Pergamon

S0360-5442(96)00018-7

Energy Vol. 21, No. 7/8, pp. 655–665, 1996

0360-5442/96 $15.00 + 0.00

NONLINEAR DYNAMICS AND THE GREAT SALT LAKE: A PREDICTABLE INDICATOR OF REGIONAL CLIMATE

HENRY D. I. ABARBANEL,†‡ UPMANU LALL,§ YOUNG-IL MOON,§ MICHAEL E. MANN¶ and TAIYE SANGOYOMI‖

‡Department of Physics and Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, Mail Code 0402, La Jolla, CA 92093–0402, U.S.A.; §Utah Water Research Laboratory, Utah State University, Logan, UT 84322–8200, U.S.A.; ¶Department of Geology and Geophysics, Yale University, P.O. Box 208109, New Haven, CT 06520–8109, U.S.A. and ‖Hydrosphere Resource Consultants, 1001 Walnut Street, Suite 200, Boulder, CO 80302, U.S.A.

(Received 15 December 1995)

Abstract—Using methods from nonlinear dynamics, we examine a long climatological record of measurements of the volume of the Great Salt Lake in Utah. These observations, recorded every 15 days since 1847, provide direct insight into the effect of large-scale atmospheric motions in climatological studies. The lake drains nearly 100,000 km^2, and it thus acts as a spatial filter for the finest degrees of freedom for climate. In filtering out a very large number of atmospheric and climatological motions, it reduces its complexity but retains its effectiveness as a climate sensing system. We demonstrate that there are four degrees of freedom active in the Great Salt Lake volume record, that these data reside on a strange attractor of dimension slightly larger than three, and that these data are predictable with a horizon of order a few years. We then show that predictive models based on local properties on the attractor perform remarkably well in reproducing the observations when trained on earlier observations. The ability to predict using earlier observations on the attractor suggests very strongly that over the period of the record, the system has been stationary and that it is a record of the natural variation of the climate. If there is anthropomorphic influence leading to changes in climate, this record suggests it has not made its effect measurable in such large-scale integrating observations. Copyright © 1996 Elsevier Science Ltd.

1. INTRODUCTION

In addition to the relatively well understood impact of the El Nino/Southern Oscillation (ENSO) on global climate,[1,2] there is increasing evidence of organized climatic variability on decadal[3–7] and interdecadal[8–12] timescales. The recognition of such climatic features provides hope for long-term forecasting of droughts and protracted wet periods. Large closed basin lakes provide a natural, regional average of the spatially distributed hydroclimate signal. They are excellent indicators of low frequency climatic variability, and may provide a low dimensional description of climate which can be more readily predictable than local, point observations and useful as a baseline for detecting changes in climate due to human activities. The Great Salt Lake (GSL) of Utah (latitude 40° to 42° N, longitude 112° to 113° W, with a drainage area of 90,000 km^2) is the fourth largest closed lake in the world. It draws on the N. Pacific for its immediate atmospheric catchment. In this article, we present an analysis of the 1847–1993 GSL volume time-series (Fig. 1) from the point of view of nonlinear dynamics using tools which have proven successful in the study of many other systems.[13] We demonstrate the relationship between structured low-frequency climatic variability, low-order dynamical behavior of the GSL, and the enhanced long-term predictability of the GSL volume.

Interannual (2, 3–5 yr) and decadal (10–12, 15–20 yr) signals are found[4,5] from spectral analyses of time series of the GSL volume, Western U.S. precipitation, and from multivariate time-series analysis of northern hemisphere sea level pressure[5] and temperature[5,12] fields. These signals are best described qualitatively as intermittent oscillations at these timescales with slowly varying amplitude and phase characteristics. The GSL volume responds with a small phase lag to regional precipitation and temperature anomalies, which are in turn forced by large-scale atmospheric circulation anomalies. The hemispheric scale and wave pattern of the pressure and temperature patterns that correspond to the high

†Author for correspondence.

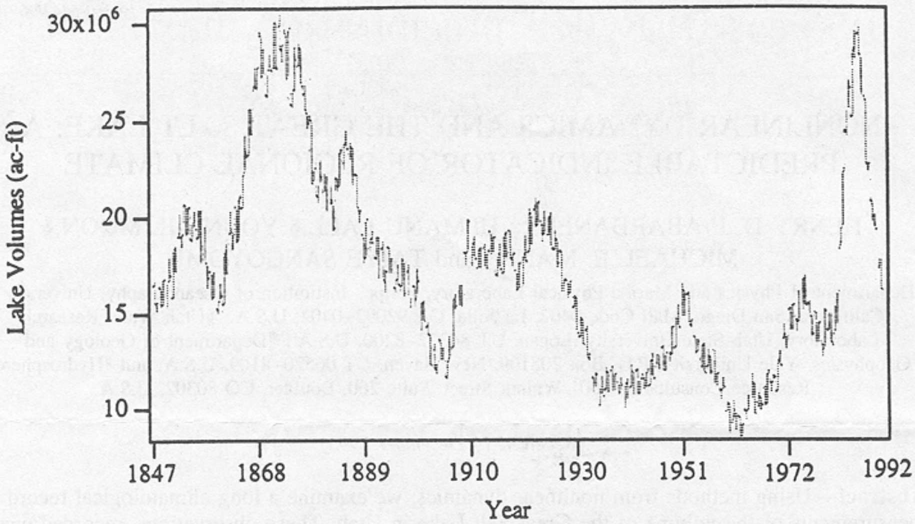


Fig. 1. Great Salt Lake volume time series from 1847–1993 (in units of acre feet). The data of GSL levels, and a relationship between level and volume was provided by the U.S. Geological Survey. Details of the bi-weekly volume time series constructed from this data are in Ref. 14.

western U.S. precipitation phase of the quasi-decadal (10–12 yr) signal (Fig. 2) is consistent with coupled atmosphere-ocean model studies that predict similar circulation anomaly patterns excited on this timescale.

Can the dynamics of continental hydrologic processes that are forced by such signals be described by a few degrees of freedom? The interannual and interdecadal signals identified above account for[4] more than 60% of the over-year variance of the GSL. Investigations[14–16] of the GSL time-series suggest that it is one of a few geophysical series whose dynamics can be described by a low dimensional, nonlinear model with limited predictability. These are summarized below. Remarkable multi-year predictions during droughts and wet periods based on the data-based approximation of the dynamics are then presented.

2. DYNAMICS OF GSL TIME-SERIES

One of the key lessons in the analysis of nonlinear systems is that from a one dimensional time-series such as the GSL volume $v(t) = v(t_0 + n\tau_s) = v(n)$, where t_0 is the time of the initial measurement and τ_s is the sampling time (here $\tau_s = 15$ days), we can reconstruct a multivariate space in which the dynamics unfolds. This is done by creating vectors in dimension d out of the measurements $v(t)$ and its time lags:

$$\begin{aligned} y(n) &= [v(t_0 - n\tau_s), v(t_0 - (n+T)\tau_s), \ldots v(t_0 - (n + (d-1)T)\tau_s)] \\ &= [v(n), v(n-T), \ldots, v(n-(d-1)T)]\,. \end{aligned} \tag{1}$$

We find that *the data can be discussed in an unambiguous fashion as a sequence of multivariate states from a dynamical system with dimension four*. We evaluate the Lyapunov exponents of the GSL volume from these data working in $d = 4$, and learn that (i) the largest exponent is $\lambda_1 \approx (100 \text{ days})^{-1}$, and (ii) the second exponent is zero. The first piece of information tells us that the horizon for accurate predictions is on average a few times a hundred days while the second tells us that differential equations are appropriate for modeling the data.[13]

Using average mutual information[13, 17] we determine that a good time lag (at which the successive values are somewhat independent) to use in this state space reconstruction is $T \approx 8 - 17$. In striking comparison to this characteristic *nonlinear* measure of 'correlation' between measurements, the traditional first zero of the autocorrelation of the GSL volume signal suggests choosing $T \approx 400$. This is a time of nearly 20 years and would obscure the dynamics represented by the interannual and interdecadal signals we consider of interest.

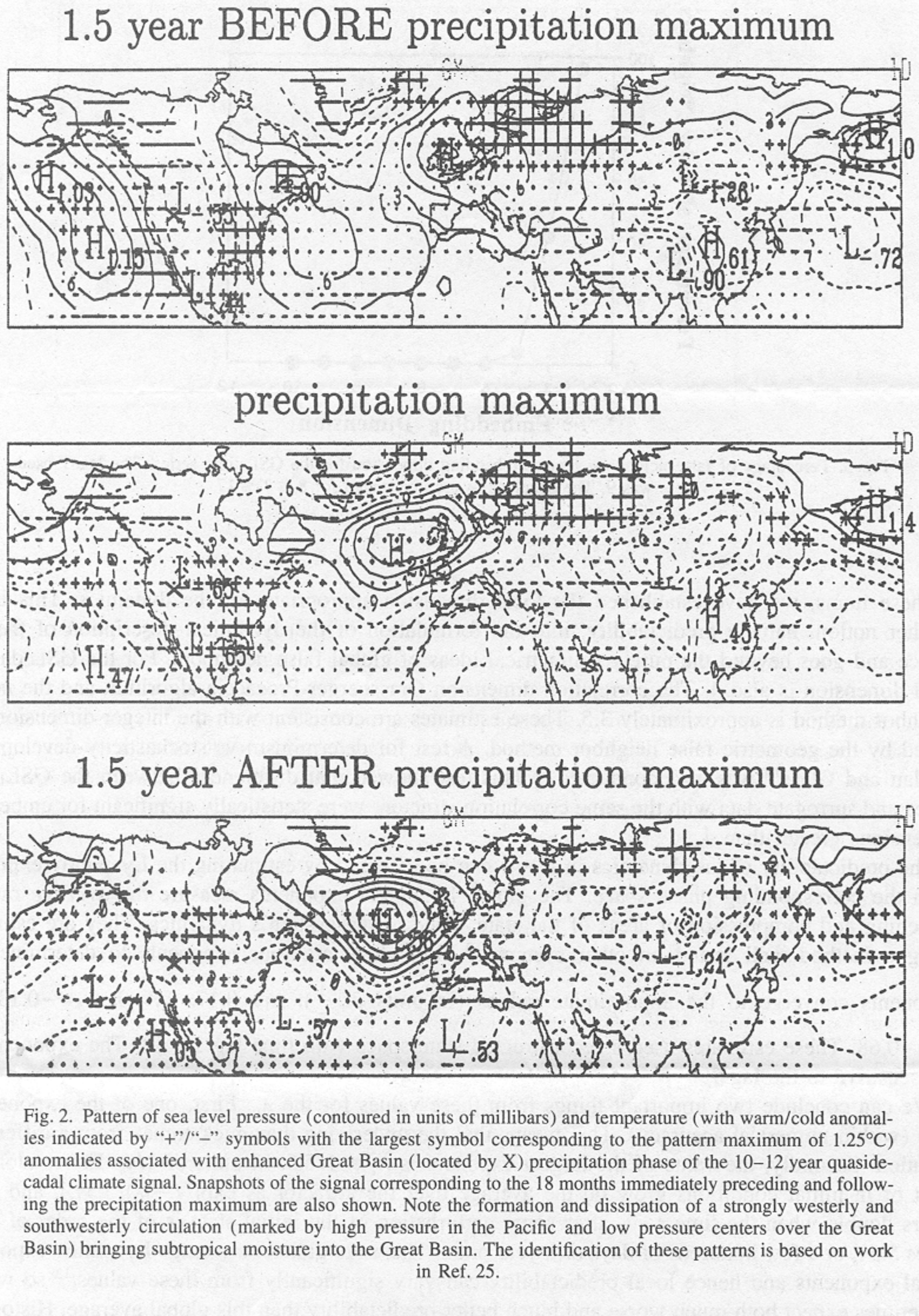


Fig. 2. Pattern of sea level pressure (contoured in units of millibars) and temperature (with warm/cold anomalies indicated by "+"/"–" symbols with the largest symbol corresponding to the pattern maximum of 1.25°C) anomalies associated with enhanced Great Basin (located by X) precipitation phase of the 10–12 year quasidecadal climate signal. Snapshots of the signal corresponding to the 18 months immediately preceding and following the precipitation maximum are also shown. Note the formation and dissipation of a strongly westerly and southwesterly circulation (marked by high pressure in the Pacific and low pressure centers over the Great Basin) bringing subtropical moisture into the Great Basin. The identification of these patterns is based on work in Ref. 25.

To choose the dimension d appropriate for describing the data in multivariate state space, we use the method of global false nearest neighbors.[13] This examines, in dimension d, the nearest neighbor of every vector $y(n)$ as it behaves in dimension $d+1$. If the nearest neighbor moves far away from $y(n)$ as we move up in dimension, then it is declared a false neighbor as it arrived in the neighborhood of $y(n)$ in dimension d by projection from a distant part of the attractor. When the percentage of these false neighbors drops to zero, the attractor is unfolded. To establish the dimension of the dynamics on the attractor, we use the method of local false nearest neighbors.[18] This moves to the neighborhood of every point $y(n)$, computes all distances in the global dimension established by the previous algorithm, and asks in what dimension does the ability to predict one, two, . . . steps ahead become independent of dimension and of the number of neighbors chosen. When this predictability becomes independent

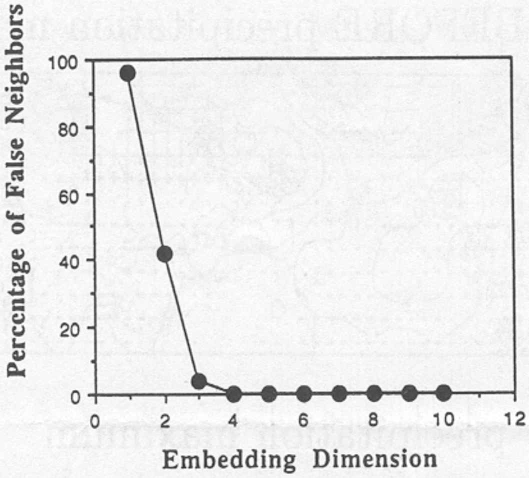


Fig. 3. Percentage of false neighbors versus embedding dimension for the GSL time series. The lag T used here was 9. The conclusions are unchanged for $8 \leq T \leq 17$.

of these items, we have established the local dimension appropriate for the dynamics. This injects another notion, namely predictability, into the formulation of the dynamical description of the data source and goes beyond the purely geometrical ideas of global false neighbors. For the GSL data the local dimension is also 4. The correlation dimension Grassberger-Procaccia algorithm and the nearest neighbor method is approximately 3.5. These estimates are consistent with the integer dimension estimated by the geometric false neighbor method. A test for determinism vs stochasticity developed by Kaplan and Glass[19] was also conducted.[14] This test showed that differences between the GSL time-series and surrogate data with the same correlation structure were statistically significant for embedding dimensions greater than 4.

The predictability of the dynamics in $d = 4$ can be assessed by estimating the Lyapunov exponents from the corresponding phase space. The global Lyapunov exponents measure the *average* rates of stretching and compression of areas of the state space of the system's dynamics. They are estimated using a method well tested on other systems.[13] Here there are four exponents to determine. The exponents converge to the approximate values (in units of $\frac{1}{\tau_s}$): $\lambda_1 = 0.17$, $\lambda_2 = 0$, $\lambda_3 = -0.13$, and $\lambda_4 = -0.68$. These estimates used a reconstructed state space with time lag $T = 12$. The exponents are not sensitive to the lag T.

We can conclude two important things from these values for the λ_a. First, one of the exponents is zero (within numerical accuracy). This means that the underlying dynamics comes from a differential equation. Secondly, the value of the largest exponent is $\lambda_1 \approx 1/5.5$, in units of $1/\tau_s$. Errors along the orbit or in initial conditions grow on the average over the attractor as $\exp[(t - t_0)/5.5\tau_s]$, and initial errors double when the time $t - t_0$ after some perturbation of the initial state is of the order of a few times $5.5\tau_s$ or a few hundreds of days or a year or so. These are global or average Lyapunov exponents. Local exponents and hence local predictability can vary significantly from these values,[13] so we can sometimes expect both much worse and much better predictability than this global average. Histograms of estimated local Lyapunov exponents[13] based on iterating the time-series up to 2.6 and 5.2 years are presented in Fig. 4. We see that for forecasts nearly two years ahead the spread in predictability is quite large. There are states of the system where the largest Lyapunov exponent associated with predicting ahead two or so years is near zero as well as states where the exponent is more than double the global average. Small variations in the Lyapunov exponents imply sizable variations in the prediction horizon. For a dynamical systems, the spread in the estimated local Lyapunov exponents about their mean decreases as the number of iterates forward is increased. This is seen in Fig. 4. Local Lyapunov exponents computed for data from a stochastic process will not do this.

We can estimate the Lyapunov dimension[20] from these exponents, and we find it to be about 3.05. This dimension estimate is consistent with the numbers reported earlier for other fractal dimensions.

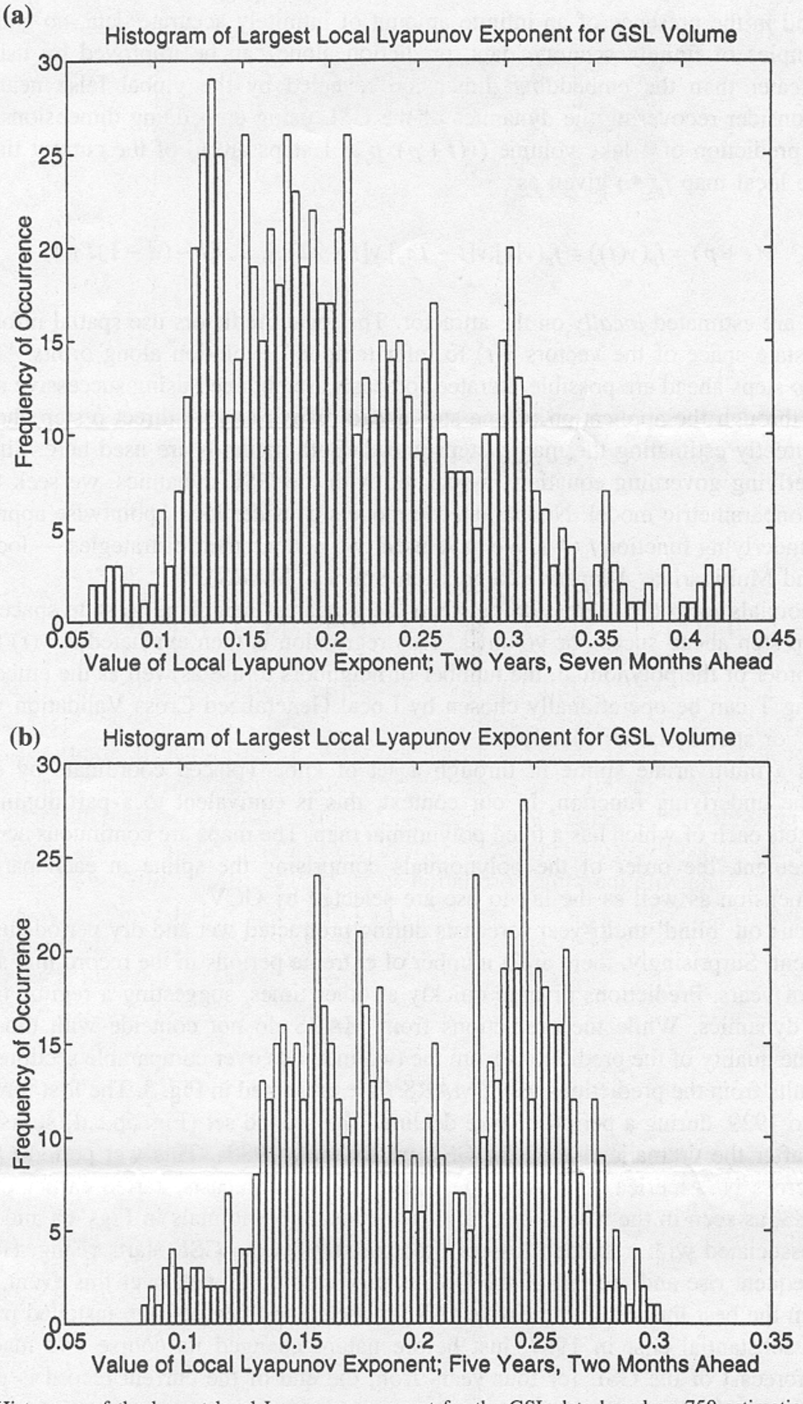


Fig. 4. Histogram of the largest local Lyapunov exponent for the GSL data based on 750 estimation points, based on iterating forward (a) two years and seven months and (b) five years and two months.

3. PREDICTION

The active dimension of the dynamics has been estimated as four, and the feasibility of modeling it through a set of ordinary differential equations has been established. One can now develop a nonlinear model for forecasting the lake volume. The geometrical embedding theorem[13] suggests that in $d = 4$ we have unfolded the attractor and no orbit crossings which would impede prediction remain. There are always biases in estimating the local prediction function we shall construct in a moment. We have found that using dimensions larger than $d = 4$ can improve predictions, while using dimensions less than four degrades them substantially. In principle, of course, there is no harm in going to larger dimensions, and in the presence of an infinite amount of infinitely accurate data, no gain either. With real, finite samples of finitely accurate data, prediction alone can be improved by using models of dimensions greater than the embedding dimension revealed by the global false nearest neighbors method. We consider recovering the dynamics of the GSL using embedding dimensions between four and nine. For prediction of a lake volume ($v(t+p)$; $p>1$ steps ahead of the current time t, we need to estimate the local map $f_p(\bullet)$ given as

$$v(t+p)=f_p(y(t))=f_p(v[t],v[t-T\tau_s],v[t-2T\tau_s],\ldots v[t-(d-1)T\tau_s])\ .$$

These maps are estimated *locally* on the attractor. The local predictors use spatial information in the reconstructed state space of the vectors $y(t)$ to infer temporal evolution along orbits.[13] Two methods for predicting p steps ahead are possible. Iterated forecasts are obtained using successive one step ahead forecasts, i.e., through the application of one step ahead maps $f_1(\bullet)$. A direct p step ahead forecast is available by directly estimating the map p steps ahead. Both methods are used here. Since we do not know the underlying governing equations associated with the GSL dynamics, we seek to recover the map using a nonparametric model. Nonparametric models consider local, pointwise approximations of the unknown underlying function $f_p(\bullet)$. We have used two nonparametric strategies — local polynomial regression[21] and Multivariate Adapative Regression Splines (MARS).[22]

Local polynomials form the regression between the k nearest neighbors in state space of the vector $y(t)$ and their p step ahead successor volumes. This regression is then evaluated at $y(t)$ to provide the forecast. The order of the polynomial, the number of neighbors to use as well as the embedding dimension and the lag T can be operationally chosen by Local Generalized Cross Validation with Leverage (LGCVLEV)[23] or specified heuristically.

MARS uses a multivariate spline fit through a set of knots (placed coordinate by coordinate) to approximate the underlying function. In our context, this is equivalent to a partitioning of the state space into subsets each of which has a fixed polynomial map. The maps are continuous across partitions. The knot placement, the order of the polynomials comprising the spline in each partition and the embedding dimension as well as the lag to use are selected by GCV.

We shall focus on 'blind' multi-year forecasts during protracted wet and dry periods using only data prior to the event. Surprisingly, there are a number of extreme periods in the record that are predictable for a number of years. Predictions diverge quickly at other times, suggesting a regime transition, or a switch in the dynamics. While the predictions from MARS do not coincide with those using local polynomials, the quality of the predictions from the two methods over comparable sections of the record is similar. Results from the predictions using MARS[24] are presented in Fig. 5. The first forecast (Fig. 5a) is from 1925 to 1929, during a period of lake decline. The second set (Fig. 5b,c,d) starts in Oct. 1983, nearly a year after the dramatic rise of the GSL in the early 1980s. This wet period (1982–87) was experienced across N. America (including the Great Lakes). Forecasts started earlier than Oct. 1982 diverge by 1983, as seen in the results of applying the local polynomials in Figs. 6a and b. The 1982–83 period is associated with a sudden loss of predictability as the GSL starts rising. However, in all cases the subsequent rise and the fall of the GSL is predicted quite well over this event, in contrast to the results from the best linear autoregressive (AR) model. The State of Utah installed pumps to lower the GSL at a substantial cost in 1987, just before nature changed its course and made the pumps redundant. A forecast of the GSL for four years from the end of the current record is presented with estimated confidence limits in Fig. 6c.

4. CONCLUSIONS

Recently developed tools for analysing nonlinear dynamics appear quite useful when used with geophysical time-series. There have been a number of other applications of similar ideas to describe the behavior of geophysical time series and also for short term predictions. Results have been mixed, and the issue of whether the time-series represents deterministic chaos or stochastic behavior has been debated. We expect that such discriminations cannot be done unambiguously given a finite record. It is more useful to focus on how these techniques improve scientific understanding and can be used operationally.

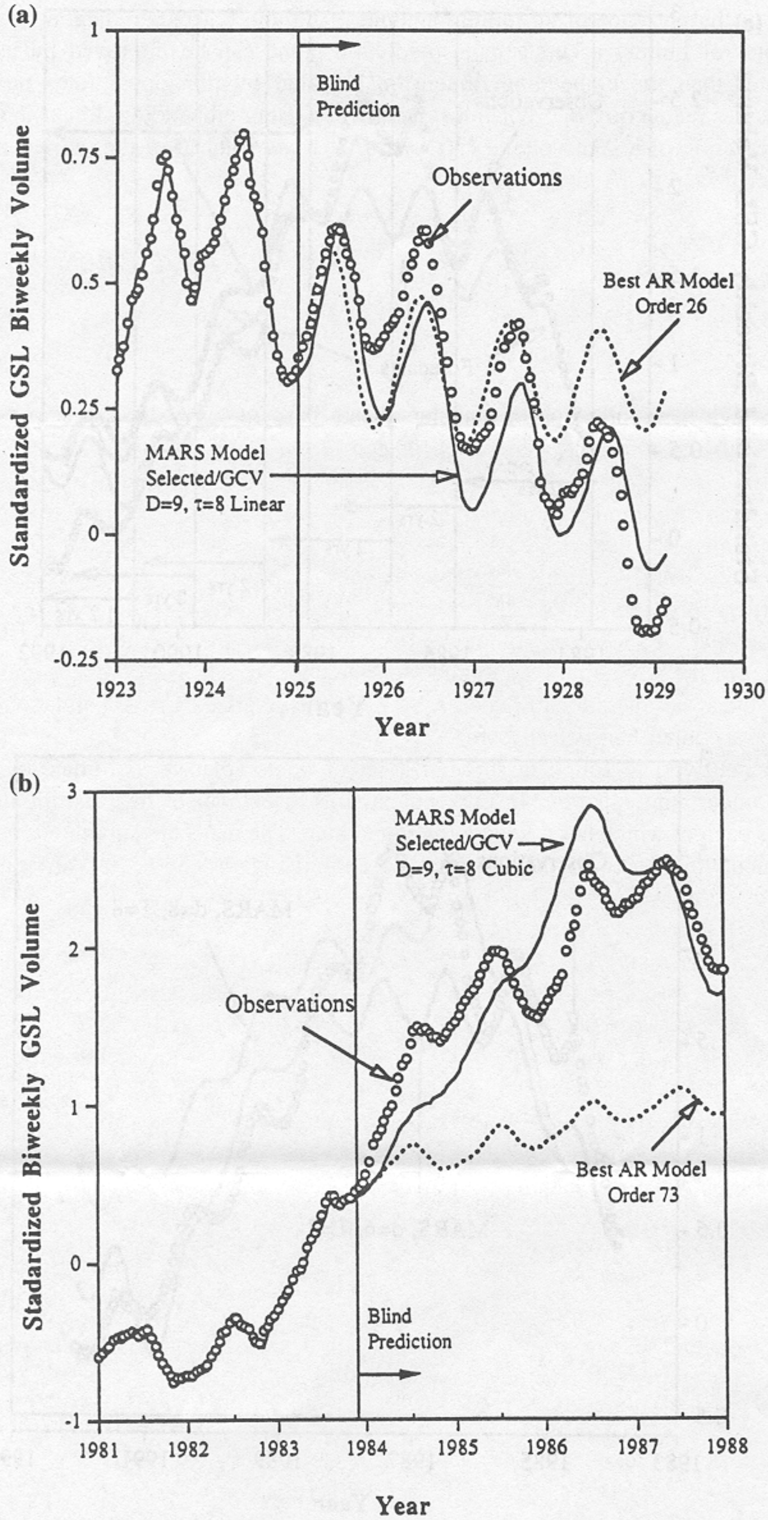
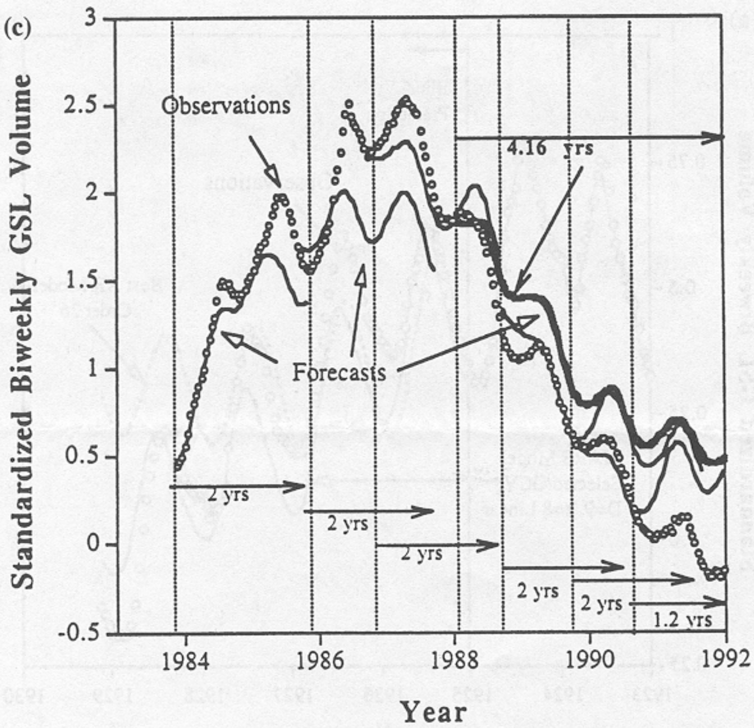


Fig. 5. Multi-year, bi-weekly iterated forecasts using MARS from different points in the time-series. The model is calibrated on data prior to the first forecast date, and forecasts during a given horizon use only the same data — no new data is used. The embedding dimension d and the lag T, are chosen using GCV for each forecast. The GSL data is scaled to have 0 mean and standard deviation 1, using a mean volume of 17.1×10^6 acre-feet and a standard deviation of 5.0×10^6 acre-feet. Figures 5(a) to (c) show results based on parameters chosen using GCV. The forecasts shown in Fig. 5(d) are not based on the optimal parameters selected by GCV. They are included because it is remarkable that some reasonable but not optimal parameter choices provide good, eight year blind forecasts.



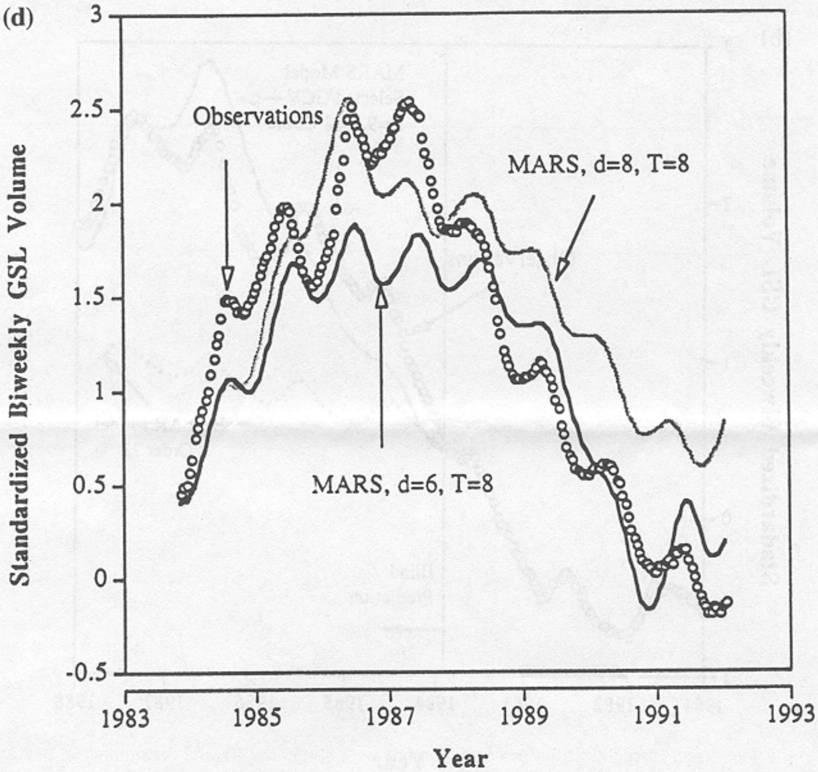


Fig. 5. Continued

The GSL record appears to represent low dimensional dynamics, have predictability that is state dependent, and in certain situations has a long predictability horizon. These situations appear to correspond to what are generally considered extreme events, and are thus of considerable societal interest. While the onset of such an event appears to be unpredictable, the behavior during the event may be

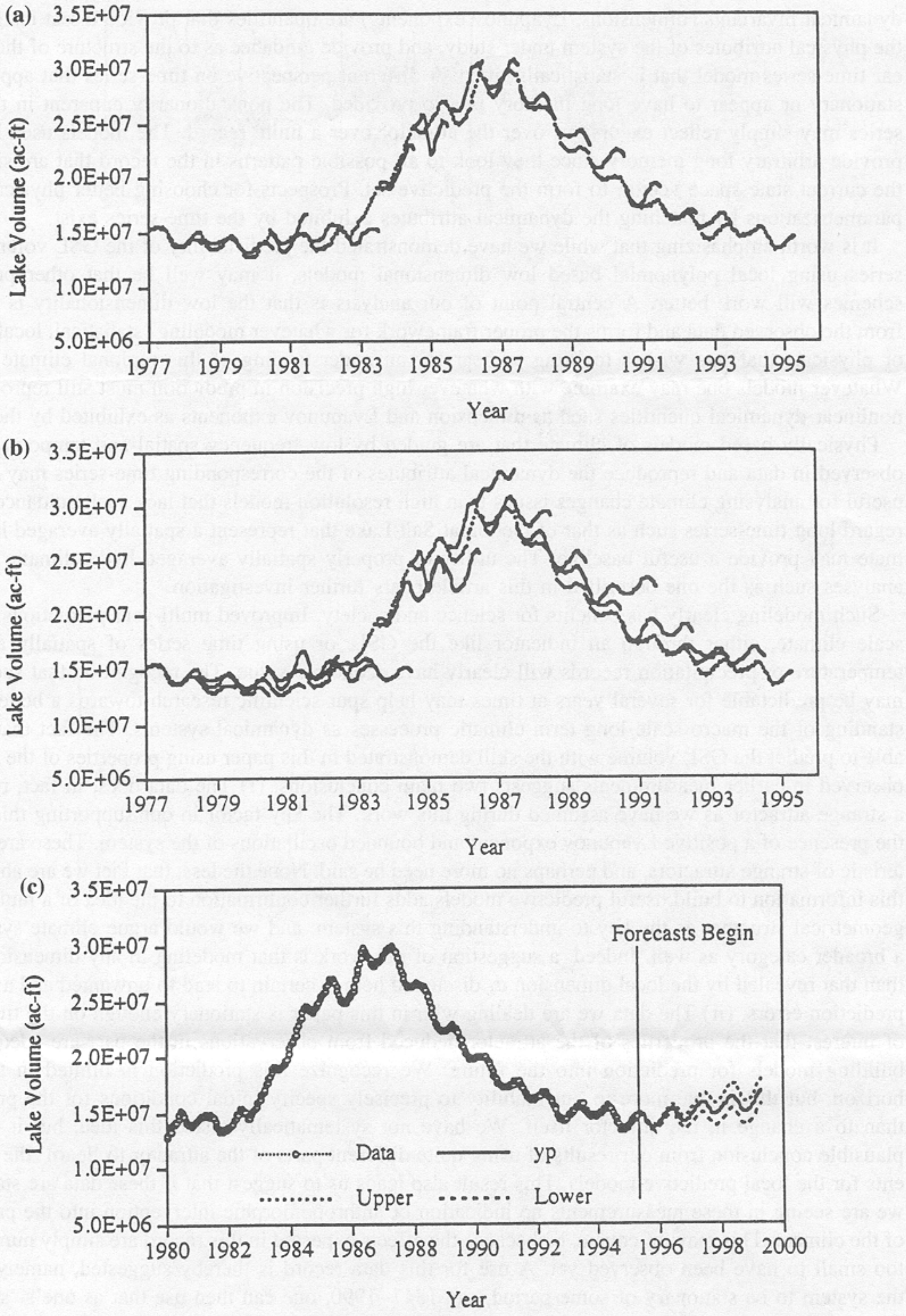


Fig. 6. Multi-year direct forecasts using local polynomials with d = 5, T = 10, and the number of nearest neighbors, and polynomial order chosen by LGCVLEV, at each step of the forecast. (a) one year blind forecasts from August 1977 to July 1995. (b) two year blind forecasts from August 1977 to July 1995. (c) four year blind forecast with approximate 95% confidence limits from August 1995.

predictable. Better methods to predict the likelihood of a regime transition or loss of predictability need to be developed.

Is this just a fancier statistical time-series model? The methods used to recover the map $f_p(\bullet)$ are indeed firmly based on statistical theory. The general structure of the models based on the nonlinear dynamics is somewhat different than the Box-Jenkins models traditionally used by statisticians. The dynamical invariants (dimensions, Lyapunov exponents) are quantities that provide direct insights into the physical attributes of the system under study, and provide guidance as to the structure of the nonlinear time-series model that is statistically fitted. A different perspective on time-series that appear nonstationary or appear to have long memory is also provided. The nonstationarity apparent in the time-series may simply reflect excursions over the attractor over a finite record. The models used here can provide arbitrary long memory since they look to all possible patterns in the record that are similar to the current state space vector, to form the predictive set. Prospects for choosing better physical model parametrizations by matching the dynamical attributes exhibited by the time-series exist.

It is worth emphasizing that while we have demonstrated the predictability of the GSL volume time-series using local polynomial based low dimensional models, it may well be that other prediction schemes will work better. A central point of our analysis is that the low dimensionality is deduced from the observed data and forms the proper framework for whatever modeling (statistical, local, global, or physical) that one wishes to bring to bear for an understanding of this regional climate system. Whatever models one may examine with whatever high precision in prediction must still reproduce the nonlinear dynamical quantities such as dimension and Lyapunov exponents as exhibited by these data.

Physically based models of climate that are guided by low frequency spatial and temporal features observed in data and reproduce the dynamical attributes of the corresponding time-series may be more useful for analysing climate changes issues than high resolution models that lack such guidance. In this regard long time-series such as that of the Great Salt Lake that represent a spatially averaged hydroclimate may provide a useful baseline. The utility of properly spatially averaged hydroclimatic data for analyses such as the one described in this article bears further investigation.

Such modeling clearly has benefits for science and society. Improved multi-year predictions of large scale climate, either through an indicator like the GSL, or using time series of spatially averaged temperature or precipitation records will clearly have economic value. The recognition that such fields may be predictable for several years at times may help spur scientific research towards a better understanding of the macro scale long term climatic processes as dynamical systems. The fact that we are able to predict the GSL volume with the skill demonstrated in this paper using properties of the attractor observed in earlier measurements suggests two main conclusions: (i) The data does, in fact, reside on a strange attractor as we have assumed during this work. The key factor in our supporting this idea is the presence of a positive Lyapunov exponent and bounded oscillations of the system. These are characteristic of strange attractors, and perhaps no more need be said. None the less, that fact we are able to use this information to build useful predictive models adds further confirmation to the idea of a multivariate geometrical structure as the key to understanding this system, and we would argue climate systems in a broader category as well. Indeed, a suggestion of this work is that modeling in any dimension lower than that revealed by the local dimension d_L discussed here is certain to lead to unwanted and avoidable prediction errors. (ii) The data we are dealing with in this paper is stationary enough on the timescales of interest that the properties of the attractor deduced from observations in the past are adequate for building models for prediction into the future. We recognize this prediction is limited in temporal horizon, but this is due more to our inability to precisely specify initial conditions for the prediction than to a change in the attractor itself. We have not systematically tested this idea, but it seems a plausible conclusion from our results of using quite different parts of the attractor to 'learn' the ingredients for the local predictive models. This result also leads us to suggest that if these data are stationary, we are seeing in these measurements no indication of anthropomorphic intervention into the properties of the climate. This may, of course, be because the effects expected in this record are simply numerically too small to have been observed yet. A use for this data record is thereby suggested, namely, taking the system to be stationary of some period, say 1847–1990, one can then use that as one's 'standard' for natural variation of the climate and seek to observe deviations of future climate readings by monitoring how it behaves compared to the expectations established by the standard. Direct comparison of time-series may actually not be the way to perform this monitoring. Rather comparison of the geometry of the attractor would be the goal, as the attractor is static when a system is stationary even though every orbit visits it in a different way because of the presence of positive Lyapunov exponents.

REFERENCES

1. M. S. Halpert and C. F. Ropelewski, *J. Clim.* **5**, 577 (1992).
2. T. P. Barnett, *J. Clim.* **4**, 269 (1991).

3. C. Deser and M. Blackmon, *J. Clim.* **6**, 1743 (1993).
4. U. Lall, and M. E. Mann, "The Great Salt Lake: a Barometer of Low Frequency Climatic Variability", *Water Resources Research*, in press (1995).
5. M. E. Mann, U. Lall, and B. Saltzman, *Geophysical Research Letters* **22**, 937 (1995).
6. M. R. Allen and L. A. Smith, *Geophys. Res. Lett.* **21**, 883 (1994).
7. R. Houghton and Y. Tourre, *J. Clim.* **5**, 765 (1992).
8. C. K. Folland, D. E. Parker, and F. E. Kates, *Nature* **310**, 670 (1984).
9. K. E. Trenberth, *Bull. Am. Met. Soc.* **71**, 988 (1990).
10. M. Ghil and R. Vautard, *Nature* **350**, 324 (1991).
11. M. E. Mann and J. Park, *Geophys. Res. Lett.* **20**, 1055 (1993).
12. M. E. Mann and J. Park, *J. Geophys. Res.* **99**, 25819 (1994).
13. H. D. I. Abarbanel, R. Brown, J. J. ("Sid") Sidorowich, and Lev Sh. Tsimring, *Rev. of Mod. Phys.* **65**, 1331 (1993); see H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer-Verlag, Heidelberg (1996), for more details on the methods used.
14. T. B. Sangoyomi, "Climatic Variability and Dynamics of Great Salt Lake Hydrology", PhD Dissertation, Department of Civil and Environmental Engineering, Utah State University, Salt Lake City (1993).
15. H. D. I. Abarbanel and U. Lall, *Climate Dynamics*, in press (1995).
16. T. B. Sangoyomi, U. Lall and H. D. I. Abarbanel, *Water Resources Research*, in press (1995).
17. A. M. Fraser and H. L. Swinney, *Phys. Rev. A* **33**, 1134 (1986).
18. H. D. I. Abarbanel and M. B. Kennel, *Phys. Rev. E* **47**, 3057 (1993).
19. D. Kaplan and L. Glass, *Phys. Rev. Lett.* **68**, 427 (1992).
20. J. L. Kaplan and J. A. Yorke, *Lecture Notes in Mathematics* **730**, 228 (1979).
21. U. Lall, Y. Moon and K. Bosworth, "Locally Weighted Polynomial Regression: Parameter Choice and Application to Forecasts of the Great Salt Lake", UWRL Working Paper UWRL-WP/95-016-UL, Utah State University (1995).
22. J. H. Friedman, *Annals of Statistics* **19**, 1 (1991).
23. A. C. Craven and G. Wahba, *Numerische Mathematik* **31**, 377 (1979).
24. U. Lall, T. B. Sangoyomi and H. D. I. Abarbanel, *Water Resources Research*, in press (1995).
25. M. E. Mann and J. Park, *Journal of Climate*, in press (1996).