The impact of time–averaging on the detectability of nonlinear empirical relations

By Yuval[1]*, William W. Hsieh[1]

[1] *University of British Columbia, Canada*

SUMMARY

This paper studies how time–averaging of observations can affect the detectable nonlinear empirical relations in the data. The example used is the simulation of the precipitation rate as a daily, weekly, and monthly variable. The feedforward Neural Network (NN) model is employed to simulate the precipitation rate. A measure of the nonlinearity of the NN relation is introduced and is used to calculate the nonlinearity of the NNs. It is found that the use of data averaged over periods longer than the inherent time scale of the involved variables can result in a dramatic weakening of the detected nonlinearity. A suggested theoretical explanation asserts that averaging of independent samples of the data records yields distributions approaching the multi–variate normal, in which case the relations among the variables are closer to linear.

KEYWORDS: Nonlinearity Neural Network Averaging data

1. INTRODUCTION

Physical processes in the ocean and atmosphere are often nonlinear. Some relationships between the physical variables are known and are studied using the dynamical governing equations. In other cases, the relationships are not very well understood but can be estimated by statistical models developed empirically from observations. Virtually all these empirical studies use digitized data and this involves an averaging of the true continuous values. It is common to further average data in space and in time to reduce the size of datasets and smooth out noise. For example, climatological studies almost always use monthly means of observed data, and statistical weather analyses usually use daily averaged values.

Due to its nonlinear modeling capability, Neural Network (NN) has become the tool of choice for many studies where a nonlinear relationship is expected. In the last decade, NN models have been applied as nonlinear multiple regression (NMR) to climatological variables (Tangang *et al.* 1998a,1998b; Tangang *et al.* 1997; Hastenrath *et al.* 1995; Navone and Ceccatto 1994), and weather related phenomena (Gardner and Dorling 1999; Kuligowski and Barros 1998a,1998b; Marzban and Stumpf 1996; McCann 1992). There have also been a few attempts to use NN models for the analysis of climatological systems (Tangang *et al.* 1998b; Grieger and Latif 1994) and for climate downscaling (Hewitson and Crane 1996).

NN prediction and analysis of climatological data have been at the focus of our research efforts in the last few years. Several NMR prediction problems on the climatological time scale at mid–latitudes were formulated and studied using the NNs developed by the strategies of Yuval (2000) and Yuval (2001). Examples are the prediction of the Pacific North American pattern, or the prediction of temperatures over land using the Pacific sea surface temperature (SST) as a predictor. Special interest is in studying the relationship between the 500mb geopotential heights and the temperatures over Canada at lead times of a season or longer. This is a problem of practical interest which was studied by Shabbar and Barnston (1996) using the Canonical Correlation Analysis (CCA) method. They found that the Pacific SST was more important than the 500mb geopotential data for temperature prediction. The CCA is a linear method and could not model

* Corresponding author: Department of Earth and Ocean Sciences, University of British Columbia, 1461-6270 University Blvd., Vancouver, BC, Canada V6T 1Z4

the expected nonlinear relation between the 500mb heights and the temperature. There was hope that this nonlinearity will be captured by the NNs and enable improved predictions. It was puzzling to find out that no statistically significant predictive advantage of NMR over linear methods could be detected in this prediction problem, nor in the prediction of other mid–latitude variables on the monthly or seasonal time scale. The failure of our attempts to improve upon results achieved using linear methods, in problems that involved what we believed to be nonlinear relationships, pointed out to some shortcoming in the investigations and prompted a study to find out the reasons.

This paper offers a possible explanation by suggesting that averaging observed data over periods that are longer than the inherent time scale of the involved physical phenomena reduces the nonlinearity of the estimated empirical relationships. The Appendix provides theoretical support to this idea by evoking the central limit theorem. We developed a methodology to quantitatively calculate the nonlinearity of NN empirical relationships to show that reduction in their nonlinearity may indeed take place as the time–averaging window increases, and that this effect can be very significant.

To demonstrate our claim, precipitation rate at various locations is simulated. The generation of precipitation involves vertical lifting, latent heat release, phase changes, and many other physical processes that are inherently nonlinear and thus the connection between the precipitation rate and other physical variables is expected to serve as a good example of a nonlinear relationship. This reasoning lead many studies (e.g., Hall *et al.* 1999, Koizumi 1999, Kuligowski and Barros 1998a,1998b; Navone and Ceccatto; 1994; Hastenrath *et al.* 1995) to try applying NNs to prediction of precipitation rates at various time scales.

The NN simulation in this paper is by models trained using the methodologies of Yuval (2001) and Yuval (2000). The original daily predictor and predictand data are weekly and monthly averaged. NN models are trained to simulate the precipitation at various geographical locations using the daily, weekly, and monthly data. The nonlinearity of the resultant NN models is calculated and presented.

The paper begins with a section describing the data used and defining the simulation problem. It is followed by a section describing the feedforward NN model. The next section develops the measure of the nonlinearity of the feedforward NN. Examples of nonlinearity in observed data relationships that is weakened or modified as the time–averaging window lengthens are shown in the results section.

2. Data and the simulation problem

The data in this study are from the NCEP/NCAR reanalysis project (Kalnay *et al.* 1996) that uses a state of the art global assimilation model to create a comprehensive dataset that is as complete as possible. The output variables are given on a $2.5^o \times 2.5^o$ grid. Forty–one years of mean daily data (1959-1999), totaling 14 975 daily time points were available. NCEP precipitation is purely model based so the NN models only simulate the dynamic model that generates it. This should not affect the conclusions of the paper and their relevance to studies using real data.

The predictand is the precipitation rate (kg m^{-2} s^{-1}) at various locations on earth. The predictors were chosen among the available physical variables, with no missing data, that can be measured on the surface or by satellites. The attempt here is to simulate the predictand at zero lead time ($\delta t = 0$), and thus the concurrent predictor variables are used as predictors. To simulate precipitation at each location, only the predictor variables at that location were used. A better simulation can result from the use of predictors of

past times and of locations in the vicinity of the predictand, but this was not attempted here in order to keep the example simple.

The predictor variables are: 1000mb geopotential height (gpm), precipitable water (kg m^{-2}), surface pressure (Pa), horizontal winds at 10 meters above the surface (m s^{-1}), downwards solar radiation flux at the surface (W m^{-2}), latent heat flux at the surface (W m^{-2}), net long wave radiation flux at the surface (W m^{-2}), sensible heat flux at the surface (W m^{-2}), upwards long wave radiation flux at the surface (W m^{-2}), upwards long wave radiation flux at top of the atmosphere (W m^{-2}), and upwards short wave radiation flux at top of the atmosphere (W m^{-2}).

The NCEP data provide complete global coverage. Several locations in the northern hemisphere with different types of climate were chosen for the examples. The locations are off the coast of British Columbia, Canada ($50^oN, 130^oW$), in the Middle East ($33^oN, 35^oE$), and northeastern China ($50^oN, 123^oE$).

The original daily data were weekly and monthly averaged. The weekly averaging was carried out by taking the mean predictor values of the seven days periods starting from the first day. The monthly averages are means of predictor values in the true calendar months. The predictand is the corresponding total precipitation (i.e. weekly and monthly precipitation amounts). In all cases, the predictors series were normalized before training by the means of their absolute values as the raw data values encompass many orders of magnitude. The lengths of the daily, weekly, and monthly data records are 14 975, 2139, and 492 data points correspondingly. The records were divided into two equal parts, the first was used to develop the models, and the second for testing and calculating the nonlinearity.

3. Two–layer feedforward neural network

An NN connects between some input and a desired output. In applications in the ocean and atmospheric sciences, the input is composed of the predictor variables, and the output, the predictands. In all the examples given in this paper, the NN's output consists of only one predictand. The type of NN considered here is the two–layer feedforward NN. This type of NN has been very successful for practical applications in various fields (Hsieh and Tang 1998) and is the most commonly used. The special class of two–layer feedforward NN is the least complicated class that can perform sufficiently well for the purposes of prediction and analysis of oceanic and atmospheric data.

Consider a set of predictors $\mathbf{P}(t = \tilde{t})$ and a predictand $T(t = \tilde{t} + \delta t)$. The independent variable t is time in most applications, and we consider now predictors and a predictand at a specific time point $\tilde{t}$. In this paper, δt is set to zero to simulate the predictand with concurrent predictors. The two–layer NN connecting $\mathbf{P}(\tilde{t})$ and $T(\tilde{t} + \delta t)$ takes the following form

$$h_i = F(\sum_j \widehat{W}_{ij} P_j(\tilde{t}) + \widehat{b}_i), \quad j = 1, ..., N \tag{1}$$

$$T(\tilde{t} + \delta t) = \sum_i \widetilde{W}_i h_i + \tilde{b}, \quad i = 1, ..., M \tag{2}$$

where $P_j(\tilde{t})$ are the elements of $\mathbf{P}(\tilde{t})$, $\widehat{W}_{ij}$ are elements of $\widehat{\mathbf{W}}$, an $M \times N$ matrix, $\widehat{b}_i$ elements of the $M \times 1$ vector $\widehat{\mathbf{b}}$, $\widetilde{W}_i$ elements of $\widetilde{\mathbf{W}}$, a $1 \times M$ vector, and $\tilde{b}$ is a scalar.

In the NN terminology, the elements of $\widehat{\mathbf{W}}$ and $\widetilde{\mathbf{W}}$ are called the weights. The elements of $\widehat{\mathbf{b}}$ and $\tilde{b}$, are called the biases. The mathematical manipulation in Eq. (1) is referred to as the first, or hidden, layer of the NN. M, the number of rows of $\widehat{\mathbf{W}}$, is

usually called the number of hidden neurons. The function F is the hidden layer transfer function, usually chosen to be the hyperbolic tangent. Equation (2) forms the second, or linear output layer of the NN. The hidden layer can perform manipulations of any degree of nonlinearity but, due to the hyperbolic tangent range, quashes the values into $[-1, 1]$. The output layer linearly scales the products of the hidden layer to any desired value. Given a set of predictors $\mathbf{P}$, the simulation power and flexibility of the NN depends on the number of hidden neurons M which controls the total number of model parameter $Q = (N + 2)M + 1$.

Another, simple way to describe the relation of Eqs. (1) and (2) is by the matrix form

$$T(\tilde{t} + \delta t) = \mathcal{F}[\mathbf{W}, \mathbf{P}(\tilde{t})] = \widetilde{\mathbf{W}} \left(F \left(\widehat{\mathbf{W}} \mathbf{P}(\tilde{t}) + \widehat{\mathbf{b}} \right) \right) + \tilde{b} \tag{3}$$

where $\mathbf{W}$ is a vector containing all the NN weights and biases. Cast in the form of Eq. (3), it is clear that the NN is just a nonlinear operator $\mathcal{F}[\mathbf{W}, \mathbf{P}(t)]$ manipulating the input $\mathbf{P}(\tilde{t})$ through some simple matrix multiplications that results in the output $T(\tilde{t} + \delta t)$.

The actual values of the model parameters, i.e. the elements of $\widehat{\mathbf{W}}$, $\widehat{\mathbf{b}}$, $\widetilde{\mathbf{W}}$, and $\tilde{b}$ in Eqs. (1)-(2) or Eq. (3), are found using a dataset $[\{\mathbf{P}(t_i), T(t_i + \delta t)\}, \quad i = 1, 2, \cdots, L]$, where L is the number of data points. An optimization problem is formulated by constructing a cost function that measures the mean squared error between the model output and the observed data. (In certain problems, the cost function should also include additional constraints). A nonlinear search algorithm is employed in order to find the model parameters that minimize the cost function. In NN terminology, this process is called training. Detailed descriptions of NN training and considerations regarding its implementation are given in text books like Bishop (1995) and Ripley (1996).

An NN study must start with decisions about the NN models architecture, and about the type of NN training to be used in order to recover the optimal NN model parameters. It was verified that for all the examples given in this paper, using one hidden neuron ($M = 1$ in Eq. (2)) was sufficient to adequately simulate the relationship. That results in total of 15 model parameters in each of the models. Given that the number of independent examples in each dataset was more than an order of magnitude larger, the NN trainings in all cases can safely be considered overdetermined and resulted in unique NN models. The MATLAB Levenberg-Marquardt routine (Demuth and Beale 1998) was used to train the models. In all cases, the first half of the dataset was used for training the model and the second half was used for independent testing of the results.

4. Nonlinearity of feedforward neural network model

The main advantage of the NN as a modeling tool is the ability to simulate any relationship with no need for prior assumptions about its degree of nonlinearity. If the NN simulation generalizes a relationship to a satisfactory level, then the degree of nonlinearity of the NN operator can serve as a clue while attempting to model this relationship from basic physical principles. In this section, the meaning of the statement that an NN is a *nonlinear* model is explored and a method is set forth for a simple estimation of that nonlinearity. The nonlinearity measure is a function of the independent variable, usually time, and thus can serve as an analysis tool to study the temporal variability in the nonlinearity of the data relationships.

(*a*) *Definition of the nonlinearity measure of a feedforward NN*

By definition, an operator $\mathcal{F}$ is nonlinear if, and only if,

$$\mathcal{F}[a_1\mathbf{P}_1 + a_2\mathbf{P}_2] \neq a_1\mathcal{F}[\mathbf{P}_1] + a_2\mathcal{F}[\mathbf{P}_2] \ . \tag{4}$$

To satisfy the above condition, $\mathcal{F}$ must possess a non–zero derivative with respect to $\mathbf{P}$ of order higher than the first. That means that in the Taylor expansion

$$\mathcal{F}[\mathbf{P} + \delta\mathbf{P}] = \mathcal{F}[\mathbf{P}] + \mathcal{F}'[\mathbf{P}]\delta\mathbf{P} + \cdots + \frac{1}{(n)!}\mathcal{F}^n[\mathbf{P}](\delta\mathbf{P})^n + R_n(\delta\mathbf{P}), \tag{5}$$

at least one of the higher–than–first order terms must be different from zero. Note that the derivatives in Eq. (5) are directional derivatives in the direction $\delta\mathbf{P}$.

Equation (5) facilitates a simple test to assess the nonlinearity of an operator. Define the ratio

$$\mathcal{R} = \frac{\mathcal{F}[\mathbf{P}] + \mathcal{F}'[\mathbf{P}]\delta\mathbf{P}}{\mathcal{F}[\mathbf{P} + \delta\mathbf{P}]}. \tag{6}$$

For a linear $\mathcal{F}[\mathbf{P}]$ all the higher order terms in Eq. (5) equal zero and $\mathcal{R} = 1$. For a nonlinear $\mathcal{F}[\mathbf{P}]$, some of the higher order terms must exist and $\mathcal{R} \neq 1$. This nonlinearity test follows the idea of estimating a trust region around a point in which a function can be considered linear. Construction of trust regions was studied in the context of designing Newton method algorithms for nonlinear optimization. More information can be found in numerical optimization texts (e.g., Dennis and Schnabel 1996; Kelley 1999; Nocedal and Wright 1999).

(*b*) *Practical considerations for calculating $\mathcal{R}$ values*

The choice of the perturbation $\delta\mathbf{P}$ is an important issue for practical purposes. A reasonable choice is

$$\delta\mathbf{P} = \nu\mathbf{P}, \tag{7}$$

where ν is some small value. This means that $\delta\mathbf{P}$ is a small step in the input space in the direction of the input. The choice of perturbation in Eq. (7) renders $\mathcal{R}$ a measure of the nonlinearity of the NN in the direction of $\mathbf{P}$. Other options exist and can serve for different purposes. An important example is

$$\delta\mathbf{P} = \nu\mathbf{P}.*\mathbf{e} \tag{8}$$

where $\mathbf{e}$ is a vector whose elements equal zero excepting the ith element that equals one, and $.*$ is the vector multiply operation (Hadamard element by element product). Using the perturbation in Eq. (8), the value of $\mathcal{R}$ is a measure of the nonlinearity of the predictand in the direction of the ith predictor. This measure can be used to study the nonlinearity of the connection between the predictand and a specific predictor.

The magnitude of $\delta\mathbf{P}$ is controlled by the value of ν. A remaining question is what this value should be. The Taylor expansion of Eq. (5) is valid only for $\delta\mathbf{P}$ which is small enough such that the $R_n(\delta\mathbf{P})$ term is negligible. On the other hand, choosing very small values for $\delta\mathbf{P}$ while evaluating $\mathcal{R}$, might result in very small values for the high order terms and $\mathcal{R}$ very close to unity. In that case it is not clear whether the result points to actual lack of nonlinearity in the relationship, or is an artifact of an excessively small choice for the value of ν. A similar problem is encountered while estimating the nonlinearity for the construction of a suitable trust region. Various solutions were suggested, all of whom rely on some experimentation to arrive at an arbitrary small tolerance value rendering a perturbation that is considered small enough. For the purpose of this paper, a suitable

value for ν is chosen by the following considerations. An observation noted in all the cases that we studied (including many not described in this paper), is that the pattern of $\mathcal{R}$ as a function of t is almost identical in form for all ν values below a certain limit, i.e., $\mathcal{R}(\nu, t_i)/\mathcal{R}(\nu, t_j) \approx Constant$ for any (i, j) and any $\nu < Lim$, where Lim is some limit found experimentally. This means that the conclusions about the nonlinearity at a time point, *relative* to the nonlinearity in other points, using $\nu \longrightarrow 0$ are similar to those using any other choice of $\nu < Lim$. Thus any value of $\nu < Lim$ is an appropriate choice and we chose one that resulted in most of the $\mathcal{R}$ values in our examples falling in the range $[-5, 5]$. Obviously this consideration can be applied only for comparisons of nonlinearity of models $\mathcal{F}$ defined on a space spanned by the same predictors $\mathbf{P}$. A comparison between the nonlinearity of models fed by different predictor variables must be ruled out unless some appropriate normalization of $\mathcal{R}$ is found. Such a comparison is not needed for this paper.

5. Results

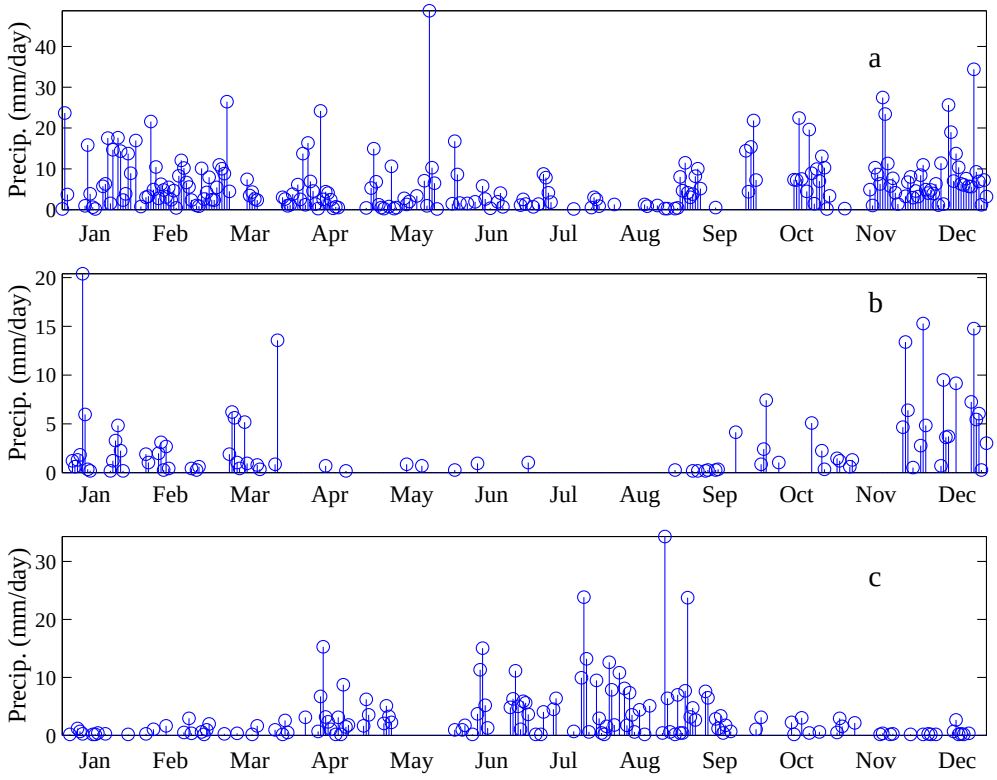

Figure 1. Precipitation patterns in locations at (a) British Columbia coast ($50^o N, 130^o W$) (b) Middle East ($33^o N, 35^o E$) (c) northeastern China ($50^o N, 123^o E$). Shown is the observed daily precipitation during 1979 (the first full year of the testing record).

Figure 1a provides a one–year record of the observed daily precipitation at the location along the British Columbia coast. It shows the typical pattern of year round precipitation, with higher and more frequent accumulations during the winter months. The upper panels of Figure 2 show scatter plots of the NN simulation of the daily, weekly,

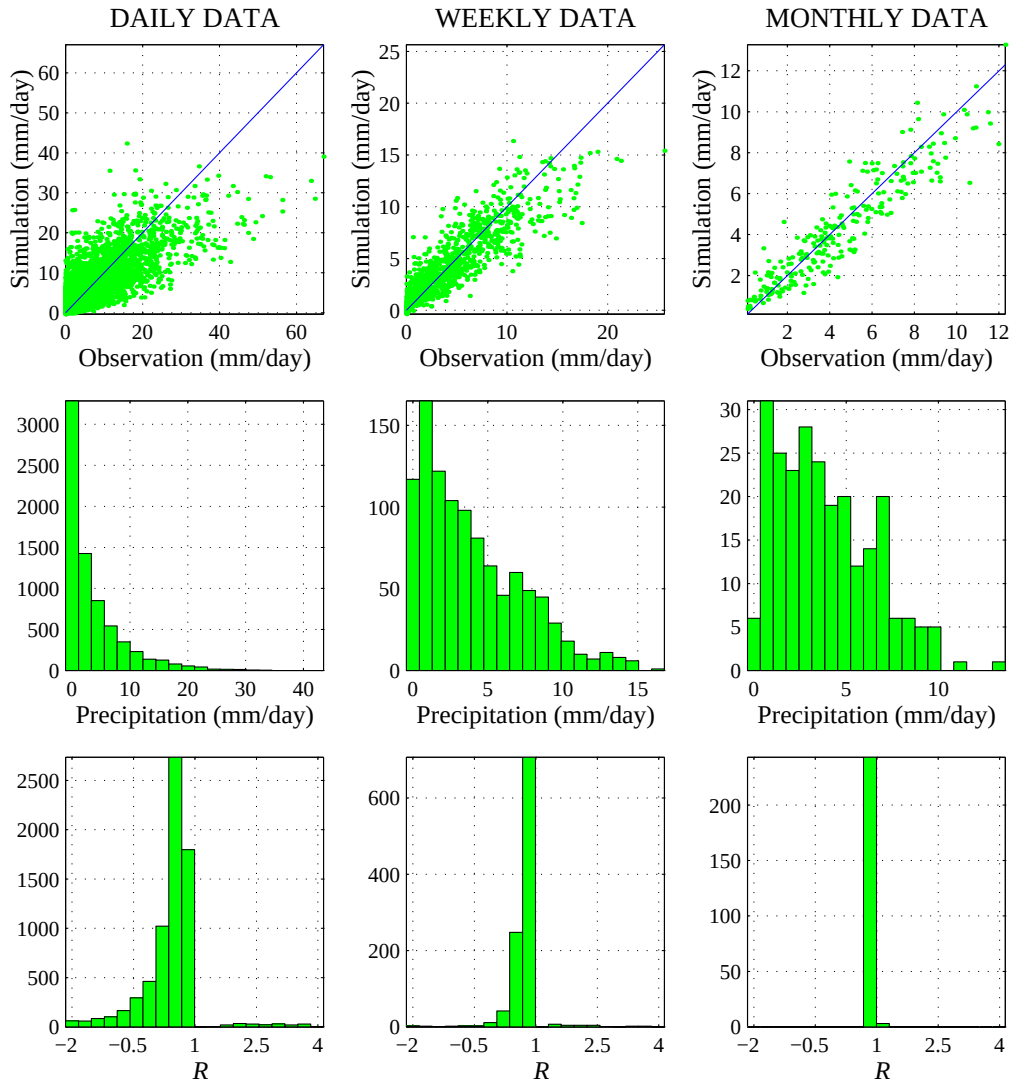

Figure 2. Results from the location off the British Columbia coast. The upper panels show scatter plots of NN simulation of the daily, weekly, and monthly precipitation as a function of the observed values. The diagonal line is the perfect one-to-one relationship. The weekly and monthly precipitation amounts were normalized to mm/day units. The middle panels show the density distribution histograms of precipitation amounts and the lower panels the corresponding plots of $\mathcal{R}$ values. The histograms are counts of the number of occurrences of precipitation and $\mathcal{R}$ values that fall within the limits of equally spaced bins dividing the value ranges.

and monthly precipitation at that location. The correlation and root mean square error (RMSE) skills of the simulations, calculated for the testing period (second half of the record), are given in Table 1. The upper panels of Fig. 2 show a clear trend of reduced scattering of the simulations around the one-to-one line, implying relative improvement of simulations with increasing length of the time–averaging window. This improvement is reflected in the increased correlation and decreased RMSE.

Histograms of the $\mathcal{R}$ nonlinearity values are shown in the lower panels of Fig. 2. The distribution of $\mathcal{R}$ values calculated for the daily data is wide and strongly skewed towards

left (lower $\mathcal{R}$ values). The highest density is around $\mathcal{R} = 0.7$. The corresponding distribution for the weekly data is densest around $\mathcal{R} = 1.0$ (i.e., linear) and is not as wide. The $\mathcal{R}$ values calculated for the monthly data relationship are all very close to unity. These results show a clear trend of diminished nonlinearity with increasing time-averaging window in the simulated relationship between the predictors and the predictand. Corroborating this conclusion is the comparison in Table 1 between the correlation and RMSE skills of the NN and linear regression simulations. The significant advantage of the NN in the daily data simulation (assumed to be due to its nonlinear capability) is decreased in the weekly data simulation, and is almost non existent in the monthly data. The actual time series of simulated monthly precipitation by the NN and the linear regression (not shown) are hardly distinguishable. The similarity of the NN and linear regression simulations of the monthly data, and their almost identical correlation and RMSE skills suggest that the NN has degenerated in this case to linear regression.

TABLE 1. Correlation (CC) and RMSE skills of precipitation simulations by NN and linear regression (LR). The skills were calculated on the second halves of the data records. The simulating models were developed using the first halves. The NN/LR columns give the ratio between the corresponding NN and regression skills. Note the clear trend of decrease in this ratio as the time–averaging window increases. This decrease is not as significant in the northeastern China location.

Location	Skills	Daily			Weekly			Monthly		
		NN	LR	NN/LR	NN	LR	NN/LR	NN	LR	NN/LR
British Columbia	CC	0.80	0.73	1.10	0.88	0.86	1.02	0.93	0.93	1.00
	RMSE	0.93	1.06	0.88	0.44	0.48	0.92	0.24	0.25	0.96
Middle East	CC	0.86	0.76	1.13	0.90	0.88	1.02	0.94	0.93	1.01
	RMSE	1.40	1.80	0.78	0.67	0.76	0.88	0.38	0.40	0.95
Northeast China	CC	0.81	0.73	1.11	0.91	0.88	1.03	0.97	0.96	1.01
	RMSE	1.37	1.60	0.86	0.62	0.69	0.90	0.32	0.35	0.91

Our interpretation of these experimental results is based on the modification that takes place in the forms of the the predictors and predictand data distributions as a result of the weekly and monthly averaging. The middle panels of Fig. 2 show the distributions of daily, weekly, and monthly simulated precipitation values. The daily data distribution is one–sided and long–tailed. The distribution of the weekly and monthly means are less one-sided, and are denser at increasingly higher values. Although far from being normal–like, it seems that the weekly and monthly distributions are getting closer to the normal distribution. This is in agreement with the central limit theorem that states that the sampling distribution of the means of a variable with an arbitrary distribution approaches the normal distribution. The weekly and monthly means of the predictors (not shown) exhibit a trend similar to that of the predictand. As discussed in the Appendix, relation between normally distributed variables is linear and thus the predictor–predictand relations tend to get closer to linear when the data are averaged and their distributions closer to normal.

Both predictors and predictand data are serially correlated. In the case of the British Columbia coast weather, the dominant autocorrelation is about 3-5 days. Thus taking both weekly and monthly means is expected to significantly average the data, strongly modify their distribution, and significantly lower the nonlinearity in the relationship between the predictors and the predictand.

A one year time series of precipitation in the location in the Middle East is given in Fig. 1b. Precipitation in this region is usually associated with winter synoptic troughs. The summer and autumn are almost totally dry. The upper panels of Fig. 3 show the scatter plots of NN simulations. The lower panels show the calculated nonlinearities. Similar to the results in the British Columbia coast, there is a general tendency of de-

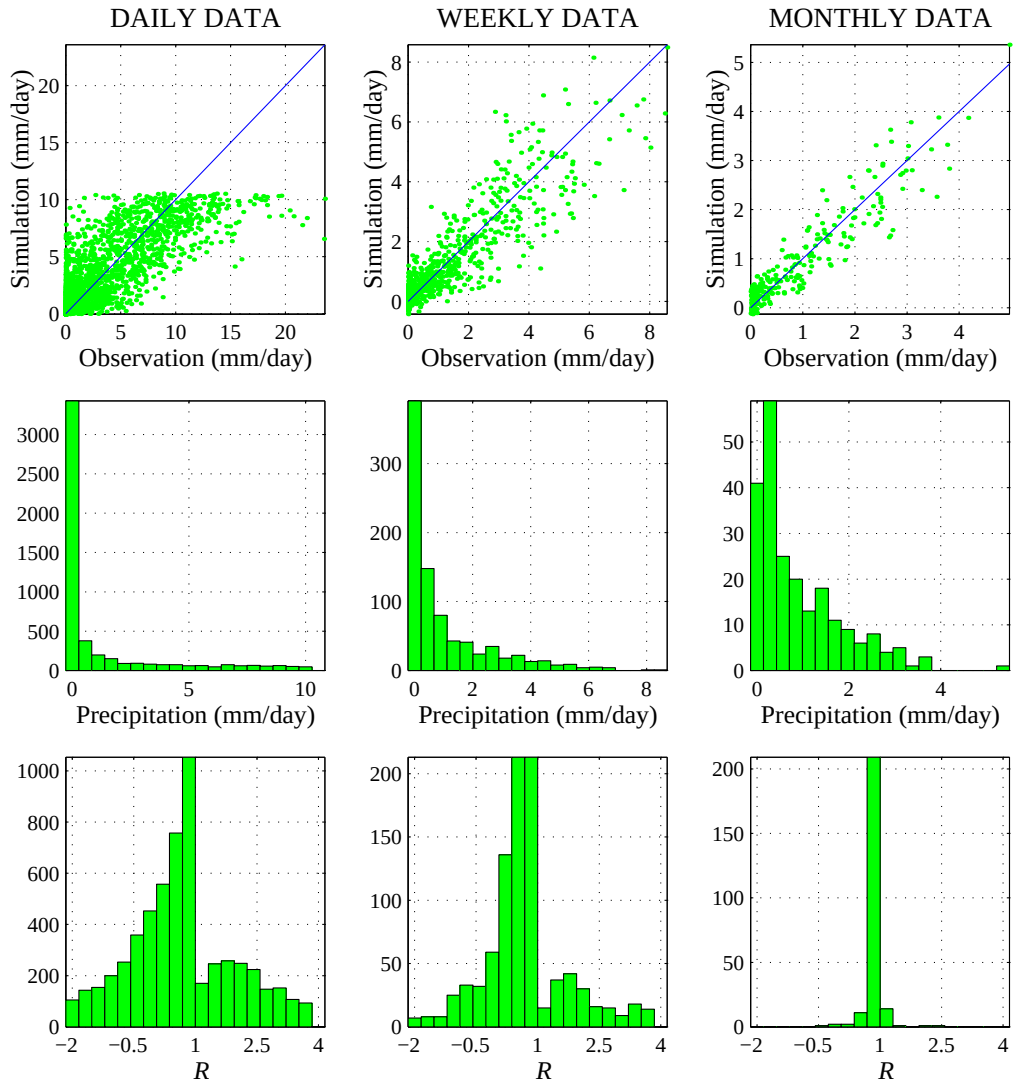

Figure 3. Similar to Fig. 2 but for the location in the Middle East.

creasing scatter in the upper plots, and improvement of the simulation skills in Table 1, with increasing length of the time–averaging window. This trend is associated with a narrowing of the $\mathcal{R}$ histograms in the lower panels of the figure which means a decrease of the nonlinearity.

The nonlinearity reduction is less noticeable in the weekly data simulation, where the $\mathcal{R}$ distribution is similar in shape but slightly more narrow than that of the daily data simulation. The nonlinearity in the monthly data relationship is clearly diminished but not as significantly as in the British Columbia location. The impact of time–averaging of data in this location is different than that in the British Columbia Coast. This can be explained by examining the distributions of the daily, weekly, and monthly precipitation shown in the middle panels of Fig. 3. The weekly data distribution is slightly wider but still one–sided and long–tailed. The monthly data distribution is narrower but still very

strongly skewed to the right. As can be seen in Fig. 1b, the precipitation pattern in the Middle East location includes many completely dry periods intercepted (in the winter) by sporadic precipitation events about week long. Weekly averaging the data in this location is thus not always significant, and as a result the change in the nonlinearity is not as large.

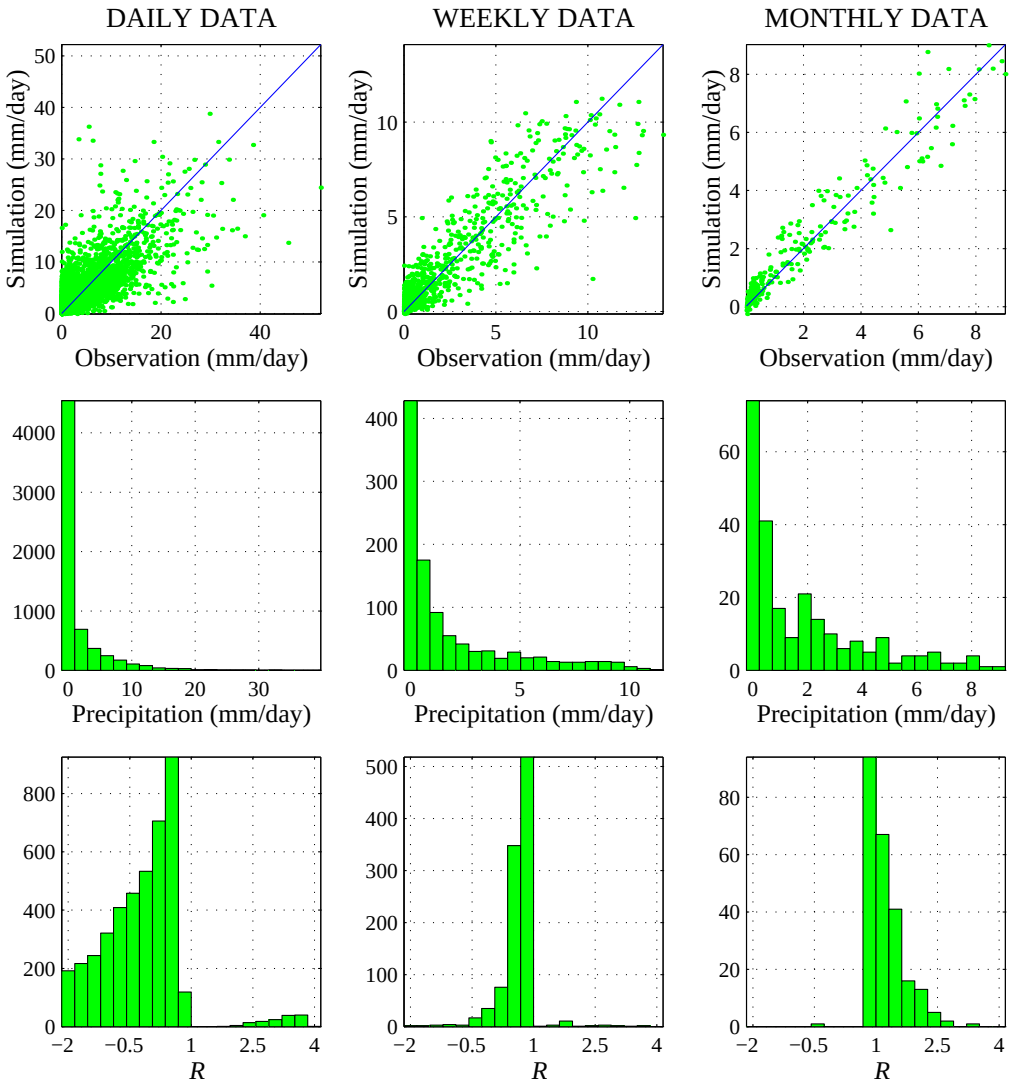

Figure 4. Similar to Fig. 2 but for the location in northeastern China.

The last example is of precipitation simulation at a location in northeastern China, where most of the precipitation occurs in the summer but occasional rain is possible in any season. Precipitation events at this location are very variable in terms of length and intensity. The scatter plots in the upper panels of Fig. 4 reveal a trend, similar to that in the two other locations, of decreasing scatter with increasing length of time–averaging window. The change in the scattering seem to be less substantial than in the first two examples. The comparison of the NN and linear regression simulation skills in Table 1

show that in this location the reduction in the advantage of the NN simulation is not as significant. This suggests that less reduction in the nonlinearity with averaging occurs in the predictor–predictand relationship at this location.

Examination of the $\mathcal{R}$ histograms in the lower panels of Fig. 4 reveals that indeed the reduction in the nonlinearity, especially in the monthly data, is not as strong as in the corresponding plots in Figs. 2 and 3. This is in agreement with the modification occurring in the precipitation distributions shown in the middle panels of Fig. 4. The weekly and monthly distributions seem to be less affected by the averaging compared to the corresponding distributions of data in the other locations.

An interesting point is the change in the general shape of the distribution of nonlinearities in the monthly data relationship at the northeastern China location. The daily and weekly ones are strongly skewed to the left while that of the monthly data is strongly skewed to the right. That means a change in the sign of the sum of high order terms in the Taylor expansion (Eq. (5)), and probably a radical change in the predictand–predictor relationship. A possible explanation does not readily come to mind but it might be related to the large variability in the length and intensity of precipitation events at that location, and its impact on the averaging process. The data distributions might not have been affected much but the effect of averaging on the predictand might have been different than on the predictors. The tendency to modify the distributions towards normal is not as strong, reflected by smaller reduction in the nonlinearity, but the nature of the predictor–predictand relation might have been affected, resulting in different nonlinearity patterns.

6. Conclusions

This study was prompted by the realization that using monthly means of physical variables, it is hard to detect nonlinearity in NN climatological studies. Here we define a quantitative measure of the nonlinearity of an NN. Using as an example the simulation of precipitation, it is shown that the nonlinearity of a relationship generally decreases as the data time–averaging window lengthens. The significance of the change in the nonlinearity depends on the length of the time–averaging windows, and the time scales on which the predictors and the predictand vary. The nonlinearity of a relationship between variables like the 500mb geopotential height and temperature, that vary on the daily or hourly time scale, cannot be captured while using data given as monthly means of the variables. Nonlinearities between such variables on the climatological time scale cannot be ruled out, but may not be detectable from the short and noisy data records.

The results demonstrated here are similar to those found in many other simulation and prediction problems that we studied. Their implications for future empirical studies are important. The earth sciences community is increasingly becoming aware of the need to use nonlinear statistical tools in data analysis. It is important to bear in mind that the advantages of using these tools cannot be realized unless the data are properly sampled. Using data averaged over periods longer than the inherent time scales of the predictors or predictands, the nonlinearity in their relationships are smoothed out. The skills of the simulation or prediction at that time scale might be misleadingly high but they only reflect the strength of the linear relationship. No wonder that advantages of NNs were convincingly demonstrated only by researchers who used properly sampled data. Good examples are the results achieved by Gardner and Dorling (1999) who used hourly data to predict nitrogen oxides pollution, and Kuligowski and Barros (1998a) who used six–hourly data for precipitation forecasts.

Using NN and other nonlinear statistical methods is usually not as straightforward as using traditional linear methods. In particular, the short length of the climate records form a formidable problem (Yuval 2000; Hsieh and Tang 1998). Before attempting a study involving a nonlinear empirical method, it must be considered whether the data enable nonlinear relationships that the method can potentially find. If this is not the case, a simpler linear method should be applied. Data availability almost always dictates some compromises, however, their implications should be kept in mind. For example, precipitation processes usually develop on a time scale that is shorter than one day. Thus we are aware that the results shown here could have probably been more dramatic using hourly data. We believe that they are convincing enough as they are.

The nonlinearity measure introduced in this paper proved to be a useful tool to investigate the failure of NN to improve the prediction and simulation of some relationships assumed to be nonlinear. This measure should be considered as an additional analysis tool also for other purposes. Provided that the NN prediction or simulation is adequate, its nonlinearity, and the variation of this nonlinearity with time, can provide hints about the relationship captured by the NN. In most cases, NNs are applied when the physical problem at hand is not well understood; their nonlinearity should be born in mind while trying to arrive at a proper analytical relationship developed from basic physical laws.

Acknowledgement

We thank Eldad Haber for suggesting the form of the nonlinearity measure and for numerous helpful discussions. This work was supported by research and strategic grants to William Hsieh from the Natural Sciences and Engineering Research Council of Canada.

Appendix

For simplicity, consider the relation between two variables x and y. If $y = f(x)$ is a nonlinear function, then even if x is a normally distributed random variable, y will in general not have a normal distribution. Now consider the effects of time-averaging on the (x, y) data. The bivariate central limit theorem (Bickel and Doksum, 1977, Theorem 1.4.3) says that if $(x_1, y_1), \ldots, (x_n, y_n)$ are independent and identically distributed random vectors with finite second moments, then (X, Y), obtained from averaging $(x_1, y_1), \ldots, (x_n, y_n)$, will, as $n \to \infty$, approach a bivariate normal distribution $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, where μ_1 and μ_2 are the means of X and Y, respectively, σ_1^2 and σ_2^2 are the corresponding variances, and ρ the correlation between X and Y. The conditional probability distribution of Y given X is also a normal distribution (Bickel and Doksum, 1977, Theorem 1.4.2), with an expectation

$$E[Y] = \mu_2 + (X - \mu_1)\rho\sigma_2/\sigma_1. \tag{A.1}$$

This linear relation explains why time-averaging tends to linearize the relationship between the two variables. The bivariate theorems readily generalize to the multi–variate case concerning this paper.

In the case of a relationship between variable series with autocorrelations longer than the time–averaging window, the averaged values are not independent and the bivariate central limit theorem does not apply. For example, variations in physical variables related to the El Niño Southern Oscillation (ENSO) phenomenon can be considered as random variables where each ENSO event is one realization. The length of autocorrelation of ENSO time series is about one year. Thus, any ENSO–related detectable nonlinearity

will not be diminished by monthly averaging of daily observations. However, mid–latitude weather variables have an autocorrelation length of a few days. Monthly averaging their daily observations will tend to modify their distribution towards multi–variate normal, and their mutual relationships towards linear.

References

Bickel, P. J., and Doksum, K. A. 1977 *Mathematical Statistics*, Holden–Day

Bishop, C. M. 1995 *Neural Networks for Pattern Recognition*, Clarendon Press

Demuth, H., and Beale, M. 1998 *Neural Network Toolbox Version 3.1*, The Math Works, inc.

Dennis, J. E., and Schnabel, R. B. 1996 *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM

Gardner, M. W., and Dorling, S. R. 1999 Neural network modeling and prediction of hourly NO_x and NO_2 concentrations in urban air in London. *Atmospheric Environment*, **33,** 709-719

Grieger, B., and Latif, M. 1994 Reconstruction of the El-Niño attractor with neural networks. *Climate. Dyn.*, **10,** 267-276

Hastenrath, S., Greischar, L., and van Heerden, J. 1995 Prediction of the summer rainfall over South Africa. *J. Climate*, **8,** 1511-1518

Hewitson, B. C., and Crane, R. G. 1996 Climate downscaling: techniques and application. *Clim. Res.* **7**, 85-95

Hall, T., Brooks, H. E., and Doswell, C. A. 1999 Precipitation forecasting using a neural network. *Wea. and Forecasting*, **14,** 338-345

Hsieh, W. W., and Tang, B. 1998 Applying neural networks models to prediction and data analysis in meteorology and oceanography. *Bull. Am. Met. Soc.*, **79,** 1855-1870

Kelley, C. T. 1999 *Iterative Methods for Optimization*, SIAM

Kalnay, E., and coauthors 1996 The NCEP/NCAR 40–year reanalysis project. *Bull. Am. Met. Soc.*, **77,** 437-471.

Koizumi, K. 1999 An objective method to modify numerical model forecasts with newly given weather data using an artificial neural network. *Wea. Forecasting*, **14,** 109-118

Kuligowski, R. J., and Barros, A. P. 1998 Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks. *Wea. Forecasting*, **13,** 1194-1204

Kuligowski, R. J., and Barros, A. P. 1998 Experiments in short–term precipitation forecasting using artificial neural networks. *Mon. Wea. Rev.*, **126,** 470-482

Marzban, C., and Stumpf, G. J. 1996 A neural network for tornado prediction based on Doppler radar–derived attributes. *J. of App. Met.*, **35,** 617-626

McCann, D. W. 1992 A neural network short–term forecast of significant thunderstorms. *Forecasting Techniques*, **7,** 525-534

Navone, H. D., and Ceccatto, H. A. 1994 Predicting Indian monsoon rainfall—a neural network approach. *Climate Dyn.*, **10,** 305-312

Nocedal, J., and Wright, S. J. 1999 *Numerical Optimization*, Springer Series in Operation Research, Springer

Ripley, B. D. 1996 *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge

Shabbar, A., and Barnston, A. G. 1996 Skill of seasonal forecast in Canada using Canonical Correlation Analysis. *Mon. Wea. Rev.* **124,** 2370-2385

Tangang, F. T., Hsieh, W. W. and Tang, B. 1997 Forecasting the equatorial Pacific sea surface temperature by neural network models. *Climate Dyn.* **13,** 135-147

Tangang, F. T., Hsieh, W. W. and Tang, B. 1998a Forecasting the regional sea surface temperature of the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. *J. Geophys. Res.* **103,** 7511-7522

Tangang, F. T., Tang, B. Monahan, A. H., and Hsieh, W. W. 1998b Forecasting ENSO events: A neural network—extended EOF approach. *J. Climate* **11,** 29-41

Yuval 2000 Neural network training for prediction of climatological time series, regularized by minimization of the Generalized Cross Validation function. *Mon. Wea. Rev.*, **128,** 1456-1473

Yuval 2001 *Strategies for implementing Neural Networks in ocean and atmosphere studies*, Ph.D. thesis., The University of British Columbia