

# Nowcasting for improving $^{222}\text{Rn}$ forecasting at LSC

T. Sánchez-Pastor<sup>1,\*</sup>

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

Miguel Cárdenas-Montes<sup>2</sup>

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

?<sup>3</sup>

*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  $^{222}\text{Rn}$  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 five years of weekly  $^{222}\text{Rn}$  levels using deep learning techniques such as Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN) and Long-Short Term Memory Nets (LSTM). 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  $^{222}\text{Rn}$  levels at LSC as input, we predict the Rn time series using Neural Networks and we estimate the number of past weeks where the temperature doesn't improves the performance.

CAMBIAR TODO

*Keywords:* Neural Network ·  $^{222}\text{Rn}$  Measurements · Canfranc Underground Laboratory · Forecasting.

---

\*Corresponding author

Email addresses: [tomas.sanchez@ciemat.es](mailto:tomas.sanchez@ciemat.es) (T. Sánchez-Pastor), [miguel.cardenas@ciemat.es](mailto:miguel.cardenas@ciemat.es) (Miguel Cárdenas-Montes), [miguel.cardenas@ciemat.es](mailto:miguel.cardenas@ciemat.es) (?), [author.three@mail.com](mailto:author.three@mail.com) (?)

<sup>1</sup>This is the first author footnote.

<sup>2</sup>Second author footnote.

<sup>3</sup>Yet another author footnote.

## 1. Introduction

The study of the  $^{222}\text{Rn}$  Time Series has been increasingly gaining importance in the past decades. The main reason is that  $^{222}\text{Rn}$  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.

$^{222}\text{Rn}$  Time Series are produced by the  $^{238}\text{U}$  decay chain that emits  $\alpha$  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:  $^{222}\text{Rn}$  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  $^{222}\text{Rn}$  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  $^{222}\text{Rn}$  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 used for forecasting the signal. Also, it is important to note that  $^{222}\text{Rn}$  signal is correlated with ambiance variables, as demonstrated in [10]. This last result opened a new methodology: the nowcasting of Radon Time Series using close LSC locations (Fig.1) ambiance data, which is the main goal of this work.

In this paper, we explain in Section 2 how the data acquisition, data cleaning, gaps filling with ARIMA and preprocessing has been done for 5 years of data between July 2013 and September 2019. Once the data is prepared, we study the correlations between  $^{222}\text{Rn}$  at LSC and temperature, pressure and wind velocity at 4 LSC surrounding cities: Barcelona, Pamplona, Huesca and Zaragoza. Then, Random Forest algorithm is applied for variable importance analysis.

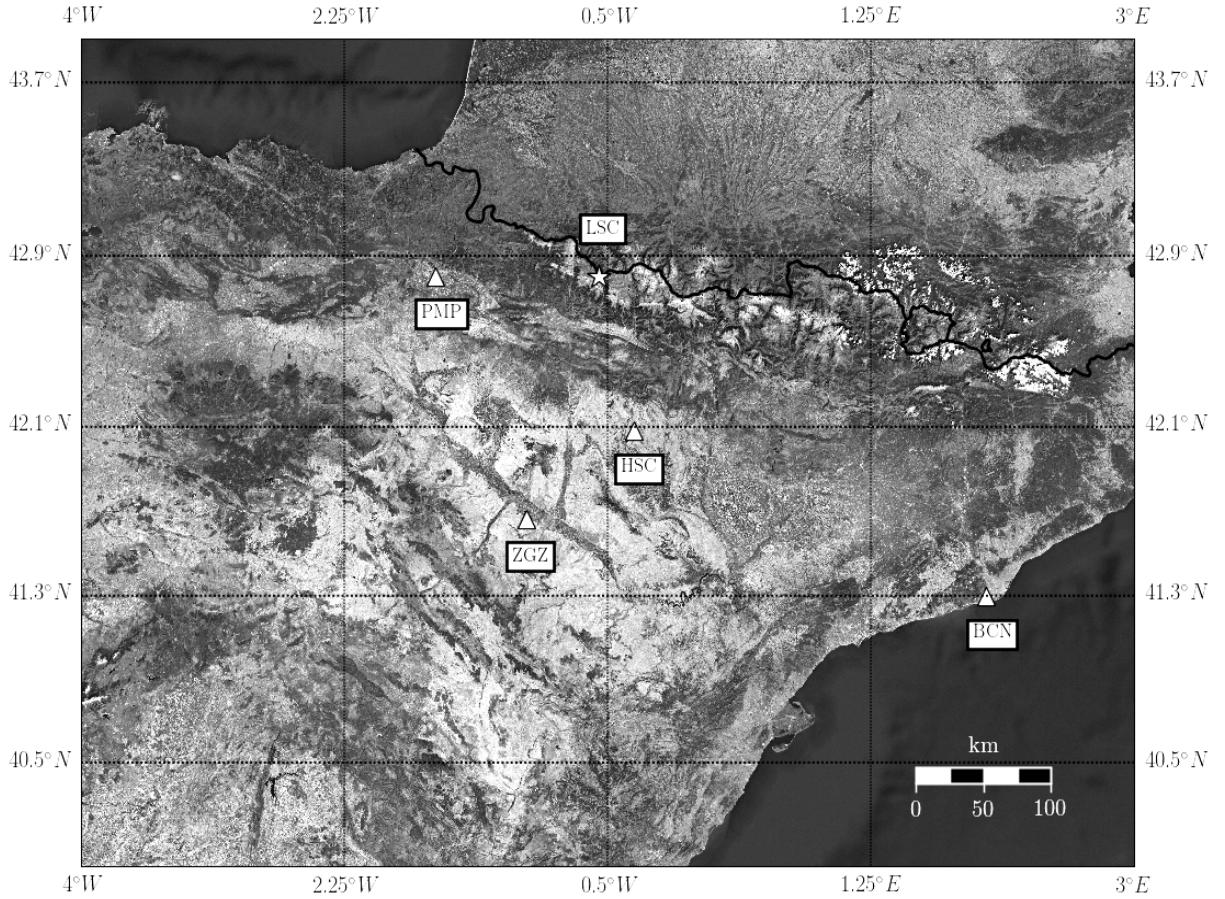


Figure 1: GPS map of the Spanish Peninsula area of interest. Triangles mark the location of the meteorological stations and the star marks the LSC station. Pamplona station is labeled as PMP, Barcelona station as BCN, Huesca station as HSC and Zaragoza station as ZGZ.

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 Bidirectional Long-Short Term Memory networks (LSTM). Finally, we focus on statistical analysis of the mean absolute errors (MAE) for 25 independent runs for each architecture and each city. This results are discussed in Section 3 and the conclusions correspond with 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  $^{222}\text{Rn}$  may pass, while the radon daughter nuclei are prevented to enter the ionization chamber. The ingoing  $^{222}\text{Rn}$  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 in the cathode and the detector counts the resulting electric intensity as one hit of  $^{222}\text{Rn}$ .

Regarding the ambiance data, it has been downloaded from AEMET (Agencia Estatal de Meteorología) Open Data spanish service.

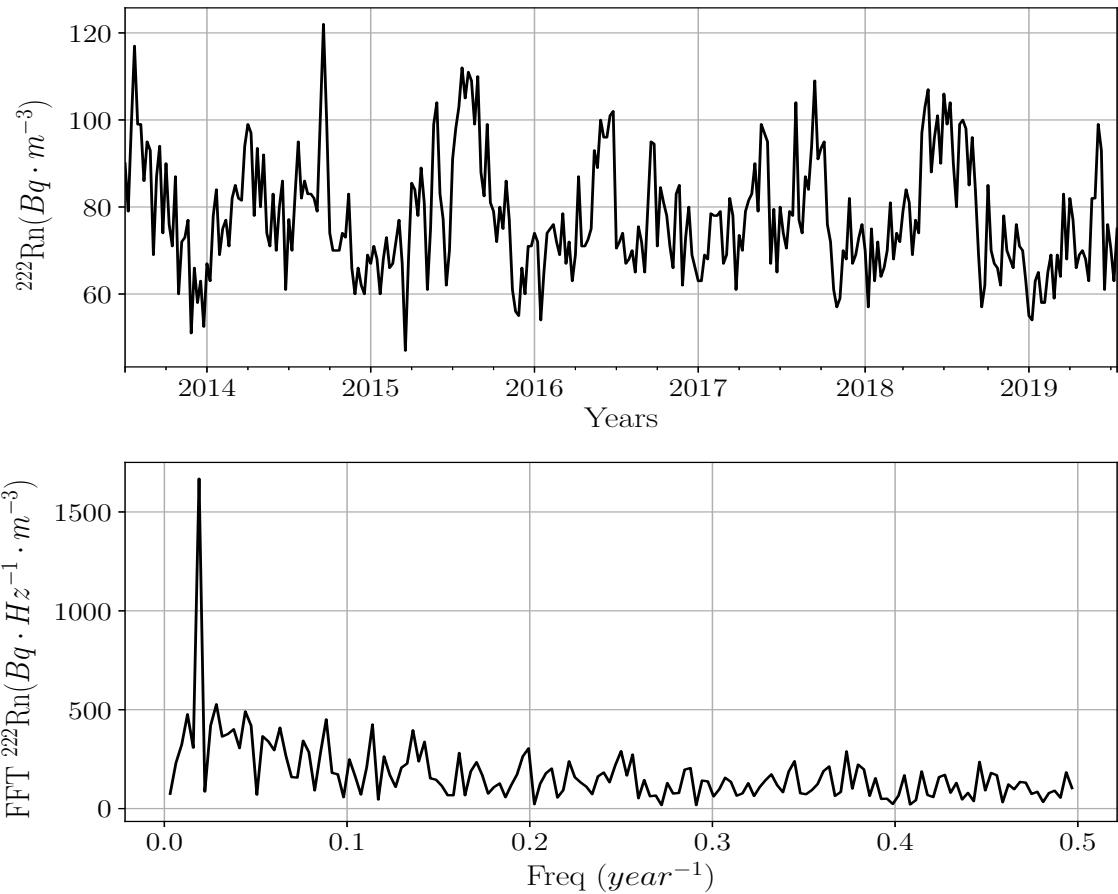


Figure 2:  $^{222}\text{Rn}$  concentrations medians per week from July 2013 to July 2019 at Hall A of the LSC (upper) and its frequency domain obtained after applying a Fast Fourier Transform (FFT) (down).

## 2.2. Correlation with meteorological variables

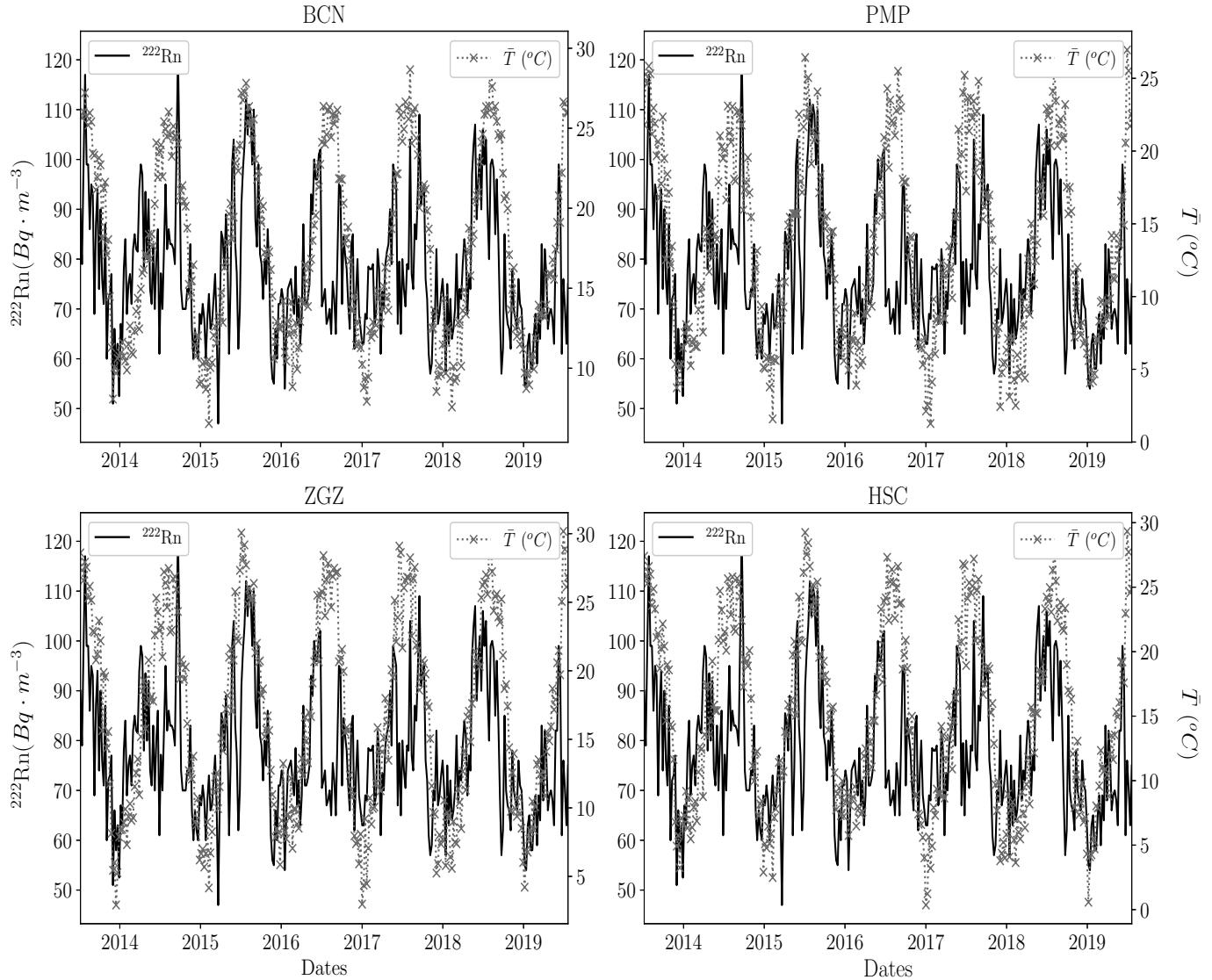


Figure 3: Weekly  $^{222}\text{Rn}$  concentrations medians (solid black) and weekly average temperature in  $^{\circ}\text{C}$  (dashed gray) from July 2013 to July 2019.

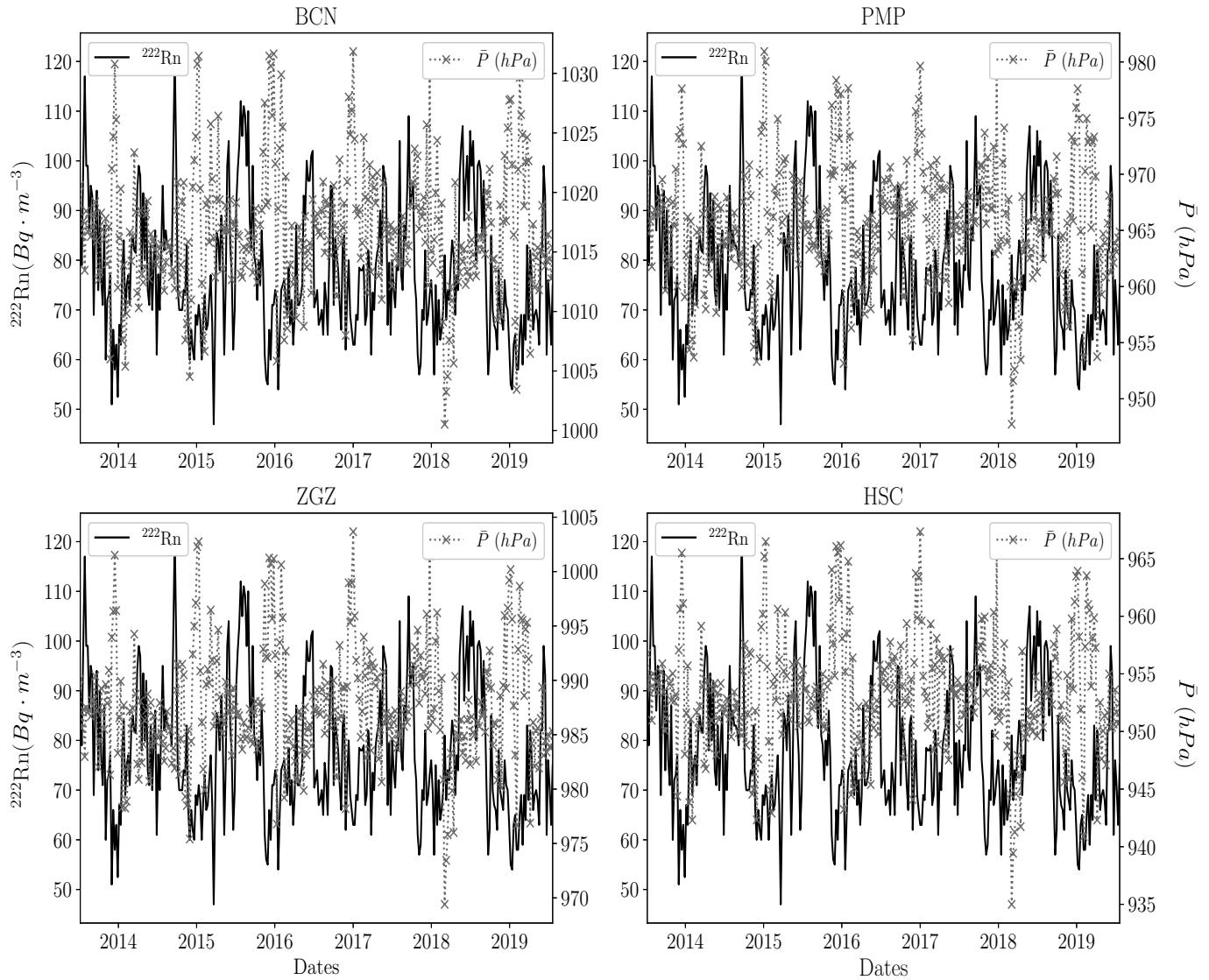


Figure 4: Weekly  $^{222}\text{Rn}$  concentrations medians (solid black) and weekly average pressure in  $\text{hPa}$  (dashed gray) from July 2013 to July 2019.

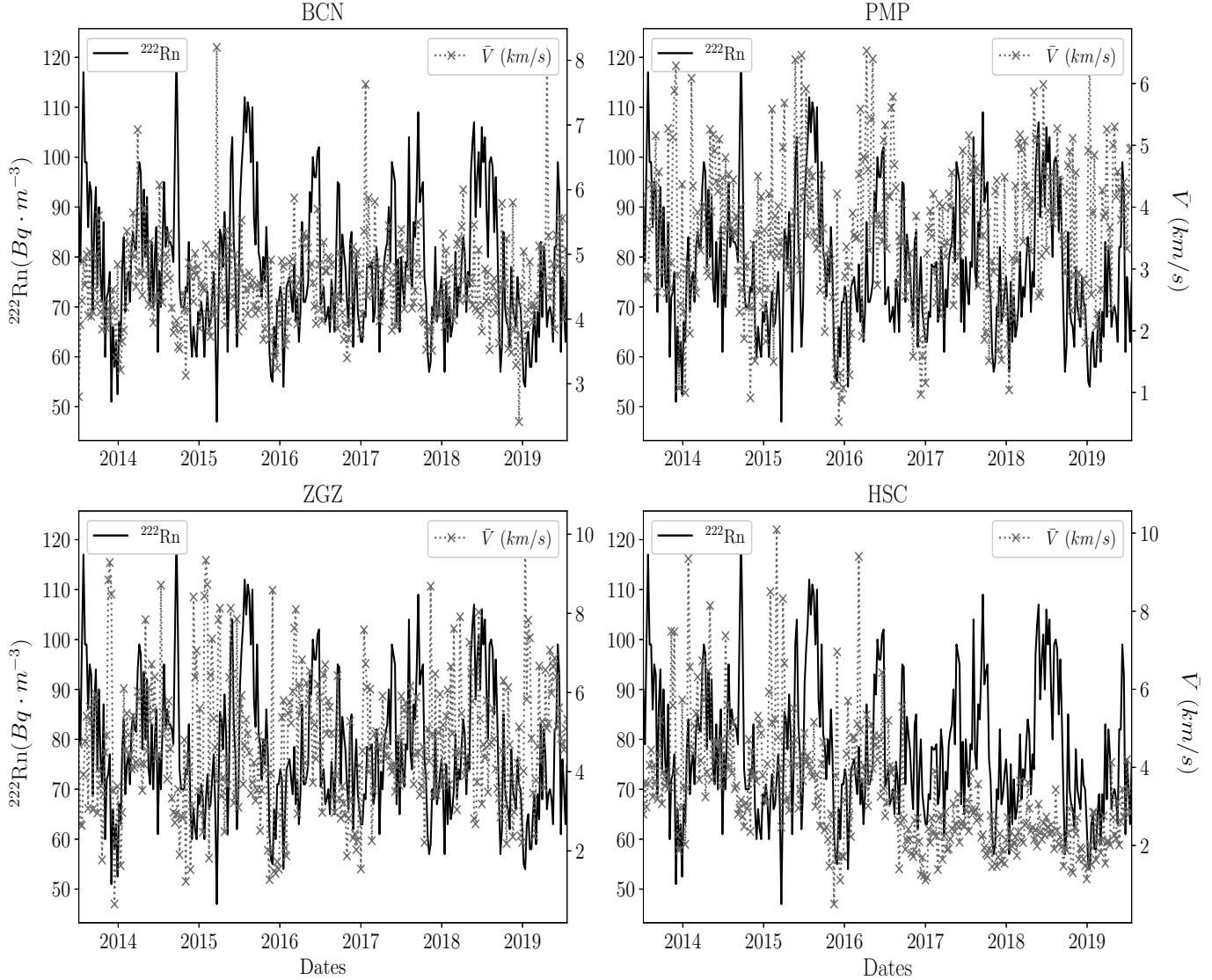


Figure 5: Weekly  $^{222}\text{Rn}$  concentrations medians (solid black) and weekly average wind velocity in  $\text{km} \cdot \text{s}^{-1}$  (dashed gray) from July 2013 to July 2019.

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: Cross correlations of the weekly median  $^{222}\text{Rn}$  levels at LSC and several meteorological variables: average pressure, average temperature and average wind velocity recorded at different stations (labeled with the city name).

### 2.3. Random Forest

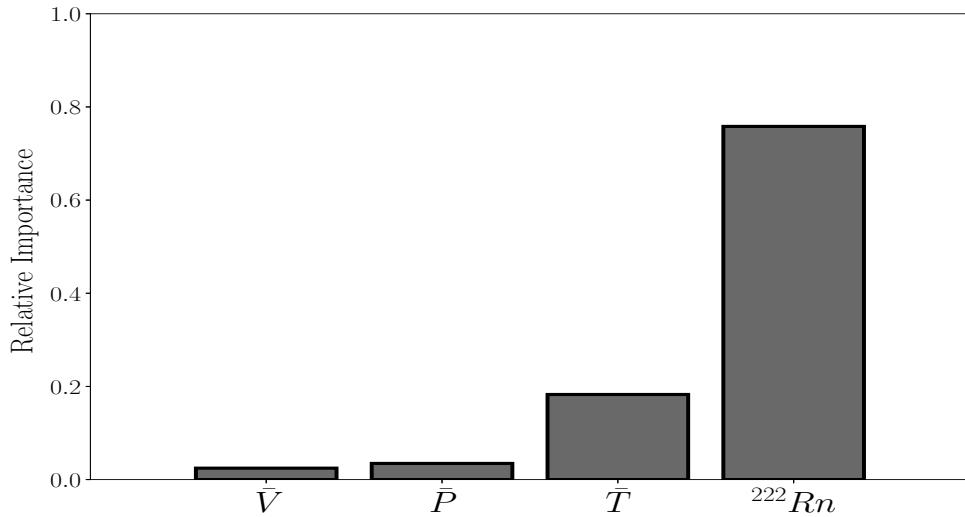


Figure 6: Variable importances to forecasting the signal obtained with Random Forest algorithm. Let  $i$  be  $\{\bar{V}, \bar{P}, \bar{T}, {}^{222}\text{Rn}\}$ , the importances  $I_i$  satisfies the following condition:  $\sum_i I_i = 1$ .

## 2.4. Convolutional Neural Networks

### 3. Results

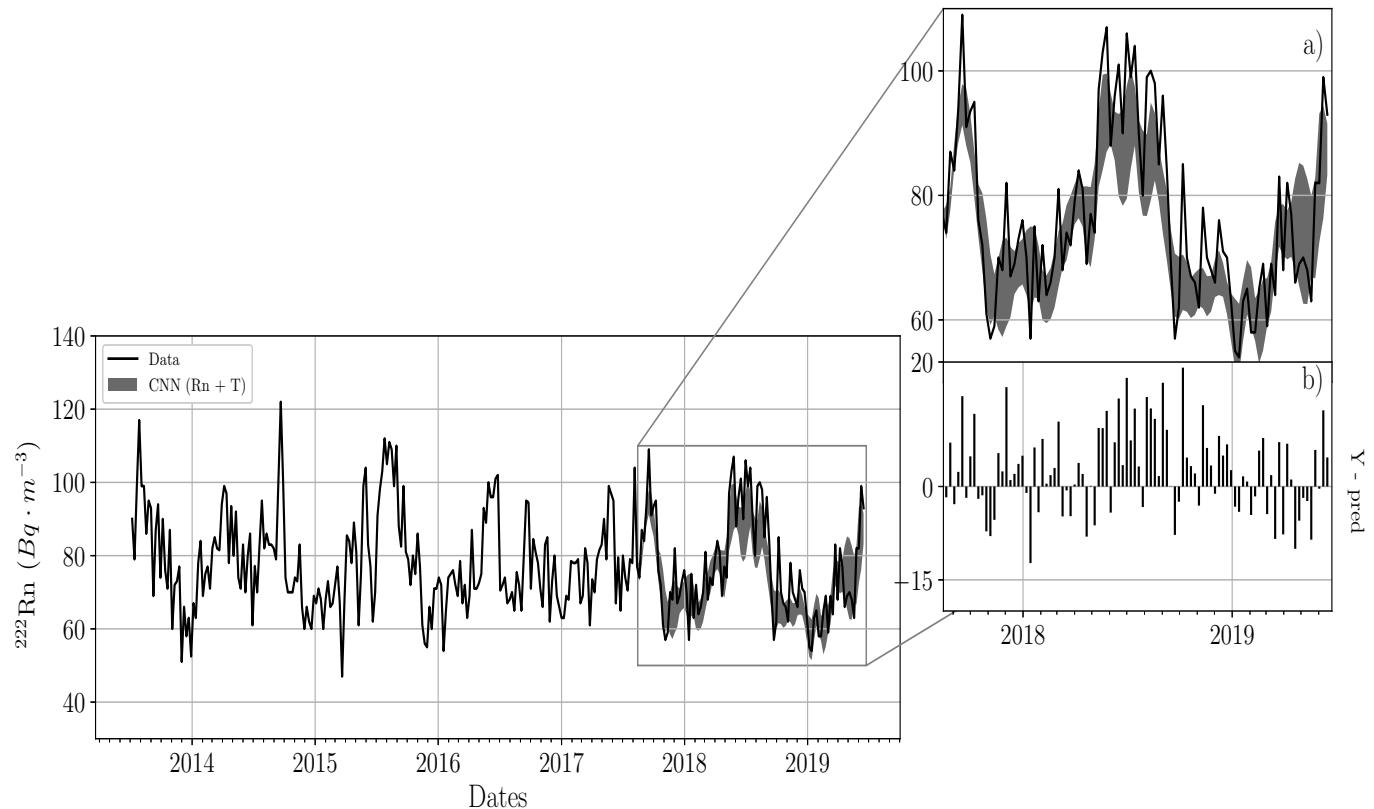


Figure 7: Weekly  $^{222}\text{Rn}$  signal (solid black) and the CNN prediction on the test set (gray band). a) Zoomed plot of the prediction. b) Distances between data and average values of the prediction band.

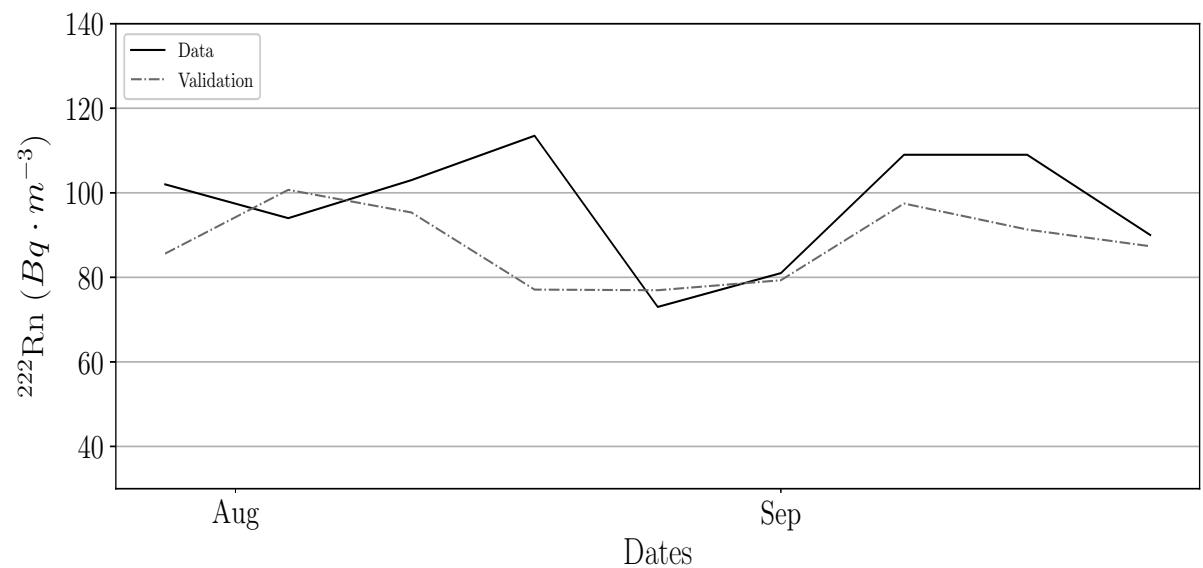


Figure 8: 8 weeks of  $^{222}\text{Rn}$  signal and nowcasting for the validation set - 2 months between 2019 August and 2019 September.

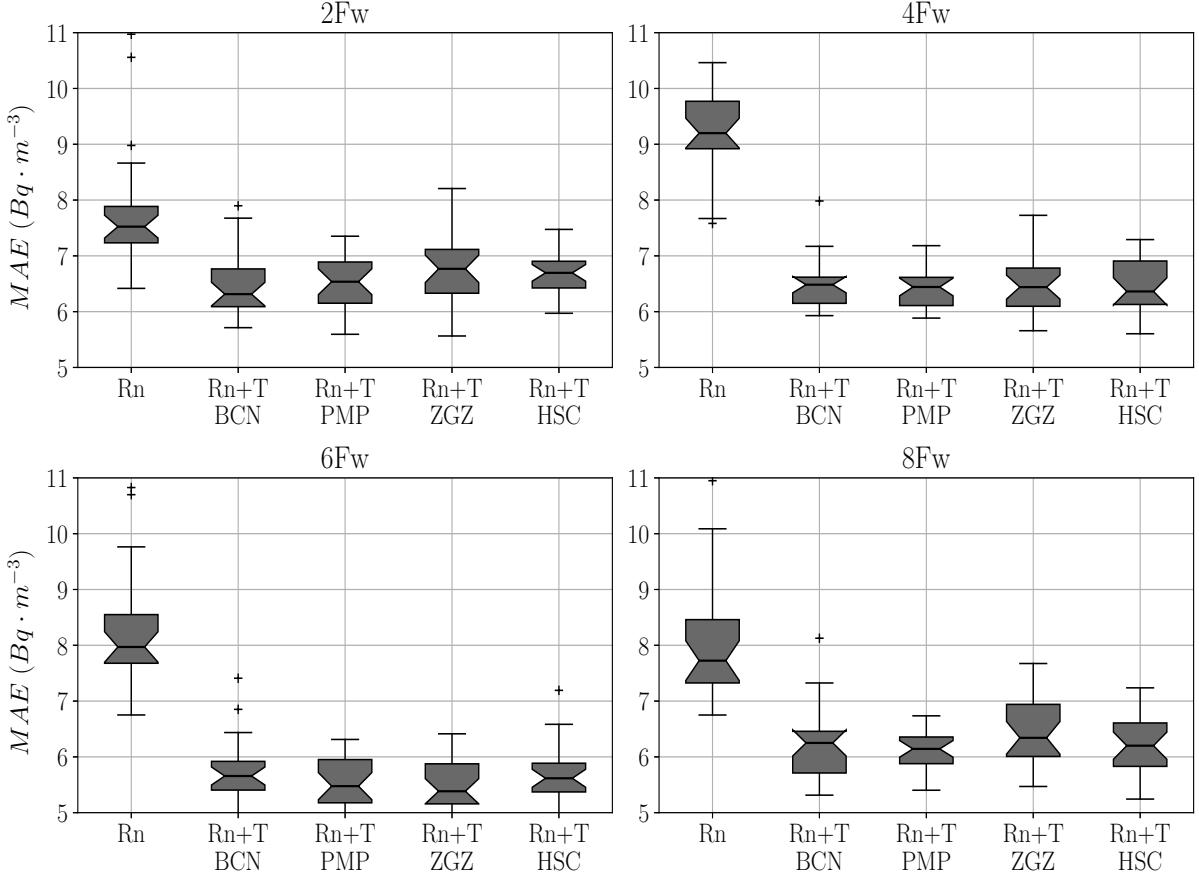


Figure 9: Mean Absolute Error Boxplots for  $N = 2, 4, 6, 8$  future values predicted in two cases: Rn as input of the ANN and Rn + Temperature as input for each city.

	$^{222}\text{Rn}$	$(\bar{\Delta}y \pm \sigma)_{BCN}$	$(\bar{\Delta}y \pm \sigma)_{PMP}$	$(\bar{\Delta}y \pm \sigma)_{ZGZ}$	$(\bar{\Delta}y \pm \sigma)_{HSC}$
2 Forward	$9.24 \pm 0.77$	$6.51 \pm 0.46$	$6.41 \pm 0.36$	$6.45 \pm 0.5$	$6.47 \pm 0.49$
4 Forward	$8.24 \pm 1.07$	$5.7 \pm 0.59$	$5.56 \pm 0.44$	$5.5 \pm 0.44$	$5.65 \pm 0.54$
6 Forward	$8.05 \pm 1.07$	$6.25 \pm 0.62$	$6.14 \pm 0.36$	$6.48 \pm 0.62$	$6.21 \pm 0.52$
8 Forward	$7.8 \pm 1.06$	$6.5 \pm 0.61$	$6.54 \pm 0.5$	$6.73 \pm 0.61$	$6.64 \pm 0.41$

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

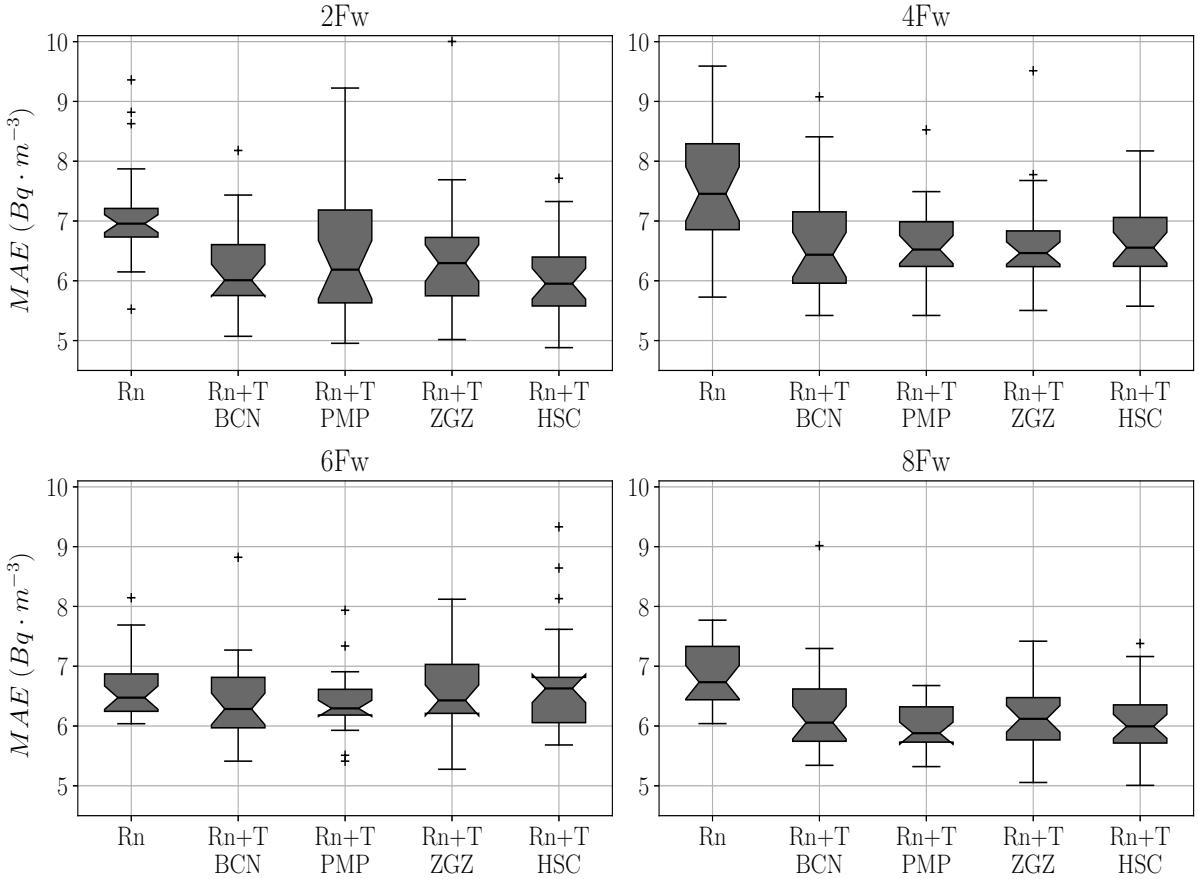


Figure 10: Mean Absolute Error Boxplots for  $N = 2, 4, 6, 8$  future values predicted in two cases: Rn as input of the CNN and Rn + Temperature as input for each city.

	$^{222}Rn$	$(\bar{y} \pm \sigma)_{BCN}$	$(\bar{y} \pm \sigma)_{PMP}$	$(\bar{y} \pm \sigma)_{ZGZ}$	$(\bar{y} \pm \sigma)_{HSC}$
2 Forward	$7.79 \pm 1.37$	$6.58 \pm 0.89$	$6.61 \pm 0.72$	$6.62 \pm 0.8$	$6.63 \pm 0.63$
4 Forward	$6.64 \pm 0.53$	$6.44 \pm 0.72$	$6.41 \pm 0.52$	$6.55 \pm 0.7$	$6.71 \pm 0.89$
6 Forward	$6.82 \pm 0.54$	$6.33 \pm 0.82$	$5.99 \pm 0.39$	$6.14 \pm 0.55$	$6.08 \pm 0.54$
8 Forward	$7.12 \pm 0.84$	$6.17 \pm 0.71$	$6.47 \pm 1.07$	$6.38 \pm 1.0$	$6.03 \pm 0.7$

Table 3: Mean and standard deviation ( $\sigma$ ) of Mean Absolute Error ( $\bar{y}$ ) 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.

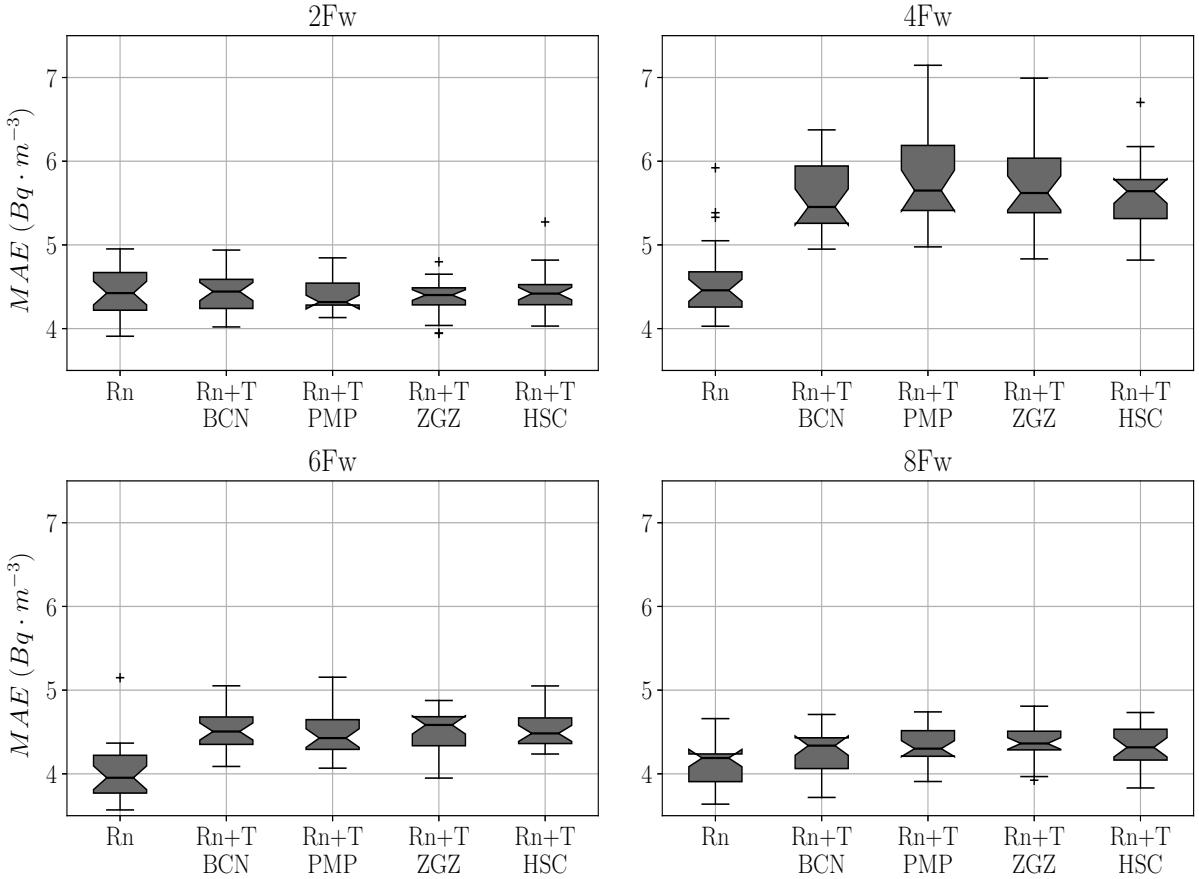


Figure 11: Mean Absolute Error Boxplots for  $N = 2, 4, 6, 8$  future values predicted in two cases: Rn as input of the LSTM and Rn + Temperature as input for each city.

	$^{222}Rn$	$(\bar{y} \pm \sigma)_{BCN}$	$(\bar{y} \pm \sigma)_{PMP}$	$(\bar{y} \pm \sigma)_{ZGZ}$	$(\bar{y} \pm \sigma)_{HSC}$
2 Forward	$4.57 \pm 0.44$	$5.56 \pm 0.41$	$5.8 \pm 0.63$	$5.75 \pm 0.54$	$5.63 \pm 0.42$
4 Forward	$4.00 \pm 0.33$	$4.50 \pm 0.22$	$4.48 \pm 0.28$	$4.51 \pm 0.25$	$4.54 \pm 0.24$
6 Forward	$4.12 \pm 0.24$	$4.28 \pm 0.26$	$4.35 \pm 0.22$	$4.39 \pm 0.24$	$4.33 \pm 0.25$
8 Forward	$4.42 \pm 0.28$	$4.42 \pm 0.23$	$4.40 \pm 0.18$	$4.37 \pm 0.21$	$4.44 \pm 0.27$

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

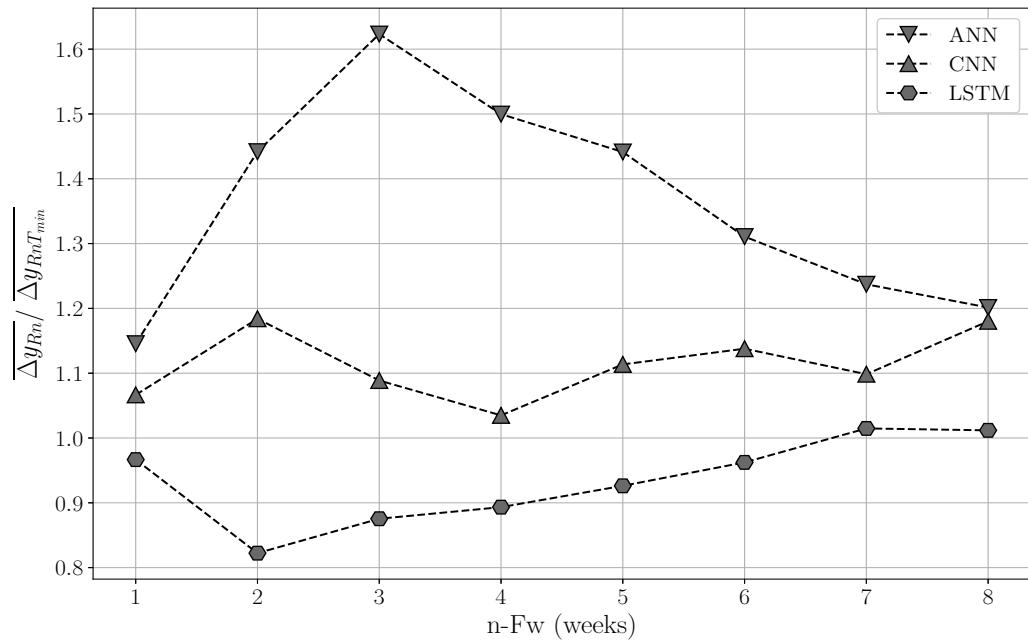


Figure 12: Mean Absolute Error Boxplots for  $N = 9$  future values predicted in two cases:  $R_n$  as input of the CNN and  $R_n + \text{Temperature}$  as input for each city.

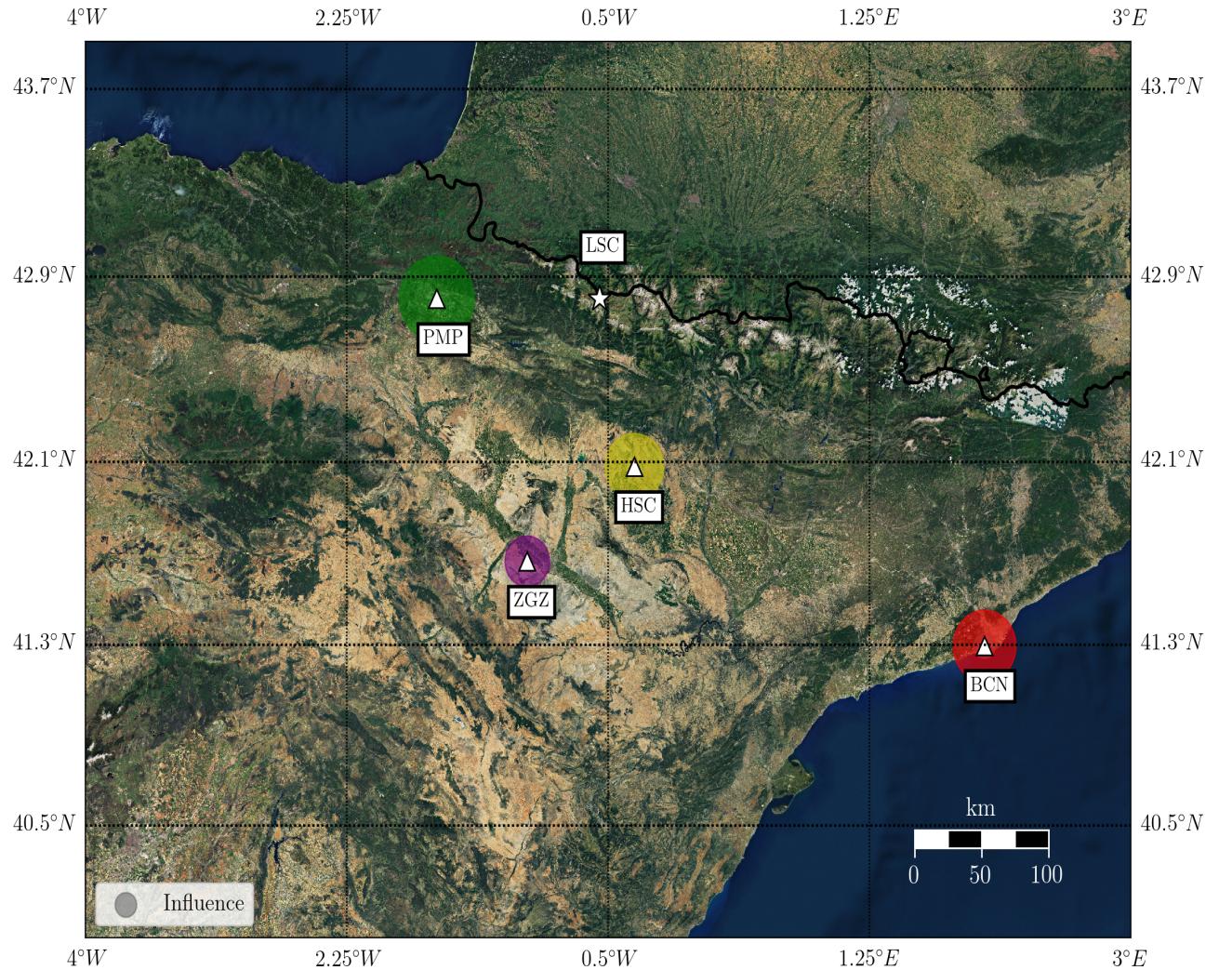


Figure 13: GPS map of the Spanish peninsula area of interest with influences represented by bubbles. The greater the bubbles are, the better the forecasting improvement is. (3 Forward case)

#### 4. Conclusions

[11] [12] [10] [9] [13] [14] [15]

#### Appendix A. Section in Appendix

#### 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, Variations on the concentration of  $^{222}\text{Rn}$  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. Hirsisiko, 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. Bhattacharai, 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 acquisition 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] A. Bettini, New underground laboratories: Europe, asia and the americas, *Physics of the Dark Universe* 4 (2014) 36 – 40, dARK TAUP2013. doi:<https://doi.org/10.1016/j.dark.2014.05.006>.
- [12] M. Cárdenas-Montes, I. Méndez-Jiménez, Ensemble deep learning for forecasting  $^{222}\text{Rn}$  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>.
- [13] R. Hyndman, G. Athanasopoulos, Forecasting: principles and practice., OTexts (2014).
- [14] I. Goodfellow, Y. Bengio, A. Courville, Deep learning., MIT Press, Massachusetts (2016).
- [15] Gamboa, J.C.B, Deep learning for time-series analysis., CoRR (2016) abs/1701.01887.