CS460 project midway presentation:Weather forecast by time series forecast method

Haraprasad Dhal(sps 18)   Ravi Prakash Singh(sps 18)
Instructor: Dr.Subhankar Mishra

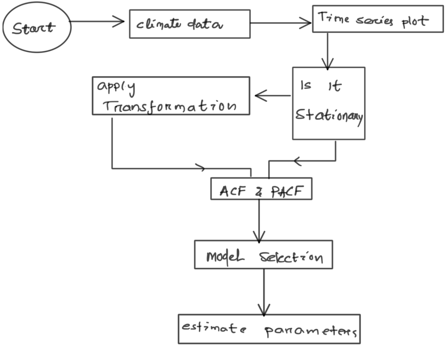
NISER,Bhubaneswar

October 17, 2021

ARIMA based daily weather forecasting:A case study for Varanasi,NIKITA SHIVHARE,ATUL KUMAR RAHUL,SHYAM BIHARI DWIVEDI and PRABHAT KUMAR SINGH DIKSHIT

•Varanasi:65 Years of daily meteorological data (rainfall,min temp ,max temp) from IMD
•(1951-1995):Training set ,monitoring.(1995-2015)validating set,testing
• They worked on ARIMA(2,0,2) for rainfall and ARIMA(2,1,3) for temperature data.With root mean squared error values 0.0948 and 0.085 for rainfall data and temperature data respectively.This accuracy shows that their algorithm worked successfully.

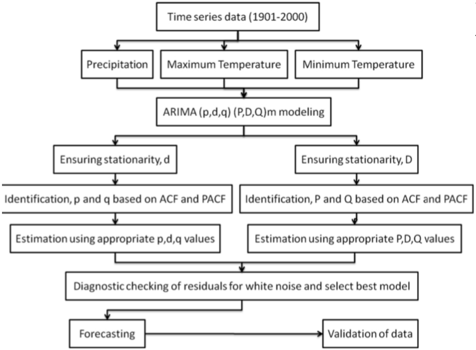
1. plotting data as times eries plot
2. Check data for any trend and seasonality
3. predict the model ARIMA(p,d,q)
4. Apply that ARIMA(p,d,q)
5. Forecasting



Paper-2

Time series analysis of climate variables using seasonalARIMA approach,TRIPTI DIMRI,SHAMSHAD AHMAD andMOHAMMAD SHARIF. j.Earth Syst.Sci.(2020)129 149

- Time series and seasonal analysis of monthly mean minimum and maximum Temperature and the precipitation for Bhagirathi River basin(Uttarkashi and Teheri)
- Box -jenkins approach to find a good model for forecasting



- Their result :-
Precipitation :SARIMA(0,1,1)$(0,1,1)_{12}$
Temperature:SARIMA(0,1,0)$(0,1,1)_{12}$
- The forecast result for precipitation were found to over predict for extreme rainfall events(For both Below and above normal precipitation) Thats they got high RMSE even though in normal rainfall days the forecast and observed data do agree.
- Forecast result for temperature in good agreement.Increasing trend for Teheri station(767 m) and Decreasing trend for Uttarkashi station(1071 m)

Paper 3

Int. Agrophys.,2018,32,253-264 Forecasting daily meteorological time series using ARIMA and Regression models. Małgorzata Murat, Iwona Malinowska, Magdalena Gos, and Jaromir Krzyszczak

- Daily air temperature and precipitation at Four European sites(Jokioinen, Dikopshof, Lleida and Lublin) .Jan 1 1980 to Dec 31 2010.Prediction from 2005 to 2011
- Box-Jenkins Approach
- ARIMA with an external regressors in the form of Fourier Terms.These would allow seasonal parameters to added to the Data.The Model looks like as follows:ARIMAF(p,d,q)[K]

$$Y_t = c + \sum_{l=1}^{K}[\alpha_l sin(\frac{2\pi lt}{m}) + \beta_l cos(\frac{2\pi lt}{m})] + U_t \tag{1}$$

U_t is an ARIMA process ,m is length of period ,The Value of K is chosen by minimising forecast error measures.

- Temperature

Site	Model	RMSE
Dikopshof	ARIMAF(3,0,1)[k=1]	3.633
	SARIMA(0,0,1)(0,1,0,365)	4.914
Jokioinen	ARIMAF(3,0,2)[K=7]	4.536
	ARIMAF(2,0,3)[K=7]	4.536
	SARIMA(3,0,1)(0,1,0,365)	5.582

- PRECIPITATION

Site	Model	RMSE
Dikopshof	ARIMA(2,0,1)	3.749
	ARIMA(1,0,3)	3.749
	ARIMAF(3,0,1)[k=7]	3.581
Jokioinen	ARIMAF(3,0,1)[K=7]	3.581
	SARIMA(3,0,0)(0,1,0,365)	5.903

Stationarity of data And its Importance.

Stationarity is a statistical property.

- Constant mean
- Constant Variance
- No Seasonality
  - Repeating trend or pattern over time.

To convert a Non-Stationary Data into Stationary.
There are different ways to kill Non-Stationarity of different types.

- Differencing , when Y(t) is following a linear trend. Y(t)=Y(t)-Y(t-1)
- Sometime we need to do multiple time Differencing.
- Sometimes we take Y(t)=LOG(Y(t)) in case where Y(t) is of exponential type forms.
- Seasonal differencing Y(t)=Y(t)-Y(t-1)

MODEL

A simple combination of Auto regression and Moving Average model
Auto Regression: Uses Past Values to make a prediction

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \longrightarrow \mathrm{AR}(1)$$

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t \longrightarrow \mathrm{AR}(2)$$

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_P y_{t-P} + \varepsilon_t \longrightarrow \mathrm{AR}(P)$$

Moving Average : Uses Past errors to make a prediction

$$y_t = \beta_0 + \theta_1 \varepsilon_{t-1} + \varepsilon_t \longrightarrow \mathrm{MA}(1)$$

$$y_t = \beta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varepsilon_t \longrightarrow \mathrm{MA}(2)$$

$$y_t = \beta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_Q \varepsilon_{t-Q} + \varepsilon_t \longrightarrow \mathrm{MA}(Q)$$

ARMA model:

$$y_t = B_0 + B_1 y_{t-1} + ... + B_P y_{t-P} + \theta_1 \varepsilon_{t-1} + ... + \theta_Q \varepsilon_{t-Q} + \varepsilon_t \longrightarrow \text{ARMA(P,Q)}$$

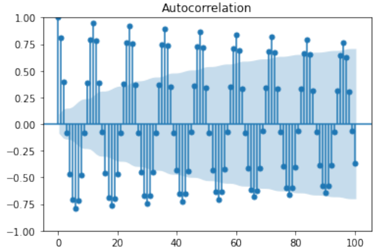
- ε_n is the error in y_n prediction.
- $\beta_i and \theta_j$ are coefficients.

Finding parameters for model

We have used ACF plots, PACF plots and AIC scores to set the model function and find p,q,d,P,D,Q and s values.
We use ACF And PACF plots, which measure the correlation between current time period and previous time lags.

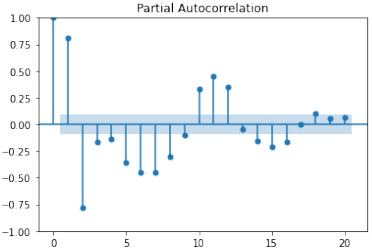
Auto Correlation Function

Measure direct and indirect effect of previous time lags on current value.Used to find order of Moving Average Model.



Partial Auto Correlation Function

Measure only direct effect of previous time lags on current value value.Used to find order of Auto Regressive Model.



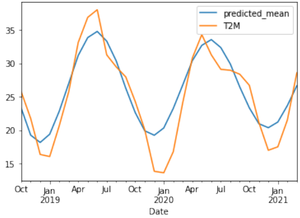
AIC

We mainly used AIC score for model fitting upto ARIMA model but for SARIMAX we used the ACF and PACF plots to set the model function.AIC lets us to test how well our model fits the data set without over-fitting it. The AIC score rewards models that achieve a high goodness-of-fit score and penalizes them if they become overly complex.

Predictions

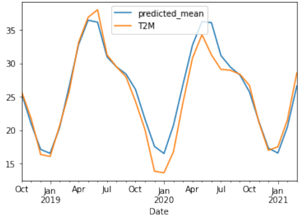
Monthly Average Temperature Prediction using ARIMA(4,0,4).

Trained for 450 months and Tested for next 30 months. It has a RMSE of 2.935 for Mean temp being 25.051 and AIC = 1878.252



Monthly Average Temp Prediction SARIMA((2,0,2),(3,0,3,12)).

Trained for 450 months and Tested for next 30 months. It has a RMSE of 1.820 for Mean temp being 25.051 and AIC = 1456.189



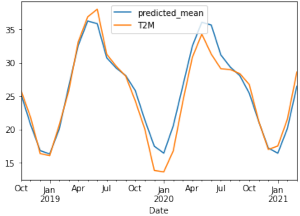
Our model function look like this but we have to put P=4 and Q=4 for ARIMA

$$y_t = B_0 + B_1 y_{t-1} + \ldots + B_P y_{t-P} + \theta_1 \varepsilon_{t-1} + \ldots.. + \theta_Q \varepsilon_{t-Q} + \varepsilon_t$$

Prediction

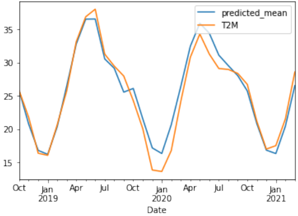
Monthly Average Temp Prediction SARIMAX (Surface Pressure)((2,0,2),(3,0,3,12)).

Trained for 450 months and Tested for next 30 months. It has a RMSE of 1.75 for Mean temp being 25.051 and AIC = 1450.676



Monthly Average Temp Prediction SARIMAX (Precipitation)((2,0,2),(3,0,3,12))

Trained for 450 months and Tested for next 30 months. It has a RMSE of 1.658 for Mean temp being 25.051 and AIC = 1405.498



Conclusion and Future Plans

Mean of Temperature over Test data is 25.051 celcius.

Monthly Average Temp Prediction			
TYPE	**MODEL**	**AIC**	**RMSE**
ARIMA	ARIMA(4, 0, 4)	1878.25	2.935
SARIMA	SARIMAX(2, 0, 2)x(3, 0, [1, 2, 3], 12)	1456.19	1.821
SARIMAX (Precipitation)	SARIMAX(2, 0, 2)x(3, 0, [1, 2, 3], 12)	1405.50	1.658
SARIMAX (Surface Pressure)	SARIMAX(4, 0, 4)x(3, 0, [1, 2, 3], 12)	1450.68	1.756

We see here that the model SARIMAX with exogenous variable as Precipitation is most accurate of all.
Using SARIMAX model we were unable to tune it directly using AIC value so plan to do it.
Our prediction for day to day average temperature using any of the above model was not good. We are thinking to use RNN (mainly LSTM) and see how it works.
We will try to change the model function of SARIMAX and make it to depend on past values of Exogenous variable.


thanks!