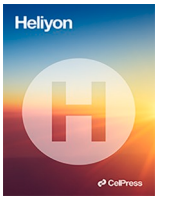
Heliyon 10 (2024) e30820

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Applying the Kolmogorov–Zurbenko filter followed by random forest models to ^{7}Be observations in Spain (2006–2021)

Ander Nafarrate [a], Susana Petisco-Ferrero [a,*], Raquel Idoeta [a], Margarita Herranz [a], Jon Sáenz [b,c], Alain Ulazia [d], Gabriel Ibarra-Berastegui [a,c]

[a] Energy Engineering Department, University of the Basque Country, UPV/EHU, Plaza Torres Quevedo, s/n, Bilbao, 48013, Spain
[b] Department of Physics, University of the Basque Country, UPV/EHU, Barrio Sarriena, s/n, Leioa, 48940, Spain
[c] Plentzia Itsas Estazioa (PIE), University of the Basque Country, UPV/EHU, Areatza Hiribidea 47, Plentzia, 48620, Spain
[d] Energy Engineering Department, University of the Basque Country (UPV/EHU), Otaola, Hiribidea, 29, Eibar, 20600, Spain

ARTICLE INFO

Keywords:
^{7}Be
Kolmogorov–Zurbenko filter
Random forests
Spain
Signal-noise decomposition
Fluid mechanics

ABSTRACT

In this study, we analysed ^{7}Be weekly surface measurements from six Spanish laboratories from 2006 to 2021. The Kolmogorov–Zurbenko filter was applied to the six ^{7}Be time series, and following an iterative process, the original data were divided into two fractions: one related to variations characterized by periods above 33 days (including, among others, the seasonal cycle) and the second noisier fraction related to mechanisms originating from variations with periods below 33 days. Both fractions were independent at the six locations. The second machine-based step using random forest models was applied with the aim of identifying the most influential inputs to the observed ^{7}Be concentrations, and machine learning-inspired regression models were fitted. With respect to seasonal components, the results indicated that the memory of the system was the most influential input, as expected by the large fraction of variance explained by the seasonal cycle, followed by that of humidity and wind-related variables. For the fraction corresponding to periods below 33 d, precipitation-, humidity-, and radiation-related variables were the most influential. This methodology has made it possible to successfully describe the major mechanisms known to be involved in the generation of the surface ^{7}Be concentrations observed in Spain.

1. Introduction

Since the discovery of ^{7}Be in 1955 [1], extensive research has been conducted to understand its formation and transport mechanisms [2]. The associated results show that it is a naturally occurring gamma-emitting radionuclide produced through the spallation processes of galactic cosmic rays and solar energetic particles [3,4] hitting light atomic nuclei (nitrogen and oxygen). With a half-life $T_{1/2}$ of 53.22 days [5], it is prevailingly formed in the stratosphere (~2/3) and the remainder (~1/3) is produced in the upper

Abbreviations: C3S, Copernicus Climate Change Service; CDS, Climate Data Store; CSN, Consejo de Seguridad Nuclear (Spanish Nuclear Safety Council); ECMWF, European Centre of Medium-Range Weather Forecasts; EREMSN, European Radioactivity Environmental Monitoring Sparse Network; HPGe, High Purity Germanium; IFS, Integrated Forecasting System; IMS, International Monitoring System; KZ, Kolmogorov–Zurbenko filter; RF, Random Forests technique; ROB, Royal Observatory of Belgium; SSNES, Spanish Sparse Network for Environmental Surveillance.

* Corresponding author.
E-mail address: susana.petisco@ehu.eus (S. Petisco-Ferrero).

https://doi.org/10.1016/j.heliyon.2024.e30820
Received 29 September 2023; Received in revised form 18 April 2024; Accepted 6 May 2024
Available online 7 May 2024

troposphere [6–10], with estimated global average production rates of 0.041 and 0.027 atoms $cm^{-2}s^{-1}$, respectively, for the solar minimum [11]. Once formed, the nuclei are quickly oxidized; they then attach themselves to accumulation mode aerosol particles (<1 μm) [12–14] and follow their transport in the atmosphere [15]. Nevertheless, the most recent studies on health effects associated to particulate matter [16] just ignore any additional or cumulative impact of the ^{7}Be attached. ^{7}Be provides integrated information on the both production areas and transport [17] on a short time scale [4], and it has thus become a proxy that enables quantitative descriptions of fundamental dynamic atmospheric processes [18]. Moreover, long-term variations in ^{7}Be have recently been used as tracers of climate change [19] and global warming [20].

The observed airborne ^{7}Be activity concentration is the result of a balance between sources that increase the activity concentration and sinks that have the opposite effect [21]. Accordingly, among the sources of ^{7}Be, the inverse correlation between sunspot number and an increased radiative budget has been widely documented in literature [7,18,22–28]. Furthermore, tropopause height [29], altitude [11,17,18,30], decreasing latitudes [23,24,27] and sudden intrusions of ^{7}Be from the upper atmosphere [31] have been found to play similar roles. However, natural radioactive decay, wind scattering and precipitation washout from the air are mechanisms that reduce ^{7}Be concentrations at the local level [32–34]. Due to its close correlation with precipitation, other variables, such as relative humidity, typically appear to be inversely correlated with 7Be concentrations [26,35–37]. This inverse relationship has also been reported in remote locations, such as the Arctic [38]. In the particular case of Spain, several studies conducted at a local level [28] and within a limited time frame [39] have also highlighted this negative correlation between ^{7}Be and precipitation related variables.

In this respect, finding a model that can relate the observed airborne ^{7}Be activity concentration to local meteorological conditions is the aim of environmental modelling using radionuclides as tracers [40–42]. As such, a variety of techniques such as multiple linear regression [27], factor regression [22], time series analysis [3], Fourier analysis [7] and regression trees [43] have been used to determine relationships or analyse individual correlations between surface ^{7}Be concentrations and meteorological variables. Models based on machine-learning algorithms, such as random forest (RF) [25,44], have also succeeded in explaining more than 70 % of the overall variability in ^{7}Be surface concentrations using meteorological inputs. In the case of Spain, it is worth mentioning a previous study based on artificial neural networks to predict ^{7}Be values at a monthly time scale [45].

Considering the above, the current study was conducted with the aim of providing, for the first time, a complete picture of the ^{7}Be field throughout Spain corresponding to the 2006–2021 period. For consistency, the same methodology was applied to the data from six laboratories that comprise the Spanish Sparse Network for Environmental Surveillance [46], and the procedure employed was as follows: the Kolmogorov–Zurbenko (KZ) filter was used to identify the most relevant periodicities involved in the generation of ^{7}Be in Spain, and RF was used to determine the relationship between ^{7}Be observations and meteorological variables for the broadband signals obtained in the former step.

This paper is organised as follows: Section 2 presents the data and methodology, including a description of the ^{7}Be measurement stations in Spain and the parameters selected from ERA5 reanalysis. The KZ filter (Section 2.2.1) and random forest (Section 2.2.2) are also introduced. The results are presented in Section 3, and Sections 4 and 5 present the main results and discuss the relevant conclusions, respectively.

2. Material and methods

2.1. Data

2.1.1. ^{7}Be data and study area

The activity concentration of ^{7}Be in airborne particulates is routinely measured on a weekly basis by the European Radioactivity Environmental Monitoring Sparse Network, which conducts high-sensitivity measurements. In Spain, six surveillance stations comprise the Spanish Sparse Network for Environmental Surveillance under the commitment of the Spanish Nuclear Safety Council (CSN). These six stations are spread along the national territory including the Iberian Peninsula and the Canary Islands, and they are associated with the following institutions indicated in decreasing latitudinal order: Universidad del País Vasco (UPV_1), Bilbao; Universidad Politécnica de Cataluña (UPC_1), Barcelona; CIEMAT (ICI_28), Madrid; Universidad de Extremadura (UCC_1), Cáceres; Universidad de Sevilla (USE_1), Sevilla and Universidad de La Laguna (ULL_3), Santa Cruz de Tenerife. It is of note that ^{7}Be data from UPV_1 were obtained directly from measurements conducted at the Bilbao station and that these (together with those from other stations that comprise the Spanish Sparse Network corresponding to the 2006–2021 period) are publicly available and can be downloaded from the website maintained by the CSN (https://www.csn.es/kprgisweb2/index.html?lang=es).

Table 1
2006–2021 ^{7}Be activity concentration averages, coordinates of laboratories, and ERA5 nearest grid points used to retrieve meteorological variables. (*) Only available from July 2008 onwards.

Lab. name	Average ± SD [mBq/m^3]	Longitude °E	Latitude °N	LonERA5 °E	LatERA5 °N
upv_1	3.30 ± 1.41	−2.94955	43.26136	−3	43.25
upc_1	3.70 ± 1.4	2.11606	41.38384	2	41.5
ICI_28	4.30 ± 1.55	−3.72454	40.45647	−3.75	40.5
ucc_1 (*)	4.60 ± 1.63	−6.34887	39.48154	−6.25	39.5
use_1	4.10 ± 1.47	−5.98677	37.35919	−6	37.25
ULL_3	4.50 ± 1.33	−16.29109	28.45667	−16.25	28.5

Table 1 summarizes the ^{7}Be activity concentration averages and corresponding standard deviations measured within the 2006–2021 period at each laboratory, as well as the associated coordinates. In the particular case of the UCC_1 laboratory, weekly ^{7}Be data were only available from July 2008 onwards. For ease of visualization, laboratory locations are displayed in Fig. 1 together with the average ^{7}Be concentrations.

After a brief examination of the averaged values of the six locations in Table 1, a value for the entire Spanish territory of 4.00 ± 0.50 mBq^{-3}can be obtained. However, differences between stations are apparent from the standard deviations.

2.1.2. Measurement stations

Aerosol samples are obtained using a high-volume aerosol sampling station ASS-500 device with a nominal flow rate of 500 m^{3} h^{-1}. The unit contains a polypropylene filter (44 × 44 cm^{2}) that is replaced weekly. Details of the sample handling are described in another study [35]. The dry weight of the dust trapped by the filter is determined from the difference in the weight of the filter before and after exposure. The ^{7}Be activity concentration is measured using high purity Germanium (HPGe) detectors with a counting time longer than two days. The relative detector efficiencies are in the range of 40–90 %, and at a resolution between 1.77 and 1.92 keV at 1.33 MeV of ^{60}Co. The background measurements are obtained using blank filters. The mean measurement uncertainty (coverage factor k = 2) for ^{7}Be is approximately 6 % at all six stations. It must be mentioned that before submitting observations to the CSN and being then made public, a quality control of data is carried out at each laboratory with strict data validation standards (ENAC ISO 17025) [47]. As a result, a very high percentage of reliable weekly observations (>99 %) have been available for the six laboratories.

2.1.3. Sunspot number

Previous studies worldwide, and particularly those on surface 7Be observations conducted in Northern Spain [28,35], indicate that the sunspot number is a proxy for the upper-atmosphere and is negatively correlated with the ^{7}Be concentrations observed. Therefore, the daily total sunspot number (*SSN*) over the entire solar disk for 2006–2021 was downloaded from the Royal Observatory of Belgium [48] and incorporated into the analysis.

2.1.4. ERA5 reanalysis

ERA5 is the 5th generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis of

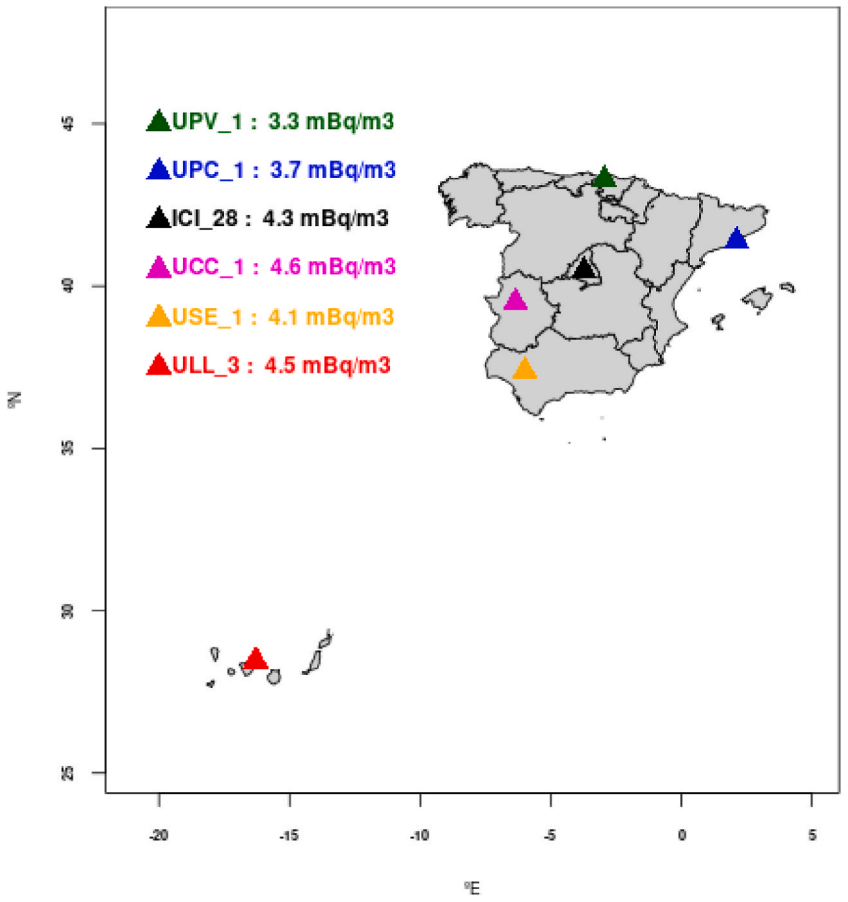


Fig. 1. Location of laboratories and ^{7}Be activity concentration average corresponding to the 2006–2021 period.

global climate product, and it covers the period January 1940 to the present [49]. It was produced by the Copernicus Climate Change Service (C3S) at the ECMWF [50] and it provides hourly estimates of numerous atmospheric, land, and oceanic climate variables. The data cover the Earth on a 30 km grid and resolve the atmosphere using 137 levels from the surface up to a height of 80 km. ERA5 includes information regarding the uncertainties of all variables at reduced spatial and temporal resolutions. It is based on 4D-Var data assimilation using Cycle 41r2 of the Integrated Forecasting System (IFS), which was operational at the ECMWF in 2016. Compared with its predecessor, ERA-Interim [51], ERA5 has a significantly enhanced horizontal resolution (31 km grid spacing vs. 79 km for ERA-Interim), and it benefits from a decade of development in model physics, core dynamics, and data assimilation. A complete set of hourly ERA5 data is available in the C3S Climate Data Store (CDS) [52], and the primary properties are listed in Table 2.

To model the correlation between measured ^{7}Be activity concentrations and meteorological observations, surface and upper atmosphere variables are considered. In addition, shortwave radiation routines used by the ECMWF in ERA5 [53] consider the interannual variability of solar activity using yearly varying values of Total Solar Irradiance.

For each laboratory, the following 13 daily meteorological variables at the nearest ERA5 grid point (Table 1) were selected for analysis:

- Surface variables
  1 Surface pressure sp (Pa)
  2 10 m zonal wind component *u10* (m s^{-1})
  3 10 m meridional wind component *v10* (m s^{-1})
  4 2 m dewpoint temperature *d2m* (K)
  5 2 m temperature *t2m* (K)
  6 Downward UV radiation at the surface *uvb* (J m^{-2})
  7 Top net solar radiation *tsr* (J m^{-2})
  8 Evaporation *e* (m of water equivalent)
  9 Total column rain water *tcrw* (kg m^{-2})
  10 Total precipitation *tp* (kg m^{-2})
  11 Total cloud cover *tcc* (−) [0 − 1]
- Upper atmosphere variables (500 hPa)
  1 Ozone mass mixing ratio O_3 (kg kg^{-1}). The reason for including ozone levels at such a height is that intrusions from the stratosphere can be detected if the concentration of ozone rises. It can be expected that in this case, ^{7}Be can also be transported downwards and then be detected on the surface.
  2 Vertical velocity ω (Pa s^{-1}). This parameter measures the downward velocity from the upper atmosphere. Since ^{7}Be is generated there, positive values might be associated to surface observations due to transport from the upper layers of the atmosphere.

2.2. Methodology

2.2.1. Kolmogorov–Zurbenko filter

The KZ filter is a nonparametric smoothing technique used to detect distinct changes in a highly noisy time series. It belongs to a class of low-pass filters that can be used with missing values. The KZ filter has two parameters: m, a positive odd integer which represents the length of the moving average window, and p, the number of iterations of the moving average obtained according to Equation (1),

$$Y_i = \frac{1}{m} \sum_{j=-\frac{(m-1)}{2}}^{j=\frac{(m-1)}{2}} X_{i+j} \tag{1}$$

As in regular moving averages, in this equation X represents any original time series of any type. The i index corresponds to the position in the original vector X. j is the number of positions before and after position i that are used to calculate the moving average. Y is the resulting time series after applying equation (1). The same equation (1) can be applied p times to the resulting Y time series in an iterative manner.

KZ filters are used in various fields, including those relating to atmospheric pollutants [54–56]. In the particular case of ^{7}Be, they

Table 2
ERA5 dataset properties.

Property	Value
Temporal resolution	1-hourly
Spatial resolution	30 km
Vertical resolution	137 levels up to 80 km
Coverage	Global
Time period	January 1940 to present
Produced by	Copernicus Climate Change Service (C3S) at ECMWF
Data variables	Atmospheric, land, and oceanic climate variables
Data access	C3S Climate Data Store (CDS)

have been used to smooth the original data and obtain a superior graphical representation of observations [25], and to detect the synoptic fraction of ^{7}Be at a referential station [18] at Mt Cimone (Italy, 2165 *m.a.s.l.*). This filter can be used to identify trends or changes in data that may be difficult to detect because of noise. Furthermore, the KZ filter can be used to separate the contributions of cycles below and above a given threshold periodicity according to Equation (2) [56] as follows,

$$mp^{0.5} \leq N \tag{2}$$

In this respect, the window length m and number of iterations p of the filter can be modified to determine the cut-off period N for what is passed through the filter.

Following previous studies conducted by the authors with other atmospheric elements [56], the KZ filter was applied in the current study to split the original data from the six Spanish laboratories (Fig. 1) into a filtered time series and random residuals associated with the noisy fraction of the original measurements. To achieve this within the framework of an iterative process with many cut-off periods obtained from different combinations of m and p, the KZ filter was applied to logarithmic data from six laboratories according to standard methodology [54–56]. In all laboratories, only a small fraction <1 % of missing ^{7}Be weekly values were present. In these cases, the standard KZ methodology was applied, and the moving averages (Equation (1)) were calculated using the available data.

Original observations can be understood as the sum of apportions associated to different periodicities. Under this approach, by selecting the m and p parameters (Equation (2)), the KZ filter allows splitting the original observations into:

1. A filtered fraction associated to the contributions of the rest of the periodicities above a given threshold N given by Equation (2).
2. The residuals, associated to the contributions of the rest of periodicities not included in the filtered fraction.

By simply adding the filtered fraction and the residuals, the original time series of the observations can be reconstructed.

The objective of this work was to obtain for each laboratory a division into two meaningful fractions, each associated to two different groups of independent mechanisms and periodicities. An indicator that this type of partition has been achieved and an optimal cut-off period has been identified, had to be that the filtered and residual fractions were statistically independent. To achieve this, in the frame of an iterative process, for each tentative combination of m, p and N the residuals and filtered series were tested to check whether the following two conditions were met:

1. The two fractions, the residuals, and the filtered time series were independent. For this purpose, the correlations and covariances between the two fractions were calculated. Correlations near zero (below 0.05) values for the covariances below 5 % of the overall variance were clear indicators of independence between both fractions.
2. The residuals were Gaussian with a zero average, as expected for the fractions corresponding to noise. Or put in other words, the net apportion during the period analysed of a great number of high-frequency, rapidly varying mechanisms is zero. In this study, the Pearson chi-square normality test [57,58] was applied to the residuals, and their averages were computed.

For the ease of interpretation of the conclusions derived from the statistical indicators analysed in this paper, interested readers are referred to the work by Crawley [59], where the basics of correlation coefficients, (co)variances, normality tests averages and other statistical tools are explained in detail.

2.2.2. Random forest

Random forest (RF) is a popular machine-learning algorithm used for both classification and regression problems. It is based on regression trees and comprises many trees that operate as an ensemble, and the algorithm uses bagging and feature randomness to construct regression trees. RF is typically used in regression problems where highly nonlinear relationships between a given output and a set of candidate inputs can be expected. An important aspect of RF is that it computes the deterioration in the mean square error (*MSE*) associated with the removal of each input. This technique therefore yields an objective quantitation of the relative importance of each input (in percentage terms) for predicting the output [60]. The mathematical aspects of RF have been extensively described [61–63]. In this study, RF models were applied to each dataset obtained from the six laboratories.

With the same aim of relating ^{7}Be observations to meteorological variables, Długosz et al. [22] recently applied a factor regression to each year of their study in central Poland. Seasonal RF models were fitted to the entire dataset in Poland using weekly averages in other studies [25,44]; the results were used to identify the most influential meteorological variables on ^{7}Be, and they also focused on analysing their individual correlations with ^{7}Be concentrations. In addition, Ioannidou et al. [27] built a multiple regression model (at a daily scale) between ^{7}Be observations and meteorological variables, and they were able to explain about 40 % of the variance using this.

Returning to the aim of the current study, which was to build a regression model for each Spanish laboratory between ^{7}Be, sunspots, and meteorological variables at their nearest locations, RF was selected because highly nonlinear relationships were expected. To implement the RF algorithm, it is necessary to remember that models need to be built at a compatible time scale between inputs and outputs. However, this was not the case for the weekly sampling rate of ^{7}Be and the time scales of ERA5 meteorological variables, which ranged from hourly to daily. To overcome this difficulty, several options were considered, and many tests conducted to obtain the optimal solution that maximised the explanatory power of the RF models. Finally, we attributed the accumulated weekly ^{7}Be concentrations to the midpoint of the week and utilised the daily inputs of the 13 variables (see Section 2.1.4) corresponding to that specific day at the nearest grid point from the ^{7}Be measuring laboratories (Table 1). During this initial process, it was observed that the

^{7}Be concentrations retained up to two weeks memory from past observations of the ^{7}Be concentrations, upper-atmosphere variables (O_3 and ω) and sunspot numbers, in particular. Taking this observation into consideration, these delayed variables were also incorporated into the RF models resulting in a total of 22 variables as candidate inputs to explain the ^{7}Be concentrations at each laboratory (i.e. 13 + 1 (daily sunspot number) + 2 × 4 (one and two weeks' memory of: ^{7}Be concentrations, O_3, ω and sunspot numbers)).

A second major observation from the preliminary study was that the explanatory power of random forests was maximised when applied separately to the two independent fractions obtained from the KZ filter. Thus, a total of 6 × 2 = 12 RF models were fitted independently.

To gain further insights into the quality of the fitting model, different sets of data were used for 1) training the RF models and 2) testing the RF models. In this respect, data corresponding to the entire 2006–2021 period were divided into two parts: the first half to train the RF models (years 2006–2013: 400 weeks) and the second half (years 2014–2021: 400 weeks) to test the RF models. In doing so, possible bias associated with using the same dataset for both the training and fitting of the RF models was avoided, and the performance of the quality of fit of the models could be assessed.

Previous studies have successfully applied RF and its associated methodology to regression problems when using other geophysical variables [64,65]. Following such previous experience and inspired by the RF standard procedures in regression applications [61], the number of selected variables at each splitting node was seven, representing roughly one third of the number of inputs ($7 \approx 22/3$). Each RF model was built using 1000 regression trees.

3. Results

3.1. KZ filter

The KZ filter was applied to the ^{7}Be time series for the 2006–2021 period of each laboratory, and the initial datasets were split into two fractions: one corresponding to the filtered time series and the other to the residuals. It is of note that for the combination of $m = 33$ and $p = 1$, the obtained fractions met the two conditions mentioned in Section 2.2.1. This meant that for a cut-off periodicity of 33 days (Equation (1)), the two fractions turned out to be independent and the residuals obtained were Gaussian with average zero. The final statistical indicators of the time series for the six laboratories are summarized in Table 3: the correlation between the filtered time series and the residuals are listed in column 2, and columns 3 and 4 show the proportions of the original variance represented by the filtered time series and the residuals, respectively. For the filtered time series, the corresponding proportion values roughly range between 47 % and 74 %, whereas the values for the residuals span between 26 % and 52 %. The sum of columns 3 and 4 is displayed in column 5, and this shows a total of nearly one for each pair of fractions, which suggests that the covariances are negligible. These observations show that the filtered time series and residuals are independent for the six laboratories. Column 6 shows that Pearson Chi-squared results are $p > 0.05$ for all cases, which implies that the residuals are Gaussian for all laboratories (at the 95 % confidence level). Finally, column 7 shows that the averages of the residuals are zero, indicating that their overall contribution is null.

To summarize this idea, the 33 day cut-off period was identified under the assumption that the residuals obtained should be Gaussian with 0 average and independent from the KZ-filtered time series. As mentioned above, the idea was to split the driving mechanisms (KZ filtered time series, associated to periodicities above the 33 day threshold) and the additional low impact factors with a null contribution in the long-term (associated to periodicities below the same threshold). To that purpose, in the framework of an iterative loop several cut-off periods were tested and the statistical indicators associated to each solution were computed. Only the 33 day period solution met those conditions for the residuals (correlation near 0, covariances near 0, Gaussian residuals with average almost 0) as can be seen in Table 3.

In consideration of all the results presented above, the following assumptions were made: the cut-off period of 33 d clearly divided the ^{7}Be time series of the original signal into two fractions. As such, two independent contributions were obtained, one corresponding to cycles with periodicities above 33 d (filtered fraction) and the other (residuals) with periodicities below 33 d (Fig. 2). A statistical analysis indicated that the two fractions at all the laboratories were independent, thus confirming the different nature of the mechanisms involved.

However, it is also of note that a different approach was used by Tositti et al. [18], who employed a KZ filter with $m = 21$ and $p = 3$ to split the time series into contributions associated with seasonality and synoptically influenced deviations from the seasonal cycle, respectively. Although the authors did not provide further information about the variability associated with each fraction, and they did not specifically calculate the cut-off frequency, by substituting into Equation (2), a value of 36 days can be found. Despite the singular location of Mt. Cimone and the differences in the sampling rates employed (48 h at Mt. Cimone vs. 1 week employed in this study), the

Table 3
Statistical indicators of the KZ filtered time series and residuals with m = 33 and p = 1 for 2006–2021.

Lab. name	Corr. KZ & Resid	Variance proportion KZ.	Variance proportion Resid	Variance proportion KZ + Resid.	Pearson ChiSquare Resid. pvalue	Resid. Average
UPV_1	0.00	0.55	0.49	1.04	0.148	0.00
UPC_1	0.03	0.55	0.43	0.98	0.055	0.00
ICI_28	0.04	0.62	0.35	0.97	0.162	0.00
UCC_1	0.01	0.74	0.26	1.01	0.065	0.00
USE_1	0.04	0.47	0.52	0.99	0.078	0.00
ULL_3	0.04	0.51	0.47	0.98	0.246	0.00

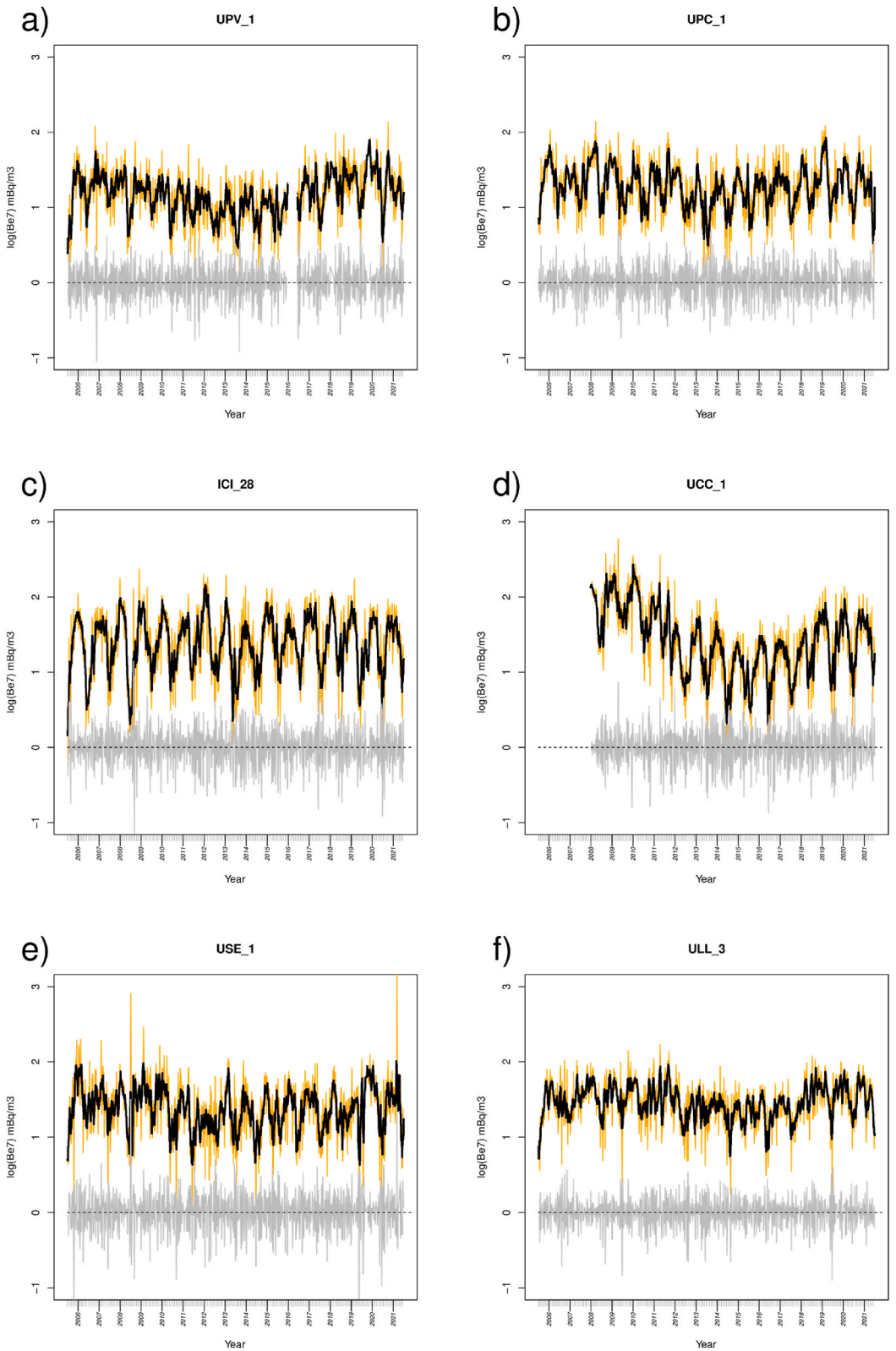


Fig. 2. 2006–2021 ^{7}Be concentrations at the six Spanish laboratories (orange): a) UPV_1, b) UPC_1, c) ICI_28, d) UCC_1, e) USE_1, f) ULL-1. The KZ filtered time series are depicted in black. Residuals are depicted in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 4
RF model performance for the test dataset and most influential inputs (2014–2021) (KZ filtered fractions).

Lab	R^2	1st	ΔMSE [%]	2nd	ΔMSE [%]	3rd	ΔMSE [%]	4th	ΔMSE [%]	5th	ΔMSE [%]	6th	ΔMSE [%]
UPV_1	0.87	*^{7}Be_kz-1*	74.43	*tcrw_kz*	51.64	*ω_kz_2*	51.11	*sp_kz*	46.48	*tp_kz*	44.22	*ω_kz*	42.61
UPC_1	0.88	*^{7}Be_kz-1*	75.82	*ω_kz*	46.41	*tcrw_kz*	46.32	*sp_kz*	45.7	*ω_kz_2*	42.87	*tp_kz*	40.83
ICI_28	0.90	*^{7}Be_kz-1*	53.43	*sp_kz*	48.85	*u10_kz*	46.37	*e_kz*	43.8	*tp_kz*	42.94	*SSN_kz_1*	42.16
UCC_1	0.87	*^{7}Be_kz-1*	55.16	*d2m_kz*	39.65	*u10_kz*	36.83	*e_kz*	34.9	*ω_kz*	33.39	*^{7}Be_kz_2*	32.97
USE_1	0.79	*^{7}Be_kz-1*	81.44	*e_kz*	54.42	*tcrw_kz*	48.43	*u10_kz*	48.33	*sp_kz*	43.01	*ω_kz*	42.28
ULL_3	0.85	*^{7}Be_kz-1*	101.34	*tcc_kz*	46.43	*t2m_kz*	45.86	*v10_kz*	43.56	*tp_kz*	42.49	*d2m_kz*	42.07

Table 5
RF model performance for the test dataset (2014–2021) and most influential inputs (Residuals).

Lab	R^2	1st	ΔMSE [%]	2nd	ΔMSE [%]	3rd	ΔMSE [%]	4th	ΔMSE [%]	5th	ΔMSE [%]	6th	ΔMSE [%]
UPV_1	0.32	*tp_res*	14.28	*ω_res*	12.36	*v10_res*	10.48	*uvb_res*	9.8	*SSN_res*	9.26	*^{7}Be_res_1*	8.8
UPC_1	0.19	*sp_res*	17.43	*^{7}Be_res_2*	9.48	*v10_res*	8.58	*tsr_res*	8.36	*ω_res*	7.45	*t2m_res*	8.4
ICI_28	0.32	*tp_res*	36.29	*tcrw_res*	26.48	*ω_res*	20.97	*sp_res*	18.69	*tsr_res*	17.54	*ω_res_1*	16.7
UCC_1	0.24	*tp_res*	25.83	*SSN_res_1*	29.42	*SSN_res_2*	23.6	*SSN_res*	20.99	*tcrw_res*	15.28	*sp_res*	14
USE_1	0.21	*tcrw_res*	20.21	*tp_res*	14.63	*tsr_res*	10.98	*uvb_res*	10.21	*tcc_res*	9.96	*^{7}Be_res_1*	7.9
ULL_3	0.17	*tp_re*	25.52	*v10_res*	13.69	*tsr_res*	11.11	*e_res*	10.31	*uvb_res*	9.53	*tcc_res*	17.4

same cut-off frequency was obtained in both studies.

To interpret the fraction representing the residuals in Table 3 in terms of the variables from their Gaussian distribution at all locations, a diversity of random factors is called. Some of them are of synoptic origin, according to Ref. [18], together with others such as precipitation or advection to/from other locations. Furthermore, as their combined net contribution to local ^{7}Be averages is zero, none of these factors can be said to be dominant in the long term. However, as with other atmospheric elements like pollen [66] further research is needed in order to elucidate the specific synoptic mechanisms involved.

The raw data of the ^{7}Be activity concentrations, KZ-filtered time series, and residuals for the six Spanish laboratories are shown in Fig. 2. From these plots, a clear seasonality can be distinguished from the KZ-filtered fractions in the six laboratories, with the maxima in summer and the minima in winter. However, in addition to this, the filtered fractions feature persistent differences among the six locations, as indicated by the averaged ^{7}Be activity concentrations and the corresponding standard deviations shown in Table 1. Therefore, as highlighted in previous studies, it is essential to note the importance of local conditions [24]. In this respect, one study [20] analysed daily data sets from 70 global IMS monitoring stations covering almost 20 years (from September 2003 to September 2019) and compared the annual average ^{7}Be concentrations with the annual average cosmic rays. The results indicated that only 18 out of the 62 stations showed significant positive correlations ($\alpha = 0.05$) for the years 2004–2018. In the other stations (37) either weakly positive results or negative (7 stations) correlations were noted but results were not significant at a 95 % confidence level. At these stations, the authors concluded that the effect of cosmic rays on the activity concentration of ^{7}Be in the air was overprinted by atmospheric processes as well as by the mixing of old and fresh air masses during transport.

3.2. Random forest

An improved understanding of the most influential variables on the observed ^{7}Be activity concentrations was obtained by applying the RF models. Based on the observed maximised explanatory power of the RF models mentioned in Section 2.2.2., fitting was conducted separately for each resulting fraction after applying the $KZ_{33,1}$ filter to the 22 candidate input variables in two steps. The training step was accomplished by fitting the models for both fractions of the data corresponding to the 2006–2013 period, and their performances were then verified when the trained models from the previous step were fed data from the 2014–2021 period. The results for each of the six laboratories are summarized in Tables 4 and 5, where R^2 (the square of the correlation coefficient) represents the fraction of the overall variability in the ^{7}Be concentrations explained by the RF models. The relevance of each input is described by the mean square error increase (ΔMSE, in percentage terms) as it represents the ratio relating to the effect of subtracting the aforesaid input on the overall RF model. Accordingly, the six most influential variables can be depicted from their associated ΔMSE values in Tables 4 and 5.

The correspondence between the RF models and the variances of the $KZ_{33,1}$ filtered fractions of ^{7}Be ranged between 47% and 74 % (Table 3) and are shown in Table 4. The high R^2 values registered in all laboratories indicate that between 80 % and 90 % of the observed ^{7}Be filtered fractions can be explained by the $KZ_{33,1}$ filtered fractions of the inputs. In all the cases, the filtered memory of the ^{7}Be concentration is the most important input. Meteorological variables with clear seasonal behaviour (*t2m_kz*, *u10_kz and v10_kz*), and those more closely related to the yearly water cycle (*tcrw_kz*, *tcc_kz*, *d2m_kz*, *e_kz and sp_kz*) are the most influential. In all cases, the averages of the residuals are zero (Table 3), which explains the clear seasonality exhibited by the ^{7}Be concentrations (driven by their $KZ_{33,1}$ filtered fractions) at the six laboratories (Fig. 2).

Notably, the filtered or seasonal component of the upper atmosphere variable ω was also an important input at 4 laboratories. In this respect, ω represents the downward vertical velocity from the upper atmosphere, and it was only analysed at the nearest vertical location from each laboratory. ω_kz can be understood as a (probably imperfect) proxy of the seasonal mechanisms involved in ^{7}Be transportation from the upper atmosphere. The influence of $KZ_{33,1}$ filtered fraction also featured memory (Table 4). The changes in SSN with periodicities above 33 days (*SSN_kz*) appeared to only be relevant in one laboratory (ICI_28), whereas their influence at the other laboratories was not among the most important (Table 4) ^{7}Be $KZ_{33,1}$ filtered fraction inputs.

The RF models fitted with the residuals obtained after applying the $KZ_{33,1}$ filter (Table 5) could only explain between 17 % and 32 % of the variability observed in this fraction of ^{7}Be. The R^2 values were far smaller than those of the $KZ_{33,1}$ fractions because the residuals were comparatively noisy and associated with greater rapidly changing ^{7}Be concentrations with cycles below 33 days (Table 5). Additionally, at the six labels, the residuals follow a Gaussian distribution with zero average (Table 3); therefore, a high number of small impact effects can be expected to act jointly without a clear dominant mechanism over the rest.

However, when the small part of the variability explained by the RF models was analysed, two major groups of variables were found to be the most important at all laboratories: humidity or precipitation-related variables (*tp_res*, *tcc_res*, *tcwr_res*, *sp_res* and *e_res*)

Table 6
Proportion of overall variability of ^{7}Be explained by RF models.

Lab	Variance proportion KZ	Variance proportion Resid.	Variance proportion RF KZ	Variance proportion RF Resid.	Combined explained variance
UPV_1	0.55	0.49	0.87	0.32	0.64
UPC_1	0.55	0.43	0.88	0.19	0.56
ICI_28	0.62	0.35	0.90	0.32	0.67
UCC_1	0.74	0.26	0.87	0.24	0.70
USE_1	0.47	0.52	0.79	0.21	0.48
ULL_3	0.51	0.47	0.85	0.17	0.51

and radiation-related variables (*SSN_res*, *tsr_res* and *uvb_res*). The first group of variables is associated with the ^{7}Be washing-out mechanisms, and the second group is known to be related to ^{7}Be generation mechanisms in the upper atmosphere. Specifically, SSN played an important role at one laboratory (UCC_1). Surface and downward wind speed (*v10_res* and *ω_res*) were also influential because of ^{7}Be transportation to/from other locations or even from the upper atmospheric layers. Again, the ^{7}Be concentration memory also played an explanatory role in this fraction.

As the filtered and residual fractions were independent (Table 3), the R^2 values of the RF models for both fractions (Tables 4 and 5) were combined (Eq. (3)) to obtain the total amount of variance that can be explained by this approach, and the results are shown in the final column of Table 6. The total proportion of the variability that can be explained by this combined approach ranged from 48 % (USE_1) to 70 % (UCC_1).

$$Variance_{explained} = Var_{KZ} \bullet R^2_{RF_{KZ}} + Var_{Res} \bullet R^2_{RF_{Res}} \quad (3)$$

In this equation, "$Variance_{explained}$" represents the overall variability explained by the combined models. $Var_{KZ/res}$ is the proportion of the variability explained by the two fractions (KZ filtered/residuals). $R^2_{KZ/res}$ is the proportion of the variability explained by the two random forests models for each of the two fractions (KZ/res). This applies to all the locations.

4. Discussion

This current study analysed weekly sampling data of ^{7}Be activity concentrations from six laboratories in Spain that comprise the Spanish Sparse Network for Environmental Surveillance.

Local surface ^{7}Be concentrations alter because of sudden upper-atmosphere intrusions or washing-out from the atmosphere due to precipitation or advection. It is difficult to capture these rapidly changing mechanisms and determine the relationship between ^{7}Be and variables that explain the observed surface concentrations throughout Spain when using only weekly averages.

To overcome this problem, we first employed a KZ filter. The results indicated that the original ^{7}Be time series from the six laboratories could be divided into two types of fractions: filtered fractions associated with seasonal effects with cycles above 33 days, and residuals originating from mechanisms with periodicities below 33 days. At all the laboratories, the two fractions were independent thus indicating two groups of different driving mechanisms acting on the observed ^{7}Be levels.

Except for the similar cut-off frequency previously reported for Mt. Cimone (Italy) [18], no other studies have focused on this division. However, although this division was found to be common based on data from all six Spanish laboratories, it needs to be confirmed in other locations. If it can be mathematically determined in a wide variety of environments, it may well represent an intrinsic natural division of ^{7}Be measurements between seasonal effects and the large numbers of mechanisms that act together, but which provide zero contribution in the long-term.

In the second step, RF, a machine-learning technique with a high capacity to capture nonlinear relationships, was used to fit regression models between ^{7}Be and the inputs. RF models were independently fitted to six laboratories for the two fractions obtained after applying the $KZ_{33,1}$ filter. Using this combined approach, the explanatory power of the RF models ranged from 48 % to 70 %. It must be emphasised that, unlike many previous studies, these results corresponded to the performance of the models fed with data belonging to the 2014–2021 period, which differed from the data used to fit them. As such, this provides a reliable indication of the actual performance and the inputs involved.

As in most previous studies, the sunspot number and surface meteorological variables at the nearest locations were used as inputs. However, to the best of our knowledge, this is the first time that i) the memory of the system and ii) upper-atmosphere meteorological variables have been incorporated into the analysis.

The memory of the system was among the most influential inputs in the seasonal fraction, whereas short-term changes in radiation, wind speed, and precipitation were captured by the residuals and corresponded to mechanisms with periodicities below 33 days.

Therefore, it would be possible to ascertain the impact of past observations on the current ^{7}Be values if the sampling rate was increased over a sufficiently long period.

The relevance that the group of precipitation-related variables play confirms that the mechanisms largely identified in the literature (solar radiation, washing out, advection) by other observational means appear in this study.

In the case of the upper-meteorological (500 hPa) variables selected here, ω, (the vertical downwards velocity) which describes the ^{7}Be movements towards the surface, was among the most relevant inputs in both fractions of the ^{7}Be series, and it exhibited memory in both fractions, thus suggesting a rather complex mechanism. We suggested that this could perhaps be better captured using ERA5 data at additional intermediate levels over an expanded area and not only at the nearest locations. Therefore, to standardize this methodology to the greatest extent possible, ERA5 data were used in all cases instead of using data from different local observational networks. Using ERA5 data is particularly relevant for upper-atmosphere variables, which are otherwise not easily available, and it allows better reproducibility in other environments.

5. Conclusions

This study analysed the 7Be time series measured at six Spanish laboratories corresponding to the 2006–2021 period. Applying the KZ filter to all time series with a cut-off frequency of 33 days enabled independent analysis of both the seasonal component and the noisiest fraction of the observations. As both fractions were found to be independent, they were analysed separately using RF models.

Overall, the RF models described the major mechanisms involved in the generation, transport, and washout of ^{7}Be in the

atmosphere. There were no significant variations between the overall averages of the laboratories.

In all laboratories, for the KZ-filtered fractions ^{7}Be observations one week before turned out to be the most influential inputs. This indicates the importance of the memory of the system at explaining the observations throughout the period analysed, and this was linked to a high fraction of the variance associated with the seasonal cycle.

Apart from the memory of ^{7}Be, it can be seen that the KZ-filtered fraction is driven by seasonal mechanisms, mainly those associated to the seasonal evolution of radiation and precipitation-related variables. These are exactly the mechanisms known to be involved in the generation and washing-out of ^{7}Be.

Based on these results, it can be concluded that Spain has a general background concentration of around 4 mBq/m^3 with clear seasonal oscillations associated to the KZ fractions plus sudden apportions and/or eliminations from the atmosphere linked to a great number of low-energy mechanisms driven by periodicities below 33 days. In the long-term, these rapidly changing mechanisms contribute with zero mBq/m^3 to the overall average at the six locations.

In this study, for the first time, upper-atmosphere radiative (O_3) and transportation variables (ω) have been used as explanatory variables for the ^{7}Be models for both fractions. In general terms, these variables have turned out to be more influential than a radiation proxy like SSN. This also means that in this and also future studies, the portion of the solar cycle analysed –another proxy after all-is likely not to have a great relevance since these new variables can more directly account for the mechanisms involved.

The combined approach employed in this study (the KZ filter followed by the application of RF models to both fractions) enabled explanation of between 48 % and 70 % of the overall variability in weekly ^{7}Be observations.

Under this novel approach, the main mechanisms known to be involved in the generation of ^{7}Be now also appear using this new methodology. The analysis in terms of relevant periodicities (KZ filter) followed by random forests has allowed a more specific identification of the variables responsible for the surface ^{7}Be observations.

It is of note that, unlike many previous studies, the models' performances were based on the use of different data from that employed to fit the models. Therefore, considering the high number of cases used here, it is considered that the results represent a realistic assessment of the methodology presented here.

Research is currently being conducted in two directions:

1. Determining whether dividing the ^{7}Be time series into two fractions with a cut-off frequency of 33 days, which has been detected in Spanish laboratories, is also applicable at other latitudes. Many factors may be responsible for this division, and among them, the so-called Carrington rotation, a synodic rotation period of the sun of approximately one month [67], is currently under consideration by the authors. In this respect, a harmonic analysis [68] of the series was conducted to identify the amplitudes A_i and phases φ_i in the origin of the ^{7}Be observations. The 95 % confidence interval of the parameters, the R^2 value of each fit and the distribution of the p-value of the fit were estimated by means of bootstrap, and the contributions of the corresponding harmonics were assessed. Preliminary results (not shown here) indicate that the seasonal cycle is dominant, and additional harmonics make relatively minor contributions. Through the use of a different methodology, these results verify those of the current study and identify the existence of a dominant seasonal fraction plus noisy residuals. However, this requires further investigation.
2. The results reported in the current study were based on data obtained at 500 hPa at the nearest vertical location from each laboratory. However, using ERA5 at additional pressure levels over a much wider area would provide more information about ^{7}Be transportation to the surface. This could be achieved using a dimensionality reduction process via principal component analysis to obtain a reasonable number of variables, and this implies the need to conduct a more accurate evaluation of the seasonal mechanisms involved. As additional inputs are required to increase the overall fraction explained, incorporating such information in the RF models could increase their performance.

Data availability statement

Data are freely and publicly available at: https://www.csn.es/kprgisweb2/index.html?lang=es.

CRediT authorship contribution statement

Ander Nafarrate: Writing – original draft, Visualization, Resources, Investigation, Data curation. **Susana Petisco-Ferrero:** Writing – original draft, Visualization, Resources, Investigation, Data curation, Conceptualization. **Raquel Idoeta:** Writing – review & editing, Project administration, Methodology, Funding acquisition, Conceptualization. **Margarita Herranz:** Supervision, Methodology, Funding acquisition, Conceptualization. **Jon Sáenz:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Alain Ulazia:** Writing – review & editing, Validation, Software, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Gabriel Ibarra-Berastegui:** Writing – review & editing, Validation, Supervision, Software, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study is part of projects PID2020-116153RB-I00 and TED2021-132109B- C21 (HOBE) funded by MCIN/AEI/10.13039/501100011033, Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación, and the European Union NextGenerationEU/PRTR (BlueAdapt).

The authors also acknowledge funding from SENyRAD (IT1694-22), a research group at the Basque University System of the Basque Government and GIU20/08 (EHU).

References

[1] J.R. Arnold, H.A. Al-Salih, Beryllium-7 produced by cosmic rays, Science 121 (1955) 451–453, https://doi.org/10.1126/science.121.3144.451.
[2] M.R. Islam, P. Sanderson, T.E. Payne, A.K. Deb, R. Naidu, Role of beryllium in the environment: insights from specific sorption and precipitation studies under different conditions, Sci. Total Environ. 838 (2022) 155698, https://doi.org/10.1016/j.scitotenv.2022.155698.
[3] M.C. Bas, J. Ortiz, L. Ballesteros, S. Martorell, Analysis of the influence of solar activity and atmospheric factors on 7Be air concentration by seasonal-trend decomposition, Atmos. Environ. 145 (2016) 147–157, https://doi.org/10.1016/j.atmosenv.2016.09.027.
[4] M. Yoshimori, Production and behavior of beryllium 7 radionuclide in the upper atmosphere, Adv. Space Res. 36 (2005) 922–926, https://doi.org/10.1016/j.asr.2005.04.093.
[5] International Atomic Energy Agency - IAEA, Live chart of nuclides, n.d. https://www-nds.iaea.org/relnsd/vcharthtml/VChartHTML.html. (Accessed 6 June 2023).
[6] J.S. Gaffney, N.A. Marley, M.M. Cunningham, Natural radionuclides in fine aerosols in the Pittsburgh area, Atmos. Environ. 38 (2004) 3191–3200, https://doi.org/10.1016/j.atmosenv.2004.03.015.
[7] M.M. Rajačić, D.J. Todorović, J.D. Krneta Nikolić, M.M. Janković, V.S. Djurdjević, The Fourier analysis applied to the relationship between 7Be activity in the Serbian atmosphere and meteorological parameters, Environ. Pollut. 216 (2016) 919–923, https://doi.org/10.1016/j.envpol.2016.06.068.
[8] R.L. Lozano, M.A. Hernández-Ceballos, E.G. San Miguel, J.A. Adame, J.P. Bolívar, Meteorological factors influencing the 7Be and 210Pb concentrations in surface air from the southwestern Iberian Peninsula, Atmos. Environ. 63 (2012) 168–178, https://doi.org/10.1016/j.atmosenv.2012.09.052.
[9] C. Papastefanou, A. Ioannidou, S. Stoulos, M. Manolopoulou, Atmospheric deposition of cosmogenic 7Be and 137Cs from fallout of the Chernobyl accident, Sci. Total Environ. 170 (1995) 151–156, https://doi.org/10.1016/0048-9697(95)04608-4.
[10] M.A. Hernández-Ceballos, E. Brattich, R.L. Lozano, G. Cinelli, 7Be behaviour and meteorological conditions associated with 7Be peak events in Spain, J. Environ. Radioact. 166 (2017) 17–26, https://doi.org/10.1016/j.jenvrad.2016.03.019.
[11] H. Nagai, W. Tada, T. Kobayashi, Production rates of 7Be and 10Be in the atmosphere, Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. Atoms 172 (2000) 796–801, https://doi.org/10.1016/S0168-583X(00)00124-5.
[12] A. Rodríguez-Perulero, A. Baeza, J. Guillén, Seasonal evolution of 7,10Be and 22Na in the near surface atmosphere of Cáceres (Spain), J. Environ. Radioact. 197 (2019) 55–61, https://doi.org/10.1016/j.jenvrad.2018.11.015.
[13] E. Gordo, C. Dueñas, M.C. Fernández, E. Liger, S. Cañete, Behavior of ambient concentrations of natural radionuclides 7Be, 210Pb, 40K in the Mediterranean coastal city of Málaga (Spain), Environ. Sci. Pollut. Res. 22 (2015) 7653–7664, https://doi.org/10.1007/s11356-014-4039-5.
[14] D. Lal, B. Peters, Cosmic ray produced radioactivity on the Earth, in: Handb. Der Phys., Springer Verlag, New York, NY, 1967, pp. 551–612.
[15] I. Sýkora, K. Holý, M. Ješkovský, M. Müllerová, M. Bulko, P.P. Povinec, Long-term variations of radionuclides in the Bratislava air, J. Environ. Radioact. 166 (2017) 27–35, https://doi.org/10.1016/j.jenvrad.2016.03.004.
[16] K. Psistaki, S. Achilleos, N. Middleton, A.K. Paschalidou, Exploring the impact of particulate matter on mortality in coastal Mediterranean environments, Sci. Total Environ. 865 (2023) 161147, https://doi.org/10.1016/j.scitotenv.2022.161147.
[17] O. Masson, J. Bieringer, E. Brattich, A. Dalheimer, S. Estier, I. Penev, W. Ringer, C. Schlosser, T. Steinkopff, P. Steinmann, L. Tositti, P. Van Beek, A. de Vismes-Ott, Variation in airborne 134Cs, 137Cs, particulate 131I and 7Be maximum activities at high-altitude European locations after the arrival of Fukushima-labeled air masses, J. Environ. Radioact. 162–163 (2016) 14–22, https://doi.org/10.1016/j.jenvrad.2016.05.004.
[18] L. Tositti, E. Brattich, G. Cinelli, D. Baldacci, 12 years of 7Be and 210Pb in Mt. Cimone, and their correlation with meteorological parameters, Atmos. Environ. 87 (2014) 108–122, https://doi.org/10.1016/j.atmosenv.2014.01.014.
[19] L. Jiwen, V.N. Starovoitova, D.P. Wells, Long-term variations in the surface air 7Be concentration and climatic changes, J. Environ. Radioact. 116 (2013) 42–47, https://doi.org/10.1016/j.jenvrad.2012.08.015.
[20] L. Terzi, G. Wotawa, M. Schoeppner, M. Kalinowski, P.R.J. Saey, P. Steinmann, L. Luan, P.W. Staten, Radioisotopes demonstrate changes in global atmospheric circulation possibly caused by global warming, Sci. Rep. 10 (2020) 1–13, https://doi.org/10.1038/s41598-020-66541-5.
[21] P.K. Pandey, K.S. Patel, P. Šubrt, Trace elemental composition of atmospheric particulate at Bhilai in central-east India, Sci. Total Environ. 215 (1998) 123–134, https://doi.org/10.1016/S0048-9697(98)00111-9.
[22] M. Długosz-Lisiecka, S. Nowański, Factor regression as a tool to predict Be-7 concentration in the air, Atmos. Environ. 297 (2023) 1–8, https://doi.org/10.1016/j.atmosenv.2023.119612.
[23] A. Kulan, A. Aldahan, G. Possnert, I. Vintersved, Distribution of 7Be in surface air of Europe, Atmos. Environ. 40 (2006) 3855–3868, https://doi.org/10.1016/j.atmosenv.2006.02.030.
[24] M.A. Hernández-Ceballos, G. Cinelli, M. Marín Ferrer, T. Tollefsen, L. De Felice, E. Nweke, P.V. Tognoli, S. Vanzo, M. De Cort, A climatology of 7Be in surface air in European Union, J. Environ. Radioact. 141 (2015) 62–70, https://doi.org/10.1016/j.jenvrad.2014.12.003.
[25] T. Zalewska, D. Biernacik, M. Marosz, Correlations between 7Be, 210Pb, dust and PM10 concentrations in relation to meteorological conditions in northern Poland in 1998–2018, J. Environ. Radioact. 228 (2021), https://doi.org/10.1016/j.jenvrad.2020.106526.
[26] H. Wang, Q. Meng, Y. Ma, Y. Lou, B. Bai, W. Zhu, Y. Kong, H. Wang, Z. Feng, S. Zhai, J. Yu, Temporal variations of 7Be and 210Pb activity concentrations in the atmospheric aerosols during 2018–2019 in Beijing, China and their correlations with meteorological parameters, J. Environ. Radioact. 262 (2023), https://doi.org/10.1016/j.jenvrad.2023.107162.
[27] A. Ioannidou, M. Manolopoulou, C. Papastefanou, Temporal changes of 7Be and 210Pb concentrations in surface air at temperate latitudes (40°N), Appl. Radiat. Isot. 63 (2005) 277–284, https://doi.org/10.1016/j.apradiso.2005.03.010.
[28] N. Alegría, M. Herranz, R. Idoeta, F. Legarda, Study of 7Be activity concentration in the air of northern Spain, J. Radioanal. Nucl. Chem. 286 (2010) 347–351, https://doi.org/10.1007/s10967-010-0710-6.
[29] M.A. Hernández-Ceballos, G. Cinelli, T. Tollefsen, M. Marín-Ferrer, Identification of airborne radioactive spatial patterns in Europe - feasibility study using Beryllium-7, J. Environ. Radioact. 155–156 (2016) 55–62, https://doi.org/10.1016/j.jenvrad.2016.02.006.
[30] L. Bohdalkova, M. Novak, P. Voldrichova, E. Prechova, F. Veselovsky, L. Erbanova, M. Krachler, A. Komarek, J. Mikova, Atmospheric deposition of beryllium in Central Europe: comparison of soluble and insoluble fractions in rime and snow across a pollution gradient, Sci. Total Environ. 439 (2012) 26–34, https://doi.org/10.1016/j.scitotenv.2012.08.089.
[31] A. Stohl, N. Spichtinger-Rakowsky, P. Bonasoni, H. Feldmann, M. Memmesheimer, H.E. Scheel, T. Trickl, S. Hübener, W. Ringer, M. Mandl, The influence of stratospheric intrusions on alpine ozone concentrations, Atmos. Environ. 34 (2000) 1323–1354, https://doi.org/10.1016/S1352-2310(99)00320-9.
[32] M.P. Mohan, R.S. D'Souza, S.R. Nayak, S.S. Kamath, T. Shetty, K.S. Kumara, Y.S. Mayya, N. Karunakara, Influence of rainfall on atmospheric deposition fluxes of 7Be and 210Pb in Mangaluru (Mangalore) at the southwest Coast of India, Atmos. Environ. 202 (2019) 281–295, https://doi.org/10.1016/j.atmosenv.2019.01.034.

[33] F. Zhang, M. Yang, J. Zhang, Beryllium-7 in vegetation, soil, sediment and runoff on the northern Loess Plateau, Sci. Total Environ. 626 (2018) 842–850, https://doi.org/10.1016/j.scitotenv.2018.01.156.

[34] J. Juri Ayub, F. Lohaiza, H. Velasco, M. Rizzotto, D. Di Gregorio, H. Huck, Assessment of 7Be content in precipitation in a South American semi-arid environment, Sci. Total Environ. 441 (2012) 111–116, https://doi.org/10.1016/j.scitotenv.2012.09.079.

[35] N. Alegría, M.Á. Hernández-Ceballos, M. Herranz, R. Idoeta, F. Legarda, Meteorological factors controlling 7Be activity concentrations in the atmospheric surface layer in northern Spain, Atmosphere 11 (2020), https://doi.org/10.3390/atmos11121340.

[36] A.C. Carvalho, M. Reis, L. Silva, M.J. Madruga, A decade of 7Be and 210Pb activity in surface aerosols measured over the Western Iberian Peninsula, Atmos. Environ. 67 (2013) 193–202, https://doi.org/10.1016/j.atmosenv.2012.10.060.

[37] T. Zalewska, D. Biernacik, Be-7 and Pb-210 in fallout and aerosols in 2000–2016 in central Europe — deposition velocity and dependence on meteorological parameters, Sci. Total Environ. 826 (2022) 154205, https://doi.org/10.1016/j.scitotenv.2022.154205.

[38] A. Ioannidou, J. Paatero, Activity size distribution and residence time of 7Be aerosols in the Arctic atmosphere, Atmos. Environ. 88 (2014) 99–106, https://doi.org/10.1016/j.atmosenv.2013.12.046.

[39] F. Piñero García, M.A. Ferro García, M. Azahra, 7Be behaviour in the atmosphere of the city of Granada January 2005 to December 2009, Atmos, Environ. Times 47 (2012) 84–91, https://doi.org/10.1016/j.atmosenv.2011.11.034.

[40] H.W. Feely, R.J. Larsen, C.G. Sanderson, Factors that cause seasonal variations in Beryllium-7 concentrations in surface air, J. Environ. Radioact. 9 (1989) 223–249, https://doi.org/10.1016/0265-931X(89)90046-5.

[41] H. Hötzl, G. Rosner, R. Winkler, Correlation of7BE concentrations in surface air and precipitation with the solar cycle, Naturwissenschaften 78 (1991) 215–217. https://api.semanticscholar.org/CorpusID:35733474.

[42] S. Rehfeld, M. Heimann, Three dimensional atmospheric transport simulation of the radioactive tracers 210Pb, 7Be, 10Be, and 90Sr, J. Geophys. Res. 100 (1995), https://doi.org/10.1029/95jd01003.

[43] R. Uhlář, P. Haroková, P. Alexa, M. Kačmařík, 7Be atmospheric activity concentration and meteorological data: statistical analysis and two-layer atmospheric model, J. Environ. Radioact. 219 (2020) 106278, https://doi.org/10.1016/j.jenvrad.2020.106278.

[44] D. Biernacik, T. Zalewska, 7Be, 210Pb, airborne particulate matter and PM10 concentrations in relation to meteorological conditions in southern Poland in 1998–2016, J. Environ. Radioact. 259–260 (2023) 107122, https://doi.org/10.1016/j.jenvrad.2023.107122.

[45] E. Chham, F. Piñero-García, E. Brattich, T. El Bardouni, M.A. Ferro-García, 7Be spatial and temporal pattern in southwest of Europe (Spain): evaluation of a predictive model, Chemosphere 205 (2018) 194–202, https://doi.org/10.1016/j.chemosphere.2018.04.099.

[46] Consejo de Seguridad Nuclear (CSN), Mapa de valores radiológicos ambientales, n.d. https://www.csn.es/sistema-de-vigilancia-ambiental-en-espana. (Accessed 6 June 2023).

[47] Technical Committee CTN 66, ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories, Int. Organ. Stand. (2018) 2–44.

[48] Royal Observatory of Belgium, Solar influences data analysis center, n.d. https://www.sidc.be/silso/datafiles. (Accessed 2 May 2023).

[49] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R.J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, J.N. Thépaut, The ERA5 global reanalysis, Q. J. R. Meteorol. Soc. 146 (2020) 1999–2049, https://doi.org/10.1002/qj.3803.

[50] Copernicus climate change service, Copernicus, n.d. https://www.copernicus.eu/en/copernicus-services. (Accessed 6 June 2023).

[51] D.P. Dee, S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, L. Haimberger, S.B. Healy, H. Hersbach, E.V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A.P. Mcnally, B.M. Monge-Sanz, J.J. Morcrette, B.K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.N. Thépaut, F. Vitart, The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc. 137 (2011) 553–597, https://doi.org/10.1002/qj.828.

[52] Climate data store (CDS), Climate data store (CDS), n.d. https://cds.climate.copernicus.eu/#!/home. (Accessed 6 June 2023).

[53] H. Hersbach, C. Peubey, A. Simmons, P. Berrisford, P. Poli, D. Dee, ERA-20CM: a twentieth-century atmospheric model ensemble, Q. J. R. Meteorol. Soc. 141 (2015) 2350–2375, https://doi.org/10.1002/qj.2528.

[54] S.T. Rao, I.G. Zurbenko, Detecting and tracking changes in ozone air quality, Air Waste 44 (1994) 1089–1092, https://doi.org/10.1080/10473289.1994.10467303.

[55] M.L. Milanchus, S.T. Rao, I.G. Zurbenko, Evaluating the effectiveness of ozone management efforts in the presence of meteorological variability, J. Air Waste Manag. Assoc. 48 (1998) 201–215, https://doi.org/10.1080/10473289.1998.10463673.

[56] G. Ibarra-Berastegi, I. Madariaga, A. Elías, E. Agirre, J. Uria, Long-term changes of ozone and traffic in Bilbao, Atmos. Environ. 35 (2001) 5581–5592, https://doi.org/10.1016/S1352-2310(01)00210-2.

[57] H.C. Thode, Testing for Normality, first ed., CRC Press, Boca Raton, 2002. https://books.google.es/books?id=gbegXB4SdosC.

[58] J. Gross, U. Ligges, Nortest: Tests for Normality, 2015.

[59] M.J. Crawley, The R Book, second ed., Wiley, Chichester, West Sussex, United Kingdom, 2013., 2013. https://search.library.wisc.edu/catalog/9910152827402121.

[60] U. Grömping, Variable importance assessment in regression: linear regression versus random forest, Am. Statistician 63 (2009) 308–319, https://doi.org/10.1198/tast.2009.08199.

[61] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1007/978-3-030-62008-0_35.

[62] A. Liaw, M. Wiener, Classification and regression by random forest, R. News 2 (2002) 18–22.

[63] D.S. Siroky, Navigating random forests and related advances in algorithmic modeling, Stat. Surv. 3 (2009) 147–163, https://doi.org/10.1214/07-SS033.

[64] G. Ibarra-Berastegi, J. Saénz, A. Ezcurra, A. Elías, J. Diaz Argandoña, I. Errasti, Downscaling of surface moisture flux and precipitation in the Ebro Valley (Spain) using analogues and analogues followed by random forests and multiple linear regression, Hydrol. Earth Syst. Sci. 15 (2011) 1895–1907, https://doi.org/10.5194/hess-15-1895-2011.

[65] G. Ibarra-Berastegi, J. Saénz, G. Esnaola, A. Ezcurra, A. Ulazia, Short-term forecasting of the wave energy flux: analogues, random forests, and physics-based models, Ocean Eng. 104 (2015) 530–539, https://doi.org/10.1016/j.oceaneng.2015.05.038.

[66] A.K. Paschalidou, K. Psistaki, A. Charalampopoulos, D. Vokou, P. Kassomenos, A. Damialis, Identifying patterns of airborne pollen distribution using a synoptic climatology approach, Sci. Total Environ. 714 (2020) 136625, https://doi.org/10.1016/j.scitotenv.2020.136625.

[67] E.H. Schröter, The solar differential rotation: present status of observations, Sol. Phys. 100 (1985) 141–169, https://doi.org/10.1007/BF00158426.

[68] R.E. Thomson, W.J. Emery, Data Analysis Methods in Physical Oceanography, Elsevier BV, 2014, https://doi.org/10.1016/C2010-0-66362-0.