Combining Fossil and Sunspot Data: Committee Predictions

Eric A. Wan
ericwan@ee.ogi.edu

Department of Electrical Engineering
Oregon Graduate Institute
P.O. Box 91000 Portland, OR 97291

Abstract

It is hypothesized that 680 million years ago solar magnetic storms producing ultraviolet and X-radiation affected the earths ozone layer, which in turn influenced the variations in the silt deposition from glacial run-off. Preserved as fossils discovered in South Australia, the striation widths constitute clues to ancient solar activity. Utilizing this noisy data, we have improved our ability to predict the modern sunspot series. In this paper, we detail how the prediction results were achieved through training on the fossil data and committee predictions with the sunspots. Through this exercise, we develop general methods for combining predictors and also time series that may be related but separated in time.

1. Introduction: Fossils and Sunspots

One of the most studied time series corresponds to the record of sunspots dating back to the 1700's. New insight into this series comes from fossilized rocks formed during the Precambrian period in South Australia [10]. Striation widths in the fossils are believed to constitute over a thousand year record of solar activity. The shape and spectral characteristics are strikingly similar to the sunspot series. Figure 1 shows the two series next to each other on the same scale. Note that over 680 million years separate the two time series. The goal of this work is to investigate the utilization of the fossil data to improve the prediction of the modern sunspots. In general, it provides a test case for investigating methods which relate multiple series that may be related but also separated in time. In addition, simple yet effective methods will also be devised for forming and training committees of predictors.

1.1. Benchmark results

Early attempts at sunspot prediction date back to 1927 with the seminal work of Yule [11]. More recent significant results are summarized in Table 1.1. It has been customary to use the dates 1700-1920 for training and then test on various segments of years after 1920. Note that if we consider

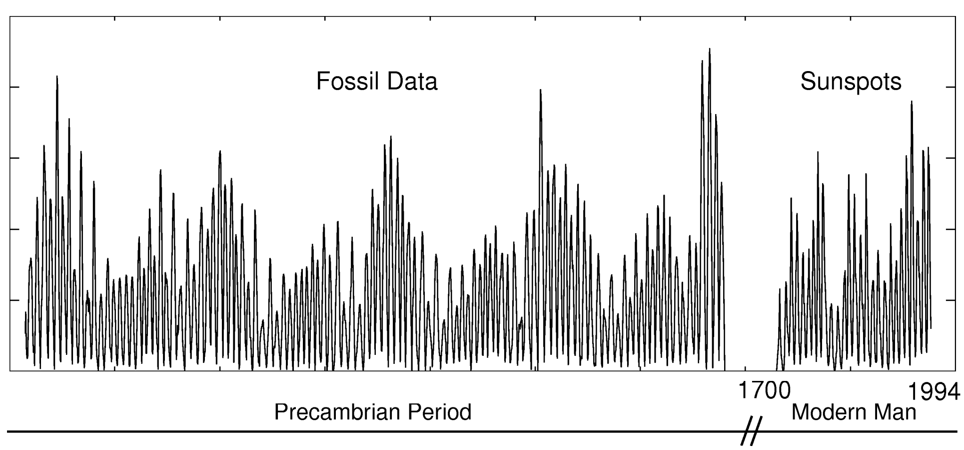

Figure 1. Fossil data and sunspot numbers.

performance on all test data up to the present, the simple Threshold Autoregression (TAR) model of Tong and Lim [6] still outperforms reported results utilizing various neural network techniques [5, 8].

	1700-1920(T)	1921-55	1956-79
Trivial	0.292	0.416	0.94
Linear AR(12)	0.128	0.126	0.36
TAR	0.097	0.097	0.28
WNet	0.082	0.086	0.35
SSNet	-	0.077	-
		1980-94	1921-94
Trivial		0.785	0.661
Linear AR(12)		0.306	0.238
TAR		0.306	0.197
WNet		0.313	0.219
SSNet		-	-

Table 1. Benchmark single step MSE/1535 sunspot predictions. 1700-1920 used for training (T), remaining data used for testing. Trivial - prediction is previous value of series. Linear AR(12) -12th order linear autoregression. TAR - Threshold Autoregressive (Tong and Lim, 1980). WNet - feedforward neural network using weight elimination (Weigend, 1990). SSNet - Soft weight sharing network (Nowlan and Hinton, 1992). (WNet test for 1980-1994 were found by simulating networks with the published parameters)

2. Data Representation

A more natural, though less conventional, representation of both time series is achieved by "de-rectifying" the data (see Figure 2). In other words, at every cycle minimum the sign of the signal is switched. This is well motivated since the approximate 11 year solar cycle actually consists of a 22 year magnetic cycle that flips polarity every 11 years (see Bracewell 1988 [1]). (Taking the absolute value of the new representation returns us to the original data.) The de-rectified time series appears more like a sum of sinusoids with an apparent beat phenomena. In the fossil series, two longer cycles (314 and 315 years) have also been observed. These cycles act as an additive undulation and may explain alternating strengths of successive 11 year solar cycles in the modern data. For purposes of neural network training, both data sets are normalized to match variances (over 1700-1920).

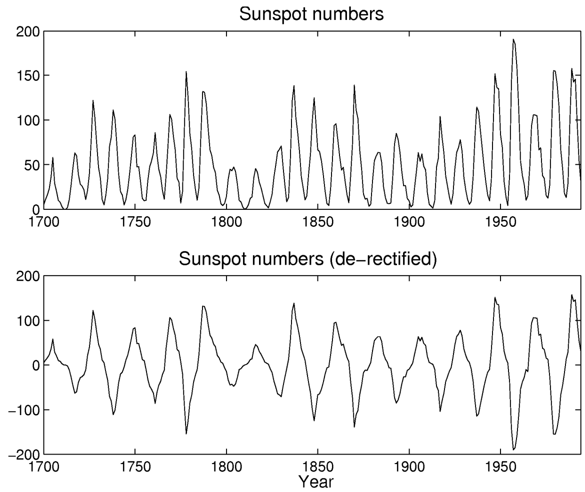

Figure 2. De-rectification of sunspot numbers

3 Methodology

The methodology developed for combining fossil predictions with sunspot predictions involves three steps:

1. *Fossil Prediction:* Train a neural network predictor on the fossil data only. Test on sunspot data.
2. *Committee Prediction:* Form a committee prediction using the fossil trained network with existing sunspot predictors.
3. *Committee Tuning:* Tune the weights of the fossil network on the sunspot data to improve the committee prediction.

Each of these steps will be explained in turn below.

3.1. Fossil Prediction

The first step is to simply train a standard feedforward network on the fossil data alone. A two layer network, with 11 lagged inputs and 6 hidden units, is trained with standard backpropagation. (200 points of the 1336 point fossil series are reserved for cross-validation and model selection). Next, this network is evaluated on the sunspot data. These results are summarized in the table below. Note in this case the standard 1700-1920 "training set" is actually a test set. While the performance is poorer than some of the better predictors, results are actually comparable to the linear predictor which was explicitly trained on the sunspots. This provides strong evidence to support the hypothesis that the fossil data is an indication of solar activity[1]. As a control ex-

[1] As opposed to competing theories regarding tidal origins of the fossil series [9].

periment, networks were also trained on pure sinusoids with frequency equal to the fundamental sunspot cycle (and 1/3 the fundamental). The control networks evaluated on the sunspot series (see Table 2) do not result in comparable performance.

	1700-1920(T)	1921-55	1956-79
FNet	**0.143**	**0.158**	**0.37**
Control1	0.317	0.401	0.98
Control3	1.23	1.89	3.72
		1980-94	1921-94
FNet		**0.376**	**0.269**
Control1		0.865	0.674
Control3		4.64	3.00

Table 2. Fossil network single step MSE/1535 sunspot predictions.

	1700-1920(T)	1921-55	1956-79
WNet+Control1	0.088	0.082	0.32
TAR+Control1	0.100	0.098	0.28
TAR+WNet	0.081	0.075	0.27
FNet+WNet	0.084	0.078	0.26
FNet+TAR	0.094	0.107	0.28
FNet+TAR+WNet	**0.082**	**0.078**	**0.25**
		1980-94	1921-94
WNet+Control1		0.25	0.190
TAR+Control1		0.25	0.187
TAR+WNet		0.238	0.172
FNet+WNet		0.208	0.161
FNet+TAR		0.310	0.202
FNet+TAR+WNet		**0.203**	**0.159**

Table 3. Single step MSE/1535 for committee predictions.

3.2. Committee Prediction

The fossil network predictor was trained on only the fossil data. It is thus likely that the errors made in predicting the sunspots will be different than the errors made by a network that was trained on the sunspot data. This motivates the use of a simple committee prediction as illustrated in the Figure below.

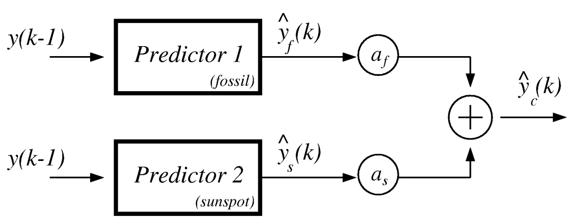

Figure 3. A simple committee prediction. The committee output is the weighting of two predictors with weights chosen to be optimal assuming errors are orthogonal.

The committee prediction is given by the weighted sum of two predictors.

$$\hat{y_c} = a_f \hat{y}_{fossil} + a_s \hat{y}_{sunspot} \quad (1)$$

There is an extensive body of literature on various established methods for choosing the weighting coefficients for both regression and classification committees, including work by Efron and Morris (1973), Jacobs and Jordan *et. al.* (1991), Brieiman (1992), and Tresp and Taniguch (1995) [3, 4, 2, 7]. Here we take a simple constant weighting approach where the coefficients are selected to be optimal such that $a_s + a_f = 1$ under the assumption that the errors made by each committee member are orthogonal. This leads to the simple expressions

$$a_f = \frac{\sigma_s^2}{\sigma_f^2 + \sigma_s^2} \quad \text{and} \quad a_s = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_s^2} \quad (2)$$

where σ_f^2 and σ_s^2 correspond to the prediction error variance from 1700-1920 for the fossil network and sunspot predictor respectively. If the errors are orthogonal and equal then the committee prediction error variance will be reduced by a factor of two. (This can also be extended in a straight forward manner to multiple member committees.)

While the variance based committee coefficients are suboptimal to the least-squares solution of the coefficients over the training set, performance on the test set is observed to be substantially better. Effectively, the non-negativity of the coefficients insures that the committee prediction is always between the minimum and maximum of either committee member (as opposed to the least-squares solution where the generalization performance may vary considerably). We will also motivate the use of this committee weighting scheme in the next section.

The results of forming the committee predictions are summarized in Table 3. It is observed that all committees perform better than any single predictor. Committees formed with the control network indicate a minor advantage to including a network trained on only the fundamental frequency of the sunspot cycle. The best prediction is achieved with the committee formed from the fossil network plus TAR plus WNet. Over the complete test set (1921-1994) the normalized MSE has been reduced by approximately 20% from the best single predictor.

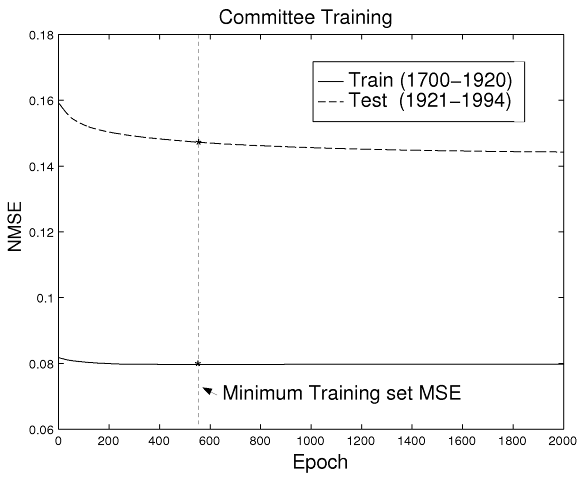

Figure 4. Typical illustration of committee training. Note, there is an actual minimum of the training set. This occurs when additional minimization of the MSE (step 1) leads to increased error correlations. At this point the committee weighting coefficients found (step 2) are suboptimal and in turn tend to increase the prediction error. This point is used as a stopping criteria.

3.3. Committee Tuning

Up to this point, the weights of the fossil network have never been optimized using the actual sunspot data. In the final stage, we consider *tuning* these weights to the sunspot data. Starting with the fossil trained network weights, we alternately perform stochastic gradient descent on the committee error (over one epoch) and reestimate the committee weighting coefficients. This represents a new procedure for training the weights in a committee, and is summarized below:

1. Over one epoch

$$\min_{W} \sum_{k=1700}^{1920} [a_f e_f(k) + a_s e_s(k)]^2$$

2. Reestimate a_f and a_s based on new prediction errors.
3. Loop.

	1700-1920(T)	1921-55	1956-79
FNet+TAR+WNet	0.082	0.078	0.25
Tuned Committee	**0.079**	**0.065**	**0.24**
Standard Dev.	*0.0005*	*0.0008*	*0.0047*
		1980-94	1921-94
FNet+TAR+WNet		0.203	0.159
Tuned Committee		**0.188**	**0.148**
Standard Dev.		*0.0045*	*0.0013*

Table 4. Single step MSE/1535 for committee prediction before and after tuning.

During minimization of the committee error, only the weights W in the original fossil network are adjusted. While the errors from the other committee predictors (TAR and WNet) affect the optimization, the weights of these networks are kept fixed. Note this procedure requires only a slight modification of the backpropagation algorithm.

Selecting the committee weighting coefficients in the manner above is key to this algorithm. Had we used a least-squares solution, then the committee could be viewed as one large network where all parameters (coefficients and network weights) are adjusted to minimize the same error. Effectively, we would simply be training a new network on the data with no guarantee that individual committee members act as actual predictors of the series. Our original assumption that errors should be orthogonal would not hold.

The final results for the tuned committee is summarized in Table 4. As can be seen, these figures represent the best performance in all test sets (the standard deviation for the results are also calculated by repeating the experiment 10 times). A typical learning curve where both the performance of the training set and test set are monitored is illustrated in Figure 4.

We can evaluate to what extent the weights of the original fossil network have changed after committee tuning. The average change in weight magnitude ($E\left[\frac{|w_i^{final} - w_i^{fossil}|}{|w_i^{fossil}|}\right]$) is approximately 10 percent, which would indicate that the network starts out at a near optimal solution. Only a few hundred epochs through the training set were necessary until a minimum is found. A slight variation involves adding to the cost function the "weight-decay" like term $\lambda||W - W^{fossil}||^2$ which explicitly keeps the weights W close to W^{fossil}. However, this does not appear necessary in this experiment, as the committee training did not exhibit overfitting problems.

3.4. Minor variations to the methodology

Several variations are possible on the overall methodology. For example, during the initial phase of training (step 1), various validation sets may be chosen to determine when to stop training and avoid overfitting to the fossil data. Options include: 1) a short segment of fossil data (as reported above), 2) the sunspot series from 1700-1920 (*i.e.* train on fossil data, validate on sunspots), 3) the sunspot series from 1700-1920 using the committee performance to validate. While using committee validation leads to slightly better performance for the untuned committee, the differences in the final performance after tuning appear negligible.

4 Sunspot Only Training

Three factors appear to contribute to our final prediction results: fossil data, de-rectified sunspot representation, and committee predictions. Finally, we investigate to what extent the fossil data contributes to the improved performance. To do this we simply train a committee with a new network on the de-rectified sunspot data starting with random weights (i.e., throw out the fossil trained weights and start over).

The results of this experiment are summarized in Table 5 (TAR + WNet + De-rectified trained Net) and indicate that performance is nearly as good as the committees using the fossil data. This would support the value of our procedure for committee training itself. In this case, however, the method is much more sensitive to early-stopping criteria since we do not start out near a solution.

Note, if we use the standard representation for the sunspots (not de-rectified) we do not show any improvement. This would indicate that the de-rectified representation is also a key factor for this problem.

	1700-1920(T)	1921-55	1956-79
Sunspot committee	0.082	0.066	0.26
		1980-94	1921-94
Sunspot committee		0.200	0.156

Table 5. Single step MSE/1535 for sunspot only trained committee.

5. Conclusions

We cannot as yet draw definitive conclusions as to the relationship between the fossil data and solar activity. However, using simple committees and committee training, we were able to improve our ability to predict sunspots. Incorporating priors into the network by initially training on one series and then tuning on a second, allows for combining data which may be related but separated in time. In addition, we devised a general method for training committees of networks which proved to be effective at improving predictions (even without the explicit use of the fossil data). As to the importance of predicting sunspots, recall that Skylab was brought to an early demise in 1979 due to inadequately forecasting increased atmospheric drag accompanying a sunspot maximum.

Acknowledgements

This work was sponsored by the NSF under grant ECS-9410823.

References

[1] R. Bracewell. Spectral analysis of the elatina series. *Solar Physics*, 116:179–194, 1988.

[2] L. Brieiman. Stacked regressions. Technical Report No. 367, Dept. of Statistics, Berkeley, 1992.

[3] B. Efron and C. Morris. Combining possibly related estimation problem. *J. Royal Statistic. Soc.*, B, 35, 1973.

[4] R. Jacobs, M. Jordan, S. Nowlan, and G. Hinton. Adaptive mixture of local experts. *Neural Computation*, 3:79–87, 1991.

[5] S. Nowlan and G. Hinton. Simplifying neural networks by soft weight sharing. *Neural Computation*, 4(4):473–493, 1992.

[6] H. Tong and K. Lim. Threshold autoregression, limit cycles and cyclical data. *J. Royal Statistic. Soc.*, 42:245–292, 1980.

[7] V. Tresp and M. Taniguch. Combining estimators using non-constant weighting functions. In *Advances in Neural Information Processing Systems*, pages 419–426. 1995.

[8] A. Weigend, B. Huberman, and D. Rumelhart. Predicting the future: a connectionist approach. *International Journal of Neural Systems*, 7(3-4):403–430, 1990.

[9] G. Williams. Cyclicity in the late precambrian elatina formation, south australia: Solar or tidal signature? *Climatic Change*, pages 117–128, 1988.

[10] G. E. Williams. Precambrian varves and sunspot cycles. In B. McCormac, editor, *Weather and Climate Response to Solar Variations*, page 517. Colorado Assoc. Universities Press, Boulder, 1983.

[11] G. Yule. On a method of investigating periodicity in disturbed series with special reference to wolfer's sunspot numbers. *Philos. Trans. Roy. Soc. London*, A 226:267–298, 1927.