



2020 - STUDY

MONGOLIA'S RAINFALL FORECAST USING RSTUDIO

A study to help small farmers in emerging countries
fight climate change

Mongolia's Rainfall Forecast using Rstudio

Morais, Manoela N. Moraimn@hotmail.com

Abstract—One of the essential sectors of Mongolia's economy is agriculture, which is sensitive to climate variation. The most important climatic element which impacts agriculture productivity is the rainfall. Therefore, rainfall prediction becomes an important issue in Mongolia, especially for small farmers that lack technological resources. In this paper, I propose to model the daily rainfall prediction over Dornod Province to help farmers to forecast rain over 1 year. I applied several modeling methodologies, which include time series regression and Machine learning. The method that presented the lower MAE and RMSE was the Arima Model, followed by the default Simple Times series.

Keywords—Time Series Regression, Rainfall Forecasting, Statistical forecasting, Mongolia weather, Farming Technology

I. INTRODUCTION

The agricultural industry is one of the most vulnerable to climate change as it directly depends on rainfall and temperature. Rainfall information is normally used by farmers for crop selection and water resource management in agriculture. The occurrence of prolonged dry periods or heavy rain at the critical stages of crop growth and development may lead to a significant reduction of the farmer's crop yield. Although climate change affects differently different crops and regions, it is expected that a decrease in farming productivity (Lobell, 2007). In fact, some decline can already be seen. In the last years, developed countries with a good farm infrastructure had more than 17% yield decrease because of climate changes (Nelson et al, 2014).

Climate variability is a major source of risk for agriculture and food systems and its changes are making it harder for the farmers to guess/predict the weather patterns by just looking for last year's results. The climate change impact is even more accentuated to small-low-income farmers in emerging countries, given their small scale, the burden of economic exclusion, and lack of access to critical resources. Around 500 million smallholder farms in the developing

world are supporting almost 2 billion people, and 70 % depending partly or completely on agriculture for their livelihoods (IFAD, 2011). Those farmers do not have adequate networking or tools from which they can learn and easily exchange know-how. Therefore, temperature and rainfall prediction become a significant factor in agriculture to reach the best crop performance.

To this study, I selected the Dornod province, in Mongolia, and an emerging country, which its economy is largely based upon agriculture. The focus of this study is to provide to the small-house-hold-farmers the rainfall forecast with a good prediction and low resources.

1.2 Research Goal

The main goal of this research is to propose a solution to help farmers selecting their crop and its start farming period by giving them useful rainfall prediction. Therefore, I am going to test several simples' models to identify one that can predict with lower errors the rainfall forecast for Dornod Province.

Dornod Province		References:
Soum (sum/district)	Climate	North
1 Chuluut Khoroosui Province	wet	Central
2 Durbatal	wet	South
3 Bayanbulag	wet	
4 Bayanbulag (Gavanshiy)	wet	
5 Gershiy	wet	
6 Tsagaan Ovoo	dry	
7 Sengelen	dry	
8 Chuluut	dry	
9 Bayanbulag (Togov)	dry	
10 Cherkon	dry	
11 Bulgan	dry	
12 Halanbulag	moderately mild	
13 Matal	moderately mild	
14 Khaltset	wet	

Figure 2. Dornod Provinces used for weather model study

DATA

I am using 5 years of meteorological data between January 2015 and May 2020 from the Nasa website from Dornod Province (see all Soums in Figure 01). From the Nasa available indices, a set of 14 indices of climate variables were selected for this study (Figure 2). For long-range forecast (LFR), which is more than 2 weeks and up to 2 years, generally, the rain forecast is consistent with fundamental variables such as previously temperature and precipitation (Doblas-Reyes & all, pp. 12). However, we are using other variables to identify the interference and importance in the rainfall prediction results.

Of the selected variables, seven indices refer to temperature, one to precipitation, and 4 to other variables. The temperature indices describe cold extremes as well as warm extremes. The precipitation indices describe wet extremes.

Parameter	Units	Description
T2M	°Celsius	Temperature at 2 Meters
T5	°Celsius	Earth's Temperature
rs	mm	Surface Rainfall
T2M_NH	°Celsius	Temperature at 2 Meters
SLP	hPa	Sea Level Pressure at 2 Meters
WSW	m/s	Wind Speed at 2 Meters
ALBID_TGA_2M_2M_NH	W/m²	Top-of-atmosphere Radiation
ALBID_SF_2M_2M_NH	W/m²	Surface Radiation at 2 Meters
PRECIP	mm/day	Precipitation
T2M_NH2M	°Celsius	Temperature at 2 Meters
T2M_NH1	°Celsius	Temperature at 2 Meters
T2M_NH2	°Celsius	Temperature at 2 Meters
ALBID_SF_2M_2M_NH	W/m²	Surface Radiation at 2 Meters
T2M_NH3	°Celsius	Temperature at 2 Meters

Figure 1. 14 Indices selected from the Nasa climatology website available indices.

II. Modeling

The methodology used will be first, get historical data and random from the Nasa website. If the model is not times series related, I will divide it in 2 categories in 80% for training (build the models) and 20% for test the models (Figure 04). If the model contains time-series I will divided for the training 2015-2019 data and 2020 results will be used for testing. Then, I am going to run several regression models for rain forecast for autoregressive and multivariable regressions. Finally compare RMSE and MAE results and select the one, that present lower results. For illustration of the models I going to represent initially Dashbalhar results, for the others soum please vide table at attachment.

A quantitative forecast of rainfall is extremely difficult and realizable. Generally, it is done only a couple hours of their occurrence with a Doppler. However, for agriculture operations, a quantitative forecast of rain is not as important as a forecast of the (i) non-occurrence of rains and (ii) type of rain spell that can be expected. Therefore, after modeling and get the results in "millimeters of rainfall", I am going to classify the rainfall based on the USG definition of raining (Figure 03). USG defines that Absent of raining ≤ 0 mm/hh, Slight rain ≤ 0.5 mm/hh, Moderate

$\leq 4\text{mm/hh}$ and heavy rain (High) $> 8\text{mm/hh}$ rainfall. Although the information is given by USG in mm/hour, I am going to consider the same value for the whole day (mm/day) to classify. The main goal of doing this classification is to give farmers the information they need and decrease forecast errors. The main goal of doing this classification is to give to farmers the information they need and decrease the forecast errors.

Classification from website		
No rain	0 - 0.001	mm/hour
Slight rain:	0.001 - 0.5	mm/hour
Moderate rain:	0.5 - 4	mm/hour
Heavy rain	4 - 8	mm/hour

Figure 3. Rain data classification extracted from the USG, Gov Resource: <https://www.usgs.gov/mission-areas/water-resources>

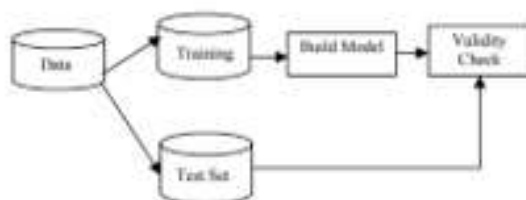


Figure 4. Overview of the forecasting Planning methodology

III. Single Regression Models

The autoregressive process is a regression on itself. Therefore, Y_t is a linear combination of the p most recent past values of itself plus the term "et" that incorporates the error (Equation 01). We are going to start this stage by establishing a Baseline model, which is basically the mean of the historical data, to be used as a reference and compared with the linear Times series model, Times series *Seasonal naïve Method*, and *Arima model*. The $Y(t)$ will be rainfall (PRECTOT).

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

Baseline Model

Assuming...that all variables are independent and there is no rainfall prediction (no time series), I constructed a dependent guess (not random) based on the mean of the value itself, using the training data. The Results obtained were RMSE= 2.187456 and MAE= 1.190749, these results will be used as a reference to compare with other models.

```
> Base Model <- mean(trainSPRECTOT)
```

Times Series linear model (TS)

The rainfall observations collected through daily data is sequential over time. Therefore, we are going to use time-series to model the stochastic mechanism that gives rise to an observed series. I am going to predict and compare it with the test set. If one of those models has the best fit, I will forecast it. We will start by analyzing the data extracted from NASA, if the variables' time series are (non) stationarity and possess a unit root.

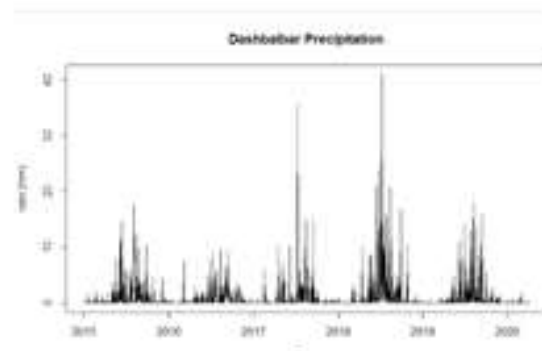


Figure 5. Precipitation for ~~Qashqabazar~~ through the last 5 years. Extracted from RStudio. We can see times series and a seasonal trend.

Checking UnitRoot

To make statistical inferences about the structure of a stochastic process on the basis of an observed record of that process,

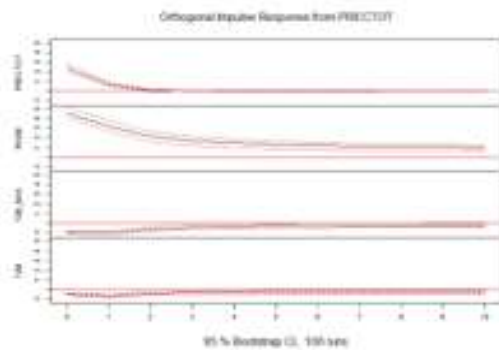


Figure 16. Impulse response for VAR Model

V. Machine Learning Models

Neural Network Model – library ([nnet](#))

The Artificial Neural Networks (ANN) are a set of algorithms, designed to recognize patterns. They interpret sensory data through a kind of machine perception, labeling or clustering raw input. Artificial ANNs have become very popular, and prediction using ANN is one of the most widely used techniques for rainfall forecasting.

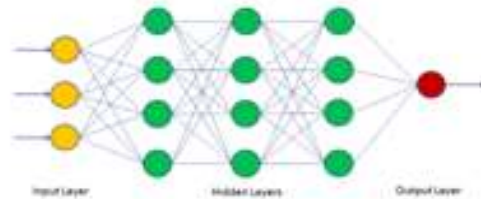


Figure 17. Artificial Neural Network [structure](#) Source: <https://www.datacamp.com/community/tutorials/neural-network-models-r>

To perform the Neural Network forecast I used the [nnet](#) function in the [forecast](#) package for R, that fits a neural network model to a non-linear time series. Used with big data sets. The NN model is organized in multiples layers, the simplest networks contain no hidden layers and are

equivalent to linear regressions. The coefficients attached to these predictors are called “weights”. The forecasts are obtained by a linear combination of the inputs.

```
forecast method: nnet(17,3,100[100])
Model Information:
Average of 10 networks, each of which is
a 18-10-5 network with 201 weights
options were: nlinear output units

Error measures:
Training set: -0.0213389 1.422740 0.0147008 RMSE 2.47 0.1001360 0.0000000
```

Regression trees

Regression tree for continuous outcome variables, is a simple and popular machine learning algorithm. In contrast with previous linear models it makes no assumptions about the relation between the outcome and predictors. It is the basis of a very powerful method that we will also use in this tutorial, called random forest



We can interpret that as when the downward Thermal Infrared (longwave) is lower than 26 we have lower than 0.23mm of precipitation (71% of the days). In the same way when the downward Thermal Infrared is between 26-31 and the humidity is lower than 65% (16% of the days), predict a wet day of 0.69mm.

Entry

Nasa Data .Retrieved from: <https://power.larc.nasa.gov/data-access-viewer/>

Alexy, V. (2018). USA advanced retail sales 24 month forecasting analysis. *RPubs*. Retrieved from: <https://www.rpubs.com/alev2301/421553>

Das, H. P., Doblas-Reyes F. J., Garcia, A., Hansen, J., Mariani, L., Nain, A., Ramesh, K., Rathore L. S. & Venkataraman, R. Weather and Climate Forecasts for Agriculture. *Agrometeorology*. Retrieved from: http://www.agrometeorology.org/files-folder/repository/gamp_chapt4.pdf

Pedro M. (2015). Modelling – predicting the amount of rain. *R-bloggers*. Retrieved from: <https://www.r-bloggers.com/part-4a-modelling-predicting-the-amount-of-rain/>

Chai, T. and Draxler1, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. Retrieved from: <https://www.geosci-model-dev.net/7/1247/2014/gmd-7-1247-2014.pdf>

FAO. retrieved from: <http://www.fao.org/3/y2722e/y2722e0y.htm>

Lobell D.B., Field C.B. Global scale climate–crop yield relationships and the impacts of recent warming. *Environ. Res. Lett.* 2007;2:014002. doi: 10.1088/1748-9326/2/1/014002. Retrieved from: <https://iopscience.iop.org/article/10.1088/1748-9326/2/1/014002>

IFAD. 2011. Rural poverty report 2011. New realities, new challenges: new opportunities for tomorrow's *génération* (available at <http://www.ifad.org/rpr2011/report/e/rpr2011.pdf>).

Nelson, G., van der Mensbrugghe, D., Abammad, H., Blanc, E., Calvin, K., Hasegawa, T., Havlik, P., Heyhoe, E., Kyle, P., Lotze-Campen, H., von Lampe, M., Mason d'Croz, D., van Meijl, H., Müller, C., Reilly, J., Robertson, R., Sands, R., Schmitz, C., Tabeau, A., Takahashi, K., Valin, H. & Willenbockel, D. 2014b. Agriculture and climate change in global scenarios: why don't the models agree? *Agricultural Economics*. 45(1): 85–101.