Weather forecasting using probabilistic graphical models

Felipe Hernández[[1]](#footnote-1) and Amos Ng[[2]](#footnote-2)

|  |  |
| --- | --- |
| Professor: Eric Xing | February 10, 2014 |

# Overview

Weather forecasts are broadly used, from personal activity planning, to large-scale economic decision-making, to emergency preparation and response. 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 and regional scales are used to forecast future conditions. In this project, we will instead take a machine learning probabilistic approach focused on a regional scale, and attempt to predict atmospheric variables at specific geographic locations.

The forecasts will be based on prior atmospheric states in the neighborhood of the selected location. In particular, we will attempt to predict the probability distribution of variables such as pressure, precipitation, and temperature based on the previous values of these variables as provided online on assimilated land products by NASA.

# Literature survey

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. Machine learning approaches have been explored to construct simplified models based on atmospheric measurements, but are not very popular among meteorologists.

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 [1]. 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. Other algorithms used in weather forecasting include linear regression, discriminant analysis, logistic regression [4].

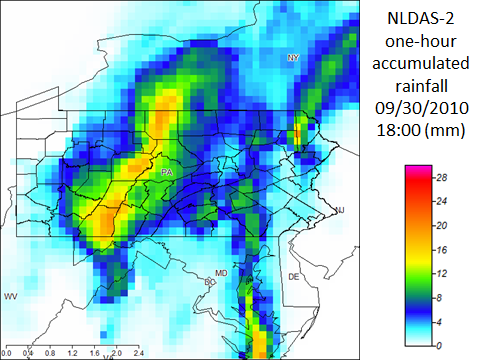
Recently, meteorological observations from satellite and Doppler radar have become widely available, providing ubiquitous coverage at the cost of decreased precision. These sources of information add spatial dimensions to the forecasting problem. For example, Fourier spectrum, structure function, and moment-scale analyses are used to understand radar precipitation in [5]. 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 [6] and [7].

# Data sets

As mentioned previously, there are multiple sources of meteorological information available online. Usually these sources are provided by government agencies such as NOAA and NASA, and have different levels of post-processing of raw data from gauges, land radars, and satellites.

The North American Land Data Assimilation System (NLDAS), a service hosted by the Goddard Earth Sciences Data and Information Services Center at NASA, is a multi-variable source of information produced through the assimilation of land measurements. We will focus mainly on this dataset due to both its relative high precision and the multiple variables it reports: Precipitation, atmospheric pressure, humidity, temperature, wind speed, convective potential energy, and radiation flux. The NLDAS product provides hourly weather data for the US beginning in 1980. Each sampled time in the data set includes the following values per cell in a grid of resolution 1/8 of a degree in both latitude and longitude directions.

The following image shows an example of a rainfall image from a severe storm in 2010 on the state of Pennsylvania. Each cell in the rainfall image has a size of roughly 10 km x 10 km. More information on the NLDAS-2 data products can be found on the following URLs: <http://ldas.gsfc.nasa.gov/index.php> and <http://disc.sci.gsfc.nasa.gov/hydrology/data-holdings>.



# Activity plan

We plan to use different probabilistic graphical models to try to determine the probability distributions of weather variables maps given the observed values from previous time steps. We will first use variable discretization so that discrete models can be used. Afterwards, we will attempt using models that are able to handle continuous variables. Similarly, we will first use simple graph representations, such as Naïve Bayes networks. Later, we will attempt to use models with more complex topologies that take into account the interdependencies of variables.

Some of the activities that we will perform include:

* Selection of the dataset and the study area
* Data download and pre-processing
* Variable discretization strategy
* Naïve Bayes formulation
* Review of probabilistic graphical models for continuous variables

# References

1. 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.
2. Hong, W. C. (2008), “Rainfall forecasting by technological machine learning methods”. Applied Mathematics and Computation 200(1): 41-57.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.

1. felipeh@andrew.cmu.edu [↑](#footnote-ref-1)
2. ajng@andrew.cmu.edu [↑](#footnote-ref-2)