A learning method for exploring noisy chaotic dynamics illustrated with solar weather

Chaotic systems such as earth's weather or solar weather provide significant challenge for longer range prediction. Delay maps, sequences of measurements with specific relationships between variables and time delays, provide an empirical method of reconstructing dynamics, hence can be used as a basis for empirical prediction. However most forms of error that accompany measurement of these systems make the target variable measurement unidentifiable, making prediction problematic. Here we use multiple delay maps to construct multiple sequences of predictions which can be used to construct predictive probability densities across time. Such predictive densities can be created from a training sample to construct likelihoods of a test sample which can be used for both testing of variable sets for better predictive learning, and to test for learning across time by weighting each sequence of predictions by how close it is to the test data as it appears. The approach will be examined both with artificial chaotic noisy data and with real data on Ap, sunspots, and F10.7 solar measurements. The existence of learning in the solar data is a necessary condition for the implementation of a spreadsheet structure where different strategies could be evaluated against an unfolding sequence of events through time, for example providing a way to evaluate different paths of inner solar system space craft against postulated time sequences of solar storms.

Items in red show command line code in R to produce results. Code is raw r code developed through a sequence of problems.

A statistical learning method of identifying predictive association in noisy chaotic systems is proposed here. The method sets up a tapestry of predictions stretching from 1 to k time units ahead. The approach is to construct density estimates for the variable being predicted using randomly selected delay maps, for each set of predictors, a given initial delay map of predictor data can be used to construct a trajectory of predictions for 1 up to k time units ahead. Then as observations appear over time, the predicted likelihood given the density can be calculated. As time proceeds the accuracy of each trajectory can be reweighted by closeness of that trajectory to actual points and the density can be recalculated according to that weighting. So we can have an unweighted density 4 seasons ahead constructed by development of a model over a training sample. Then when an observation appears in the next season we can reweight the 4 season ahead density by the closeness to the observation and get a new likelihood for it, when it is now three seasons ahead, and so forth. There is learning occurring if the reweighted density results consistently in a higher likelihood.

In this paper, this approach is applied to an artificial data set constructed from a seasonalized version of the Lorenz attractor, and a solar measurement data set [1] that has been reduced to seasonal averages. The solar data set includes AP, sunspot number and F10.7 measurements and here we use data from 1963-2012. The theoretical framework being proposed is spelled out below along with the computational methodology.

The work of Lalley [2] pointed out that it is not possible to estimate even the path of a chaotic system if measurement error is unbounded (even gaussian). In systems without noise with sufficiently large embedding dimension [3], the set of diffeomorphic embeddings is prevalent [3,4], so local linear regression [5] provides approximate linear prediction for each view within the range of the Lyapunov coefficient. Based on Lalley's results this also hold for closed chaotic systems with appropriately bounded [2] measurement error. Consider the predictions constructed from different diffeomorphisms[6,7]. With no measurement error and sufficiently short lag to the prediction time the predictions are nearly identical. With appropriately bounded measurement error, the average prediction based on Lally's filter estimate would converge to the same result. With unbounded error, we consider nearest neighbor linear regression estimates [5]. With unbounded measurement error, assuming an appropriate generalization of the chaoticity hypothesis [8,9], we conjecture that the first theorem of Ruelle [10] can be extended to cover sub delay maps of larger delay maps as projections, so Neumanns [11] result can be applied to show convergence to the mixture of densities that occur at that point. In particular using h and g to denote trajectories from which the observation (error + trajectory) intersect at that point, the probability that trajectory h is responsible for the observed delay map being used to create the neighborhood for regression asymptotically is $\pi_{h(x)} = \frac{f_h(x)}{\sum_{f_g} f_g(x)}$ as the number of observations in the shrinking neighborhood increases with increasing sample sizes. Weighting the algorithmically produced trajectories by closeness to the new observations will as time proceeds gradually accumulate higher weight on the closer trajectories, which will result in subsequently improved density estimates and improved predictions. Since this depends on a chain of asymptotic assumptions with unknown convergence rates, we will check for the property experimentally in this paper. If it is correct the likelihood should increase in each chain of measurements along a trajectory, where for a given season there are 4 chains, a 4 season ahead chain of 4 predictions a 3 season ahead chain of 3 predictions and so on. For example predicting in fall, we should have a 4 season ahead prediction of fall, but each thread of the 4 season ahead tapestry for fall includes a 3 season ahead prediction of summer, a two season ahead prediction of spring, and a 1 season ahead prediction of winter. Once winter is observed, we reweight the thread for fall by closesness of its winter prediction to actual winter, and similarly for spring and summer. We will test whether learning occurs (at all) in a sequence by testing whether any of the reweighted 4 season ahead predictions improve on the original 4 season ahead prediction, and similarly with 3 and 2 seasons ahead.

If for example this works for a particular measure, then the tapestry can be the basis for a spreadsheet structure looking forward in time. So by postulating a high or low response in the next time period, we can see how the probability of high or low response in the following time period changes and evaluate how to modify infrastructure responses to better meet the long term trajectory of the response. Observing what variable sets/delay maps get emphasized by reweightings also allows more insights into the dynamics of the system and how different variables reweight.

.METHOD OVERVIEW: For each predicted season for each variable we can estimate an initial predictive likelihood, by just estimating the density as the derivative of the weighted empirical distribution function of predictions in a small neighborhood of the observed value in the test season. As each season appears, we can reweight the threads for following seasons. The original data is scaled to have mean 0 and variance 1. Here a gaussian function centered at the new observation with unit variance is used for reweighting the threads as the data comes in. Inference was done using paired t tests on the log predictive likelihoods in sequence through time, using the auto covariance between the log likelihood differences to account for any cross-year correlation (likelihoods calculated by predicted season), and multiple tapestry creation to account for the variation between random sampling of delay maps.

The computational procedure is to choose a variable to predict (dependent variable) and a set of variables to predict with (independent variables) . The variables to predict with are always lagged behind the variable to predict so prior observations of the dependent variable is typically included in the set of "independent" variables. A dimension is chosen using the Levina-Bickel [12] algorithm for identifying the dimension of a manifold and the delay map dimension is chosen as 2 times the estimated dimension (as an estimate of the size of the delaymap necessary for it to produce a diffeomorphic mapping of the attractor[3]). For this paper we used approximately 600 random delay maps in each run for 2 runs.

The delay maps are selected and then they are used to build predictive local linear regressions for each of the k terms being selected (in this case for 1 to 4 seasons ahead) where the actual regression is chosen using a LARS approach minimizing Cp [13]. 600 Predictions are constructed, then a random residual from the regressions based on a smaller neighborhood in the (sqrt of the window size) of the regressions is added back to each prediction (remember we are predicting the distribution of the observations). The density is constructed a neighborhood of the 600 of these pseudo observations, currently constructing an empirical weighted CDF, and calculating a regression based on the a small neighborhood of the new observation. If the new observation falls outside the range of the predictions, a Laplace distribution is fit to the residuals(between Gaussian and Laplace, Laplace seems a better fit to the residuals) and the density is scaled appropriately by its distance from where the density is estimated. We will examine two data sets, one an artificial data set constructed using the Lorenz attractor, and a real data set constructed from measurements of the solar weather

For these examples,real data was available for each data set from 1963 though 2012. The training data was through 1993, the test data (for which "predictive pseudo likelihoods were constructed" was from 1994-2012. The likelihood in this case is constructed separately for predictions for each season and the sum becomes the "pseudo predictive likelihood". Because the correlations between seasons a year apart are not taken into account these sums can not be considered full predictive likelihoods. To test for differences, between different weightings as new observations accumulate, the usual autocorrelation function based estimate of the sum of pseudo likelihood differences can be used to construct a t like statistic with a conservative set of degrees of freedom. Because of the many asymptotic approximations being used in this, extra conservativeness is being used to identify statistically significant differences. A p value is constructed from the t statistic, and for a given comparison, the false discovery rate [14] assuming correlated tests is applied to all the tests in each sequence of tests.

The seasonalized Lorenz model is calculated by taking the usual calculation, from time 0 to 50 in .01 time increments, each column is then standardized (subtract mean divide by standard deviation). Then the third variable is duplicated in a 4^{th} column and 1^{st} and 2^{nd} columns then had independent normal(0,.1) errors added to it while the third and fourth had independent sequences of normal(0, 1) errors added to them. Then the data was seasonalized by adding the vector 1,2,1,0 repeated 1250 times (with a 1 to cap the end) to each column. Finally every 13^{th} data point was taken until 198 times were included to create a chaotic data set with seasonality some separation and random error, and the same number of data points as our real data set spanning seasons from 1963 through 2012.

To display the results, we first show the raw pseudo likelihoods, with vertical lines separating each season, and a number representing how many seasons ahead the unweighted prediction is being made and then in a second plot we display the differences, with those that pass the false discovery rate criterion (with Q=.1) having 3 red astericks plotted over the number.

In particular in figure 1 where the modified Lorenz attractor column 3 is being predicted using columns 3, and 4 , the 1^{st} season is fall, and the numbers 4,4,4,4,first appear in sequence. The first 4 is the likelihood predicting 4 seasons ahead without reweighting. The second is that after reweighting with the next observed winter, the 3^{rd} with the next observed winter and spring, and the final with reweightings from winter spring and summer. We can see there is no regular rising along each numbered sequence, in the second plot the likelihood difference (eg subtracting the first unweighted prediction from weighted prediction in each sequence we see a random walk around 0 with very little departure from 0.

In contrast in figure 2 column 1 is being predicted using columns 1 and 2 and we see that that many of the subsequent predictions in each sequence are rising. In particular in part b of the figure we see 9 of the 24 differences meet the .1 false discovery rate criterion we have set to indicate statistically interesting results.

```
liklrnlorenz.12.34<-
superloopSP(n=2,modvec=c(1,2),y=1,cvar=1,odat1=lorenz5,odat2=lorenz7)
lik.sn.summary2new(c("liklrn.lorenz12.34"),"Lorenz high noise",F,F)
```

Figure 1a: Lorenz high noise likelihood

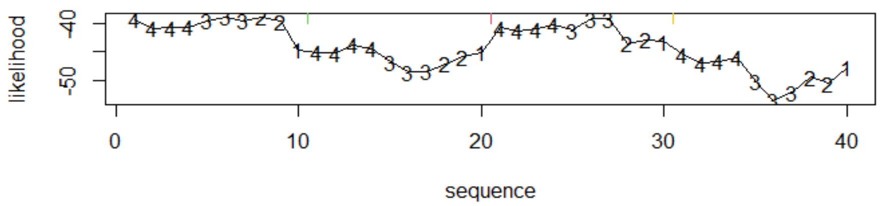
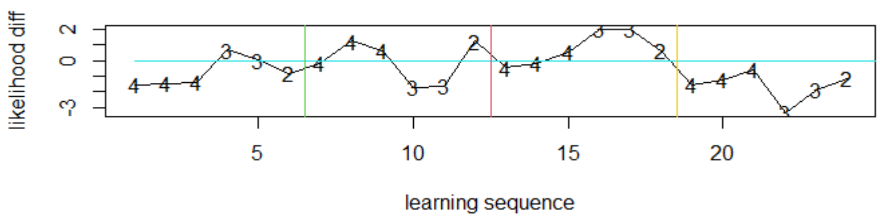


Figure 1b: Lorenz high noise likelihood differences



Legend Figures 1 A and B: Here we test to see if there is any learning going on in a high noise situation predicting 4, 3, 2, and 1 season ahead for each season, fall, winter, spring, and summer. Although there are a few seasons where some rise over the initial likelihood occurs there is none where they exceed the false discovery rate criteria set for an interesting difference.

```
lik.sn.summary2new(c("liklrn.lorenz12.34"),"Lorenz low noise",F,T)
```

Figure 2a: Lorenz low noise likelihood

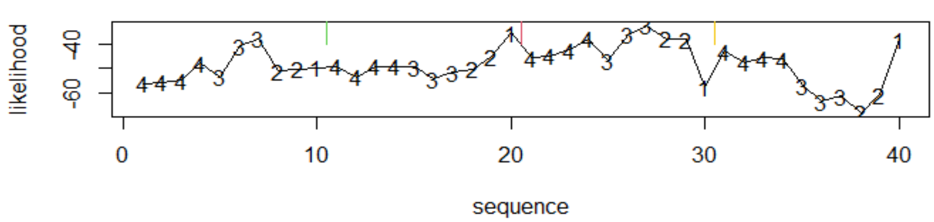
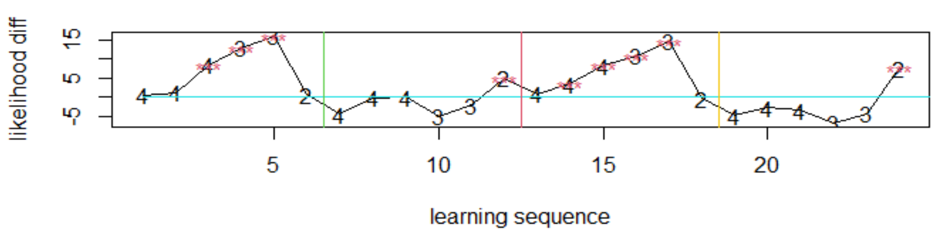


Figure 2b: Lorenz low noise likelihood differences



Legend figure 2 a and b: With columns 1 and 2 being used to predict the future of column 1 where there was low noise we see that the reweighting does seem to be working and allowing some learning to occur hence enabling a framework for spreadsheet like future prediction of the system.

Below we show the application of the method to prediction of the seasonal average AP, Sunspot number, and F10.7 index. For modeling we build our delay maps from all three solar indexes, seasonal average AP index, Sunspot number, and F10.7, with the one being predicted, with lags respectively of 1 through 4 seasons.

AP.calc<-superloopSP(n=2,modvec=c(7,8,9),y=1,cvar=c(1:3),odat1=Fsp.nat0full0plus,odat2=Fsp.nat0full0plus)

lik.sn.summary2new(c("Ap.calc"),"AP ",F,T)

Figure 3a: Ap likelihood

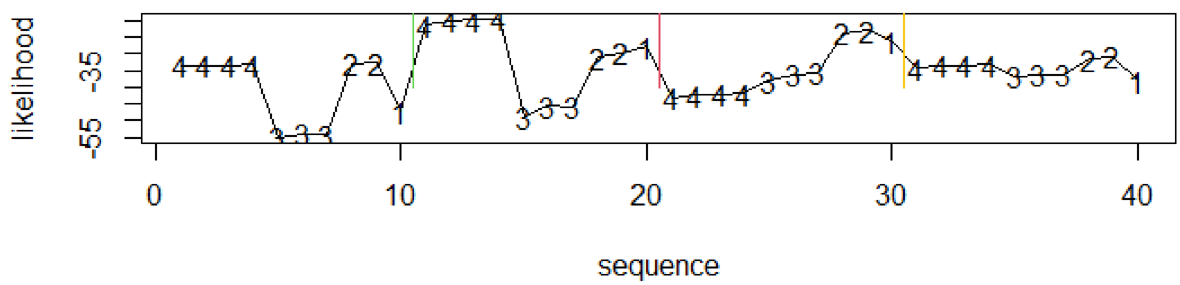
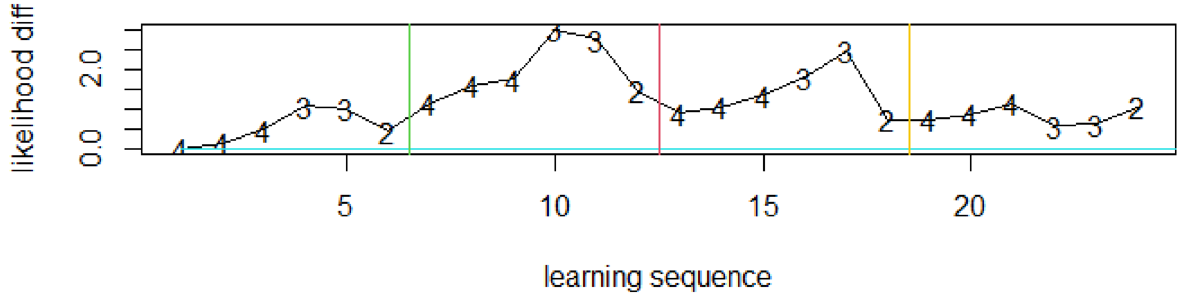


Figure 3b: Ap likelihood differences



Legend Figure 3a and b: For the measurement of the AP, in the raw likelihood data we see some fairly large jumps occurring as we go from 4 to 3 seasons ahead in fall which corresponds to at least one of the observations actual sitting outside the range of 600 predictions in at least one of the replications. We see from the likelihood differences that there is apparent learning occurring for each prediction but none of the points exhibit statistically interesting learning.

SP.calc<-superloopSP(n=2,modvec=c(8,7,9),y=1,cvar=c(1:3),odat1=Fsp.nat0full0plus,odat2=Fsp.nat0full0plus)

lik.sn.summary2new(c("Sp.calc"),"Sunspot ",F,T)

Figure 4a: sunspot likelihood

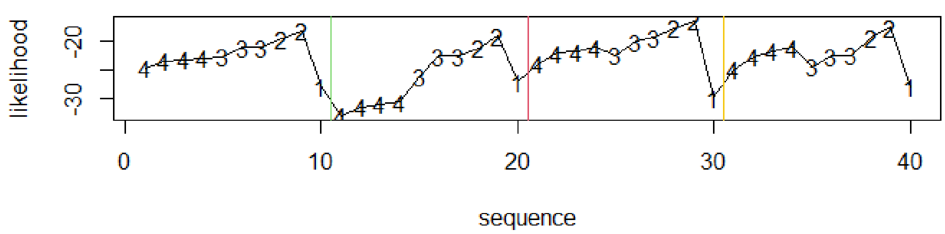
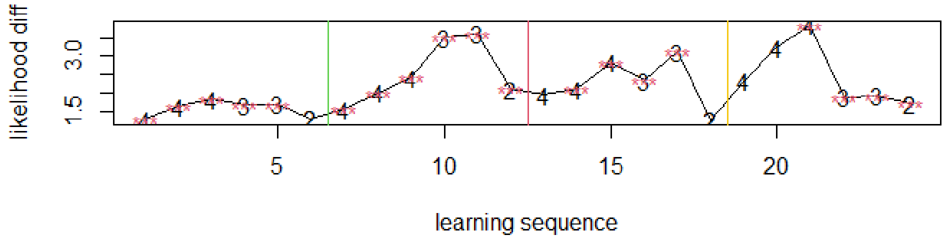


Figure 4b: sunspot likelihood differences



Legend Figure 4a and b: For the measurement of the sunspot number, none of the jumps occurred in the raw data we saw the Ap data. Twenty of the twenty four differences showed statistically interesting learning. Another interesting aspect of this is the drop when we go to 1 season ahead prediction.

```
F10.7.calc<-superloopSP(n=2,modvec=c(9,7,8),y=1,cvar=c(1:3),odat1=Fsp.nat0full0plus,odat2=Fsp.nat0full0plus)
lik.sn.summary2new(c("F10.7.calc"),"F10.7",F,T)
```

Figure 5a: F10.7 likelihood

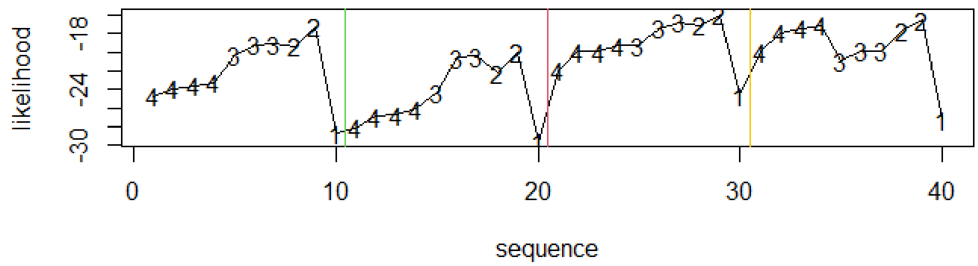
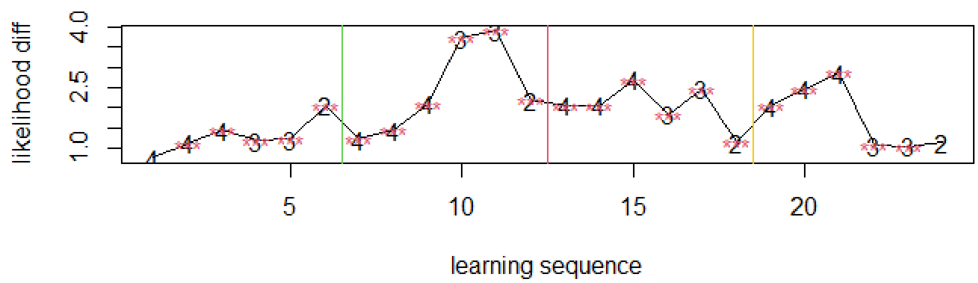


Figure 5b: F10.7 likelihood differences



Legend Figure 5a and b: For the measurement of the F10.7 index, again there are no jumps, in this case 22 of the differences qualify as interesting and there is a drop in likelihood when we go to 1 season ahead prediction.

We see apparent learning in all three indeces, however the AP index does not reach our criterion of statistical interest (a false discovery rate of .1) while both the sunspot number and F10.7 prediction show statistically interesting learning through almost all seasons.

Another interesting point is that the likelihood for the 1 season ahead predictions are for the most part less than the 2, 3 and 4 season ahead predictions (in both sunspot number and F10.7 measurements where the results are significant. So prediction of these two indeces counter intuitively are better from 2 seasons out than from 1 season out.

Given this learning exists, we see that for solar data we could use postulated trajectories of solar weather to play scenarios with a spreadsheet like framework for planning maintenance for systems affected cumulatively by the solar weather. To reach the level of predicting solar storms one approach would be constructing conditional trajectories for smaller time intervals based on the weighted set chosen in the seasonal time intervals used here. The results with the artificial data indicates that this learning process may be predicated on how well measured a particular chaotic system of interest is.

The general statistical method proposed here is certainly open for significant improvement. The thread reweighting as new observations appear, the method of density estimation and extrapolation, and the statistical inference approach can all probably be improved, as well as exploring any variable selection that comes out be inspecting how the weights improve the likelihood.

References

1. Matzka, J., Stolle, C., Yamazaki, Y., Bronkalla, O. and Morschhauser, A., 2021. The geomagnetic Kp index and derived indices of geomagnetic activity. Space Weather, *https://doi.org/10.1029/2020SW002641* (DATA https://kp.gfz-potsdam.de/en/data
2. Lalley, 1991, "Beneath the noise, chaos" Annals of Statistics 1999, V27, #2, 461-479
3. Sauer, T., Yoreck, J. and Casdagli, M., 1991 "Embedology", Journal of Statistical Physics, 65, 579-616
4. Hunt, Sauer, York, 1992, "Prevalence, a translation invariant "almost every" on infinite dimensional spaces", Bulletin of the AMS, 1992, v27, n2, October 1992, 217-238.
5. Charles J. Stone, *The Annals of Statistics*, Vol. 5, No. 4 (Jul., 1977), pp. 595-620 https://www.jstor.org/stable/2958783
6. Ye, H., and Sugihara G.,(2016), "Information leverage in interconnected ecosystems, overcoming the curse of dimensionality", Science, Vol 353, issue 6502, 922-925
7. Deyle E.R. Sugihara, G (2011), Generalized Theorems for Nonlinear State Space Reconstruction, PLoS ONE 6(3), e18295, doi:10.1371/journal.poine.0018295
8. Bonetto, F., Galavotti, G., Giuliani, A. and Zamponi, F. (2006),"Chaotic Hypothesis,Fluctuation Theorem and Singularities", Journal of Statistical Physics.V123, no 1. DOI: 10.1007/s10955-006-9047-5
9. Galavotti, G. (1996), "Chaotic Hypothesis: Onsager Reciprocity and Fluctuation-Dissipation Theorem", Journal of Statistical Physics, V84, No5/6, 899-925
10. Ruelle, D, (1976), "A measure associated with axiom A attractors", American Journal of Mathematics, V98, no 3, 619-654
11. Neumann, M.H., (1998),"Strong approximation of density estimators for weakly dependent observation by density estimators from independent observations", Annals of Statistics, V26, no 5, 2014-2048
12. Levina, E. and Bickel, P. J. (2005). "Maximum likelihood estimation of intrinsic dimension". Advances in NIPS 17. MIT Press
13. Trevor Hastie and Brad Efron (2013). Lars: Least Angle Regression, Lasso and Forward Stagewise. R package version 1.2. https://CRAN.R-project.org/package=lars
14. Benjamini, Y. and Yekutieli, D. (2001). "The control of the false discovery rate under dependency", *Ann. Stat.* **29**, 1165-88.