ANNEX I

Detailed specifications of the presented pollution level maps

Spatial maps are constructed from the results of measurements at individual locations, using and combining a wide range of information (CHMI 2022d). Uncertainties of individual maps depend mainly on the density of the network of monitoring stations and the uniformity of coverage of the national territory by these stations, as well as on the uncertainties of individual measurements, model inputs, model calculations and the methods 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 are associated with 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 2021 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 dispersion models**. Outputs from the following models are used:

CAMx — Eulerian model, resolution 2.3 × 2.3 km, 2021:

- meteorology: ALADIN 2021 model in 2.3 × 2.3 km resolution
- anthropogenic emissions for the territory of the CR for 2020, unless otherwise stated: REZZO 1 and 2 stationary sources — reporting for 2020 updated by reporting for 2021 available as of 3 April 2022; REZZO 3 areal sources — local heating (background data 2020, degree-days 2021), agriculture — breeding (2019 and 2020) and agriculture activities, brown coal and black coal mines (2021), quarries — surface mining, fugitive emissions from production of coke, iron and steel, foundries and other resources, landfills, construction activities, use of solvents; REZZO 4 mobile sources — road transport according to the Road and Motorway Directorate census (2016), off-road transport, Václav Havel Airport in Prague (2016, updated for 2021 according to CO_2 emissions ratio for the Czech Republic (EUROCONTROL 2022)), other international airports (2020, updated for 2021 according to CO_2 emission ratio for the Czech Republic (EUROCONTROL 2022))
- anthropogenic emissions for Poland for 2019: Central emission database for air quality modelling in Poland (KOBiZE 2022)
- anthropogenic emissions outside the CR and Poland: basic substances – CAM S-REG-AP v4.2-ry[1] for 2019 (Kuenen et al. 2021); benzo[*a*]pyrene, cadmium, and lead (2019) (EMEP/CEIP 2022; these cadmium and lead emissions were also used for the territory of Poland)
- biogenic VOC emissions from plants and NO from soil: the MEGAN v2.1 model (GUENTER et al. 2012)
- boundary conditions – in time and space, variables from the global WACCM model (NCAR 2022)

SYMOS — Gaussian model, resolution 1 × 1 km (reference points in a 250 × 250 m grid in built-up areas and a 500 × 500 m grid outside built-up areas, averaged into a grid of 1 × 1 km); outside the CR 1 × 1 km, 2021 (meteorology: wind roses 2021 from the ALADIN model in the 2.3 × 2.3 km grid and at 120, 330, 500, and 700 m height levels according to the effective height of the source, anthropogenic emissions: as for the CAMx model (except for emissions from the EMEP/CEIP database)). For $PM_{2.5}$, the model was smoothed in a 3 × 3 km grid.

CAMS ensemble forecast[2] – median of the Eulerian model ensemble, resolution 0.1 x 0.1°, year 2021 (meteorology: ECWMF 2021, emission: CAMS-REG-AP; see COPERNICUS (2022) for details)

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

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

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

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

1 https://permalink.aeris-data.fr/CAMS-REG-AP
2 https://www.regional.atmosphere.copernicus.eu/

2. Estimates of uncertainty

The uncertainty in relation to the relevant map was assessed using the cross-validation method; see Horálek et al. (2007). Estimations of the concentrations at measuring sites was 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).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\dot{Z}(s_i) - Z(s_i))^2}$$

$$RRMSE = \frac{RMSE}{\frac{1}{N}\sum_{i=1}^{N} Z(s_i)} \cdot 100$$

Where $Z(s_i)$ is the measured value of the concentration at the i^{th} point,
$\dot{Z}(s_i)$ is the estimate at the ith 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 generally 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 Tables 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), 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), 87 urban and suburban background and 27 traffic stations. The results of measurements at six 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 29 rural (without distinguishing background and industrial), 52 urban and suburban background and 19 traffic stations. The results of measurements at seven 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 the 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 42 urban and suburban stations (without distinguishing background, traffic, and industrial ones), which were supplemented by eight rural and eight urban and suburban stations whose values were estimated using the measured values of previous years. Concerning the city and rural map layers, an exponential interrelation with the $PM_{2.5}$ city and rural map layer, respectively, was applied. Due to the low number 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 31 rural (without distinguishing background and industrial), 48 urban and suburban background and 22 traffic stations. The results of measurements at 4 urban and suburban industrial stations were taken into account only in their immediate vicinity. The maps for NO_X were constructed using 29 rural, 46 urban and suburban background and 22 traffic stations (Tab. 4, Annex I).

e. **Tropospheric ozone:** The maps of the 26 highest maximum daily 8-hour running averages were constructed on the basis of 24 rural and 29 urban and suburban stations. The measurement results of 3 transport and 2 urban and suburban industrial stations were taken into account only in their immediate vicinity. The maps for AOT40 were constructed using 27 rural and 34 urban and suburban background stations (Tab. 5, Annex I).

f. **Benzene:** The maps were constructed using 6 rural, and 24 urban and suburban background stations. The results of measurements at 2 industrial and 6 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 16 rural and 40 urban and suburban stations (without distinguishing between background, traffic and industrial stations). The cadmium map was constructed using 56 stations (without distinguishing according to type). 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 highest 24-hour concentrations was constructed using 29 rural (without distinguishing background and industrial) and 30 urban and suburban background stations. The results of measurements at 3 traffic and 3 industrial stations were taken into account only in their immediate vicinity. The maps of the annual or winter averages were constructed using 38 and 35, respectively, rural (without distinguishing background and industrial) and 30 urban and suburban background stations. The results of measurements at 3 and 2, respectively, traffic stations and 3 industrial stations were taken into account only in their immediate vicinity (Tab. 8, Annex I).

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)	3.5	13.9	10.7	−5.0	24.2	18.6
a1 (model CAMx)	1.72	0.69	0.88	1.82	0.61	0.73
a2 (altitude)	-0.0055	-0.0073			-0.0139	
range [km]	45	90	5	47	25	5
nugget	0	4.6	0	0	19.0	0
partial sill	2.9	2.2	4.5	11.6	1.3	13.8
weight IDW		1			1	
RMSE [$\mu g \cdot m^{-3}$]	**1.8**	**2.7**	**2.3**	**4.2**	**5.2**	**3.5**
relat. RMSE [%]	**11**	**14**	**10**	**14**	**15**	**9**

Tab. 2 $PM_{2.5}$ map parameters

Linear regression model + interpolation of residuals	Annual average		
	rural areas	urban background	traffic
c (constant)	0.7	−2.4	−2.1
a1 (rural map of PM_{10})	0.62		
a2 (urban background map of PM_{10})		0.86	
a3 (traffic map of PM_{10})			0.79
a4 (model SYMOS)	0.95		
range [km]	10	100	2
nugget	0	1.1	0
partial sill	1.2	0.2	1.9
weight IDW		1	

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

The urban and rural layers were combined using the limits of the classification intervals (CHMI 2022d): $\alpha1 = 200$ inhabitants·km^{-2}, $\alpha2 = 1000$ inhabitants·km^{-2}. The background and traffic layers were combined using the limits of the classification intervals (CHMI 2022d): $\tau1 = 3$ t·km^{-2}·year^{-1}, $\tau2 = 8$ t·km^{-2}·year^{-1} (for PM_{10} and $PM_{2.5}$ maps), or $\tau1 = \tau2 = 10$ t·km^{-2}·year^{-1} (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[3].

3 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 $NO_X > 5$ t·km^{-2}·year^{-1} 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)	0.1	0.1
b1 (constant)	0.2	
b2 (constant)		0.2
a1 (exp(b1*rural map $PM_{2,5}$))	0.04	
a2 (exp(b2*urban map $PM_{2,5}$))		0.03
range [km]	30	6
nugget	0	0
partial sill	0.02	0.25
RMSE [µg·m^{-3}]	**> 0.3**	**0.5**
relat. RMSE [%]	**> 30**	**35**

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)	8.0	16.7	18.8	8.9	25.7	31.3
a1 (model SYMOS NO_2)	4.43	1.50				
a2 (model SYMOS NO_2 – REZZO4)			2.75			
a3 (model SYMOS NO_x)				2.35	0.62	
a3 (model SYMOS NO_x – REZZO4)						1.93
a4 (altitude)	-0.01	-0.01		−0.01	-0.03	
weight IDW	1	1	1	1	1	1
RMSE [µg·m^{-3}]	**1.1**	**2.4**	**5.5**	**2.3**	**5.2**	**16.3**
relat. RMSE [%]	**14**	**16**	**22**	**22**	**23**	**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)	111.7	23.0	16146	8114
a1 (model CAMS)		0.90		0.84
a2 (altitude)	0.01		1.12	
weight IDW	1	1.1	1	1
RMSE [µg·m^{-3}]	**4.0**	**4.2**	**2727**	**2617**
relat. RMSE [%]	**3**	**4**	**16**	**16**

Tab. 6 Benzene map parameters

Linear regression model + interpolation of residuals	Annual average	
	rural areas	urban background
c (constant)	−1.2	−1.1
a1 (model CAMx)	6.87	7.15
weight IDW	1	1.6
RMSE [μg·m⁻³]	**0.1**	**0.3**
relat. RMSE [%]	**16**	**22**

Tab. 7 Arsenic and cadmium map parameters

Linear regression model + interpolation of residuals	Arsen – annual average		Kadmium – annual average
	rural areas	urban background	whole map
c (constant)	−0.8		0.1
a1 (rural map PM_{10})	0.109		
a2 (model CAMx)			1.68
range [km]	130	10	16
nugget	0	0	0
partial sill	0.2	0.4	0.1
RMSE [μg·m⁻³]	**0.3**	**0.6**	**0.2**
relat. RMSE [%]	**38**	**47**	**88**

Tab. 8 SO_2 map parameters

Linear regression model + interpolation of residuals	4^{th} highest daily average		Annual average		Winter average	
	rural areas	urban background	traffic	rural areas	urban background	traffic
c (constant)	0.4	5.6	1.1	2.5	1.2	2.0
a1 (model CAMx)	0.99	0.46	0.89	0.40	0.71	0.43
weight IDW	1.5	1.7	1	1	1	1
RMSE [μg·m⁻³]	**4.7**	**6.3**	**1.1**	**1.5**	**1.2**	**1.0**
relat. RMSE [%]	**43**	**45**	**31**	**35**	**31**	**22**