The effectiveness of the short-run smog alert policy of Budapest – Online Appendix

Máté Kormos[¶]
CEU Economics MA

Econometrics 2 Term Paper
for Prof. Lychagin, Prof. Rooney

A. More on Methodology

Let us introduce the following expressions for further notation in this section: y_{st} – PM_{10} level at station s at time t ($\mu g m^{-3}$); x_{kt} – explanatory variable k at time t.

A.1. Imputation of missing data

Let $\bar{s}$ be the index of a station such that $y_{\bar{s}t}$ is missing. Then the imputed value, $\tilde{y}_{\bar{s}t}$, is given by $\tilde{y}_{\bar{s}t} := \sum_{s:\exists y_{jt}} \omega_s y_{st}$, with $\omega_s = \frac{r_{\bar{s}s}}{\sum_{j:\exists y_{st}} r_{\bar{s}j}}$, where $r_{\bar{s}s}$ is the Pearson correlation coefficient between PM_{10} levels at station $\bar{s}$ and s measured in the training period, neglecting missing observations. For values of r see Appendix Figure 3.

A.2. Neural Network

Variables The variables included in $\boldsymbol{x_t}$: minimum, maximum, average temperatures; wind blow; their lags up to 4; their dynamic variance, which for variable k at time t is given as $\mathbf{Var}_{kt} := 8^{-1} \sum_{\tau=0}^{7} (x_{kt-\tau} - \bar{x}_{kt})^2$, where $\bar{x}_{kt} = 8^{-1} \sum_{\tau=0}^{7} x_{kt-\tau}$.

Neural Network Networks for each station consist of the input, a hidden, and the output layer, with 4 units in the hidden one. Minimisation of $L = 2^{-1} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2$ was done with stochastic mini-batch gradient descent using a learning rate of 0.5 and a batch size of 241. Relatively quick convergence was witnessed.

[¶]E-mail: mate.kormos33@gmail.com.

B. Tables

B.1. Information on PM_{10} stations

Table 1: Station types and intervals of systematically missing observations

Station[a]	Represented air-quality cluster[b]	Intervals of missing observations[c]
Csepel	suburb industry	• 2013: [06-18, 07-16] • 2014: lot of one-offs
Erzsébet	city traffic	• 2016: [08-29, 09-22]
Gergely	city industry	• 2013: [09-17, 12-05] • 2015: [01-01, 04-10] • 2016: [08-02, 10-03]
Gilice	suburb background	• 2014: [04-29, 06-26]
Honvéd	city background	• 2013: [04-23, 07-28] • 2014: [02-04, 04-15]; [05-15, 09-04] • 2015: [05-06, 05-20]; [06-01, 06-11]; [07-21, 08-04]; [09-27, 11-04] • 2016: lot of one-offs: [07], [09], [10], [11] • 2017: [01-04, 01-31]
Káposztás	city background	• 2014: [01-22, 07-23] • 2017: [01-07, 02-08]
Kőrakas	city background	none
Kosztolányi	city traffic	• 2013: [06-30, 10-03] • 2015: some in: [04], [07] • 2016: [07-24, 10-03]
Pesthidegkút	city background	none
Széna	city traffic	none
Teleki	city traffic	• 2014: [04-04, 07-08]; [08-07, 11-04] • 2016: [06-13, 07-15]; [07-28, 09-06]
Tétény	suburb industry	• 2014: [09-12, 10-31]; [11-28, 12-31] • 2015: [01-01, 01-31]; lot of one-offs • 2016: [09-21, 10-05]

[a] Proper names of stations are: Csepel, Csillaghegy, Erzsébet tér, Gergely utca, Gilice tér, Hondvédtelep, Káposztásmegyer, Kőrakás park, Kosztolányi D. tér, Pesthidegút, Széna tér, Teleki/Baross tér, Budatétény. Highlighted bold names indicate kept stations.
[b] Taxonomy of air-quality clusters are given at the OLM website. Link to the website.
[c] The notation $[m_1m_1 - d_1d_1,\ m_2m_2 - d_2d_2]$ stands for inclusive time interval from $year - m_1m_1 - d_1d_1$ to $year - m_2m_2 - d_2d_2$; while $[m]$ indicates whole month, m, of the given year.

Table 2: Locations of PM_{10} stations, coordinates

Station[a]	Latitude	Longitude
Csepel	**47.404735**	**19.091183**
Erzsébet	**47.497526**	**19.052748**
Gergely	**47.467413**	**19.155894**
Gilice	**47.430904**	**19.181157**
Honvéd	47.521723	19.068420
Káposztás	47.585304	19.114766
Kőrakás	**47.543335**	**19.146578**
Kosztolányi	**47.475968**	**19.040069**
Pesthidegkút	**47.562097**	**18.961434**
Széna	**47.507872**	**19.026794**
Teleki	47.493421	19.084785
Tétény	47.406124	19.009180

[a] Highlighted bold names indicate kept stations. Coordinates are represented in decimal degrees. Conversion from addresses on OLM site to coordinates is done by MapsEasy.com (2017).

B.2. Descriptive statistics

Table 3: Descriptive statistics of the PM_{10} concentration at the kept stations in the training period

	Csepel	Erzsébet	Gergely	Gilice	Kőrakas	Kosztolányi	Pesthidegkút	Széna	Overall
Mean	27.8272	36.0347	24.6619	30.7384	28.0043	34.7470	26.3957	33.0029	30.2130
Median	24.0000	34.0000	22.0000	27.0000	25.0000	30.0000	22.5000	30.0000	27.0000
Mode	16	32	15	19	25	25	14	22	22
Maximum	102	95	87	126	93	107	104	91	126
Minimum	3	9	3	6	4	4	4	8	3
Standard deviation	15.9433	15.5716	12.4006	15.9410	12.9959	18.5286	14.9971	14.1606	15.6033
No. of observations	625	721	624	669	698	589	690	685	5301
No. of days in the period	730	730	730	730	730	730	730	730	5840

Period: from 2013-01-01 to 2014-12-31. Measurement unit: $\mu g m^{-3}$.

Table 4: Descriptive statistics of the PM_{10} concentration at the kept stations in the test period

	Csepel	Erzsébet	Gergely	Gilice	Kőrakas	Kosztolányi	Pesthidegkút	Széna	Overall
Mean	33.2071	37.5242	26.8571	26.4756	27.9375	19.3624	21.5974	35.0815	28.4813
Median	31.0000	34.0000	23.0000	22.0000	24.0000	11.0000	18.0000	30.0000	25.0000
Mode	31	30	16	13	16	8	12	27	16
Maximum	90	119	98	101	93	93	85	115	119
Minimum	7	13	8	6	5	3	5	9	3
Standard deviation	15.1105	16.5208	14.8517	16.0657	16.7347	18.2491	14.0939	17.5670	17.2720
No. of observations	198	227	231	225	224	218	231	233	1787
No. of days in the period	235	235	235	235	235	235	235	235	1880

Period: from 2015-12-01 to 2016-07-22. Measurement unit: $\mu g m^{-3}$.

Table 5: Descriptive statistics of the PM_{10} concentration at the kept stations in the Evaluation 1 period

	Csepel	Erzsébet	Gergely	Gilice	Kőrakas	Kosztolányi	Pesthidegkút	Széna	Overall
Mean	78	80	90	71	57	89	33	77	71.8750
Median	78	80	90	71	57	89	33	77	77.5000
Mode	78	80	90	71	57	89	33	77	33.0000
Maximum	78	80	90	71	57	89	33	77	90
Minimum	78	80	90	71	57	89	33	77	33
Standard deviation	0	0	0	0	0	0	0	0	17.6241
No. of observations	1	1	1	1	1	1	1	1	8
No. of days in the period	1	1	1	1	1	1	1	1	8

Period: 2015-11-08. Measurement unit: μgm^{-3}.

Table 6: Descriptive statistics of the PM_{10} concentration at the kept stations in the Evaluation 2 period

	Csepel	Erzsébet	Gergely	Gilice	Kőrakas	Kosztolányi	Pesthidegkút	Széna	Overall
Mean	128.7500	70.7500	91.5000	71.5000	80.2500	82.7500	75.5000	88.0000	86.1250
Median	131.5000	67.5000	90.5000	70.5000	81.5000	76.0000	69.5000	85.5000	79.0000
Mode	75	53	57	44	53	60	61	66	75
Maximum	177	95	128	101	105	119	102	115	177
Minimum	75	53	57	44	53	60	61	66	44
Standard deviation	37.0700	15.5302	25.4214	20.6942	18.9918	22.2078	16.1632	17.5926	28.6637
No. of observations	4	4	4	4	4	4	4	4	32
No. of days in the period	4	4	4	4	4	4	4	4	32

Period: from 2017-01-23 to 2017-01-26. Measurement unit: μgm^{-3}

Table 7: Descriptive statistics of the imputed meteorological data.

	Average temp.	Max. temp.	Min. temp.	Wind blow	Wind power
Mean	11.5184	16.7689	6.6875	22.2634	2.5003
Median	11.8000	16.9000	6.7000	20.0000	2.0000
Mode	15.1000	24.5000	10.3000	20.0000	1.0000
Maximum	29.5000	39.1000	21.8000	93.0000	19.0000
Minimum	-12.4000	-8.6000	-18.3000	0	0
Standard deviation	8.3680	10.0831	7.2705	10.4151	2.4786
No. of imputed observations	11	11	11	57	53
No. of days in the period	1492	1492	1492	1492	1492

Period: from 2013-01-01 to 2017-01-31. Abbreviations: temp.=temperature; max.=maximum, min.=minimum. Measurement units: temperature $^{\circ}C$; wind: kmh^{-1}

Table 8: Distances between stations

Station[a]	**Csepel**	***Csillaghegy***[b]	**Erzsébet**	**Gergely**	**Gilice**	Honvéd	Káposztas	**Kőrakas**	**Kosztolányi**	**Pesthidegkút**	**Széna**	Teleki	Tétény
Csepel	0.0000	21.5755	10.7150	8.5007	7.3685	13.1205	20.1563	15.9641	8.8042	20.0309	12.4482	9.8732	6.1733
Csillaghegy	21.5755	0.0000	10.9596	16.6841	21.1464	8.4895	5.6690	9.8479	13.3200	7.0305	9.8285	11.8489	21.2193
Erzsébet	10.7150	10.9596	0.0000	8.4435	12.1679	2.9367	10.8138	8.6944	2.5796	9.9274	2.2637	2.4498	10.6783
Gergely	8.5007	16.6841	8.4435	0.0000	4.4821	8.9252	13.4677	8.4711	8.7575	18.0035	10.6930	6.0762	12.9696
Gilice	7.3685	21.1464	12.1679	4.4821	0.0000	13.1822	17.8780	12.7689	11.7322	22.0294	14.4178	10.0406	13.2295
Honvéd	13.1205	8.4895	2.9367	8.9252	13.1822	0.0000	7.8791	6.3408	5.5155	9.2003	3.4849	3.3785	13.6035
Káposztás	20.1563	5.6690	10.8138	13.4677	17.8780	7.8791	0.0000	5.2417	13.3887	11.7883	10.8503	10.4618	21.4449
Kőrakás	15.9641	9.8479	8.6944	8.4711	12.7689	6.3408	5.2417	0.0000	10.9594	14.0502	9.8205	7.2344	18.4233
Kosztolányi	8.8042	13.3200	2.5796	8.7575	11.7322	5.5155	13.3887	10.9594	0.0000	11.2513	3.6851	3.8804	8.1062
Pesthidegkút	20.0309	7.0305	9.9274	18.0035	22.0294	9.2003	11.7883	14.0502	11.2513	0.0000	7.7737	12.0037	17.7106
Széna	12.4482	9.8285	2.2637	10.6930	14.4178	3.4849	10.8503	9.8205	3.6851	7.7737	0.0000	4.6433	11.3911
Teleki	9.8732	11.8489	2.4498	6.0762	10.0406	3.3785	10.4618	7.2344	3.8804	12.0037	4.6433	0.0000	11.2493
Tétény	6.1733	21.2193	10.6783	12.9696	13.2295	13.6035	21.4449	18.4233	8.1062	17.7106	11.3911	11.2493	0.0000

[a] Highlighted bold names indicate kept stations.
[b] Meteorological station, source of meteorological data
Distances are measured in km. Source: own computation based on addresses to coordinates converter (MapsEasy.com, 2017) and Python implementation of spherical distance haversine formula of Dunn (2011)

C. Figures

Figure 1: Violin plots of PM_{10} level in the training period

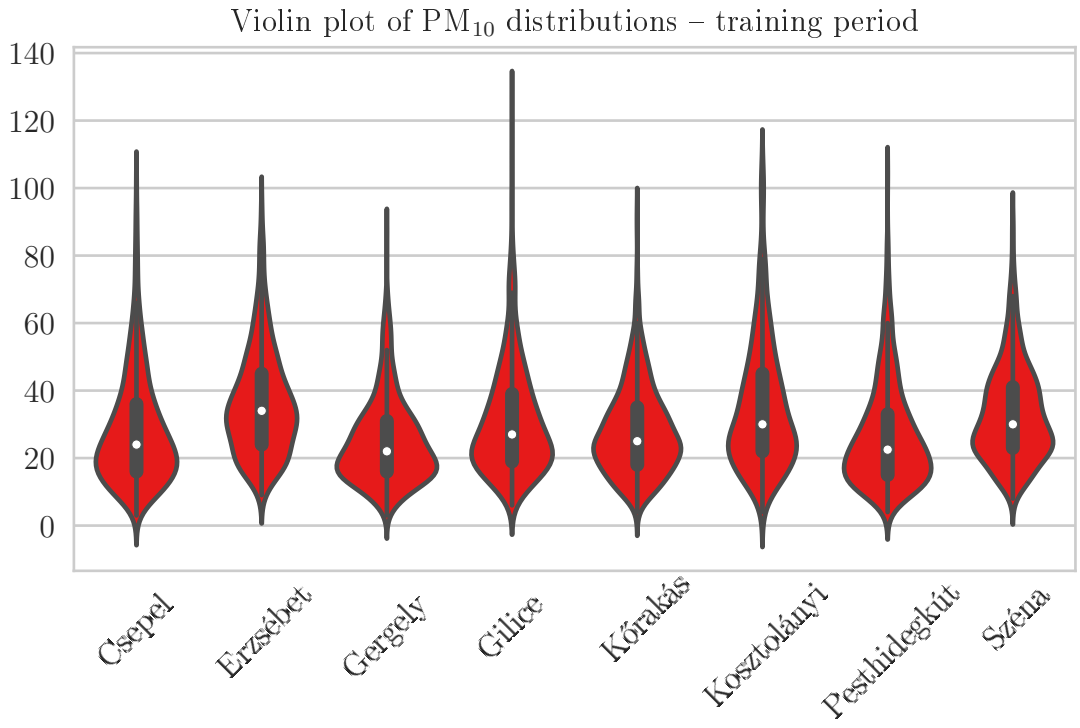


Figure 2: Violin plots of PM_{10} level in the test period

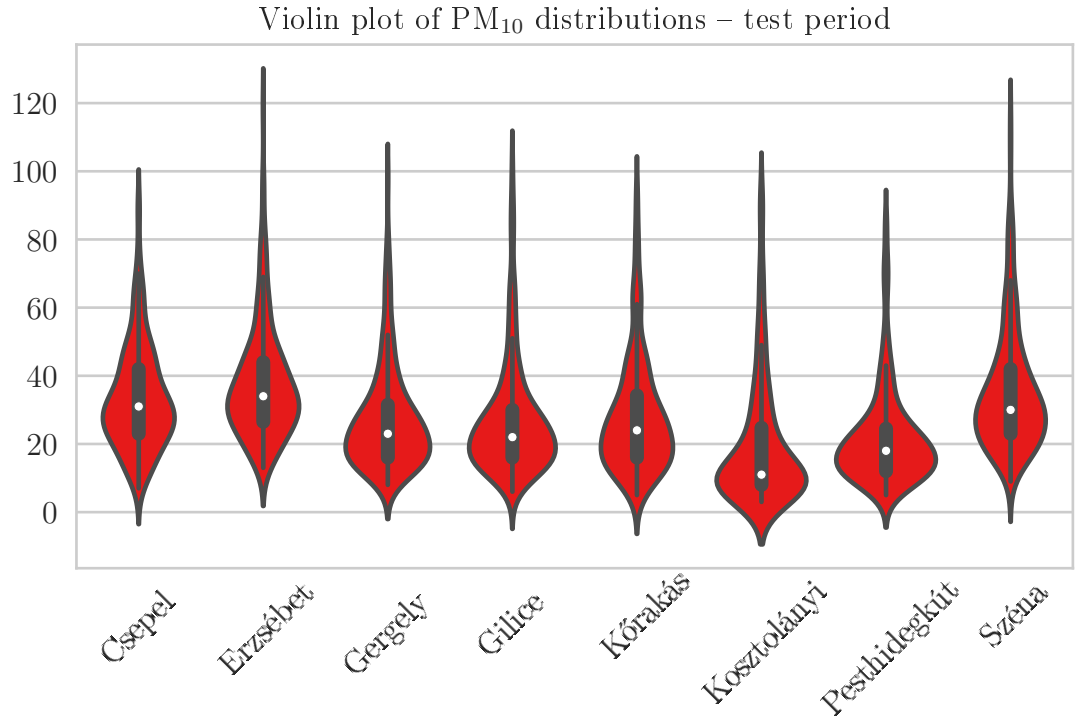


Figure 3: Between-station correlation of PM_{10} level in the training period

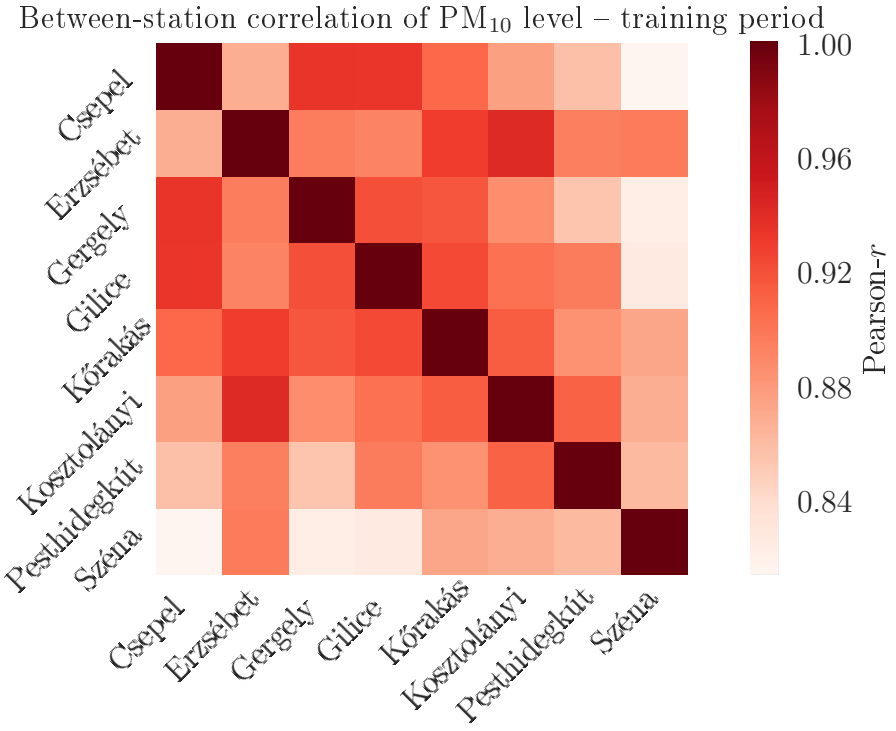


Figure 4: Distances between stations

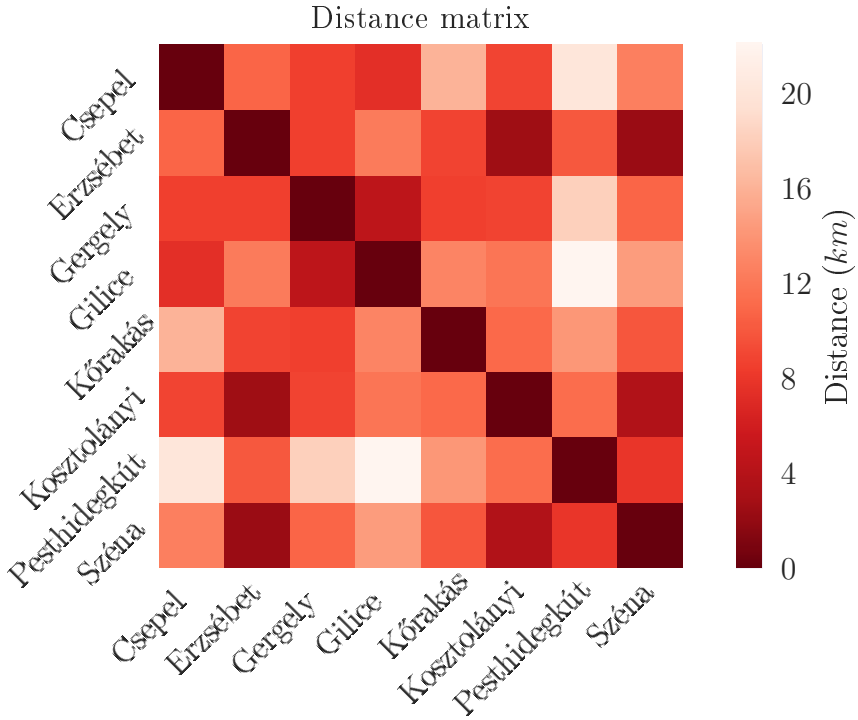


Figure 5: Temperature time series

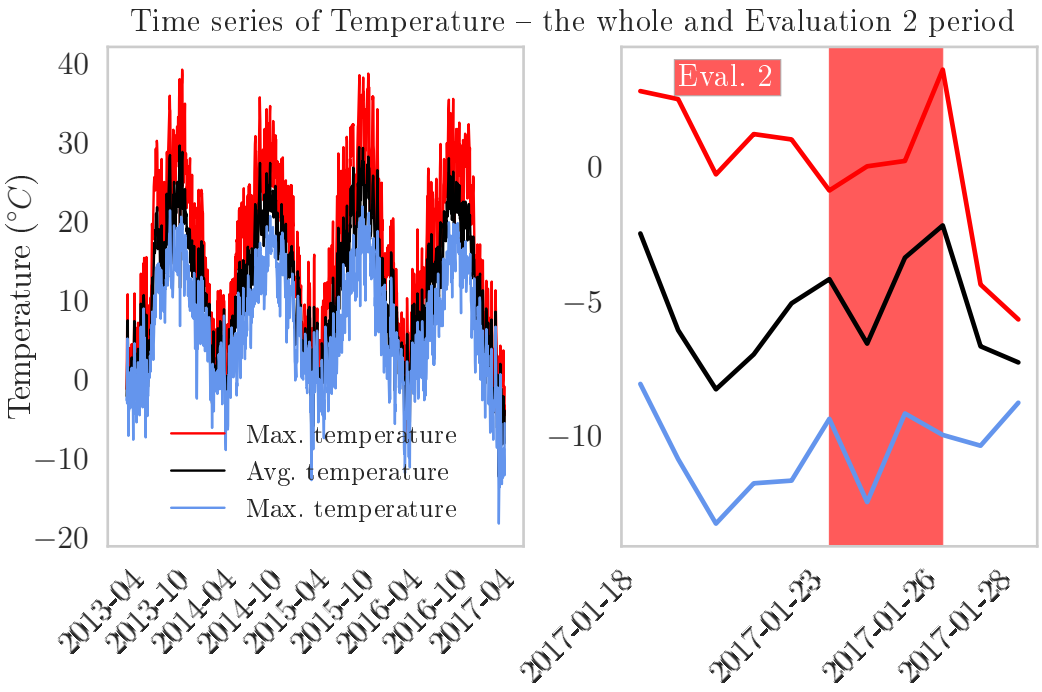
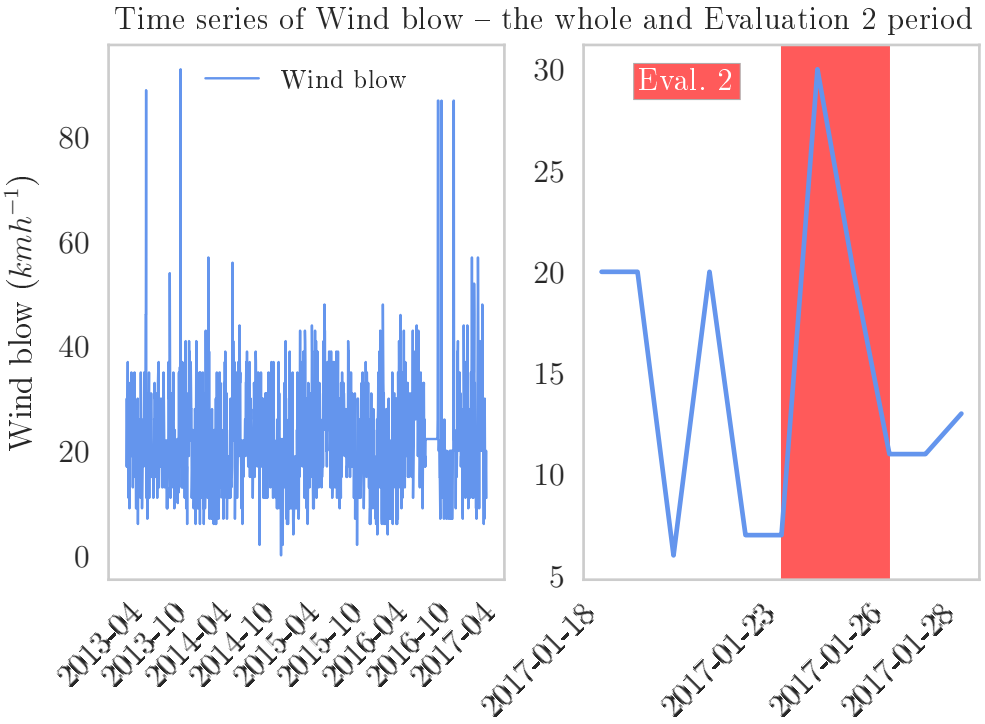


Figure 6: Wind blow time series



References

Dunn, M. (2011). Haversine formula Python implementation. http://stackoverflow.com/questions/4913349/haversine-formula-in-python-bearing-and-distance-between-two-gps-points. Last access: March 23, 2017.

MapsEasy.com (2017). Addresses to coordinates. http://www.mapseasy.com/adress-to-gps-coordinates.php. Last access: March 22, 2017.