Dust Storm Prediction Using ANNs Technique
(A Case Study: Zabol City)


Abstract—Dust storms are one of the most costly and destructive events in many desert regions. They can cause massive damages both in natural environments and human lives. This paper is aimed at presenting a preliminary study on dust storms, as a major natural hazard in arid and semi-arid regions. As a case study, dust storm events occurred in Zabol city located in Sistan Region of Iran was analyzed to diagnose and predict dust storms. The identification and prediction of dust storm events could have significant impacts on damages reduction. Present models for this purpose are complicated and not appropriate for many areas with poor-data environments. The present study explores Gamma test for identifying inputs of ANNs model, for dust storm prediction. Results indicate that more attempts must be carried out concerning dust storms identification and segregate between various dust storm types.

Keywords—Dust Storm, Gamma Test, Prediction, ANNs, Zabol.

I. INTRODUCTION

Dust storms are natural recurrent events common in many desert regions of the world. They can cause significant damage to properties and may lead to the loss of lives. In spite of the fact that dust storms occur in many arid environments around the world, such as north China [8], Mongolia [12], Central Asia [17], Bodélé Depression in northern Chad [3], Middle East (Walter and Wilkerson, 1991) and Australia [7], this phenomenon has still been insufficiently explored mainly due to the lack of surface observations which are especially sparse in hyper-arid low-populated desert regions.

Most studies on dust storms aim at understanding and characterizing the frequency of this phenomenon and its effects on environment and health. A Review on dust storm regimes in Central Asia based on observations over a 5-years period (1951–1955) was produced by Romanov (1961). In addition to frequency and duration of dust storms for this period, Romanov (1961) presented a classification of synoptic pre-conditions for the onset of dust storms and developed recommendations for dust storm forecasting [17]. Orlovsky et al. (2005) represented the general characteristics of frequency, distribution and seasonality of dust storms in Central Asia [15]. Other works on dust storm prediction have been done mainly on the basis of satellite images and complicate models (Burning et al., 2004), or by artificial intelligence techniques such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), that haven't had full efficiency [9] & [10]. The purpose of this study is to provide the basic attempt to analyze distribution, frequency, seasonality and conditions for dust storm generation in Sistan region and predict dust storm based on observations from Zabol meteorological station over a period of 25 years (1980-2005).

Dust storms occurring in Sistan region (Iran) were described in the scientific literature by Middleton (1986a) [11].

II. STUDY AREA

Sistan region is located in the Southeast of Iran, between latitudes 30° 18' N and 31° 20' N and longitudes 61° 10' E and 61° 50' E and covers an area of 15197 km². It borders Afghanistan to the north, Chehl-dokhtar (Forty girls) and Malek-siah-kuh (Black King) Mountains to the west and south-west, and Pakistan to the south and south-east. Sistan region lies within a dry-temperate zone in Demartonne climatic zonation system.
The main part of Sistan region is occupied by lowlands which are bordered to the west by a mountain range going from North to South (name of mountain range). The aridity of the climate is manifested by very low annual precipitation (61 mm), low air humidity, low cloudiness, high evaporation rates (4420 mm), and frequent droughts and dry winds. There are two distinct seasons for each year in Sistan region: a very dry and warm period in northern hemisphere summer and a relatively cold period in northern hemisphere winter (Figure 1). During the cold period, Sistan is influenced by the northern periphery of the Siberian High (SH) Pressure System and experiences air mass intrusions from the north-west and north. Severe droughts in the past decades have significantly reduced vegetation cover in the Sistan region and caused drying the main parts of lakes located in north of this region. These lakes dries in many years, completely as observed in figure 3, that resulted into the exposure of major sources for solid particles in dust storms (Figure 2). The Hamund and River, the main river in this region, discharges into an inland depression which, when sufficient water is available, forms the Hamoun Lake(s). The Hamoun Lakes are located in the northern and western parts of Sistan Plain. The Hamoun Lakes are one of the main and most valuable aquatic ecosystems throughout Iran. The Hamoun system around Sistan consists of several lakes that during high water periods form an extensive wetland of about 5,000 km2 [4].

Livelihood in the Sistan region is strongly linked with and dependent on the wetland products and services. The reed beds provide fodder for livestock, fuel for cooking and heating, and raw materials for handicraft and construction. Fishing and hunting are the important sources of income for many households. This fundamental dependence on the wetlands has resulted in the collapse of the local economy during the recent drought period. Severe water shortages have destroyed the ecological system of the wetlands and caused damage to agriculture in the region, which is primarily based on irrigation from the Helmand River. The approximate population in the Sistan region is several hundred thousand people. Prolonged droughts, when the rivers fail to bring sufficient water to fill up the lakes and wetlands, and hence supply the irrigation-based agriculture have occurred in the late 1960s, mid-1980s, and between 1999 and 2005 [4]. The first September 2004 MODIS image (Figure 3), indeed, reveals that the primary source of the dust plumes is the lakebed of the Hamoons.

III. METHODOLOGY OF DUST STORM PREDICTION

For optimal prediction of a dust storm event, all meteorological variables (minimal cloud cover, low relative humidity, maximum surface winds, frontal passage, upper trough passage, maximum temperature, maximum winds aloft, and an upward vertical velocity field) must come together [13].
Blowing dust from arid environments can be transported downwind over an extended area. Visibility forecasting for dust storm events is generally very difficult [13]. Visibility may remain at 5 – 9 km in dust haze and re-suspended dust for days after an intense dust event. Intense dust storms reduce visibility to near zero in and near source regions with visibility improving away from the source. Dust settles when winds drop below the speed necessary to carry the particles, but some level of dust haze persists nearly constantly in the region during the dry season. Small particles restrict visibility more than large particles. In general, however, the worst visibility occurs within 6 meters of the surface [2]. Strong winds by themselves are not sufficient for a significant dust storm event: the wind must be sufficiently turbulent to loft dust, and it must occur in a reasonably unstable environment. Friction velocity, a single parameter incorporating wind speed, turbulence, and stability, is a useful predictor of dust emission. Friction velocity is currently computed for many numerical weather prediction models, such as NOGAPS (Navy Operational Global Atmospheric Prediction System). NOGAPS is a global forecast model that is spectral in the horizontal and energy-conserving finite difference (sigma coordinate) in the vertical [2]. It is cited that NOGAPS friction velocity products are thus very useful for dust forecasting [13], but in large dust storms are not appropriate because particles carried from far distances.

IV. ANALYSIS OF DUST STORMS DATA SET

In this study, data sets were collected from the meteorological station of Zabol (31° 02´ N, 61° 29´ E) for a period of 1980-2005. Meteorological codes include 31, 32, 33, 34 and 35 cited for dusty weather. Also other codes include 6, 7, 8, 9 and 98 indicating that some dust exists in air but we consider first-mentioned codes only because these codes indicates severe dust storms and are more highlighted in synoptic records. In the records, sand–dust weather is classified as floating dust, blowing sand and dust storm. In the present study, we only examine the dust storm events. Dust storm is defined as a weather phenomenon when dust on land surface is blown into air by strong winds, resulting in air turbidity with visibility less than 1 km. The last drought was exceptionally long, transforming the lakes into barren desert exposing large quantities of easily deflatable lacustrine and fluviatil deposits. The summers in the region are characterized by winds known as '120-day winds' resulting windblown sand originating from the lakebeds covering the surrounding villages by the end of the season. In the study period, there were some missed data that were discarded.

A. Climatology

In Sistan region, it appears that no relation exists between wind speed and minimum visibility, in spite of the low visibility which is associated with the occurrence of high wind speeds. We have experienced some cases that in a high wind speed, no dust storm has been recorded; even when the Hamoon Lake bed had been dry. The bare soil with poor vegetation that associates with much interference by people has been made loose soil. Sometimes, a dusty weather after a light rain that is associated with a normal wind speed about 6-7 m.s-1, cause a dusty weather. The records of large scale dust storms can be confused with not real dust storms.

B. Relative Humidity

Novlan et al. (2007) showed that relative humidity in the 15-20% range is the most frequently observed during dust storms in El Paso, Texas [13]. This is in comparison to El
Paso’s average annual relative humidity of 41% and an average relative humidity of 28% during the driest month of April: dust events are clearly associated with drier than average air. In Zabol mean relative humidity reaches about 38%. Minimum relative humidity is 23%, 24%, 25%, in August, September and July, respectively. It is notable that most dust storm events occur in these months. It seems that this climatic factor could justify dusty events (Figure 4).

Fig. 4 The distribution of the mean relative humidity (HUM) observed in Sistan region over 1980-2005 period.

**C. Wind**

Qiang et al (2007) showed that the wind regime (strength and variability) is one of the key factors on the sand and dust deposition during dust storm events but this relationship is poor with a correlation coefficient about 0.6 [16]. Figure 5 indicate, somewhat, there is no relation between mean of maximum wind speed and number of dusty days. The reason for this may be that changes in the surface conditions are more important for dust emission, but these drastic changes doesn’t present in synoptic data. Figures 6 and 7 indicate mean speed and direction of wind during study period, respectively.

Fig. 5 The relation between mean maximum wind speed with number of dusty days in Sistan region over 1980-2005 period.

Fig. 6 The distribution of the mean wind speed (mean FF) observed in Sistan region over 1980-2005 period.

Fig. 7 Histogram of wind direction (DD) frequencies in Sistan region over 1980-2005 period.

**D. Precipitation**

Figure 8 shows the graphs of precipitation amounts and dust storm events simultaneously from 1985 to 2005. Note that there is a 95% correlation with years having below normal rainfall amounts and above normal dusty days. Also it is noted that approximately one year after some El Niño events [14], dust events have possibly increased in relative frequency in the El Paso.

Fig. 8 Mean annual rainfall and number of dusty days (some abnormal cases identified by circles).

**E. Visibility**

Figure 9 shows a plot of average annual visibility associated and rainfall. It is appeared that no relation is
between sum annual rainfalls and means annual minimum visibility.

Fig. 9 Relation between annual minimum visibilities associated with sum of annual rainfall (some abnormal cases identified by circles).

**F. Seasonality**

The seasonal distribution of dust storms at Sistan region is presented in Figure 10. A drastic rise in temperature in spring season may bring about the rapid evaporation of precipitation and drying of the upper sandy layer resulting in a high frequency of dusty days between June and September. High wind speeds occurring during this time of year also favor rapid drying of the surface sediments. Thus, the weak fixedness of the sands, inefficient amount of precipitation, rapid drying of the upper layer of the sands and high wind speeds are all factors in creating favorable conditions for dust storm formation by springtime.

![Fig. 10 Percentages of dusty days for different months over 1980-2005 period.](image)

**V. DUST STORM EVENTS CLASSIFICATION**

Several Dust Storm classifications have been used in several studies. Numerical case studies used horizontal visibility in dust storm time as a dust storm severity index (Novlan, 2007; Huang et al. 2006 [9]; Extröm, 2004). A dust storm index that combines the records of three types of dust events has presented by squires (2001) [22], as severe dust storms (SD), moderate dust storms (MD) and local dust events (LDE). Where dust storm severity is defined according to the weather codes. SD, when visibility is reduced to <200 m; MD, visibility is <1000 m and LDE when total suspended particulate matter >150 µg/m3/hr. Because lack of data of suspended particles measurements, visibility between 1000-3000 m, selected as light dust storm events. We proposed same classification but different visibility boundaries as severe dust storm when visibility is reduced to <100 m; MD, when visibility is between 100 m and 500 m, and LDE when visibility is between 500-1000 m.

**VI. SELECTION OF MODEL INPUTS**

The meteorological variables of interest in this study include minimum daily temperature, maximum daily temperature, mean daily temperature, dew point, and grass minimum temperature rounded to nearest a whole degree. Dominant parameters such as soil moisture and daily sun radiation influencing dust storm occurrence process are included in this study. Moreover, pressure gradients play a dominant role on air mass movements; hence pressure values at station level and equal sea pressure that could affect dust storm formation. Also, mean and maximum daily wind speed, daily rainfall and relative humidity could be effective in this process.

Data-driven modeling is particularly useful when we do not understand the inner workings of the system we wish to model, thus we treat the system as a so called ‘black box’. In some cases, we may understand the physical laws that govern certain entities within the black box, but there are so many of them that their combined effect is impossible to model. The Gamma test is one tool to estimate the variance of the noise module, the best possible smooth model (with bounded derivatives) present in observations of the real system, and therefore quantifies the extent to which the system can be described by such a smooth model.

Based on the Gamma test [20], these non-parametric methods enable us to quickly evaluate, prior to model construction, an estimate for the best mean-squared error that can be achieved by a smooth model on unseen data for a given selection of inputs. By examining this estimate for alternative selections of inputs, we show how the best choice of inputs for modeling a particular target output can be selected. We also discuss how the distribution of the data in input space and the complexity of the unknown function that we seek to model, influence the size of the required data set. Some following definitions are given for readers about this method.

**Model:** A smooth data model is a differentiable function from inputs \( x = (x_1,...,x_m) \) to each output \( y \). It is assumed that the data can be represented by an unknown model \( f \) so that:

\[
y = f(x_1,...,x_m) + r
\]

where \( r \) is a stochastic variable which represents noise.

**Gamma Test:** An algorithm to estimate the variance of the noise \( \text{Var}(r) \) associated with a particular output. Not to be confused with the variance of the output. This test is used to show how the Gamma statistics (and the other results returned by the Gamma test) estimate varies as more data is used to compute it. Eventually, if enough data is used the Gamma...
statistic should asymptote to the true noise variance on the output for which it has been computed.

Near Neighbor: The near neighbor search problem can be defined as follows; given a set of M data points where each data point consists of m attributes, we wish to find the nearest set of points to a query point. Each attribute can be measured on some quantifiable scale, and there may be many metrics to determine an attributes size. For our purposes we will assume this metric space is m-dimensional real space, Rm.

Gamma Statistic: Often referred to as a "Gamma value". It is the vertical intercept of the regression line plot and represents our best estimate for Var(r).

Embedding: A selection of past values of a time series used to predict the current value.

Mean Squared Error (MSE): If y(i) (1 ≤ i ≤ M) is a set of values of an output and y*(i) is a set of predictions for y(i) then the MSE of the predictions is given by:

\[
MSE = \frac{1}{M} \sum_{i=1}^{M} (y^*(i) - y(i))^2
\]

(2)

Standard Error (SE): This is the standard error about a regression line and is calculated as:

\[
SE(\Gamma) = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} (\Gamma(i) - \bar{\Gamma})^2}
\]

(3)

where (i) is the ith Gamma regression point value and \(\bar{\Gamma}\) is their mean.

Genetic Algorithm: A useful feature associated with a full embedding or Genetic Algorithm (GA) search is the Embbeding Histogram, which shows the frequency of embeddings with a specific Gamma statistic: If the choice of embedding is largely determined by statistical variations in the data, this histogram tends to have a Gaussian or Normal distribution. If on the other hand there are clear underlying dynamics in the data, then the histogram often shows a bimodal or multimodal distribution. This option searches the space of all masks using a (GA) to find good embeddings.

VII. LINEAR MODELS

The Gamma test (Dune and Everitt, 2001) is a non-linear analysis tool which allows quantification of the extent to which a smooth relationship exists within a numerical input/output data set. Because of this fact it generally needs far more data than parametric analysis where the model is presumed to have a particular form. It is akin to a (least squares) linear regression, but can be applied to data that has any underlying smooth non-linear function. It is able to estimate the lowest attainable mean squared error by any smooth model. The relationship between an input x and its corresponding output y can be expressed as: y = f(x) + r. Where, f is some smooth unknown function representing the system, f: C R, C Rm, and r is a random variable having expectation zero representing noise.

The Gamma test allows the variance of the noise variable r (Var(r)) to be estimated, despite the fact that f is unknown. We do not assume anything about the parametric equations governing the system, only that the underlying function is smooth with bounded derivatives.

The Gamma test can estimate Var(r) directly from the data, even though the underlying function f is unknown. This estimate is calculated by the following equations:

\[
\delta_M^2(k) = \frac{1}{M} \sum_{i=1}^{M} \left| x_{N[i,k]} \right| - x_i \right|^2
\]

(4)

where:

1 ≤ k ≤ p

\(\left| \cdot \right|\) denotes Euclidean distance,

\(N[i,k]\) denotes the index of the kth nearest neighbor to \(x_i\),

p is the number of near neighbors, typically p = 10.

Thus \(\delta_M^2(k)\) is the mean square distance to the kth nearest neighbor.

\[
y_{M,k} = \frac{1}{2M} \sum_{i=1}^{M} \left| y_{N[i,k]} - y_i \right|^2
\]

(5)

\(y_{N[i,k]}\) is the corresponding output of \(x_{N[i,k]}\).

\(\left| \cdot \right|\) denotes Euclidean distance.

The critical success index (CSI) is used to evaluate the performance of ANN.

The CSI is usually used in the meteorology to judge the performance of the sand-dust storm forecasting model. It is estimated by the equation:

\[
CSI = \frac{C_F}{C_F + W_F} \times 100
\]

where \(C_F\) is the number of dust storm events that can be predicted correctly, \(W_F\) is the total of the dust storm events left out and the events having a wrong dust storm predicting. The CSI is a conservative estimate of skill since it does not consider the correct null events [5].

VIII. RESULTS

Gamma Test identifies more important factors that influence on desired output. In this study, after gamma test, some parameters discarded in both classifications. Regarding to this, Relative humidity, mean daily temperature, dew point, and air pressure in station level, were discarded in squires classification. Figures 11 and 12 indicates results achieved from GA process in squires dust storm classification and revealed there is no same effective parameters on dust storm in a day and next day. Gamma test parameters for this method are given in table 1.
TABLE I GAMMA STATS IN BEST EMBEDDING OF INPUTS IN CLASSIFICATION AFTER SQUIRES

<table>
<thead>
<tr>
<th>Stat</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>0.027</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.0014</td>
</tr>
<tr>
<td>Gradient</td>
<td>0.029</td>
</tr>
<tr>
<td>V ratio</td>
<td>0.715</td>
</tr>
</tbody>
</table>

V ratio values is some deal high and indicate high variance in model results and therefore much uncertainty, although other values are some what reasonable.

Fig. 11 Minimum Gamma value come to 10 indicate best mask for parameter identification.

Fig. 12 Multimodal distribution in Gamma test data representing no clear underlying dynamics in the data sets.

In proposed classification, Maximum wind speed, daily rainfall, wet thermometer value, minimum daily temperature, dew point, and air pressure in station level, were discarded. Figures 13 and 14 indicate results achieved from GA process in proposed dust storm classification and revealed proposed classification is more effective in dust storm events diagnosis therefore more precise predictions (Table 2).

Fig. 13 Minimum Gamma value come to 10 indicating best mask for parameter identification.

Fig. 14 A near normal distribution in Gamma test data representing underlying dynamics in the data sets.

TABLE II GAMMA STATS IN BEST EMBEDDING OF INPUTS IN PROPOSED CLASSIFICATION

<table>
<thead>
<tr>
<th>Stat</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>0.013</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.0023</td>
</tr>
<tr>
<td>Gradient</td>
<td>0.021</td>
</tr>
<tr>
<td>V ratio</td>
<td>0.496</td>
</tr>
</tbody>
</table>

V ratio value in this classification is improved, albeit maximum wind speed is discarded, but we experienced many cases that maximum wind speed cause dust storms, even in dry seasons.

Finally, a three-layer artificial Neural Network model is used to predict dust storm events. The results are given in tables 3 and 4.

TABLE III ANN RESULTS IN SQUIRES DUST STORM CLASSIFICATION

<table>
<thead>
<tr>
<th>Type</th>
<th>CSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe</td>
<td>0.880</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.394</td>
</tr>
<tr>
<td>Light</td>
<td>0.543</td>
</tr>
</tbody>
</table>

TABLE IV ANN RESULTS IN PROPOSED DUST STORM CLASSIFICATION

<table>
<thead>
<tr>
<th>Type</th>
<th>CSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe</td>
<td>0.964</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.583</td>
</tr>
<tr>
<td>Light</td>
<td>0.657</td>
</tr>
</tbody>
</table>

The performance of the dust storm prediction model would be improved when the classification method is modified. This improvement is more obvious in severe dust storm. It seems that this is for more reliability in operator of synoptic station in recording. This result shows that other criteria must be defined for more accurate prediction of dust storm events.

IX. CONCLUSIONS AND DISCUSSION

Dust storms are more obstinate and pernicious events than other extreme events, and quantitative dust storm forecasting remains a difficult but vitally important task for climatologists and environment scientists. A number of efforts have been made to improve the accuracy of forecasts using statistical and
nonlinear models. Traditional statistical forecasting methods like linear multiple regressions are unable to capture non-linearity process in climatic variables related to dust storms. In this study a forecasting strategy based on the neural networks is applied to forecast dust storm at 1-day lead time.

With increasing population density and a more mobile society, dust storms may pose an increasing natural hazard. More identification of their formation and extension can be effective in their prediction. Due to the dust storm events were recorded by visibility scope, some errors may be occurred in these records. Most meteorological events can influence on visibility such as local dusts and other aerosols, especially in urban areas. If a light dust storm originated from a source with dark minerals such as Biotite, the visibility can be more decreased if these minerals be light, because low albedo coefficient. Also recording visibility value in night is very erroneous. Thus most reviews and corrections made by operators in synoptic stations can decrease these errors. It seems that presence of more environmental samplers such as dust and sand samplers must be setup in these stations for related reviews. These instruments must be setup and calibrated in different elevations for area under consideration.

X. SUMMARY

The Sistan region of Iran is one of the most active sources of dust storms in Asia. Dust storms occur in this area, all the year round with a maximum frequency of dust storms occurring in spring and summer. A diurnal variation in dust storm frequencies is also noted: dust storms over Sistan occur in most cases during daylight, reaching their peak at 003–006 h which may be explained by increased convective processes in the boundary layer as a result of surface heating during day time. The main factors causing dust storm activity are meteorological, such as wind speed and direction, droughts and high temperatures. These factors are compounded by physiographic conditions: weak fixedness of the soils by vegetation cover, light mechanical texture of the soils. A drastic decrease in frequency of dust storms was observed in Sistan region in the early 1980s. These data are in agreement with similar reports from the other regions of Central Asia (Orolovsky et al., 2004). Such variation was possibly due to periodic changes in global atmospheric circulation. Thus, it is preferable that the frequency of dust storms as an indicator of desertification or rehabilitation processes should be used together with other indicators of land/vegetation degradation as well as long-term observation data. An evaluation of the synoptic climatology of blowing dust in a desert city such as Zabol, main city in this region, may aid the meteorological community in the diagnosis and prognosis of these events, and raise the public’s awareness of potential dust hazards and their causes in order to help residents be better prepared for action when needed, especially in the area of ground and air transportation. A better understanding of the synoptic climatology of dust events and their local source areas could lead to better dust storms forecasting with their positive economic effects. The overall climatological distribution of the various meteorological variables associated with blowing dust events can serve as a basis to formulate forecasting templates for these storms and even potentially act as a guide for long term planning to avoid scheduling of dust sensitive events during times of high probability. We suggest that dust storm events’ codes in meteorological observations must be justified. Also dust storm reports received from other stations, have to be compared to each other and considered in order to correct inaccurate and false codes identified at the station of interest that can improve next dust storm forecasts.

REFERENCES

as inferred from a series of observations in the northeastern Qinghai–Tibetan Plateau, China. Powder Technology 175: 82–89.


