ANNEX I

Detailed specification of the presented pollution level maps

Spatial maps are constructed from the results of measurements at the individual locations using and combining a wide range of information (ČHMÚ 2020d). Uncertainties of individual maps depend mainly on the density of the network of monitoring stations and the uniformity of coverage of the territory of the Czech Republic by stations, as well as on the uncertainties of individual measurements, model inputs, model calculations and a way used in constructing the spatial maps. Maps have the least uncertainty near measuring stations. Although the uncertainties of some particular maps are quite high, these relate to estimates of the air pollution field that adequately correspond to the background data used and the state of current knowledge. The uncertainties of maps must be taken into account when interpreting them.

The following paragraphs describe the background sources used for construction of the air pollution maps for 2019 and the specifications of the individual maps presented in this yearbook.

1. Data employed

a. **Measured air pollution data**; The annual characteristics of the measured data from the AQIS database are used.

b. **Outputs from the dispersion models**; Outputs from the following models are used

CAMx – Eulerian model, resolution 2.3 x 2.3 km, 2019:

- meteorology: ALADIN 2019 model in 2.3 x 2.3 km resolution
- anthropogenic emissions for the territory of the Czech Republic: REZZO 1 and 2 stationary sources – reporting for 2018 updated by reporting for 2019 available as of 4 February 2020; REZZO 3 areal sources – local heating (background data 2018, degree-days 2019), agriculture – breeding and agriculture activities (2018), surface brown coal mines (2018), black coal mines (2017), quarries – surface mining (2017), fugitive emissions from production of coke, iron and steel, foundries and other resources in 2017, landfills (2018), construction activities (2018), use of solvents (2018); REZZO 4 mobile sources – road transport according to the Road and Motorway Directorate census (2016), off-road transport (2017), Václav Havel Airport in Prague (2016)
- anthropogenic emissions for the territory of Poland: detailed emissions for 2015 provided under the LIFE-IP MAŁOPOLSKA[1] project by GIOS (Główny Inspektorat Ochrony Środowiska) – area sources and KOBiZE (Krajowy Ośrodek Bilansowania i Zarządzania Emisjami) – point sources
- anthropogenic emissions for the rest of the territory: basic substances – CAMS-REG-AP v3.11[2] for 2016 (Granier 2019); benzo[*a*]pyrene (2017) (EMEP/CEIP 2019)
- biogenic VOC emissions from plants and NO from soil: the MEGAN v2.1 model (GUENTER et al. 2012)
- boundary conditions – minimum values from the CAMx model

CAMS ensemble forecast[3] – median of nine Euler models, resolution 0.1 x 0.1°, year 2019 (meteorology: ECWMF 2019, emission: CAMS-REG-AP v2.2.1 2015; see METEO-FRANCE (2019) for details)

SYMOS – Gaussian model, resolution 1 x 1 km (reference points in 250 x 250 m grid in a built-up area and 500 x 500 m grid outside a built-up area averaged into a grid of 1 x 1 km), 2019 (meteorology: wind roses 2019 from the ALADIN model in the 2.3 x 2.3 km grid and four altitude levels, anthropogenic emissions: for the Czech Republic as for the CAMx model (emissions from construction activities were not included); outside the Czech Republic CAMS-REG-AP v3.1);

The latest outputs that were available from the particular models at the time of preparing the yearbook were always used.

c. **Emissions from traffic**: resolution 1 x 1 km, source: the Road and Motorway Directorate census (2016)

d. **Elevation**: resolution 1 x 1 km, source: ZABAGED, SALSC.

e. **Population density**: resolution 1 x 1 km, source: CSO.

1 Project LIFE14 IPE/PL/000021. WWW: https://powietrze.malopolska.pl/en/life-project/
2 https://permalink.aeris-data.fr/CAMS-REG-AP
3 https://www.regional.atmosphere.copernicus.eu/

2. Estimate of uncertainty

The uncertainty in relation to the relevant map was assessed using the cross-validation method, see Horálek et al. (2007). Estimation of the concentrations at the measuring sites is always created by leaving out the given measurement using the other data, thus objectively estimating the quality of the map outside the measuring site. This approach was used repeatedly for all the measuring sites. The estimated values were compared with the measured values using the **root-mean-square error (RMSE)** or the **relative root-mean-square error (RRMSE)**.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{Z}(s_i) - Z(s_i))^2} \qquad RRMSE = \frac{RMSE}{\frac{1}{N}\sum_{i=1}^{N} Z(s_i)}.100$$

where

$Z(s_i)$ is the measured value of the concentration at the i[th] point,
$\hat{Z}(s_i)$ is the estimate at the i[th] point using the other data,
N is the number of monitoring stations.

For calculation reasons, the estimate of the uncertainty was calculated only for interpolation of the residuals; thus the overall uncertainty of the map is somewhat greater. It should also be noted that this is the median uncertainty of the whole map; the spatial distribution of the uncertainty was not estimated.

3. Parameters of the individual maps

For the maps of the individual pollutants, the Tab. 1–8 below present the supplementary quantities used in the linear regression model and their parameters (c, a1, a2, ...), the interpolation parameters using kriging (range, nugget, partial sill) and the inverse distance values (IDW — inverse distance weighted) and, for most maps, the root-mean-square of the error (RMSE) in the map is also given. These parameters are always given for the individual pollution layers (rural, urban, traffic).

a. Suspended particulate matter PM_{10}: The maps were constructed using 55 rural (without distinguishing background and industrial), 88 urban and suburban background and 25 traffic stations. The results of measurements at seven urban and suburban industrial stations were taken into account only in their immediate vicinity (Tab. 1, Annex 1).

b. Suspended particulate matter $PM_{2.5}$: The maps were constructed using 26 rural (without distinguishing background and industrial), 52 urban and suburban background and 18 traffic stations. The results of measurements at four urban and suburban industrial stations were taken into account only in their immediate vicinity. The uncertainty in the map was not calculated because of the mapping methodology (Tab. 2, Annex I). This is because PM_{10} maps were used as supplementary quantities — due to strong regression relation between PM_{10} and $PM_{2.5}$ the uncertainty estimates would be underestimated.

c. Benzo[*a*]pyrene: The maps were constructed using 11 rural, and 36 urban and suburban background and traffic stations. The results of measurements at six industrial stations were taken into account only in their immediate vicinity. Due to the lack of measuring stations in small settlements, the estimation of uncertainty in rural areas is only indicative (Tab. 3, Annex I).

d. Nitrogen dioxide and nitrogen oxides: The maps for NO_2 were constructed using 25 rural (without distinguishing background and industrial), 45 urban and suburban background and 21 traffic stations. The results of measurements at 8 urban and suburban industrial stations were taken into account only in their immediate vicinity. The maps for NO_X were constructed using 24 rural, 45 urban and suburban background and 21 traffic stations (Tab. 4, Annex I).

e. Tropospheric ozone: The maps of the 26[th] highest maximum daily 8-hour running average were constructed on the basis of 24 rural and 31 urban and suburban stations. The maps for AOT40 were constructed using 23 rural and 25 urban and suburban background stations (Tab. 5, Annex I).

f. Benzene: The maps were constructed using 6 rural, and 22 urban and suburban background stations. The results of measurements at 4 industrial and 7 traffic stations were taken into account only in their immediate vicinity (Tab. 6, Annex 1).

g. Heavy metals: The maps for arsenic were constructed using 14 rural and 44 urban and suburban stations (without distinguishing between background, traffic and industrial stations). The cadmium map was constructed using 58 stations (without distinguishing according to type). The uncertainty in the cadmium map was estimated without the Tanvald municipality and its immediate vicinity because the high absolute values at this location would cause distortion of the overall uncertainty of the map. The high relative uncertainty of the cadmium map is related to the low cadmium values over most of the territory (Tab. 7, Annex I).

h. Sulphur dioxide: The map of the 4[th] highest 24-hour concentration was constructed using 25 rural (without distinguishing background and industrial) and 27 urban and suburban background stations. The results of measurements at 2 traffic and 7 industrial stations were taken into account only in their immediate vicinity. The maps of the annual or winter averages were constructed using 27 and 25, respectively, rural (without distinguishing background and industrial) and 28 and 25, respectively urban and suburban background stations. The results of measurements at 2 traffic stations and 7 and 4, respectively, industrial stations were taken into account only in their immediate vicinity (Tab. 8, Annex I).

The numbers of stations also include foreign (German and Polish) stations that were used in the creation of some maps.

Tab. 1 PM_{10} map parameters

Linear regression model + interpolation of residuals	Annual average			36th highest daily average		
	rural areas	urban background	traffic	rural areas	urban background	traffic
c (constant)	7.2	19.7	11.0	8.4	35.0	19.5
a1 (model CAMx)	1.73	0.54	1.13	1.65	0.49	0.95
a2 (altitude)	-0,0053	-0.0136		-0.0054	-0.0276	
range [km]	26	18	25	34	28	0
nugget	0	3.6	0	0	17	19
partial sill	3.6	5.6	5.8	12	7	9
weight IDW		1			1	
RMSE [μg.m^{-3}]	**1.8**	**2.6**	**1.8**	**4.1**	**5.2**	**4.1**
relat. RMSE [%]	**11**	**13**	**8**	**14**	**14**	**11**

Tab. 2 $PM_{2.5}$ map parameters

Linear regression model + interpolation of residuals	Annual average		
	rural areas	urban background	traffic
c (constant)	-0.2	-1,1	0.9
a1 (rural map of PM_{10})	0.55		
a2 (urban background map of PM_{10})		0.79	
a3 (traffic map of PM_{10})			0.66
a4 (model CAMx)	0.56		
range [km]	90	110	150
nugget	0.7	0.7	0
partial sill	0.0	0.2	3.2
weight IDW	1	1	

The urban and rural layers were combined using the limits of the classification intervals (ČHMÚ 2020d): α1 = 200 inhabitants per km^2, α2 = 1000 inhabitants per km^2. The background and traffic layers were combined using the limits of the classification intervals (ČHMÚ 2020): τ1 = 3 tonnes p.a. per km^2, τ2 = 8 tonnes p.a. per km^2 (for PM_{10} and $PM_{2.5}$ maps), or τ1 = τ2 = 10 tonnes p.a. per km^2 (for NO_2 and NO_x maps), where the PM_{10} and $PM_{2.5}$ maps were based on SPM emissions, while the NO_2 and NO_x maps were based on NO_x emissions[4].

4 For the spatial maps of NO_2 and NO_x, the traffic layer was used only in cities, while outside of cities in territories with NO_x > 10 tonnes p.a. per km^2 the layers were used from all the urban, suburban, rural and traffic stations.

Tab. 3 Benzo[*a*]pyrene map parameters

Linear regression model + interpolation of residuals	Annual average	
	rural areas	urban background
c (constant)	-0.5	-2.4
a1 (urban map of $PM_{2,5}$)		0.17
a2 (model CAMx)	1.76	0.71
a3 (model SYMOS – local heating emission only)		0.73
range [km]	70	8
nugget	0	0
partial sill	0.12	0.2
weight IDW		
RMSE [$\mu g.m^{-3}$]	**> 0.3**	**0.5**
relat. RMSE [%]	**> 40**	**43**

Tab. 4 NO_2 and NO_x map parameters

Linear regression model + interpolation of residuals	NO_2 – annual average			NO_x – annual average		
	rural areas	urban background	traffic	rural areas	urban background	traffic
c (constant)	8.4	18	21.5	11.1	28.6	87.5
a1 (model SYMOS NO_2)	4.5	2.1				
a2 (model SYMOS NO_2 – REZZO 4)			4.2			
a3 (model SYMOS NO_x)				1.9	0.9	
a3 (model SYMOS NO_x – REZZO 4)						34.9
a4 (altitude)	-0.01	-0.02		-0.01	-0.03	
weight IDW	1	1	1	1	1	1
RMSE [$\mu g.m^{-3}$]	**1.3**	**3.1**	**6.1**	**2.2**	**7.1**	**18,4**
relat. RMSE [%]	**15**	**19**	**22**	**20**	**28**	**34**

Tab. 5 Ground-level ozone map parameters

Linear regression model + interpolation of residuals	26th highest maximum daily 8-hour average		AOT40 exposure index	
	rural areas	urban background	rural areas	urban background
c (constant)	-5.3	32.2	10 915	11 238
a1 (model CAMS)	1,2	0.9	0.7	0.5
weight IDW	1	1	1	1
RMSE [$\mu g.m^{-3}$]	**4.1**	**3.4**	**2 789**	**2 939**
relat. RMSE [%]	**3**	**3**	**15**	**17**

Tab. 6 Benzene map parameters

Linear regression model + interpolation of residuals	Annual average	
	rural areas	urban background
c (constant)	0.3	-0.1
a1 (model CAMx)	4.3	9.8
weight IDW	1	1
RMSE [μg.m⁻³]	**0.3**	**0.3**
relat. RMSE [%]	**29**	**25**

Tab. 7 Arsenic and cadmium map parameters

Linear regression model + interpolation of residuals	Arsenic – annual average		Cadmium – annual average
	rural areas	urban background	whole map
c (constant)	-0.6		
a1 (rural map of PM_{10})	0.094		
range [km]	320	15	15
nugget	0	0	0
partial sill	0.1	0.5	0.3
weight IDW			
RMSE [μg.m⁻³]	**0.2**	**0.6**	**0.2**
relat. RMSE [%]	**23**	**41**	**92**

Tab. 8 SO_2 map parameters

Linear regression model + interpolation of residuals	4th highest daily average		Annual average		Winter average	
	rural areas	urban background	traffic	rural areas	urban background	traffic
c (constant)	10.1	5.8	2.6	2.6	2.8	2.1
a1 (model CAMx)	0.4	0.5	0.6	0.5	0.6	0.5
weight IDW	3	2	1	1	2.4	1.6
RMSE [μg.m⁻³]	**7.9**	**6.9**	**2**	**1.7**	**2.1**	**1.6**
relat. RMSE [%]	**45**	**41**	**42**	**33**	**40**	**30**