer to 100% than to 1%. There is a critical need to develop strategies for the collection of image data for the analysis of model capacity and quality.

The goal of this paper is to report on work done in the development of techniques to assist modellers evaluate the quality of their model in mimicking real world conditions.

2 THEORETICAL BASIS OF THE WORK

Two sets of FAPAR data were used in this work, a set estimated from the AVHRR data and a set derived from the Lund-Potsdam-Jena model [1]. The remotely sensed estimates of FAPAR were derived from the Pathfinder global NDVI data set derived from AVHRR data. This data set contained monthly NDVI values for degree resolution cells. For this study data was used from July 1981 - June 1994 and from July 1995 to June 1998. The dataset formed a 192-image stack consisting of 16 annual cycles of monthly data.

The images have been rectified, calibrated and corrected for some atmospheric effects [2,3]. The calibrated data are then used to compute NDVI images. To create 1 degree by 1-degree pixels, a geometric subset of the pixels in this window is used. The pixel value with the highest NDVI values within this geometric subset was then selected to represent that pixel in that month. This NDVI data was converted into an estimate of FAPAR using the method reported in [4]. This work showed that FAPAR is highly linearly correlated with NDVI, but that the gradient and offset of this line varies with cover type. The IGBP landcover map at 30 seconds of arc resolution was used to derive a percentage landcover map at 1-degree resolution. Pixels within the derived 1-degree landcover percentage data were then found that were more than 90 percent of the one class. These were used as being representative of that class. The temporal NDVI data set was then searched for each of these pixels to find the maximum NDVI values. These values were taken as representing a maximum FPAR value of 0.95 for that cover type. These maximum NDVI values were then used with the cover proportions for each pixel to compute pixel maximum values using a linear mixture model. These maximum pixel values were assumed to match a maximum FPAR value of 0.95. A linear model was then computed for the pixel by assuming that an NDVI value of 0.1 represented a FPAR value of 0.0.

The modeled data is derived from the Lund-Potsdam-Jena (LPJ) vegetation model [1]. This model starts with no vegetative cover. It uses nine functional plant groups. The model replicates the known climate patterns many times to allow the vegetation to evolve. The model derives output data at 30 minutes of arc resolution. This data was scaled to 1-degree resolution pixels using the average value from each of the four source pixels to derive the comparable 1-degree pixel values.

The global datasets will be cyclic in nature, where the mean, amplitude, phase shift and wavelength will vary from place to place. For the theoretical development, this cyclic pattern was assumed to be a sine wave. In this development, the two datasets of FAPAR are compared along the time axes to remove the time parameter from the resulting relationship between the datasets. The shape of the resulting figure depends on the mean, amplitude, wavelength or frequency and phase shift of both source datasets. Consider the different combinations of the two sine waves.
2.1 Similar wavelength and phase shift, different amplitudes and means

\[ W = A_0 \sin(B_0 t + C_0) + D_0 \]
\[ X = A_1 \sin(B_1 t + C_0) + D_1 \]

Remove \( t \) from these equations to give;

\[ W * A_1 - X * A_0 = D_0 * A_1 - D_1 * A_0 \] \( \text{(1)} \)

or \[ W = X * (A_0 / A_1) + (D_0 - D_1 * (A_0 / A_1)) \] \( \text{(2)} \)

The figure resulting from (1) or (2) is a line.

2.2 Similar wavelength, but different phase shift and amplitude

\[ W = A_0 \sin(B_0 t + C_0) + D_0 \]
\[ Y = A_1 \sin(B_0 t + C_1) + D_1 \]

Eliminate \( t \) from these equations, to give:

\[ (Y - D_1)^2 / (A_1)^2 - 2 (Y - D_1) (W - D_0) \cos(\delta C) / (A_0 A_1) + (W - D_0)^2 / (A_0)^2 = \sin^2(\delta C) \] \( \text{(3)} \)

When \( \delta C = 0 \) then this equation becomes the same as (1). The function depicted by (3) is an ellipse with its elliptical eccentricity being a function of \( \delta C \) and the amplitudes, \( A_0 \) and \( A_1 \).

2.3 Different wavelength, phase shift, and amplitude

\[ W = A_0 \sin(B_0 t) + D_0 \]
\[ Z = A_1 \sin(2B_0 t) + D_1 \]

Eliminate \( t \) from these equations, to give:

\[ K_0(W - D_0)^8 + K_1(W - D_0)^6 + K_2(W - D_0)^4 + K_3(W - D_0)^2 + K_4(Z - D_1)^4 + K_5(Z - D_1)^3 \]
\[ + K_6(Z - D_1)^2 + K_7(Z - D_1) + K_8(W - D_0)^6(Z - D_1)^2 + K_9(W - D_0)^4(Z - D_1)^2 + K_{10}(W - D_0)^2(Z - D_1)^2 \]
\[ + K_{11}(W - D_0)^2(Z - D_1) + K_{12}(W - D_0)^2(Z - D_1)^2 + K_{13}(W - D_0)^2(Z - D_1)^3 + K_{14} = 0 \] \( \text{(4)} \)

where the constant terms \( K_0 \) to \( K_{14} \) are complicated functions of \( C_0, C_1, A_0 \) and \( A_1 \). The shape of such a figure, for specific values of \( C_0, C_1, A_0 \) and \( A_1 \) is shown in Figure 1.

In general, when the two sine waves only vary in mean value and amplitude, then the comparison yields a linear relationship (2). The gradient of this line is a function of the ratio of the two amplitudes as shown in (2). The offset of this line depends on both the mean values and amplitudes of both signals. If there is only a difference in mean values between the signals, then the gain will be about 1.0 and the offset will be the difference between the mean values. If the signals have similar values at the troughs, then the offset will be about zero. Thus both the gain and the offset provide useful information on the two signals.

When the two waves have different phase shifts, then the resulting figure is an ellipse (3). When the phase shift is exactly 180 degrees, or the amplitudes of the two signals are the same, then the figure is a circle. This transformation can be reversed by introducing a phase shift that is opposite in sign to the first phase shift. This concept is used to find the phase shift at which a linear relationship holds between the two signals. Thus in the comparison, a phase shift of one value is introduced by comparing the first value in one signal with the value at the phase shift in the other signal. In annual data, a phase shift of zero months, one month, two months, up to eleven months are used.
and the data at these phase shifts are then compared. The maximum $R^2$ values are then used to indicate the best fit between the datasets, and this phase shift is then taken as the phase shift that exists between the datasets.

The shape of the frequency-doubling figure is always a figure of eight, but its shape and orientation depend on the amplitudes and phase shifts of the two signals. In Figure 2(b), three examples of frequency doubling are shown. These show the effects of changes in amplitude and phase shift as well as the frequency doubling.

The situations depicted in Figure 1 are the more common situations that arise in long term global image data. They are not the only cyclic components that may contribute to the signal at a

![Figure 1. Comparison of a reference sine wave with a second sine wave. (a) - the various source signals for the comparison, (b) - the comparison of each signal with the reference signal.](image)
location, but these other causes will be ignored in this development. If the goal is to discriminate between the types of comparative signals shown in Figure 1(b), then it is necessary to develop a technique for doing this.

The way chosen to achieve this goal was to take the two signals that are to be compared, and conduct the comparison by introducing a phase shift in one signal relative to the other. A comparison was conducted at each phase interval within a full cycle of the data. A linear regression was conducted, and the best solution was taken to be that with the highest $R^2$ squared value from the regression. If the $R^2$ value remains low at all phase shifts, then the figure is assumed to be a Figure of Eight, caused by a change in wavelength.

This methodology provides a mechanism for discriminating differences in the mean, amplitude and phase shift between the two time series, and discriminating these from differences in frequency or wavelength by use of values in the Coefficient of Determination derived from linear regression. As currently developed it does not provide a mechanism for discriminating differences in the variation in frequency or wavelength.

3 CONDUCT OF THE ANALYSES

Two comparisons were conducted. The first was a comparison of eight years of image data against the second eight years of image data. The second comparison was between the sixteen years of data in the two data sets. In each comparison the work was conducted in two stages. The first stage took each cycle of twelve months of data and conducted the analysis to derive the best linear regression and its $R^2$ value. The best linear regression was that one with the maximum $R^2$ value, of the twelve different phase shift comparisons that were conducted within each twelve-month set of data. This stage yielded four parameters for each twelve monthly cycle: gain, offset, phase shift and $R^2$ value. The second stage simplified this dataset by deriving statistical data from these derived parameters and grouping these to form a temporal image. In this work, the combinations chosen included:

- The average gain setting for the cycles that had Coefficient of Determination values greater than an operator set threshold. This threshold was set at 0.975.
- The average offset value for the same cycles as in 1 above.
- The average phase shift for the same cycles as above.
- The average phase shift for the cycles that did not meet the threshold value in their Coefficient of Determination.
- The proportion of the cycles that met the threshold in their Coefficient of Determination.

4 RESULTS OF THE ANALYSIS

A summary of the results are shown in Figure 2.

4.1 Comparison between the two AVHRR based FPAR datasets

Much of the map is red, indicating that there is close correspondence between the two datasets. Most of the remainder of the map is green, indicating close correspondence, but with a shift in phase between the datasets. The strongest anomalies are in the Thar desert of India – Pakistan, and the Tikla Makan to Gobi desert in China. Since NDVI values will usually be very low in these areas, so a small change can create a large response. The anomalies in these areas can probably be expected. The results show that the technique is working as expected.
4.2 Comparison between the AVHRR and LPJ based FPAR datasets

There was little red in this map, extensive green areas and a significant amount of blue and black. The red areas show good correspondence, occurring in parts of the Sahel (Tropical Savannah), South east USA (Humid Sub-tropical), North East China (Continental), Japan (Continental) and the Balkans (Marine West Coast).

Moderately good approximation was shown in green for South east Asia (Humid Sub-Tropical), southern India (Tropical Savannah), southern Africa (Tropical Savannah) and most of South America Figure (Tropical Savannah). Significant anomalies have been found in northern India (Humid Sub-Tropical), northern Sahel (Tropical Steppe), northern Africa (Desert), parts of South America (Tropical Savannah) and central USA (Continental, Steppe and Sub-Arctic). The comparison immediately suggests that the LPJ model is most successful at mimicking warm moist to wet climates and not as good at mimicking dry and cold climates. There are some anomalies to this observation, notably northern India and parts of South America. One potential explanation for this is the impact of man, as this will be significant in northern India. The causes of the anomaly in South America are not known.

![Comparison of Datasets](image)

**Comparison of Datasets**

Modelled data using the LPJ model versus AVHRR derived data (1981 - 1998)


**Colour assignment**
- **Red** - Areas that fit the linear model with R-squared values better than 0.975 and zero phase step.
- **Green** - Areas that fit the linear model with R-squared values better than 0.975, but with a phase step.
- **Blue** - Areas that do not fit the model at this R squared value.

**Tonal assignment**
- **Red** - Percentage of cycles that fit the above criteria.
- **Green** - Percentage of cycles that fit the above criteria.
- **Blue** - Most frequent phase shift.

Figure 2. Comparison of datasets, showing a summary of the results of the two comparisons reported in this paper.
CONCLUSIONS

Long-term temporal image data sets contain a wealth of information on the spatio-temporal characteristics of those dynamic processes that are captured in the image data. Earlier work has shown that very general information on these dynamic processes can be collapsed from the temporal sequence into a temporal image that is then amenable to visual interpretation and analyses.

This paper has shown, that in comparing image data with model estimates, very specific information can be derived and then depicted as temporal images on the comparison. Again, such images can be used visually for interpretation and analysis so as to improve our understanding of when, under what conditions, and how the model mimics the real world. Such analyses is an essential component, with appropriate numerical analyses, in the development and subsequent testing of hypotheses concerning the characteristics of the model and its capacity to mimic the real world.

REFERENCES


