Use of a Nonlinear Dynamic Limit-Cycle Model to Identify Perturbations Embedded in Surface Energy Flux Data

M.-D. Cheng and W. F. Lawkins
Oak Ridge National Laboratory
Environmental Sciences Division
Oak Ridge, Tennessee

Introduction

With the advent of real-time in-situ measurement capabilities, the dynamics (evolution of the system states) of natural (e.g., atmosphere) or manmade (e.g., industrial engineering) processes can be probed with an unprecedented temporal and spatial resolution. Extremely large volumes of ambient data describing atmospheric states (meteorology, radiation, clouds/aerosols, and water) and evolution of the states can be produced. Generally, the data describing the evolution of atmospheric states are time series of measurement variables that may be contaminated with noise. Many complex phenomena, including atmospheric processes, are nonlinear and sensitive to perturbations. Development and application of methods to analyze large volumes of nonlinear time series data have been limited. Analytical time series models developed to investigate nonlinear atmospheric processes based on the chaos theory (e.g., Tsonis et al. 1994; Wang 1995) are recent but promising. However, most of the techniques are sensitive to noise in the data and to the size of the data set.

To identify noisy segments in the time series, we applied a technique based on the theory of nonlinear dynamics time series (NDTS). A segment consists of a single datum (i.e., a single outlier or a spike) or a block of contiguous data points. We focused on the surface energy transfer processes using data produced by the Energy Balance Bowen Ratio (EBBR) stations of the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Program from January 1, 1994, to December 31, 1995. This analytical model can, however, be applied to data other than those of EBBR. The choice of EBBR data was based on the convenience and possibility of verifying our results.

Objectives

The objective of this research was to develop a physics-based analytical technique for identifying perturbations in an atmospheric time series data.

Description of Data

Currently, ten EBBR stations at Cloud and Radiation Testbed (CART) sites are producing continuous measurements for the energy flux calculation. In this meeting, we present the analysis results using the data produced by the EBBR station at the Central Facility (CF). The sample resolution of the heat fluxes was 30 minutes. Thus, there were 48 data points per day and 35,040 points from January 1, 1994, to December 31, 1995. It is well known that the surface heat flux calculated by the Bowen Ratio (BR) technique is unreasonable when the BR is near -1. Wesely et al. (1995) suggest that heat fluxes whose corresponding BR values were in the range of [-0.75, -1.5] were unreliable. BR was calculated as the ratio of sensible to latent heat fluxes. We used all the data points (including bad BRs) in the analysis such that these data points could provide a benchmark to test our perturbation analysis technique.

Brief Theory for Process Modeling and Perturbation Analysis

To use the perturbation analysis technique developed at the Oak Ridge National Laboratory to identify anomalous data in a time series, one has to model the underlying process using the observable time series data. The evolution of the process involving surface-atmosphere energy exchange was modeled by constructing a limit-cycle attractor. Perturbation can exist when 1) an anomaly occurs in the underlying process because of external forcings, and/or 2) measurement error contaminates the observable values. We adopted a theory based on Broomhead and King (1986) and refined by Lawkins et al. (1996) to identify anomalous observations. Once the attractor is built, anomalous energy fluxes can be identified by examining a trajectory projected into the state space containing the attractor. A trajectory is a mathematical description of the time-dependent evolution of the surface radiative energy transfer processes. Most of the time the trajectory is expected to reside on the attractor. If occasionally the trajectory leaves that attractor and moves about in an extended region of phase space before falling back to the attractor, that trajectory segment will be called a perturbation or change block. A block can consist of a single datum or several data points contiguous in time. Such a change block is significant and may represent an anomaly in the atmospheric data. Natural variability of atmospheric processes is one cause of anomaly in the observations. Measurement noise can also cause the trajectory to fly off the attractor, resulting in anomalies in the observations. In theory, either type of anomaly would be identified by our analytical technique. Suppose a correlation matrix M is defined as

$$M = \frac{1}{N} \sum_{i=0}^{N} y_i y_i^T \tag{1}$$

where the superscript T is the transpose operation. If we denote the eigenvalues (or squares of the singular values) and principal modes of M in pairs by $[(\sigma_j^2, \phi_j)]$ where j is from 1 to n and the singular values are in descending order, and we define the n by n matrix ψ as $\psi = (\psi_1, \psi_2, ..., \psi_n)$ so that the principal mode ψ_j is the j-th column of ψ, then

$$\Theta = \Psi^T y \tag{2}$$

is an orthonormal transformation Ψ: $E_y^n \rightarrow E_\Theta^n$ that transforms $A_y = \{y_i\}$ to $A_\Theta = \{\Theta_i\}$, I = 1, N. In general, we expect the j-th mode to have approximately j/2 waves over the time span of a point y_i, which is the window time scale t_w as defined above. Consequently, the j-th coordinate in E_Θ^n corresponds approximately to information in the time series resolved by the frequency

$$f_j = \left[\frac{t_w}{j/2}\right]^{-1} = j \text{ x } (½\ f_w) \tag{3}$$

where f_w is the windows frequency. The singular values are then estimates of the second moments of each coordinate value in the set A_Θ and determine the length scales in the phase space $E_\Theta{}^n$ for A_Θ.

Our experience indicated that candidate values for the reconstruction parameters (n, k) can be based on the singular values σ. To resolve information in the time series, we want the lowest order of the principal component frequencies f_j, j =1, ..., m, where $m \le n$, which are considered to be most important. Alternatively, we can choose the largest singular values of m to resolve the state-space range of length scales associated with the frequencies f_j, j = 1, ..., m. Recall the ordering of singular values. If σ_{m+1} is sufficiently small, then the projection

$$P^m : E_\Theta^n \rightarrow E_\Theta^m \tag{4}$$

where E_Θ^m is the linear subspace of E_Θ^n spanned by the first m principal components will result in eliminating small scale, noisy detail in the highest order coordinates. The resulting attractor is then defined by the ordered triple of parameters (n, k, m), where m is the number of principal components retained in the model.

Using the constructed attractor we could identify any trajectory segments that had traveled off-course for a certain period of time because of a data anomaly. Given the parameter values (n, k, m), we can then define a projection:

$$P_\Theta^{m1,m2} : E_\Theta^n \rightarrow E_\Theta^{m2-m1+1} \tag{5}$$

where

$$P_\Theta^{m1,m2}(\Theta) = (\Theta_{m1},......,\Theta_{m2})^T \in L[\psi_j]_{j=m1}^{m2} \tag{6}$$

and $L[\psi_j]$ is the linear subspace of E_Θ^n spanned by the eigenvectors. We assume that data produced by a natural radiative transfer process in the atmosphere are generally concentrated in a highly dense region B for a small value of m2 - m1 + 1. The region B is defined as $B = P_\Theta^{m1,m2}(\Theta_i)$, I=1,N. Thus, B is the projection into the (m2 - m1 + 1)-dimensional subspace L of the trajectory Θ in E_Θ^n. In other words, an anomaly in the time series will produce a trajectory segment in $L\{\psi_j\}_{j=m1}^{m2}$ that moves outside of the region B. Let $\Gamma_j = \{\Theta_i\},\ i = i_j$ to $i_j + l_j - 1$ where l_j is the length in time steps of Γ_s. Let T_s be a characteristic time scale of the process, describing a minimum time frame for discriminating a normal from abnormal event. If the following conditions are met,

$$\begin{gathered} P_\Theta^{m1,m2}(\Theta_{ij-1}) \notin B \\ P_\Theta^{m1,m2}(\Gamma_j) \subset B \end{gathered} \tag{7}$$

then the trajectory segment Γ_j is not an anomaly and the time segment corresponding to Γ_j is a normal data block. In turn, define $\Delta_j = \{\Theta_i\}$, I is from $i_j + l_j$ to $i_{j+1} - 1$ that separates Γ_j and Γ_{j+1} to be the jth

$$\begin{gathered} P_\Theta^{m1,m2}(\Theta_{i_j+l_j}) \notin B \\ l_j \text{ x } t_s \ge T_s \end{gathered} \tag{8}$$

anomaly segment. The length in time steps of Δ_j is $p_j = (i_{j+1} - 1) - (i_j + l_j) + 1 = i_{j+1} - (i_j = l_j)$. This definition allows a trajectory segment to pass through the region B so long as the time it takes is less than the time scale T_s. The time series segment Δ_j is by definition an anomalous segment with respect to the time scale T_s. It is important to remember that the choice of T_s is application-specific and would have significant influence on the results of the perturbation analysis.

Results and Discussion

The analysis used a time scale of T_s = 6 hours to define data anomalies (i.e., perturbations). This time scale was fine enough to allow us to resolve data anomalies that might have resulted from diurnal variations. We investigated the use of other time scales ranging from 9 to 72 hours and found that the 6-hour time scale was the probe that provided the most details. A time scale finer than 6 hours did not provide any further details in perturbation analysis, indicating a saturation of information transmission beyond 6 hours. If a trajectory segment remained longer than 6 hours in the attractor, the data points on this segment would not be identified as anomalous. Figure 1 is a plot for sensible heat flux data from the April 1-30, 1996, segment. The darker color points were those identified as anomalous, and the lighter ones were normal. Using the BRs calculated for this time section (not included in this extended abstract because of page limitations), we could confirm that most of the darker colors were the points whose BR values were in the range of [-1.5, -0.75]. There were points whose BR values were not in that range (i.e., the bad data points) but which were caught in the time segments and color-coded red. This result is a limitation of the current NDTS identification technique, which is, e.g., those embedded in a bad segment. We will present more results in the poster.

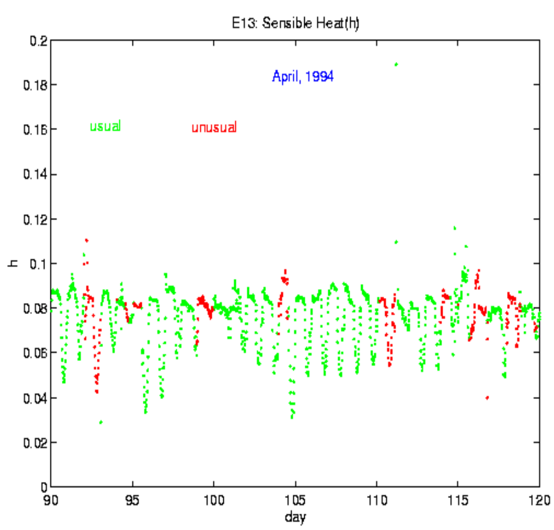

Figure 1. Plot of sensible heat flux data obtained April 1-30, 1996.

Conclusions

The NDTS methodology developed at the Oak Ridge National Laboratory was employed to identify perturbations in a large data set of surface heat flux obtained by the ARM Program. Contaminated data points produced by the EBBR stations were identified by the analysis technique and confirmed by the BRs. These data points showed anomalous behavior in the phase space into which the trajectory was projected, which we used to our advantage to identify them.

References

Broomhead, D. S., and G. P. King, 1986: Extracting Qualitative Dynamics from Experimental Data, *Physica*, **20D**:271.

Lawkins, W. F., N. E. Clapp, Jr., C. S. Daw, L. M. Hively, V. Protopopescu, and M. L. Eisenstadt, 1996: *Modeling and Analyzing Non-Seizure EEG Data for Patients with Epilepsy*, Technical Report ORNL/TM, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Tsonis, A. A., G. N. Triantafyllou, J. B. Elsner, J. J. Holdzkom II, and A. D. Kirwan, Jr., 1994: An Investigation of the Ability of Nonlinear Methods to Infer Dynamics from Observables, *Bull. Amer. Meteo. Soc.*, **75**:1623-1633.

Wang, Q., 1995: Correlation Dimension Estimates of Global and Local Temperature Data, *J. Appl. Meteo.*, **34**:2556-2564.

Wesely, M. W., D. R. Cook, and R. L. Coulter, 1995: Surface Heat Flux Data from Energy Balance Bowen Ratio Systems, *Preprints of the 9th Symposium on Meteorological Observations and Instrumentation*, pp. 486-489, Charlotte, NC, March 27-31, 1995, *Amer. Meteo. Soc.*, Boston, Massachusetts.

Acknowledgment

This research was supported in part by the DOE ARM Program. Oak Ridge National Laboratory is managed by the Lockheed Martin Energy Research Corp. for the U.S. Department of Energy under contract number DE-AC05-96OR22464.