

Multivariate Wind Speed Forecasting with LSTMs for Wind Farm Performance Estimation

Fredy H. Martínez S.^{#1}, César A. Hernández S.^{#2} and Fernando Martínez S.^{#3}

[#] Facultad Tecnológica, Universidad Distrital Francisco José de Caldas
Carrera 13 #40-53, Bogotá D.C., Colombia

¹ fhmartinezs@udistrital.edu.co

² cahernandezs@udistrital.edu.co

³ fmartinezs@udistrital.edu.co

Abstract—One of the most important renewable energy sources today is wind power generation. However, this energy source depends on the flow of air in the area where the wind generators are installed, and as with many other renewable resources, this is a natural resource whose magnitude is not directly controlled by man. The prediction of wind speed then becomes a key problem when we want to project the energy performance of a wind farm. The behaviour of the wind, as a climatic variable, can be estimated from various atmospheric parameters such as temperature, humidity and air pressure. In this paper we propose a multivariate wind speed estimation model from the history of these atmospheric parameters using a deep neural network. The performance of the model is then evaluated against the same historical data, which produces a fairly small error. The code is implemented in Keras deep learning library with TensorFlow backend.

Keyword - Atmospheric parameters, Deep learning, Forecasting, Kinematics, Multivariate model, Wind speed

I. INTRODUCTION

Many developing countries such as Colombia must focus their resources correctly in order to support the development of their individuals [1]. Electricity is a fundamental element for the well-being of individuals and the industrial development of a country, therefore it is also widely used as an indicator of development. Colombia has many areas of the country without electricity service [2]. This is mainly due to the costs of bringing the traditional power system to remote areas with low population density. However, these areas are characterized by own resources that can be redirected to favor the population. There are areas of the country with a very good wind resource, which can be used to generate electricity as an energy solution in specific areas [3, 4, 5]. However, the implementation of this strategy entails many challenges. One of them is related to the projection of wind energy available in the areas. The planning, projection and operation of a wind power plant needs to quantify a priori the available wind resource [6]. This is also important in cases of interconnection with the traditional power grid.

However, predicting wind speed is a fairly complex problem due to its fluctuating behaviour and the number of variables that affect it [7, 8]. Most of the strategies currently used to predict wind speed have faults, and even sometimes incur large errors. Among the conventional strategies used to estimate wind speed is the Numerical Weather Prediction (NWP) method [9]. This method consists in an interpolation of values from known data in nearby places. In order to improve the behaviour of these methods a correction strategy is usually used, one of the best results are recurrent neural networks given their natural ability to learn time series [10].

Another conventional method applied uses the Persistence model [11, 12]. In this case the method is able to provide very good short-term values due to its short-term prediction capability. However, when we want to make projections with a great data history the method fails. Neural networks are also widely used, particularly recurrent ones [13, 14]. These strategies offer very good results, particularly when hybrid methods are implemented in conjunction with statistical tools.

We propose the use of machine learning techniques for the prediction of wind speed, particularly deep neural networks [15]. We use Long Short-Term Memory (LSTM) networks together with a climatological history of 10 years to estimate the short and medium term behaviour of the wind in the region of Guajira (Colombia), the place where the Jepirachi wind farm is installed [16], and which corresponds to one of the areas of the country with the lowest coverage of electrical service [17, 18].

The following part of the paper is arranged in this way. Section 2 presents preliminary concepts and problem formulation. Section 3 illustrates the design profile and development methodology. Section 4 we present the preliminary results. And finally, in Section 5 we present our conclusions.

II. PROBLEM FORMULATION

Jepírachi means northeastern winds (winds coming from the northeast in the direction of Cabo de la Vela) in Wayuunaiki, the native Wayuu language of the indigenous population of the Guajira (Colombia). This is also the name given to the first wind farm for the generation of wind energy built in the country. The wind farm is located in the northeastern region of the Colombian Atlantic Coast, between the towns of Cabo de la Vela and Puerto Bolívar, in the vicinity of Bahía Portete, in the municipality of Uribia. It has an installed capacity of 19.5 MW of nominal power, with 15 wind turbines of 1.3 MW each, subjected to the trade winds that blow almost all year round in this part of the peninsula, at an average of 9.8 meters per second. The machines are distributed in two rows of eight and seven machines respectively, in an area approximately one kilometer long parallel to the beach and 1.2 kilometres wide to the north of the Kasiwolinrancheria and to the west of the Arutkajuirancheria.

This pilot facility also functions as an experimental research center of the Empresas Públicas de Medellín (EPM). Numerous research projects have been proposed on this plant, its operation and development. In our case, we seek to establish the projection and future performance in terms of electricity generation capacity. As part of this work we are developing models of wind behaviour and prediction in the area. From this idea, we propose the use of machine learning strategies for the construction of these models from the historical climatic behaviour.

We are interested in making a multivariable model using a Long Short-Term Memory (LSTM) network. From previous research, and the nature of historical climate data, it is logical to think about using recurrent networks for the development of the model. This persistence of information or memory brings the behaviour of the network to a sequence, a characteristic present in climate data that we want to replicate to estimate the wind speed.

LSTMs are a very special type of recurrent neural networks. In general they work much better than traditional recurrent networks, particularly when it comes to long-term dependencies, in fact, they were specifically designed to avoid this problem. In short, LSTMs are networks capable of remembering information for long periods of time, and they do so naturally. They do this because instead of using a single connection layer for the inputs, she has four interacting with each other in a special way.

Using the LSTM we seek to find a behaviour model for wind speed in the area of operation of the Jepírachi wind farm that contemplates the history of several climatic parameters, including: temperature, humidity, precipitation and air pressure.

III. METHODOLOGY

The Long Short-Term Memory (LSTM) recurrent neural networks are able to almost seamlessly model problems with multiple input variables. For this reason this type of networks become a great tool in time series forecasting, particularly when the problem is characterized by dependence on multiple variables. This is also a great advantage over classical linear methods.

A. Wind Speed Forecasting

In the estimate we are using a weather dataset. These data correspond to daily readings for ten years at the Rioacha airport Almirante Padilla (Guajira, Colombia). The data includes date, temperature, wind speed and weather. The complete feature list in the raw data is as follows:

- No: Row number
- year: Year of data
- month: Month of data
- day: Day of data
- wind: Wind speed [km/h]
- tmax: Maximum temperature in degrees Celsius [C]
- tmin: Minimum temperature in degrees Celsius [C]
- humi: Humidity [%]
- prec: Absolute precipitation [mm]
- Pres: Air pressure [mb]

We use this data to estimate the future behaviour of the wind. That is, given some weather conditions in previous days, what will be the wind speed in the next few days. This dataset can also be used to estimate other variables, however the particular interest of this research is the future estimation of the electrical performance of the Jepírachi wind park.

B. Data Preparation

The raw dataset must first be prepared for use with Pandas. In particular, we must first consolidate the date information in a single column in order to use it later as an index. The original raw dataset is shown in Fig. 1.

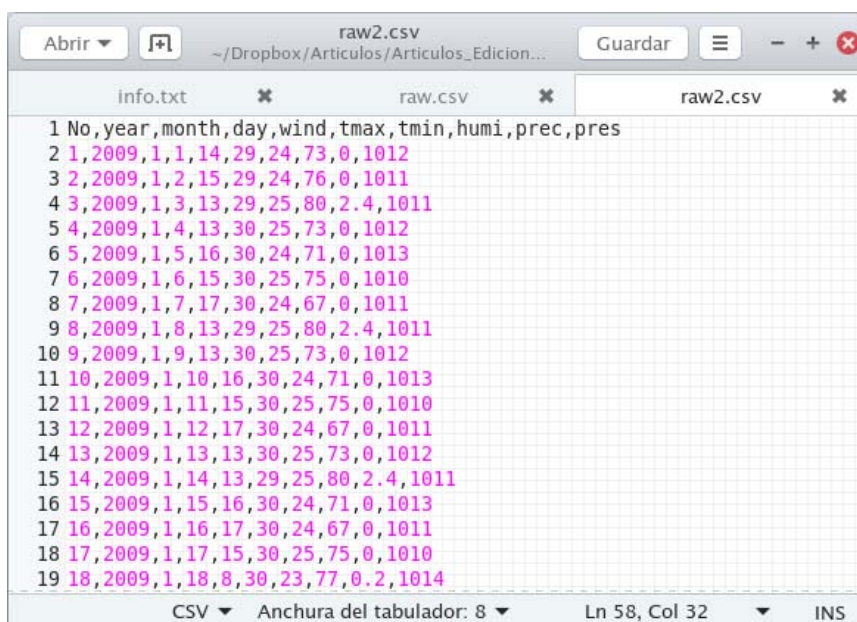


Fig. 1. Original dataset conformed by daily readings of weather of the last 10 years

The script in Fig. 2 loads the raw dataset and parses the date information as the Pandas DataFrame index. Then we remove the "No" column and specify the names for each column. We keep the same names defined before, and print the first 10 resulting rows to verify the result (Fig. 3). The processed dataset is saved again with a new name (raw_2b.csv), but maintaining its format.

```

7 from pandas import read_csv
8 from datetime import datetime
9
10 # Load raw dataset
11 def parse(x):
12     return datetime.strptime(x, '%Y %m %d')
13 dataset = read_csv('raw2.csv', parse_dates = [['year', 'month', 'day']], index_col=0, date_parser=parse)
14 dataset.drop('No', axis=1, inplace=True)
15
16 # Manually specify column names
17 dataset.columns = ['wind', 'tmax', 'tmin', 'humi', 'prec', 'pres']
18 dataset.index.name = 'date'
19
20 # Show first 10 rows to verify results
21 print(dataset.head(10))
22
23 # save to file
24 dataset.to_csv('raw_2b.csv')
    
```

Fig. 2. Script for raw dataset pre-processing. Adjustment of date information and data columns

| date | wind | tmax | tmin | humi | prec | pres |
|------------|------|------|------|------|------|------|
| 2009-01-01 | 14 | 29 | 24 | 73 | 0.0 | 1012 |
| 2009-01-02 | 15 | 29 | 24 | 76 | 0.0 | 1011 |
| 2009-01-03 | 13 | 29 | 25 | 80 | 2.4 | 1011 |
| 2009-01-04 | 13 | 30 | 25 | 73 | 0.0 | 1012 |
| 2009-01-05 | 16 | 30 | 24 | 71 | 0.0 | 1013 |
| 2009-01-06 | 15 | 30 | 25 | 75 | 0.0 | 1010 |
| 2009-01-07 | 17 | 30 | 24 | 67 | 0.0 | 1011 |
| 2009-01-08 | 13 | 29 | 25 | 80 | 2.4 | 1011 |
| 2009-01-09 | 13 | 30 | 25 | 73 | 0.0 | 1012 |
| 2009-01-10 | 16 | 30 | 24 | 71 | 0.0 | 1013 |

Fig. 3. Result of dataset adjustment for later analysis

To facilitate the analysis of behaviour, we made plots of each series within the dataset (Fig. 4). Again the curves have been organized in the same order, from top to bottom: wind, tmax, tmin, humi, prec and pres. The figure shows the repetitive cyclic behaviour of the data, some dependence that relates them, and some atypical data.

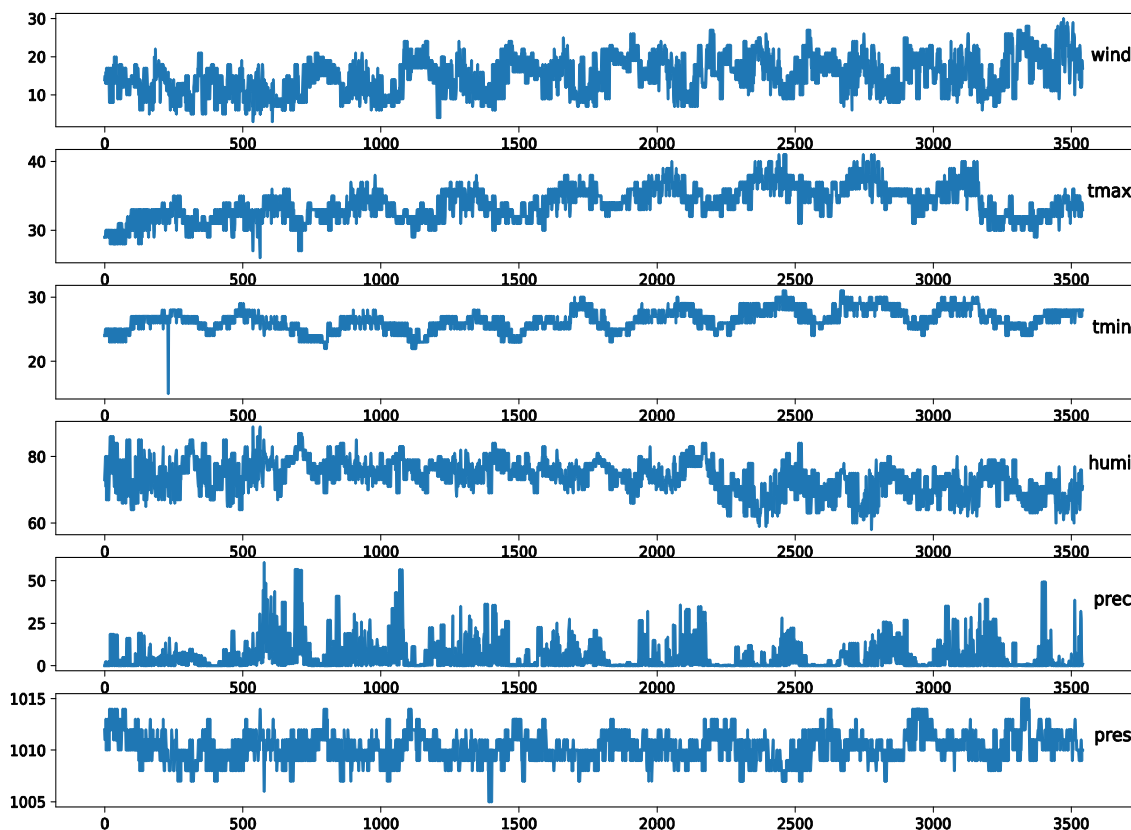


Fig. 4. Behaviour of each column of data within the dataset. From top to bottom: wind, maximum temperature, minimum temperature, humidity, precipitation and pressure

C. Multivariate LSTM Forecast Model

In order to make the model fit it is first necessary to formulate the prediction as a supervised learning problem. As an initial test we used the data to predict the behaviour of the wind (wind speed) on the current day (instant t) according to the values of wind and weather of the previous day (instant $t-1$). This is only one option that allows to establish a solution, and that coincides with the objective of establishing the daily performance of the wind farm. Other possible formulation alternatives (and maybe more interesting) could be to make the prediction but considering a longer data history (one month, one year, for example), or project some other weather parameter.

To process the data, we first normalize all input variables in the range 0 to 1. Fig. 5 shows the first five rows of the transformed dataset and the prediction of the wind variable. We show the six input variables at instant $t-1$ (input values), and the output variable at instant t (wind speed on the current day).

| | var1($t-1$) | var2($t-1$) | var3($t-1$) | var4($t-1$) | var5($t-1$) | var6($t-1$) | var1(t) |
|---|---------------|---------------|---------------|---------------|---------------|---------------|-------------|
| 1 | 0.407407 | 0.200000 | 0.5625 | 0.483871 | 0.000000 | 0.700005 | 0.444444 |
| 2 | 0.444444 | 0.200000 | 0.5625 | 0.580645 | 0.000000 | 0.599998 | 0.370370 |
| 3 | 0.370370 | 0.200000 | 0.6250 | 0.709677 | 0.088889 | 0.599998 | 0.370370 |
| 4 | 0.370370 | 0.266667 | 0.6250 | 0.483871 | 0.000000 | 0.700005 | 0.481481 |
| 5 | 0.481481 | 0.266667 | 0.5625 | 0.419355 | 0.000000 | 0.800003 | 0.444444 |

Fig. 5. Prediction of one day with respect to the previous day. The first six columns show the normalized variables at instant $t-1$. The last column shows the prediction at instant t

From this point the model behaviour analysis is performed, examining wind speed predictions using as input variables data from more than one day in the past. The objective is to exploit the time backpropagation of the LSTM network for the projection of the time sequence. The LSTM-based model has multiple input variables, and the dataset is used for both training and model evaluation.

We fit the model on the first eight years of data, and then evaluate it on the remaining three years of data. In this sense we divide the dataset into training and test sets. Then, we divide these two sets into input and output variables. Finally, the inputs are reshaped into the 3D format expected by LSTMs, i. e., [samples, timesteps, features]. The script in Fig. 6 takes care of these tasks. The final format of the inputs is (Eq. 1):

$$((2920, 1, 6), (2920,)), (621, 1, 6), (621,)) \quad (1)$$

That is, 2920 days of data for training, and 621 days of data for testing.

```

55 # Split dataset into train and test sets
56 values = reframed.values
57 n_train_days = 365 * 8
58 train = values[:n_train_days, :]
59 test = values[n_train_days:, :]
60
61 # Split set into input and outputs
62 train_X, train_y = train[:, :-1], train[:, -1]
63 test_X, test_y = test[:, :-1], test[:, -1]
64
65 # Reshape input to be 3D [samples, timesteps, features]
66 train_X = train_X.reshape((train_X.shape[0], 1, train_X.shape[1]))
67 test_X = test_X.reshape((test_X.shape[0], 1, test_X.shape[1]))
68 print(train_X.shape, train_y.shape, test_X.shape, test_y.shape)
    
```

Fig. 6. Splitting data in training and testing, and into input and output variables

We defined the LSTM network with 200 neurons in the first hidden layer, and one neuron in the output layer to predict wind speed. The input shape will be one time step with six features. We use the Mean Absolute Error (MAE) loss function and the efficient Adam version of stochastic gradient descent. The model will be fit for 100 training epochs with a batch size of 72. The script in Fig. 7 takes care of these tasks.

```

74 # Design of the LSTM network
75 model = Sequential()
76 model.add(LSTM(200, input_shape=(train_X.shape[1], train_X.shape[2])))
77 model.add(Dense(1))
78 model.compile(loss='mae', optimizer='adam')
79
80 # Fit of the network
81 history = model.fit(train_X, train_y, epochs=100, batch_size=72, validation_data=(test_X, test_y), verbose=2, shuffle=False)
    
```

Fig. 7. Design of the deep neuronal network type LSTM

IV. RESULTS AND DISCUSSION

We developed a wind speed estimation model for the Jepirachi wind farm (Guajira, Colombia) using the Keras (2.1.5) deep learning library with TensorFlow (1.6.0) backend. We use an environment with Python 2.7.14 (GCC 7.2.0), SciPy (1.0.0), scikit-learn (0.19.1), Pandas (0.22.0), NumPy (1.14.1) and Matplotlib (2.0.0). The raw dataset (in total 3542 daily readings corresponding to the last 10 years, from 1 January 2009 to 12 September 2018) correspond to measurements made at the Rioacha airport Almirante Padilla (Guajira, Colombia). The site is located 13 m above sea level to the southwest of the city.

We use the model described in the previous section to test the entire dataset. We combine the forecast with the test dataset and invert the scaling. We also invert scaling on the test dataset with the expected wind speed values. With forecasts and actual values in their original scale, we can calculate an error score for the model. In this case, we calculate the Root Mean Squared Error (RMSE) that gives error in the same units as the variable itself. The script in Fig. 8 takes care of these tasks.

```

90 # Invert scaling for forecast
91 inv_yhat = concatenate((yhat, test_X[:, 1:]), axis=1)
92 inv_yhat = scaler.inverse_transform(inv_yhat)
93 inv_yhat = inv_yhat[:,0]
94
95 # Invert scaling for actual
96 test_y = test_y.reshape((len(test_y), 1))
97 inv_y = concatenate((test_y, test_X[:, 1:]), axis=1)
98 inv_y = scaler.inverse_transform(inv_y)
99 inv_y = inv_y[:,0]
100
101 # Calculate RMSE
102 rmse = sqrt(mean_squared_error(inv_y, inv_yhat))
103 print('Test RMSE: %.3f' % rmse)
    
```

Fig. 8. Adjusting data scales and calculating Root Mean Squared Error (RMSE)

In the end, the model's RMSE was 4.223 km/h. Given the speed values, this is a fairly small error for the model. We also plot the curves of train and test loss from the LSTM during training (Fig. 9).

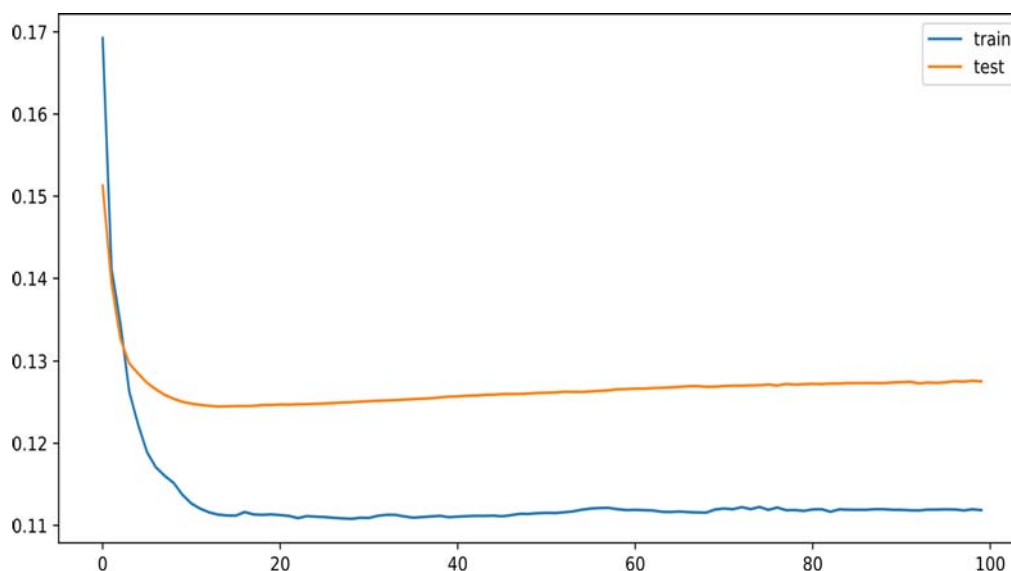


Fig. 9. Plot of train and test loss from the multivariate LSTM model during training

V. CONCLUSION

In this paper we developed a multivariate model for the estimation of wind speed using LSTM networks. The model was trained and validated using weather data from the last 10 years recorded at the Rioacha airport Almirante Padilla (Guajira, Colombia). The wind prediction is performed as an input for the performance estimation of the Jepírachi wind farm located in this city. The model was validated with the same training data (18% of the data) and the results were submitted to the Root Mean Squared Error (RMSE) calculation. The RMSE value of the model was 4.223 km/h, a value that guarantees a good performance of the model.

ACKNOWLEDGMENT

This work was supported by the Universidad Distrital Francisco José de Caldas, in part through CIDC, and partly by the Facultad Tecnológica. The views expressed in this paper are not necessarily endorsed by Universidad Distrital. The authors thank the research group ARMOS for the evaluation carried out on prototypes of ideas and strategies.

REFERENCES

- [1] K. Surendra, D. Takara, A. Hashimoto and S. Kumar, Biogas as a sustainable energy source for developing countries: Opportunities and challenges, *Renewable and Sustainable Energy Reviews*, vol. 31(1), 2014: pp. 846–859.
- [2] E. Gaona, C. Trujillo and J. Guacaneme, Rural microgrids and its potential application in Colombia, *Renewable and Sustainable Energy Reviews*, vol. 51(1), 2015: pp. 125–137.
- [3] T. Ackermann, G. Ancell, L. Borup, P. Eriksen, B. Ernst, F. Groome, M. Lange, C. Mohrlen, A. Orths, J. O'Sullivan and M. Torre, Where the wind blows, *IEEE Power and Energy Magazine*, vol. 7(6), 2009: pp. 65–75.
- [4] P. Maegaard, Wind energy development and application prospects of non-grid-connected wind power, 2009 World Non-Grid-Connected Wind Power and Energy Conference, 2009, pp. 1–3.

- [5] A. Kane, Prospects and status of non-grid-connected wind power all over the world, 2010 World Non-Grid-Connected Wind Power and Energy Conference, 2010, pp. 1–3.
- [6] Z. Haoran, W. Qiuwei, H. Shuju, X. Honghua and N. Claus, Review of energy storage system for wind power integration support, *Applied Energy*, vol. 137(1), 2015: pp. 545–553.
- [7] N. Ivo and L. Jing, Modeling and forecasting the mean hourly wind speed time series using gmdh-based abductive networks, *Renewable Energy*, vol. 34(7), 2009: pp. 1686–1699.
- [8] M. Wani and H. Bhat, One-against-all and one-against-one multiclass support vector machine algorithms for wind speed prediction, *International Journal of Renewable Energy Research IJREER*, vol. 8(2), 2018: pp. 909–915.
- [9] T. Wardah, A. Kamil, A. Sahol and W. Maisarah, Statistical verification of numerical weather prediction models for quantitative precipitation forecast, 2011 IEEE Colloquium on Humanities, Science and Engineering, 2011, pp. 88–92.
- [10] Y. Chanjong, A. Heewoong and S. Junhee, Coordinate-rnn for error correction on numerical weather prediction, *International Conference on Electronics, Information, and Communication (ICEIC 2018)*, 2018, pp. 1–3.
- [11] N. Korprasertsak and T. Leephakpreeda, Comparative investigation of short-term wind speed forecasting models for airborne wind turbines, 8th International Conference on Mechanical and Aerospace Engineering (ICMAE 2017), 2017, pp. 451–454.
- [12] C. Walt and N. Botha, A comparison of regression algorithms for wind speed forecasting at alexander bay, *Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, 2016, pp. 1–5.
- [13] S. Balluff, J. Bendfeld and S. Krauter, Short term wind and energy prediction for offshore wind farms using neural networks, *International Conference on Renewable Energy Research and Applications (ICRERA 2015)*, 2015, pp. 379–382.
- [14] T. Senjyu, A. Yona, N. Urasaki and T. Funabashi, Application of recurrent neural network to long-term-ahead generating power forecasting for wind power generator, *IEEE PES Power Systems Conference and Exposition*, 2006, pp. 1260–1265.
- [15] Q. Xiaoyun, K. Xiaoning, Z. Chao, J. Shuai and M. Xiuda, Short-term prediction of wind power based on deep long short-term memory, *IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC 2016)*, 2016, pp. 1148–1152.
- [16] Y. Ulianov and S. Lafn, Small signal stability analysis of Jepirachi wind park, *IEEE International Symposium on Alternative Energies and Energy Quality (SIFAE 2012)*, 2012, pp. 1–6.
- [17] E. Ojeda, H. Hernandez, L. Bedoya, A. Barrios and J. Candelo, Strategies applied for renewable energy source adoption in indigenous communities of la Guajira, Colombia, *International Journal of Engineering and Technology*, vol. 8(6), 2017: pp. 2689–2695.
- [18] L. Rodriguez, J. Valencia, D. Rodriguez and A. Martinez, Design, implementation and operation of a solar hybrid system in a remote area in the Colombia Guajira desert, *WIT Transaction on Ecology and the Environment*, vol. 195(1), 2015: pp. 427–438.

AUTHOR PROFILE

Fredy Hernán Martínez Sarmiento. PhD in Engineering - Systems and Computation, Professor at the Universidad Distrital Francisco José de Caldas, Director and principal researcher of the ARMOS research group.

César Augusto Hernández Suarez. PhD in Engineering - Systems and Computation, Professor at the Universidad Distrital Francisco José de Caldas, Researcher of the ARMOS research group.

Fernando Martínez Santa. MSc in Electronic and Computer Engineering, Professor at the Universidad Distrital Francisco José de Caldas, Researcher of the ARMOS research group.