Tomás Sánchez Sánchez-Pastor

Centro de Investigaciones Energéticas, Medioambientales y tecnológicas.
(CIEMAT)

1. Time Series principles

A *Time Series* is defined as a set of data in time order. Most commonly, is a sequence taken at successive equally spaced points in time. the main goal that we want to achieve is the forecasting of the ^{226}Rn time series produced in the underground ArDM experiment, located in Canfranc.

In order to forecast the time series, we are going to analyze the existing different components that take part in the observed series.

1.1. STL Decomposition

The STL decomposition is a statistical task that deconstructs a time series into several components, each representing the underlying categories of patterns. STL decomposes a time series into:

- **Trend (T)**, also named secular variation, is a long-term non-periodic variation. Exists when there is a persistent increasing or decreasing direction in data.
- **Seasonal (S)**, corresponds with the periodic behaviour or pattern influenced by seasonal factors.
- **Residual (R)**, describes the noisy contribution to the signal.

The calculus that we have to apply to the data is:

$$Y_{STL} = T + S + R \tag{1}$$

And the multiplicative form of the decomposition:

$$Y_{STL} = T \cdot S \cdot R \tag{2}$$

The main difference between these two methods remains in the time series properties. If the frequency and amplitude of the time series vary we would use the multiplicative way and the additive one if are constants.

2. Results

2.1. STL decomposition of the ^{226}Rn signal

We have shown below the STL decomposition of the weekly ^{226}Rn signal as a function of the recording time in the laboratory. As can be appreciated, the seasonal contribution has the same scale than the noise, this implies that there is not periodicity, therefore it corresponds with a modulated noisy signal.

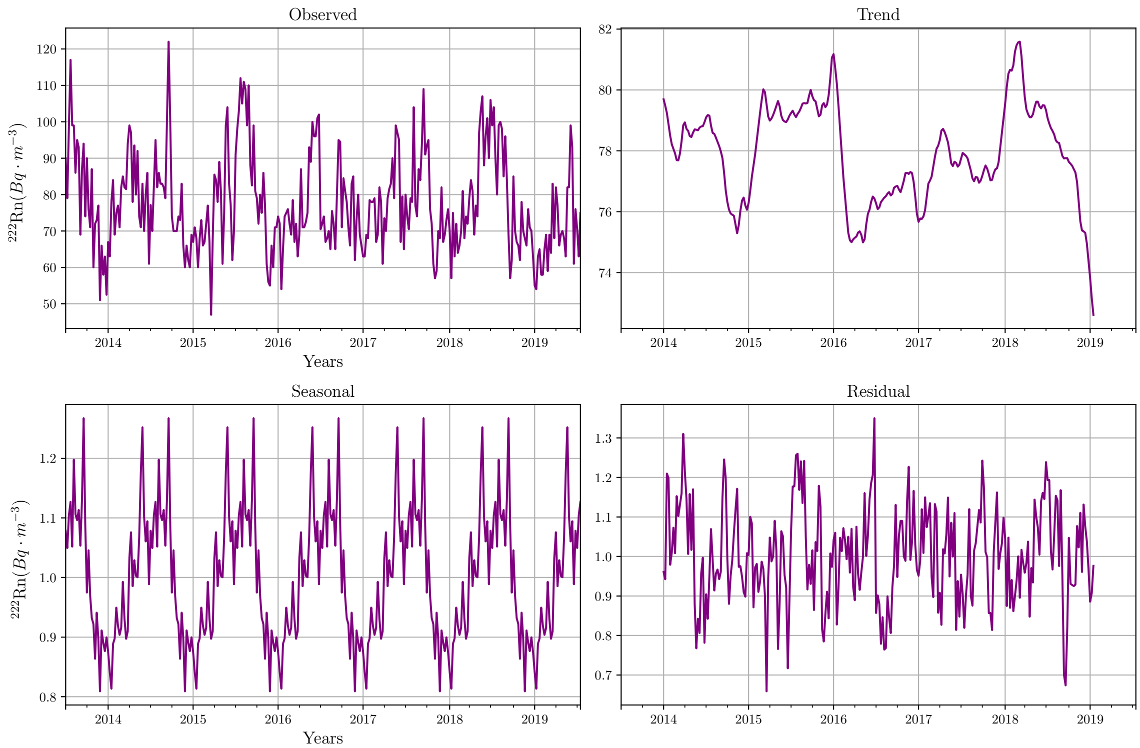

Fig. 1: STL decomposition of the weekly Radon time series at ArDM experiment . Observed data (upper left), trend (upper right), seasonal (lower left) and noisy contributions (lower right) as a function of the recording time.

2.2. Fast Fourier Transform (FFT)

Due to the noisy nature of the signal it is important to subtract the frequency domain. Once it is done applying a filter in the relevant frequency intervals would help us obtain a clearer signal

A comparision betwen de FFT and the observed signal is shown in Fig. 4. As can be seen in the lower figure, there is only one relevant frequency located at 0,019 $weeks^{-1}$.

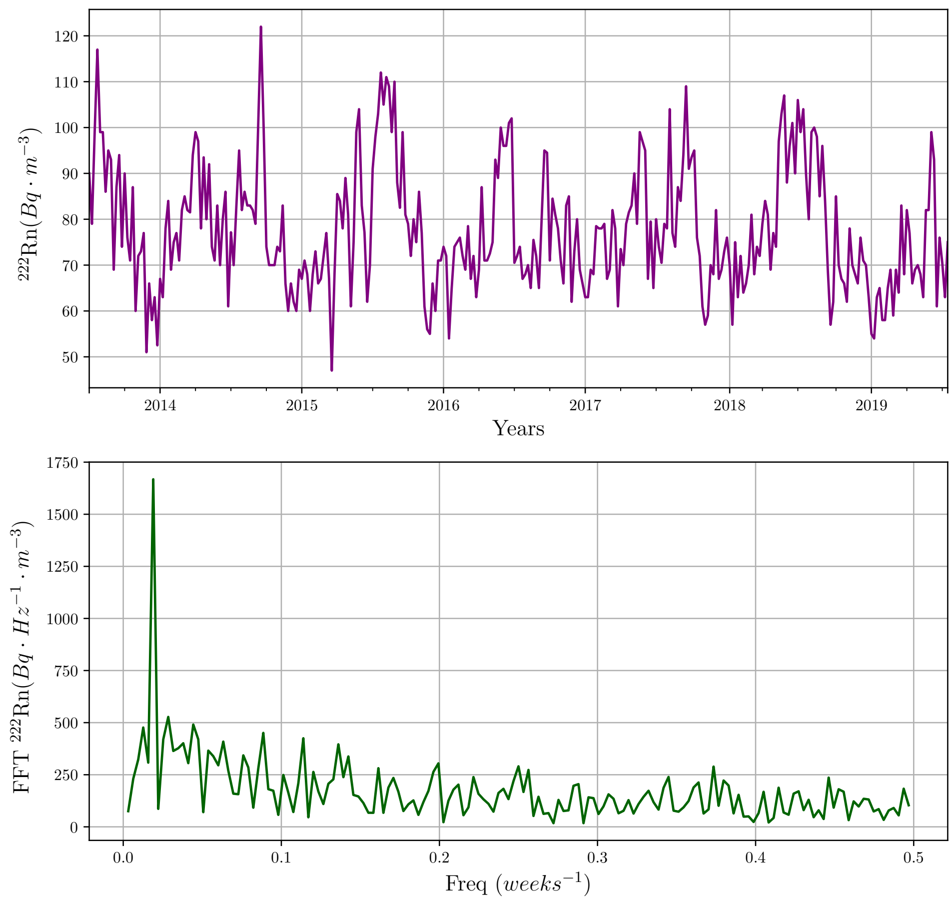

Fig. 2: Observed signal (upper) and frequency domain obtained with the FFT algorithm (lower).

The existence of a fundamental frequency in the signal implies that there is a periodic behavior on the Radon levels in the underground laboratory of Canfranc. The time period according to the frequency is approximately 53 *weeks*.

It is important to highlight that any seasonality has been found with the STL Decomposition method, nevertheless, it is possible computing the frequency domain of the signal through FFT method. As expected, the period of the seasonality is about 1 *year*.

2.3. CNN forecasting of the signal

The collected data are divided into two sets: the training data set (including the first four years, from July 2013 to June 2017) and the test data set, which includes the last year.

The Convolutional Neural Network model employed is composed of one convolutional layer with 64 filters with relu as activation function, a Max Pooling layer and two dense layers with 64 and 32 neurons respectively and relu activation function.

In Fig. 3, the predicted values for the test set as long as the complete set of data are shown. The CNN method is able to capture information of the previous values to predict the behavior of the time series. Even though, the accuracy the model gets is of a 3,6 %

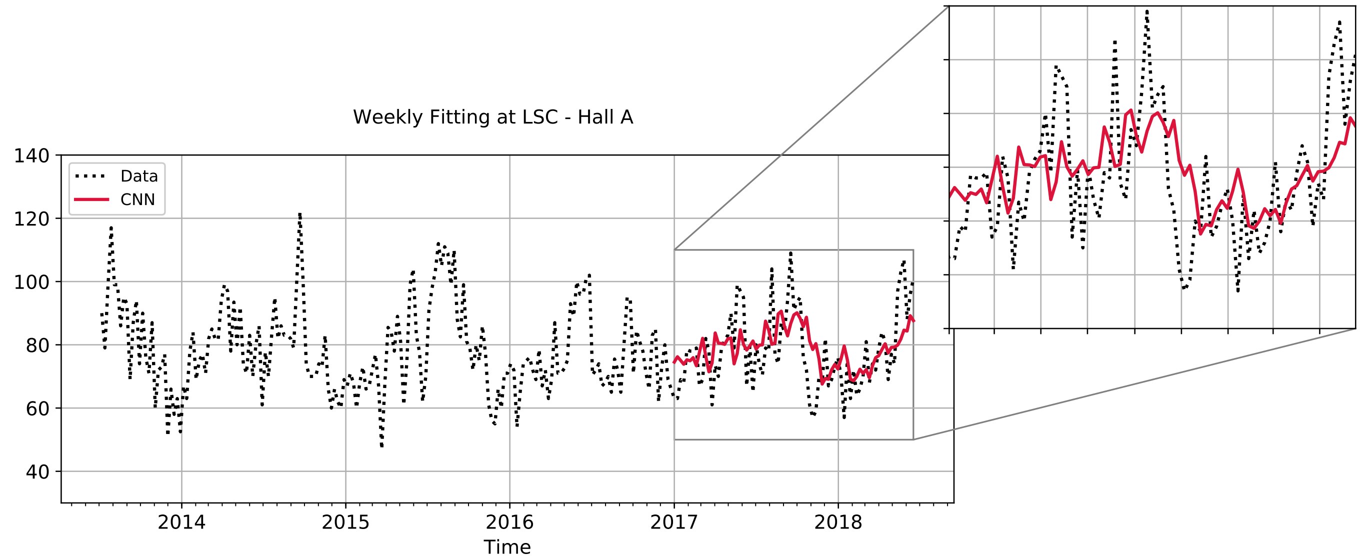

Fig. 3: Signal (black dots) and prediction of the Convolutional Neural Network model (solid red line) for the two last years of data recording.

2.4. Study of the performance of the model

In order to check if there is a bias or variance problem in the cognitive model, we have shown the values of the Loss function as a function of the epochs used to train the model in Fig. 4. As we can see, the model quickly reaches a minimum value in the two data sets, then there is not a bias problem. Moreover, the values of train and test loss function are roughly equal at the end of the training, the variance problem is ruled out.

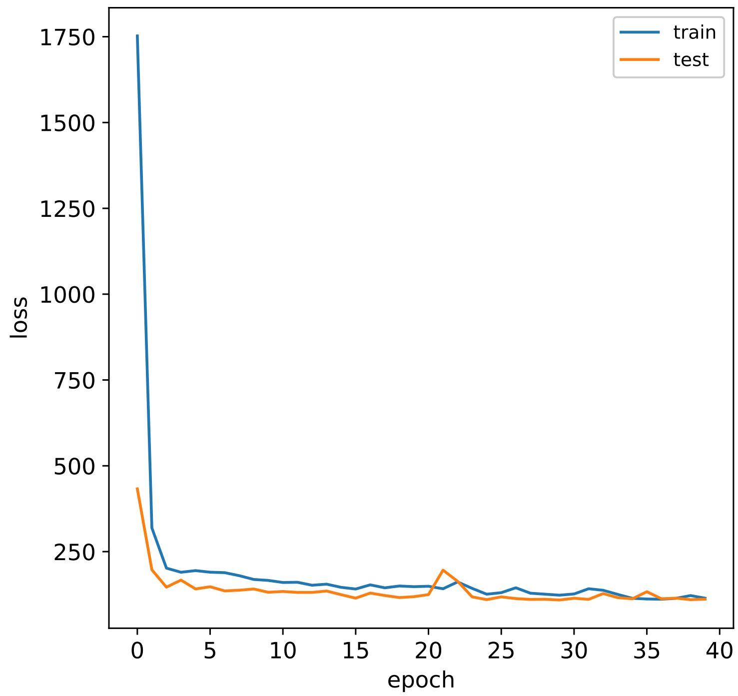

Fig. 4: Train and test loss as a function of the epochs used to train the CNN.