atmosphere

MDPI

Supplementary Materials: Change Points Detection and Trend Analysis to Characterize Changes in Meteorologically Normalized Air Pollutant Concentrations

Roberta Valentina Gagliardi [1,*] and Claudio Andenna [2]

Model evaluation parameters

The statistical indicators used to assess the performances of the selected optimal RF model by comparing predicted and observed pollutants concentrations values are, the coefficient of determination (R^2), the mean bias error (MBE), the mean absolute error (MAE), the root mean square error (RMSE) and the index of agreement (IoA). High accuracy (R^2 and IoA close to 1) and minimal errors (MBE, MAE, and RMSE close to 0) are the desired performances for an optimal prediction model. MBE provides a good indication of the mean over or under estimate of predictions, while MAE provides a good indication of the mean error regardless of whether it is an over or under estimate. The RMSE provides an overall measure of how close modelled values are to predicted values. R2 represents the amount of the variance for an air pollutant concentration that is explained by meteorological and time variables used. Moreover, the dimensionless IoA is also included in the analysis to illustrate the agreement between the values. All these metrics together explain whether the models have taken into account the most appropriate variables to describe the air pollutant concentrations patterns and whether they perform reasonable predictions.

Table S1. Statistical indicators.

Statistic name	Equation
Mean Bias Error	$MBE = \frac{1}{N}\sum_{i=1}^{N} M_i - O_i$
Mean Absolute Error	$MAE = \frac{1}{N}\sum_{i=1}^{N} \lvert M_i - O_i \rvert$
Root Mean Squared Error	$RMSE = \sqrt{\left(\frac{\sum_{i=1}^{N}(M_i - O_i)^2}{N}\right)}$
Coefficient of Determination	$R^2 = \left(\{\sum_{i=1}^{N}(M_i - \bar{M})(O_i - \bar{O})\} \Big/ \{\sum_{i=1}^{N}(M_i - \bar{M})^2 (O_i - \bar{O})^2\}^{\frac{1}{2}} \right)^2$
Index of Agreement	$IoA = 1 - \frac{\sum_{i=1}^{N}\lvert M_i - O_i \rvert}{c\sum_{i=1}^{N}\lvert O_i - \bar{O} \rvert}$, when $\sum_{i=1}^{N}\lvert M_i - O_i \rvert \le c\sum_{i=1}^{N}\lvert O_i - \bar{O} \rvert$ $IoA = \frac{c\sum_{i=1}^{N}\lvert O_i - \bar{O} \rvert}{\sum_{i=1}^{N}\lvert M_i - O_i \rvert} - 1$, when $\sum_{i=1}^{N}\lvert M_i - O_i \rvert > c\sum_{i=1}^{N}\lvert O_i - \bar{O} \rvert$ with c=2

Where:

N = total number of measurements; M_i = ith predicted value; O_i = ith observed value; $\bar{M}$ = mean of the predicted values; $\bar{O}$ = mean of the observed values

Statistical results

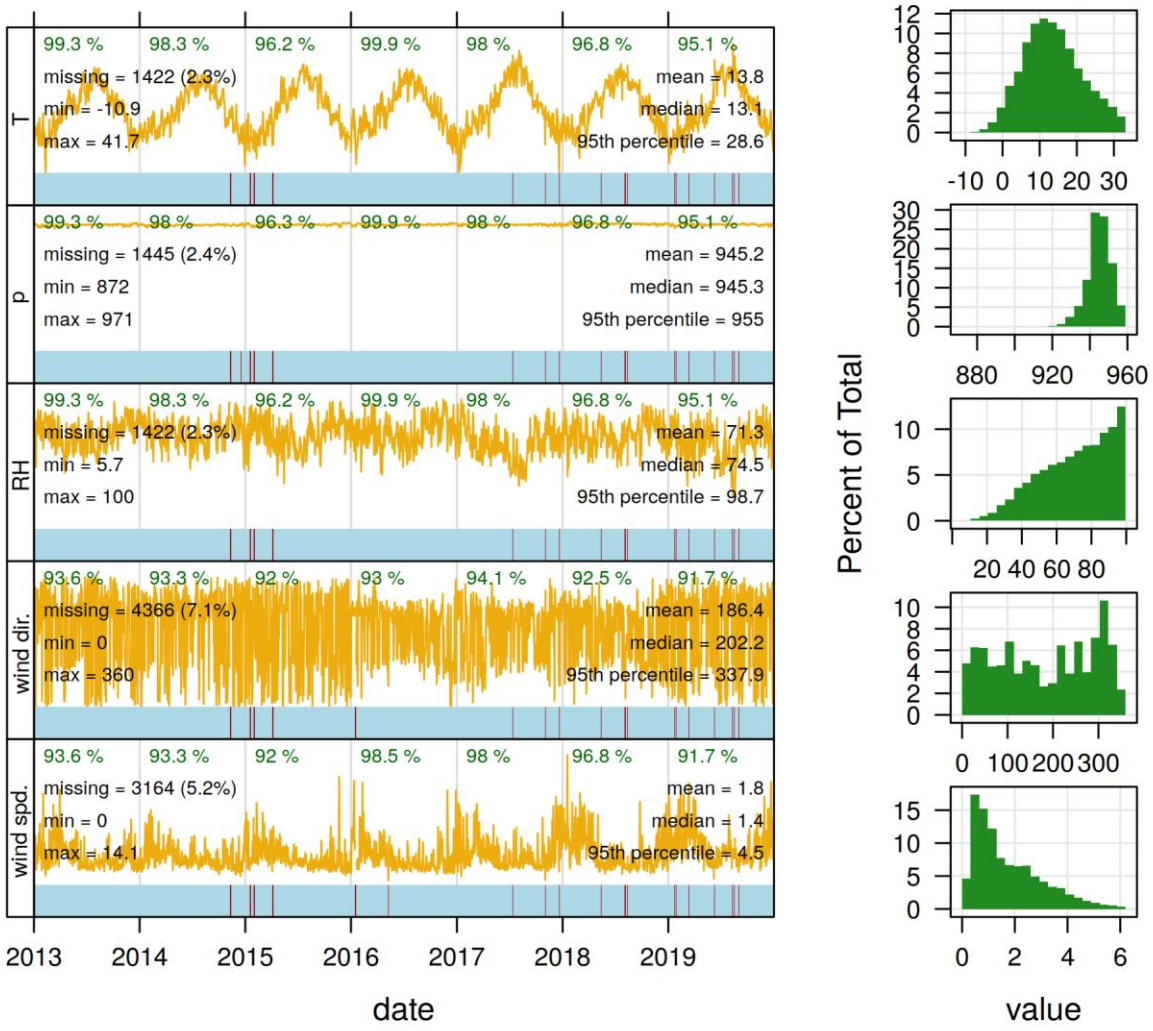
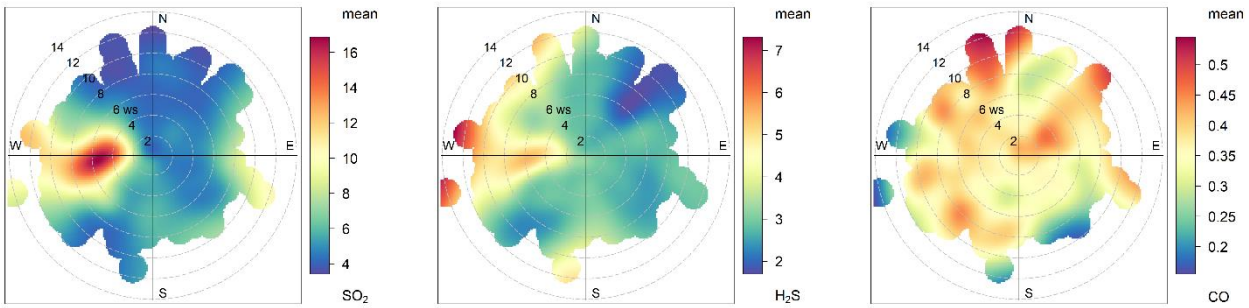


Figure S1. Left panel: the time series data of hourly T, p, H, wd and ws values (yellow lines,) and the availability and non-availability of the data (respectively the blue and red colors on the rectangular bar at the bottom of the plots). The minimum, maximum, number and percent of missing data, mean, median and the 95th percentile for each variable plotted (in black). The percentage of the data captured for every year (in green) on the upper part of each year data plot. Right panel: the distribution of each variable using a histogram plot over the selected periods.



a) **b)** **c)**

Figure S2. Polar Plot of hourly concentrations of (**a**) SO2, (**b**) H2S and (**c**) CO over the 2013-2019 period.

Table S2. Change points in the meteorologically normalised concentrations of the selected pollutants.

id	NOx	CO	SO_2	H_2S
1	09/03/2016	19/10/2015	24/04/2018	28/03/2014
2	08/10/2013	19/05/2017	31/12/2014	12/11/2014
3	04/10/2015	09/04/2019	02/08/2016	09/02/2019
4	14/08/2016	08/09/2013	12/06/2015	02/08/2016
5	09/02/2018	05/05/2016	27/02/2014	20/10/2016
6	13/10/2018	06/03/2014	20/10/2016	15/05/2017
7	13/10/2019	18/07/2018	16/07/2017	16/11/2018
8	14/02/2019	15/10/2019	07/05/2019	15/01/2018
9	14/01/2017	18/04/2013	24/09/2019	07/02/2018
10	14/10/2017	09/06/2014	05/04/2016	22/06/2015
11	12/10/2016	19/09/2016	02/09/2014	09/05/2019
12	30/03/2016	10/12/2019	16/12/2019	06/03/2017
13	15/02/2016	19/05/2015	31/12/2015	21/11/2013
14	11/11/2015	02/05/2014	23/02/2018	02/08/2018