Predicting Meteorological Values on a Spatial Grid using the North American Land Data Assimilation System

Felipe Hernández
University of Pittsburgh
felher.c@gmail.com

Ben Humberston
Carnegie Mellon University
bhumbers@cs.cmu.edu

Abstract

In this project we examine the use of Vector Auto-Regression (VAR) models for weather forecasting. Using satellite-based hourly observations, we evaluate the performance of single and multi-cell models as well as those with different model input history lengths, or lag orders. Lasso regularization is used to prevent overfitting of high complexity models. We find that a baseline linear regression model is as effective for one-hour-ahead forecasting as any of the tested VAR models, indicating that simple models are sufficient for predicting weather system dynamics over the short time scales being considered.

1 Introduction

1.1 Motivation

Weather forecasts are broadly used; from personal activity planning (What clothes should I wear today? Will it be a good time to plan outdoor events?); to large-scale economic decision-making (What activities should be prioritized on a crop next week? What precautions should the air-traffic controllers enforce on a given day?); to emergency preparation and response (What alternate routes should become available due to snow?, Where should the emergency vehicles be sent during a flood event?). The availability and accuracy of forecasts thus have a profound impact on human activities at many levels, both in measurable and unmeasurable aspects. However, predicting weather is a difficult research problem. Most often, physically-based models with global scale are used to forecast future conditions. In this project, we instead take a machine learning approach focused on a local scale in order to predict atmospheric variables in a small geographic neighborhood. Using vector-autoregressive (VAR) models, we predict future atmospheric system state, which includes variables such as air pressure, precipitation, and temperature, based on prior system states provided by historical satellite observations. The effectiveness of these models for predicting system state one hour in the future is compared with a baseline linear regression and found not to be significantly more accurate for our particular performance measures.

1.2 Background and Related Work

Researchers in the atmospheric sciences have investigated a variety of methods for weather forecasting. Many subtle physical phenomena affect the weather at any given time, including energy and mass transfer between the sun and different layers of the atmosphere, ground, and ocean [1]. The use of physically-based forecasting models requires detailed measurements of these phenomena, but such data is generally not available. Given this data scarcity along with the prohibitive complexity of these models, the use of simplified models is often preferred or even mandatory [2].

Machine learning approaches have been explored to construct simplified models based on meteorological measurements. Many of these techniques are tailored to fit the nature of the observations available. Weather monitoring stations are the predominant data source, providing point measurements with high accuracy and varying temporal resolution.

Artificial neural networks have proven to be effective in such cases for forecasting rainfall amounts in the near future given a time series of previously observed values ([3] and [4]). More recent works have attempted to combine NNs with other techniques to improve the performance of predictions. In [2], a recursive NN is trained using a support vector regression together with a chaotic particle swarm optimizer, while in [5], a genetic algorithm is used to calibrate the network. Other algorithms used in weather forecasting include linear regression, discriminant analysis, logistic regression [6], wavelets, neuro-fuzzy models [7], and genetic programming [8].

Recently, meteorological observations from satellite and Doppler radar have become widely available, providing ubiquitous coverage at the cost of decreased precision. These sources of information add spatial dimensions to the forecasting problem. For example, Fourier spectrum, structure function, and moment-scale analyses are used to understand radar precipitation in [9]; decision trees are used on a Lagrangian reference framework to learn rainfall behavior from satellite images in [10]; and several sources of information, including satellite and radar images, are proposed in [11] for the spatial estimation of multiple weather variables.

Forecasting models have also been created to take advantage of interdependencies between atmospheric variables that are measured or estimated using other models. Rain-gauge data and outputs from atmospheric models are used for forecasting precipitation in [12] and [13]. Additional upper air soundings are also used in [14], and radar and satellite data is incorporated in [15].

1.3 Dataset

In this work, we use data from the satellite-based NLDAS-2 (North American Land Data Assimilation System), a service hosted by the Goddard Earth Sciences Data and Information Services Center at NASA [16]. It provides hourly weather data for the US beginning in 1980. The data is provided on a regular cell grid with a resolution of 1/8th of a degree in latitude-longitude coordinates. Figure 1 shows an example of the weather images available from NLDAS-2, and Table 1 lists the full set of variables available at each spatial cell.

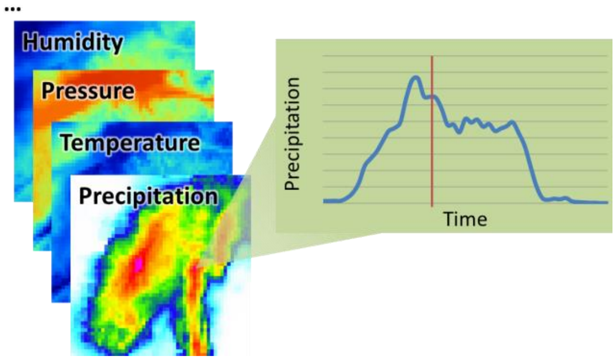

Figure 1. Spatial and temporal dimensions of the NLDAS-2 meteorological observations [16]: hourly images of each variable are available on a regular latitude-longitude cell grid.

Table 1. Meteorological variables available on NLDAS-2 [16]

Name	Description	Units
precipitation	Accumulated height of precipitated water column	millimeters
available potential energy	180-0 mb above ground Convective Available Potential Energy	Joules per kilogram
% convective precipitation	Fraction of total precipitation that is convective	none
LW radiation	Long wave radiation flux downwards (surface)	watts per square meter
SW radiation	Short wave radiation flux downwards (surface)	watts per square meter
potential evaporation	Potential evaporation	millimeters
surface pressure	Surface pressure	pascals
specific humidity	2 m above ground Specific humidity	none
temperature 2 m	2 m above ground Temperature	kelvin
zonal wind speed	10 m above ground Zonal wind speed	meters per second
meridional wind speed	10 m above ground Meridional wind speed	meters per second

2 Methods

We represent the system being observed as $\boldsymbol{X}$, an N-length vector formed by reshaping and concatenating all of the geographic grids of observed meteorological variables (temperatures, air pressures, and so on). The system is thus comprised of $N = SV$ values, where $S = s_x s_y$, and s_x and s_y are the number of cells in the horizontal and vertical directions, respectively. V is the number of meteorological variables of interest that is observed at each cell. Note that each variable is offset to zero-mean and normalized to unit variance over all time steps in order to negate the effects of different unit measurements.

2.1 Linear Regression

Our baseline model is a linear regression in which the system state at one prior time step is used to estimate the new system state. For each meteorological variable v, we estimate the value of that variable at the current time t at some grid cell (i, j) as a linear combination of the *prior* system state at the same cell:

$$x_{v,i,j,t} = w_v^T \boldsymbol{X}_{i,j,t-1}, \quad \boldsymbol{X}_{x,y,t-1} = \left[x_{1,i,j,t-1} \ \dots \ x_{V,i,j,t-1}\right]^T \tag{1}$$

Here, w_v is a V-length vector of coefficients. This model is very similar to a VAR(1) model which targets a single grid cell (see below), except that w_v may be regressed using data from *all* system cells and timesteps at once. To do so, we observe the system state at all S cells over T one-hour steps. This gives TS observed $x_{v,i,j,t}$ values for each variable v. These values are arranged in a vector $\boldsymbol{x}_{v,T}$ of size TS and a matrix $\boldsymbol{X}_{T-1}$ of size TS × V. Each row of $\boldsymbol{X}_{T-1}$ is the system state at a single cell just prior to the corresponding time for a row of $\boldsymbol{x}_{v,T}$. This yields the linear system $\boldsymbol{x}_{v,T} = \boldsymbol{X}_{T-1} w_v$, which we solve using the normal equations:

$$w_v = \left(\boldsymbol{X}_{T-1}{}^T \boldsymbol{X}_{T-1}\right)^{-1} \boldsymbol{X}_{T-1}{}^T \boldsymbol{x}_{v,T} \tag{2}$$

Additionally, using the same formulation, a polynomial regression can be computed by adding higher degree terms into the set of features, but we found that doing so only marginally improved performance.

2.2 VAR model

We also developed a set of vector autoregressive (VAR) models in order to simulate and predict the evolution of the meteorological system over time. We adopt the formulation and estimation strategy given in [17] and [18]. The model describes the system state at time t given some number p of prior system states, where p is known as the *lag order* of the model:

$$\boldsymbol{X}_t = \boldsymbol{c} + \boldsymbol{\Pi}_1 \boldsymbol{X}_{t-1} + \cdots + \boldsymbol{\Pi}_p \boldsymbol{X}_{t-p} + \boldsymbol{\varepsilon}_t \tag{3}$$

Here, $\boldsymbol{X}_t$ is the system state at the t^{th} time step, $\boldsymbol{c}$ is a constant offset, and each $\boldsymbol{\Pi}_l$ specifies an $N \times N$ model coefficient matrix which defines weighted sums of the observed data at time step t - l. In a model of lag order p, the new state $\mathbf{X}_t$ is a sum of linear combinations of prior system states $\mathbf{X}_{t-l}$ from $l = 1$ to $l = p$. Finally, $\boldsymbol{\varepsilon}_t$ is a zero-mean noise generator which is uncorrelated between time steps which introduces innovations to the model.

This system can be solved for the unknown $\boldsymbol{\Pi}_l$ matrices using ordinary least squares by rewriting it as a system of equations where each row is the equation for a variable v at a particular grid cell for one time step:

$$\boldsymbol{x}_{v,i,j} = \boldsymbol{Z}\boldsymbol{\pi}_{v,i,j} + \boldsymbol{e}_{v,i,j} \tag{4}$$

Since we train on T time steps, $\boldsymbol{x}_{v,i,j}$ is a T-length vector giving at each time step the values for variable v at a single grid cell (i, j). For example, $\boldsymbol{x}_{v,i,j}$ might represent the atmospheric pressure at (20° N, 34° W) over a 24 hour period. Z is a $T \times k$ matrix ($k = Np + 1$), where the t^{th} row is defined as $Z_t = (1, \boldsymbol{X}^{\mathrm{T}}_{t\,-1}, \boldsymbol{X}^{\mathrm{T}}_{t-2}, \dots , \boldsymbol{X}^{\mathrm{T}}_{t-p})$, e_i is a zero-mean noise vector, and $\boldsymbol{\pi}_i$ is a k-length vector of coefficients to be determined. Using a typical least squares approach, we estimate each value of $\boldsymbol{\pi}_{v,i,j}$ over $v = [1, \dots, V]$, $i = [1, \dots, s_x]$, and $j = [1, \dots, s_y]$, yielding our estimation for the full model:

$$\hat{\boldsymbol{\Pi}} \in (k \times N) = \begin{bmatrix} \hat{\boldsymbol{c}} \\ \hat{\boldsymbol{\Pi}}_1^{\mathrm{T}} \\ \vdots \\ \hat{\boldsymbol{\Pi}}_p^{\mathrm{T}} \end{bmatrix} = \left[\hat{\boldsymbol{\pi}}_{1,1,1} \dots \hat{\boldsymbol{\pi}}_{V,s_x,s_y}\right] \tag{5}$$

2.3 Lasso

Particularly for large systems or lag orders, there should significant sparsity in the VAR model, as we expect the true system dynamics to have a somewhat parsimonious description. We use lasso regularization [19] to enforce sparse feature selection. This changes our least squares problem slightly. We now wish to minimize the quantity

$$\sum_{v=1}^{V}\sum_{i=1}^{s_x}\sum_{i=1}^{s_y}\left(\boldsymbol{x}_{v,i,j} - \boldsymbol{Z}\boldsymbol{\pi}_{v,i,j}\right) + \lambda\sum_{v=1}^{V}\sum_{i=1}^{s_x}\sum_{i=1}^{s_y}\left\|\boldsymbol{\pi}_{v,i,j,}\right\|_1 \tag{6}$$

Here, λ is the Lagrange multiplier on the lasso constraints. This is a convex optimization problem and is solvable using standard tools. Cross-validation testing showed a value of about $\lambda = 0.05$ gave the most consistent performance when lasso regularization was used. Models grew overly sparse at higher values.

Prior updates for this project proposed that a *fused* lasso model [20] may also be employed to constrain seasonal variations of the model over an annual cycle, but we decide against including this component due to time constraints.

2.4 Evaluation

We evaluate and compare the predictive performance of several model variations, including the baseline linear regression and VAR models of lag orders $p = 1$, 2, and 10. Note that the baseline linear regression and VAR(1) single-cell models are structurally equivalent, but are regressed using slightly different inputs and techniques, as described previously.

Since the VAR models encode a linear coefficient between all system terms at a given timestep, their complexity, as measured by the number of free parameters, scales as the *square* of the size N of the system state. Thus, we found it helpful to examine the performance with VAR models trained either on a *single* grid cell ($S_{single} = 1$) or on *multiple* grid cells ($S_{multi} > 1$). Figure 2 shows the difference between these two variations.

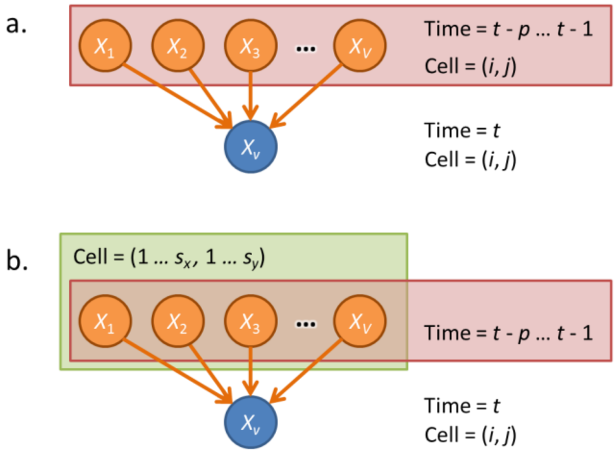

Figure 2. Variables used for estimating system state $\boldsymbol{X}$ at time t at cell (i, j);
a. Single cell model; b. Multi-cell model.

The success of each model depends on its ability to correctly forecast (predict) the new system state $\boldsymbol{X}_t$ at time t based on system states at prior timesteps. Since there is some inherent noise in the variables of the system, the primary goal is to minimize the residual of predicted system state $\widehat{\boldsymbol{X}}_t$ relative to the actual historical observation $\boldsymbol{X}_t$ to within some reasonable percentage of this variation. The size of our data set allows us to train on a subset of the full data and test on another independent set.

3 Results

A small region around Pittsburgh, PA was selected as the data set for testing. It spans a 6 x 6 grid of cells, each cell being 10 km to a side. Hourly records for each cell were used from 9/15/2010 to 10/15/2010. Our data is thus comprised of $S = 36$ grid cells, each containing $V = 10$ meteorological variables of interest measured over $T = 720$ hours, yielding 259,200 individual data values. It is worth noting that the multiple-cell VAR models include a factor of roughly $S^2 = 1296$ more free parameters compared to their corresponding single-cell VAR models, a very prominent increase in complexity which had implications for model overfitting, as discussed in Section 4.

Each model was trained on a random 20% of the data set and tested on the remaining 80%. We report the RMSE averaged over all system variables and cells included in the test set.

3.1 Model Complexity

The complexity of each trained model is given in Table 2, as measured by the number of non-zero elements in the trained model (the set of w_v vectors for linear regression, or $\widehat{\boldsymbol{\Pi}}$ for the VAR models). The VAR models include both non-regularized and lasso-regularized versions in order to demonstrate the effect of variable selection; the baseline model with lasso regularization showed poor performance and is not given here.

Table 2. Number of non-zero parameters in each model

Input Region	Model Type	# of Non-zero Parameters in Model	
		Non-regularized	Lasso-regularized
Single Cell	Linear Regression (LR)	132	-
	VAR(1)	132	44
	VAR(2)	253	51
	VAR(10)	1,221	81
Multi-Cell	Linear Regression (LR)	132	-
	VAR(1)	56,628	3,068
	VAR(2)	56,232	3,559
	VAR(10)	53,064	5,106

3.2 Performance

The predictive performance of each model on the test data set is shown in Figure 3. Test set RMSE, averaged across all variables for each of the models. Note that 100% RMSE indicates that a model's error magnitude is comparable to the total variance of the system over the test period.

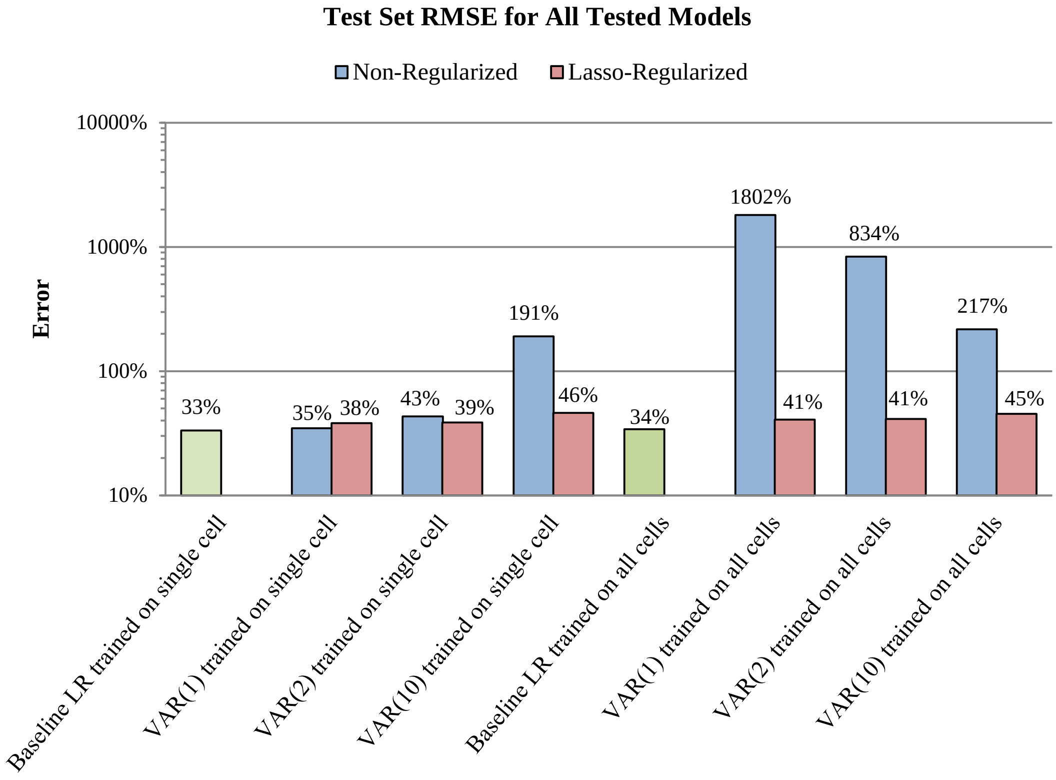

Figure 3. Test set RMSE, averaged across all variables for each of the models.

4 Conclusions

Surprisingly, the results in Figure 3 show that predictive performance was not significantly improved over the linear regression baseline using any of the VAR models. This implies that predicting a weather system's state one hour into the future is no more accurate when using several prior system states compared to using only a single prior state, at least for the class of linear models which we examined. Even with regularization, there is some evidence of overfitting for increasing lag orders, as shown in Figure 4.

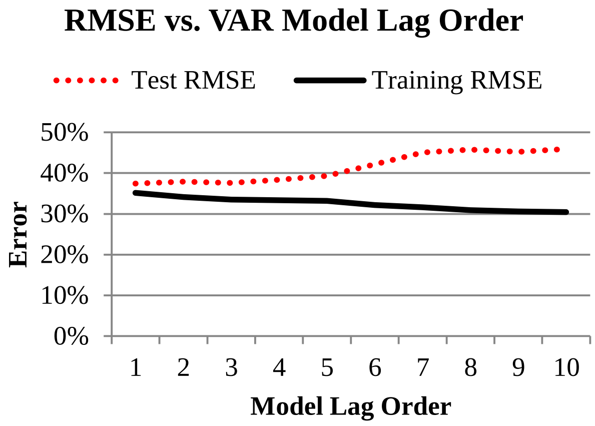

Figure 4. Comparison of test and training errors for lasso-regularized VAR models with lags from $p = 1$ to $p = 10$.

The gap between test and training error increases steadily with the model lag order. During informal testing, this general trend also held true for other values of the lasso Lagrange multiplier, λ. We conclude that weather prediction at this short time scale (hour-by-hour) does not require a long history of prior observations; a single prior timestep is sufficient.

Another finding apparent from Figure 3 is that models which are trained on only a single grid cell (but evaluated on all grid cells) are at least as successful as those trained on all of the available grid cells. For the baseline linear regression model, which has the same complexity regardless of the system size, this suggests that additional data will yield no further performance gains. For the VAR models, which scale in complexity very quickly when modeling larger systems, this indicates that adding coefficient terms between different cells has, at best, no performance effect, or, at worst, adds many irrelevant parameters which degrade the performance of the model.

Finally, we note that predicting variables such as temperature or air pressure is well-served by our linear model, but other variables like precipitation show poor performance regardless of model complexity (see Figure 5). Critically, the former variables exhibit dense, periodic trends, while the latter class includes sporadic phenomena with no clear periodicity. This suggests that, using our approach, effective weather forecasting is possible for phenomena which already show some degree of predictability, but other aspects may require physically-based modeling.

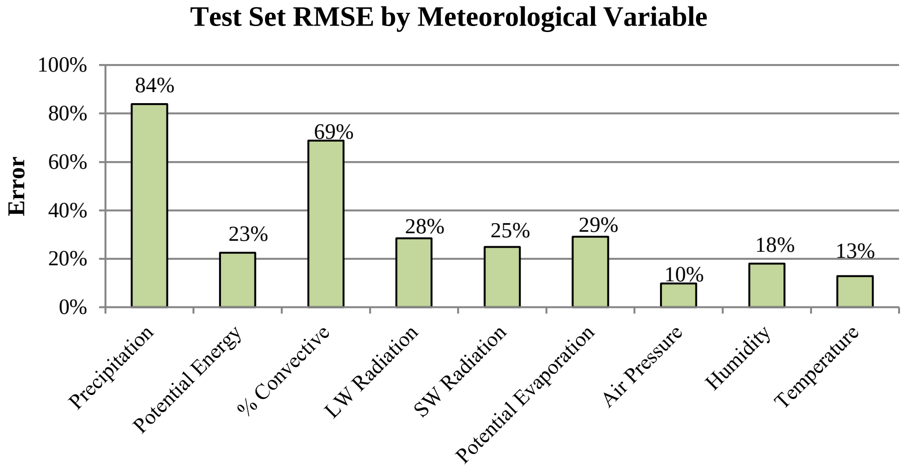

Figure 5. RMSE for a subset of variables predicted using the baseline model

References

[1] Shuttleworth, J. (2012), *Terrestrial Hydrometeorology*, Wiley-Blackwell.

[2] Hong, W. C. (2008), "Rainfall forecasting by technological machine learning methods". *Applied Mathematics and Computation* **200**(1): 41-57.

[3] French, M. N., Krajewsky, W. F. & Cuykendall, R. R. (1992), "Rainfall forecasting in space and time using a neural network". *Journal of Hydrology* **137**(1): 1-31.

[4] Maier, H. R. & Dandy, G. C. (2000), "Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications". *Environmental Modelling & Software* **15**: 101-124.

[5] Nasseri, M., Asghari, K. & Abedini, M. (2008), "Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network". *Expert Systems with Applications* **35**: 1415-1421.

[6] Applequist, S., Gahrs, G. E., Pfeffer, R. L. & Niu, X.-F. (2002), "Comparison of Methodologies for Probabilistic Quantitative Precipitation Forecasting". *Weather and Forecast.* **17**(4): 783-799.

[7] Partal, T. & Kisi, Ö. (2007), "Wavelet and neuro-fuzzy conjunction model for precipitation forecasting". *Journal of Hydrology* **342**: 199-212.

[8] Kisi, O. & Shiri, J. (2011), "Precipitation Forecasting Using Wavelet-Genetic Programming and Wavelet-Neuro-Fuzzy Conjunction Models". *Water Resources Management* **25**(13): 3135-3152.

[9] Harris, D.; Foufoula-Georgiou, E.; Droegemeier, K. K. & Levit, J. J. (2007), "Multiscale Statistical Properties of a High-Resolution Precipitation Forecast". *Journal of Hydrometeorology* **2**(4): 406-418.

[10] Yang, Y., Lin, H., Guo, Z. & Jiang, J. (2007), "A data mining approach for heavy rainfall forecasting based on satellite image sequence analysis". *Comput. and Geosciences* **33**(1): 20-30.

[11] Bartok, J.; Habala, O.; Bednar, P.; Gazak, M. & Hluchy, L. (2010), "Data mining and integration for predicting significant meteorological phenomena". *Procedia Computer Science* **1**(1): 37-46.

[12] Kuligowski, R. J. & Barros, A. P. (1998), "Localized Precipitation Forecasts from a Numerical Weather Prediction Model Using Artificial Neural Networks". *Weather and Forecasting* **13**(4): 1194-1204.

[13] Ramírez, M. C. V.; de Campos Velho, H. F. & Ferreira, N. J. (2005), "Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region". *J. of Hydrology* 301: 146-162.

[14] Hall, T.; Brooks, H. E. & Doswell, C. A. (1999), "Precipitation Forecasting Using a Neural Network". *Weather and Forecasting* **14**(3): 338-345.

[15] Koizumi, K. (1999), "An Objective Method to Modify Numerical Model Forecasts with Newly Given Weather Data Using an Artificial Neural Network". *Weather and Forecast.* **14**(1): 109-118.

[16] National Aeronautics and Space Administration, "Land Data Assimilation Systems (LDAS)", available online at http://ldas.gsfc.nasa.gov/index.php.

[17] Lutkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

[18] Zivot, E. and Wang, J. (2006), "Vector autoregressive models for multivariate time series", in *Modeling Financial Time Series with S-PLUS*, pages 385–429. Springer New York.

[19] Tibshirani, R. (1994), "Regression shrinkage and selection via the lasso". *Journal of the Royal Statistical Society, Series B* **58**: 267-288.

[20] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005), "Sparsity and smoothness via the fused lasso". *Journal of the Royal Statistical Society: Series B* **67**(1): 91-108.