Research article

Study of wind speed and relative humidity using stochastic technique in a semi-arid climate region

Suhail Mahmud[1],*, Md Al Masum Bhuiyan[1], Nusrat Sarmin[2] and Sanjida Elahee[3]

[1] Department of Computational Science, University of Texas at El Paso, 500 w university avenue, El Paso, TX 79912, USA
[2] Department of Computer Science, University of Texas at El Paso, 500 w university avenue, El Paso, TX 79912, USA
[3] Department of Chemistry, University of Delaware, Newark, DE 19716, USA

* Correspondence: Email: smahmud@miners.utep.edu; Tel: +19152532230.

Abstract: This paper deals with the stochastic analysis of wind speed based on relative humidity data. We propose a stochastic regression technique to estimate the time-varying parameters of wind speed in a semi-arid climate region. The modeling of stochastic parameters of atmospheric data with consistent properties facilitates prediction with higher precision. In order to compare the estimation, we used simulated atmospheric time series and observational time series. The atmospheric time series was generated by the Weather Research and Forecasting (WRF) model, whereas the observational time series was obtained from the surface weather stations. The time-varying parameters of the model used are estimated by Maximum Likelihood process. The results obtained suggest that relative humidity exhibits a stochastic effect to predict stationary wind speed data. This type of analysis helps to characterize some key meteorological variables, which would be useful in forecasting irregular wind speed.

Keywords: wind speed; relative humidity; maximum likelihood estimation; stochastic regression; bootstrapping; WRF model

1. Introduction

Forecasting and analyzing of wind speed are always complicated and challenging to weather scientists and meteorologists. Weather based and time series based are the two methods to forecast

wind speed [1]. The former uses hydrodynamic atmospheric models which incorporate physical phenomena such as frictional, thermal, and convective effects. The latter uses only historical data recorded at the site to build statistical models from which forecasts are derived. In this paper, we apply a combination of both approaches to forecast the wind speed based on the relative humidity time series data. Forecasting of atmospheric time series with stochastic parameters is very imperative in the field of weather research, climate forecasting, and prediction analysis. Currently, NWP models are widely used to forecast the weather, as they can capture many statistical properties of data. However, those models have some limitations like complex topography, horizontal resolution, initial, and boundary condition problem [2,3]. In this study, atmospheric data shows dynamic behavior as they evolve. Therefore, stochastic technique is useful in predicting with higher accuracy for those data series [4].

The study area for this research, Paso del Norte (PdN) region, has semi-arid climate, which is the next driest climate after the desert climate. Also, the rainfall is slightly higher than the desert climate and receives precipitation of between 10 and 20 inches annually. It is often considered the intermediate state between the desert and humid climates. Semi-arid climates characterize the tropics and sub-tropics located in the 20º and 30º latitudes. The countries with this type of climate conditions are mostly located in Africa, South Asia, some parts of Europe, particularly Spain, Mexico, Southwestern United States, and parts of South America [6].

PdN region is characterized by a unique geographical location. This region is comprised of three counties in southwestern Texas and southern New Mexico of the United States, and the municipality of Ciudad Juarez in the northern part of Mexico. The two largest cities, El Paso and Ciudad Juarez are separated by a river called the Rio Grande and connected by five land bridges. It is located at the virtual midpoint of the 1500-mile border shared by the United States and Mexico; 1700 miles southwest of Washington, DC and 970 miles northwest of Mexico City [9].

Figure 1. Geographical representation of PdN Region [10].

The meteorology and topographic factors of the Paso Del Norte region play a significant role on the climate. This region is intersected by the Franklin Mountains, containing versatile geographical components like The Chihuahuan desert, Rio Grande river, Kilbournes Maar and Franklin mountains Volcano peaks. The most well-known feature of the area is the Rio Grande river which passes around

the southern end of the Franklin mountains, West of Juarez and El Paso. This binational river, flows through three states of United State, which are Texas, New Mexico and Colorado.

In this semi-arid region, the range of the temperature varies around 88–95 degrees Fahrenheit. Fahrenheit in the summer, while the relative humidity varies between 28–50 percent throughout that season [5]. We observed that relative humidity is more stochastic compared to temperature over time. Due to the dynamic behavior of relative humidity, a stochastic regression model has proposed to analyze the wind speed based on relative humidity. In this case we use the asymptotic and bootstrapping estimation of the model parameters to explain the dynamic behavior of atmospheric time series.

Previous studies were conducted to determine air quality standard and meteorological parameters using various numerical weather prediction models and air quality models. Based on the 1996 ozone study campaign, several research papers were published [11,12]. Global and regional atmospheric chemistry models like Community Multistate Air Quality (CMAQ) or Comprehensive Air Quality Model Extensions (CAMx) were used with the combination of WRF to calculate the effects of emission on global oxidizing capacities and develop ozone reduction strategies [13,14]. The use of remote sensing technology and models [15–17] was also one of the methods that applied, but the accuracy and efficiency of these models forecast were less correlated with the observational data. However, in this study, a statistical approach like linear regression of wind speed was analyzed based on the stochastic effect of relative humidity data. A significant feature of stochastic regression is that the consequences of the ordinary least square estimator is unbiased and it is efficient to predict the target variable by allowing random variation in predictors. Moreover, the bootstrapping and the maximum likelihood method are followed to adequately estimate the parameters and to improve the accuracy. We determine the adequacy and stationarity of the data by computing the asymptotic error, bootstrapping error, and some robust tests, and these are discussed in the later section of this paper.

This paper is organized as follows: Section 2 describes the weather research and forecast model and its required parameters. The techniques of stochastic regression and the estimation of its time-varying parameters are also discussed. Section 3 highlights the background of atmospheric data and their sources used in this study. Section 4 discusses the descriptive statistics and distribution of data that will be useful in estimating the model parameters. In section 5, we perform tests that analyze the stationary behavior of data. Section 6 provides the results of models when applied to the data sets. This section also includes the suitability of our model regarding the variation of parameters using asymptotic errors and bootstrapping errors. Finally, section 7 contains the conclusion and validity of our study.

2. Methodology

This section describes a numerical weather prediction model, WRF, and stochastic regression model applied to meteorological data sets. We will first discuss some techniques to understand the background of our methodology. Later we will study the models and estimation procedures to determine the stochastic effects and time-varying parameters of data.

2.1. WRF model

WRF or Weather Research Forecast, is a community numerical weather prediction model developed by a collaborative partnership of different scientific institutes like National Oceanic and Atmospheric Administration (NOAA), National Center for Atmospheric Research (NCAR), and National Center for Environmental Protection (NCEP), etc. This model is a next generation mesoscale NWP system which is designed for both atmospheric research and operational forecasting applications [18]. It contains two computational or dynamical cores known as Advanced Research WRF or ARW and Non-hydrostatic

Mesoscale Model or NMM. It also has a data assimilation system and a software architecture system, which allows users to run in a parallel computation system [19].

WRF allows users and researchers to create simulation reflections of either real data (observational, analysis) or idealized conditions. This model provides forecasting a strong and flexible platform while offering many developments in physics, numerical analysis, and data assimilation by the user and researcher from around the world [20]. Many organizations, including but not limited to the National Weather Service and The National Severe Storm Laboratory, use the WRF model to predict and forecast weather at different scales. Furthermore, this model has created a large worldwide community of users, which includes members from more than 150 countries and expanding. To analyze any atmospheric event, first, we need to choose the initial condition from any external meteorological data source. We have selected the analysis data of the Global Forecast System which, is a weather forecast model produced by the National Center for Environmental Prediction. The Grid/Scale resolution of the data sets is 0.5-degree domain. The details of the WRF simulation are given below in Table 1.

Table 1. Description of the WRF model.

Parameter	Description
Period	May–July 2017
Initial Condition Meteorology	GFS-ANL 0.5 degree
Vertical levels/Eta Levels	34
Horizontal Grids	172 x 172
Grid resolution	10, 3, 1
Time Steps	180s, 180s, 180s
Microphysics	WSM or WRF single moment
Planetary Boundary layer	YSU (Yonsei University)
Cumulus Parametrization	Kain-Fritsch scheme
Shortwave and Longwave	RRTM Longwave Scheme
Land Surface	Noah Land surface Scheme
Surface Layer Option	Monin Obukhov Similarity Scheme
Projection	Lambert
Boundary Conditions Meteorology	GFS-ANL 0.5 degree

We used three model domains with two-way nesting condition (Figure 1). The outermost domain, D01, covers the southwestern part of the United States and Mexico, while intermediate domain, D02, is focused on the southwestern part of Texas, Northern part of Mexico and some lower part of New Mexico state. The smallest domain, which is denoted by D03, is focused on our research area of interest which is The Paso Del Norte region, comprised of El Paso city, some of the counties from New Mexico State and Juarez city in Mexico. We executed the simulations for the 36-hour interval run, where first 12 hours were the spin-up or warm-up run, and the next 24 hours were our simulation day.

WRF version 3.9.1 was used for the simulation with ARW (Advance Research Wrf) core. Grid nudging was switched on for every domain and the simulation was restarted every 1440 minutes, which is 24 hours.
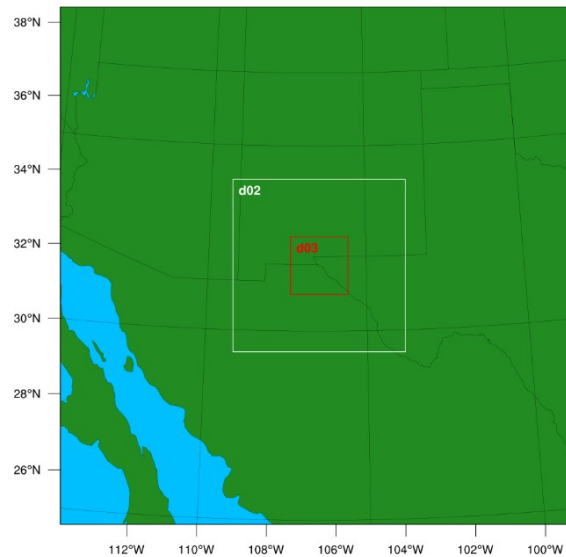
Figure 2. Three model domain with the grid spacing of 10 km, 3 km and 1 km respectively

2.2. Stochastic regression (SR)

Stochastic regression is a technique for estimating the probability distribution of potential outcomes by allowing for random variation in input variables [21]. In this study, we use SR model to forecast the wind speed (y_t) of atmosphere based on the relative humidity (z_t) data. So the relative humidity is considered as a stochastic regressor in this model. An advantage of stochastic regressor is that the consequences of ordinary least square estimator is unbiased and it is efficient to predict the target variable. We now define the SR model as follows:

$$y_t = \alpha + \beta_t z_t + v_t \tag{1}$$

where α is a fixed constant, β_t is a stochastic regression coefficient, and v_t is a white noise with variance σ^2 [28]. The distribution of estimators $\hat{\beta}_t$ depends on the distribution of v_t and z_t. To estimate the stochastic regression term β_t, we use a first order autoregression as follows:

$$(\beta_v - b) = \phi(\beta_{t-1} - b) + \omega_t \tag{2}$$

where b is a constant, and ω_t is a white noise with variance σ^2. The both noise terms, v_t and ω_t in the SR model are considered as uncorrelated. Our approach is to estimate the likelihood of parameters φ, α, b, σ_w, σ_v and to compare the estimated asymptotic standard errors and bootstrapping errors of each estimation.

In order to estimate the time-varying parameters in the SR model, we define Eq 1and Eq 2 as a state space model [29]. At this point, Eq 1 is considered as an observation equation and Eq 2 is considered as a state equation, where

$$(\beta_v - b) = \phi(\beta_{t-1}) + \omega_t \qquad \beta_0 \sim N_p(\mu_o, 0) \tag{3}$$

Since the observations contain noises and the systems are continuously changing, we use the filtering technique to estimate the unknown variables by following three steps: forecasting, updating,

and parameter estimation. At the first step, we forecast the unobserved state vector β_t using the state Eq 2, where the predicted state estimators $\beta_{t-1} = E(\beta_t | y_1, \ldots \ldots y_{t-1})$. We update the results while we have a new observation of y_t at time t. The prediction errors z_t of the likelihood function are computed as follows:

$$z_t = y_t - E(y_t | y_1, \ldots, y_{t-1}) = y_t - z_t \beta_t^{t-1} - 1 \tag{4}$$

Now we consider the corresponding error covariance matrix of prediction errors (innovation covariances) $\sum_t = z_t M_t^{t-1} Z_t + \sigma_v^2$, where M_t^{t-1} is the variance-covariance matrix of state estimators. The form of Kalman filter is used in this process as follows:

$$K_t = [\phi M_t^{t-1} z_t'] \sum_t^{-1} \tag{5}$$

The Kalman filter is an efficient recursive filter that estimates the internal state of a linear dynamic system from a series of noisy measurements. The advantage of Kalman filter is that it specifies how to update the filter from β^{t-1} to β^t, once a new observation y_t is obtained, without reprocessing the entire dataset $y_1, y_2, \ldots, y_t$ For a detail of Kalman filtering, the reader is referred to the reference in [21]. We then update the stochastic regression effect using the Kalman filter as follows

$$\beta_{t+1}^t = \sigma \beta_t^{t-1} + (1 - \phi)\beta + K_t \epsilon_t \tag{6}$$

We also use the Kalman filter to measure the estimates precision of filter error covariance matrix as:

$$M_{t+1}^t = \phi M_t^{t-1} \phi' + \Theta \sigma_\omega^2 \Theta' - K_t \sum_t K_t' \tag{7}$$

In this case, Θ is a coefficient matrix of σ_ω^2. In order to estimate the parameters, we initialize the procedure by selecting initial values for the model parameters as $\Theta_0 = (\mu_0, \Sigma_0, \phi, \sigma_w, \sigma_v)^t$. So the model coefficients and the correlation structure of the model are uniquely parameterized as follows $\phi = \phi(\Theta_0), (1 - \phi)b(\Theta_0), \sigma_\omega^2 = \sigma_\omega^2(\Theta_0), \sigma_v^2 = \sigma_v^2(\Theta_0)$ .Then the parameters $(\Theta = (\phi, \alpha, b, \sigma_w, \sigma_v)^t$ are estimated by maximizing the expected likelihood as follows:

$$lnL_y(\Theta) = -\frac{1}{2}\sum[\ln|\Sigma(\Theta)| + \epsilon_t(\Theta)'\Sigma(\Theta)^{-1}\epsilon_t(\Theta)] \tag{8}$$

Using the MLE, we estimate the model parameters with asymptotic standard errors. Since the datasets do not behave exactly as Gaussian distribution, we also use the bootstrapping technique by resampling the data to obtain better approximation. In this case, since the innovations ϵ_t are uncorrelated, and

when we resample the data, the innovations are standardized as. $e_{t=} \sum_t^{-\frac{1}{2}} \epsilon_t$ So we obtain the MLE from Eq 8 as follows:

$$lnL_y(\Theta) = -\frac{1}{2}\sum[\ln|\Sigma(\Theta)| + e_t(\Theta)'e_t(\Theta)] \tag{9}$$

3. Data Background

Surface observational data is required to validate the WRF model. For this study, different surface stations operated by TCEQ have been used. TCEQ, with the help of the Environmental Protection Agency, set up a grid of observational data collection stations throughout the state of Texas (USA), these stations are known as Continuous Ambient Monitoring Station, or CAMS, are used for measuring both air and water pollutant across the state of Texas. In addition to measuring air pollutants, CAMS also contain instruments to measure local meteorological parameters like outdoor temperature, wind speed, wind direction, relative humidity, dew point temperature, solar radiation, precipitation, etc. [22]. CAMS contains equipment that measures ambient gaseous materials and particulate matter, ambient concentration of ozone, carbon monoxide and oxides of nitrogen. Particulate matter is measured in two classifications: PM10 (less than 10 microns in aerodynamic diameter) and PM2.5 (particles with an aerodynamic diameter of 2.5 microns or less).

Figure 3. Location of the observation ground station.

Table 2. Information about observational site.

Site characteristics	Description	Description
Site name	UTEP	Chamizal
EPA site number	481410037	481310044
Sites coordinates	31.7628 N, −106.501W	31.7656 N, −106.455 W
Elevation	1158 m	1122.0 m
CAMS	0012, 0125, 0151	0041, 0126, 3001
Activation date	January 01, 1981	April 01, 1988
County	El Paso	El Paso
City	El Paso	El Paso
Zip	79902	79905
Sampler type (Meteorological)	Preci., RH, Temp, Wind	Preci., RH, Temp, Wind
Sampler type (Pollutant)	O3, NOx, CO, PM2.5	O3, NOx, CO, PM2.5

For downscaling the WRF simulation to compare with the observational data site, several methods have been taken. We used statistical downscaling, which uses observed relationships between variables at different spatial scales to predict regional-scale model fields from coarser data [23]. Another approach we applied to downscaling WRF specifically for this region is the high-resolution domain,

where the smaller domain consists of 1km resolution. Use of Ncar Command Language (NCL) to convert the nearest grid point to specific latitude and longitude also applied.

We specifically choose two different locations around the Paso del Norte. These areas are geologically unalike, and contain divergent environmental significance. Relevant information is shown in Table 2 [24].

4. Descriptive statistics of atmospheric data

We analyzed the dynamic behavior of atmospheric data as they evolve over time. The analyses were performed by MATLAB and R programs. We used 432 data points on an hourly basis for each observational and simulated datasets used in this study. The descriptive statistics of the datasets are as follows.

Table 3. Descriptive statistics of atmospheric data from UTEP source.

Statistics	WS (obs)	WS (Sim)	RH (obs)	RH (Sim)
Mean	7.3016	6.0754	18.8227	25.3035
Std. dev	3.1895	3.1165	10.5697	12.8907
Minimum	0.6000	0.3200	3.9000	6.3000
Maximum	17.6000	16.69000	52.6000	68.2400
Skewness	0.4756	0.7484	1.1478	0.9489
Kurtosis	3.2673	3.3611	3.7166	3.1938
Index of agreement	0.50	0.50	0.60	0.60
Mean bias error	−1.22	−1.22	6.46	6.46

Table 4. Descriptive statistics of atmospheric data from Chamizal source.

Statistics	WS (obs)	WS (Sim)	RH (obs)	RH (Sim)
Mean	8.0276	7.9014	17.5429	25.6249
Std. dev	4.8007	4.0854	8.8070	12.7455
Minimum	0.4000	0.4200	4.2000	5.9800
Maximum	28.0000	19.0100	46.000	66.3800
Skewness	0.9080	0.4245	1.0368	0.9882
Kurtosis	4.2161	2.2079	3.5049	3.6249
Index of agreement	0.62	0.62	0.40	0.40
Mean bias error	−0.172	−0.172	8.06	8.06

The mean, standard deviation, minimum, and maximum values of data sets are introduced as summary measures of the location and variability of data distribution (see Tables 3 and 4). The skewness and kurtosis give the summary information about the shape of a distribution. As we see that the skewnesses vary from 0.4 to 1.14, which supports the non-normality distribution of data. The kurtosis measures the attribute of tailedness of distribution. Tables 3 and 4 show that most of the kurtosis of datasets is greater than 3.0, meaning that they have leptokurtic distribution, i.e., more peaked than a normal distribution with longer tails. Furthermore, we compute the index of agreement and bias error to see how much those predictions and real values matched in between them. As we can see from those two tables, the index of agreement of Wind speed and Relative humidity varies between

0.4 and 0.6, which indicates better resemblance. Mean bias error for relative humidity is higher than wind speed in both locations.

4.1. Stationarity

To analyze the atmospheric data used in this paper, we test their stationary behavior by using unit root test. In a stationary time series, the second-order behavior of data, namely mean, variance, and covariance does not change with time. A unit root test tests whether an autoregressive process is a random walk as opposed to a stationary process. We computed the test statistics and corresponding p-values of atmospheric data by using two powerful unit root tests, namely the Augmented Dickey Fuller (ADF) test and the Phillips-Perron (PP) test).

4.1.1   ADF test

To test the stationarity of data, we first use the ADF test. The test of Dickey Fuller (DF) checks the null hypothesis that a time series y_t is a unit root against the alternative that it is stationary, assuming that the dynamics in the data have an Autoregressive Moving Average (ARMA) structure [25]. The ADF test is an augmented version of the DF test, where the t-statistic is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some significance level. The computed t-statistics and p-values of this test are given below:

Assumption: Null hypothesis (H_0): There is a unit root for the time series. Alternative hypothesis (H_a): There is no unit root for the time series, i.e., the series is stationary.

Table 5. ADF t-statistics test.

Variables	UTEP source		Chamizal source	
	Test statistics	p-value	Test statistics	p-value
Wind speed (obs)	−4.84	0.01	−4.25	0.01
Wind speed (sim)	−4.67	0.01	−4.41	0.01
Rel. humidity (obs)	−6.21	0.01	−5.89	0.01
Rel. humidity (sim)	−6.35	0.01	−5.26	0.01

The p-values suggest that whether the null hypothesis is acceptable or not, based on the significance level. In Table 5, the computed p-value is lower than the significance level $\alpha = 0.05$. We reject the null hypothesis H_0 for all atmospheric time series used in this paper and accept the alternative hypothesis H_a. Thus the data under study are all stationary time series.

4.1.2.   PP test

The PP test is an alternative method to correct the serial correlation in unit root testing. It basically uses the standard ADF test, but modifies the t-ratio so that the serial correlation does not affect the asymptotic distribution of the test statistic. In particular, where the ADF test uses a parametric auto-regression to approximate the ARMA structure of the error in the test regression, the PP test ignores any serial correlation in the test regression. For a detail of PP test, the reader is referred to the reference in [26].

Table 6. PP t-statistics test.

Variables	UTEP source		Chamizal source	
	Test statistics	p-value	Test statistics	p-value
Wind speed (obs)	−68.74	0.01	−56.78	0.01
Wind speed (sim)	−60.21	0.01	−49.30	0.01
Rel. humidity (obs)	−43.63	0.01	−35.56	0.01
Rel. humidity (sim)	−47.57	0.01	−31.24	0.01

Assumption: Null hypothesis (H_0): There is a unit root for the time series. Alternative hypothesis (H_a): There is no unit root for the time series, i.e., the series is stationary. Since the computed p-values in Table 6 is lower than the significance level $\alpha = 0.05$, we reject the null hypothesis H_0 that has a unit root. We therefore accept the alternative hypothesis H_a that the data used in this work are all stationary time series.

4.2. *Autocorrelation of data*

In this subsection, we present the autocorrelation (ACF) of atmospheric data used in this paper. The autocorrelation function gives a complete characterization of a stationary time series. The shape of the ACF shows how the autocorrelations behave as the distance between observations increases. Figure 4, we see that the autocorrelations follow an oscillation with gradual damping, meaning that they oscillate in sign but decrease in magnitude. Eventually, the autocorrelations goes to zero, which confirms that the data are stationary at some specific lags.
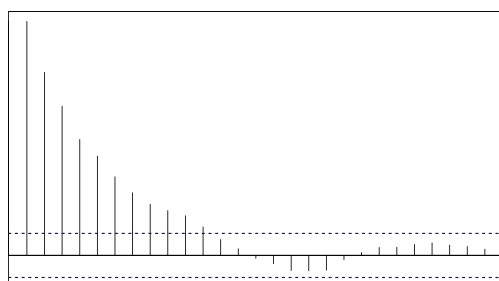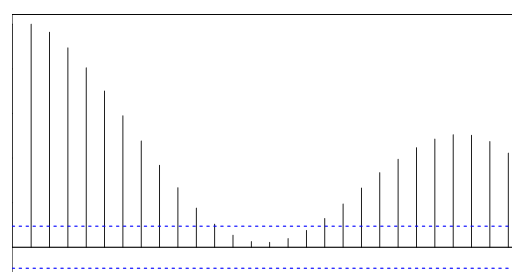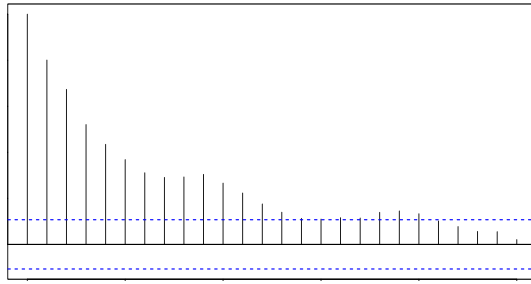

(a) Wind speed (UTEP).
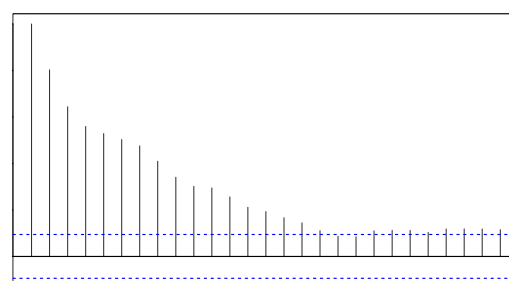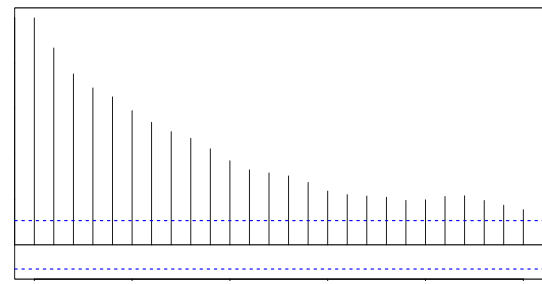

(b) Wind speed simulated (UTEP).
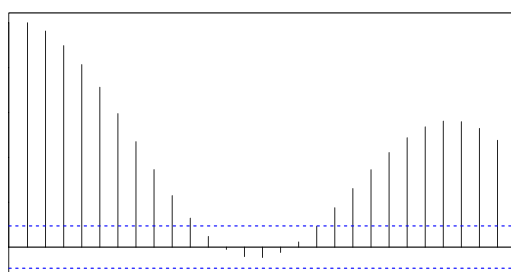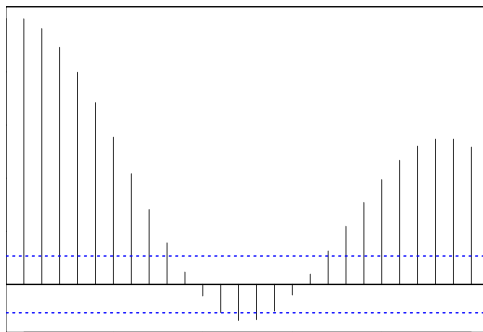

(c) R.Humidity (UTEP).
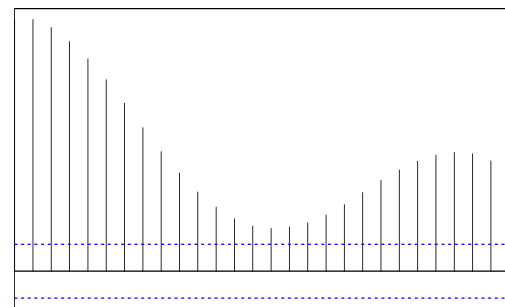

(d) R.Humidity Simulated (UTEP).

(e) Wind speed (Chamizal).

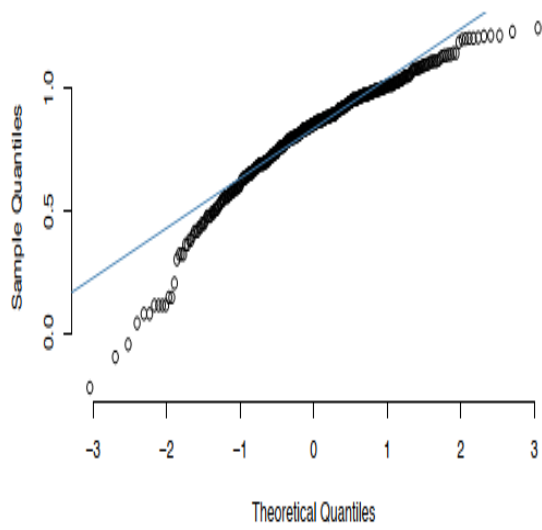(f) Wind speed simulated (Chamizal).

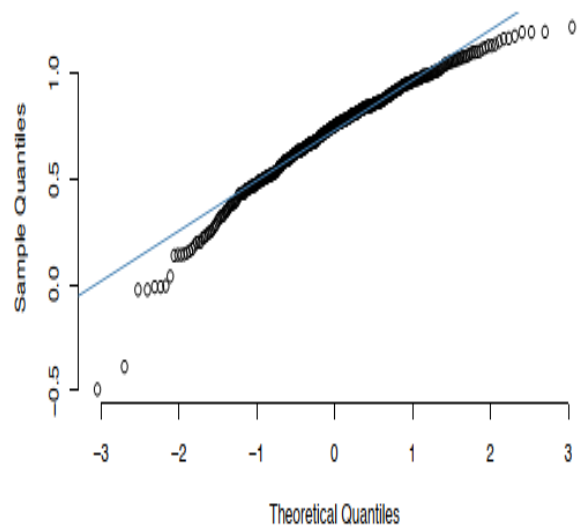(g) R. Humidity (Chamizal).

(h) R. Humidity simulated (Chamizal).

Figure 4. Autocorrelation of atmospheric data.

4.3.   Q-Q plot analysis

We now present the Quantile-Quantile (Q-Q) plot of wind speed and relative humidity data. A Q-Q plot is used to compare the shapes of data distributions, providing a graphical view of how properties such as location, scale, and skewness are similar or different with the normal distributions. In Figure 5, we see that the data are not distributed as fully Normal, but slightly left or right skewed.
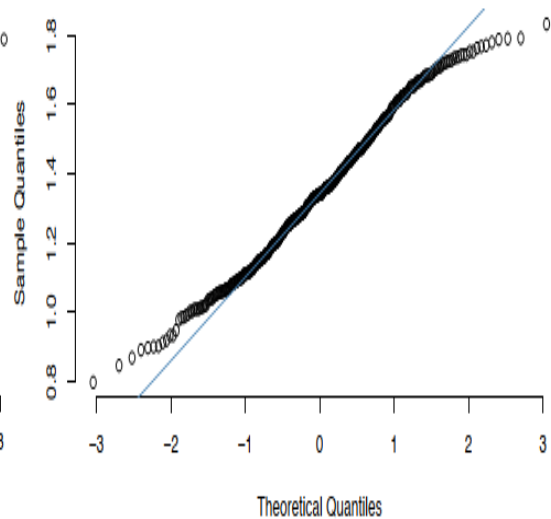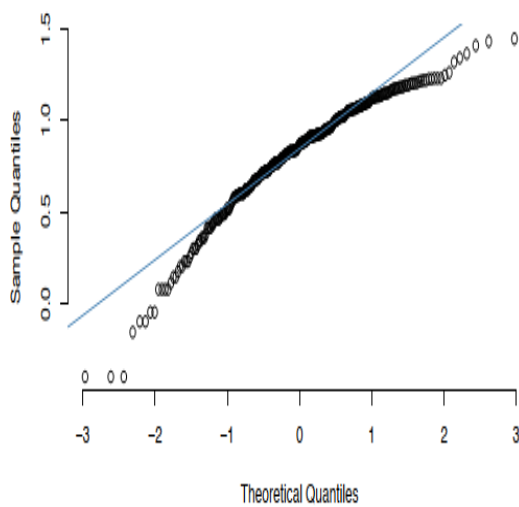
(a) Wind Speed (UTEP).

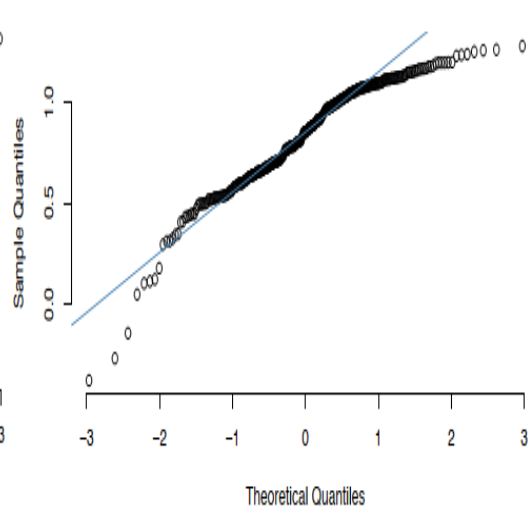(b) Wind Speed simulated (UTEP).
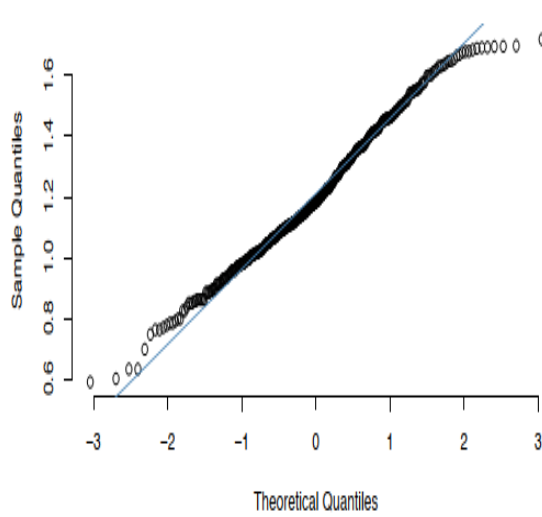
(c) R. Humidity (UTEP).

(d) R. Humidity simulated (UTEP).

(e) Wind speed (Chamizal).
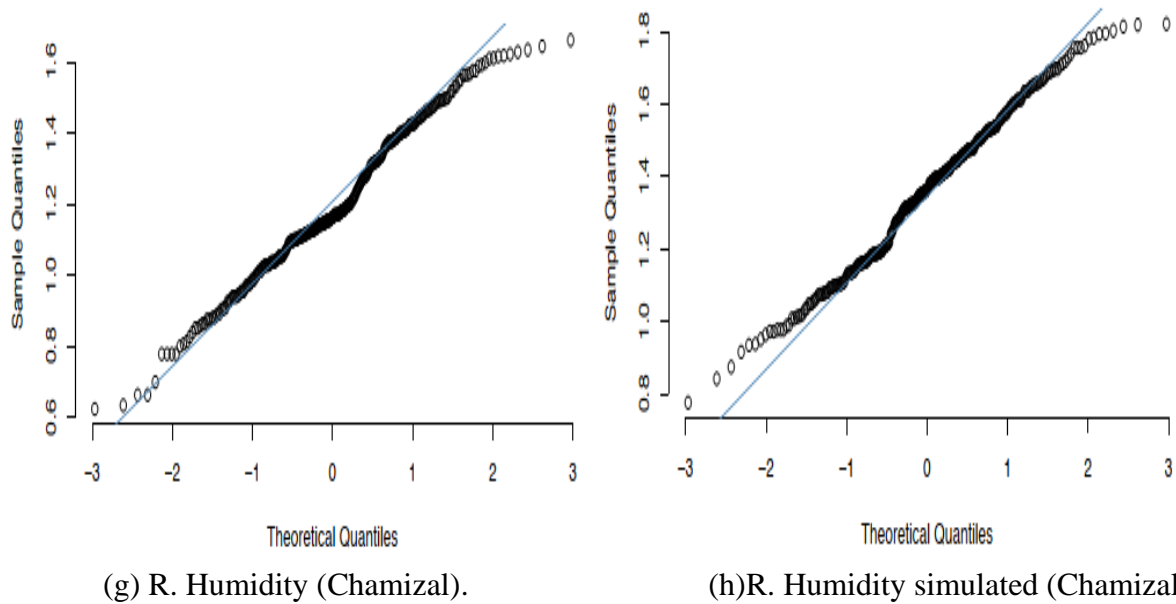
(f) Wind speed simulated (Chamizal).

| (g) R. Humidity (Chamizal). | (h)R. Humidity simulated (Chamizal). |

Figure 5. Normal Q-Q plot of atmospheric data.

5.   Results and discussion

This section presents the results and analyzes of parameter estimation for two meteorological components, namely wind speed and relative humidity. In order to estimate the time-varying parameters, we used maximum likelihood estimation in the stochastic regression model. We compared the noise terms in the model in two ways, i.e., asymptotic and bootstrapping approaches. The estimation procedure has been performed by R statistical program.

We analyzed 432 data points for each dataset obtained from two sources, such as UTEP and Chamizal CAMS. Because the datasets were not large, we used the bootstrapping technique to estimate the errors of fitted model. An advantage of bootstrapping is that it does not require any distributional assumptions, such as normally distributed errors. The time-varying parameters in Eqs 1 and 2 were initialized in order to observe the stochastic regression effect during a set of wind speed for relative humidity in the atmosphere. We set the initial values as $\sigma_0 = 0.02$, $\phi = 0.80$, $\alpha = -0.065$, $b = 0.75$, $\sigma_\omega = 0.09$, and $\sigma_y = 1.50$. The bootstrapping has been replicated 500 times with relative tolerance 0.001 to obtain the convergence of numerical optimization.

Table 7. Parameter estimation of Experimental Data from UTEP source.

Parameter	Estimate	Asymptotic error	Bootstrapping error
φ	0.903	0.022	0.009
α	6.993	0.517	0.961
b	−0.008	0.059	0.073
σ_ω	0.093	0.006	0.008
σ_v	0.944	0.095	0.394

Tables 7–10 summarize the estimation of parameters ϕ, α b, σ_ω, σ_y, asymptotic standard errors, and bootstrapping standard errors for the data sets from UTEP and Chamizal CAMS sources. We see that the variations of time-varying parameters are very low, so the model has good predictive ability.

At this point, we can say that estimates are close to the true parameters. In these tables, the asymptotic standard errors are typically much smaller than the bootstrapping errors. For most of the cases, the bootstrapped standard errors are at least 40% larger than the corresponding asymptotic value, which shows our stochastic regression model fits well into the atmospheric data.

Table 8. Parameter estimation of simulated data from UTEP source.

Parameter	Estimate	Asymptotic error	Bootstrapping error
φ	0.912	0.021	0.007
α	6.168	0.544	0.677
b	−0.009	0.046	0.037
σ_ω	0.067	0.004	0.006
σ_v	0.712	0.086	0.341

Table 9. Parameter estimation of experimental data from chamizal source

Parameter	Estimate	Asymptotic error	Bootstrapping error
φ	0.905	0.024	0.019
α	10.664	0.874	1.376
b	−0.194	0.106	0.352
σ_ω	0.147	0.009	0.007
σ_v	1.196	0.136	0.513

Table 10. Parameter estimation of simulated data from chamizal source

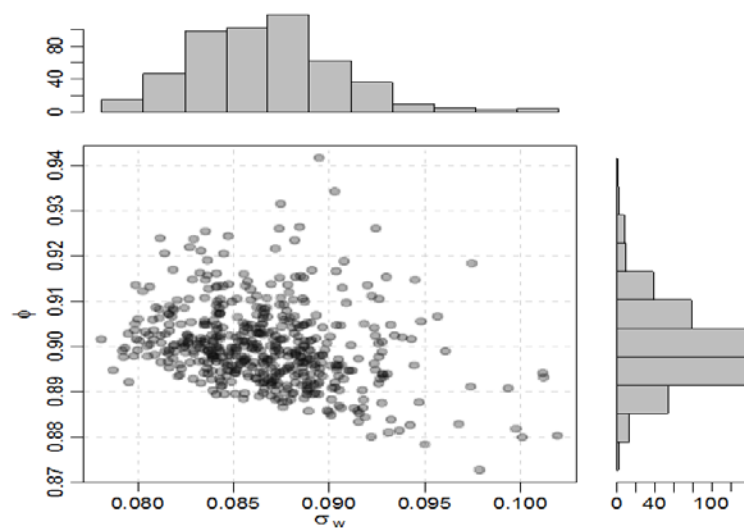| Parameter	Estimate	Asymptotic error	Bootstrapping error
φ	0.933	0.021	0.018
α	7.850	0.742	1.343
b	−0.019	0.072	0.302
σ_ω	0.076	0.005	0.007
σ_v	0.628	0.114	0.528

Figure 6. Joint and marginal bootstrap distributions for experimental data from UTEP source.

To explain the stochastic concepts of our model with data, we present the plot of joint bootstrap distribution of estimated parameters ϕ and σ_ω (see Figures 6–9). We see that the parameter $\hat{\sigma}_w$ is clearly away from zero, which suggests that $\hat{\sigma}_w$ is a stochastic regression parameter. However, we notice that when $\hat{\sigma}_w$ increases, then ϕ reduces over the time. So the parameter σ_w is bigger than zero, corresponds to $\hat{\phi} \approx 0$. At this point, the state dynamics of Eq 2 are followed as $\beta_t = b + \omega_t$. In this case, when β b, the dynamics of the data have fixed regression effect. However, Tables 7–10 represent that σ_w is greater than b, which suggests that the system is stochastic. Thus we conclude that the dynamics of the atmospheric data evolve over the time and follow a stochastic regression effect of wind speed based on relative humidity.
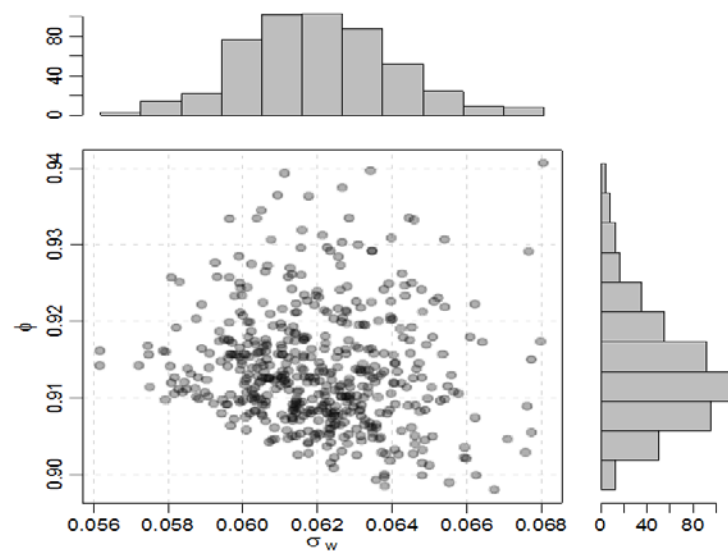

Figure 7. Joint and marginal bootstrap distributions for simulated data from UTEP source.
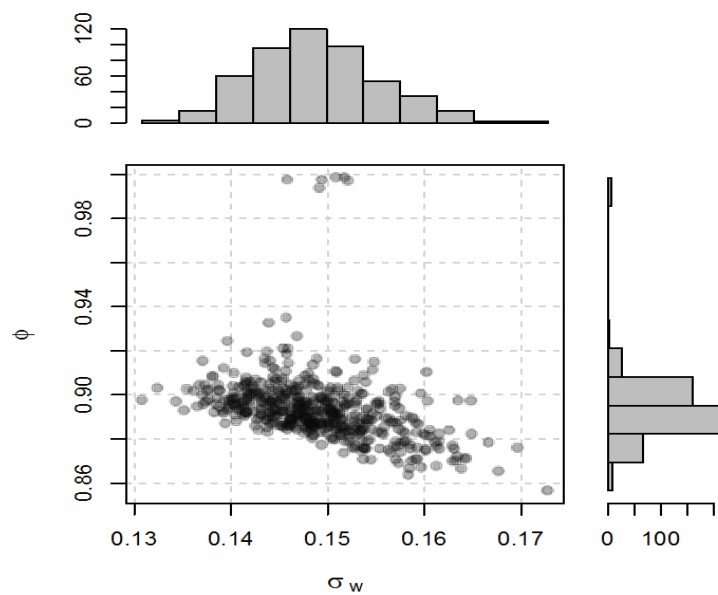

Figure 8. Joint and marginal bootstrap distributions for experimental data from chamizal CAMs source.
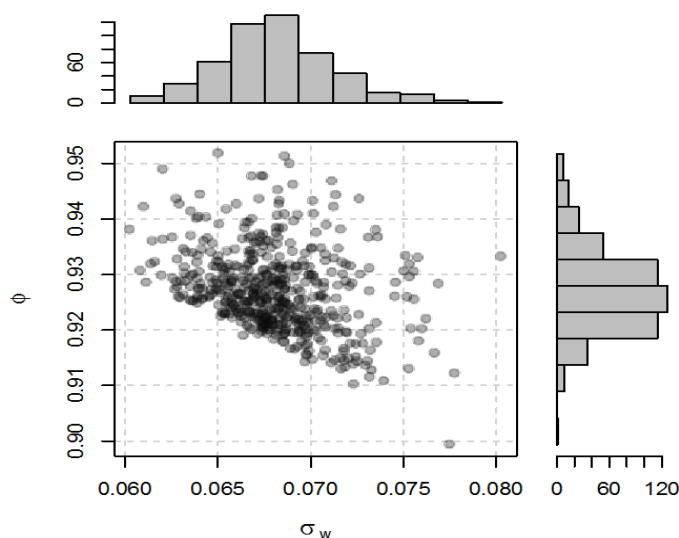
Figure 9. Joint and marginal bootstrap distributions for simulated data from chamizal CAMs source.

6. Conclusion

In this study, we used the stochastic regression technique to analyze meteorological components like wind speed and relative humidity in the Paso del Norte region. This region is very crucial for the geographical location, as it is comprised of three counties in southwestern Texas and southern New Mexico of the United States, and the municipality of Ciudad Juarez in Mexico. Since the region is closer to the border between the United States and Mexico, its weather and climates are mostly influenced by the city of Juarez and other cities of Mexico [30,31]. We notice that relative humidity in this region varies a lot throughout the year among other meteorological components. For example, the range of the temperature is around 88–95 degrees Fahrenheit in the summer, while the relative humidity varies between 28–50 percent throughout that season. We used four data sets from two different locations, namely UTEP and Chamizal CAMS in the summer season. Observational time series data were collected from those surface stations, and simulated time series data were obtained using WRF model.

We used some robust tests to analyze the stationary behavior of data (see section 4.1). Results of these tests suggest that the data is stationary, meaning that their second-order behavior-e.g., mean, variance, and covariance do not change with time. We also noticed that the sample atmospheric data is not distributed as Normal distribution (see section 4). So we applied both asymptotic and bootstrapping estimation of the model parameters since the bootstrapping provides better approximation accuracy than the asymptotic distribution. Our results show that the errors of estimated model parameters are very low; hence the estimates are close to the actual parameters. We test our claim that the system is stochastic (see subsection 4.1), and the maximum likelihood estimation converges well into the data sets studied here. We, therefore, conclude that the dynamics of the wind speed evolve over time with a stochastic regression effect of relative humidity in the region.

at El Paso and the Texas Commission of Environment Quality for all the support.

Conflict of interest

The authors declare no conflict of interest.

References

1. Kavasseri RG, Seetharaman K (2009) Day-ahead wind speed forecasting using f-ARIMA models. *Renew Energ* 34: 1388–1393.

2. Cassola F, Burlando M (2012) Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. *Appl Energ* 99: 154–166.

3. Warner TT, Peterson RA, Treadon RE (1997) A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *B Am Meteorol Soc* 78: 2599–2618.

4. Franzke CL, O'Kane TJ, Berner J, et al. (2015) Stochastic climate theory and modeling. *Wiley Interdis Rev: Climate Change* 6: 63–78.

5. Yu Media Group. El Paso, TX. Detailed climate information and monthly weather forecast. Available from https://www.weather-us.com/en/texas-usa/el-paso-climate.

6. Misachi J (2017) What Are The Characteristics Of A Semi-arid Climate Pattern. Available from: https://www.worldatlas.com/articles/what-are-the-characteristics-of-a-semi-arid-climate pattern.html

7. Novlan DJ, Hardiman M, Gill TE (2007) A synoptic climatology of blowing dust events in El Paso, Texas from 1932–2005. In Preprints, *16th Conference on Applied Climatology, American Meteorological Society J*

8. Breshears DD, Kirchner TB, Whicker JJ, et al. (2012) Modeling aeolian transport in response to succession, disturbance and future climate: Dynamic long-term risk assessment for contaminant redistribution. *Aeolian Res* 3: 445–457.

9. Regional Stakeholders Committee (2009) The Paso Del Norte Region, US-Mexico: Self-Evaluation Report, OECD Reviews of Higher Education in Regional and City Development, IMHE. Available from: https://www.oecd.org/unitedstates/44210876.pdf

10. Baumbach JP, Foster LN, Mueller M, et al. (2008) Seroprevalence of select blood borne pathogens and associated risk behaviors among injection drug users in the Paso del Norte region of the United States–Mexico border. *Harm Reduct J* 5: 33.

11. Lu D, Reddy R, Fitzgerald R, et al. (2008) Sensitivity modeling study for an ozone occurrence during the 1996 Paso del Norte ozone campaign. *Int J Environ Res Pub He* 5: 181–203.

12. Pearson R, Fitzgerald R (2005) Application of a wind model for the El Paso-Juarez airshed. *J Air Waste Manage Assoc* 51: 669–680.

13. Cai C, Kelly JT, Avise Stockwell WR, et al. (2001) Photochemical modeling in California with two chemical mechanisms: model intercomparison and response to emission reductions. *J Air Waste Manage Assoc* 61: 559–572.

14. Mahmud S, Wangchuk P, Fitzgerald R, et al. (2016) Study of Photolysis Rate Coefficients to Improve Air Quality Models. *B Am Phy Soc* 61.

15. Mahmud S (2016) The use of remote sensing technologies and models to study pollutants in the Paso del Norte region. The University of Texas at El Paso. Available from: https://scholarworks.utep.edu/open_etd/685/

16. Ullwer C, Sprung D, Sucher E, et al. (2019) Global simulations of Cn2 using the Weather Research and Forecast Model WRF and comparison to experimental results. In Laser communication and Propagation through the Atmosphere and Oceans *VIII*: 111330I

17. Brown MJ, Muller C, Wang W (2001) Costigan, K. Meteorological simulations of boundary layer structure during the 1996 Paso del Norte Ozone Study. *Sci Total Environ*. 276: 111–133.

18. Michalakes J, Dudhia J, Gill D, et al. (2005) The weather research and forecast model: software architecture and performance. *Use High Perform Comput Meteorol* 2005: 156–168.

19. Michalakes J, Chen S, Dudhia J, et al. (2001) Development of a next-generation regional weather research and forecast model. *Dev Teracomput* 2001: 269–276.

20. Skamarock WC, Klemp J B, Dudhia J, et al. (2005) A description of the advanced research WRF version 2 (No. NCAR/TN-468+ STR*). National Center for Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology Div.

21. Islam MR, Peace A, Medina D, Oraby T (2020) Integer versus Fractional Order SEIR Deterministic and Stochastic Models of Measles. *Int J Env Res Pub He* 17: 2014.

22. Allen DT, Torres VM (2010) TCEQ Flare Study Project, Final Report. The University of Texas at Austin The Center for Energy and Environmental Resources.

23. Wilby RL, Charles SP, Zorita E, et al. (2004) Guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA 27.

24. Raysoni AU, Sarnat JA, Sarnat SE, et al. (2011) Binational school-based monitoring of traffic-related air pollutants in El Paso, Texas (USA) and Ciudad Jurez, Chihuahua (Mxico). *Env Pol* 159: 2476–2486.

25. Said SE, Dickey D (1984) Testing for Unit Roots in Autoregressive Moving-Average Models with Unknown Order. *Biometrika* 71: 599–607.

26. Phillips PCB, Perron Pierre (1988) Testing for a Unit Root in Time Series Regression. *Biometrika* 75: 335–346.

27. Wellner Jon A (2003) Gaussian White Noise Models: Some Results for Monotone Functions. *Lecture Notes-Monograph Series* 2003: 87–104.

28. Kitagawa G (1994) State Space Modeling of Time Series. *The Institute of Statistical Mathematics* 43–64.

29. Grineski SE, Collins TW, McDonald YJ, et al. (2015) Double exposure and the climate gap: changing demographics and extreme heat in Ciudad Jurez, Mexico. *Local Env* 20: 180–201.

30. Wilder M, Garfin G, Ganster P, et al. (2013) Climate change and US-Mexico border communities. In Assessment of Climate Change in the Southwest United States, Island Press, Washington DC: 340–384.