International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

December 2024	Issue 61	Volume 16	Number 4	Pages 129-138

HYBRID ARIMA-RANDOM FOREST MODEL FOR GEOMAGNETIC K_P INDEX PREDICTION USING HILBERT TRANSFORMATION

B. Chafik M. El Malki F. Miskane H. Nebdi

Laboratory of Innovation in Science, Technology and Modeling (LISTM), Faculty of Sciences, University of Chouaib Doukkali, El Jaida, Morocco
chafik.b@ucd.ac.ma, fmiskane@gmail.com, nebdi-hamid@yahoo.fr, elmalki.med@ucd.ac.ma

Abstract- The objective of this article is to study the geomagnetic index K_{pa} with a view to modelling its evolution over a long period and predicting its daily values. The K_{pa} index is a key indicator for assessing geomagnetic disturbances that seriously affect modern technological infrastructures such as GPS systems, communication networks, satellites, power grids and human activities. Predicting geomagnetic activity is extremely important for taking the necessary precautions during intense geomagnetic events. In this article, we imported and analyzed daily K_{pa} index data collected over 60 years from a specialized database. We used artificial intelligence techniques, in particular ARIMA model and combined Random Forest and ARIMA model, to model and predict K_{pa} index variations. The first part of our study consists of using an ARIMA model in order to model the evolution of the K_p index, which reveals non-linear characteristics and complex interactions at different time scales. In the second part, to overcome the limitations of the ARIMA model, we introduce an adjustment by a Random Forest model, a robust machine learning technique that captures non-linear relationships in the data. Before implementing Random Forest, we apply a Hilbert transformation to time series in order to extract additional temporal and frequency characteristics, which improves the representation of the data and the predictive ability. The significant results obtained will contribute to better anticipation of geomagnetic disturbances and protection of critical infrastructures.

Keywords: Geomagnetic Index, K_{pa} Index, Prediction, Artificial Intelligence, ARIMA, Random Forest, Extreme Events, Hilbert.

1. INTRODUCTION

The K_p geomagnetic index is a key measure of the Earth's geomagnetic activity, used to evaluate disturbances as a result of wind interactions with the terrestrial magnetic field. It was introduced in 1939 by Julius Bartels and is based on observations made by magnetometers located in several stations distributed geographically around the globe [1]. This index quantifies the intensity of geomagnetic storms on a scale from 0 to 9, where 0 represents very calm conditions and 9 extreme geomagnetic storms [2, 3]. The K_p index is based on measurements of variations in the Earth's magnetic field that are registered every three hours. These variations are expressed as a function of the amplitude of disturbances in two main components: the horizontal component (north-south) and the vertical component (east-west). In this way, the K_{pa} index integrates global observations to provide a general indicator of the intensity of geomagnetic activity in the Earth's atmosphere [4].

The dependences on the K_{pa} index are many and varied, covering scientific and technical fields that are sensitive to space conditions. For space weather forecasting and infrastructure protection, the K_{pa} index plays a very important role in predicting solar activity and geomagnetic storms. These events, often linked to coronal mass ejections (CMEs) or solar flares, have a significant impact on the Earth's technological infrastructures. A major geomagnetic storm can generate induced geomagnetic currents (GICs) in power grids, leading to serious disruption and even blackouts [5].

Power companies and communication network operators therefore depend on indexes such as K_p to monitor the risks of surcharges and protect their critical systems. In addition, in polar zones, induced currents can also affect pipelines, accelerating their corrosion [6]. Positioning System) systems, are strongly affected by geomagnetic activity. During periods of strong solar storms, the Earth's ionosphere becomes disrupted, creating irregularities in the propagation of GPS signals. This results in positioning errors, affecting sectors that rely on precise geolocation, such as aviation, military operations and maritime navigation [7-9]. The K_{pa} index therefore serves as a predictive indicator enabling operators of these systems to take preventive measures to mitigate positioning errors during periods of high geomagnetic activity [10].

The K_{pa} index is commonly used to predict and observe the polar aurora. Solar wind charged particles can enter into interactions with the Earth's atmosphere, creating disturbances in the magnetosphere. These interactions can also generate polar aurorae, which are visible around the Earth's magnetic poles during days of intense geomagnetic

Registered: 240717 Article Info: 1907 Accepted: 241129
Reserved: 241024 Available: 241231

activity. The K_{pa} index can therefore be used to predict the conditions that are favorable for the appearance of auroras, which is invaluable not only for scientists but also for astronomy enthusiasts and tourists. The K_{pa} index is also used to assess the risks associated with increased exposure to radiation during high-altitude commercial flights, particularly in the polar regions. During periods of high geomagnetic activity, more ionizing radiation can enter the Earth's atmosphere, affecting long-haul flights and astronauts in orbit [11]. In this context, space agencies and airlines monitor the K_{pa} to adjust routes and protect crew and equipment [12].

The geomagnetic K_{pa} index is an indispensable tool for understanding and managing the complex interactions between the Sun and the Earth. From the study of auroras to the protection of critical infrastructures and navigation systems, this index plays a vital role in preventing the risks associated with geomagnetic disturbances. The K_{pa} index predictions are essential for assessing risk of geomagnetic storms affecting vulnerable technologies such as satellites, navigation and communication systems based on radio signals, and space communications. Its importance in space surveillance is undeniable, it is a key parameter in space weather, where it helps monitoring solar wind interactions with the Earth's magnetosphere, and space weather research will continue to refine its use to better protect our future technologies and infrastructures [13].

Previous studies have revealed significant correlations between the K_p index and the different phases of each solar cycle [14]. The combination of the ARIMA and Random Forest models for predicting the K_p geomagnetic index offers a unique approach compared with other domains such as finance or human activity. ARIMA is good at capturing linear trends and seasonality, which are fundamental to the prediction of recurring geomagnetic phenomena, while Random Forest is better at detecting sudden, non-linear variations, which are characteristic of disturbances resulting from geomagnetic events. Combined, these models improve prediction accuracy by minimizing errors, adjusting to complex variability and exploiting larger volumes of data, which makes this approach especially useful for geomagnetic phenomena.

Incorporating the Hilbert transform into the modelling of geomagnetic disturbances with the ARIMA and Random Forest models offers significant advantages compared to applying them without pre-processing. The Hilbert transform extracts important information about the instantaneous phase and amplitude of the signals, which makes the models more able to capture non-linear and non-stationary signals, typical of variations in the K_{pa} index. It also helps to better detect abrupt changes and reduce noise in the data, which improves the global accuracy of predictions by enabling ARIMA to better capture long-term trends and Random Forest to better model complex relationships. We have tested multiple combined models, such as LSTM, KNN, RNN, Boost, etc., to predict K_p index over time, and evaluated their accuracy using metrics. We found that the hybrid model incorporating the Hilbert transform, Random Forest and ARIMA performed best, offering greater accuracy in capturing linear and non-linear trends in the K_{pa} index. In this contribution we will mathematically model the K_{pa} index based on data provided by a GFZ Potsdam, and we will study these variations in order to establish an empirical model, with the aim of understanding how this physical parameter varies on the one hand and predicting future values on the other.

In the Materials and methods section, we will describe the computer tools and mathematical equations that we used to process our data, as well as the metrics for evaluating the performance of the chosen models. In the Results and Discussion section, we will present the results we found and analyses, discuss and interpret them. At the end, we will conclude and give perspectives of our work.

2. MATERIALS AND METHODS

We used data from GFZ Potsdam [15]. These data have enabled us to process K_{pa} data from 1964 to 2024.

The K_{pa} values are measured with a time step of 3 hours, giving 8 values per day. To predict the K_p index for each day, we will take the maximum value measured during each day. To achieve our objective, we will use the Python language, which allows us to test different models adapted to our problem. We have tested several methods for analyzing the data and predicting future K_p values. These include LSTM, Boost, RNN, KNN, linear regression, logarithmic regression, etc., but the best results were obtained using the ARIMA and Random Forest models combined with the Hilbert transformation.

ARIMA (Autoregressive Integrated Moving Average) is an artificial intelligence and statistical modelling method developed for studying and forecasting time series (TM). This model is composed on 3 key elements combined in a structured way to analyze and predict TM [16, 17]:

- Autoregression (*AR*)
- Integration (*I*)
- Moving Average (MA)

ARIMA is widely used in various scientific and industrial fields to model time series, such as geomagnetic indexes like K_p [18]. Autoregression (*AR*) is an approach founded on an idea where the predicted future values are dependent on their past values. The *AR*(*p*) model predicts every value by using a linear combination of the time series historical (past) data. The Equation (1) expresses the values prediction by this autoregressive method [19]:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \tag{1}$$

where, X_t is data at instant *t*, c is correction coefficient, φ_i is autoregression coefficient, and ε_t is the random error or white noise.

Integration (*I*) in ARIMA is used to make a time series stationary. When the statistical parameters (mean, variance) of a time series remain constant over time, it is considered to be stationary. The ARIMA model includes an integration process of order *d*, which consists of differentiating the data several times to make the series stationary. The Equation (2) describes the differentiation operation [16]:

$$Y_t = X_t - X_{t-1} \tag{2}$$

In the case of a non-stationary time series, the model can be differentiated d times to make it stationary. The moving average (MA) model considers past error in future prediction. The MA (q) model depends on past errors (white noise) and predicts values based on residuals based on Equation (3) [19]:

$$X_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \tag{3}$$

where, μ is mean of the series, and θ_i is moving average coefficients.

The described ARIMA(p, d, q) model combines the three main component parts: the AR(p) representing the autoregressive part, $I(d)$ corresponding to the integration (differentiation) component and the MA(q) for the moving average part [17]. As expressed in Equation (4), the model can effectively represent the temporal dependencies in the time series data [16, 19]:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \sum_{i=1}^{p} \varphi_i X_{t-i} \tag{4}$$

The ARIMA requires three parameters to be specified:
- p: number of AR terms
- q: number of terms MA

ARIMA modelling involves three steps:
- Identifying stationarity: The first step is to verify the condition of time series stationarity. If that is not true, we then proceed to differentiation.
- Parameter selection: Select orders p, d and q by employing statistical factor like the Akaike Information Criterion (AIC) [20].
- Parameter estimation: The model estimates the AR and MA coefficients using methods such as maximum likelihood.
- Model validation: Verification of the residuals (the prediction errors) to ensure that they approximate white noise, which means a good fit of the model.
- Prediction: Once the model has been fitted, it can be used to predict future values in the time series.

In the context of the K_{pa} index, the ARIMA model can be applied for predicting the future variation of this geomagnetic index, which is essential for various applications. The Hilbert transform is used to transform a real signal into a complex analytical signal, making it easier to analyze characteristics such as instantaneous amplitude and frequency. The Equation (5) expresses the analytical signal $\hat{x}_a(t)$ [21]:

$$\hat{x}_a(t) = x(t) + j\hat{x}(t) \tag{5}$$

where, $x(t)$ represents the real part of the input signal and Hilbert transform is represented by $\hat{x}(t)$ [22].

The original signal is decomposed in three steps:
- The Fourier Transform: This stage consists of transforming the time signal $x(t)$ into the frequency range using the Discrete Fourier Transform (DFT), often calculated digitally in accordance with the Fast Fourier Transform (FFT) which transforms a temporal signal to a frequential signal based on Equation (6) [23]:

$$X(f) = FFT[x(t)] = \sum_{k=0}^{N-1} x(k).e^{-j.2\pi f.k/N} \tag{6}$$

where, N is the signal length, $x(k)$ is the time signal at index k, and $e^{-j.2\pi f.k/N}$ is the complex exponential element that decomposes the signal into which decomposes the signal into a sinusoidal sum.
- Frequency filtering: The aim is to create an analytical signal (no negative components). This step removes the negative frequencies, while they are multiplied twice in the case of positive frequencies based on Equation (7) [21]:

$$H(f) = \begin{cases} 2X(f)\ ,\ 0<f<\dfrac{N}{2} \\ X(f)\ ,\ f = 0 \text{ and } f = \dfrac{N}{2} \\ 0\ ,\ \dfrac{N}{2} < f < N \end{cases} \tag{7}$$

- Inverse Fast Fourier Transform (IFFT): The Equation (8) represents the inverse Fourier transform (IFFT) which re-converts the filtered spectrum back into a temporal signal. It recreates the signal in the temporal domain, by merging the transformed frequencies in the previous steps. This transform is used to find the Hilbert transform $\hat{x}(t)$ of the initial signal $x(t)$ as shown in Equation (8) [24]:

$$\hat{x}(t) = IFFT[H(f)] = \frac{1}{N}\sum_{f=0}^{N-1} H(f).e^{j2\pi ft/N} \tag{8}$$

To summaries this approach, the Hilbert transform consists of converting a signal to frequency domain, eliminating negative frequencies and then returning to the time domain. This process creates an analytical signal that reveals the instantaneous amplitude and phase, which is advantageous for studying complex signals such as the evolution of geomagnetic activity and, in particular, the K_p index.

To evaluate the model's performance, we will need the metrics such as r, MAE, R^2, RMSE and MBE [25, 26]. The Pearson correlation coefficient r is the correlation used to measure the linear relationship over time that exists between two variables. The formula used to calculate the correlation coefficient between the X_{me} and X_{pr} series is expressed by Equation (9) [27]:

$$\rho = \frac{\sum_{i}^{N} (X_{pr}i - \bar{X}_{pr})(X_{me}i - \bar{X}_{me})}{\sqrt{\left[\sum_{i}^{N} (X_{pr}i - \bar{X}_{pr})^2 \times \sum_{i}^{N} (X_{me}i - \bar{X}_{me})^2\right]}} \tag{9}$$

where, $X_{me}i$ are measured values, $X_{pr}i$ are predicted values, $\bar{X}_{me}$ is the mean of measured values, $\bar{X}_{pr}$ is the mean value of predicted values, and N is the points data number.

- Determination coefficient (R^2): The determination coefficient (R^2) calculates the fraction of variance represented a given by model. It is obtained by comparing the sum of the residual squares to the sum of the total squares as described in Equation (10) [28, 29]:

$$R^2 = 1 - \frac{\sum (X_{me}i - X_{pr}i)^2}{\sum (X_{me}i - \bar{X}i)^2} \tag{10}$$

where, $X_{me}i$ is measured value, $X_{pr}i$ is predicted value, and $\bar{X}i$ is the measured values average.

- *MAE*: The average absolute error of predicted and observed values is measured using *MAE* metric with the Equation (11) [29]:

$$MAE = \left(\frac{1}{N}\right) \times \sum \left| X_{me}i - X_{pr}i \right| \tag{11}$$

- *RMSE*: This metric is determined by calculating the square root of *MAE* and indicates the residuals dispersion. The formula (12) show how calculate this indicator [29]:

$$RMSE = \sqrt{\left[\left(\frac{1}{N}\right) \times \sum \left(X_{me}i - X_{pr}i \right)^2\right]} \tag{12}$$

- *MBE*: measures the average bias between measured and predicted values. It is calculated as shown in formula (13) below [29]:

$$MBE = \left(\frac{1}{N}\right) \times \sum \left(X_{me}i - X_{pr}i \right) \tag{13}$$

To develop our work, we implemented an algorithm in Python according to the following instructions:
- Import the K_{pa} data from the Excel file;
- Plot the distribution of the values given by the imported K_{pa} index;
- Test several learning methods and rank them according to the correlation coefficient;
- Testing, evaluating and choosing the best parameters;
- Give the correct mathematical equation for each method;
- Plot the evolution of the predicted and measured K_{pa} values over time on the same graph;
- Perform a total statistical analysis and comparing of the predicted to measured values for each of the last five solar cycles;
- After validating certain criteria, the algorithm predicts the K_{pa} index for the year 2024;
- Perform the Hilbert decomposition
- Adjust the predictions using the combined Random Forest and ARIMA model;
- Calculate statistical metrics.

The algorithm used for testing and selecting the best parameters for ARIMA is described in the flowchart on Figure 1.

3. RESULTS AND DISCUSSIONS

As we mentioned already, our objective in this article is divided into two main areas: the first is to model the evolution of the K_p index mathematically, and the second is to predict the future values of this index. Before using the (p, d, q) parameters of ARIMA, we created an algorithm that automatically searches for the combination of parameters that gives the best determination coefficient found (Figure 1). After executing the loop, we have found that the best combination is (6, 0, 1). We first studied the distribution of the K_{pa} index according to frequency of occurrence. The following graph shows the distribution of measured K_p values.

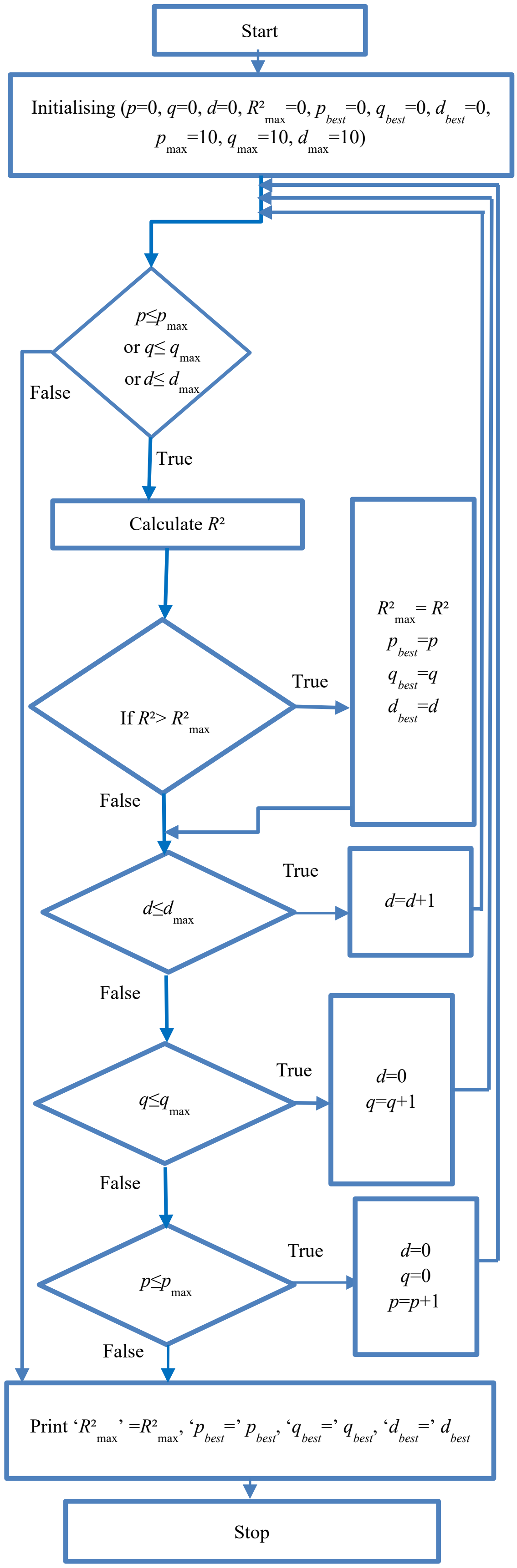


Figure 1. Flowchart of loops implemented to find the best combination of ARIMA parameters (p, q, d)

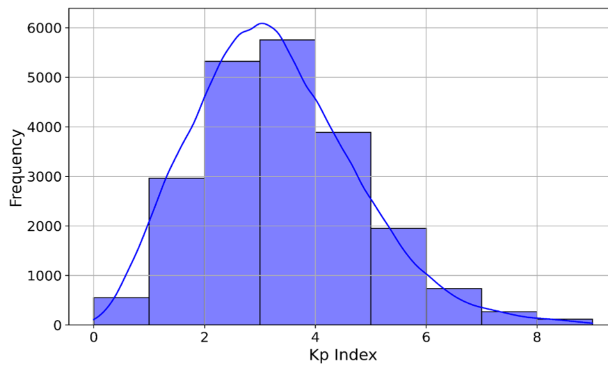


Figure 2. Distribution of measured K_{pa} index values

Table 1. Numerical distribution of K_{pa} values

K_p interval	Percentage (%)
0.0 - 1.0	5.75
1.0 - 2.0	17.77
2.0 - 3.0	27.25
3.0 - 4.0	24.23
4.0 - 5.0	14.72
5.0 - 6.0	6.69
6.0 - 7.0	2.34
7.0 - 8.0	0.90
8.0 - 9.0	0.35

The Table 1 shows the percentage of K_{pa} index values classified according to ranges with a scale of 1. Analysis of the graph above and table 1 shows that the distribution of K_{pa} values is centered on a range of 1-5. Previous studies have shown that extreme events accompanied by intense geomagnetic disturbances are observed on days when the K_{pa} index is greater than or equal to 7 [7]. We find that in the database we have 392 active days from 1/10/1964 to 31/08/2024. The algorithm implemented according to the ARIMA model obtained gives Equation (14):

$$Kp(t) = 1.9171 + \sum_{1}^{6} a_i \times Kp(t-i) \tag{14}$$

where, i=1, 2, 3, 4, 5 and 6.

The coefficients as a function of the K_{pa} values over the past 6 days are given in Table 2. After plotting the curve obtained by a numerical application of the Equation (14), we found that there was a time difference of one day in relation to the measured data.

Table 2. ARIMA Coefficients

i	Coefficients α_i
1	0.85623
2	−0.33274
3	0.05366
4	−0.01726
5	0.01566
6	−0.00520

To reduce this difference, we chose to make the following adjustment: after calculation using Equation (14), we will replace the values by $K_p(t_i) = K_p(t_{i+1})$, $K_p(t_{i+1}) = K_p(t_{i+2})$. Table 3 summarizes the values of the statistical metrics obtained by comparing the measured values of the K_{pa} index with those predicted during the last 5 cycles and the current cycle.

Table 3. Metrics between measured and calculated values

Indicators	Values
r	0.94
R^2	0.71
MAE	0.78
MBE	0.51

The Table 4 gives the values of the statistical metrics comparing the values of the K_{pa} index measured and those calculated by the mathematical model for each solar cycle (20th, 21st, 22nd, 23rd, 24th and 25th cycle).

Table 3. Statistical indicators by solar cycle

Solar cycle	r	R^2	MAE	$RMSE$	MBE
20	0.93	0.72	0.62	0.74	0.47
21	0.93	0.79	0.53	0.64	0.31
22	0.94	0.77	0.57	0.69	0.36
23	0.94	0.70	0.68	0.81	0.54
24	0.93	0.50	0.84	0.95	0.78
25	0.93	0.55	0.78	0.88	0.71

Analysis of Table 4 shows a significant correlation for the predicted K_{pa} values calculated by the mathematical equation based on the ARIMA method using the historical data compared with those measured. The following figures show the values calculated by the ARIMA model compared with the measured values.

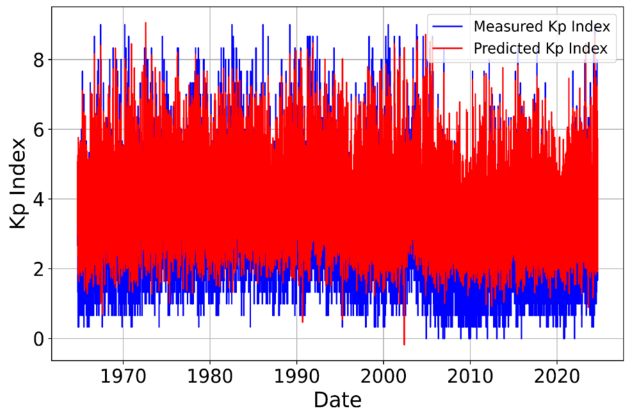


Figure 3. Comparison of predicted values $K_{pa}(pred)$ by the ARIMA model and measured values $K_{pa}(meas)$ over the last 5 solar cycles

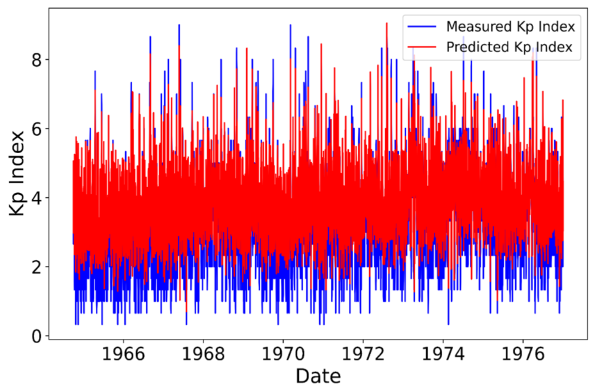


Figure 4. Comparison of $K_{pa}(meas)$ and $K_{pa}(pred)$ by the ARIMA model during the 20th solar cycle

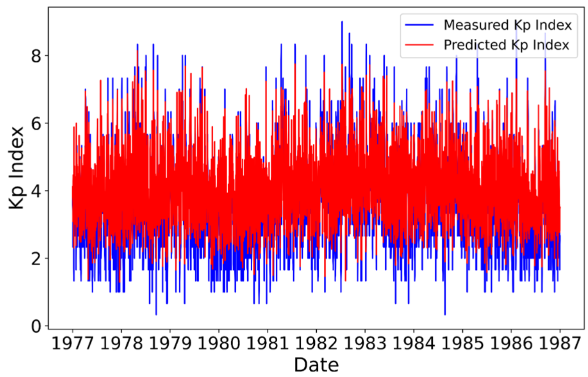


Figure 5. Comparison of $K_{pa}(meas)$ and $K_{pa}(pred)$ by the ARIMA model during the 21st solar cycle

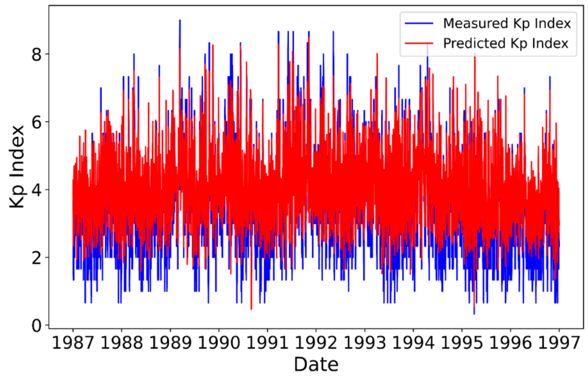


Figure 6. Comparison of $K_{pa}(meas)$ and $K_{pa}(pred)$ by the ARIMA model during the 22nd solar cycle

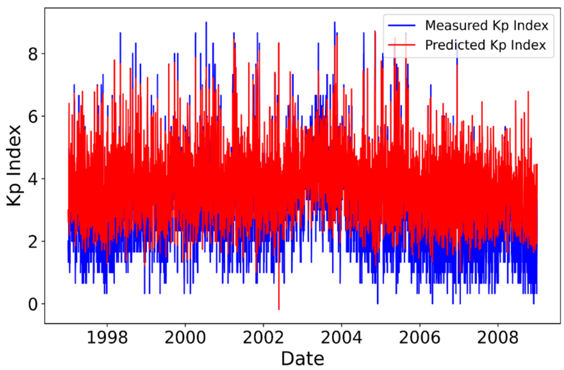


Figure 7. Comparison of $K_{pa}(meas)$ and $K_{pa}(pred)$ by the ARIMA model during the 23rd solar cycle

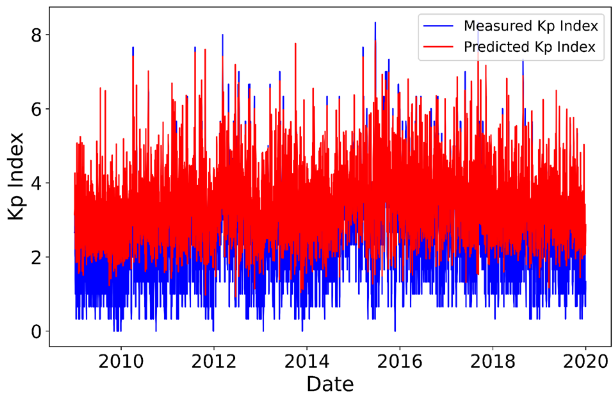


Figure 8. Comparison of $K_{pa}(meas)$ and $K_{pa}(pred)$ by the ARIMA model during the 24th solar cycle

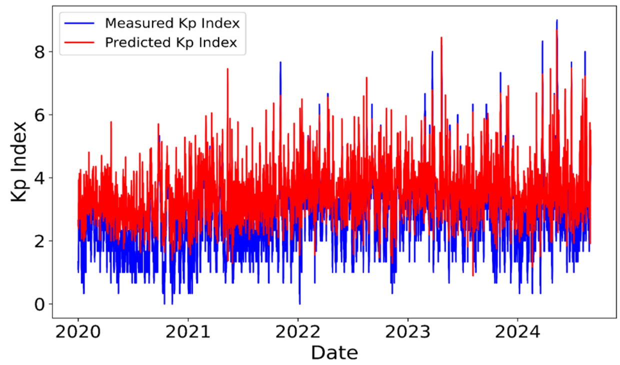


Figure 9. Comparison of $K_{pa}(meas)$ and $K_{pa}(pred)$ by the ARIMA model during the 25th solar cycle

As already explained, to reduce the difference between the predicted and the measured values, we used an ARIMA-adjusted Random Forest model. This combined approach to predict the K_{pa} index with a daily time step is to first process the measured data by Hilbert Transform to extract information about the instantaneous amplitude and phase. This allows the capturing the fine details of the signal dynamics that would be ignored otherwise. This transform generates a complex analytical signal defined by Equation (5). The Random Forest model is then implemented to train predictions on this transformed data. The Random Forest, composed of multiple decision trees, models the complex non-linear relations between historical K_{pa} values and future values. We used Grid Search to optimise the random forest model's hyperparameters. The adjusted hyperparameters include n_estimators (number of trees), max_depth (maximum tree depth), and min_samples_split and min_samples_leaf to ensure robustness [30].

The Random Forest model contains 200 trees (n_estimators=200) with a maximum depth of 12 (max_depth=12), limiting the complexity of the trees to avoid overfitting. The parameters (min_samples_leaf=3) and (min_samples_split=10) ensure that each leaf contains enough samples, reinforcing generalisation. The model uses bootstrap sampling with 80% of the data (max_samples=0.8) to reduce variance. Cross-validation is used to evaluate performance and avoid overfitting, ensuring a good compromise between accuracy and computational cost.

The final prediction of the Random Forest is calculated by the average of results from all the trees. However, as Random Forest may omit some linear trends underlying the K_{pa} index, we apply the Equation (14) given by ARIMA model to adjust these predictions.

Before running the values through the training of the Random Forest model, we will use a Hilbert transformation of the Scipy library in Python (open-source software). The Hilbert decomposition separates the geomagnetic signal into two components: a sinusoidal part and a fluctuating part. Using a Random Forest model, these two components are integrated as distinct variables, allowing the model to better capture the non-linear dynamics and fluctuations of the geomagnetic K_{pa} index. This approach improves the prediction accuracy by considering the complex variations and irregular behavior of the K_{pa} signal. Figure 10 shows results of transformation.

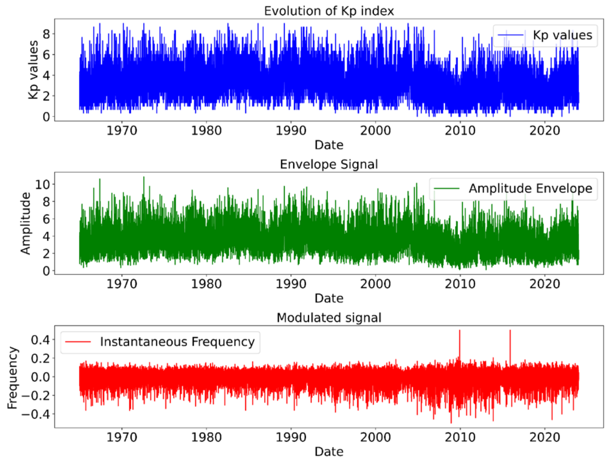


Figure 10. Hilbert transform of the K_{pa} signal

The model is validated using 2023 data as test data. After training on data from 1965 to 2022, performance is evaluated on the new data. The model is validated based on test data from 2023. Two main metrics, MSE and R^2, are calculated to measure the quality of predictions during the training and test stages. This method makes it possible to objectively evaluate the model's ability to generalize and generate accurate predictions on data which the model has not seen previously, so ensuring robustness and reliability for future forecasts. The signal results are trained with the combined Random Forest model imposed by the Hilbert transformation. The training results are evaluated on days in 2023. Figure 11 compares the measured and predicted values of the K_{pa} index during 2023.

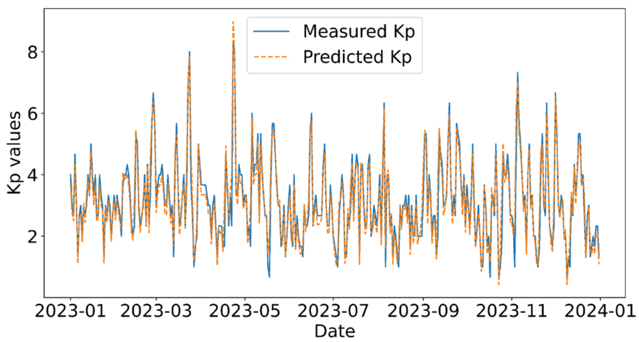


Figure 11. Comparison of measured and predicted K_{pa} during 2023

We used the performance metrics to assess the quality of our model's predictions for the year 2023. The results are $MSE = 0.14$ and $R^2 = 0.96$.

The MSE shows a small margin of error between the predicted and measured values. The determination coefficient R^2 shows that the combined model predicts the Kpa values with a considerable accuracy. The combined model is used to predict Kpa index values from 01/01/2024 to 31/03/2025. The following figure shows the prediction results of the combined model.

We can see that the predicted results from the ARIMA tests produce a different equation to those found previously. To solve this problem, we opted to adjust the predicted values using equation 14, in order to ensure that the predicted values have the same mathematical evolution as the historical data. The following figure compares predicted values generated using the Random Forest model and those predicted and adjusted by the ARIMA Random Forest model with the real K_{pa} index. The metrics on the K_{pa} index values for 9 months in 2024 give values shown in Table 5.

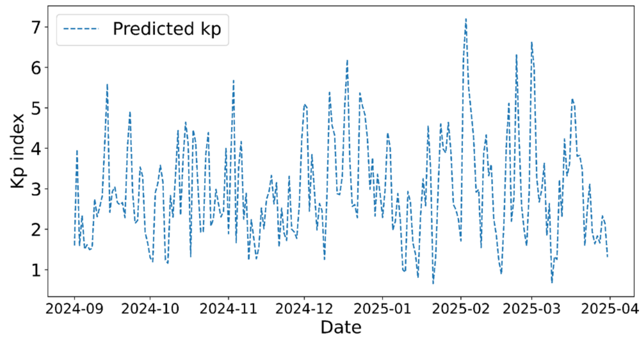


Figure 12. K_{pa} predictions from September 2024 to March 2025 by Random Forest model

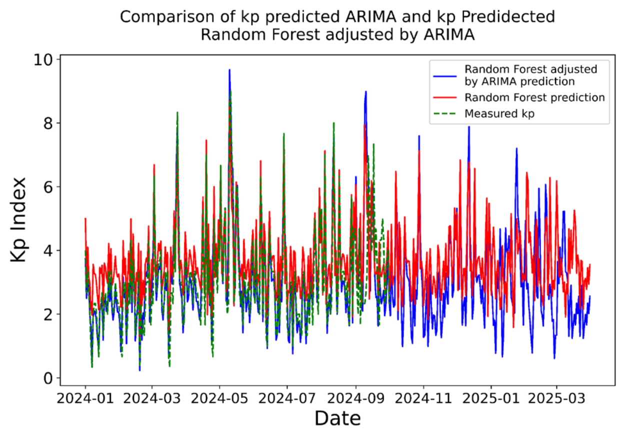


Figure 13. Comparison of K_{pa} index predicted by Random Forest model and K_{pa} predicted by Random Forest adjusted by ARIMA model

Table 4. metrics comparing K_{pa} index measured and predicted values for different periods

	Time Period	ρ (%)	R^2 (%)	RMSE	MBE
Months	9	82.18	62.39	0.92	0.08
	8	94.10	88.21	0.51	0.05
	7	93.71	87.36	0.52	0.06
	6	93.65	87.27	0.53	0.06
	5	93.12	86.13	0.56	0.07
	4	90.21	80.57	0.55	0.07
	3	93.25	86.49	0.45	0.05
	2	92.06	84.73	0.34	-0.01
	1	96.84	90.39	0.24	-0.11
Weeks	3	96.65	91.84	0.24	-0.06
	2	97.07	92.66	0.26	-0.03
	1	98.01	93.63	0.30	0.03
Days	6	97.24	88.97	0.32	0.00
	5	96.96	89.78	0.22	-0.12
	4	94.23	80.77	0.24	-0.14
	3	99.62	77.71	0.26	-0.25
	2	100	88.69	0.22	-0.22

Comparing the values predicted directly by Random Forest and those predicted by Random Forest adjusted by ARIMA, we can see that this last model gives more and more significant predictions by reducing the prediction interval.

4. CONCLUSION AND PERSPECTIVES

In this paper we have highlighted our interest in the study and prediction of the K_{pa} geomagnetic index in various fields. Detailed analysis of the data collected from the GFZ Potsdam databases has allowed us to understand how this index evolves. We found that the K_{pa} index fluctuates significantly over time. We also focused on the most active days during which the K_{pa} index measured is greater than or equal to 7, when extreme geomagnetic events such as geomagnetic storms and the aurora borealis occur, leading directly to the malfunction or degraded performance of vulnerable technologies.

Analyzing the data manually or using conventional tools was still insufficient to model and predict the behavior of geomagnetic activity measured by geomagnetic indices such as K_{pa}. To help solve this problem, we used artificial intelligence in two ways: We did mathematical modelling of the evolution of the K_{pa} index, which gave us a good correlation up to 94% between historical K_{pa} values and those calculated by the equation given by ARIMA. In the second approach, we used a model that combines ARIMA and Random Forest after making the Hilbert decomposition. Using this approach, we were able to make progressively more accurate predictions by reducing the prediction interval. These results give us a prediction of highly geomagnetic active days ($K_p \geq 7$) to take the necessary preventive measures.

Our Hybrid Arima-Random Forest Model using Hilbert Transform enables the K_{pa} index to be accurately predicted several weeks in advance and updated daily. This method is fundamental to forecasting intense geomagnetic events, providing proactive protection for critical infrastructure and vulnerable industries, such as power grids, navigation systems (GPS) and satellite communication systems. With real-time adjustments, this model will make it possible to better manage the risks associated with geomagnetic disturbances, improving the impact to systems in the face of geomagnetic storms and other extreme events.

In the perspectives of this work, we will develop other machine learning and prediction models in order to establish a mathematical model that will make it possible to predict the values of the K_p index over a long period.

ACKNOWLEDGEMENTS

The authors would like to thank GFZ Potsdam (German Research Centre for Geosciences) for making their data available for this study. Their engagement in providing quality resources greatly supports the advancement of scientific knowledge, and this work would not have been possible without their contribution.

NOMENCLATURES

1. Acronyms

ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
IFFT	Inverse Fast Fourier Transform
KNN	Neighbor's Network
LSTM	Long-Short Time Memory
RNN	Recurrent Neural Network
TM	Time series
VGAD	Very geomagnetic active days
K_{pa}	Planetary geomagnetic index

2. Symbols / Parameters

N: Total number of the measurements in the dataset

REFERENCES

[1] S. Chakraborty, S.K. Morley, "Probabilistic Prediction of Geomagnetic Storms and the K_{pa} Index", Journal of Space Weather and Space Climate, Vol. 10, No. 36, pp. 1-16, 2020.

[2] L.M.M. Myint, K. Hozumi, S. Saito, P. Supnithi, "Analysis of Local Geomagnetic Index Under the Influence of Equatorial Electrojet (EEJ) at the Equatorial Phuket Geomagnetic Station in Thailand", Advances in Space Research, Vol. 70, No. 5, pp. 1429-1440, June 2022.

[3] J. Matzke, C. Stolle, Y. Yamazaki, O. Bronkalla, A. Morschhauser, "The Geomagnetic Kpa Index and Derived Indices of Geomagnetic Activity", Space Weather, Vol. 19, No. 5, pp. 1-21, 2021.

[4] A. Collado Villaverde, P. Munoz, C. Cid, "Classifying and Bounding Geomagnetic Storms Based on the SYM-H and ASY-H Indices", Natural Hazards, Issue 20, Vol. 120, pp. 1141-1162, January 2024.

[5] M.J. Heyns, S.I. Lotz, C.T. Gaunt, "Geomagnetic Pulsations Driving Geomagnetically Induced Currents", Space Weather, Issue 2, Vol. 19, pp. 1-19, November 2021.

[6] L. Trichtchenko, A.P. Trishchenko, P. Hejda, R. Langer, "Evaluation of Telluric-Associated Corrosion on Buried Pipelines", Journal of Atmospheric and Solar-Terrestrial Physics, Issue 22, Vol. 248, pp. 1-14, May 2023.

[7] H. Nebdi, "Study of Geomagnetic Activity During Three Solar Cycles Using a K Index Data Base", Journal of Umm Al-Qura University for Applied Sciences, Issue 1, Vol. 2, pp. 33-41, 2010.

[8] C. Marque, et al., "Solar Radio Emission as a Disturbance of Aeronautical Radionavigation", Journal of Space Weather and Space Climate, Vol. 8, No. 42, pp. 1-13, 2018.

[9] G. Li, B. Ning, B. Zhao, L. Liu, J.Y. Liu, K. Yumoto, "Effects of Geomagnetic Storm on GPS Ionospheric Scintillations at Sanya", Journal of Atmospheric and Solar Terrestrial Physics, Issue 7, Vol. 70, pp. 1034-1045, May 2008.

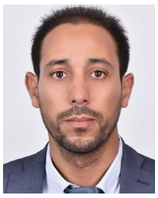
[10] H. Nebdi, "Space Weather and Link to Climate Change", Handbook of Research on Global Environmental Changes and Human Health, IGI Global, pp. 1-20, 2019.

[11] A. Mishev, S. Tuohino, I. Usoskin, "Neutron Monitor Count Rate Increase as a Proxy for Dose Rate Assessment at Aviation Altitudes During GLEs", Journal of Space Weather and Space Climate, Vol. 8, No. 46, pp. 1-11, October 2018.

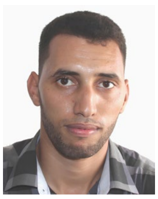
[12] U. Aksen, U.D. Goker, E. Timocin, C. Akcay, M. Ipek, "The Effect of Geomagnetic Storms on Aircraft

Accidents Between the Years 1919-2023 in Civil Aviation", Advances in Space Research, Issue 1, Vol. 73, pp. 807-830, January 2024.
[13] M.J. Owens, M. Lockwood, L.A. Barnard, C.J. Scott, C. Haines, A. Macneil, "Extreme Space-Weather Events and the Solar Cycle", Solar Physics, Vol. 296, No. 82, pp. 1-19, May 2021.
[14] B. Chafik, M. El Malki, F. Miskane, H. Nebdi, "Study of Geomagnetic Activity According to Kpa Index and Its Variability in Relation to Sunspots Over the Last Five Solar Cycles", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 59, Vol. 16, No. 2, pp. 331-341, June 2024.
[15] H. Zentrum, P, Deutsches, "Geo Forschungs Zentrum (GFZ), Geomagnetic Indices (Kpa, ap)", "Hybrid of Nowcast and Definite Kpa and ap Values, Simple ASCII Format", The Directory Contains Real Time Files, Files with all Values Since 1932, Year Files, and Files with ap and Solar Activity Indices SN and F10.7, 03 October 2024, https://kp.gfz-potsdam.de/en/data.
[16] E. Chodakowska, J. Nazarko, L. Nazarko, H.S. Rabayah, R.M. Abendeh, R. Alawneh, "ARIMA Models in Solar Radiation Forecasting in Different Geographic Locations", Energies, Issue 13, Vol. 16, No. 5029, June 2023.
[17] V. Arumugam, V. Natarajan, "Time Series Modeling and Forecasting Using Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average Models", Instrumentation Mesure Metrologie, Vol. 22, No. 4, pp. 161-168, August 2023.
[18] R. Lawton, "Time Series Analysis and its Applications", International Journal of Forecasting, Issue 2, Vol. 17, No. 2, pp. 299-301, 2001.
[19] D.F. Hendry, M.S. Morgan, "The Foundations of Econometric Analysis", Cambridge University Press, Cambridge, pp. 1-576, 1995.
[20] H. Akaike, "A New Look at the Statistical Model Identification", IEEE Transactions on Automatic Control, Issue 6, Vol. 19, pp. 716-723, December 1974.
[21] G. Zhang, Y. Li, T. Wang, H. Du, F. Luo, Y. Zhan, "Extended Hilbert Transform and Application for Seismic Attributes", Earth and Space Science, Issue 5, Vol. 6, pp. 873-886, May 2019.
[22] H.G. Brachtendorf, C. Dalpiaz, M. Steiger, "Multiplier-Less Broadband and Linear Phase Digital Hilbert Transformers", Circuits, Systems, and Signal Processing, Vol. 43, No. 10, pp. 6475-6486, June2024.
[23] I. Kanatov, D. Kaplun, D. Butusov, V. Gulvanskii, A. Sinitca, "One Technique to Enhance the Resolution of Discrete Fourier Transform", Electronics, Vol. 8, No. 330, pp. 1-16, March 2019.
[24] G. Hill, "IP and Cellular Networks, and Mathematical Techniques", The Cable and Telecommunications Professionals Reference, 3rd ed., pp. 1-560, Routledge, New York, USA, 2007.
[25] M. Vyas, N. Hemrajani, "Predicting Effort of Agile Software Projects Using Linear Regression, Ridge Regression and Logistic Regression", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 47, Vol. 13, No. 2, pp. 14-19, June 2021.
[26] R.J. Hyndman, G. Athanasopoulos, "Forecasting: Principles and Practice", OTexts, 3rd ed., pp. 1-442, New York, USA, May 2021.
[27] M. Wisniewski, et al., "Thirteen Ways to Look at the Correlation Coefficient", The American Statistician, Vol. 42, No. 1, pp. 59-66, February 1988.
[28] J.O. Rawlings, S.G. Pantula, D.A. Dickey, "Applied Regression Analysis: A Research Tool", Springer, 2nd ed., pp. 1-660, New York, USA, 1998.
[29] H. Taneja, M.C. Taplamacioglu, V. Jain, A. Jain, A. K. Dubey, M. Demirci, "Weather Prediction Using Regression Algorithm and Neural Network Technique", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 59, Vol. 16, No. 2, pp. 303-312, June 2024.
[30] P. Probst, M.N. Wright, A.L. Boulesteix, "Hyperparameters and Tuning Strategies for Random Forest", WIREs Data Mining and Knowledge Discovery, Issue 3, Vol. 9, pp. 1-19, June 2019.

BIOGRAPHIES



Name: **Bouchaib**
Surname: **Chafik**
Birthday: 20.12.1984
Birthplace: Sidi Bennour, Morocco
Bachelor: Renewable Energies, Faculty of Sciences, University of Chouaib Doukkali, El Jaida, Morocco, 2016
Master: Materials and Radiation, Energy and Environment, Faculty of Sciences, University of Chouaib Doukkali, El Jaida, Morocco, 2018
Doctorate: Student, Department of Physics, Faculty of Sciences, University of Chouaib Doukkali, El Jaida, Morocco, Since 2018
Research Interests: Materials Physics, Theoretical Physics, Experimental Physics, Geophysics
Scientific Publications: 2 Papers



Name: **Mohammed**
Surname: **El Malki**
Birthday: 05.10.1991
Birthplace: Settat, Morocco
Bachelor: Renewable Energies, Faculty of Sciences, University of Chouaib Doukkali, El Jaida, Morocco, 2016
Master: Materials and Radiation, Energy and Environment, Faculty of Sciences, University of Chouaib Doukkali, El Jaida, Morocco, 2018
Doctorate: Student, Department of Physics, Faculty of Sciences, University of Chouaib Doukkali, El Jaida, Morocco, Since 2018
Research Interests: Materials Physics, Theoretical Physics, Experimental Physics
Scientific Publications: 3 Papers

Name: **Fatiha**
Surname: **Miskane**
Birthday: 22.06.1960
Birthplace: El Jaida, Morocco
Bachelor: Physics of Materials, Department of Physics, Faculty of Sciences, Mohammed V University, Rabat, Morocco, 1985
Masters: Condensed Matter Physics, Department of Physics, Faculty of Sciences, Mohammed V University, Rabat, Morocco, 1986
Doctorate: Physics of Plasma, Department of Physics, Faculty of Science Ben M'Sik, Hassan II University, Casablanca, Morocco, 2002
The Last Scientific Position: Prof., Department of Physics, Faculty of Sciences, University of Chouaib Doukkali, El Jaida, Morocco, Since 2000
Research Interests: Physics of Plasmas, Controlled Thermonuclear Fusion, Plasma of Tokamak
Scientific Publications: 9 Papers
Scientific Memberships: International Society of Muslim Women in Science (ISMWS)

Name: **Hamid**
Surname: **Nebdi**
Birthday: 01.01.1968
Birthplace: El Jadida., Morocco.
Bachelor: Physics in Electronics, Department of Physics, Faculty of Sciences, University of Chouaib Doukkali, El Jadida, Morocco, 1993
Masters: 1. Radiation-Matter Physics, Faculty of Science Ben M'Sik, Hassan II University, Casablanca, Morocco, 1995 - 2. Atomic Physics, University Catholic of Louvain, Ottignies-Louvain-la-Neuve, Belgium, 1997
Doctorate: Atomic Physics, University Catholic of Louvain, Ottignies-Louvain-la-Neuve, Belgium, 2000
The Last Scientific Position: Prof., Department of Physics, Faculty of Sciences, University of Chouaib Doukkali, El Jaida, Morocco, Since 2011
Research Interests: Space Weather, Atomic Physics, Laser Physics, Mathematical Physics and Education
Scientific Publications: 82 Papers
Scientific Memberships: Spinet. Aerdds