Arima Modelling Based Relative Humidity Prediction Analysis

Namratha V,
Civil Engineering Department,
UVCE, JB Campus,
Bangalore University, India.

Usha N Murthy,
Professor and Head, Civil Engineering Department,
UVCE, JB Campus,
Bangalore University, India.

ABSTRACT

Extreme climate and weather events are progressively being renowned as key aspects of climate change. In general, the frequency of occurrence of hot days and hot nights showed widespread increasing trend, while that of cold days and cold nights has shown widespread decreasing trend. However, the results generally indicate that the daily maximum and minimum temperatures are becoming less variable within the season. Climate change is a problem of global proportion, requiring equally global responses. Yet climate change is a fundamental urban issue. This current research work mainly focuses on effect of changes in relative humidity. To attain this aforementioned aim of research work, the humidity data is obtained from Karnataka pollution control board (KSPCB) from 1980 to 2019. Advanced machine learning techniques such as ARIMA model and ANN are incorporated in this research and the humidity prediction is made till 2030. MATLAB is the forecasting software used for predicting the meteorological data till 2025.The version used in this research is MATLAB 2014b and this work is compatible with MATLAB 2013a and above version.

Keywords--- ARIMA, climate change, humidity, KSPCB, machine learning.

1. INTRODUCTION

Moisture problem is one of the most serious factors in building and housing industry. Over the last decade, moisture failures in building systems have reached billions of Euros in damages in Europe, many of which involved the deterioration of sheathing panels and energy efficiency. Additionally, excess moisture in envelopes can lead to the presence of molds which results in poor indoor air and causes health problems of the inhabitants [1]. Thus indoor moisture prediction becomes the part of import work prior to indoor air quality control. Over the decades, many researchers have devoted to such modelling topics. There are many models available. In our laboratory for instance, an accurate numerical model of coupled heat and moisture transfer in buildings has been developed [2]. More detailed and complicated models are Navier–Stokes equations which describe the flow of fluids for airflow, temperature and contaminant distributions. A computational fluid dynamics (CFD) technique is employed to handle these equations. Teodosiu et al. employed a CFD technique and a modified k–e turbulence model to predict indoor air moisture and its transport in a mechanically ventilated test room to estimate the level of thermal comfort. Experimental–numerical comparisons with regard to thermal comfort indices were also provided [3]. The model is very useful in studies dealing with thermal comfort predictions where an exact distribution of indoor air moisture is required. These numerical methods, called physical models, can simulate inside climate environment and airflow distribution even before building is constructed. This is one of their advantages. However, these numerical methods typically require a lot of computation and lead to time-consuming simulations [9]. Take CFD models as an example. Although CFD models can give highly detailed results, the implied accuracy of the results is defined by the assumptions inherent in the model setup, thus, there is the potential of a very costly and refined computation. For a medium-size building, it may take days to complete indoor temperature simulation in a modern personal computer (PC) [4]. Therefore, most of the CFD models are limited to steady state calculations as the model developed. Most importantly, a general drawback of these models is that the output of the model is only as accurate as input physical data, for example airflow rate was needed in the CFD model developed. Presently there are many buildings whose input physical data are poorly defined, which creates ambiguity or uncertainty in predicting and interpreting the output. Physical models fail to account for these complicated cases. For example, in a central ventilation control room where inside climate gets great impacts by ventilation machine activities, it is almost impossible for physical models to simulate indoor climate because too many unknown factors are involved [10]. On the other hand, a black-box model, such as neural networks, can deal such extreme cases without much difficulty. Unlike physical models, neural networks entirely depend on experimental data which can be made adaptive and offer a much faster computation. Compared to physical models, a neural network takes just a few minutes to finish indoor climate forecast for a medium-size building [5]. Physical and neural network models are complementary. However, indoor humidity prediction with neural network models is lacking in literature due to its more complicated mechanism involved which depends on thermal behaviours or temperature prediction. Sigumonrong et al. used historical data to predict indoor temperature and relative humidity, yet their main focus was indoor relative humidity maintenance rather than prediction. No details were provided on input variable identification and prediction results. Similar work was done by Zhang et al. Concerning indoor temperature prediction, some literatures exist. Ferreira et al. [8] adopted RBF (radial basis function) neural network model to predict indoor temperature for a green house. Using RBF, Ruano et al. predicted indoor temperature for a school building where a genetic algorithm was also employed for searching optimal structure for neural networks. Thomas et al. investigated indoor temperatures for two buildings using feed-forward neural networks. Despite the efforts these works made, there are still modelling issues that have not been touched. First, for both indoor temperature and relative humidity, the actual prediction situations involved in these works are not very complicated. Impact factors can be well identified and unknown factors have little impact on predictions. Secondly, the prediction of indoor relative humidity using neural networks is not detailed. No detailed information is provided on how to identify input variables and how to search optimal structures for indoor relative humidity. Thirdly, most papers on indoor temperature predictions pay great attention on training stage, such as optimal structure search and input variable identification, and give less attention on validation stage [6]. Criteria like MSE (the mean of square errors), MAE (the mean of square errors) and SSE (the sum of square errors) for accuracy test in validation stage are commonly adopted. However, it is insufficient for these criteria to address problems, such as whether the network is uncertain to a particular input as well as over-fitting and under-fitting problems.

2. LITERATURE REVIEW

Prediction of the future values by analyzing Temperature and humidity data is one of the important parts which can be helpful to the society as well as to the economy. Work has been done in this constrain since years. Different techniques have been applied to predict the temperature and humidity and other parameters of weather. Some of the work in this area is as follows: In data mining, the unsupervised learning technique of clustering is a useful method for ascertaining trends and patterns in data. Most general clustering techniques do not take into consideration the time-order of data. Tasha R. Inniss used a mathematical programming and statistical techniques and methodologies to develop a seasonal clustering technique for determining clusters of time series data, and applied this technique to weather and aviation data to determine probabilistic distributions of arrival capacity scenarios, which can be used for efficient traffic flow management [7]. The seasonal clustering technique is modeled as a set partitioning integer programming problem and resulting clustering's are evaluated using the mean square ratio criterion [2]. The resulting seasonal distributions, which have satisfied the mean square ratio criterion, can be used for the required inputs (distributions of airport arrival capacity scenarios) into stochastic ground holding models. In combination, the results would give the optimal number of flights to ground in a ground delay program to aid more efficient traffic flow management. S. Kotsiantis, A. Kostoulas, S. Lykoudis, A. Argiriou, K. Menagias investigate the efficiency of data mining techniques in estimating minimum, maximum and mean temperature values. Using temperature data from the city of Patras in Greece, a Regression algorithm is applied for the number of results. The performance of these algorithms has been evaluated using standard statistical indicators, such as Correlation Coefficient, Root Mean Squared Error, etc. [8] Godfrey C. Onwubolu1, Petr Buryan, Sitaram Garimella, Visagaperuman Ramachandran,Viti Buadromo and Ajith Abraham, presented the data mining activity that was employed in weather data prediction or forecasting. The approach employed is the enhanced Group Method of Data Simple temperature prediction methods mining in the past weather data records produced accurate prediction for development of intelligent control solutions. The problem was closely related to the prediction of the actual weather conditions within the immediate environment of the greenhouse, an intelligent greenhouse collects its own climate data, with time weather records from weather station localized strictly by the greenhouse were mined to the algorithm, increasing the prediction accuracy.

3. RESEARCH METHODOLOGY

The proposed methodology or steps in which the implementation is done will be explained in detail in this section. The source of data and both hardware and software used are examined.

The learning operation is performed in the database source as well as database reference. Here the predictive modeling is deployed with various algorithms like ARIMA Time series modeling; Artificial Neural networks (ANN) and so on.

Since the previous few decades, ANN a voluminous development within the application field of ANN has unfolded new avenues to the forecasting task involving environment connected development.

French et al. (1992), took a pioneering work in applying ANN for rain forecasting, that used a neural network to forecast two-dimensional rainfall, 1h prior to.

Their ANN model used present humidity information, generated by a mathematical humidity simulation model, as an input data. This work is, however, restricted in a very range of aspects. For instance, there's a trade-off between the interactions and also the training time, that couldn't be simply balanced. The amount of hidden layers and hidden nodes appear short, compared with the amount of input and output nodes, to reserve the upper order relationship required for adequately abstracting the method. Still, it's been thought-about because the 1st contribution to ANN's application and established a brand new trend in understanding and evaluating the roles of ANN in investigating complicated geophysical processes.

A series that you can model as a stationary ARMA (p, q) process after being differenced D times is denoted by ARIMA (p, D, q). The form of the ARIMA (p, D, q) model in Econometrics Toolbox™ is

$$\Delta^D y_t = c + \phi_1 \Delta^D y_{t-1} + \ldots + \phi_p \Delta^D y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \quad [1]$$

Where,

$\Delta^D y_t$ denotes a Dth differenced time series, and ε_t is an uncorrelated innovation process with mean zero.

In lag operator notation, $L^i y_t = y_{t-i}$. You can write the ARIMA (p,D,q) model as

$$\phi^*(L) y_t = \phi(L)(1-L)^D y_t = c + \theta(L)\varepsilon_t. \quad [2]$$

Here,

$\phi^*(L)$ is an unstable AR operator polynomial with exactly D unit roots. You can factor this polynomial as $\phi(L)(1-L)^D$, where $\phi(L)=(1-\phi_1 L-\ldots-\phi_p L^p)$ is a stable degree p AR lag operator polynomial (with all roots lying outside the unit circle). Similarly, $\theta(L)=(1+\theta_1 L+\ldots+\theta_q L^q)$ is an invertible degree q MA lag operator polynomial (with all roots lying outside the unit circle).

The signs of the coefficients in the AR lag operator polynomial, $\phi(L)$, are opposite to the right side of Equation 1. Specifying and interpreting AR coefficients in Econometrics Toolbox uses the form in Equation 1.

ARIMA Model Specifications

This example shows how to use the shorthand arima (p,D,q) syntax to specify the default ARIMA(p, D, q) model,

$$\Delta^D y_t = c + \phi_1 \Delta^D y_{t-1} + \ldots + \phi_p \Delta^D y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q},$$

Where,

$\Delta^D y_t$ is a D^{th} differenced time series.

This model can be modified in condensed form using lag operator notation:

$$\phi(L)(1-L)^D y_t = c + \theta(L)\varepsilon_t.$$

By default, all parameters in the created model object have unknown values, and the innovation

Specify the default ARIMA(1,1,1) model:

```
model = arima(1,1,1)
model =
  arima with properties:
    Description: "ARIMA(1,1,1) Model (Gaussian Distribution)"
    Distribution: Name = "Gaussian"
          P: 2
          D: 1
          Q: 1
      Constant: NaN
```

AR: {NaN} at lag [1]
SAR: {}
MA: {NaN} at lag [1]
SMA: {}
Seasonality: 0
Beta: [1×0]
Variance: NaN

The output shows that the created model object, model, has NaN values for all model parameters: the constant term, the AR and MA coefficients, and the variance. You can modify the created model using dot notation, or input it (along with data) to estimate. The property P has value 2 ($p + D$).

This is the number of pre-sample observations needed to initialize the AR model.

It may seem unusual to apply univariate techniques commonly used in business and economic modeling to global temperature data, but there are some persuasive arguments for doing so. Although the strengths and limitations of causal relative to non-causal modeling are well known, it may be useful to restate this debate in the context of modeling climate change. The data requirements for univariate ARIMA modeling are usually less onerous than those for other techniques such as causal modeling, and this is particularly true in the case of climate change.

Causal models of climate change often include a large number of explanatory variables to reflect the complexity of the causal relationships. A model of global temperature change, for example, might include (at least) variables for natural phenomena such as the El Nino Southern Oscillation and volcanic and solar activity, and variables for human influence such as greenhouse gas and aerosol concentrations, as well as changes in stratospheric and tropospheric ozone.

Given that global temperature changes are the product of gradually evolving processes, it is desirable to calibrate these causal models on data that go as far back in time as possible. But observations for some of the causal variables may not be reliable, and may not even be available for period's further back in time. Additionally, mixing variables observed at high frequencies (monthly, daily) with others observed only at low frequencies (annual) will generally require converting the high frequency variables to the lowest variable frequency. Thus potentially important data 'richness' may be lost in causal models.

4. RESULTS

Humidity Prediction

This section determines the humidity prediction analysis during 1980 – 2030. The relative humidity graph for different months is displayed in below Figures (Figure 1, 2, 3, 4, and 5). The humidity prediction from 2020 till 2030 is described (Table 1).

Table 1 Relative Humidity Analysis from January - December, 2019-2030

2020	RH (0830IST)	72	64	61	78	80	88	88	89	86	87	78	77
	RH (1730IST)	34	29	22	44	55	69	68	73	62	71	61	47
2021	RH (0830IST)	78	68	57	67	75	82	85	88	84	87	79	78
	RH (1730IST)	42	35	31	32	44	54	70	68	61	68	58	58
2022	RH (0830IST)	80	60	69	75	79	85	87	82	84	78	80	81

	RH (1730IST)	46	29	35	49	63	69	65	67	67	56	72	57
2023	RH (0830IST)	75	68	68	71	73	86	88	85	85	79	64	74
	RH (1730IST)	42	32	28	31	47	73	72	64	65	56	37	47
2024	RH (0830IST)	76	60	63	69	73	82	85	87	82	83	81	77
	RH (1730IST)	44	23	27	34	47	62	59	65	70	63	62	48
2025	RH (0830IST)	82	73	65	71	76	81	86	88	88	82	72	81
	RH (1730IST)	64	38	25	34	52	61	68	71	69	64	56	57
2026	RH (0830IST)	77	75	69	71	76	83	87	89	85	85	78	80
	RH (1730IST)	38	41	44	37	46	59	67	72	60	70	58	49
2027	RH (0830IST)	75	68	68	71	73	86	88	85	85	79	64	74
	RH (1730IST)	42	32	28	31	47	73	72	64	65	56	37	47
2028	RH (0830IST)	76	60	63	69	73	82	85	87	82	83	81	77
	RH (1730IST)	44	23	27	34	47	62	59	65	70	63	62	48
2029	RH (0830IST)	82	73	65	71	76	81	86	88	88	82	72	81
	RH (1730IST)	64	38	25	34	52	61	68	71	69	64	56	57
2030	RH (0830IST)	77	75	69	71	76	83	87	89	85	85	78	80
	RH (1730IST)	38	41	44	37	46	59	67	72	60	70	58	49

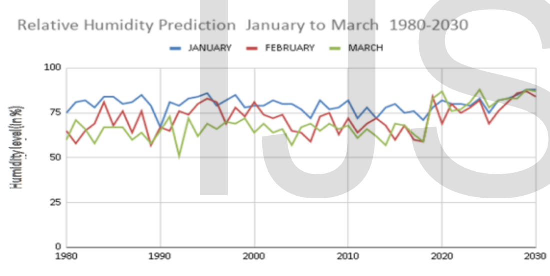

Figure 1: Relative Humidity Prediction from January - March 1980-2030

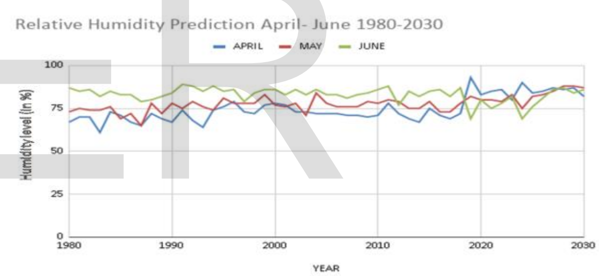

Figure 2: Relative Humidity Prediction from April - June 1980-2030

Humidity prediction analysis for the month of January to March is mentioned for the year 1980 to 2030. The humidity prediction from 2020 till 2030 is found to be changing (Figure 1). If the relative humidity is 100 percent (i.e., dew-point temperature and actual air temperature are the same), this does not necessarily mean that precipitation will occur. Saturation may result in fog (at the surface) and clouds aloft (which consist of tiny water droplets suspended in the air).

Relative humidity prediction is analyzed from April to May and the graph is described above (Figure 1). If the humidity is high, this moisture doesn't evaporate from the body very fast. In areas that are very dry, such as Arizona, the humidity is so low that when sweat occurs, the water evaporates so quickly that they may not even feel it. In this case, they must be careful to stay hydrated because the water loss goes unnoticed. Humidity can also affect plant turgor pressure, which is an indicator of the amount of water in plant cells. When humidity is low, and dew points are in the 50s and low 60s, moisture evaporates from plants very quickly.

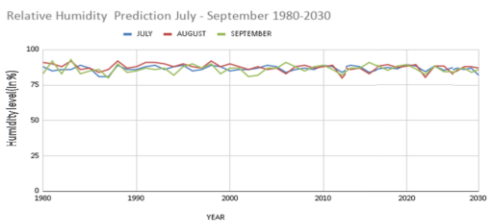

Figure 3: Relative Humidity Prediction from July - September 1980-2030

Relative humidity is analyzed for the month of July to September and it is found that they are more or less equal (Figure 3). If the heat is reduced, the temperature lowers, and the molecules no longer move around as much and they stick. They become visible, turning into water. The water can take the form of a visible plume of steam, like a cloud, or just raindrops that form pools of water on the ground. If the temperature lowers even more, the molecules become closer.

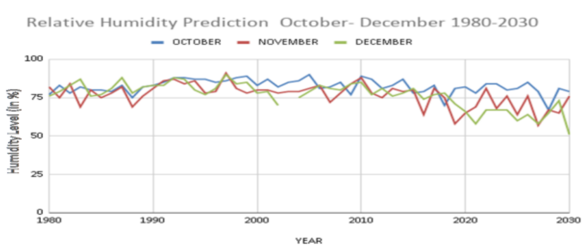

Figure 4: Relative Humidity Prediction from October - December 1980-2030

From October to December, the relative humidity value falls down for the year 2020 till 2030 as per the prediction (Figure 4). When it rains, it will increase the relative humidity because of the evaporation. The air where the rain is falling may not be completely saturated with water vapor. However, the longer it rains, the more the humidity will increase because of the air constantly drawing the water.

Abdurrahman presented the presence of significant decrease trends in relative humidity values which give rise to an application and analysis result that will draw attention to the management of water resources and draw attention to climate change.

5. CONCLUSION

This current research work mainly focused on effect of humidity. The humidity prediction graph is exhibited for past years from 1989 to 2019. With the aid of ARIMA modeling, future humidity prediction is done till 2030 and the values are presented.

Bangalore is a rapidly growing city that has undergone a profound social and economic transformation in the last two decades, associated with the rise of the information technology (IT) industry. This has had a strong impact on the urban fabric, for example by developing new communication infrastructures and the building of new developments on the edge of the city to serve the needs of a rapidly emerging industry and middle-class population. The climatic change occurs due to urbanization, deforestation, etc.

REFERENCES

1. Lu, T., & Viljanen, M. (2009). Prediction of indoor temperature and relative humidity using neural network models: model comparison. *Neural Computing and Applications*, *18*(4), 345.
2. Badhiye, S. S., Chatur, P. N., & Wakode, B. V. (2012). Temperature and humidity data analysis for future value prediction using clustering technique: an approach. *International Journal of Emerging Technology and Advanced Engineering*, *2*(1), 88-91.
3. Daskalov, P. I. (1997). Prediction of temperature and humidity in a naturally ventilated pig building. *Journal of Agricultural Engineering Research*, *68*(4), 329-339.
4. Harpold, A. A., Rajagopal, S., Crews, J. B., Winchell, T., & Schumer, R. (2017). Relative humidity has uneven effects on shifts from snow to rain over the western US. *Geophysical Research Letters*, *44*(19), 9742-9750.
5. Šundić, T., Marco, S., Perera, A., Pardo, A., Hahn, S., Bârsan, N., & Weimar, U. (2002). Fuzzy inference system for sensor array calibration: prediction of CO and CH4 levels in variable humidity conditions. *Chemometrics and Intelligent Laboratory Systems*, *64*(2), 103-122.
6. Thornton, P. E., Hasenauer, H., & White, M. A. (2000). Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. *Agricultural and forest meteorology*, *104*(4), 255-271.
7. Zou, W., Yao, F., Zhang, B., He, C., & Guan, Z. (2017). Verification and predicting temperature and humidity in a solar greenhouse based on convex bidirectional extreme learning machine algorithm. *Neurocomputing*, *249*, 72-85.
8. Erdogan, I. G., Fosso-Kankeu, E., Ntwampe, S. K., Waanders, F. B., Hoth, N., & Rand, A. (2019, July). Acid rock drainage prediction of metalliferous soils from O'kiep, Namaqualand, South Africa: A humidity cell test assessment. In *IMWA 2019 Conference "Mine Water–Technological and Ecological Challenges* (pp. 15-19).
9. Kacsoh, B. Z., Barton, S., Jiang, Y., Zhou, N., Mooney, S. D., Friedberg, I., ... & Bosco, G. (2019). New Drosophila long-term memory genes revealed by assessing computational function prediction methods. *G3: Genes, Genomes, Genetics*, *9*(1), 251-267.
10. Park, J. E., Son, W. S., Ryu, Y., Choi, S. B., Kwon, O., & Ahn, I. (2020). Effects of temperature, humidity, and diurnal temperature range on influenza incidence in a temperate region. *Influenza and other respiratory viruses*, *14*(1), 11-18.