The Analysis of Large-Scale Climate Data: Jordan Case Study

Yaser Jararweh, Izzat Alsmadi*, Mahmoud Al-Ayyoub and Darrel Jenerette+
Jordan University of Science and Technology, Irbid, Jordan
*Yarmouk University, Irbid, Jordan
+University of California, Riverside

Abstract— The analysis of large-scale data for the purpose of extracting patterns is applicable to several research fields. In this paper, a dataset of climate related historical data from Jordan is collected. The main focus is to study evolution criteria of weather attributes in Jordan over the evaluated period of time and how it is connected to climate change. Results showed that humidity and dew point weather attributes are going to face a significant increase in the future. Such an increase is expected to have a direct, as well as an indirect impact on human life.

Keywords—Big Data; Data Analysis; Weather prediction; climate change; data mining; evolution analysis; prediction models

I. INTRODUCTION

Studying historical data, for the purpose of future management and planning, has been a core concept in data mining and decision support systems. In this paper, we focus on studying historical weather data from several observing stations in Jordan. The data is collected over the span of about 10 years; from 2000 to 2011. The collection is based on the availability of such data in those stations. There are some missing data, especially in some stations which were recently assigned the task of observing weather information and attributes. This is considered as one of the challenges or issues related to weather data analysis, as some stations are recently established and hence do not include historical data for the same period as those stations that were established earlier. Missing data issues are handled with some special algorithms. They can be also ignored or given default or average values in some other cases.

The study and knowledge of how weather evolves over time in some location or country in the world can be beneficial for several purposes. Such information or knowledge can be used for future predictions. For example, knowledge of how precipitation is changing may help in planning for water resources usage and management. This is due to the fact that water is very critical for human life, agriculture, plant phenology, and most aspects of life. Temperature, air pressure, humidity, etc. may help in the analysis and prediction of climate change and some of the possible impacts of such change.

Weather and climate terms are sometimes used in interchangeable situations. Their main difference is that weather prediction refers to a short period (e.g. several days to one week), where on the other hand, climate prediction includes the process of predicting future evolution for months, years, etc. While all weather attributes are used in climate process or prediction, climate prediction may include further attributes not included in weather prediction or analysis.

Major data attributes included in the collected weather station information include: year, month, day, and attributes collected on a daily basis related to the station region including the maximums, minimums and averages for the temperature degree, dew point, humidity, sea level pressure, and wind speed. Finally precipitation attributes are included. These attributes include numeric attributes for the size or amount of precipitation measured in millimeter or centimeter. They can also include general weather condition related to rain, wind, etc. such as: calm, hot, rainy, etc. Dew point is also measured in degrees like temperature. The dew point temperature is related to the amount of moisture (or humidity) in the atmosphere.[1] It is then associated with temperature and humidity. Fog is expected to be formed when dew point is approaching certain temperatures. Humidity will also be very high when weather state is “fog,” and reaches 100% in its maximum value. Dew point formula then depends on temperature and relative humidity as:

$$A = (\ln(RH / 100) + ((17.27 * T) / (237.3 + T))) / 17.27 \quad D = (237.3 * A) / (1 - A)$$

where A is an intermediate value, RH is the relative humidity, T is the temperature, and D is the dew point temperature

The rest of the paper is organized as follows. The next section describes the big data and the climate data in Jordan. A literature review is presented in Section III. Section IV includes the goals to be achieved in this paper and the approaches used to achieve them. Section V describes experiments and analysis conducted on the dataset. The paper is then concluded with final remarks and discussion of future work.

II. BIG DATA AND CLIMATE DATA IN JORDAN

In 2012 the world produced about 2.5 Exabytes of data daily [19]. A recent study estimates that 7 Zettabytes of

[1] http://weather.unisys.com

978-1-4799-7100-8/14/$31.00 ©2016
IEEE 288

data to be generated in 2014 [20]. This data is generated due to the explosion in the use of electronic devices such as computers, sensor networks, and smart phones, as well as the use of social communication sites in several daily activities. This huge amount of generated data needs to be handled using novel and efficient data management systems giving rise to the term "big data." This term is currently used to represent such huge and complex data sets that no traditional data processing systems can handle efficiently. Big data, according to McKinsey [21], refers to datasets whose sizes are beyond the abilities of typical database software tools to capture, store, manage and analyze. Big data needs new technologies that will be able to extract value from those datasets; such processed data might be used in other fields such as artificial intelligence, data mining, etc.

In Jordan, more than 75% of the land is considered as arid or semi-arid lands with less than 120 millimeters (4.7 in) of rain per year with a very dry weather for most of the year. It is estimated that only 7% of Jordan's land is arable. Moreover, as the rest of the world, Jordan faces many environmental challenges. Examples of such challenges include limited natural fresh water resources, desertification, endangered species, climate change, etc. Factors contributing to the emerging environmental challenges in Jordan are rapidly expanding population, industrial pollution, depletion of natural resources and recent unrest and disputes in the region, namely the Gulf War. The impacts of climate change are likely to aggravate the situation in the future. Moreover, Jordanian habitats and wildlife communities have endured substantial changes due to the above mentioned threats. Fresh water shortage and the continuous threat to wildlife and vegetation are major concerns for Jordan.

Climate change presents even greater threats due to its direct and indirect impacts on human life, animals and species diversity and vegetations. Climate change can cause an increase, in frequency and intensity, of heat waves, floods, and windstorms. Moreover, recent studies show that climate change may have indirect effect on humans' health as well as effecting the quality and quantity of water resources.

Water resources in Jordan are scarce. Besides the rapid population growth, the impacts of climate change will further exacerbate the problem. Temperatures will increase, and the total annual precipitation is likely to decrease, however, with a fair share of uncertainty. Hence, existing and new activities with the objective to minimize the gap between water supply and demand are needed to adapt Jordan to tomorrow's climate. This might be accompanied by activities to improve Jordan's capacity to monitor and project meteorological and hydrological data, and assess its own vulnerability to climate change.

According to its water strategy, Jordan is among the four driest countries in the world. Due to rapid population growth, water availability per capita has declined significantly from 3,600 m^3 per capita in 1946 to only 145 m^3 in 2008. Water demand distinctly exceeds water supply. Almost two thirds (64%) of the water is used for irrigation, while municipal use accounts for 30%, industry for 5% and tourism for 1%. In order to overcome the water crisis, the Jordanian Water Strategy focuses on demand management and an increase in water supply through the utilization of treated wastewater, the exploitation of the nonrenewable Disi aquifer, and a canal from the Red Sea to the Dead Sea.

Jordan's Second National Communication suggests that the rising temperatures will lead to a decrease in surface runoff at least in the two river basins, Zarqa and Yarmouk, which were chosen as pilot study areas. If the temperature rises by 2 °C, even an increase in precipitation by 20% would not compensate the increase in evaporation resulting in a decrease of surface runoff. Given this situation, it is likely that climate change will lead to even more water scarcity in Jordan.

Addressing these environmental problems requires the involvement of several governmental and nongovernmental organizations to coordinate the actions required to contain and limit the growing environmental threats. In order for these organizations to carry the right actions, it is essential to provide accurate, reliable, and updated data of the current environment status of the various regions as well as the current trends for environment changes. Such data can be used to derive the current environment impacts and recommend possible reversing actions.

III. LITERATURE REVIEW

The field of climate analysis is a rich field in terms of publications. They are categorized based on several aspects, such as the approaches followed, case or area of the study, etc. We discuss some research papers here as examples based on relevance and citation from Google Scholar.

Klein's paper in 1971 [2] discussed using computer algorithms for weather prediction. The goal was to develop a system to automate the weather forecasting process. Specifically, the system tried to make precipitation prediction in a large number of cities in the US for about 1-3 days in advance. Precipitation is used as the output attribute or class with binary values (0 indicating "no precipitation" and 1 indicating otherwise). Three main attributes were used as inputs: location height, temperature, and moisture as well as the last precipitation state. The data was collected over a period of 7-8 years. Precipitation status or state is predicting between day and night times based on the four different year seasons.

Integrated Surface Database (ISD) website[2] is used to aggregate weather data from around the world. The goal was to aggregate and unite units of measurements for aggregated data. This will reduce possible inconsistencies in data and analysis. Such data is utilized in many research papers. The website includes data from more than 20,000 stations around the world from periods as early as the year 1900 in some stations. The work of Hughes et al. in 1993 [3] is an early water prediction paper that tried to classify important attributes related to the prediction of precipitation. Willett et

[2] http://www.ncdc.noaa.gov/oa/climate/isd/

al. [4] utilized ISD and derived a homogenized gridded dataset of surface humidity in order to examine changes in surface-specific humidity over the late twentieth century. Brown and DeGaetano paper in 2009 [5] used data from 10 stations in the United States to develop a method to detect inconsistencies in historical hourly dew point data.

Precipitation prediction was always a major focus in weather data analysis. Li and Sailor [6] used a tree structured regression to relate precipitation prediction as output with different weather input attributes. This is a decision tree classification algorithm to build a tree based on attributes that impact precipitation.

IV. Goals and Approaches

We study all climate related data from all weather stations in Jordan over a period of about 10 years. The data is collected for the years from 2000 to 2011. In some cases, data were missing from some weather stations. In the following paragraphs, we present our analysis of the collected data. Due to size limitations, we only show the figures with significant interest.

Figure 1 below shows monthly evolution over 10 years for some weather stations in Jordan. Some stations are not selected to reduce the clutter in the picture and improve visibility. Fluctuations of temperatures vary from 41 or 42 down to 4 or 5 °C (in daily highest temperature). Low degrees are largely for Amman and Irbid stations, while other stations such as Aqaba (coastal desert) or other south and east desert stations show highest degrees. Some weather stations such as Amman show slight increase over the year for the highest daily temperature values.

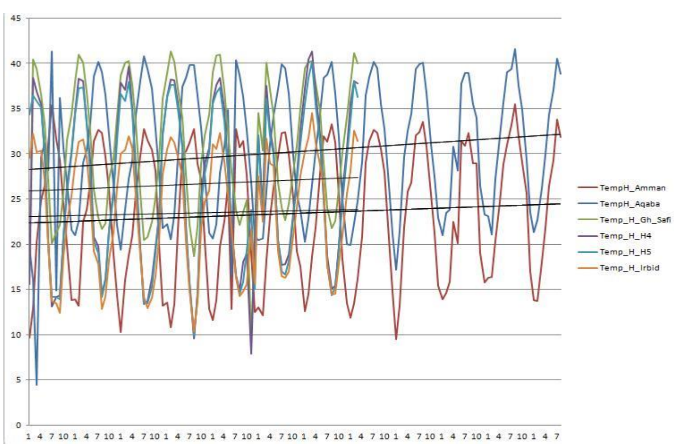

Figure 1: Evolution of Highest Daily Temperature in Jordan.

Figures of the average and the lowest daily temperatures show somewhat similar results. The dark shaded region shows 80% prediction intervals. That is, each future value is expected to lie in the dark blue region with a probability of 80%. The light shaded region shows 95% prediction intervals. These prediction intervals are a very useful way of displaying the uncertainty in forecasts. In this case, the forecasts are expected to be very accurate, hence the prediction intervals are quite narrow.

Figure 2 shows evolution of the highest daily dew point for selected stations. It shows that the value of the highest daily dew point is significantly decreasing with time over the studied years. This is clearly visible in Amman and Aqaba stations. Similar results are seen for the lowest and the average daily dew points.

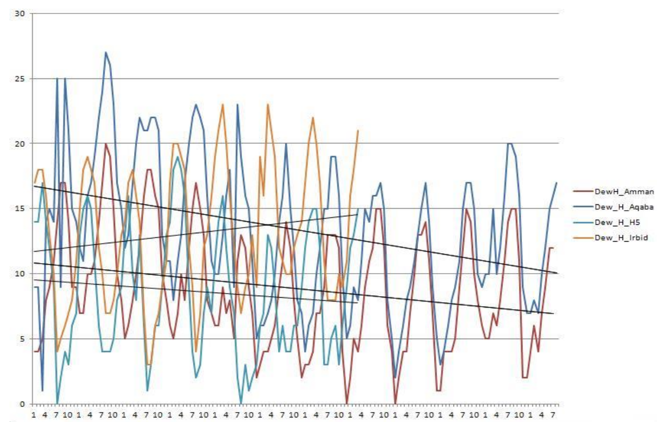

Figure 2: Evolution of the Highest Daily Dew Point in Jordan.

As described earlier, dew point is directly related to temperature and humidity. It is also indirectly related to pressure. Temperature and dew point converge when humidity is very high. The highest daily humidity for the two stations (i.e. Amman and Irbid) is summarized in Figure 3, which shows that the highest daily humidity values are increasing over the evaluated years.

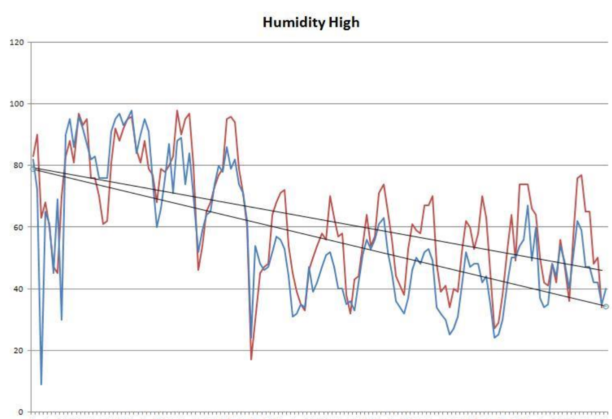

Figure 3: Evolution of the Highest Daily Humidity for selected stations in Jordan

Figure 5 shows precipitation evolution. With the exception of some days over the evaluated years, precipitation is low and not consistent in all stations. Jordan's weather varies between desert and Mediterranean with typical four seasons where rain falls mostly in winter.

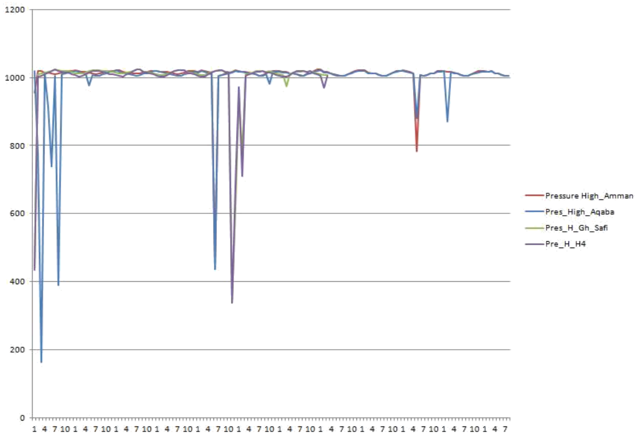

Figure 4: Evolution of Pressure in Jordan.

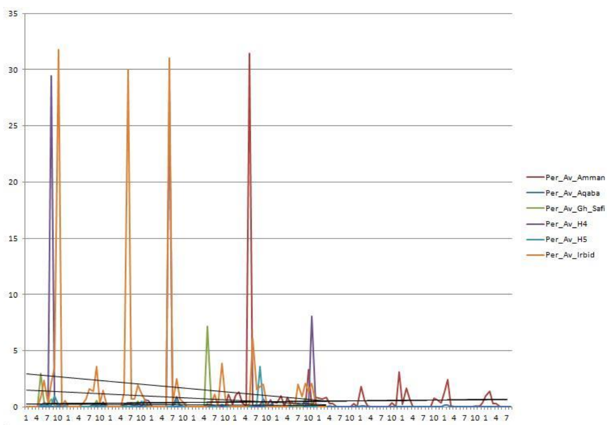

Figure 5: Evolution of Precipitation in Jordan

V. EXPERIMENTS AND ANALYSIS

For the experiments and analysis, we use time series functions from different applications including (R-Package-forecast and WEKA time series). Some of the time series methods are: ARIMA, ar, HoltWinters and StructTS. The initial analysis shows that ARIMA can be the best choice for our data and forecasting.

Autoregressive Integrated Moving Average (ARIMA) time series algorithms or models include an explicit statistical model. It can handle irregular components of time series. This allows for non-zero autocorrelations in the irregular component. ARIMA models are defined for stationary time series. Some preprocessing methods are used to make a time series stationary. An example of those is the function (diff) in R-Package.

A. *Time series analysis*

Time series algorithms include a sequence of data points separated uniformly in time. We study time series based algorithms to study evolution and weather forecasting. Typically in such algorithms, data should be homogenous. As such, we select Amman weather station for the analysis in this section as it is complete and does not include empty values.

Time series algorithms and techniques were used previously in research for forecasting. A recent master thesis evaluated the usage of time series in forecasting [12]. The author developed his own implementation of time series and made a comparison with other algorithms such as: ANN, SVM and ARIMA. He also showed that good implementations of such algorithms can lead to a prediction or forecasting accuracy which is better than many other relevant algorithms such as those mentioned earlier. The case study includes data collected from the Mississippi river daily water flow rate. Another thesis [13], which represents a more relevant paper to our research, applies time series for climate forecasting. The author evaluated the use of fractal theory in time series forecasting. She also evaluated the use of two specific algorithms in time series: Apriori-FD and CLEARMiner. Each one showed strengths for specific climate related prediction cases. There are several other examples of papers for using time series in forecasting, including: Chakrabarti and Faloutsos [14], Wang et al. [15], Nascimento et al. [16], El-Hajj et al. [17], Flamand et al. [18] and many others. As a simple example of a time series chart, Figure 6 below shows monthly temperature forecast for Amman weather station in Jordan.

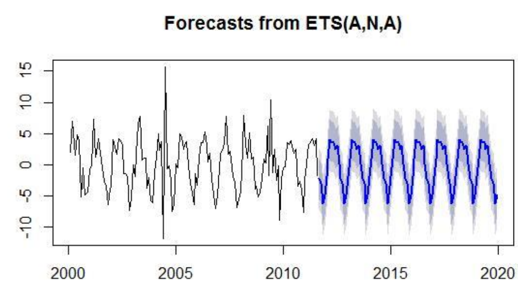

Figure 6: Amman station Temperature Forecasting

Figure 6 shows fluctuation of temperature over the different year quarters with the lowest in Quarter 1 (January, February, March) and the highest in Quarter 3 (July, August and September). This is typical in Mediterranean weather. The figure shows also evolution over the years. Highest degrees are in Quarter 3 with about 3000 total temperature or 33.3 degrees on daily average. The lowest are in Quarter 1 with 1000 total or 11.11 degree on daily average.

We run the forecasting part to produce future weather attributes based on historical data. Each attribute is studied separately based on its archival data. Figure 7 shows one sample for the highest daily temperature degree with prediction on a monthly basis for 100 future months (for Amman weather station).

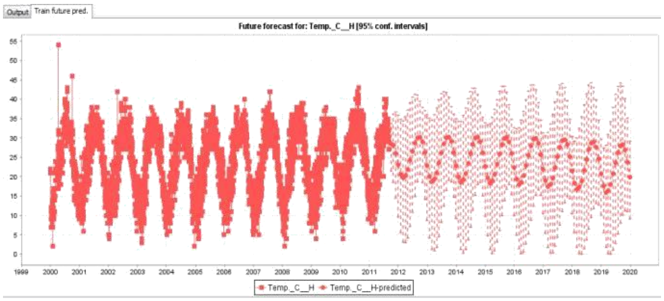

Figure 7: Highest Daily Temperature Time Series Forecasting

Figure 7 shows the average temperature for each month highlighted with maximum and minimum values for months showed in dots. Figure 8 shows the performance metrics for the forecasting process. For both mean absolute error and root mean squared error metrics, the smaller the value the better.

```
=== Evaluation on training data ===
Target                    1-step-ahead  2-steps-ahead  3-steps-ahead
=====================================================================
Temp._C__H
  N                             4162           4161           4160
  Mean absolute error         2.3308         3.5062         4.7632
  Root mean squared error     3.1426         4.4798         5.8585

Total number of instances: 4174
```

Figure 8: Temp forecasting performance metrics.

More interestingly, dew points are going to see a significant increase in future. Figure 9 shows evolution of the average dew point for the next 100 months measured on monthly basis. Similar results can be seen for the highest and lowest monthly dew point forecasting.

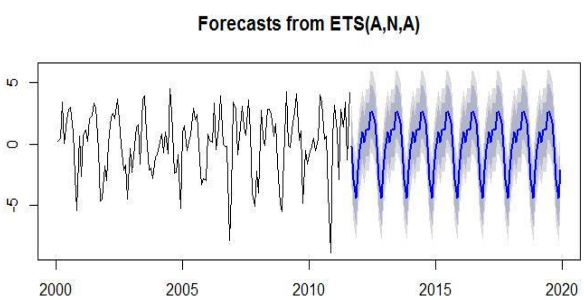

Figure 9: Amman Station Dew Point Forecasting

Table 1 shows accuracy or performance metrics for dew point forecasting. All the metrics listed below give indications of how good or accurate the forecasting was.

Table 1: Dew Point Forecasting Accuracy

ME	RMSE	MAE	MASE
-0.0132	1.694	1.34	0.656

In general, smaller numbers mean better forecasting accuracy. In each case, we ran several scenarios to improve forecasting based on those metrics.

Evolution forecasting for humidity shows similar behavior of that in dew point (Figure 10).

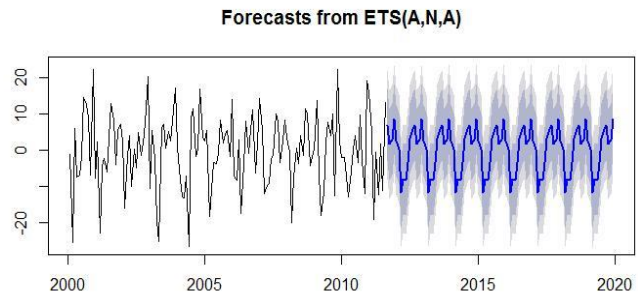

Figure 10: Amman Station Humidity Forecasting

Figure 11 shows forecasting of pressure which shows future increase. However, the increase is not as significant as that seen in dew point and humidity.

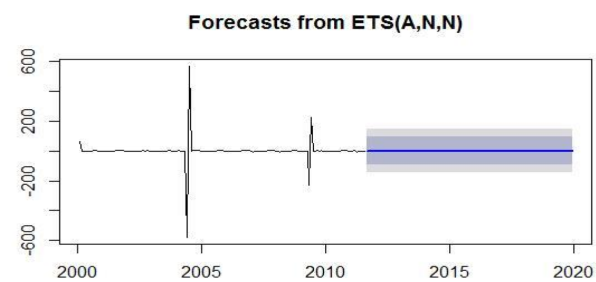

Figure 11: Amman Station Pressure Forecasting

Visibility future forecasting shows conflicting results between the highest, the average, and the Lowest values. Wind forecasting also showed conflicting results between the highest, the average and the lowest values. Figure 12 shows percipitation forecasting. Values did not show significant future change.

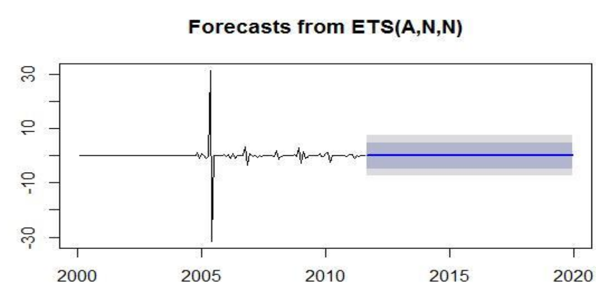

Figure 12: Amman Station Precipitation Forecasting

Relative humidity is related to the amount of water vapor in the atmosphere. It affects visibility, and is related to the global warming. It can also impact forming of fog, smoke, clouds, etc. It also has the potential to increase the intensity of tropical cyclones, and human heat stress. Furthermore, it may have important effects on the biosphere and surface hydrology (Willett et al. 2007 [6]). Dew point temperature depends on temperature and humidity. As temperature showed no significant evolutionary change, shifts in dew points can be attributed to humidity.

The papers by Robinson [7] and Gaffen and Ross [8] in 2000 and 1999, respectively, studied the evolution of humidity and dew point in the US. Their papers showed trends of increase in those parameters. A previous paper from Jordan by Abu-Taleb et al. [9] in 2007 also showed that RH is increasing in Jordan. Other papers indicated decrease of RH in some other world locations (e.g. Wijngaarden and Vincent [10] result in 2003 about Canada).

If this trend continues, the atmosphere in this region will come closer to saturation. This has implications for frequency of convective precipitation, precipitation amounts, and intensity.

With the increase in dew point temperature, agricultural animals and production processes may also be influenced. All animals have a thermal neutral zone where their performance is optimal and once outside of this zone, their performance decreases. This zone may be becoming smaller with the increase in dew points and decrease in daytime maximum temperatures.

B. *Impacts of Dew point and Humidity rise*

There are several environmental factors that may be affected by rises in dew point and humidity. One of the first impacts is on plants and animals as indicated in Stalder's paper [11] in 2006. For animals, this may cause the reduction of productivity. Product quality may also suffer a decline. Such a large overall increase in dew points may also affect crop production, soil erosion, water supplies, and even human health. Figure 13 shows average temperature vs. average dew point forecasting.

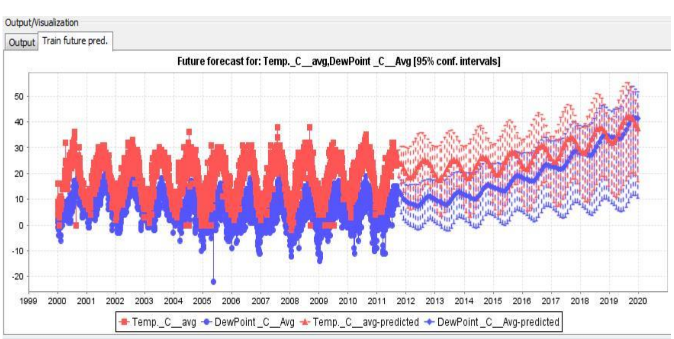

Figure 13: Temperature vs. Dew Point Forecasting

Figure 13 shows that future forecasting expects convergence in values between temperature and Dew point. Joining the three factors together (Figure 14) (i.e. temp., humidity, dew point) shows that humidity is the one that is most likely to go with the most significant change or rise in future.

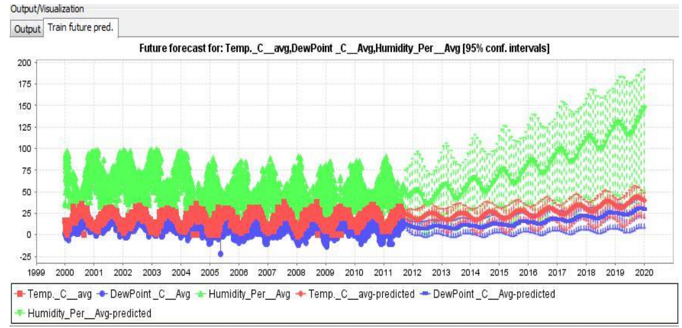

Figure 14: Temp., humidity, dew point forecast

In principle, temperature should not go lower than dew point. They are equal when humidity is 100 %. Rise in dew point and humidity will then ultimately cause a rise in temperature as well. Higher dew points may lower temperature on land and increase it in sea shores. An area unaccustomed to thunderstorms may start seeing thunderstorms based on new climate conditions.

C. *Evaluating Forecasting Accuracy*

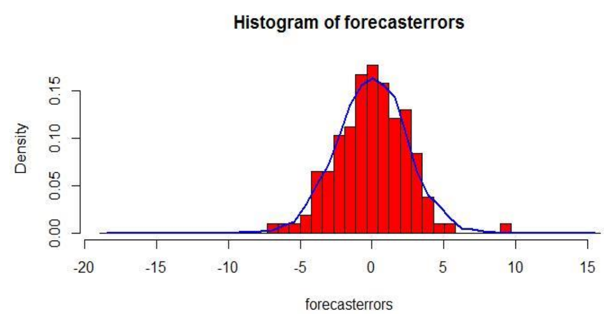

Figure 15: Dew Point Forecast Errors Histogram

Figure 15 shows that the distribution of Dew Point forecast errors is roughly centered at zero. It is hence more or less normally distributed, although it seems to have some off evaluates on the right side. However, the right deviation is relatively small, and so it is plausible that the forecast errors are normally distributed with mean zero.

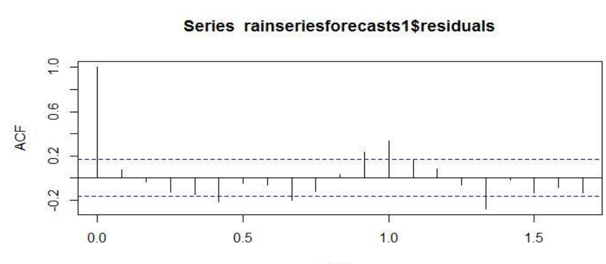

Figure 16: Dew Point correlogram auto correlation

The correlogram shows that the autocorrelations for the in-sample forecast errors do exceed the significance bounds for lags 1-20.

VI. Conclusion and Future Works

In this paper, climate change in Jordan is studied. Weather data is collected for several stations for the period of about 10 years from 2000 to 2011. Forecasting methods are used to predict future possible evolution in weather attributes. Humidity and dew point are the two weather attributes that showed significant increase. Such increase is categorized under global warming and is expected to cause impacts on the life of: plants, animal and human directly or indirectly.

Acknowledgment

This work is supported by the USAID - PEER Science program with the grant number PGA-2000003469.

References

[1] Y. Jararweh, A. Hary, Y. B. Al-Nashif, S. Hariri, A. Akoglu, D. Jenerette, Accelerated discovery through integration of kepler with data turbine for ecosystem research, in: Computer Systems and Applications, 2009. AICCSA 2009. IEEE/ACS International Conference on, IEEE, 2009, pp. 1005–1012.

[2] W. H. Klein, Computer prediction of precipitation probability in the united states, Journal of Applied Meteorology 10 (5) (1971) 903–915.

[3] J. P. Hughes, D. P. Lettenmaier, P. Guttorp, A stochastic approach for assessing the effect of changes in synoptic circulation patterns on gauge precipitation, Water Resources Research 29 (10) (1993) 3303–3315.

[4] K. M. Willett, N. P. Gillett, P. D. Jones, P. W. Thorne, Attribution of observed surface humidity changes to human influence, Nature 449 (7163) (2007) 710–712.

[5] P. J. Brown, A. T. DeGaetano, A method to detect inhomogeneities in historical dewpoint temperature series, Journal of Applied Meteorology & Climatology 48 (11).

[6] X. Li, D. Sailor, Application of tree-structured regression for regional precipitation prediction using general circulation model output, Climate Research 16 (1) (2001) 17–30.

[7] P. J. Robinson, Temporal trends in United States dew point temperatures, International Journal of Climatology 20 (9) (2000) 985–1002.

[8] D. J. Gaffen, R. J. Ross, Climatology and trends of us surface humidity and temperature, Journal of Climate 12 (3).

[9] A. A. Abu-Taleb, A. J. Alawneh, M. M. Smadi, Statistical analysis of recent changes in relative humidity in Jordan, American Journal of Environmental Sciences 3 (2).

[10] W. A. Van Wijngaarden, L. A. Vincent, Trends in relative humidity in canada from 1953–2003, Bulletin of the American Meteorological Society 85 (2004) 349.

[11] K. Stalder, Personal communication (2006).

[12] Josh Forman-Gornall , Exploring the Potential of a Novel Time Series Prediction Algorithm, Imperial College London Department of Computing, http://www.doc.ic.ac.uk/-teaching/distinguished-projects/2013/j.forman-gornall.pdf, June 2013.

[13] Luciana Romani, Integrating time series mining and fractals to discover patterns and extreme events in climate and remote sensing databases, http://www.cpa.unicamp.br-/alcscens/articles/Tese_Luciana_Alvim_Santos_Romani.pdf, USP, Brazil, Dec. 2010.

[14] Deepayan Chakrabarti and Christos Faloutsos, Large-scale Automated Forecasting using Fractals , CIKM '02 Proceedings of the eleventh international conference on Information and knowledge managementPages 2-9, 2002.

[15] Wang, J., Price, K. P. and Rich, P.M (2003), Temporal responses of NDVI to precipitation and temperature in the central great plains. International journal of remote sensing, 24:2345-2364.

[16] Nascimento, C.R., Zullo Jr, J., Romani, L.A.S. and Rodrigues, L. A., (2009). Identification of sugar cane fields in the state of Sao Paulo using a time series of AVHRR/NOAA satellite images. In: International Workshop on the Analysis of Multi-Temporal Remote Sensing Images, 5., 2009, Groton, Connecticut. Proceedings... Storrs : UConn, 2009.

[17] M. El Hajj, A. Bégué, S. Guillaume, and J. F. Martiné, "Integrating SPOT-5 time series, crop growth modeling and expert knowledge to monitor agricultural practices - Case of sugarcane harvest on Reunion Island," Remote Sensing of Environment, 2009.

[18] Claude Flamand, Mickael Fabregue, Sandra Bringay, Vanessa Ardillon, Philippe Quénel, Jean-Claude Desenclos, Maguelonne Teisseire, Mining local climate data to assess spatiotemporal dengue fever epidemic patterns in French Guiana, Journal of the American Medical Informatics Association (18 February 2014), doi:10.1136/amiajnl-2013-002348 Key: citeulike:13053772.

[19] A.Mcafee, E.Brynjolfsson, (2012) 'Big Data: The Management Revolution', Harvard Business Review.

[20] R. Villars, C. Olofson, M.Eastwood, (2011) 'Big Data: What It Is and Why You Should Care', White paper.

[21] J.Manyika, M.Chui, B.Brown, J.Bughin, R.Dobbs, C.Roxburgh, A.H.Byers, (2011)'Big Data: the next frontier for innovation, competition, and productivity', xxx