Nonlinear Processes in Geophysics (2005) 12: 195–200
SRef-ID: 1607-7946/npg/2005-12-195
European Geosciences Union

Nonlinear Processes in Geophysics

Observing extreme events in incomplete state spaces with application to rainfall estimation from satellite images

A. A. Tsonis[1] **and K. P. Georgakakos**[2, 3]

[1]Department of Mathematical Sciences, Atmospheric Sciences Group, University of Wisconsin-Milwaukee, Milwaukee, WI 53201-0413, USA
[2]Hydrologic Research Center, 12780 High Bluff Drive, Suite 250, San Diego, CA 92130, USA
[3]Scripps Institution of Oceanography, UCSD, La Jolla, CA 92093-0224, USA

Received: 11 August 2004 – Revised: 22 December 2004 – Accepted: 23 December 2004 – Published: 2 February 2005

Part of Special Issue "Nonlinear deterministic dynamics in hydrologic systems: present activities and future challenges"

Abstract. Reconstructing the dynamics of nonlinear systems from observations requires the complete knowledge of its state space. In most cases, this is either impossible or at best very difficult. Here, by using a toy model, we investigate the possibility of deriving useful insights about the variability of the system from only a part of the complete state vector. We show that while some of the details of the variability might be lost, other details, especially extreme events, are successfully recovered. We then apply these ideas to the problem of rainfall estimation from satellite imagery. We show that, while reducing the number of observables reduces the correlation between actual and inferred precipitation amounts, good estimates for extreme events are still recoverable.

1 Introduction

In inference problems concerning spatially extensive physical systems, it is often the case that available remotely-sensed spatially-extensive data do not directly measure system state variables. In such cases, a complete characterization of these systems in state space is not available and estimation of system response must be done through the available state measures. Along these lines, remotely-sensed observations are combined with in situ or remotely-sensed observations of system response to form observation-response relationships (typically based on regression analyses). These relationships are then used in areas and spatial scales where response observations are non-existent to estimate system response. For example, in the atmosphere, infrared (IR) and visible (VIS) observations have been used in this manner to estimate response, such as surface precipitation (e.g. Scofield and Oliver, 1977; Tsonis and Isaac, 1985; Arkin and Meisner, 1987; Adler and Negri, 1988).

Correspondence to: A. A. Tsonis
(aatsonis@uwm.edu)

In all such cases, the observables depend on a portion of the state vector of the system flow, while the response does not necessarily depend on the same portion of the state vector. The point is whether the observational problem as described leads to useful data and response estimates. The underlying state vector is not known but it may allow regions in state space where the system flow evolves about strange attractors. Thus the question posed is: For a nonlinear system with chaotic dynamics, are indirect observations of a part of the system state vector adequate to identify desired characteristics of the system response such as extreme response variability? We will approach this problem first by considering a known dynamical system and then by extending the methods developed to observed data related to delineation of rain amounts from satellite images.

2 Mathematical formulation

We start with the mathematical formulation of the problem, which is demonstrated and explored with a simple nonlinear dynamical system described by the following equations (Lorenz, 1963). This well-studied system approximates the behavior of a layer of fluid of infinite horizontal extent, which is subject to a temperature-difference forcing of $\Delta T(>0)$ between the lower and upper surface. As the fluid is heated in contact with the warmer surface it rises and creates convection. The system governing equations are shown next for establishing notation.

$$\frac{dX}{d\tau} = -\sigma X + \sigma Y \tag{1}$$

$$\frac{dY}{d\tau} = -XY + rX - Y \tag{2}$$

$$\frac{dZ}{d\tau} = XY - bZ, \tag{3}$$

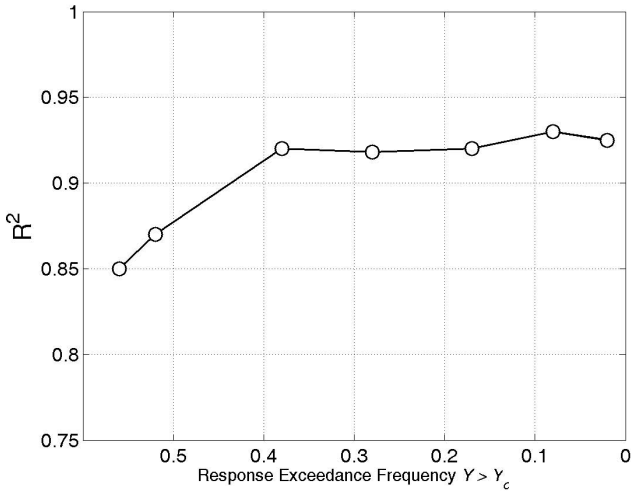

Fig. 1. Regression correlation coefficient and sample size for $Y > Y_c$ as functions of response exceedance frequency.

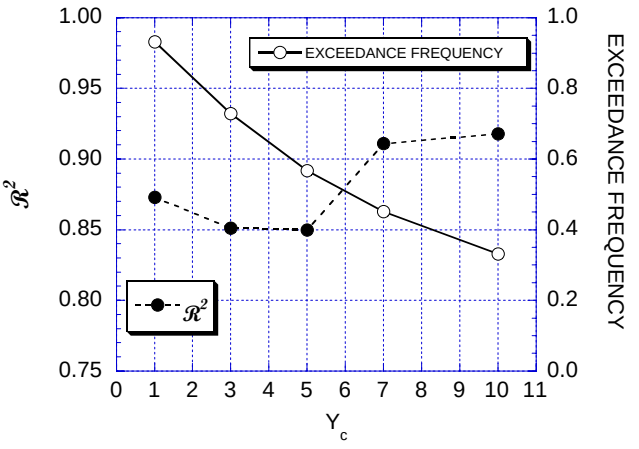

Fig. 2. Regression correlation coefficient and response exceedance frequency as functions of parameter Y_c.

where X, Y, and Z are the state variables with X being proportional to the intensity of the convective motion, Y being proportional to the temperature difference between the ascending and descending currents, and Z being proportional to the distortion of the vertical temperature profile from linearity. It is known that for σ=10, b=8/3 and r in the range 24.74 to 31.10, the Lorenz system possesses chaotic dynamics and a strange attractor for large integration times (Berge et al., 1984).

The observation problem may be stated as follows (Georgakakos and Tsonis, 2001): Given observations, possibly noisy, of one or two of the states of the Lorenz system, estimate system response which may be a function of one or more states (some unobservable). In the simplest case, we postulate observations that are linear functions of certain system states

$$O_v = \varepsilon X + v_1 \tag{4}$$

$$O_i = \delta Z + v_2, \tag{5}$$

where ε and δ are coefficients, and v_1 and v_2 are independent random processes with uniform distribution functions in the intervals $[-V_1, +V_1]$ and $[-V_2, +V_2]$, respectively. We also postulate a positive response function, which is a linear function of the remaining system state:

$$P = c_2 Y + w;\ c_2 > 0;\ Y > Y_c \tag{6}$$

with w possessing a uniform distribution function in the interval $[-P_2, +P_2]$. The noise terms v_1, v_1, and w represent the effects of non-modeled components in the observation process and the system response function. The only assumption we will make for the response noise w is that P_2 is inversely proportional to Y for large Y. That is, the contribution of non-modeled effects diminishes for high response and, for such a regime, the Lorenz system is largely driving the response function. It is then postulated that w possesses a uniform distribution function in the interval $\left[-\frac{D}{Y}, +\frac{D}{Y}\right]$, with $Y \geq Y_c$, and D a scale parameter.

Other combinations of observation and response functions are possible with results analogous to those obtained from the set (Eqs. 4–6). Note also that, either both or only one of Eq. (4) or Eq. (5) may be used as an observation equation to estimate the response P as defined in Eq. (6).

3 Numerical experiments

Numerical experiments were performed by numerically simulating the Lorenz system (Eqs. 1–3). For given V_1, V_2, D, and Y_c, observations O_v and O_i are simulated using Eq. (4) and Eq. (5) and response values P greater than zero are simulated using Eq. (6). It is supposed that there is no knowledge of the underlying nonlinear system in real cases, and we wish to estimate the response from the observations. As it is often the case in practice, a multiple linear regression relationship is then established between P and $(O_v,\ O_i)$, and the regression correlation coefficient, $\Re$, is recorded (the square of this coefficient is the portion of variance in the response explained by the observations). The relationship used for our analysis is:

$$P = \alpha_1 O_v + \alpha_2 O_i + \alpha_0 + e, \tag{7}$$

where α_0, α_1, and α_2 are regression parameters and e is the regression error. The above formulation is designed to mimic several observational problems in atmospheric sciences, for example, rainfall estimation from space, where rainfall is estimated from a combination (typically linear) of a few observables of the climate system (such as visible and infrared images, Kidder and Vonder Haar, 1995). The analysis is done for various threshold values P_T for which $P > P_T$ in order to probe the reliability of estimating extreme values of P. The sensitivity analysis examines the behavior of $\Re^2$ when varying the quantities: V_1, V_2, Y_c, and P_T. The constant coefficients used in the simulations are: σ=10, b=8/3, r=28,

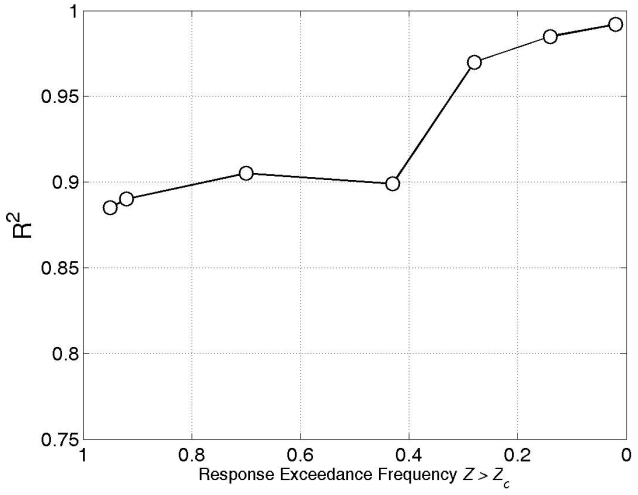

Fig. 3. Regression correlation coefficient and sample size for $Z > Z_c$ as functions of response exceedance frequency.

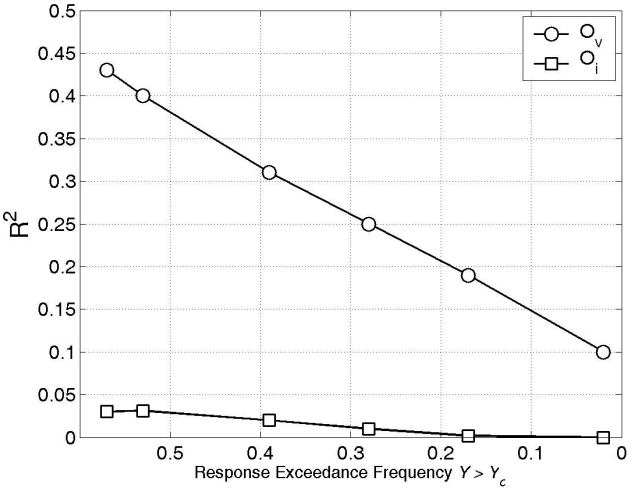

Fig. 4. Regression correlation coefficient and sample size for $Y > Y_c$ as functions of response exceedance frequency for single observations O_v or O_i.

ε=13, δ=5, c_2=3.5. We note here that the correlation coefficient measures linear dependence. However, in many nonlinear problems, for example, evaluation of linear and nonlinear prediction, the correlation coefficient between actual and predicted is a common measure of performance (Sugihara and May, 1990; Tsonis, 1992).

Figure 1 shows $\Re^2$ as a function of the response exceedance frequency for the case of V_1=V_2=P_2=0 and Y_c=5. It is evident that as the exceedance frequency decreases the observables O_v and O_i explain a larger portion of the response variability (from 85% to 92% of response variance). Dependence of the result on Y_c may be discerned from Fig. 2, which shows $\Re^2$ as a function of Y_c. The exceedance frequency resulting from a certain value of Y_c is also shown. For the results shown in Fig. 2, the rest of the parameters were set to the values used to produce the results of Fig. 1, with P_T=0.1. The increase of the explained portion of response variance with increasing Y_c is evident (from about 87% to about 92%). This result corroborates that of Fig. 1 in that in both cases for a reduction of response exceedance frequency there is an increase of $\Re^2$.

The character of the results (better reproduction of the system response by the observables for extreme cases than otherwise) was preserved when other response functions and observables were used. For example, when the response P was defined as a linear function of the system state Z, with $Z > Z_c$ in analogy to Eq. (10), and the observable O_i was defined as a linear function of Y, in analogy to Eq. (6), the analysis produced the results shown in Fig. 3 (analogous to Fig. 1).

In cases with single observables (either one of O_v or O_i), the reproduction of the response may be shown to be poor throughout the range of response magnitude, and especially for the extremes. Figure 4 shows results for the case of single observables. The parameter values used are: V_1=V_2=0, D=0, Y_c=5. As P_T increases, the exceedance frequency of the response decreases. There are two curves corresponding to O_v and O_i considered individually as observables. It is notable that the results in this case are much worse than those obtained when both O_v and O_i were used as observables (at best only about 40% of the response variance is explained in the present case). Also, O_v is a much more suitable observable than O_i for cases when a single observable is used to reproduce the system response P. Additional results were obtained (but not shown) for a number of values of the parameters. It was found that the character of the results in Fig. 4 is preserved for other values of the parameters.

Next we investigated the effect of noise on the results. Figure 5 shows the dependence of $\Re^2$ on the observation noise parameters V_1=V_2 for two values of the response noise scale parameter D (D=0 on top, D=200 on bottom) and for two values of the response threshold P_T=0.1 (circles) and P_T=50 (squares). These noise ranges cover a range of noise from zero noise (V_1=V_2=0, D=0) to rather large noise ($V_1 \rightarrow 30$, $V_2 \rightarrow 30$, D=200). The results correspond to the set of observables and response defined in Eqs. (4)–(6). We observe that increased observation noise results in deterioration of response reproduction by the observables (negative slope of curves). We also observe that large response noise dominates the estimation of the response by the observables, with very different values of P_T producing similar results (bottom). Thus, the introduction of moderate noise does not influence the results of the above experiments greatly but it does reduce $\Re^2$ somewhat, especially for the higher exceedance frequencies. However, the presence of high noise in observations or response due to non-modeled effects substantially deteriorates the ability of observables to reproduce system response variability. This is attributed to the substantial change (caused by the presence of high noise) in the morphology of the flow on the strange attractor when mapped onto the response-observables space (for more details on the effect of noise on the results see Georgakakos and Tsonis, 2001).

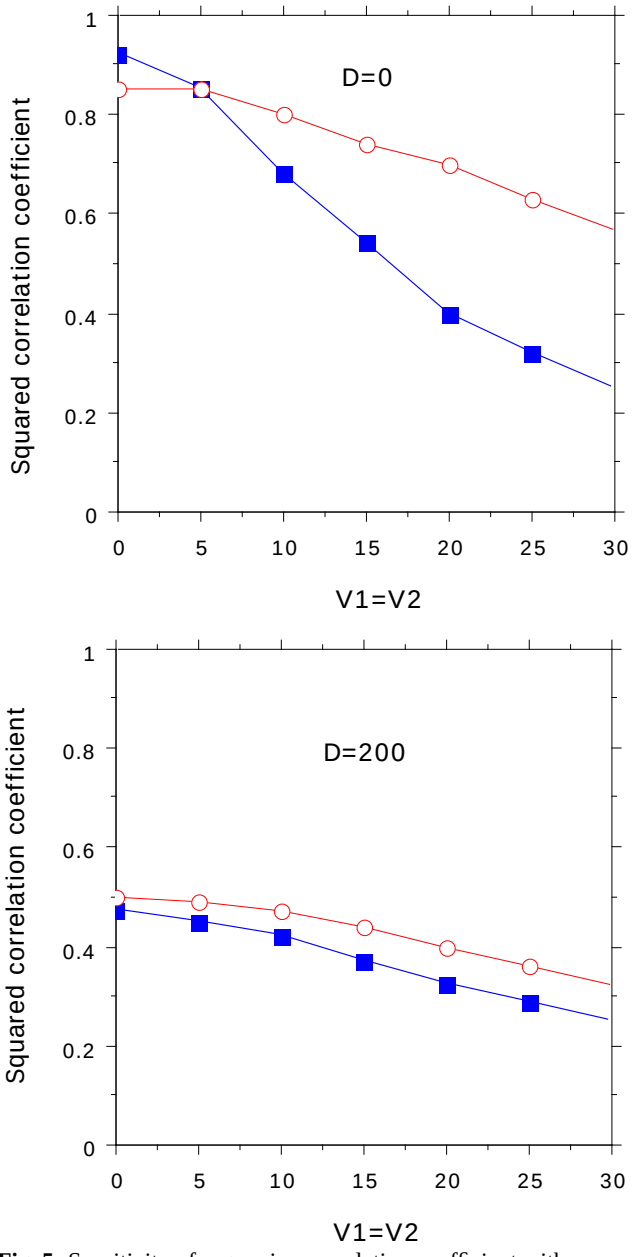

Fig. 5. Sensitivity of regression correlation coefficient with respect to observational noise strength V_1=V_2, for two values of the response noise scale parameter D (D=0 on top, D=200 on bottom) and for two different exceedance threshold P_T, P_T=0.1 (circles) and P_T=50 (squares).

We conclude that for this known dynamical system, even in a reduced state space (two out of three variables), good estimates, especially for extreme values, can be obtained. When the state space, however, is reduced too much (one out of three variables) even the extremes may not be estimated with any desired accuracy.

4 Application to rainfall estimation from satellite images

We are now considering the problem of estimating rainfall amounts from satellite visible and infrared images. Visible and infrared images are very important in estimating rain amounts over areas of the globe where conventional radars cannot be used. We assume that rainfall amount, radiation in the visible (VIS), and radiation in the infrared (IR) are three of the variables of the climate system, which is high dimensional. By trying to estimate rainfall from only visible and infrared images, in effect we are trying to estimate a response from information from just a few other observed variables of the climate system. Truly enough, the amount of rainfall is related to how thick the clouds are (visible count) and how tall the clouds are (infrared count). However, precipitation depends on many other factors. As such, estimating precipitation from information in the visible and infrared frequency domain only represents estimation in a much lower state space.

The data used here are the same as the data used in Tsonis et al. (1996). This data set includes visible (0.54–0.70 μm) and infrared (10.5–12.6 μm) images over the Des Moines River basin (an area of about 15 000 km^2) as well as extensive and prototype real-time hydrometeorological database (a network of 29 rain gauges), which can be used to compute mean areal rainfall amounts. The spatial resolution of the satellite data is 4×4 km. The temporal resolution is 30 min but we only consider images every three hours to ensure that successive images will not be correlated (Tsonis and Isaac, 1985). The period for which data were available to us is May–September 1982–1988. For each pair of VIS/IR images, their bivariate distribution is obtained and based on this distribution and the Tsonis and Issac (1985) approach, a given pair is classified as rainy or non-rainy. Note that, since visible information is involved in the estimation of rainfall, only images during daylight can be used. This introduces errors because if the rain falls at night, the satellite rainfall amount estimation will be underestimated. In order to minimize these errors, only days with at least three VIS/IR pairs available were considered (for more details on the data, see Tsonis et al., 1996). Once this has been done the daily mean areal rainfall amount was regressed with six variables, which relate to certain properties of the images (Tsonis et al., 1996). These six variables are: 1) the daily mean rain area as estimated from the images according to Tsonis and Isaac (1985), 2) the daily mean relative frequency of the peak in the bivariate distribution, which corresponds to raining clouds, 3)–4) the daily mean coordinates of that peak in the VIS/IR domain, 5) the daily mean narrowness of that peak, and 6) the daily mean cloud area (Tsonis et al., 1996). The multivariate model resulted in a multiple regression coefficient of determination between actual and predicted rainfall amounts of $\Re^2$=0.70 for all events. $\Re^2$ gives the amount of variance of the daily mean rainfall amount (estimated by the rain gauge network), which is explained by the regression model. When instead we consider only the upper half (in intensity) events $\Re^2$ is increased to 0.81. In contrast, when only the lower half (in intensity) events are considered $\Re^2$ drops to 0.52. Thus, while we get a good overall correlation, the correlation is much stronger when only strong events are considered. This will indicate that in our lower VIS/IR state space significant

information about extreme events is recoverable.

So, what will happen if we decrease our state space even further? In order to address this issue, we considered the same problem but now we used only IR images. In this case the daily mean areal rainfall amount is regressed with four variables relating to infrared images alone (Lehnes, 1996). These variables are: 1) the daily mean rain area as delineated by an optimum infrared threshold, 2) the mean relative peak of the univariate distribution in the infrared domain, 3) the mean coordinate of that peak in the infrared domain, and 4) mean the narrowness of that peak. Now the multivariate model results in a multiple regression coefficient of determination of $\Re^2$=0.52 for all events. When the upper half (in intensity) events are considered $\Re^2$ is increased to 0.62. When the lower half is considered $\Re^2$ drops to 0.40. These results are not as impressive as those in the VIS/IR domain (they represent a loss of 18%, 19% and 12% in variance explained, respectively) but still it appears that some information (especially for the extremes) exists. All the above results are summarized in Table 1.

While the above results may have been anticipated, this study looks at the problem of rainfall estimation from space from the dynamical systems point of view. The reduction in $\Re^2$ seen when only IR images are considered is consistent with the reduction observed in our simple dynamical system when the dimension of the state space becomes too small. Here as well, it appears that while some rainfall information exists in a very low dimensional space, when it comes to studying the variability of rainfall from satellite images the best bet is to study extreme variability. For weak events the properties of rainfall may not be adequately resolved.

Table 1. Multiple regression coefficient of determination, $\Re^2$, between the actual and inferred rain amount by a regression model using VIS and IR information and a regression model using only IR information. $\Re^2$ gives the amount of variance of the daily mean rainfall amount (estimated by the rain gauge network), which is explained by the regression model. This table indicates that some information, especially for strong events, exists even in low dimensional state spaces (see text for details).

	VIS/IR	IR
All events	0.70	0.52
Upper half intensity events	0.81	0.62
Lower half intensity events	0.52	0.40

5 Concluding discussion

A numerical study of an idealized model of thermally driven convection in a layer of fluid was used as the mathematical model of a low-dimensional nonlinear system. A low order observable vector and a system response were postulated as linear functions of portions of the state vector. The ability of the observables to reproduce system response variability was studied with specific focus on the reproduction of response extremes. The main conclusions drawn for the particular system studied are:

1. The presence of a strange attractor, causing a contraction in state space flow, allows for reasonably good reproduction of response variability from observable vectors of lower dimensionality than the underlying system and which are not functions of the state dominating the system response. In our study, a two-dimensional observation vector estimated well the response of the system which was a function of the remaining state.

2. However, very low dimensional observable vectors estimate poorly system response variability. For the system studied, a one-dimensional observation vector (scalar) cannot reproduce extreme variability of system response.

By extending these ideas to the problem of estimating rainfall from satellite imagery, we were able to show that estimating extreme values of response from limited information is rather adequate. This is consistent with the general conclusion that the atmosphere, even though very complex, may exhibit low dimensional attractors (Tsonis, 1996, 2001; Sivakumar, 2004). The existence of these attractors, can therefore aid us, if utilized properly, in studying and estimating the properties of extreme events. Thus, seeking and understanding the properties of these attractors should be encouraged.

Acknowledgements. The work of the second author was sponsored by NSF Award EAR-0125706.

Edited by: B. Sivakumar
Reviewed by: two referees

References

Adler, R. F. and Negri, A. J.: A Satellite Infrared Technique to Estimate Tropical Convective and Stratiform Rainfall, J. Appl. Meteorol., 27, 30–51, 1988.

Arkin, P. A. and Meisner, B.: The Relationship Between Large-Scale Convective Rainfall and Cold Cloud Over the Western Hemisphere During 1982–1984, Monthly Weather Review, 115, 51–74, 1987.

Berge, P., Pomeau, Y., and Vidal, C.: Order Within Chaos, Towards a Deterministic Approach to Turbulence, John Wiley & Sons, New York, 1984.

Georgakakos, K. P. and Tsonis, A. A.: Observing extreme variability in nonlinear systems, in Emergent Nature, edited by Novak, M. M., World Scientific, Singapore, 209–221, 2001.

Kidder, S. Q. and Vonder Haar, T. H.: Satellite Meteorology, An Introduction, Academic Press, San Diego, 1995.

Lehnes, C. A.: On the Feasibility to Estimate Mean Areal Precipitation From Infrared Satellite Images, M.S. thesis, Department of Geosciences, University of Wisconsin-Milwaukee, 1996.

Lorenz, E. N.: Deterministic Nonperiodic Flow, J. Atmos. Sc., 20, 130–141, 1963.

Scofield, R. A. and Oliver, V. J.: A Scheme for Estimating Convective Rainfall from Satellite Imagery, NOAA Technical Memorandum NESS 86, Washington, D.C., 1977.

Sivakumar, B.: Chaos theory in geophysics: past, present and future, Chaos, Solitons and Fractals, 19, 441–462, 2004.

Sugihara, G. and May, R. M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, Nature, 344, 734–741, 1990.

Tsonis, A. A. and Isaac, G. A.: On a New Approach for an Instantaneous Rain Area Delineation in the Midlatitudes Using GOES Data, J. Clim. Appl. Meteorol., 24, 1208–1218, 1985.

Tsonis, A. A., Triantafyllou, G. N., and Georgakakos, K. P.: Hydrological applications of satellite data: 1. Rainfall estimation, J. Geophys. Res., 101, 29 517–26 525, 1996.

Tsonis, A. A.: Chaos: From theory to applications, Plenum, New York, 1992.

Tsonis, A. A.: Dynamical systems as models for physical processes, Complexity, 2, 5, 23–30, 1996.

Tsonis, A. A.: The impact of nonlinear dynamics in atmospheric sciences, Int. J. Bifurcation and Chaos, 11, 881–902, 2001.