Nowcasting for improving ^{222}Rn forecasting at LSC

T. Sánchez-Pastor[1,*]

Centro de Investigaciones Energéticas Medioambientales y Tecnológicas, 28040 Madrid, Spain.

Miguel Cárdenas-Montes[1]

Centro de Investigaciones Energéticas Medioambientales y Tecnológicas, 28040 Madrid, Spain.

Roberto Santorelli[1]

Centro de Investigaciones Energéticas Medioambientales y Tecnológicas, 28040 Madrid, Spain.

Abstract

The improvement of experiments at underground laboratories is closely linked to background reduction. The main source of background is the ^{222}Rn levels at this type of experiments i.e Argon Dark Matter 1-t located at the underground laboratory of Canfranc (LSC), Spain, aimed at the dark matter direct searches. One approach that can be done in order to get rid of this background consists of modeling and forecasting the signal for efficient planning activities at the experiment. In this work, we have analyzed six years of weekly ^{222}Rn (June 2013 - September 2019) levels using deep learning techniques such as Convolutional Neural Networks (CNN), Artificial Neural Net-

*Corresponding author

Email addresses: tomas.sanchez@ciemat.es (T. Sánchez-Pastor), miguel.cardenas@ciemat.es (Miguel Cárdenas-Montes), roberto.santorelli@ciemat.es (Roberto Santorelli)

works (ANN) and Recurrent Neural Networks (RNN). Taking into account several meteorological variables from cities around LSC such as Barcelona (BCN), Huesca (HSC), Pamplona (PMP) and Zaragoza (ZGZ). We study the variable importance in the forecasting of those variables through the Random Forest algorithm and we find that temperature is the one that would make significant improvements to the forecasting. Therefore, using temperature from each weather station and ^{222}Rn levels at LSC as input, we feed the neural networks and find that the inclusion of environmental variables generally improves the forecasting. In this manner, we present a new methodology that improves the prediction of ^{222}Rn signal at every underground laboratory, being this an essential task in order to schedule maintenance operations of the detectors.
Keywords: Neural Network · ^{222}Rn Measurements · Canfranc Underground Laboratory · Forecasting.

1. Introduction

The study of the ^{222}Rn Time Series has been increasingly gaining importance in the past decades. The main reason is that ^{222}Rn is a source of background signal at underground laboratories, where very high accuracy in measurements is crucial. Time Series analysis techniques (classic and deep learning ones) requires equally-spaced (separated) points in time and have shown to be relevant in several fields, such as in seismicity [1, 2], volcanism [3], atmospheric studies [4] and gamma rays [5], among others.

^{222}Rn Time Series are produced by the ^{238}U decay chain that emits α particles with very low decay energies (5.59 MeV) and relative large half-

life (3.8 days) [6]. The way radon arrives to the underground experiment installations is explained as follows: ^{222}Rn is intrinsically present in gas form in the porous media of the Earth's crust in variable amount. Despite the large half-life, the gas migrates from the rocks to the air before its decay, diffusing on the experimental hall [7]. This is the case of the Canfranc Underground Laboratory (LSC), where the Argon Dark Matter 1T (ArDM) experiment is taking place and a deep study of the ^{222}Rn signal is needed.

The ArDM experiment is located in a former railway Tunnel 850 m under the Pyreness in Canfranc, aimed to the direct detection of very weak-type interactions between WIMPS (Weakly Interacting Massive Particles) and atomic nuclei through a double-phase (liquid-gas) Argon TPC [8]. WIMPS are hypothetical particles that fulfills the galactic medium, orbiting around the galaxy centers at non-relativistic velocities. Despite the fact that WIMP-Nucleus cross sections are very low, the great flux of particles passing through the Earth makes plausible the detection of this elusive particles. Theoretically, it can be shown that the WIMP signal presents an annual modulation due to the relative velocities of the Earth-Sun system respect the center of the galaxy. By other hand, there is also a seasonal periodicity in the case of ^{222}Rn Time Series, being the key for a precise understanding of the background in this type of experiments [9].

The existence of seasonalities in noisy data points out that deep learning algorithms could be applied for forecasting the signal. Meanwhile, other works, like [10], demonstrated that in similar systems, the main environmental factor associated with radon variability is the temperature gradient between the experimental hall and the air outside the tunnel. This results

supports the study presented in this manuscript because ^{222}Rn signal is correlated with humidity at LSC, whose external temperature is responsible. For further references on ambiance correlations at LSC and classic/deep learning techniques on this dataset see [11, 12].

Combining this last results, a new methodology for forecast the radon arise: the deep learning nowcasting of ^{222}Rn Time Series using close LSC locations (Fig.1) ambiance data, which is the main goal of this work.

Recently, other approaches has been developed to improve this time series forecasting at LSC; in [13] the authors applied a STL decomposition to the dataset before feed the neural network, obtaining significant improvements in the prediction, in [14] the authors used a evolutionary algorithm (Population-Based Incremental Learning) to automatically find the optimal set of hyperparameters in neural networks, and, finally, in [15] the authors take advantage of stochastic regularization techniques, such as Dropout regularization, to properly define the systematic errors as well as the confidence intervals.

In what concerns to this paper, we explain in Section 2 the methodology that we have built for improve the predictions. Firstly, we dig into how the preprocessing has been done: data acquisition, data cleaning, gaps filling and weekly summed data for 6 years of data between July 2013 and September 2019. Once the data is prepared, we study the correlations between ^{222}Rn at LSC and temperature, preassure and wind velocity at 4 LSC surrounding cities: Barcelona, Pamplona, Huesca and Zaragoza. Then, Random Forest algorithm is applied for variable importance analysis.

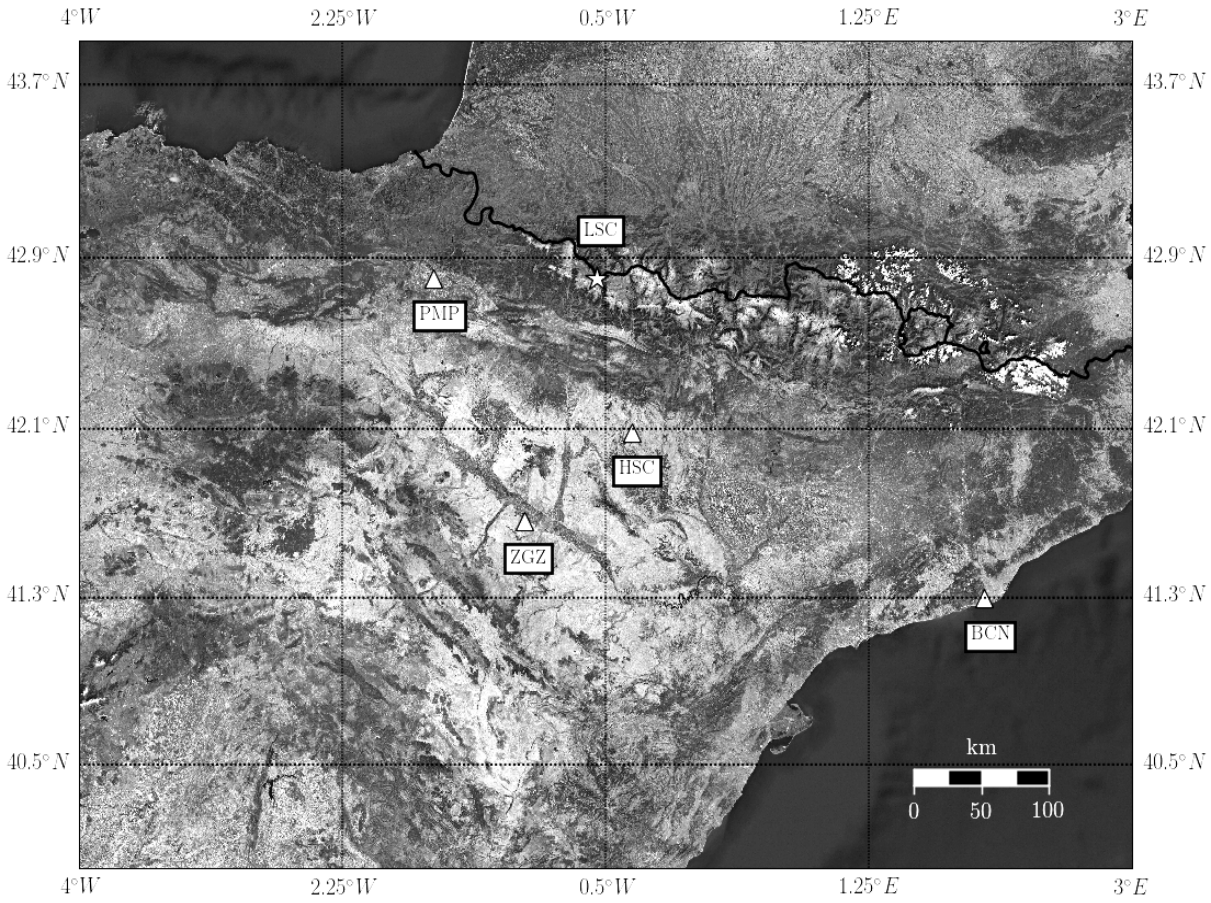

Fig. 1: GPS map of the Spanish Peninsula area of interest. Triangles mark the location of the meteorological stations and the star marks the LSC station. Pamplona station is labeled as PMP, Barcelona station as BCN, Huesca station as HSC and Zaragoza station as ZGZ.

At the end of the day, a wide spectrum of Neural Networks architectures are fed with the weekly average temperature and weekly Radon medians for predicting the Radon Time Series in the next weeks. The spectrum includes: Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN)

and Recurrent neural networks (RNN). Finally, we focus on the statistical analysis of the Mean Absolute Errors (MAE) for 25 independent runs for each architecture and each city to state whether or not the addition of ambiance variables to the dataset improves the performance of the trained model. This results are discused in Section 3 and the conclusions collected in Section 4.

2. Methodology

2.1. Data acquisition

Radon density levels at the experimental Hall in LSC goes from tens to hundreds of Bq/m^3 and has been measured every 10 minutes since July 2013 with an Alphaguard P30. The AlphaGuard is the center piece of a compact portable measuring system for the continuous determination of radon and radon progeny concentration in air, as well as selected climatic parameters. In standard operation mode, the measuring gas gets by diffusion through a large-scale glass fiber filter into the ionization chamber, i.e. through the glass filter only ^{222}Rn may pass, while the radon daughter nuclei are prevented to enter the ionization chamber. The ingoing ^{222} Rn interacts with the inert gas creating ion-electron pairs. Then, the current strong electric field drives the ions along the chamber with a constant acceleration. At some point, the electrons reach enough energy to ionize the gas, producing a Townsend discharge. Finally the avalanche of electrons arrive the cathode and the detector counts the resulting electric intensity as one hit of ^{222}Rn.

Regarding the ambiance data, it has been downloaded from AEMET (Agencia Estatal de Meteorología) Open Data spanish service. The four selected stations are located at the airports of each city to avoid most of

the missing data. Between the available data that can be downloaded, the significant variables to carry out this study are: the average temperature per day ($\bar{T}$ (^oC)), the average preassure per day ($\bar{P}$ (hPa)) and the average wind velocity per day ($\bar{V}$ ($m \cdot s^{-1}$)).

2.2. Data preprocessing

Missing values have been filled using an autoregressive integrated moving average model (ARIMA). ARIMA(p, q, d) models are stochastic processes that can be explained by its own history [16], commonly used to forecast time series. The model can be expressed as follows:

$$y(t) = \varphi_0 + \varepsilon_0 + (y_t - y_{t+d}) + \sum_p \varphi_p y_{t-p} - \sum_p \theta_q \varepsilon_{t-q} \tag{1}$$

Being y_t the measurements and ε_t the intrinsic white noise. On one hand, the terms containing φ_p are associated with an autoregressive model (AR), which takes into account the past p values to predict the next one. The terms with θ_q correspond with the moving average model (MA), that takes the past q white noise error terms and the average value of the series to predict the future value. Finally, the integration (I) term is $(y_t - y_{t+d})$, that tries to make the signal stationary and trend-less. The model we have used to fill the null values is an ARIMA(3, 0, 1).

At short time-scales, ^{222}Rn levels at LSC shows a random behaviour, in contrast to large time-scales, where the seasonal component dominates. For the Radon signal, the weekly medians has been selected as the representative variable, meanwhile, weekly averages for the meteorological variables. Hence, the accessible volume of data is 325 values. In Fig. 2 are represented the weekly medians of the signal and its Fast Fourier Transform (FFT):

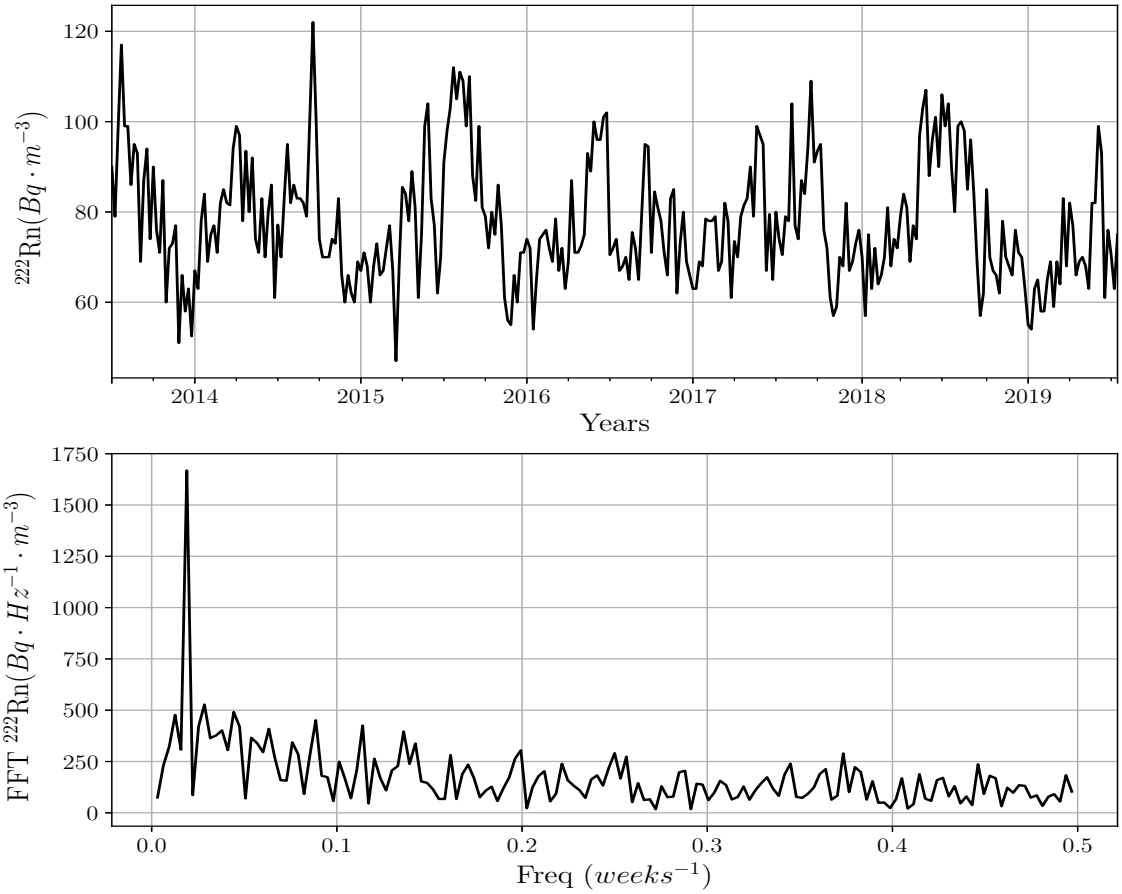

Fig. 2: ^{222}Rn concentrations medians per week from July 2013 to July 2019 at Hall A of the LSC (upper) and its frequency domain obtained after applying a Fast Fourier Transform (FFT) (down).

The ^{222}Rn levels ranges from 50 to 120 $Bq \cdot m^{-3}$ approximately, reaching its minimum on winter and its maximum on summer. As said, the signal presents a clear seasonal component at large times and a random component at short times. The FFT, which contains the information of the seasonalities presented on the signal, exhibits a main peak corresponding with a period of

52.6 weeks, which is roughly 1 year.

2.3. Correlation with meteorological variables

Previous to use forecasting neural networks algorithms, we study the existing correlations between the radon signal and the environmental variables. The latter, is essential for add or remove relevant information to the training dataset. Neural networks are able to learn about the intrinsic correlations between the inputs, nevertheless, the performance of the model may improve if spurious correlations are avoided.

The following figures shows the correlation of ^{222}Rn with temperature, pressure and wind velocity for Barcelona (BCN), Pamplona (PMP), Zaragoza (ZGZ) and Huesca (HSC). In the case of temperature in Fig. 3, we see a clear noiseless annual modulation that couples with radon. Also, temperature behaves similar for every city, i.e. the trend and the period are invariant, although the maximum and the minimum values changes.

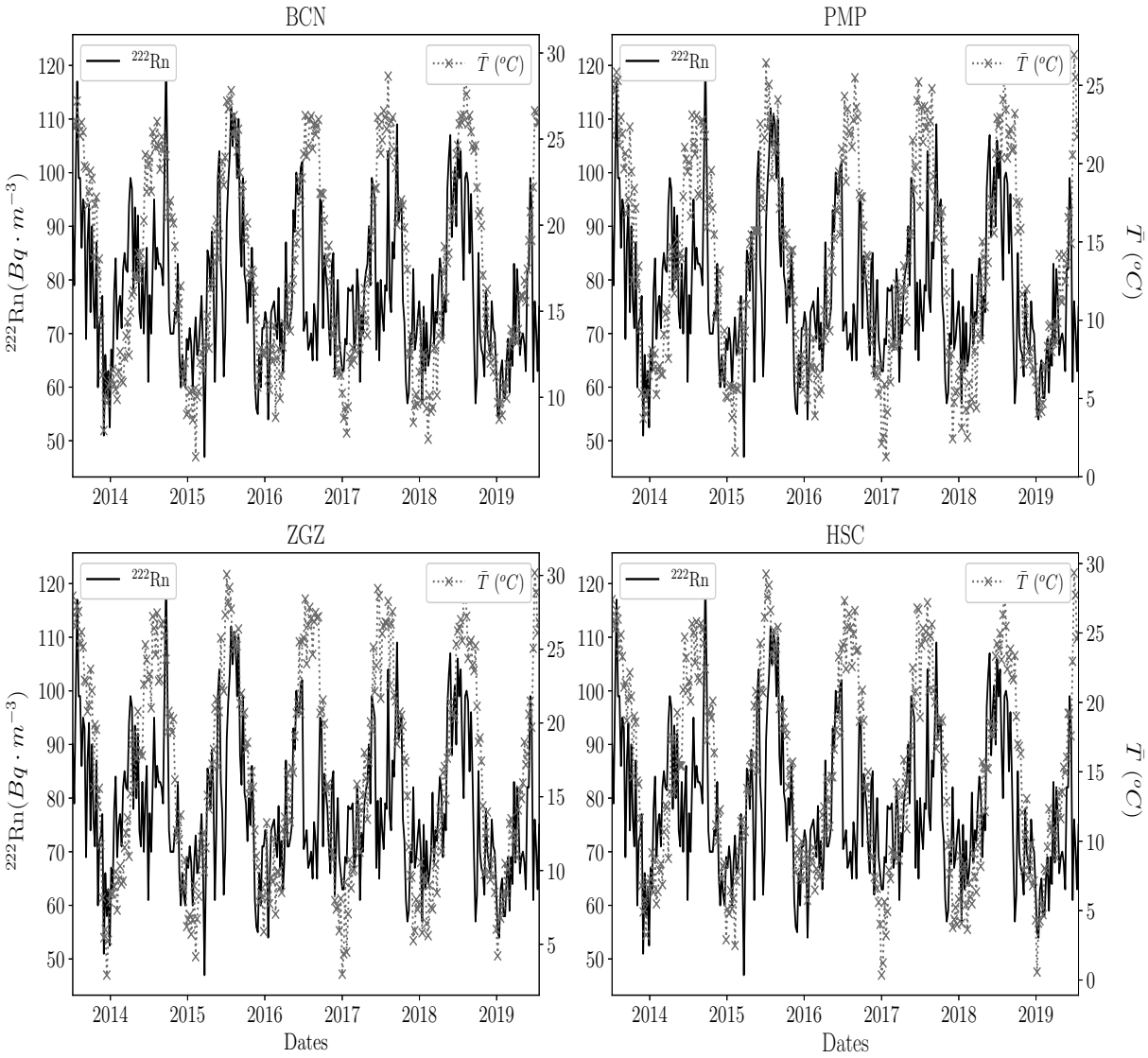

Fig. 3: Weekly ^{222}Rn concentrations medians (solid black) and weekly average temperature in ^oC (dashed gray) for each city from July 2013 to July 2019.

For the pressure (Fig. 4), we don't see a clear correlation with radon.

This time series is more noisy than the previous one. In addition, trend and period are practically invariant over the city locations.

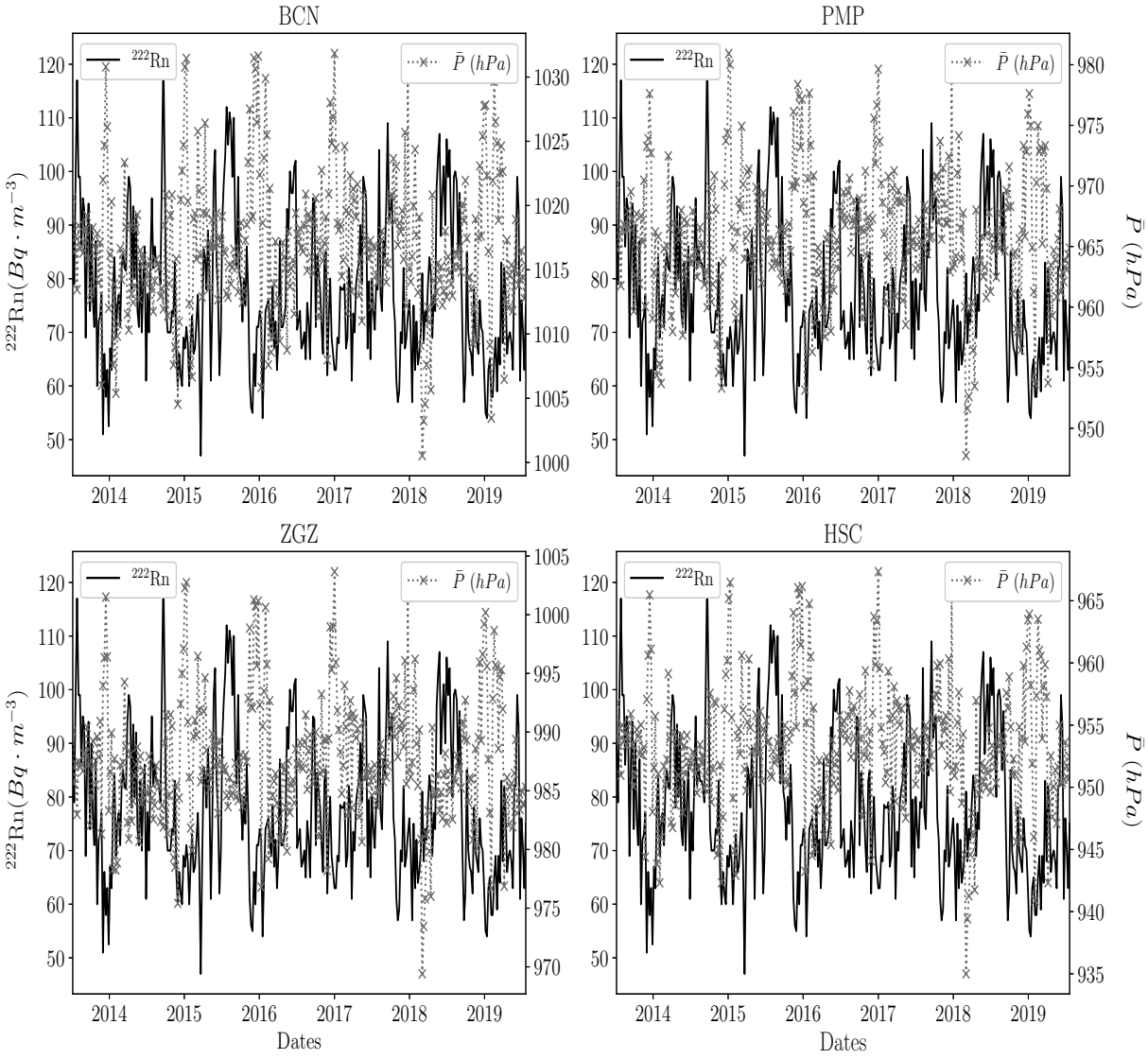

Fig. 4: Weekly ^{222}Rn concentrations medians (solid black) and weekly average preassure in hPa (dashed gray) for each city from July 2013 to July 2019.

The wind speed (Fig. 5) is the most stochastic time series and depends strongly on the city location. Moreover, it has not a clear seasonal component that couples to the radon modulation.

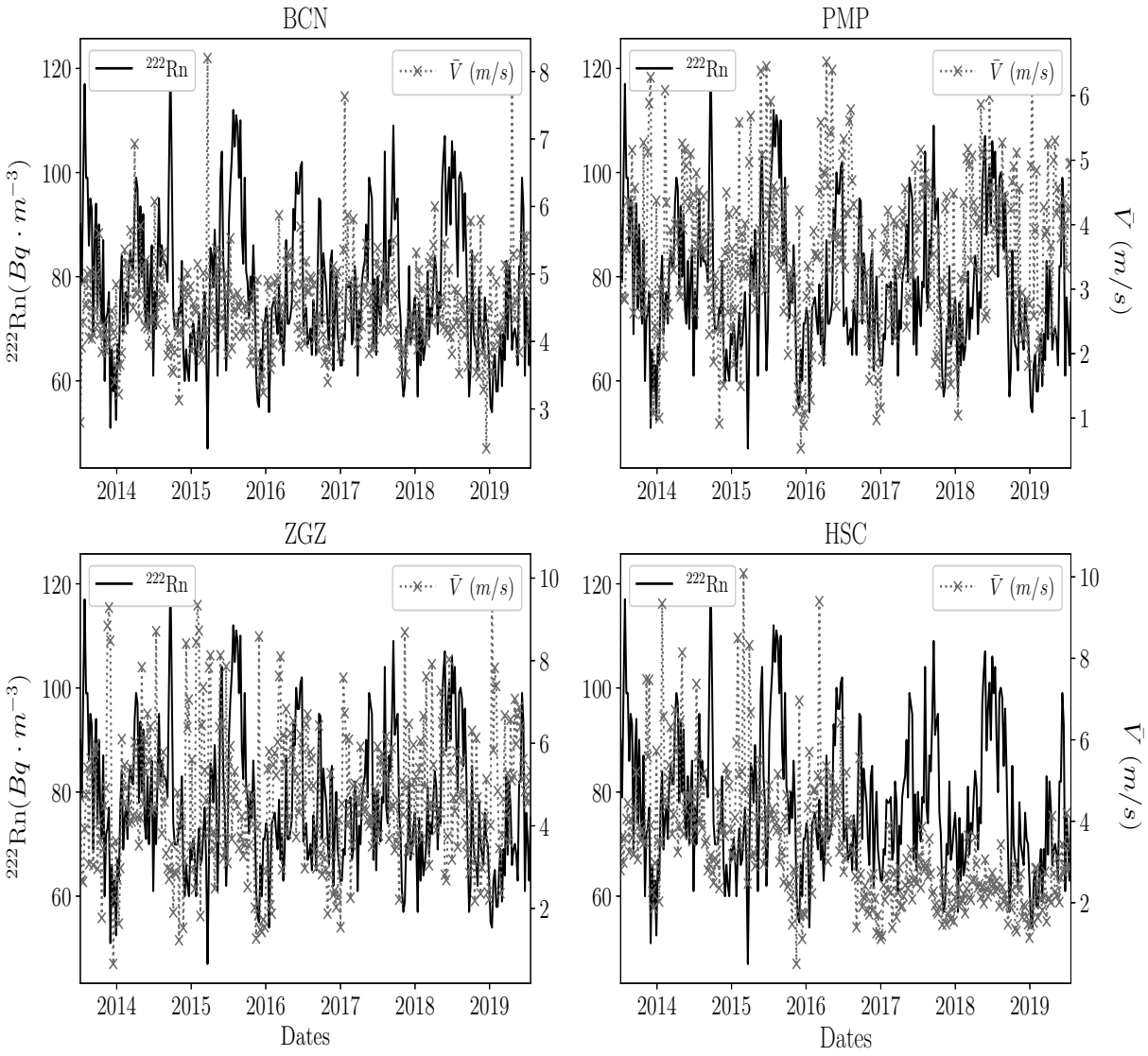

Fig. 5: Weekly ^{222}Rn concentrations medians (solid black) and weekly average wind velocity in $m \cdot s^{-1}$ (dashed gray) for each city from July 2013 to July 2019.

The Pearson coefficient is used to measure the correlation between vari-

ables and the results are collected in Table 1. As expected, temperature is the most correlated variable reaching a pearson coefficient of 0.53 in the cases of Barcelona and Zaragoza. Meanwhile, preassure is anticorrelated and wind velocity presents weak correlations for every city unless for Pamplona:

City	corr(Rn + T)	corr(Rn + P)	corr(Rn + V)
BCN	0.53	-0.21	0.06
PMP	0.52	-0.19	0.17
ZGZ	0.53	-0.27	0.10
HSC	0.52	-0.16	0.12

Table 1: Pearson coefficients between the weekly median ^{222}Rn levels at LSC and several meteorological variables: average preassure, average temperature and average wind velocity recorded at different stations (labeled with the city name).

2.4. Random Forest

There are other ways to measure the significance of the inputs at the moment of forecasting a time series. One of the most popular methods is the variable importance analysis that Random Forest algorithm offers [17].

Random Forest algorithm is applicable for classification or regression problems and operate bagging an ensemble of independent decision trees whose input are random sections of the original dataset. For regression, the algorithm ends up averaging the values obtained on each tree, instead of classification that counts the majority class. In this manner, the variable importance is carried out looking at how much a prediction error increases when a given input vary while all others remain unchanged. The relative importances are normalized, so $\sum_i I_i = 1$

Fig. 6 shows that temperature is the most important external variable to predict ^{222}Rn, whereas pressure and wind velocity importances are at least one order of magnitude smaller.

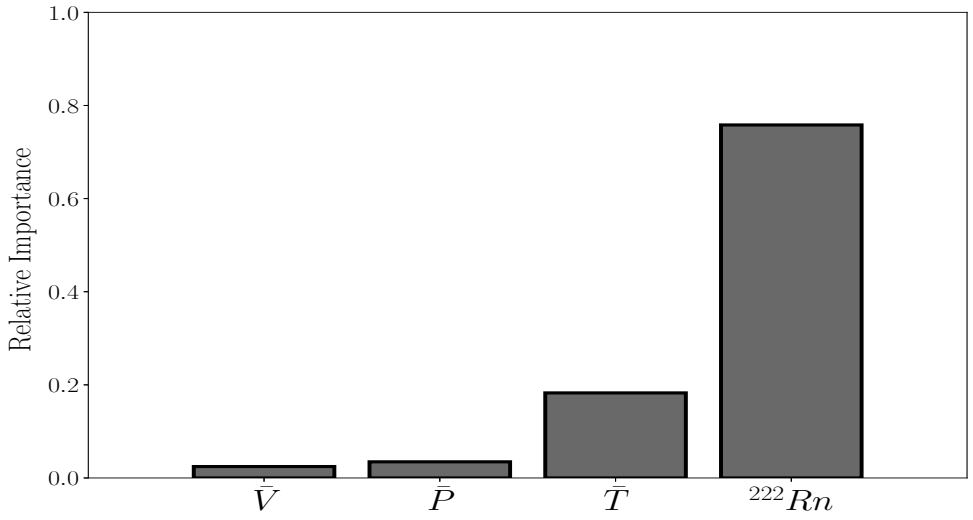

Fig. 6: Relative variable importances to forecasting the signal obtained with Random Forest algorithm.

This last result is in agreement with the correlations computed before, henceforth, in pursuit of this tests, we use only temperature as external input variable to predict the radon time series.

2.5. Data structure

Regarding the data structure that will feed the corresponding neural network architecture, we have create a moving window of one-year extent, say 52 weekly values of radon and temperature. The length of the input window

is also known as the "loop-back", as well as "loop-forward", the length of the output window. For our study, the loop-forward is the number of future weeks that the NN predicts and ranges from 1 to 8.

The schema shown in Fig. 7 represents the neural network processing. As can be seen, there are some outputs that are going to be predicted by the neural network more than once.

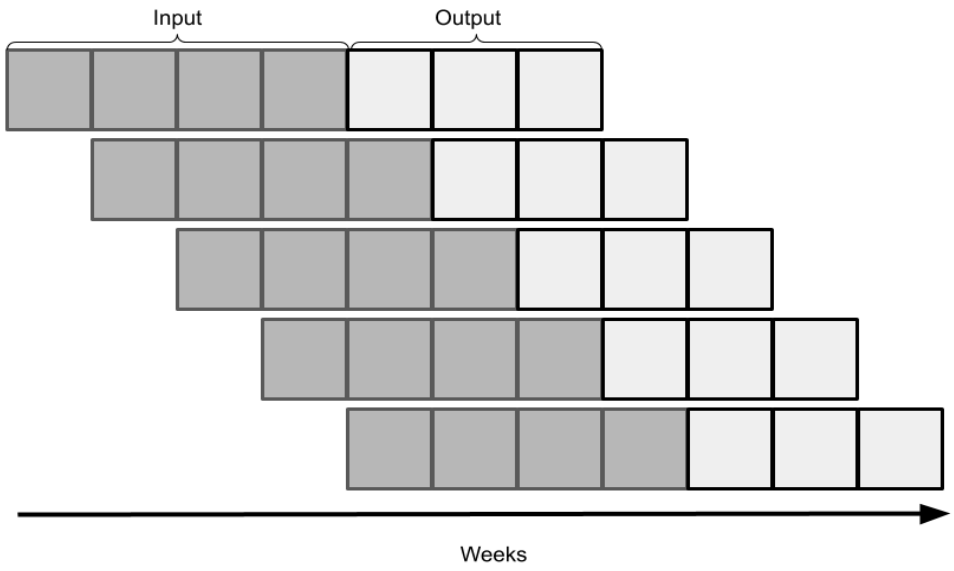

Fig. 7: Structure of input (gray) and output (white) data. In this example, the loop-back is 4 and the loop-forward 3.

We can state that the maximum number of predictions over the same value of the time series is equal to the chosen loop-forward. In this manner, extracting the max, min and the average of the repeated predictions the forecasting can be represented by a band of width $|max(pred_i) - min(pred_i)|$. We take advantage of this approach in the forecasting of the ^{222}Rn.

2.6. Neural Networks

A Neural Network is a combination of individual neurons that maps several inputs, $x_1, x_2, ..., x_n \in \mathbb{R}$ into a real output o_k through an activation function $a_j(\theta) : \mathbb{R} \to \mathbb{R}$ that tries to classify or forecast a target y. Every single neuron performs the next computation over the data: $a_j = f(\sum_i w_{ij}x_i + b_j)$, where w_{ij} are the so-called weights of the activation function and b_j the bias. This functions add complexity to the complete model due to they all are non-linear functions. There are a wide spectrum of activation functions, the most common are: *sigmoid*, *tanh* and $reLu$ functions. Moreover the neurons are collected in layers (input, hidden and output) as can be seen in Fig. 8.

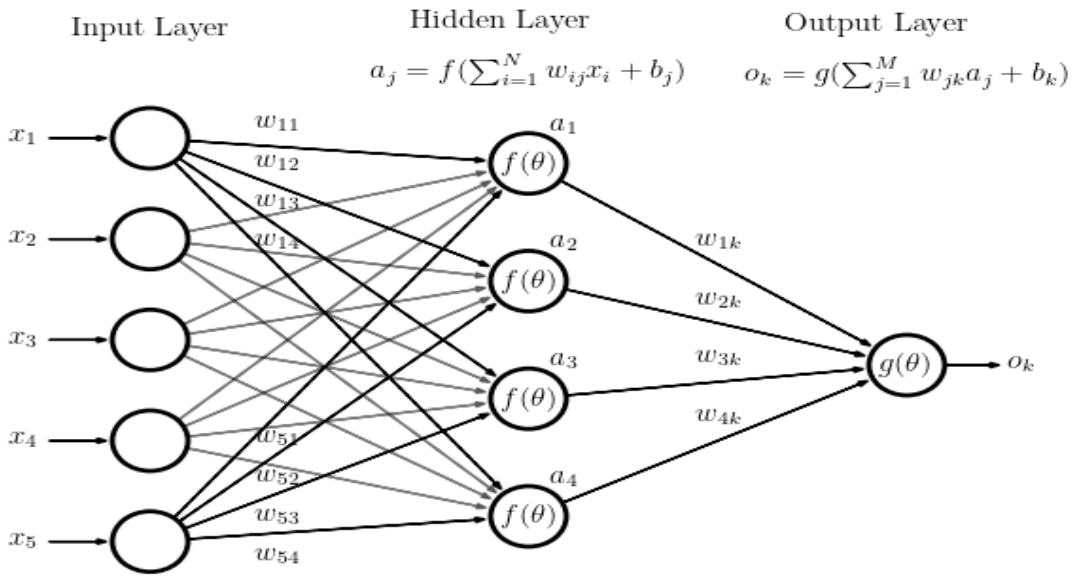

Fig. 8: Example of neural network.

The neural network connects each i-th layer forwarding the results to the

$(i+1)$-th layer until the output layer. On each neuron a Cost Function (i.e. $C(x, w, b) = ||y(x) - o(x, w, b)||^2$), which measures how close the output of the network is to the desired output, is computed.

Neural networks uses the Fast Forward method that consist on: first, the weight and the bias are randomly initialized (making the neural network an stochastic algorithm). The activation function is applied in every neuron raising a value that would feed the next layer and so on. Once the computations have crossed from the input layer to the output layer, the resultant cost function is minimized correcting the initial parameters of the model in a recursively way (Back Propagation algorithm). For more details see [18].

In this sense, we can see a neural network as an adaptable non-linear function that maps an input into an output minimizing a Cost Function in order to find the suitable combination of hyperparameters that fits the desired data, i.e. $F(x, w, b) : \mathbb{R}^n \to \mathbb{R}^m$.

Therefore, for this study we have used three different types of neural networks:

- **Artificial Neural Network (ANN).** This type of networks connect every neuron from the previous layer to the present layer applying the equations written on Fig. 7 to produce the output, this is de so-called "Dense layers". In this work, we have used one ANN to forecast the ^{222}Rn signal using itself as input and other using ^{222}Rn and temperature as input.

  The suboptimal configuration that we have found for the first NN has been: a first hidden dense layer with 256 neurons and $reLu$ activation function, a Dropout layer to prevent overfitting with a rate of 0.2,

a second hidden dense layer with 128 neurons and $reLu$ activation function, followed by a Flatten layer and $nFw \in [1, 8]$ neurons (future weeks predicted) on the output layer. The model has been trained within 40 epochs. Instead of the NN of ^{222}Rn+T, which has been trained within 60 epochs.

- **Convolutional Neural Network (CNN).** Nowadays, convolutional nets are commonly used, for example, in biomedical sciences for image processing. The only difference with ANN is that convolution operation replaces the multiplication between the weights matrix (w) and the data (x) ($a_j = f\left(\sum_i w_{ij} * x_i + b_j\right)$). Convolution provides the neural network translation invariance and parameter sharing, properties that makes CNN a good option to forecast 1D time series. Weights are now understood as filters that the model has to learn to extract the relevant information on the dataset.

  In this work, in the case of only ^{222}Rn as input, we have used a first convolutional layer with 64 filters and $reLu$ activation followed by a MaxPooling layer of size 2, a second convolutional layer with 128 filters and $reLu$ activation, a Flatten layer, a dense layer with 32 neurons and $reLu$, another dense layer with 64 neurons and the same output layer than in the ANN case, trained within 40 epochs. For ^{222}Rn+T as input, we have used the same architecture with two times more of filters, neurons and epochs. The padding is valid, the kernel size 3 and the stride 1 in every convolutional layer.

- **Recurrent Neural Network (RNN).** Recurrent Neural Networks

(RNN) are architectures used for process time sequenced data $x^{(1)}, ..., x^{(t)}$ that allows the model to learn about the past history of the time series. From this perspective, each single neuron is unfolded into t neurons that process the information sorted forward in time. Unfolding neurons can be viewed as a self-loop where the neurons are fully connected with each others. Thus RNN includes the dependence on the past history of the input to the future outputs.

RNN algorithms can't get rid of the vanishing gradient problem for large time series, in order to address it, Long Short-Term Memory (LSTM) architectures solves this problem including a self-loop conditioned on the context rather than fixed. By making this self-loop controlled by another hidden unit, the time scale can be changed dynamically [19].

In this architecture, for only ^{222}Rn as input, a first LSTM layer with 52 neurons with $\tanh(x)$ activation function and a second Dense fully connected layer with 64 neurons and reLu activation has been used. In the case of predict using ^{222}Rn+T as input the only hyperparameter that changes is the number of Dense neurons, which is 104. We have trained this architectures through 7 epochs.

2.7. Statistics

Non-parametric test are commonly used when the statistical nature of the data is unknown. In this work, statistical test like Kruskal-Wallis is applied to infer whether the environment variables improves the forecasting of the Radon signal and if additional information is included on the dataset.

This test assumes that the null hypothesis corresponds with identical medians of the groups, and the alternative hypothesis is that at least one median of one group is different from the median of other group [20].

By other hand, Wilcoxon signed-rank test is a non-parametric test used for comparisons between two populations. In this manner, we have used the Wilcoxon test to check whether the best cases using ^{222}Rn+T and ^{222}Rn are statistically different.

3. Discussion

3.1. Forecasting

The dataset has been separated in two different parts: 70% of the total amount for the training set and 30% for the test set. Differences between architectures and loop-forward are not significant in the behaviour of the forecasting, unlike for the study of the test errors. For this reason, in the following two figures we only show the results obtained with a CNN architecture, 52 values for the loop-back and 1 for the loop-forward.

In Fig. 9, the training set and the test set are represented with its associated NN forecasting. The predictions form a band using the approach explained before and fits the data very well. We can appreciate that the NN captures the past information of the time series and can reproduce both the behaviour of the signal and the levels. With the plot 7b we can conclude that in the region of high values the model outputs lower values than the real ones, in contrast to regions of low values, where the model obtains better predictions. This property of our model is essential to schedule maintenance operation at underground laboratories because the unshielding of

the detectors needs to take place when the ^{222}Rn levels are low.

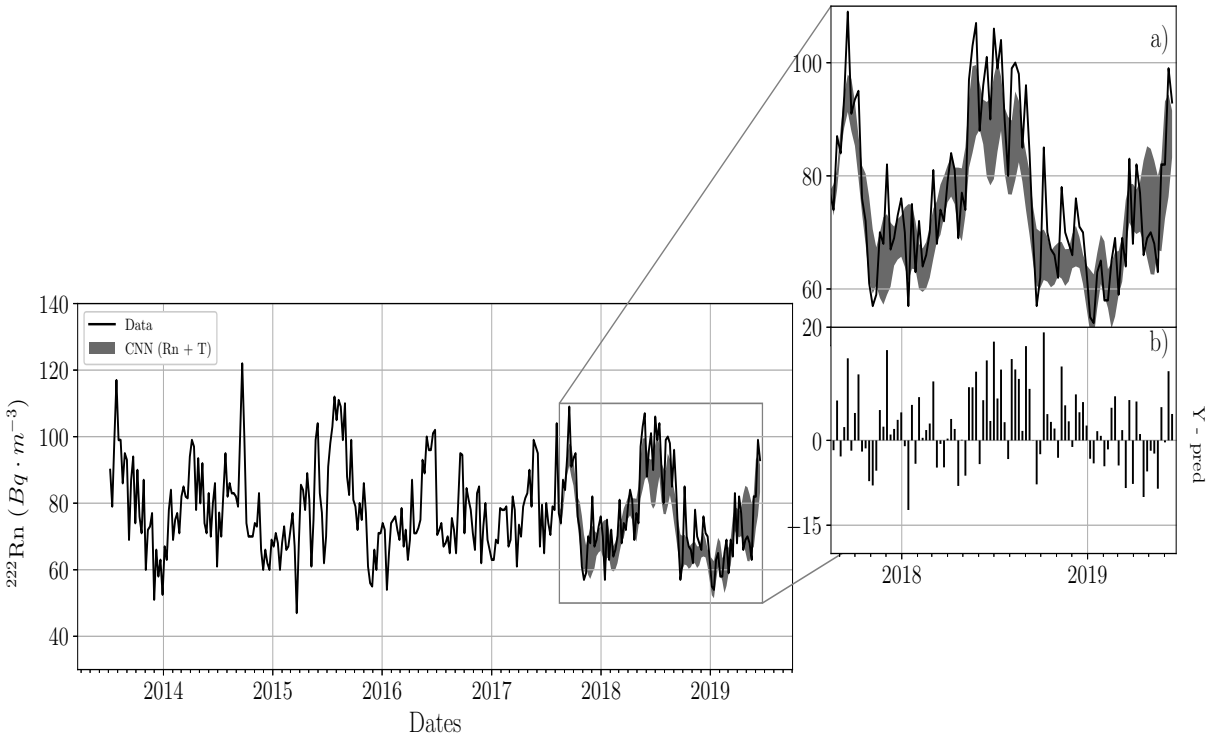

Fig. 9: Weekly ^{222}Rn signal (solid black) and the CNN prediction on the test set (gray band). a) Zoomed plot of the prediction. b) Distances between data and average values of the prediction band.

In Fig. 10 we show the validation plot of the neural network. This 8 values are the result of using the last window of the test set (52 values) to feed the model once is trained.

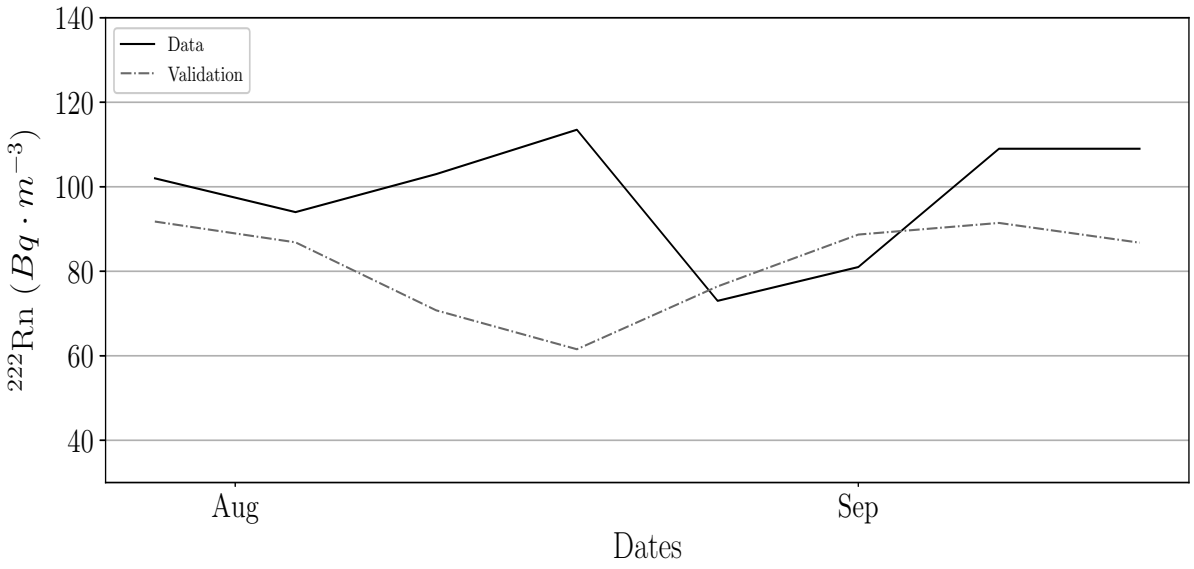

Fig. 10: 8 weeks of ^{222}Rn signal and nowcasting for the validation set - 2 months between 2019 August and 2019 September.

Collecting the forecasting and validation plots of the Radon signal we can state that we have found a non-overfitted sub-optimal architecture that is good enough to carry out this study. In the next few figures, the statistical analysis of the test errors for every architecture are represented.

3.2. Performance improvement due to enviromental variables

In Fig. 11 the Mean Absolute Errors (MAE) boxplots for 25 independent runs of the ANN and for $2, 4, 6$ and 8 values of loop-forward are shown. The X-label marks the input that feeds the neural network, see only Radon on the left and Radon plus the temperature of the corresponding city on the rest. The information of the plot is completed on Table 2, where the mean

and deviation of the 25 independent runs are described.

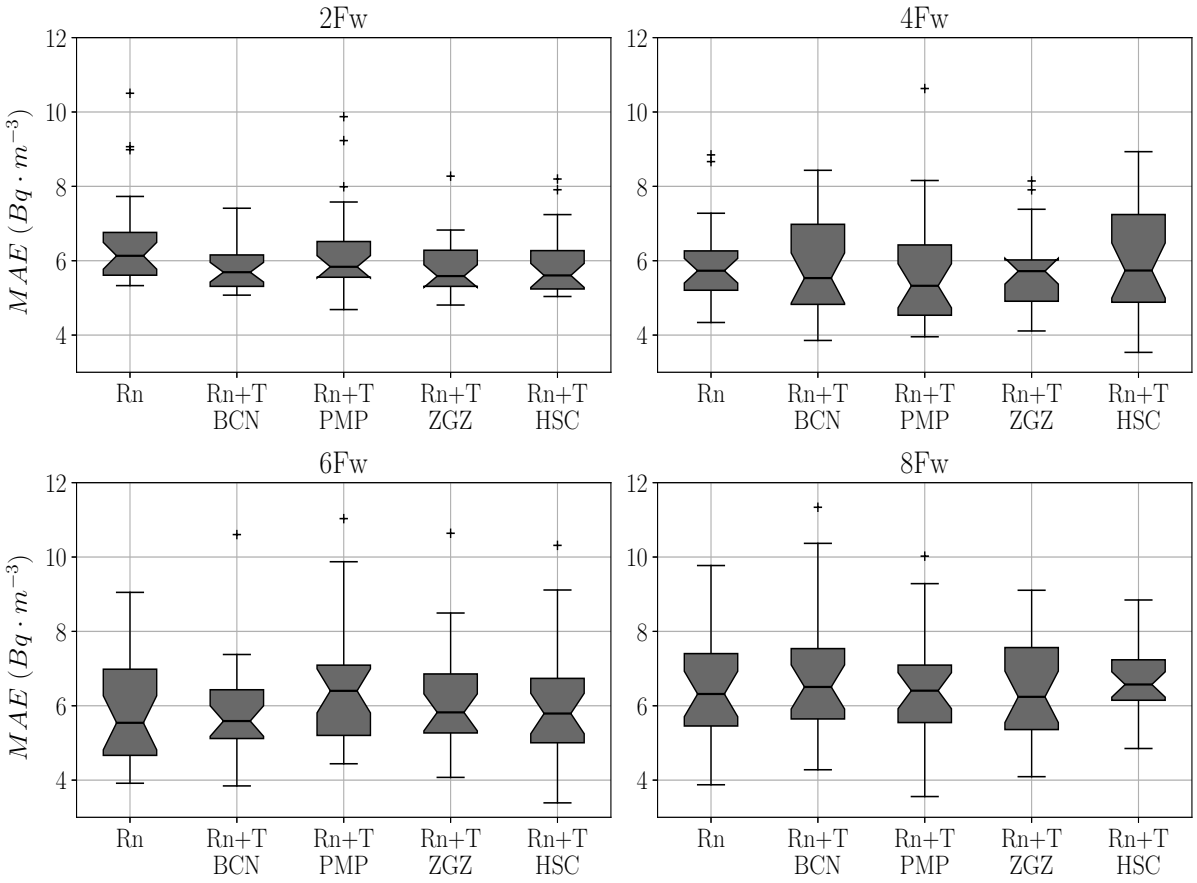

Fig. 11: Mean Absolute Error Boxplots for $N = 2, 4, 6, 8$ future values predicted in two cases: Rn as input of the ANN and Rn + Temperature as input for each city.

	^{222}Rn	$(\overline{\Delta y} \pm \sigma)_{BCN}$	$(\overline{\Delta y} \pm \sigma)_{PMP}$	$(\overline{\Delta y} \pm \sigma)_{ZGZ}$	$(\overline{\Delta y} \pm \sigma)_{HSC}$
2 Forward	6.52 ± 1.29	5.87 ± 0.72	6.23 ± 1.25	5.79 ± 0.74	5.93 ± 0.89
4 Forward	5.92 ± 1.17	5.91 ± 1.36	5.63 ± 1.56	5.70 ± 1.09	5.90 ± 1.52
6 Forward	5.85 ± 1.48	5.98 ± 1.31	6.60 ± 1.79	6.12 ± 1.50	6.12 ± 1.58
8 Forward	6.56 ± 1.63	6.72 ± 1.82	6.52 ± 1.53	6.60 ± 1.94	6.88 ± 1.56

Table 2: Mean and standard deviation (σ) of Mean Absolute Error ($\overline{\Delta y}$) for 25 independent runs for the test data set and for two different cases: ANN predicting 2, 6 and 8 future values for each input window.

As can be seen, the behaviour is pretty similar in every case, which means that the ANN is independent of loop-forward. Looking at the best case, 4Fw, the model has obtained the best prediction using the temperature of Pamplona, its minimal MAE is 3.95. Even though, the most accurate model corresponds with Zaragoza because its boxplot is the most compacted ($\sigma = 1.09$).

Following the same schema presented for the ANN, Fig. 12 and Table 3 gather the results produced by the CNN.

In this matter, we can see that the MAE are generally larger than for ANN in the both cases of feeding the NN. The standard deviation of the boxplots has considerably decreased remaining the improvement of the performance for ^{222}Rn+T for every loop-forward.

5.37 is the lowest MAE obtained for Huesca and a loop-forward equals to 8. This value is considerably larger than the minimum prediction of the ANN. Also, Huesca is the most accurate model due to the compacteness of its boxplot.

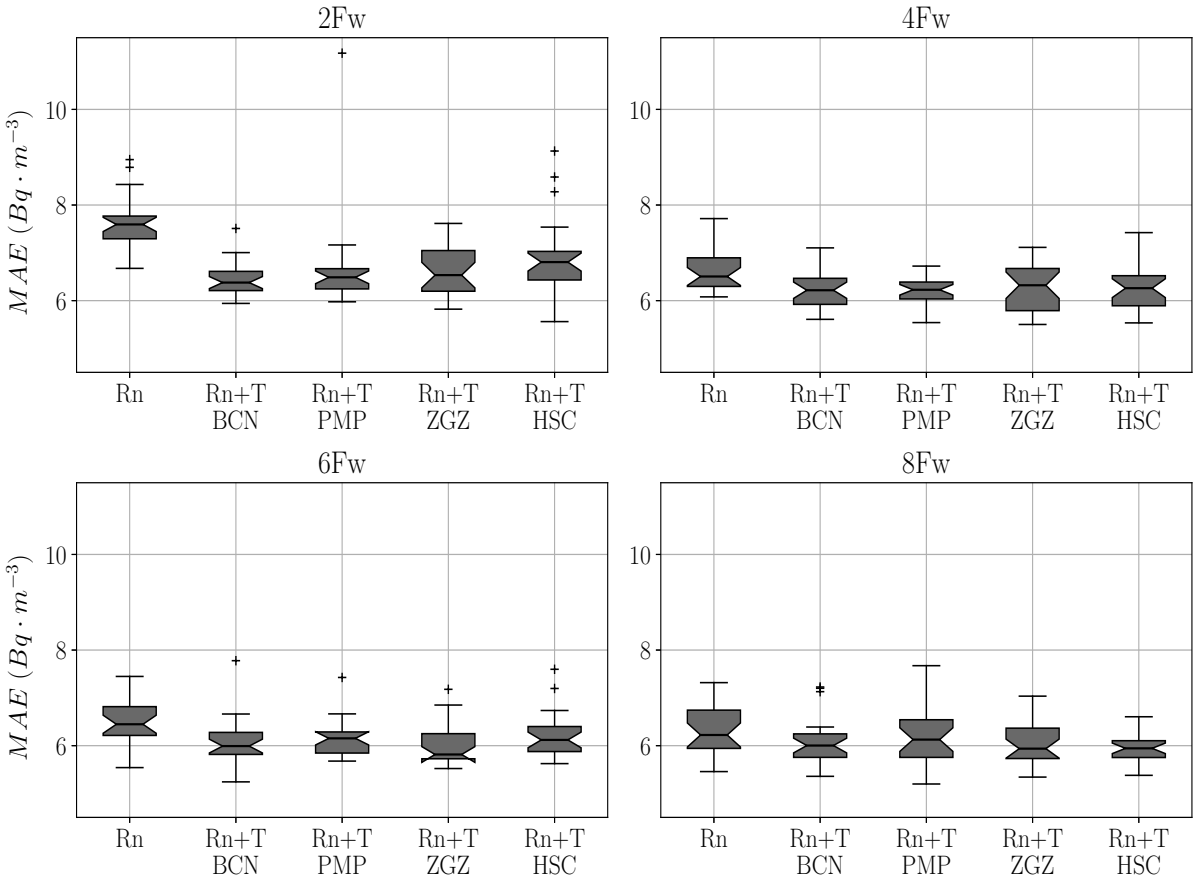

Fig. 12: Mean Absolute Error Boxplots for $N = 2, 4, 6, 8$ future values predicted in two cases: Rn as input of the CNN and Rn + Temperature as input for each city.

	^{222}Rn	$(\overline{\Delta y} \pm \sigma)_{BCN}$	$(\overline{\Delta y} \pm \sigma)_{PMP}$	$(\overline{\Delta y} \pm \sigma)_{ZGZ}$	$(\overline{\Delta y} \pm \sigma)_{HSC}$
2 Forward	7.60 ± 0.59	6.45 ± 0.36	6.66 ± 0.99	6.63 ± 0.52	6.87 ± 0.83
4 Forward	6.60 ± 0.41	6.22 ± 0.39	6.19 ± 0.28	6.28 ± 0.48	6.31 ± 0.54
6 Forward	6.51 ± 0.47	6.09 ± 0.51	6.13 ± 0.39	6.00 ± 0.40	6.23 ± 0.46
8 Forward	6.35 ± 0.52	6.06 ± 0.51	6.15 ± 0.53	6.03 ± 0.42	5.93 ± 0.30

Table 3: Mean and standard deviation (σ) of Mean Absolute Error ($\overline{\Delta y}$) for 25 independent runs for the test data test and for two different cases: CNN predicting 2, 6 and 8 future values for each input window.

Recurrent Neural network is the last architecture used in this work. As can be seen, RNN performance is much better for larger loop-forwards than for smaller ones. The best prediction is obtained for only ^{222}Rn and a loop-forward equals to 8, reaching a MAE of 6.05, nevertheless, Pamplona is the most accurate model. This architecture is the worst one in terms of the MAE for this study.

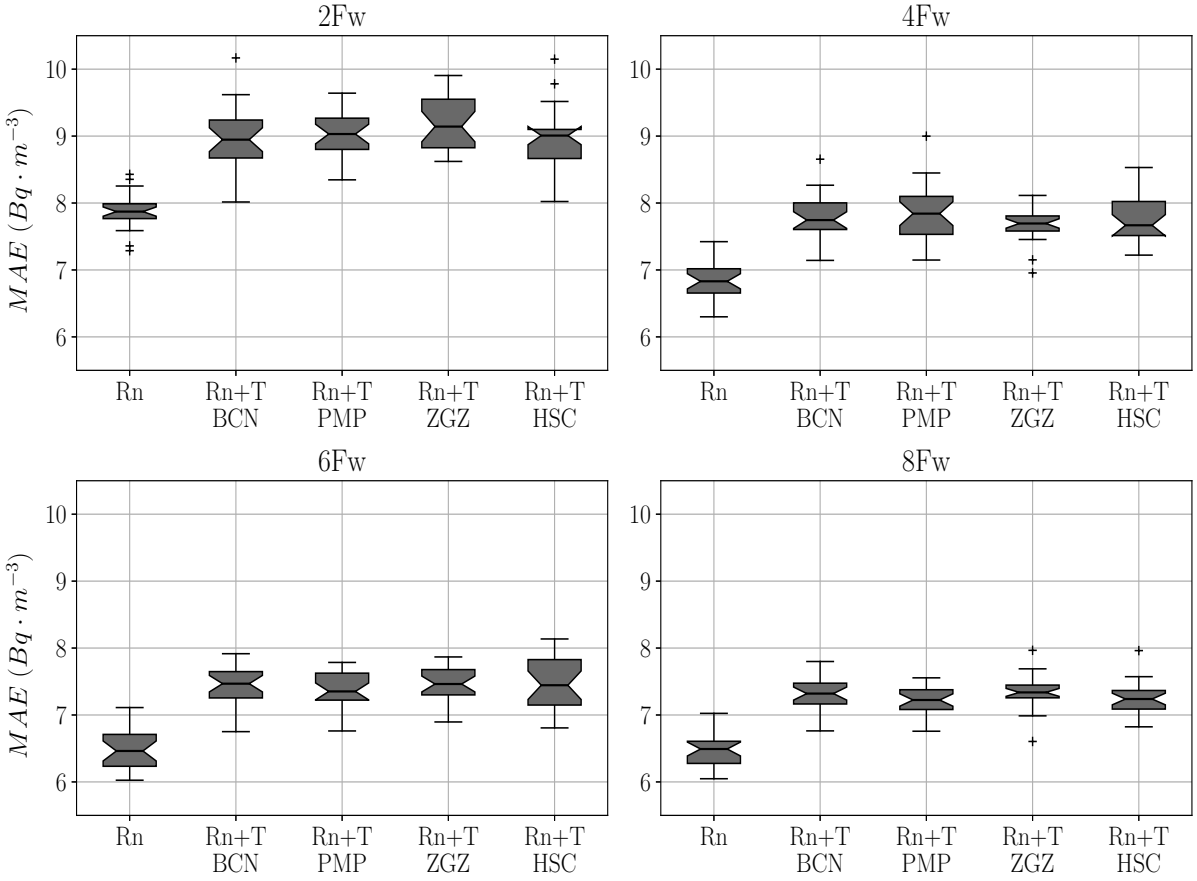

Fig. 13: Mean Absolute Error Boxplots for $N = 2, 4, 6, 8$ future values predicted in two cases: Rn as input of the RNN and Rn + Temperature as input for each city.

	^{222}Rn	$(\overline{\Delta y} \pm \sigma)_{BCN}$	$(\overline{\Delta y} \pm \sigma)_{PMP}$	$(\overline{\Delta y} \pm \sigma)_{ZGZ}$	$(\overline{\Delta y} \pm \sigma)_{HSC}$
2 Forward	7.89 ± 0.27	8.96 ± 0.47	9.03 ± 0.34	9.18 ± 0.41	8.99 ± 0.44
4 Forward	6.81 ± 0.28	7.79 ± 0.34	7.83 ± 0.43	7.68 ± 0.25	7.75 ± 0.36
6 Forward	6.48 ± 0.30	7.43 ± 0.30	7.37 ± 0.27	7.47 ± 0.25	7.46 ± 0.39
8 Forward	6.48 ± 0.23	7.31 ± 0.28	7.20 ± 0.21	7.33 ± 0.25	7.25 ± 0.25

Table 4: Mean and standard deviation (σ) of Mean Absolute Error ($\overline{\Delta y}$) for 25 independent runs for the test data test and for two different cases: RNN predicting 2, 6 and 8 future values for each input window.

In this case, there is not improvement due to adding the temperature to the dataset simply looking at the boxplots. To see this effects on the output see Table 5, where are presented the p-values of the two statistical test explained before: Kruskal-Wallis and Wilcoxon signed-rank test.

P-values lower than $5 \cdot 10^{-2}$ means that the distributions are different with more than 95% of confidence. Kruskal Wallis test has been computed for all of the cases with the best loop-forward (4 in the case of ANN and 8 for CNN and RNN), in contrast, Wilocoxon signed-rank test has been computed for the ^{222}Rn case and the city (temperature) that outputs the lowest MAE.

Adding the Bonferroni Correction [21] to the Kruskal-Wallis test the critical p-value would be $1.25 \cdot 10^{-2}$

	Kruskal-Wallis	Wilcoxon
ANN (4Fw)	$8.03 \cdot 10^{-1}$	$3.26 \cdot 10^{-1}$ (PMP)
CNN (8Fw)	$4.63 \cdot 10^{-2}$	$5.45 \cdot 10^{-4}$ (HSC)
RNN (8Fw)	$2.24 \cdot 10^{-12}$	$1.23 \cdot 10^{-5}$ (^{222}Rn)

Table 5: P-values of the Kruskal-Wallis test between the five distributions for the best loop-forward and the p-values of the Wilcoxon signed rank test between the ^{222}Rn case and the best city.

The ANN does not pass any statistical test because of its p-values; we can nor assure that the four distributions are statistically different, neither the ^{222}Rn and the Pamplona distrbutions for 4Fw. By other hand, Temperature is essential to add new information to the dataset CNN. In this case, the results don't pass the Kruskal Wallis test with the Bonferroni Correction but the Wilcoxon signed-ranked test is well satisfied. The pvalue are much smaller than $5 \cdot 10^{-2}$, which means that the improvement of the environment variables is statistically consistent. Finally, RNN result passes both statistical test: the distributions are statistically different and the temperature don't adds further information to the dataset. From these results we see a strong dependence of the further information added by ambiental variables and the type of neural network architecture; study this effect is proposed as future work.

Furthermore, a quantitative measure of the improvement of the predictions is presented in Fig. 14. The position of the triangles has been computed dividing the average MAE for only ^{222}Rn as input, labeled as $\overline{\Delta y_{Rn}}$, by the minimum average MAE for ^{222}Rn+T, labeled as $\overline{\Delta y_{RnT_{min}}}$, for a range of loop-forwards between 1 to 8; this is the so-called improvement factor. Us-

ing this approach we can distinguish the loop-forwards that don't improves the performance, measure the maximum improvement factor when adding the temperature to the dataset and study the behaviour of the different architectures used.

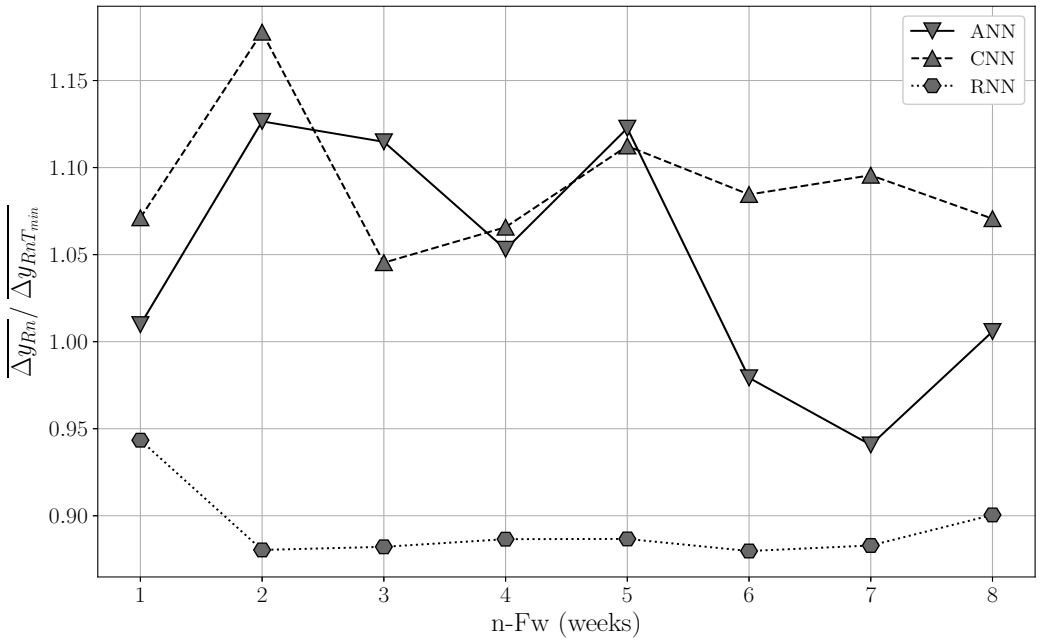

Fig. 14: Fraction between the average MAE obtained with only Radon as input and the minimum MAE of Radon plus the temperature of the corresponding city versus the loop-forward. Downside triangles for ANN, upside triangles for CNN and hexagons for RNN.

Regarding the results, the improvement factor on ANN depends strongly on the loop-forward. Following the discussion of the ANN, the maximum improvement takes place for 2 of loop-forward reaching a value of 1.13, the

second local maximum takes place in loop-forward = 5. The temperature doesn't improve the performance for 6 and 7, meanwhile, for 8 it doesn't have any effect on the improvement.

CNN presents a similar dependence on the loop-forward. Firstly, the improvement factors are mostly above the ANN ones, meaning that the temperature always improves the forecasting of the ^{222}Rn. The local maximums takes place in the same positions than the ANN and the global maximum corresponds with an improvement factor of 1.18.

The improvement factor of RNN is close to 1 in every case and doesn't have a strong dependence on the loop-forward. Even though the populations are not statistically different there is a smooth improvement when adding the temperature.

In contrast, RNN improving factor is always below the unity; meaning that there is not improvement due to adding the temperature, as we saw in the boxplots. Predicting one future week is the best case, where the factor takes a value of 0.94.

4. Conclusions

In this work, we have forecast the ^{222}Rn Time Series at the Canfranc Underground Laboratory using several neural network architectures like ANN, CNN and RNN. As explained, the ^{222}Rn levels at underground experiments contributes to the background signal that is needed to get rid of to achieve more measurements precision.

The data used ranges from July 2013 to July 2019 and has been summed up in week values. By other hand, the environment temperature, preassure

and wind velocity of 4 surrounding cities to the LSC: Barcelona, Pamplona, Huesca and Zaragoza has been downloaded and preprocessed. Missing data has been filled using ARIMA algorithm and the variable importances has been computed using Random Forest algorithm. At the end of the study, temperature was the only variable selected for being added to the dataset. As usual in every machine learning problem, the dataset has been separated in 70% of training set, 30% of test set and 8 values of validation.

We found a sub-optimal configuration of hyperparameters for the different architectures that captures the past information and reproduces both the test set and the validation set. Moreover, the statistical analysis of the Mean Absolute Errors demonstrates that the inclusion of the temperature to the dataset improves the predictions obtaining smaller errors than the case of using only ^{222}Rn as input in ANN and CNN, but not for RNN. From this results we see that the best for ANN corresponds with the temperature from Pamplona and for a loop-forward of 4, for CNN the best result is obtained for loop-forward of 8 and the temperature of Huesca and for RNN 8 of loop-forward and using only the ^{222}Rn levels. ANN obtains the lowest value of the MAE, which is, 3.95 between the three architectures, followed by the CNN (5.37) and finally, the RNN (6.05).

Furthermore, the Kruskal-Wallis test has been used to show that the errors distributions are different in this two ways of feeding the neural network and we can state that the temperature includes additional information to the ^{222}Rn forecasting and is needed to improve the performance of the training process in the CNN but, even though for ANN we don't see a statistical difference between the populations Rn and Rn+T there exist a clear improvement

in every case.

Finally, the improvement factor are computed for the three architectures. Even the improvement factor of ANN is maximum for 2 of loop-forward, the MAE are smaller for loop-forward equals to 4. Combining this results we can say that the best model to forecast the ^{222}Rn signal is the ANN with a loop-forward of 4. This means that we can predict the ^{222}Rn levels one month forward, which is enough time to schedule maintenance operations at underground laboratories.

This methodology could be applied in any other underground experiment because the radon problem is always persisting on experiments where high precision measurements are needed. Also, this study would be key for drive studies on contaminants forecasting, where the environmental variables affect directly to the density of the pollutants at a given place.

References

[1] Z. Begin, G. Steinitz, Temporal and spatial variations of microearthquake activity along the dead sea fault, 1984-2004, Israel Journal of Earth Sciences 54 (2005) 1–14. doi:10.1560/QTVW-HY1E-7XNU-JCLJ.

[2] A. D. K. Tareen, M. Nadeem, K. Kearfott, K. Abbas, M. Khawaja, K. Rafique Mir, Descriptive analysis and earthquake prediction using boxplot interpretation of soil radon time series data, Applied Radiation and Isotopes 154 (2019) 108861. doi:10.1016/j.apradiso.2019.108861.

[3] A. eff darwich, C. Martín-Luis, M. Quesada, J. Nuez, J. Coello, Vari-

ations on the concentration of 222rn in the subsurface of the volcanic island of tenerife, canary islands, Geophysical Research Letters - GEOPHYS RES LETT 29 (2002) 26–1. doi:10.1029/2002GL015387.

[4] X. Chen, J. Paatero, V.-M. Kerminen, L. Riuttanen, J. Hatakka, V. Hiltunen, P. Paasonen, A. Hirsikko, A. Franchin, H. Manninen, T. Petäjä, Y. Viisanen, M. Kulmala, Responses of the atmospheric concentration of radon-222 to the vertical mixing and spatial transportation 21 (2016) 299–318.

[5] M. Greenfield, A. Domondon, N. Okamoto, I. Watanabe, Variation in -ray count rates as a monitor of precipitation rates, radon concentrations, and tectonic activity, Journal of Applied Physics 91 (3) (2002) 1628–1633. doi:10.1063/1.1426248.

[6] M. Wang, G. Audi, F. Kondev, W. Huang, S. Naimi, X. Xu, The ame2016 atomic mass evaluation (ii). tables, graphs and references, Chinese Physics C 41 (2017) 030003. doi:10.1088/1674-1137/41/3/030003.

[7] F. Girault, B. Koirala, M. Bhattarai, S. Rajaure, P. Richon, F. Perrier, Radon emanation of rock and soil samples: A tool for stratigraphy, geology, geophysical modeling and radon health hazard, 2010.

[8] B.-R. Montes Núñez, Analysis of the first underground run and background studies of the argon dark matter experiment, Ph.D Thesis, Universidad Complutense de Madrid (2017).

[9] O. Monge, M.A, Design, scale-up and characterization of the data ac-

quisition system for the anais dark matter experiment., Ph.D Thesis, Universidad de Zaragoza (2015).

[10] S. M. Barbosa, H. Zafrir, U. Malik, O. Piatibratova, Multiyear to daily radon variability from continuous monitoring at the amram tunnel, Geophysical Journal International 182 (2010) 829–842. doi:https://doi.org/10.1111/j.1365-246X.2010.04660.x.

[11] M. Cárdenas-Montes, I. Méndez-Jiménez, Ensemble deep learning for forecasting 222 rn radiation level at canfranc underground laboratory, Springer Nature: Advances in Intelligent Systems and Computing 14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019) May 13 – 15, proceedings. doi:https://doi.org/10.1007/978-3-030-20055-8.

[12] I. Mendez-Jimenez, M. Cardenas-Montes, Modelling and Forecasting of the ^{222}Rn Radiation Level Time Series at the Canfranc Underground Laboratory, 2018, pp. 158–170. doi:10.1007/978-3-319-92639-1_14.

[13] I. Mendez-Jimenez, M. Cardenas-Montes, Time Series Decomposition for Improving the Forecasting Performance of Convolutional Neural Networks: 18th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2018, Granada, Spain, October 23–26, 2018, Proceedings, 2018, pp. 87–97. doi:10.1007/978-3-030-00374-6_9.

[14] R. A. Vasco-Carofilis, M. A. Gutierrez-Naranjo, M. Cardenas-Montes, PBIL for optimizing hyperparameters of Convolutional Neural Networks and STL Decomposition (sent), HAIS2020.

[15] M. Cardenas-Montes, Uncertainty estimation in the forecasting of the 222Rn Radiation Level Time Series at the Canfranc Underground Laboratory, SOCO2019.

[16] R. Hyndman, G. Athanasopoulos, Forecasting: principles and practice., OTexts (2014).

[17] L. Breiman, Machine learning, volume 45, number 1 - springerlink, Machine Learning 45 (2001) 5–32. doi:10.1023/A:1010933404324.

[18] I. Goodfellow, Y. Bengio, A. Courville, Deep learning., MIT Press, Massachusetts (2016).

[19] F. Gers, J. Schmidhuber, F. Cummins, Learning to forget: Continual prediction with lstm, Neural computation 12 (2000) 2451–71. doi:10.1162/089976600300015015.

[20] S. García, D. Molina, M. Lozano, F. Herrera, A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: A case study on the cec'2005 special session on real parameter optimization, J. Heuristics 15 (2009) 617–644. doi:10.1007/s10732-008-9080-4.

[21] C. W. Dunnett, A multiple comparison procedure for comparing several treatments with a control, Journal of the American Statistical Association 50 (272) (1955) 1096–1121.