
Hierarchical Clustering and Dissimilarity Polygon Analyses

Optimizing the Polish Deposition Network

Joana Soares, Wenche Aas, Sabine Eckhardt and
Cristina Guerreiro



Contents

Contents	3
Summary	5
1 Hierarchical clustering as data analysis methodology	6
1.1 Station dissimilarity	7
1.2 Modelled Dissimilarity Regions	9
1.3 Source receptor relationships based FLEXPART backward calculations for deposited material.....	9
2 Observational and model data used in the assessment	10
3 Results	13
3.1 Modelled regions (clusters): regions representing the areas of representativeness of a single station.....	13
3.1.1 SO _x deposition.....	14
3.1.2 OXN deposition.....	17
3.1.3 RDN deposition.....	20
3.1.4 Cd deposition.....	23
3.1.5 Hg deposition.....	26
3.1.6 Pb deposition.....	29
3.2 Station dissimilarity analysis (dendrograms).....	31
3.2.1 SO ₄ deposition.....	32
3.2.2 NH ₄ deposition.....	33
3.2.3 Ca deposition.....	34
3.2.4 Cl deposition.....	35
3.2.5 Na deposition.....	36
3.2.6 K deposition.....	37
3.2.7 H ⁺ deposition	38
3.2.8 Mg ²⁺ deposition	39
3.2.9 Relative ranking of stations	40
3.3 Source receptor analysis for selected stations for deposited SO _x	40
4 Analysis of the clustering results	43
4.1 Areas of representativeness of a single station	43
4.2 Station dissimilarity analysis based on observations	44
4.3 Combining modeling and observations.....	48
5 Overall conclusions	49
6 References	50
Annex	52
A.1 Modelled regions (clusters): 25 regions representing the areas of representativeness of a single station based on (1-R) x EuD metrics	52
A.1.1 SO _x	52
A.1.2 OXN.....	54
A.1.3 RDN.....	56
A.1.4 Cd.....	58
A.1.5 Hg	60

A.1.6	Pb	62
A.2	Modelled regions (clusters): 21 and 25 regions representing the areas of representativeness of a single station based on 1-R (temporal variation)	64
A.2.1	SOx	64
A.2.2	OXN	65
A.2.3	RDN	66
A.2.4	Cd	67
A.2.5	Hg	68
A.2.6	Pb	69
A.3	Modelled regions (clusters): 21 and 25 regions representing the areas of representativeness of a single station based on EuD (magnitude variation).....	70
A.3.1	SOx	70
A.3.2	OXN	71
A.3.3	RDN	72
A.3.4	Cd	73
A.3.5	Hg	74
A.3.6	Pb	75
A.4	Station dissimilarity analysis based on the temporal variation (1-R).....	76
A4.1	SO4.....	76
A4.2	NH4	76
A4.3	Ca	77
A.4.4	Cl	77
A.4.5	Na.....	78
A.4.6	K	78
A.4.7	H+	79
A.4.8	Mg	79
A.5	Station dissimilarity analysis based on the magnitude of deposition levels (EuD)	80
A5.1	SO4.....	80
A5.2	NH4	80
A5.3	Ca	81
A.5.4	Cl	81
A.5.5	Na.....	82
A.5.6	K	82
A.5.7	H+	83
A.5.8	Mg	83

Summary

This report describes the work done to assess the potential re-design of the current deposition monitoring network in Poland. Hierarchical clustering analysis was used to assess the areas (polygons) with similar deposition patterns across the country. This statistical method determines the inherent or natural groupings of datasets, and/or to provide a summarization of data into groups using different metrics to assess the (di)similarity. The metrics are based on the correlation, to assess the temporal similarity, the Euclidean distance, to assess the magnitude similarity, and the combination of both. The statistical analysis uses measurement data provided by the Chief Inspectorate for Environmental Protection (GIOS) and modelling data provided by the Meteorological Synthesizing Centre-West (MSC-W) of the European Monitoring and Evaluation Programme (EMEP) model for acidic compounds and the GLEMOS model by the Meteorological Synthesizing Centre-East (MSC-E) for the heavy metals. Additionally, source receptor relationships based FLEXPART backward calculations for deposited material was calculated.

Analysing the hierarchical clustering results based on the modelled data clearly identified regions that represent the area of representativeness of a single station (a large number of cells are part of a single cluster). This means that a single station located within this region will monitor an air mass that is similar temporally and has similar deposition levels. The methodology has also identified smaller source areas indicating point/area sources of the compounds analysed. Acidic compounds show a more homogeneous distribution of the areas representing a single station than metals. The results for the metals seem to indicate the difference between west-east and north-south, with smaller clusters south/east part of the country. Thus, indicating a higher number of sources in the south/east part of Poland. However, independently of the chemical compounds, the results seem to indicate that the north-west side of the country needs fewer monitoring sites than the south-east part of the country, independently of the compound.

Analysing the hierarchical clustering results based on the observations show which stations can be potentially monitoring similar pollution regimes (redundant) and stations that measure unique air masses. The analysis also points to the need to keep stations monitoring the Polish coast, particularly for shipping-related pollution. A station in Gdansk seems to have the most variety of air masses and could be a unique monitoring location (also indicated by the footprint from FLEXPART).

The results based on the modelling (at station location) and measurements show there may be some discrepancy between the results based on model and observations. This discrepancy may be due to most of the sites being located at urban background areas when the model does not represent urban areas accurately, especially runs concerning metals (~40km resolution).

Hierarchical Clustering and Dissimilarity Polygon Analyses

Optimizing the Polish Deposition Network

1 Hierarchical clustering as data analysis methodology

Hierarchical clustering is a well-established statistical method to determine the inherent or natural groupings of datasets, and/or to provide a summarization of data into groups. Hierarchical clustering compares data to determine how they differ (a type of *dissimilarity analysis*). The data may be the time series of observations (measured at monitoring stations) or predictions from an air-quality model. Each time series is compared to every other time series, record-wise (see Figure 1), meaning that every record of a pair of time series is compared using a "dissimilarity metric". It has been demonstrated that a combination of metrics can be beneficial for optimizing monitoring networks.

The dissimilarity metrics are:

- (1) 1-R, where R is the Pearson correlation coefficient (Eq 1.) (Galmarini and Solazzo, 2015);
- (2) the Euclidean distance (EuD, Eq 2.) (Soares et al, 2018b);
- (3) The combination of both ((1-R) x EuD, Eq. 3).

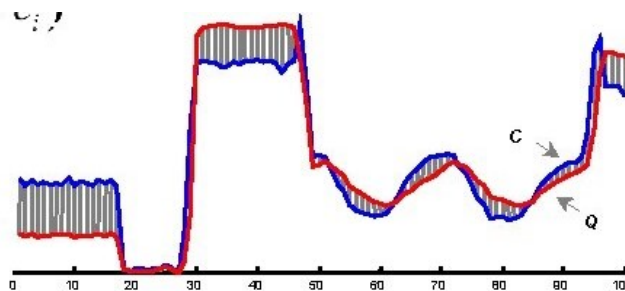


Figure 1: Comparison between two time series, record-wise. To aid in the understanding of Eq. 1, 2 and 3.

$$r_{cq} = \frac{\sum_j (c_j - m_c) (q_j - m_q)}{\sqrt{\sum_j (c_j - m_c)^2 \sum_j (q_j - m_q)^2}} \quad (1)$$

$$EuD_{cq} = \sqrt{\sum_j (c_j - q_j)^2} \quad (2)$$

$$(1 - R) \times EuD_{cq} = (1 - r_{cq}) \times EuD_{cq} \quad (3)$$

The metric based on correlation (1-R) assesses dissimilarities associated with the changes in the temporal variations in concentration, while the metric based on the Euclidian distance (EuD) assesses dissimilarities on the basis of the magnitude of the observations over the time period of the analysis. Soares et al. (2018a) has shown that the multiplication of these two metrics - (1-R) x EuD - allows simultaneously assessing correlations in terms of time variation and pollution levels, providing the

most comprehensive results for assessing air quality monitoring networks. The report will focus the analysis on the **combined metric**.

The hierarchical clustering analysis is an iterative process. The first step of the analysis consists of producing a dissimilarity matrix cross-comparing every single time series (observation or modelled results) available. This matrix results from calculating the metrics based on the available data, where each possible pair of time series will be compared record-wise. The next step is to identify the pair with the highest similarity level and re-evaluate the dissimilarity matrix for all remaining stations/grid-cells. The re-evaluation is done by averaging the latest cluster and the remaining stations/grid-cells' dissimilarity. The averaging method is the general averaging method of Næs et al. (2010). The process is iterated until all the stations/grid-cells have been paired. This iterative pairing process results in a "cluster" of similarity that includes the different stations/grid-cells. Initially, every single station/grid-cell is considered a single cluster. As the hierarchical clustering process goes on, the number of clusters is reduced. The process is completed when all stations/grid-cells have been clustered in a single cluster.

The outcome of the methodology indicates the (dis)similarity of stations/grid-cells, at a certain level of the chosen metric, across the whole data set. The higher the dissimilarity level between a station/grid-cell or cluster of sampling points/grid-cells, the more dissimilar those pairs are. The dissimilarity level value depends on the metric chosen. The values calculated based on the correlation metric can vary from 0 to 2, with values over 1 meaning that the time series anti-correlate. The values for the other two metrics range from zero to a value that strongly depends on the average deposition levels and the number of records analysed. The outcome of the analysis also strongly depends on the quality of the reported data, so that random errors in the observations can potentially change the results (Soares et al., 2018).

The methodology provides different results depending on the underlying data. Here in this report, we present:

- (1) dendrograms based on observations to discuss the similarity of stations currently monitoring deposition;
- (2) spatial distribution of the group of stations based on the dendrograms;
- (3) ranking of stations based on the dendrograms;
- (4) maps of dissimilarity regions (clusters) based on modelling results. The methodology assumes that every single grid cell of the model domain is a potential station location. Thus, the regions shown in the maps are areas of representativeness of a single station.

These means of showing the results are described in the subsections below.

1.1 Station dissimilarity

Hierarchical clustering results are often displayed as dendrograms (see Figure 2) – these are diagrams that track the value of the dissimilarity metric as each data record becomes part of a cluster. Dendrograms show the pattern of linkages between the data series while clustering occurs and their level of dissimilarity. Dendrograms thus resemble the root system of a tree, with the most similar stations forming the lowest level of the smallest roots, and the two least similar clusters being linked at the top of the diagram as the trunk of the tree. Vertical lines on the dendrogram represent the difference in the level of dissimilarity between consecutive stages of clustering. The horizontal lines show which time series or clusters of time series have been linked at a given level of the dissimilarity metric.

The dendrograms allow assessing how stations group in terms of similarity quickly. Another outcome of the analysis of dendrograms is to visualise how these groups are distributed spatially. We can request to "cut" the dendrogram across and see how many stations are linked together and map those stations colour-coded according to the cluster they fall into.

Since the clustering process is tracked (which stations clusters at which level of dissimilarity), the methodology allows ranking of the stations based on their degree of dissimilarity. The stations which join at the lowest values of the dissimilarity metric are the most similar, those with the highest values of the dissimilarity metric when joining a cluster are the least similar. The stations with the most similar records are potentially more redundant or the first stations considering relocation. The stations with the least similar records should be considered to maintain in the same location.

Note that the dendrograms depend on the individual chemical or parameter that was clustered; potential redundancies for one chemical at one station may not be the same as for another chemical at the same station. Depending on the purpose of the optimization, each chemical can be analysed separately or by combining different chemicals. Either way, the purpose of the methodology is to determine whether there are stations that may be redundant and/or extraneous and whether there are gaps in the monitoring.

We offer a simple example (see Figure 2) to better understand the methodology applied.

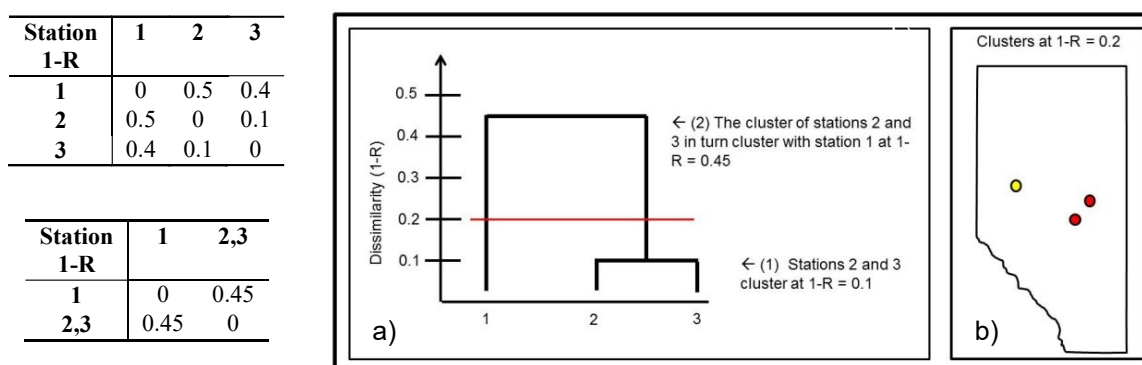


Figure 2: Example of clustering of three stations: supporting tables to establish the dissimilarity between the stations (left); a) dendrogram and b) map of spatial distribution of clusters.

In this example, data for three different hypothetical stations are collected. The values of 1-R are calculated between each pair of station records to measure their level of dissimilarity. The data from stations 1 and 2 have a 1-R value of 0.5, stations 1 and 3 have a 1-R value of 0.4, and stations 2 and 3 have a 1-R value of 0.1. The lowest level of dissimilarity is thus between stations 2 and 3, and they are combined to become the first cluster, at a 1-R level of 0.1 (a horizontal line joins stations 2 and 3 in the dendrogram of Figure 2). The averages of 1-R between this cluster and the other stations are then calculated; in this case $(0.5 + 0.4)/2 = 0.45$, the second table of Figure 2.1. Stations 2 and 3 thus cluster at $1 - R$ of 0.1, and the remaining station, 1, clusters at 1-R of 0.45. The second horizontal line of the diagram portion of Figure 2.1 shows the connection between the initial cluster between stations 2 and 3 and the final cluster with station 1. The result is a 3 station dendrogram (Figure 2).

In conclusion, the station dissimilarity analysis's main outcomes are the dendrograms and tables of relative rankings of the (dis)similarity between station records. Maps with the station's location and distribution of the clusters at a certain level of dissimilarity can also be provided.

1.2 Modelled Dissimilarity Regions

Hierarchical clustering of air quality model output (Soares et al., 2018a,b) is used to map the degree of dissimilarity in predicted values across all or part of the model's gridded domain. Each model grid cell is treated as a potential station location, and the analysis otherwise proceeds identically to the observation station application described above. At a given level of the clustering metric, the model grid-cells contained within each cluster are equally similar. A map of these clusters, colour-coded by cluster, thus shows regions over which a single monitoring station would be able to represent the entire region, to the value of the clustering metric used. Hence, two points within different regions are dissimilar, and two points within the same region are similar.

These dissimilarity maps are thus mapping the area of representativeness of a certain number of stations defined a priori, based on a number that an expert judgment considers sustainable for a monitoring network. Since the number of regions is user-defined, the statistical similarity within a region (or dissimilarity between regions) presented here is relative, as opposed to representing an absolute threshold of similarity.

A region containing multiple stations indicates the potential for redundancy, and hence provides a focus for discussion, while additional information is needed to determine the extent to which stations are truly redundant and/or extraneous. Regions that contain multiple stations suggest possible redundancy since these stations have relatively similar simulated depositions. In contrast, regions containing no stations suggest an area that could be a monitoring gap. This form of analysis thus provides information on the extent to which existing monitoring network stations are spatially representative of the region around them. The analysis also identifies potential gaps in an existing monitoring network, as model-generated clusters containing no current monitoring stations.

Note that the regions depend on the individual chemical or parameter that was clustered; potential redundancies for one chemical at one station may not be the same as for another chemical at the same station.

1.3 Source receptor relationships based FLEXPART backward calculations for deposited material

The Lagrangian particle dispersion model FLEXPART version 10.4 (Pisso et al., 2019) was used to obtain source-receptor relationships (SRRs) for some of the background stations located in different regions. FLEXPART calculates transport of virtual particles using mean winds and random motions representing turbulence. For sub-grid-scale moist convective transport, FLEXPART uses the scheme of Emanuel and Zivkovic-Rothman (1999), as described in Forster et al. (2007). As meteorological input, the operational meteorological data of the European Centre for Medium Range Weather Forecasts (ECMWF) was ingested into FLEXPART for this study. We used the reanalysis data at a resolution of $1^\circ \times 1^\circ$, every 3 hours, and with 138 vertical levels.

For the year of 2019 we performed monthly backward simulations from each location, using FLEXPART's capacity to quantify SRRs between emissions and deposition in backward mode (Eckhardt et al, 2017). This method was already used in several studies for the quantitative interpretation of ice cores (McConnell et al., 2019). One backward simulation for dry and one for wet deposition was run for each month using 500 000 particles for wet deposition. The particles were traced backward in time for 20 days. The wet scavenging scheme uses three-dimensional fields of the cloud water content and precipitation data from the re-analysis and distinguishes between in-cloud and below-cloud scavenging. Similar to the widely used backward mode for atmospheric concentrations (e.g. Seibert and Frank, 2004), we obtain spatially resolved emission sensitivities (Eckhardt et al., 2017). The

emission sensitivities (i.e., source-receptor relationships, SRR), written out on a three-dimensional grid, represent the deposition at the receptor ($\mu\text{g}/\text{m}^2/\text{s}$) that would result from a unit emission of 1 kg/s in a respective model grid cell. The output grid had a resolution of $1^\circ \times 1^\circ$ and the lowest level (often called the footprint layer) extended from the surface to 100 m a.g.l., as most emissions occur near the surface. Deposition fluxes at the receptor are then obtained by multiplying the SRR with the spatially resolved fluxes taken from a suitable emission inventory (ECLIPSE V6) and integrating the product over the entire globe. In our analysis we used monthly SRRs in combination with sulphur containing emissions.

2 Observational and model data used in the assessment

The analysis was applied to gridded output from an air quality model, to generate dissimilarity regions as a screening tool for identifying potential redundancies and gaps in the current deposition network. The deposition data available was estimated by the air quality model run by the Meteorological Synthesizing Centre-West (MSC-W) of the European Monitoring and Evaluation Programme (EMEP) model (Simpson et al., 2012) for acidic compounds and the GLEMOS model employed by the Meteorological Synthesizing Centre-East (MSC-E) for the heavy metals (Travnikov et al, 2017). Both models simulate meteorology, emissions of numerous air pollutants, pollutant transport/dispersion, atmospheric chemistry, and atmospheric deposition in three dimensions. Both models have been evaluated against deposition observations, with the latest evaluation reported in (MSC-W & CCC, 2021; Klein et al, 2021; Ilyin et al., 2021; Strizhkina et al., 2021).

The data available was retrieved from the MSC-W (2022) and MSC-E (Ilyin et al., 2022) and the time period considered spans from 2015 to 2019 to accommodate meteorology variability on the deposition levels (as discussed in a joint meeting on February 24th, 2022) . The modelling results available for the analysis are described in Table 1.

The observational data used in the analysis comprise of data measured between 2015 and 2019 at 24 stations: 22 from the Chemistry and Precipitation Network and two from the suburban and rural background network. Table 2 describes the stations with data available for this analysis. Note that the data availability extends the period here analysed but decided to assess for the same time period as the modeling data, as agreed on the joint meeting held on March 17th, 2022.

Table 1: Ddata description.

source	Type	Compounds	Time resolution	Spatial resolution
MSC-W	Model Gridded	oxidized sulfur (SO ₂ , SO ₄) oxidized nitrogen (NO ₂ , PAN, HNO ₃ , HONO, NO ₃) reduced nitrogen (NH ₃ , NH ₄)	2015-2019 Monthly	Europe 0.1x0.1°
MSC-E	Model Gridded	Cl, Hg, Pb	2015-2019 Monthly	Europe 0.4x0.4°
GIOS	Observations, 24 stations	SO ₄ ⁺² , NH ₄ ⁺ , Ca ⁺² , Cl ⁺² , H ⁺ , K, Mg ⁺² , Na ⁺²	2015-2019 Monthly	Poland

Table 2 station list used in this analysis. Stations where observations are available, or footprint runs are available are described in the last two columns.

network	Name of the locality	Station name	name in maps/dendrograms	UE code	Y WGS84	X WGS84	area	observations	footprint
precipitation	Białystok	Chemizm-Białystok	Bialystok	Białystok	53.100000	23.166667	urban	x	
precipitation	Chojnice	Chemizm-Chojnice	Chojnice	Chojnice	53.716667	17.533333	urban	x	
precipitation	Gdańsk	Chemizm-Gdańsk_Swibno	GdanskSwibno	Gdańsk	54.333667	18.934306	urban	x	
precipitation	Gorzów Wielkopolski	Chemizm-Gorzów Wlkp.	GorzowWlkp	Gorzów Wielkopolski	52.750000	15.283333	urban	x	
precipitation	Kalisz	Chemizm-Kalisz	Kalisz	Kalisz	51.783333	18.083333	urban	x	
precipitation	Zakopane	Chemizm-Kasprowy Wierch	KasprowyWierch	Zakopane	49.233333	19.983333	rural	x	x
precipitation	Katowice	Chemizm-Katowice	Katowice	Katowice	50.233333	19.033333	urban	x	
precipitation	Łeba	Chemizm-Łeba	Leba	Łeba	54.750000	17.533333	suburban	x	x
precipitation	Legnica	Chemizm-Legnica	Legnica	Legnica	51.200000	16.200000	urban	x	
precipitation	Lesko	Chemizm-Lesko	Lesko	Lesko	49.466667	22.350000	urban	x	
precipitation	Nowy Sącz	Chemizm-Nowy Sacz	NowySacz	Nowy Sącz	49.627167	20.688611	urban	x	
precipitation	Olsztyn	Chemizm-Olsztyn	Olsztyn	Olsztyn	53.766667	20.416667	urban	x	
precipitation	Poznań	Chemizm-Poznań	Poznan	Poznań	52.416667	16.850000	urban	x	
precipitation	Racibórz	Chemizm-Racibórz	Raciborz	Racibórz	50.050000	18.200000	urban	x	
precipitation	Sandomierz	Chemizm-Sandomierz	Sandomierz	Sandomierz	50.700000	21.716667	urban	x	
precipitation	Karpacz	Chemizm-Śnieżka	Sniezka	Karpacz	50.736389	15.739722	rural	x	x
precipitation	Sulejów	Chemizm-Sulejów	Sulejow	Sulejów	51.350000	19.866667	urban	x	
precipitation	Suwałki	Chemizm-Suwałki	Suwalki	Suwałki	54.133333	22.950000	urban	x	
precipitation	Świnoujście	Chemizm-Świnoujście	Swinoujście	Świnoujście	53.916667	14.233333	urban	x	
precipitation	Toruń	Chemizm-Toruń	Torun	Toruń	53.050000	18.583333	urban	x	
precipitation	Włodawa	Chemizm-Włodawa	Wlodawa	Włodawa	51.550000	23.533333	urban	x	
precipitation	Zielona Góra	Chemizm-Zielona Góra	ZielonaGora	Zielona Góra	51.933333	15.533333	urban	x	
suburban and rural background	Zielonka	Bory Tucholskie	Zielonka	PL0077A	53.662136	17.933986	rural		x
suburban and rural background	Jarczew	IMGW-Jarczew	Jarczew	PL0002R	51.814367	21.972375	rural	x	x
suburban and rural background	Diabla Góra	KMŚ Puszcza Borecka	PuszczaBorecka	PL0005R	54.124819	22.038056	rural	x	
ZMSP-IMNE	Karkonosze	KARKONOSZE	Karkonosze	14ZM	50.790699	15.5121	rural		x
ZMSP-IMNE	Parsęta	PARSĘTA	Parsęta	06ZM	53.778999	16.4974	rural		x

network	Name of the locality	Station name	name in maps/dendrograms	UE code	Y WGS84	X WGS84	area	observations	footprint
ZMSP-IMNE	Wolin	WOLIN	Wolin	11ZM	53.946899	14.4756	rural		x
ZMSP-IMNE	Kampinos	KAMPINOS	Kampinos	08ZM	52.285801	20.454399	rural		x
ZMSP-IMNE	Łysogóry	ŁYSOGÓRY	Łysogóry	09ZM	50.8619	21.052799	rural		x
ZMSP-IMNE	Puszcza Borecka	PUSZCZA BORECKA	Puszcza Borecka	01ZM	54.125099	22.038099	rural		x
ZMSP-IMNE	Poznań-Morasko	RÓŻANY STRUMIENI	Poznań-Morasko	13ZM	52.463001	16.941401	urban		
ZMSP-IMNE	Roztocze	ROZTOCZE	Roztocze	12ZM	50.590801	22.9979	rural		
ZMSP-IMNE	Wigry	WIGRY	Wigry	05ZM	54.060902	23.0135	rural		x
ZMSP-IMNE	Koniczynka	KpKoniczynka	Koniczynka	07ZM	53.080601	18.684	rural		x
ZMSP-IMNE	Szybark	MpSzybbaGorl	Szybark	10ZM	49.633499	21.1166	rural		x

3 Results

A hierarchical clustering analysis using the (1-R) x EuD metric was performed based on deposition data from 2015 to 2019 for the compounds listed in Table 1 and cover the whole Polish territory. In Annex 1, we include the same results for the two metrics separately, 1-R (Eq. 1) and EuD (Eq. 2), for interpreting how the temporal and magnitude variation individually influence the outcome presented in this Section. Note that the metrics are calculated for every pair of stations/grid-cells available, where every single record (monthly deposition) of the time series is compared between the pair applying Eq. 3.

The results presented here include:

- (1) regions (clusters) representing the area where a single station can be located. Within that region, the station should be measuring similar temporal and magnitude variation. These regions are based on modelled monthly deposition.
- (2) deposition maps showing the average of the modelled (MSC-W) monthly deposition for 5-years.
- (3) station dissimilarity assessment based on dendrograms and a relative ranking of stations considering the level of dissimilarity the stations are clustering at. The station dissimilarity assessment is based on the monthly mean deposition observations.
- (4) Monthly analysis of source regions for deposited sulfate and their seasonal variability. Combination of those with a SO₂ emission inventory to show the most relevant sources for different stations.

3.1 Modelled regions (clusters): regions representing the areas of representativeness of a single station

The regions are mapped and shown in the following subsections. The maps provide the areas representing the spatial representativeness of a single station across Poland based on the temporal and magnitude variation of modelled deposition across Poland.

As discussed in a joint meeting on February 24th, 2022, an expert judgment on the number of stations considers somewhere between 21 and 25 stations for a deposition monitoring network in Poland a reasonable number of stations. For the sake of consistency and clarity, this report depicts only the minimum number regions indicating the areas of representativeness of a single station. The figures and tables for the maximum are presented in Section A.1 in the Annex. The minimum and maximum maps based on the temporal and magnitude variation separately are available in Section A2 and A3 in Annex, respectively.

The maps are overlaid with the station location described in Table 2. The decision to overlaying ZMNE-INE stations on the map is because these stations are located in rural areas, whereas most of the stations with observations available for the analysis describe urban background deposition. A table describing the region (cluster) where a station is located is also provided to aid in the assessment. Note that some stations are right at the edge of the region's boundaries. Thus, the uncertainty related to the cluster where they fall into is higher.

The mapping of the regions (clusters) can be cross-compared with the 5-year monthly deposition average maps based on the modelling results, especially to locate potential source areas.

3.1.1 SOx deposition

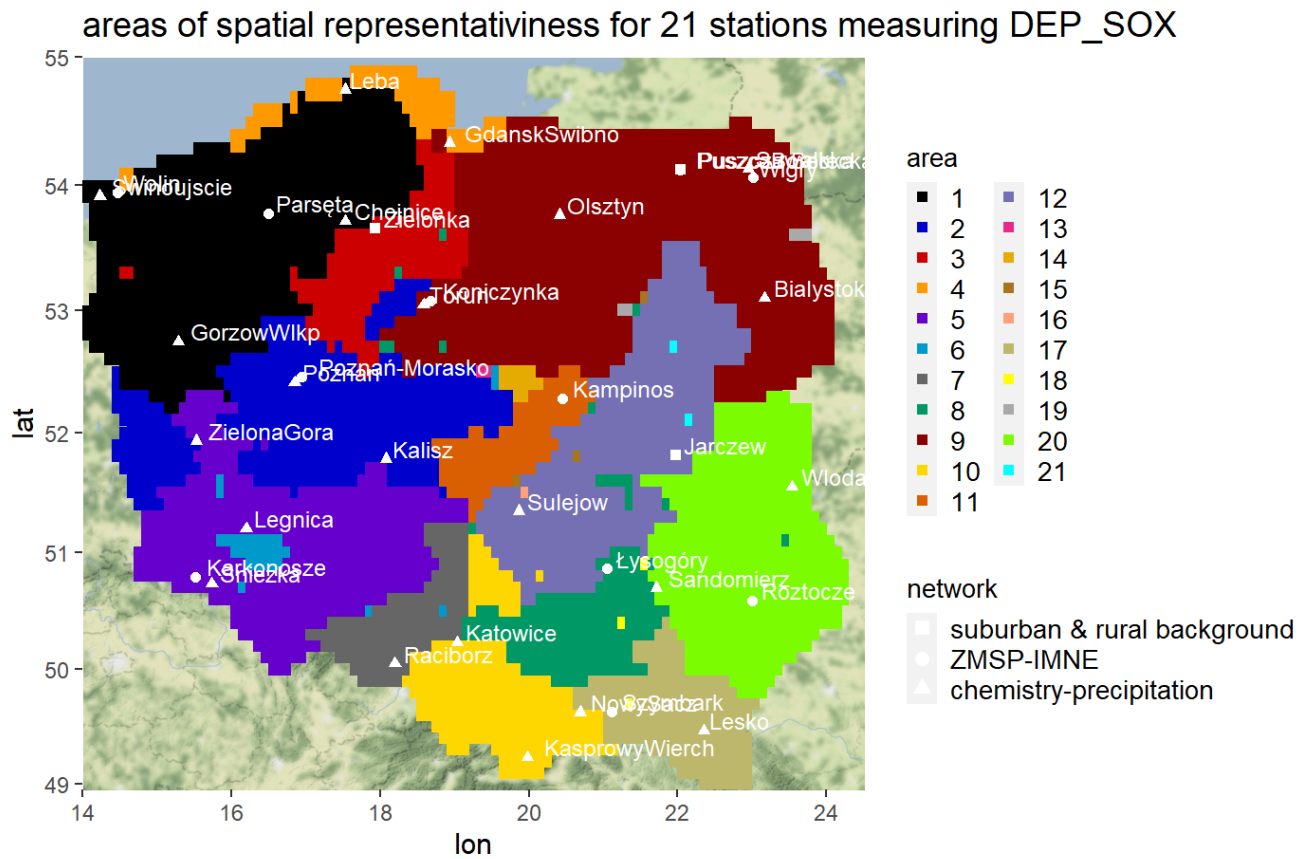


Figure 3: Areas of spatial representativeness for 21 stations measuring SOx deposition.

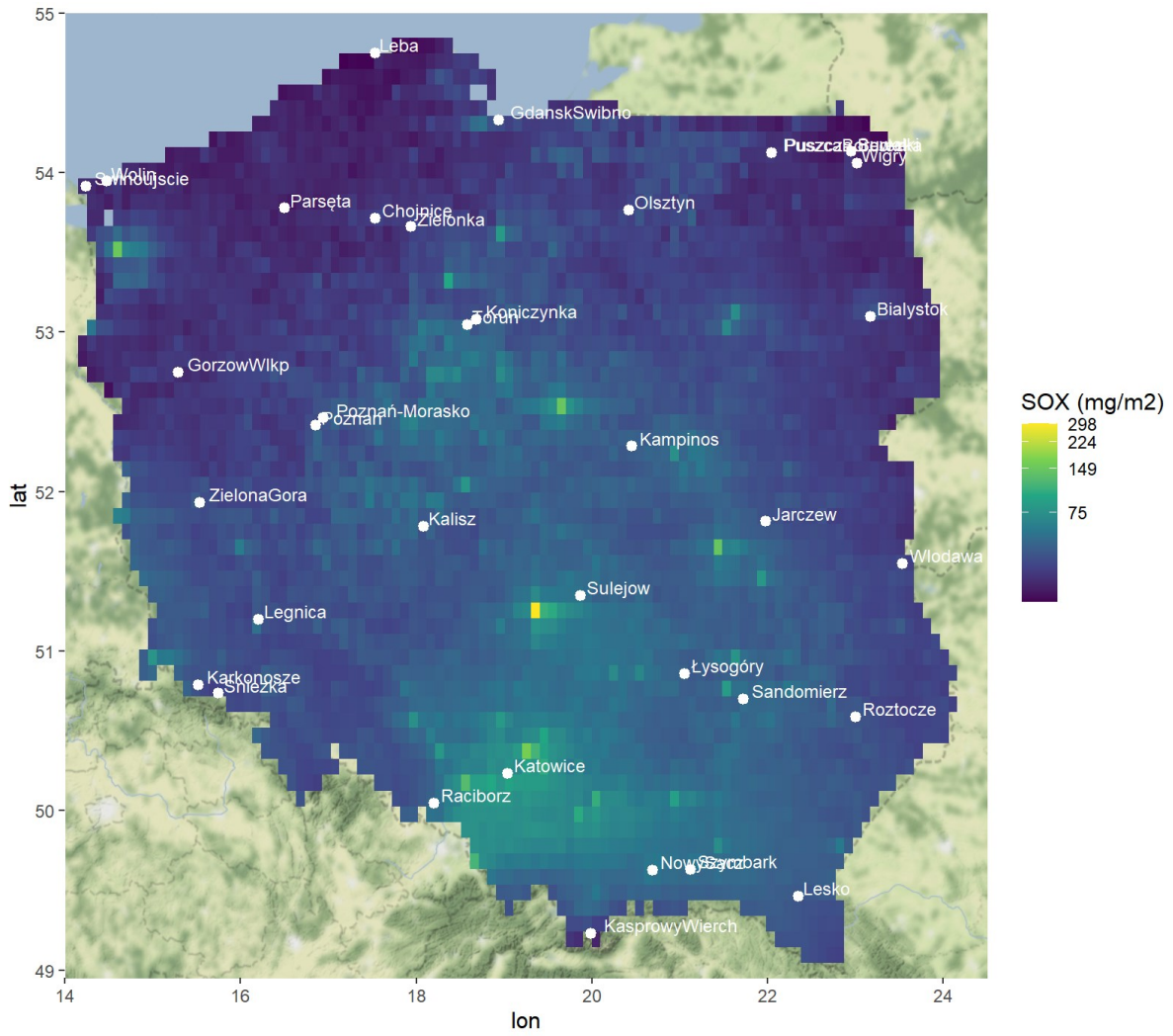


Figure 4: Average SOx monthly deposition between 2015 and 2019 (plotted in logarithmic scale).

Table 3: Cluster information based on the results presented in Figure 3 and Figure 26 for the monitoring stations listed in Table 2.

21 regions	name
1	Chojnice
1	GorzowWlkp
1	Parseta
1	Wolin
2	Poznan
2	Torun
2	ZielonaGora
2	Poznan-Morasko
3	Zielonka
4	Leba
5	Kalisz
5	Sniezka
5	Swinoujscie
5	Karkonosze
6	Legnica
7	Raciborz
8	Sandomierz
8	Lysogóry
9	Białystok
9	GdanskSwibno
9	Olsztyn
9	Suwalki
9	PuszczaBorecka
9	Puszcza Borecka
9	Wigry
9	Koniczynka
10	KasprowyWierch
10	Katowice
10	NowySacz
11	Kampinos
12	Sulejow
12	Jarczew
17	Lesko
17	Szymbark
20	Wlodawa
20	Roztocze

3.1.2 OXN deposition

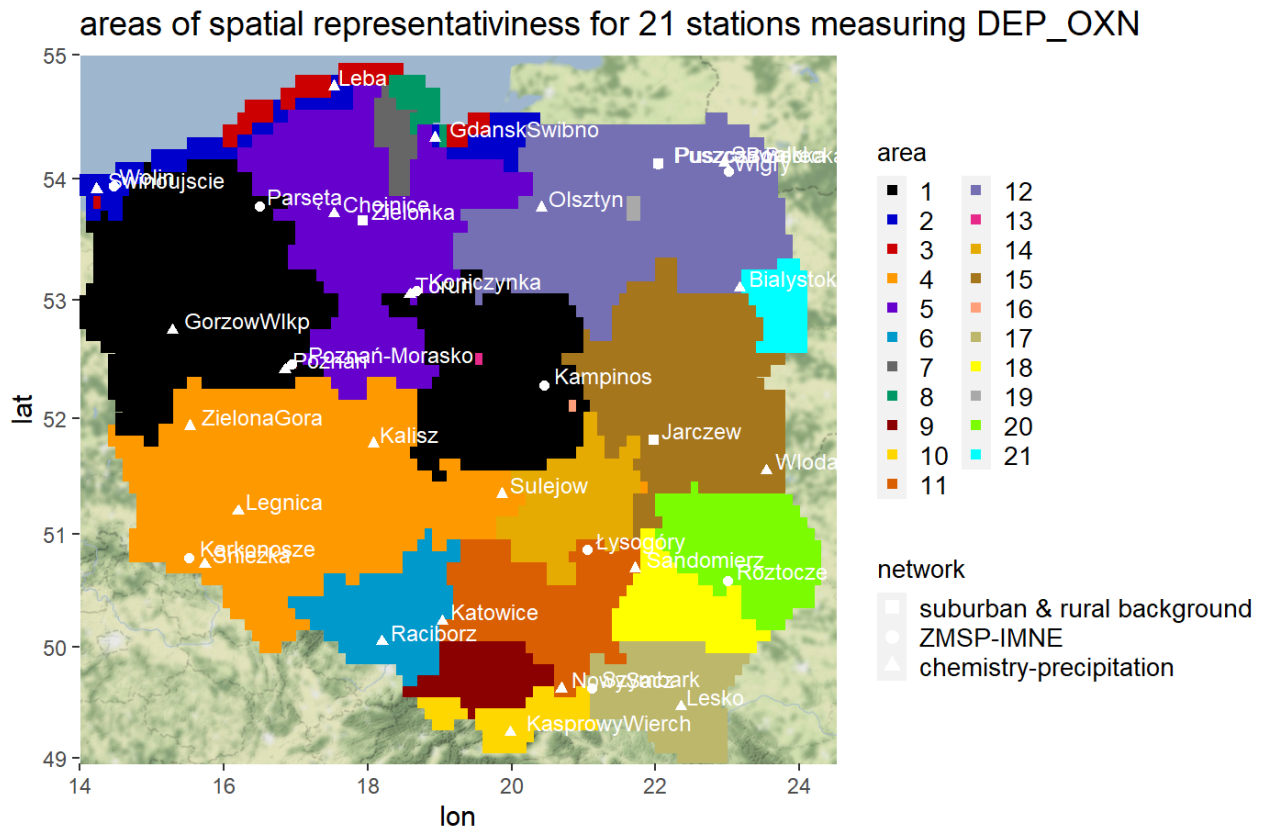


Figure 5: Areas of spatial representativeness for 21 stations measuring OXN deposition.

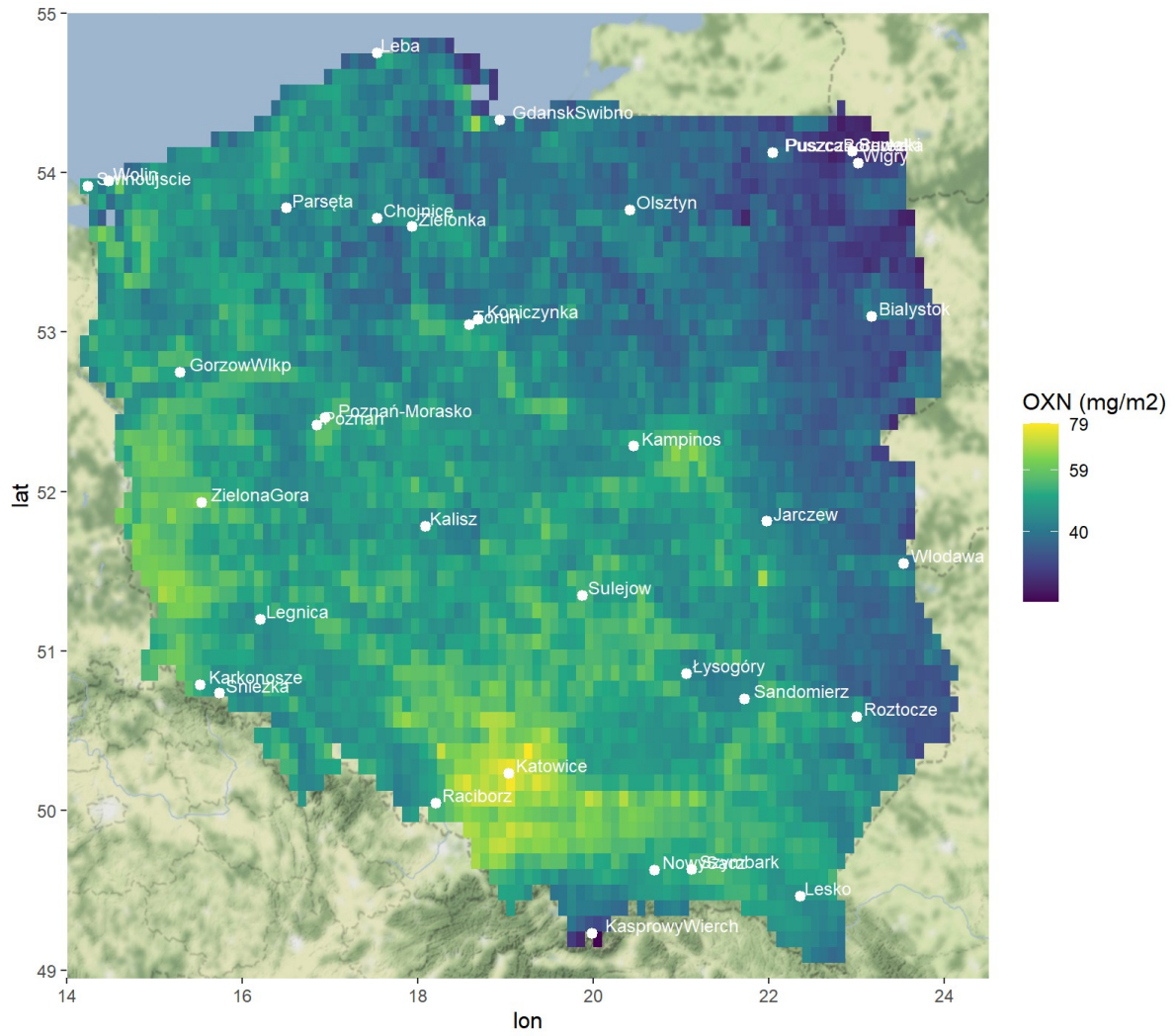


Figure 6: Average OXN monthly deposition between 2015 and 2019 (plotted in logarithmic scale).

Table 4: Cluster information based on the results presented in Figure 5 and Figure 27 for the monitoring stations listed in Table 2.

21 regions	name
1	GorzowWlkp
1	Poznan
1	Torun
1	Parseta
1	Kampinos
1	Poznan-Morasko
2	Wolin
3	Leba
4	Kalisz
4	Legnica
4	Sniezka
4	Sulejow
4	Swinoujscie
4	ZielonaGora
4	Karkonosze
5	Chojnice
5	GdanskSwibno
5	Zielonka
5	Koniczynka
6	Katowice
6	Raciborz
10	KasprowyWierch
10	NowySacz
10	Szymbark
11	Sandomierz
11	Lysogóry
12	Bialystok
12	Olsztyn
12	Suwalki
12	PuszczaBorecka
12	Puszcza Borecka
12	Wigry
15	Wlodawa
15	Jarczew
17	Lesko
20	Roztocze

3.1.3 RDN deposition

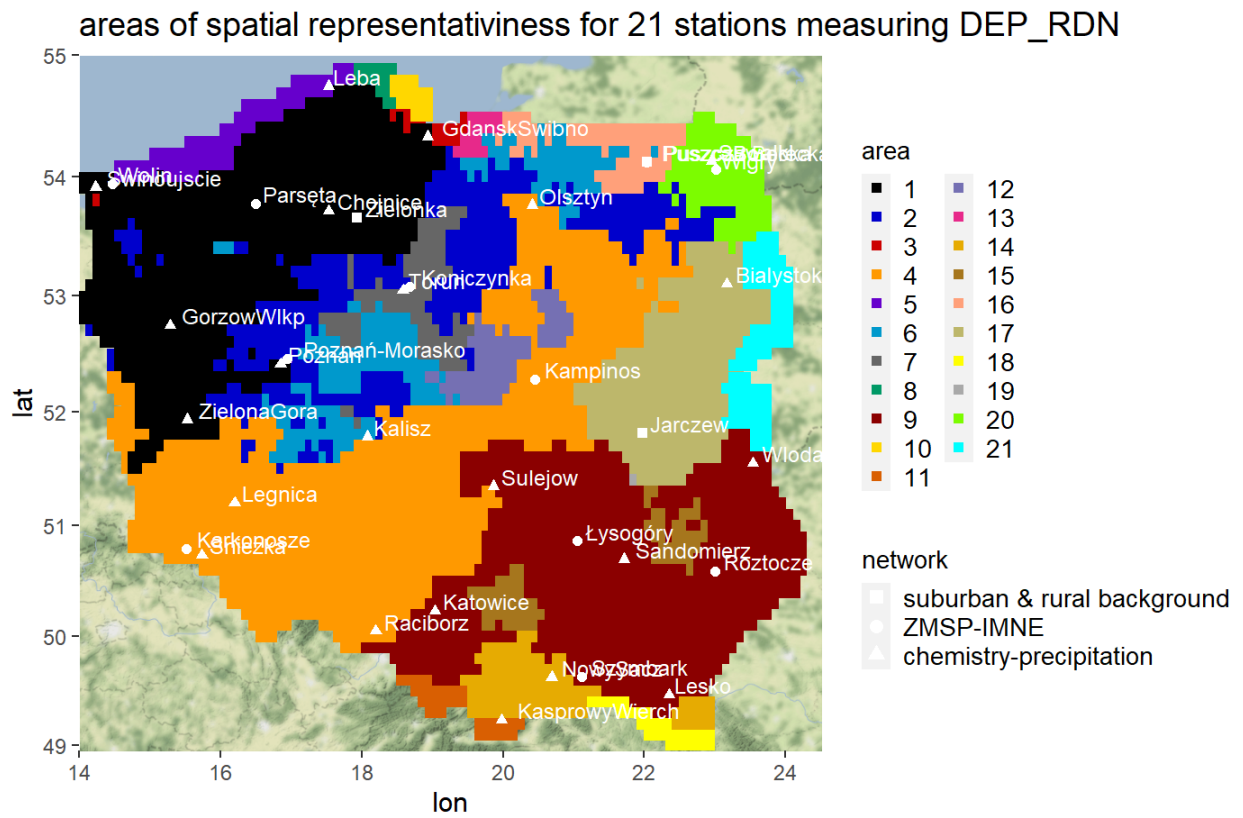


Figure 7: Areas of spatial representativeness for 21 stations measuring RDN deposition.

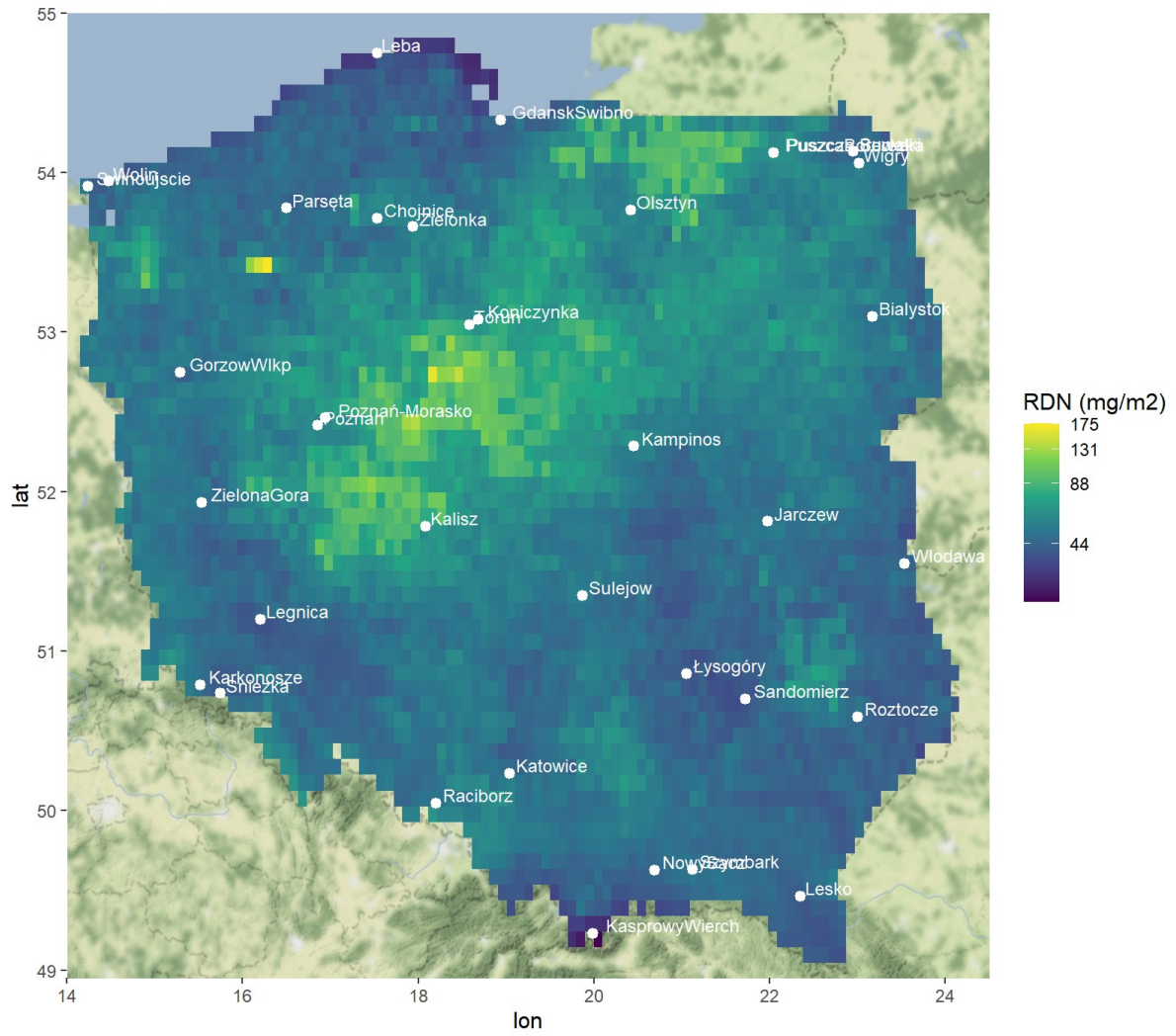


Figure 8: Average RDN monthly deposition between 2015 and 2019 (plotted in logarithmic scale).

Table 5: Cluster information based on the results presented in Figure 7 and Figure 28 for the monitoring stations listed in Table 2.

21 regions	name
1	Chojnice
1	GdanskSwibno
1	GorzowWlkp
1	Poznan
1	ZielonaGora
1	Zielonka
1	Parseta
1	Wolin
1	Poznan-Morasko
2	Koniczynka
4	Kalisz
4	Legnica
4	Olsztyn
4	Raciborz
4	Sniezka
4	Sulejow
4	Swinoujscie
4	Karkonosze
4	Kampinos
5	Leba
7	Torun
9	Katowice
9	Lesko
9	Sandomierz
9	Wlodawa
9	Lysogóry
9	Roztocze
9	Szymbark
11	KasprowyWierch
14	NowySacz
16	PuszczaBorecka
16	Puszcza Borecka
17	Bialystok
17	Jarczew
20	Suwalki
20	Wigry

3.1.4 Cd deposition

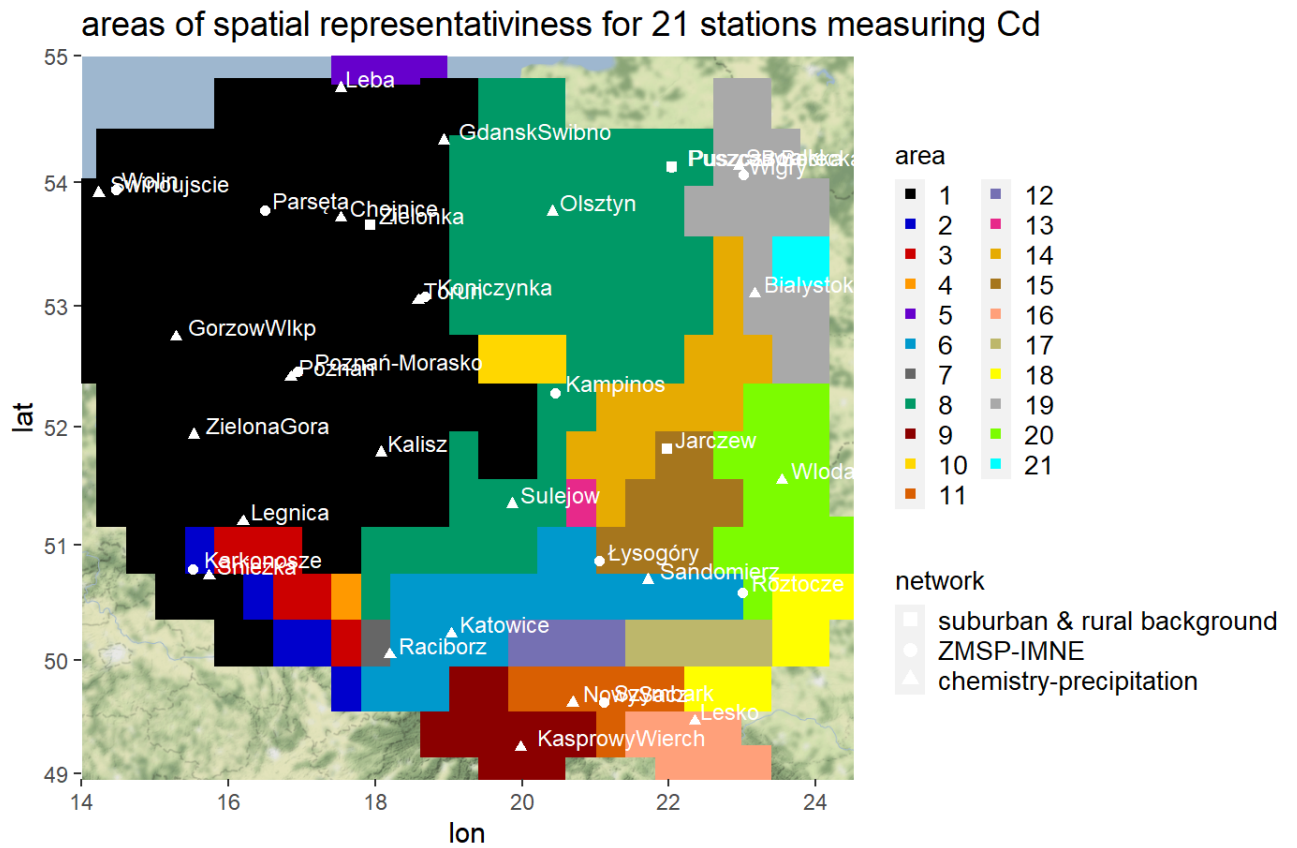


Figure 9: Areas of spatial representativeness for 21 stations measuring Cd deposition.

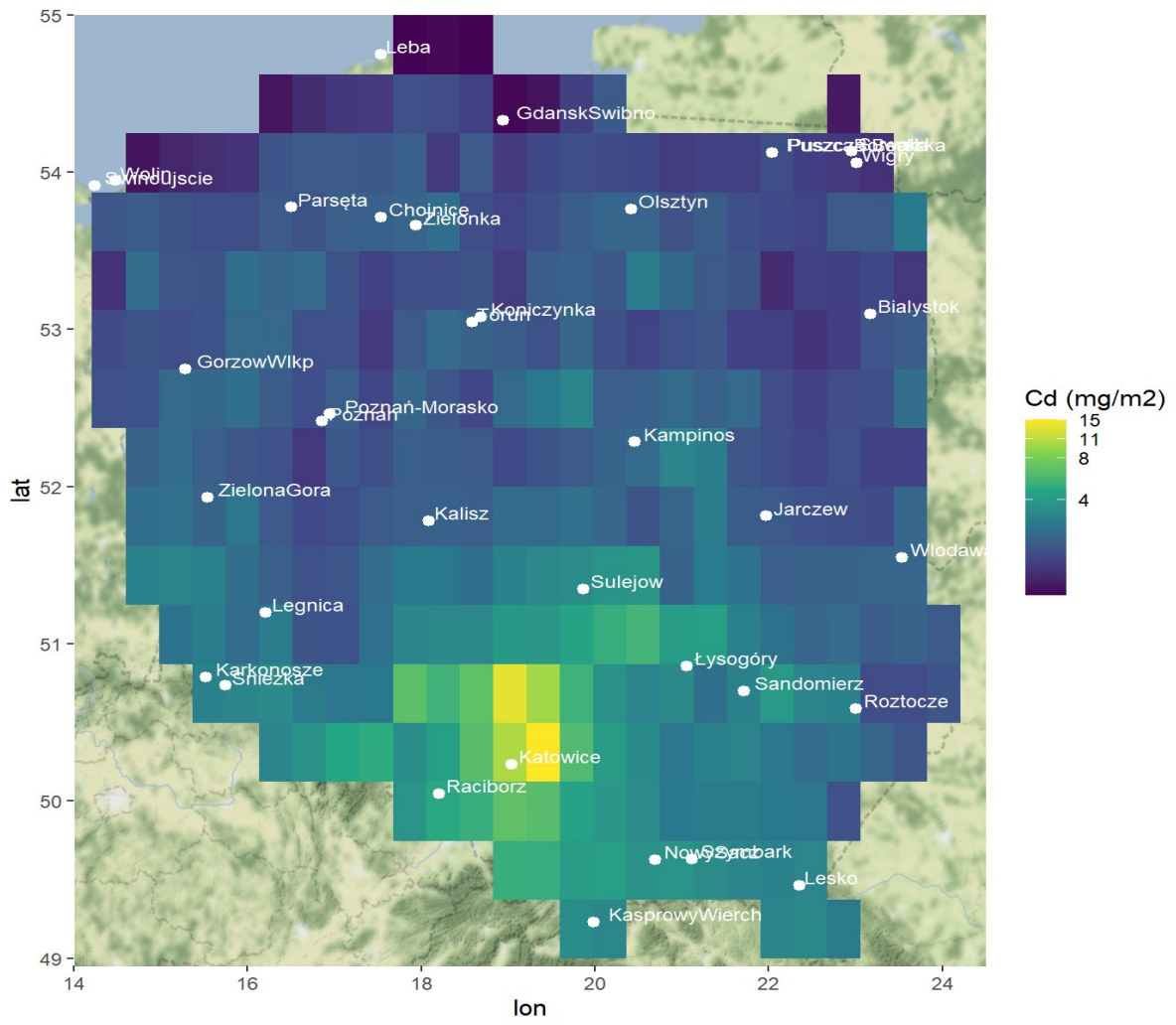


Figure 10: Average Cd monthly deposition between 2015 and 2019 (plotted in logarithmic scale).

Table 6: Cluster information based on the results presented in Figure 9 and Figure 29 for the monitoring stations listed in Table 2.

21 regions	name
1	Chojnice
1	GdanskSwibno
1	GorzowWlkp
1	Kalisz
1	Leba
1	Legnica
1	Poznan
1	Sniezka
1	Swinoujscie
1	Torun
1	ZielonaGora
1	Zielonka
1	Parseta
1	Wolin
1	Poznan-Morasko
1	Koniczynka
2	Karkonosze
6	Katowice
6	Sandomierz
6	Lysogóry
6	Roztocze
7	Raciborz
8	Olsztyn
8	Sulejow
8	PuszczaBorecka
8	Kampinos
8	Puszcza Borecka
9	KasprowyWierch
11	NowySacz
11	Szymbark
14	Bialystok
14	Jarczew
16	Lesko
19	Suwalki
19	Wigry
20	Wlodawa

3.1.5 Hg deposition

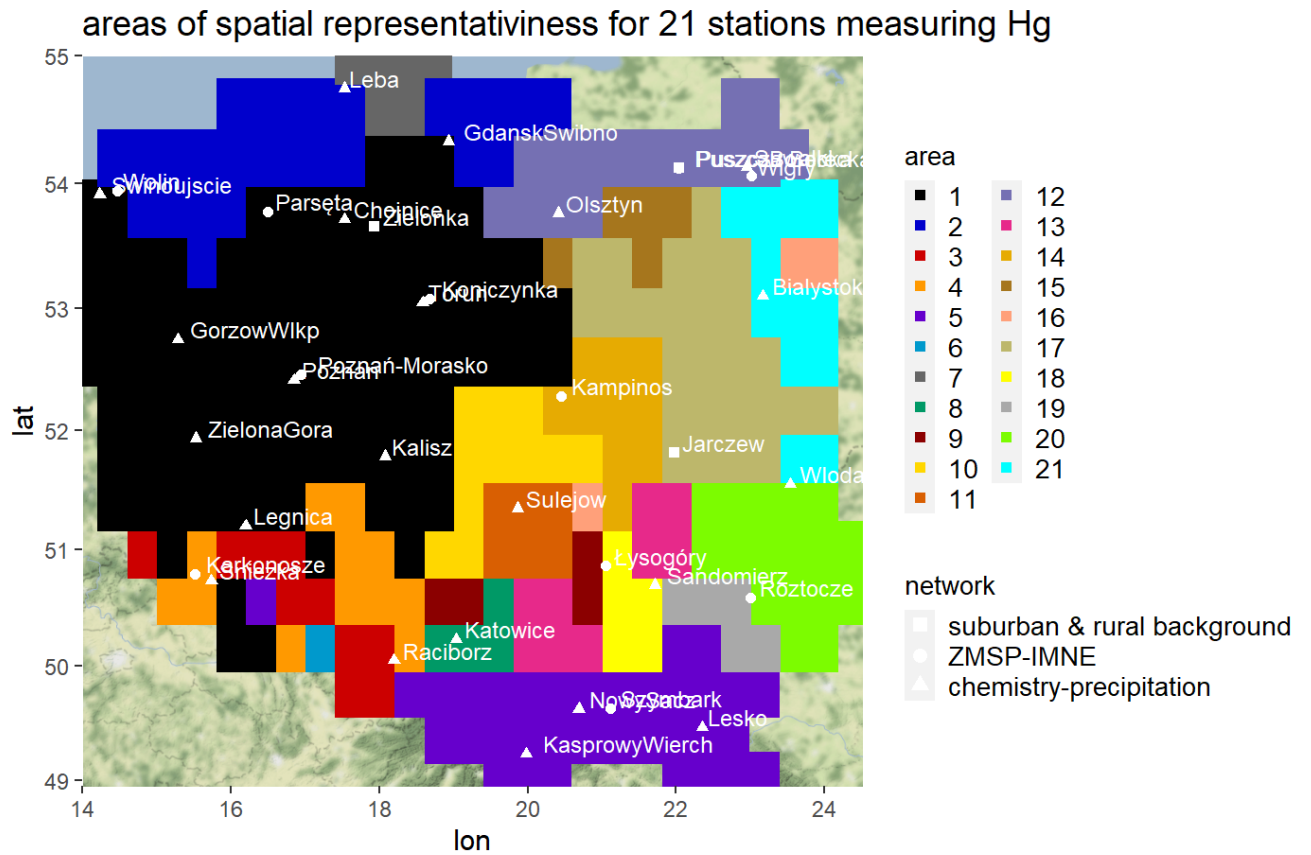


Figure 11: Areas of spatial representativeness for 21 stations measuring Hg deposition.

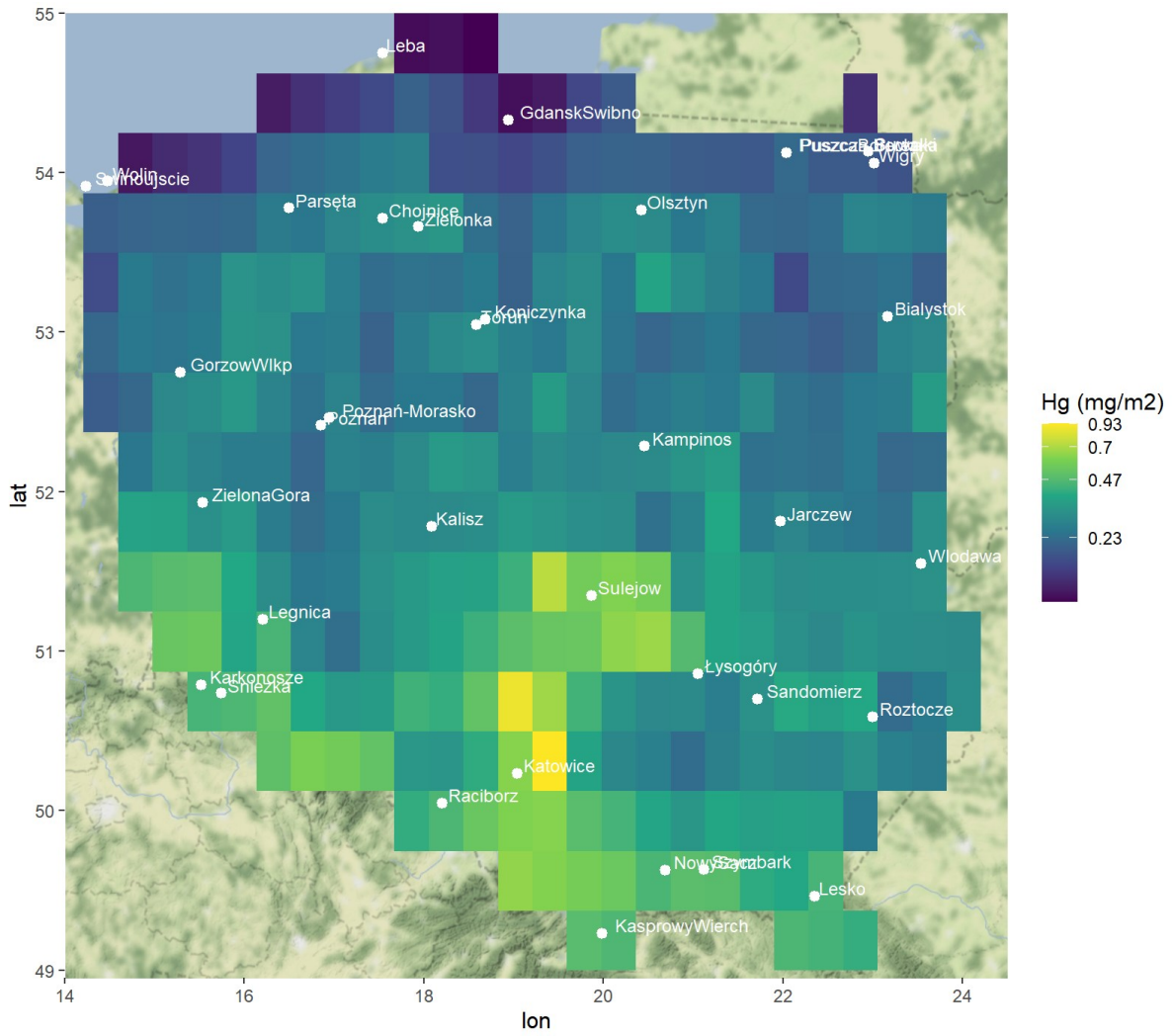


Figure 12: Average Hg monthly deposition between 2015 and 2019 (plotted in logarithmic scale).

Table 7: Cluster information based on the results presented in Figure 11 and Figure 30 for the monitoring stations listed in Table 2.

21 regions	name
1	Chojnice
1	GdanskSwibno
1	GorzowWlkp
1	Kalisz
1	Legnica
1	Poznan
1	Swinoujscie
1	Torun
1	ZielonaGora
1	Zielonka
1	Parseta
1	Wolin
1	Poznan-Morasko
1	Koniczynka
2	Leba
3	Raciborz
4	Sniezka
4	Karkonosze
5	KasprowyWierch
5	Lesko
5	NowySacz
5	Szymbark
8	Katowice
9	Lysogóry
11	Sulejow
12	Olsztyn
12	Suwalki
12	PuszczaBorecka
12	Puszcza Borecka
12	Wigry
14	Jarczew
14	Kampinos
17	Bialystok
18	Sandomierz
19	Roztocze
20	Wlodawa

3.1.6 Pb deposition

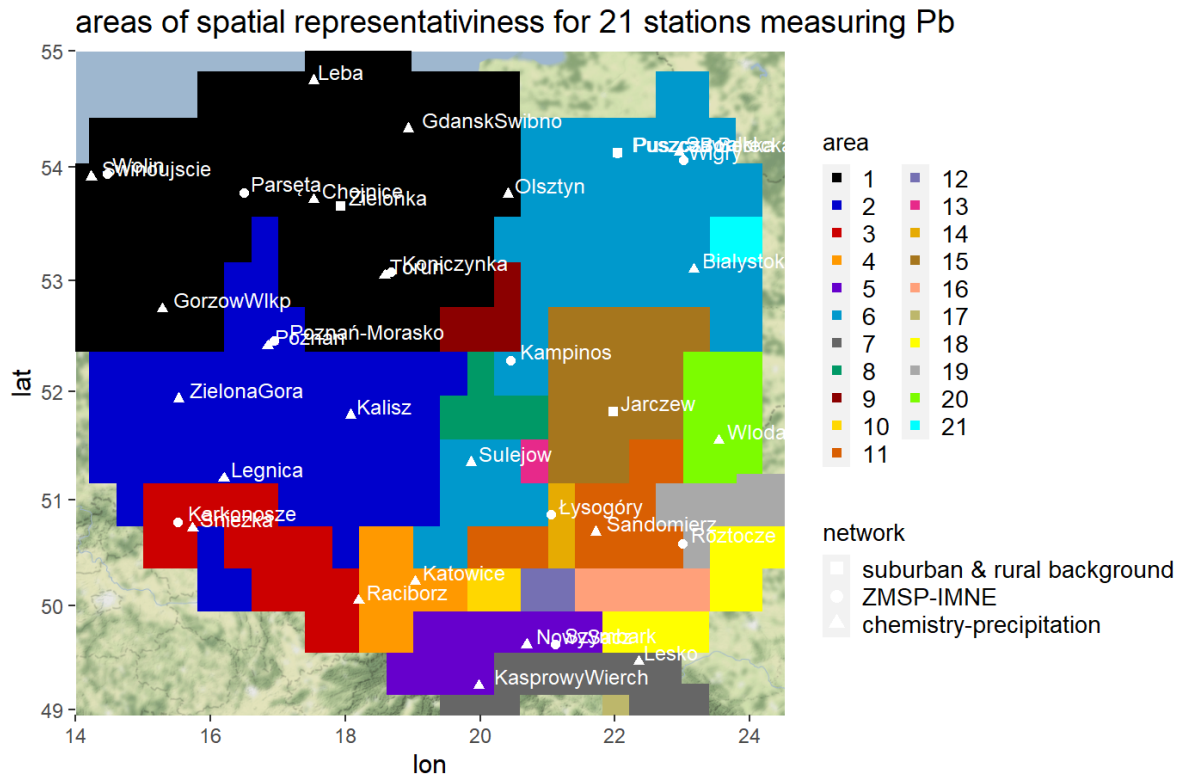


Figure 13: Areas of spatial representativeness for 21 stations measuring Pb deposition.

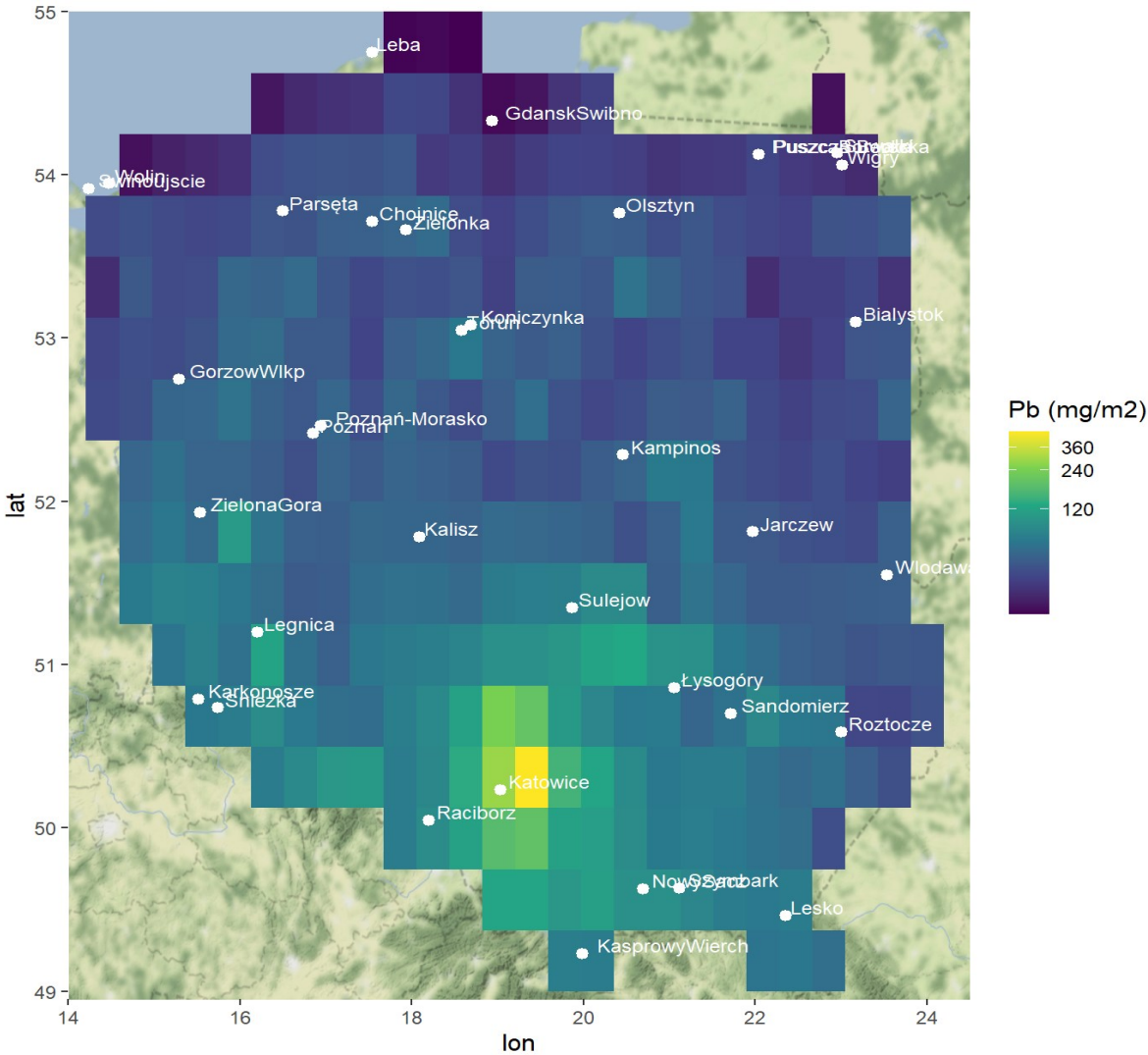


Figure 14: Average Pb monthly deposition between 2015 and 2019 (plotted in logarithmic scale).

Table 8: Cluster information based on the results presented in Figure 13 and Figure 31 for the monitoring stations listed in Table 2.

21 regions	name
1	Chojnice
1	GdanskSwibno
1	GorzowWlkp
1	Leba
1	Olsztyn
1	Torun
1	Zielonka
1	Parseta
1	Wolin
1	Koniczynka
2	Kalisz
2	Legnica
2	Poznan
2	Swinoujscie
2	ZielonaGora
2	Poznan-Morasko
3	Raciborz
3	Sniezka
3	Karkonosze
4	Katowice
5	NowySacz
5	Szymbark
6	Bialystok
6	Sulejow
6	Suwalki
6	PuszczaBorecka
6	Kampinos
6	Lysogóry
6	Puszcza Borecka
6	Wigry
7	KasprowyWierch
7	Lesko
11	Sandomierz
11	Roztocze
15	Jarczew
20	Wlodawa

3.2 Station dissimilarity analysis (dendrograms)

A hierarchical clustering analysis using the (1-R) x EuD metric was performed based on deposition data described in Table 1. The dendrograms show the relation between the observational time series available for this analysis, and the maps depict the six main groups of stations (location and which cluster they fall into) defined by the clustering process and represented via the dendrogram.

3.2.2 NH4 deposition

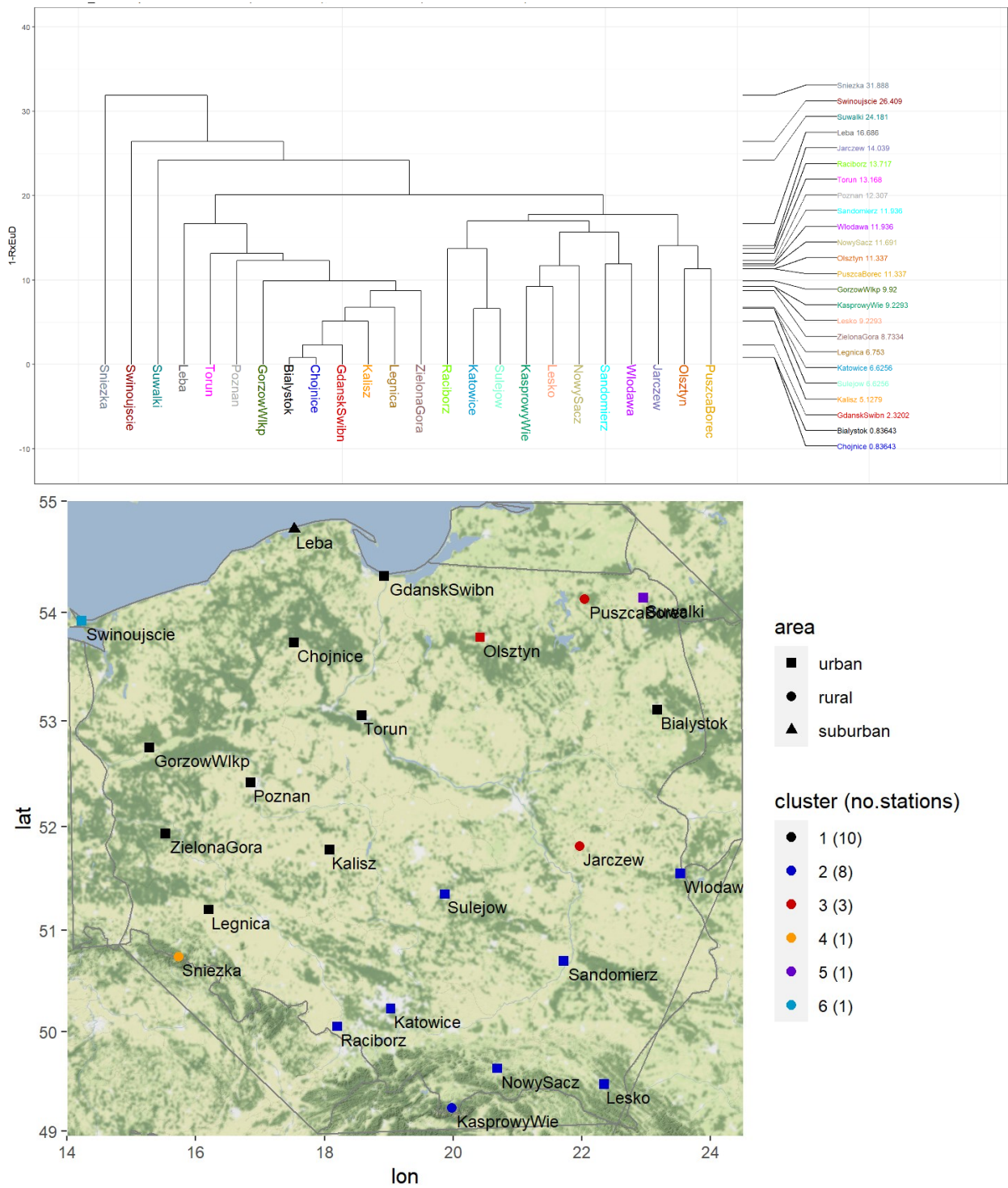


Figure 16: Dendrograms based on the NH4 deposition and mapping of the main six groups of stations (cluster colour-coded).

3.2.3 Ca deposition

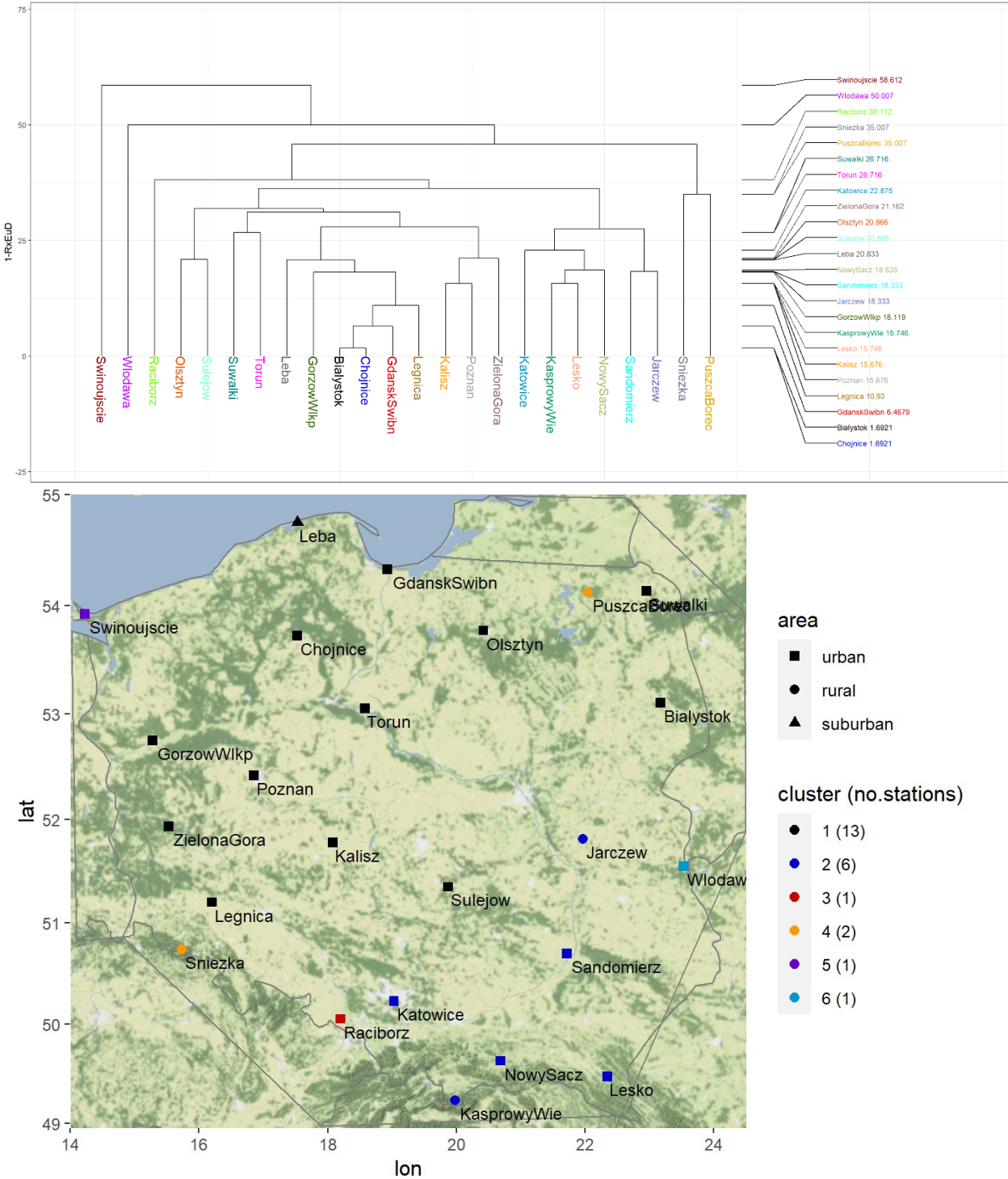


Figure 17: Dendrograms based on the Ca deposition and mapping of the main six groups of stations (cluster colour-coded).

3.2.4 Cl deposition

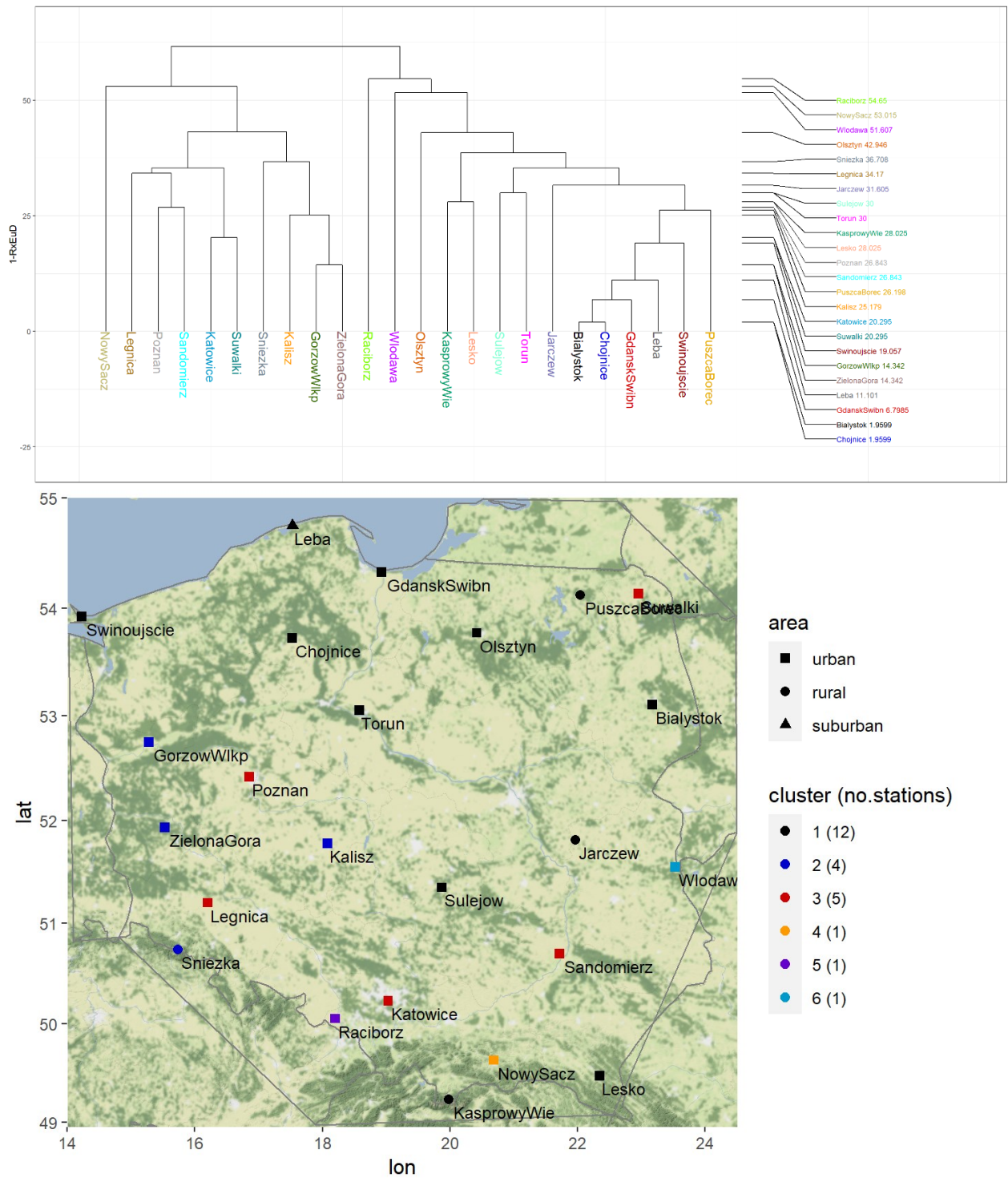


Figure 18: Dendrograms based on the Cl deposition and mapping of the main six groups of stations (cluster colour-coded).

3.2.5 Na deposition

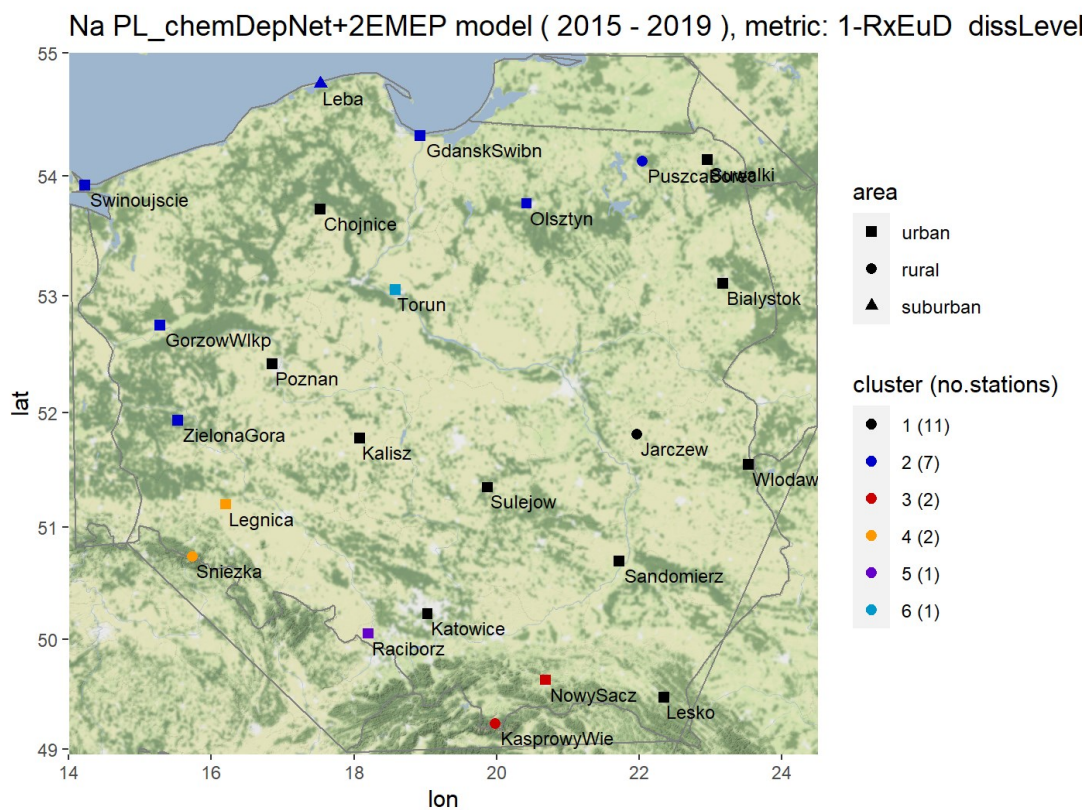
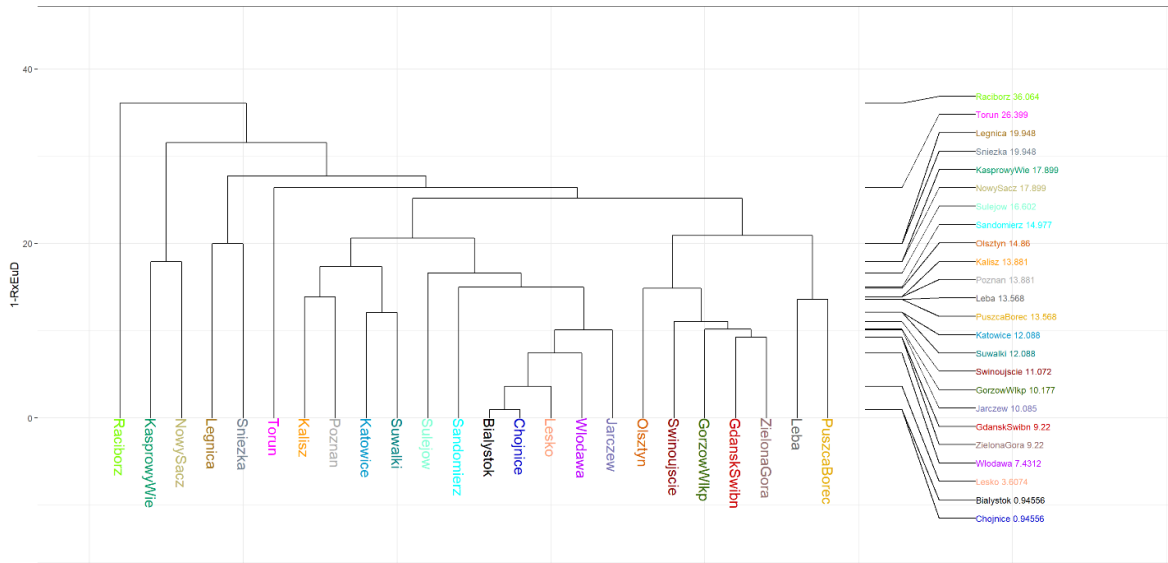


Figure 19: Dendrograms based on the Na deposition and mapping of the main six groups of stations (cluster colour-coded).

3.2.6 K deposition

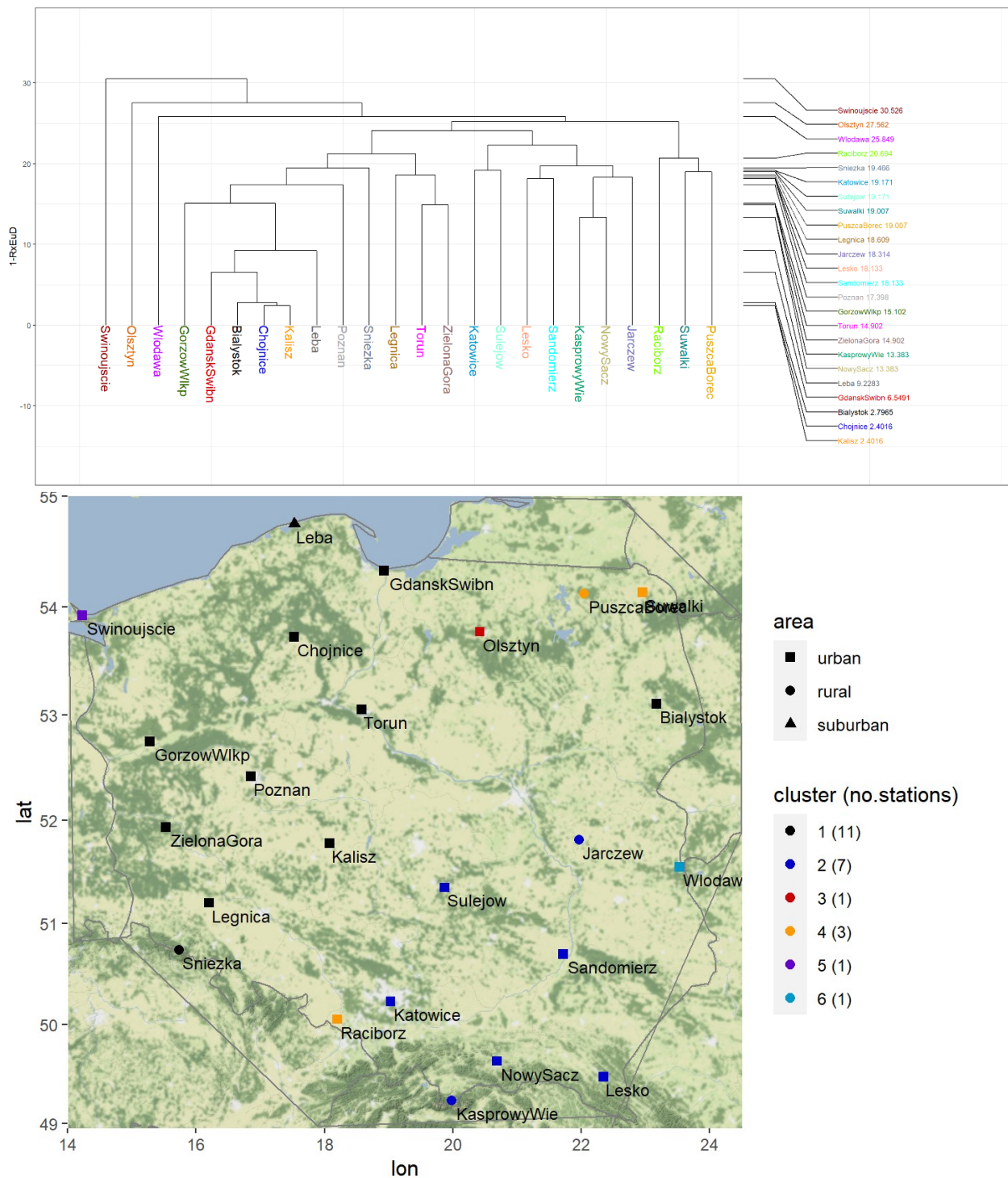


Figure 20: Dendrograms based on the K deposition and mapping of the main six groups of stations (cluster colour-coded).

3.2.7 H⁺ deposition

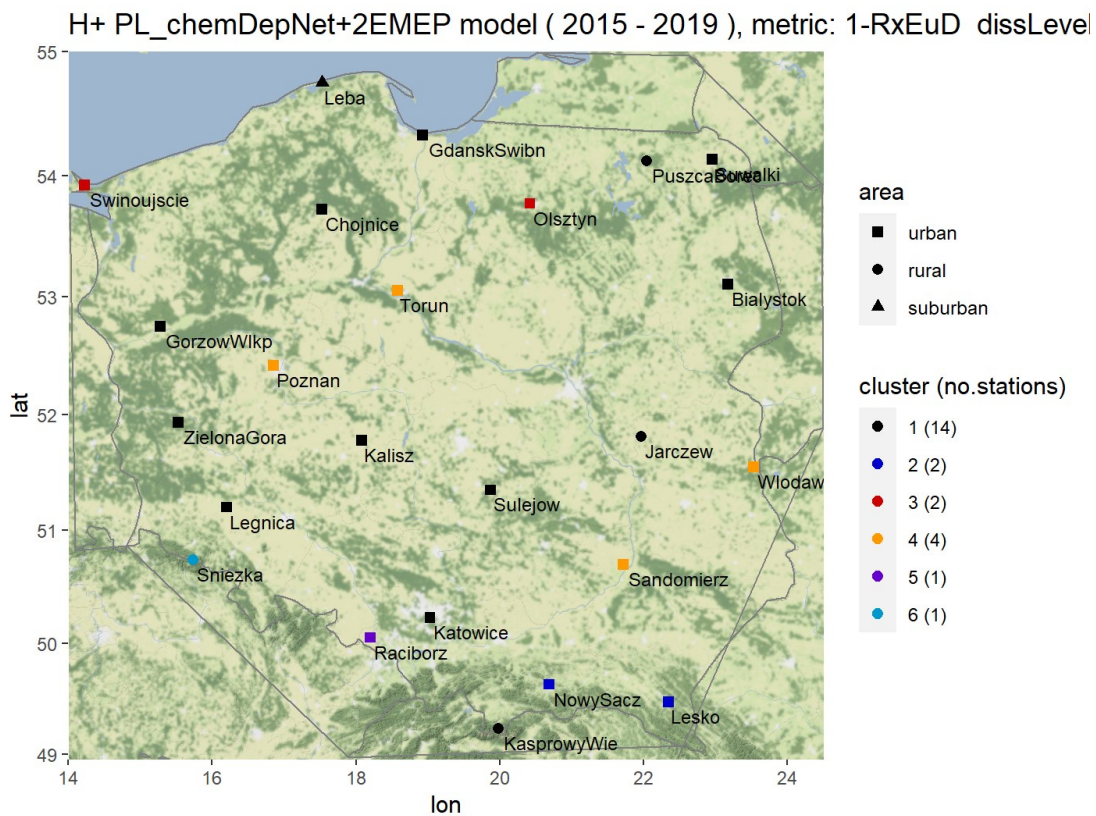
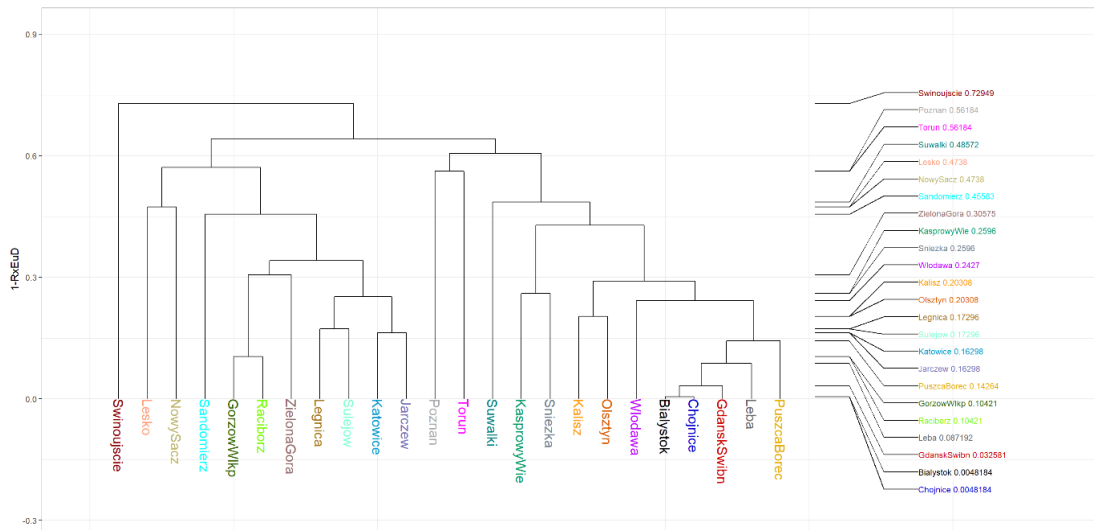


Figure 21: Dendrograms based on the H⁺ deposition and mapping of the main six groups of stations (cluster colour-coded).

3.2.8 Mg²⁺ deposition

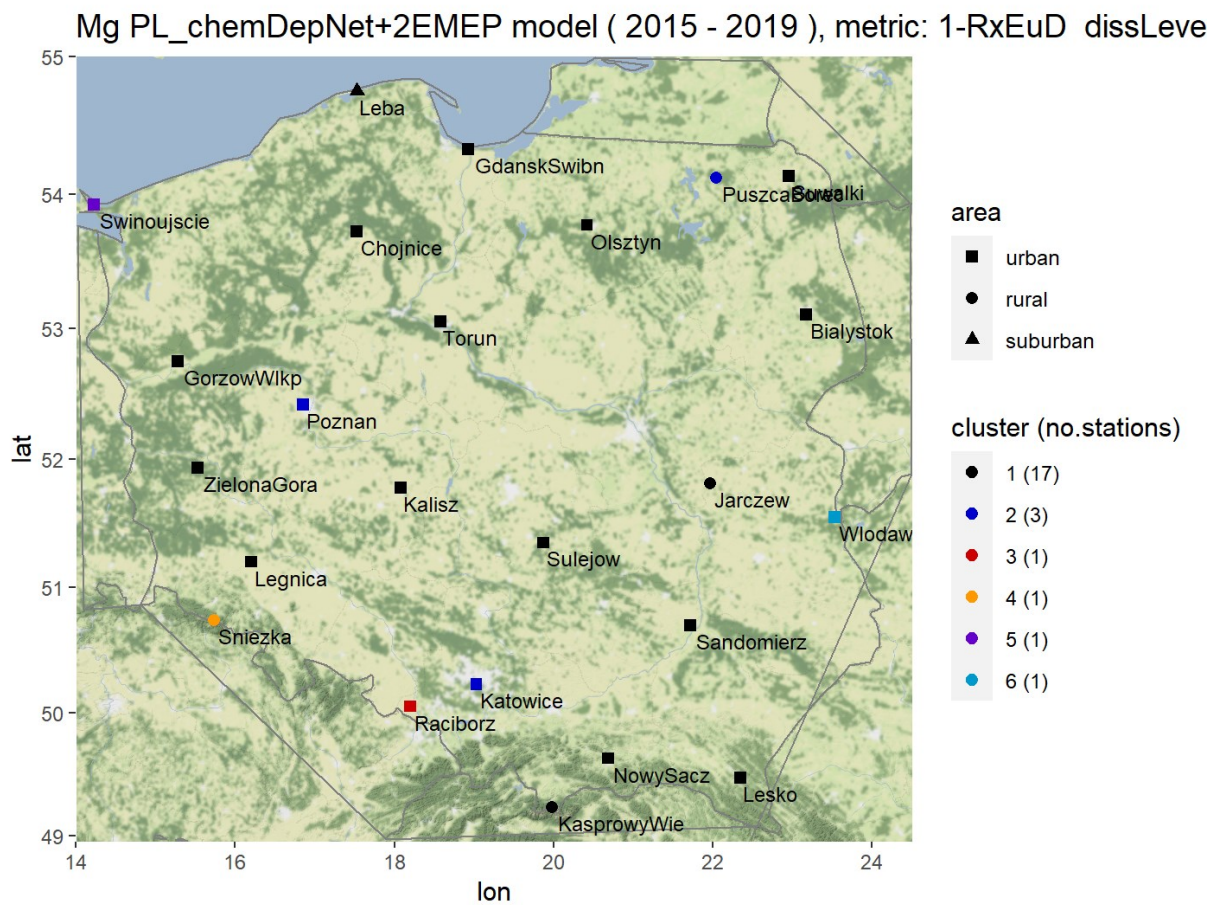
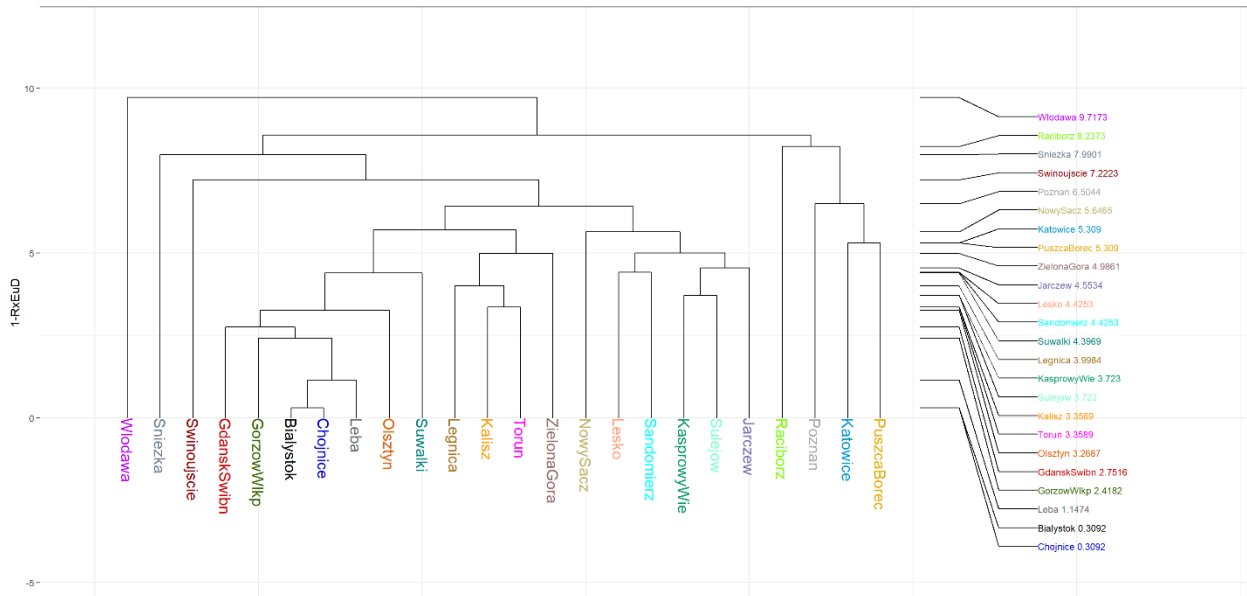


Figure 22: Dendrograms based on the Mg²⁺ deposition and mapping of the main six groups of stations (cluster colour-coded).

3.2.9 Relative ranking of stations

One of the main outcomes of the analysis is an ordering or ranking of stations, according to the degree of similarity of their observation records. Absolute thresholds for redundancy cannot be generated since the relative rankings depend on the available observation data (number of stations and chemical species observed). The analysis thus does indicate the relative ranking of monitoring record similarity, which can be used as one of the inputs for network optimization decision making by identifying stations that may show unique characteristics (top of the ranking – station clustered at a high level of dissimilarity) or may be measuring similar features (bottom of the ranking- station clustered at low levels of dissimilarity). It is also an easier way to cross-compare different parameters to be considered in the analysis.

Table 9: Relative ranking of stations based on the level of dissimilarity based on the clustering analysis.

Ca	Na	K	Cl	H+	Mg	SO4
Swinoujscie	Wlodawa	Olsztyn	NowySacz	Swinoujscie	Poznan	Wlodawa
Wlodawa	Raciborz	Katowice	Sandomierz	Poznan	Swinoujscie	Raciborz
Torun	Swinoujscie	ZielonaGora	Raciborz	Suwalki	Torun	Sniezka
Leba	Sniezka	Swinoujscie	Sniezka	Torun	Sniezka	Leba
Sniezka	KasprowyWie	Poznan	Wlodawa	Lesko	Sulejow	Torun
Suwalki	NowySacz	Sulejow	Torun	NowySacz	Lesko	Lesko
Katowice	Sandomierz	PuszczaBorec	Katowice	Sandomierz	Raciborz	Suwalki
Sulejow	Lesko	Raciborz	Kalisz	KasprowyWie	Sandomierz	Katowice
Raciborz	Sulejow	NowySacz	Sulejow	ZielonaGora	Leba	NowySacz
Poznan	Poznan	Sniezka	Olsztyn	Sniezka	Jarczew	Poznan
NowySacz	ZielonaGora	Legnica	Legnica	Wlodawa	Katowice	ZielonaGora
ZielonaGora	Katowice	GorzowWlkp	Jarczew	Kalisz	Wlodawa	Jarczew
PuszczaBorec	Leba	Suwalki	Lesko	Legnica	PuszczaBorec	GorzowWlkp
Olsztyn	Suwalki	Torun	KasprowyWie	Olsztyn	Olsztyn	Legnica
Legnica	Torun	Wlodawa	Leba	Sulejow	GdanskSwibn	Swinoujscie
Jarczew	Kalisz	Lesko	Poznan	Jarczew	Kalisz	Sandomierz
Sandomierz	GorzowWlkp	KasprowyWie	Swinoujscie	Katowice	Legnica	KasprowyWie
Kalisz	Jarczew	Jarczew	GorzowWlkp	GorzowWlkp	KasprowyWie	Olsztyn
GdanskSwibn	PuszczaBorec	Sandomierz	Suwalki	PuszczaBorec	Suwalki	Sulejow
Lesko	Olsztyn	GdanskSwibn	PuszczaBorec	Leba	ZielonaGora	Kalisz
KasprowyWie	GdanskSwibn	Leba	ZielonaGora	Raciborz	NowySacz	PuszczaBorec
GorzowWlkp	Legnica	Chojnice	GdanskSwibn	GdanskSwibn	GorzowWlkp	GdanskSwibn
Chojnice	Chojnice	Kalisz	Chojnice	Chojnice	Chojnice	Chojnice
Bialystok	Bialystok	Bialystok	Bialystok	Bialystok	Bialystok	Bialystok

3.3 Source receptor analysis for selected stations for deposited SOx

With this analysis we investigate which emission region will potentially influence specific stations of the network. Depending on the location of the station, but also on the seasonality of the data, this so-called footprint maps will show different patterns of air masses influencing the station. The analysis is similar to a trajectory analysis, but performed with a model which can quantitatively calculate deposited mass in backward mode (FLEXPART). First, we performed an analysis for finding the source regions of air masses containing later deposited material for background stations in cluster 1, 5, 9 and 17 (see Table 3). Independently of the cluster, it can be seen that the sources near to the station have the largest influence over the station (red colour). Clusters 1 and 5 potentially

receive air masses from the west, while clusters 9 and 17 receive air from the east and also from the south.

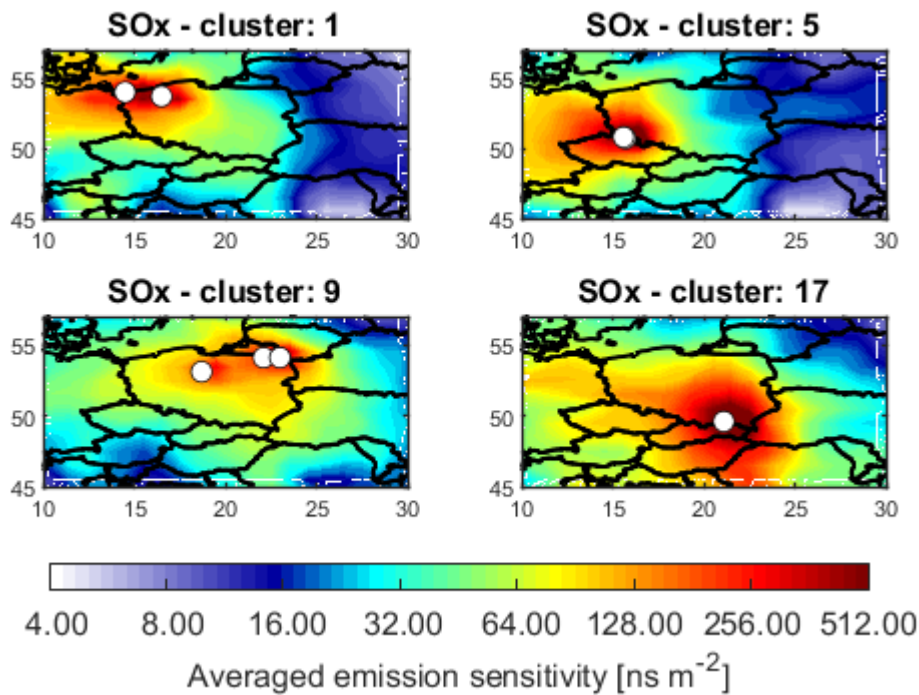


Figure 23: Residence time of air masses which lead to deposition at the stations in the NE, SE, NW and SW of Poland.

If not annual averages, but monthly footprints are considered, it can be seen that the transport patterns during e.g., April and October are significantly different. The relevant source regions are mainly over Poland in April, while in October are in the Czech Republic.

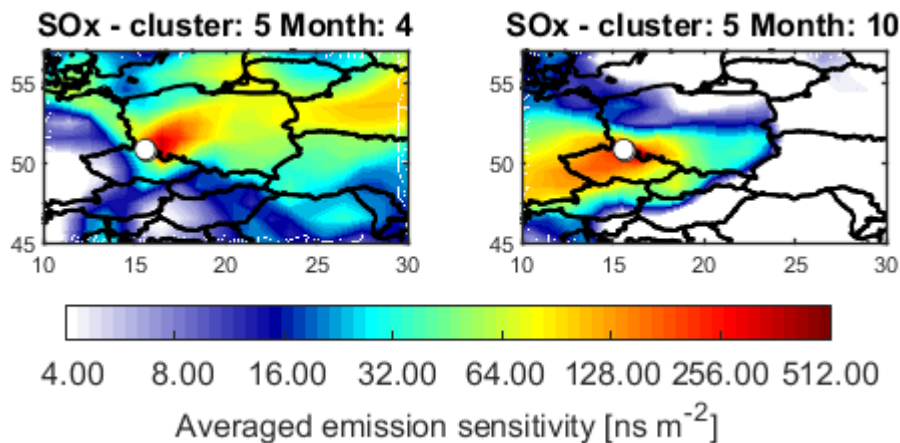


Figure 24: Residence time of air masses that lead to deposition at the stations in the SW of Poland, but for different months. The left panel shows the SRR for April, while the right panel shows the SRR for October.

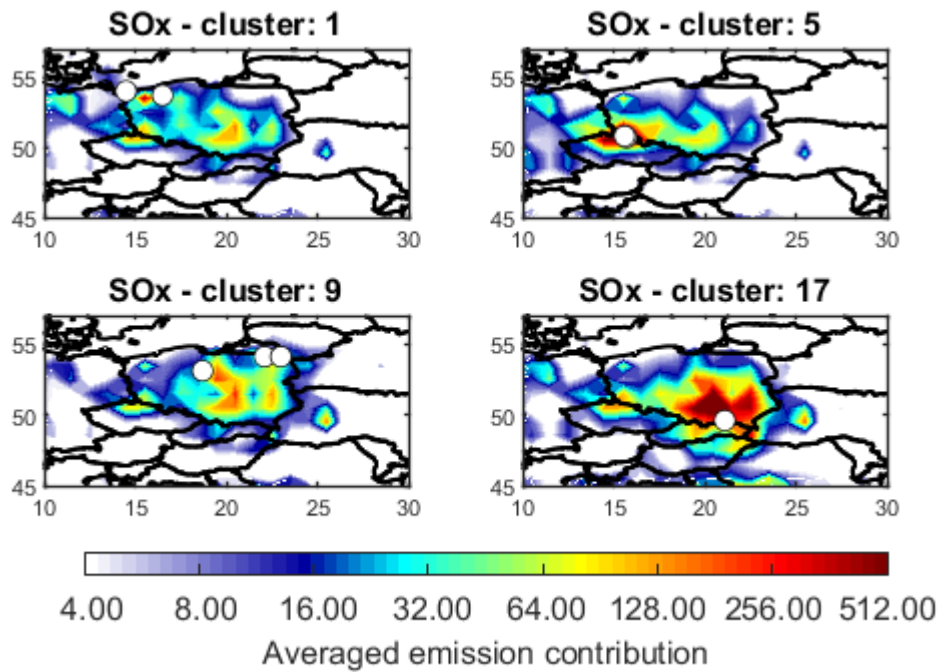


Figure 25: Same as Fig 24, but this shows the emission contribution when using the ECLIPSE SO₂ inventory.

Investigating which emissions sources are important for which cluster shows clearly, that the emission within Poland is dominating the observed deposition fluxes at the stations. However, for cluster 5 and partly for cluster 17, Czech and Slovakian emissions also play a role.

4 Analysis of the clustering results

4.1 Areas of representativeness of a single station

Analysing the hierarchical clustering results based on the modelled data has provided us several insights. The results are different depending on the pollutant, as the different compounds have different sources and seasonality. However, the methodology has clearly identified regions that represent the area of representativeness of a single station (a large number of cells are part of a single cluster). This means that a single station located within this region will monitor an air mass that is similar temporally and has similar deposition levels. The methodology has also identified smaller source areas, typically a cluster comprised of 1 to 10 grid-cells, indicating point/area sources of the compounds analysed. These source areas are clearly seen for higher resolution runs (acidic compounds), especially if there are larger clusters surrounding a few-grid-cells region.

The results show that SO_x, O₃, RDN show a more homogeneous distribution of the areas representing a single station than Cd, Hg, Pb results. The results for the metals seem to indicate the difference between west-east and north-south, with smaller clusters south/east part of the country. Thus, indicating a higher number of sources in the south/east part of Poland. However, independently of the chemical compounds, the results seem to indicate that the north-west side of the country needs a lesser degree of monitoring than the south-east part of the country, independently of the compound. These results are supported by the mapping of 25 regions (Section A.1 in the Annex) and the deposition maps show in the subsections of Section 3.1: i) the mapping of 25 regions shows that bigger clusters are broken into smaller ones, mainly in the south-east of Poland; ii) the depositions maps clearly show the location of the emission sources responsible for the deposition levels. Note that the deposition maps complement information to understand how the grid-cells clustered have clustered, e.g., source location indication. However, the map can only be compared to a certain extent, as it represents the average of monthly means over a 5-year period (average deposition maps); the clustering analysis results considers every single monthly average for the 5-year period. Hence, the clustering analysis results consider the seasonal variations across the time period.

The table below shows how many stations are located per cluster for the different compounds for 21 regions. It shows several regions with two or more stations, and some have none. However, this is due to the grid-cells where sources are located showing as a cluster. Nevertheless, there are clusters, in particular in central and east Poland, that are not covered by the stations with observations included in this analysis.

Table 10: Cluster information based on the results presented in Figure 3, Figure 5, Figure 7, Figure 9, Figure 11, and Figure 13.

cluster	SOx	OXN	RDN	Cd	Hg	Pb
1	3	3	6	11	9	7
2	3	1	0	0	1	4
3	0	1	0	0	1	2
4	1	5	6	0	1	1
5	2	2	1	0	3	2
6	1	2	0	2	0	4
7	1	0	1	1	0	1
8	1	0	0	3	1	0
9	5	0	4	1	0	0
10	3	2	0	0	0	0
11	0	1	1	1	1	1
12	2	4	0	0	3	0
13	0	0	0	0	0	0
14	0	0	1	2	1	0
15	0	2	0	0	0	1
16	0	0	1	1	0	0
17	1	1	2	0	1	0
18	0	0	0	0	1	0
19	0	0	0	1	0	0
20	1	0	1	1	1	1
21	0	0	0	0	0	0

4.2 Station dissimilarity analysis based on observations

As expected, the dendrograms give very different information depending on the parameters we are analysing. It might be useful to analyse it individually if any of the parameters becomes the driver of the optimization. Since there was no preferred parameter (compound), it is easier to use the relative ranking for the analysis to examine the dissimilarity of the stations depending on the measured compound.

Considering the number of stations with observations available for the analysis, we have chosen to look at the bottom/top five stations of the relative ranking. This allows to cross check the stations that measure different compounds. Here the word network is to assume that currently the network is comprised of the stations that we have observation for the analysis.

Table 11: Highlighting the stations that seems to be most similar across the results for individual pollutants.

Ca	Na	K	Cl	H+	Mg	SO4	NH3-N
Swinoujscie	Raciborz	Swinoujscie	Raciborz	Sniezka	Wlodawa	Wlodawa	Sniezka
Wlodawa	Torun	Olsztyn	NowySacz	Raciborz	Raciborz	Sniezka	Swinoujscie
Raciborz	Legnica	Wlodawa	Wlodawa	Wlodawa	Sniezka	Raciborz	Suwalki
Sniezka	Sniezka	Raciborz	Olsztyn	Olsztyn	Swinoujscie	Torun	Leba
PuszczaBorec	KasprowyWie	Sniezka	Sniezka	Swinoujscie	Poznan	Leba	Jarczew
Suwalki	NowySacz	Katowice	Legnica	ZielonaGora	NowySacz	Suwalki	Raciborz
Torun	Sulejow	Sulejow	Jarczew	Lesko	Katowice	Lesko	Torun
Katowice	Sandomierz	Suwalki	Sulejow	NowySacz	PuszczaBorec	Poznan	Poznan
ZielonaGora	Olsztyn	PuszczaBorec	Torun	Poznan	ZielonaGora	Jarczew	Sandomierz
Olsztyn	Kalisz	Legnica	KasprowyWie	Sandomierz	Jarczew	Kalisz	Wlodawa
Sulejow	Poznan	Jarczew	Lesko	Torun	Lesko	Legnica	NowySacz
Leba	Leba	Lesko	Poznan	KasprowyWie	Sandomierz	NowySacz	Olsztyn
NowySacz	PuszczaBorec	Sandomierz	Sandomierz	Suwalki	Suwalki	Swinoujscie	PuszczaBorec
Sandomierz	Katowice	Poznan	PuszczaBorec	Jarczew	Legnica	KasprowyWie	GorzowWlkp
Jarczew	Suwalki	GorzowWlkp	Kalisz	Katowice	KasprowyWie	Sandomierz	KasprowyWie
GorzowWlkp	Swinoujscie	Torun	Katowice	Legnica	Sulejow	Olsztyn	Lesko
KasprowyWie	GorzowWlkp	ZielonaGora	Suwalki	Kalisz	Kalisz	PuszczaBorec	ZielonaGora
Lesko	Jarczew	KasprowyWie	Swinoujscie	PuszczaBorec	Torun	Katowice	Legnica
Kalisz	GdanskSwibn	NowySacz	GorzowWlkp	GorzowWlkp	Olsztyn	Sulejow	Katowice
Poznan	ZielonaGora	Leba	ZielonaGora	Sulejow	GdanskSwibn	GorzowWlkp	Sulejow
Legnica	Wlodawa	GdanskSwibn	Leba	Leba	GorzowWlkp	ZielonaGora	Kalisz
GdanskSwibn	Lesko	Bialystok	GdanskSwibn	GdanskSwibn	Leba	GdanskSwibn	GdanskSwibn
Bialystok	Bialystok	Chojnice	Bialystok	Bialystok	Bialystok	Bialystok	Bialystok
Chojnice	Chojnice	Kalisz	Chojnice	Chojnice	Chojnice	Chojnice	Chojnice

Table 11 can be used as a guide to quickly visualize which stations are the most similar and where they stand in the relative ranking. From this quick screening, six stations that recurrently are in the bottom part of the ranking: Chojnice, Bialystok, Gdansk Swibn, ZielonaGora, Gorzow Wlkp and Kalisz. These stations should be target first if to consider changes in the monitoring network.

Table 12: Highlighting the stations that seems to be least similar across the results for individual pollutants.

Ca	Na	K	Cl	H+	Mg	SO4	NH3-N
Swinoujscie	Raciborz	Swinoujscie	Raciborz	Sniezka	Wlodawa	Wlodawa	Sniezka
Wlodawa	Torun	Olsztyn	NowySacz	Raciborz	Raciborz	Sniezka	Swinoujscie
Raciborz	Legnica	Wlodawa	Wlodawa	Wlodawa	Sniezka	Raciborz	Suwalki
Sniezka	Sniezka	Raciborz	Olsztyn	Olsztyn	Swinoujscie	Torun	Leba
PuszczaBorec	KasprowyWie	Sniezka	Sniezka	Swinoujscie	Poznan	Leba	Jarczew
Suwalki	NowySacz	Katowice	Legnica	ZielonaGora	NowySacz	Suwalki	Raciborz
Torun	Sulejow	Sulejow	Jarczew	Lesko	Katowice	Lesko	Torun
Katowice	Sandomierz	Suwalki	Sulejow	NowySacz	PuszczaBorec	Poznan	Poznan
ZielonaGora	Olsztyn	PuszczaBorec	Torun	Poznan	ZielonaGora	Jarczew	Sandomierz
Olsztyn	Kalisz	Legnica	KasprowyWie	Sandomierz	Jarczew	Kalisz	Wlodawa
Sulejow	Poznan	Jarczew	Lesko	Torun	Lesko	Legnica	NowySacz
Leba	Leba	Lesko	Poznan	KasprowyWie	Sandomierz	NowySacz	Olsztyn
NowySacz	PuszczaBorec	Sandomierz	Sandomierz	Suwalki	Suwalki	Swinoujscie	PuszczaBorec
Sandomierz	Katowice	Poznan	PuszczaBorec	Jarczew	Legnica	KasprowyWie	GorzowWlkp
Jarczew	Suwalki	GorzowWlkp	Kalisz	Katowice	KasprowyWie	Sandomierz	KasprowyWie
GorzowWlkp	Swinoujscie	Torun	Katowice	Legnica	Sulejow	Olsztyn	Lesko
KasprowyWie	GorzowWlkp	ZielonaGora	Suwalki	Kalisz	Kalisz	PuszczaBorec	ZielonaGora
Lesko	Jarczew	KasprowyWie	Swinoujscie	PuszczaBorec	Torun	Katowice	Legnica
Kalisz	GdanskSwibn	NowySacz	GorzowWlkp	GorzowWlkp	Olsztyn	Sulejow	Katowice
Poznan	ZielonaGora	Leba	ZielonaGora	Sulejow	GdanskSwibn	GorzowWlkp	Sulejow
Legnica	Wlodawa	GdanskSwibn	Leba	Leba	GorzowWlkp	ZielonaGora	Kalisz
GdanskSwibn	Lesko	Bialystok	GdanskSwibn	GdanskSwibn	Leba	GdanskSwibn	GdanskSwibn
Bialystok	Bialystok	Chojnice	Bialystok	Bialystok	Bialystok	Bialystok	Bialystok
Chojnice	Chojnice	Kalisz	Chojnice	Chojnice	Chojnice	Chojnice	Chojnice

Table 12 can be used as a guide to quickly visualize which stations are the most dissimilar and where they stand in the relative ranking. From this quick screening, five stations are recurrently at the top of the ranking: Sniezka, Raciborz, Swinoujscie, Wlodawa, and Olsztyn. These stations seem to be unique in the network.

Table 13: Highlighting the stations that seems to have a mix similarity across the results for individual pollutants.

Ca	Na	K	Cl	H+	Mg	SO4	NH3-N
Swinoujscie	Raciborz	Swinoujscie	Raciborz	Sniezka	Wlodawa	Wlodawa	Sniezka
Wlodawa	Torun	Olsztyn	NowySacz	Raciborz	Raciborz	Sniezka	Swinoujscie
Raciborz	Legnica	Wlodawa	Wlodawa	Wlodawa	Sniezka	Raciborz	Suwalki
Sniezka	Sniezka	Raciborz	Olsztyn	Olsztyn	Swinoujscie	Torun	Leba
PuszczBorec	KasprowyWie	Sniezka	Sniezka	Swinoujscie	Poznan	Leba	Jarczew
Suwalki	NowySacz	Katowice	Legnica	ZielonaGora	NowySacz	Suwalki	Raciborz
Torun	Sulejow	Sulejow	Jarczew	Lesko	Katowice	Lesko	Torun
Katowice	Sandomierz	Suwalki	Sulejow	NowySacz	PuszczBorec	Poznan	Poznan
ZielonaGora	Olsztyn	PuszczBorec	Torun	Poznan	ZielonaGora	Jarczew	Sandomierz
Olsztyn	Kalisz	Legnica	KasprowyWie	Sandomierz	Jarczew	Kalisz	Wlodawa
Sulejow	Poznan	Jarczew	Lesko	Torun	Lesko	Legnica	NowySacz
Leba	Leba	Lesko	Poznan	KasprowyWie	Sandomierz	NowySacz	Olsztyn
NowySacz	PuszczBorec	Sandomierz	Sandomierz	Suwalki	Suwalki	Swinoujscie	PuszczBorec
Sandomierz	Katowice	Poznan	PuszczBorec	Jarczew	Legnica	KasprowyWie	GorzowWlkp
Jarczew	Suwalki	GorzowWlkp	Kalisz	Katowice	KasprowyWie	Sandomierz	KasprowyWie
GorzowWlkp	Swinoujscie	Torun	Katowice	Legnica	Sulejow	Olsztyn	Lesko
KasprowyWie	GorzowWlkp	ZielonaGora	Suwalki	Kalisz	Kalisz	PuszczBorec	ZielonaGora
Lesko	Jarczew	KasprowyWie	Swinoujscie	PuszczBorec	Torun	Katowice	Legnica
Kalisz	GdanskSwibn	NowySacz	GorzowWlkp	GorzowWlkp	Olsztyn	Sulejow	Katowice
Poznan	ZielonaGora	Leba	ZielonaGora	Sulejow	GdanskSwibn	GorzowWlkp	Sulejow
Legnica	Wlodawa	GdanskSwibn	Leba	Leba	GorzowWlkp	ZielonaGora	Kalisz
GdanskSwibn	Lesko	Bialystok	GdanskSwibn	GdanskSwibn	Leba	GdanskSwibn	GdanskSwibn
Bialystok	Bialystok	Chojnice	Bialystok	Bialystok	Bialystok	Bialystok	Bialystok
Chojnice	Chojnice	Kalisz	Chojnice	Chojnice	Chojnice	Chojnice	Chojnice

Table 13 can be used as a guide to quickly visualize which stations are both at the top and bottom of the ranking and where they stand in the relative ranking. From this quick screening, nine stations have a mix signal, they either appear in the bottom or/and top: Jarczew, KasprowyWie, Leba, Legnica, Lesko, NowySacz, Poznan, PuszczBorec and Suwalki. This once again reflect how different the results are depending of the compound being considered

Some of these stations, such as Bialystok and Chojnice, are examples that one might think they are not similar. Note that the analysis considering the temporal and magnitude variation separately, these are not considered similar, see figures in A4 and A5 in the Annex for metrics 1-R and EuD respectively. It is only when we combine both metrics that these show up as very similar stations. Another note is that the analysis based on the model result does put these stations in different monitoring regions.

4.3 Combining modeling and observations

The modelling and observation results can be combined to assess the optimization of the current network. As presented in tables in Section 3, if stations fall within the same cluster based on the modeling results (regions) then further assessment is needed to indicate which of the stations are unique (most dissimilar) or may be redundant (two or more stations measuring similar patterns of temporal and magnitude variation) and, possibly, one or more stations be relocated.

The figure below shows one way of combining the observations and the modelling data results. It also serves to check the potential differences between measurements and modelling results. Note that this analysis includes mostly stations monitoring urban areas and that the model does not represent these areas accurately because of the coarse resolution (~10 and 40 km for acidic compounds and metals, respectively). This might be the reason why we see some unexpected clustering at relatively low levels of dissimilarity.

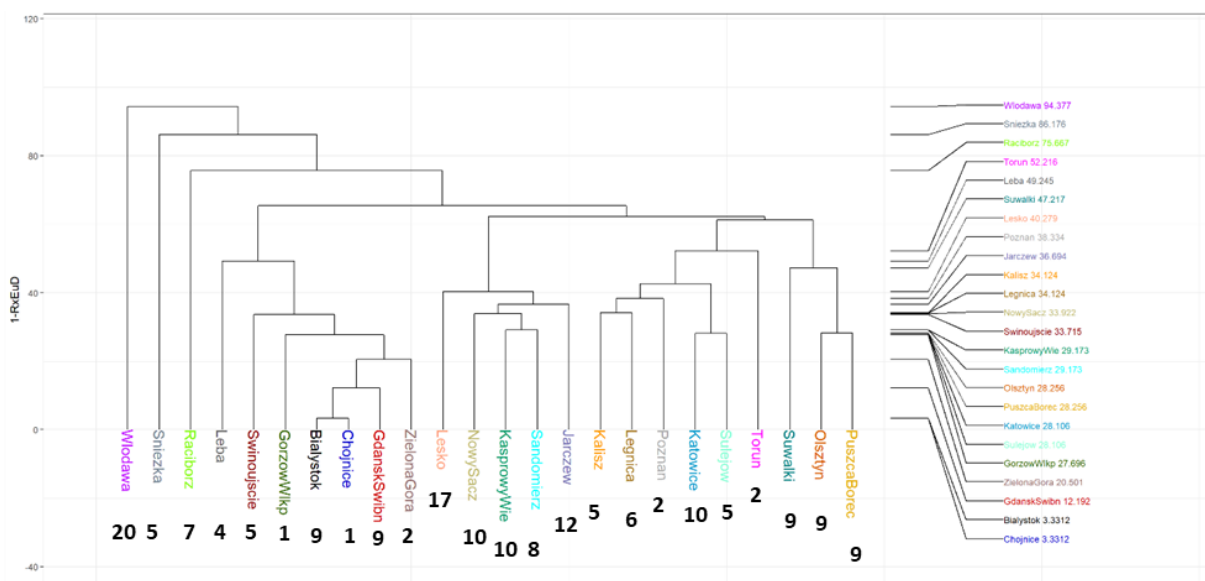


Figure 26: SO4 observation-based dendrogram and cluster number where the station falls into based on the modelled data (Figure 3).

5 Overall conclusions

Areas have been identified for a potential re-design of the current monitoring network in Poland. The locations to be considered should be within larger regions and never small. The latter indicates the presence of sources, and the monitoring should be set-up stations downwind of a source.

It was expected that the results would vary depending on the parameter analysed. It does seem to indicate that the acidic compounds need monitoring across the country, whereas the metals have a clearer distinction between north-south and east-west. However, it appears that a denser network is needed in the south-east part of the country and less monitoring in the north-west, independently of the compound.

There is a need to keep stations monitoring the Polish coast, particularly for shipping-related pollution. A station in Gdansk seems to have the most variety of air masses and could be a unique monitoring location. The footprint shows that many monitoring sites are influenced mainly by Polish sources but occasionally from neighbouring countries such as the Czech Republic and Slovakia.

If the monitoring network were comprised of the stations that we have the data available for, the centre of Poland would not be monitored at all, showing a gap in the monitoring network.

The relative ranking based on the observations indicates the following:

- six stations that recurrently are in the bottom part of the ranking: Chojnice, Bialystok, Gdansk Swibn, ZielonaGora, Gorzow Wlkp and Kalisz. These stations should be targeted first if to consider changes in the monitoring network;
- five stations are recurrently at the top of the ranking: Sniezka, Raciborz, Swinoujscie, Wlodawa, and Olsztyn. These stations seem to be unique in the network;
- Katowice, Lesko, NowySacz, Olsztyn, Poznan, Raciborz, Sandomierz, , Sulejow, Swinoujscie, Torun, Wlodawa
- nine stations have a mix signal, they either appear in the bottom or/and top: Jarczew, KasprowyWie, Leba, Legnica, Lesko, NowySacz, Poznan, PuszczaBorec and Suwalki. This once again reflect how different the results are depending of the compound being considered

The results based on the modelling and measurements show there may be some discrepancy between model results and observations. This discrepancy may be due to most of the sites being located at urban background areas when the model does not represent urban areas accurately, especially runs concerning metals (~40km resolution). Cases such as the high similarity based on observations between Chojnice and Bialystok are to be investigated. The model shows that the stations should be measuring different regimes (combination of meteorology and emissions). This is expected since they are geographically far apart. However, the observations say otherwise for all compounds available for the analysis.

6 References

- Eckhardt, S., Cassiani, M., Evangeliou, N., Sollum, E., Pisso, I., and Stohl, A. (2017). Source-receptor matrix calculation for deposited mass with the Lagrangian particle dispersion model FLEXPART v10.2 in backward mode. *Geoscientific Model Development*, *10*, 4605-4618, <https://doi.org/10.5194/gmd-10-4605-2017>.
- Emanuel, K. A., and Zivkovic-Rothman, M. (1999). Development and evaluation of a convection scheme for use in climate models. *Journal of the Atmospheric Sciences*, *56*, 1766-1782, 1999.
- Forster, C., Stohl, A., and Seibert, P. (2007). Parameterization of convective transport in a Lagrangian particle dispersion model and its evaluation. *Journal of Applied Meteorology and Climatology*, *46*, 403-422, <https://doi.org/10.1175/jam2470.1>.
- Ilyin, I., Travnikov, O., Rozovskaya, O., Strizhkina, I. (2022). Trends in deposition of heavy metals to the OSPAR Maritime Area in the period 1990-2019 (EMEP/MSC-E Technical Report 1/2022). https://oap-cloudfront.ospar.org/media/filer_public/e6/08/e608e1bd-8ed3-45a3-b748-e55fec7861de/p00897_emep_e_trends_in_heavy_metals.pdf.
- Ilyin, I., Batrakova, N., Gusev, A., Kleimenov, M., Rozovskaya, O., Shatalov, V., Strizhkina, I., Travnikov, O., Breivik, K., Halvorsen, H. L., Nizzetto, P. B., Pfaffhuber, K. A., Aas, W., Mareckova, K., Poupa, S., Wankmueller, R., Ullrich, B., Degorska, A. (2021). *Heavy metals and POPs: Pollution assessment of toxic substances on regional and global scales*. (EMEP Status Report 2/2021).
- Klein, H., Gauss, M., Nyiri, A., Tsyro, S., and Fagerli, H. (2021). *Transboundary air pollution by sulphur, nitrogen, ozone and particulate matter in 2019* (MSC-W Data Note 1/2021). Oslo: Norwegian Meteorological Institute, https://www.emep.int/publ/reports/2021/Country_Reports/report_PL.pdf.
- MSC-E (2022). *Data of HMs and POPs for the EMEP region* <https://www.msceast.org/index.php/pollution-assessment/emep-domain-menu/data-hm-pop-menu>, last accessed 28.02.2022.
- MSC-W (2022). *EMEP MSC-W modelled air concentrations and depositions* https://emep.int/mscw/mscw_moddata.html, last accessed 28.02.2022.
- McConnell, J. R., Chellman, N. J., Wilson, A. I., Stohl, A., Arienzo, M. M., Eckhardt, S., Fritzsche, D., Kipfstuhl, S., Opel, T., Place, P. F., and Steffensen, J. P. (2019). Pervasive Arctic lead pollution suggests substantial growth in medieval silver production modulated by plague, climate, and conflict. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 14910-14915, <https://doi.org/10.1073/pnas.1904515116>.
- Næs, T., Brockhoff, P.B., and Tomic, O. (2010). *Statistics for Sensory and Consumer Science*, 6th ed., Wiltshire: John Wiley & Sons.
- Pisso, I., Sollum, E., Grythe, H., Kristiansen, N. I., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Zwaafink, C. D. G., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhardt, J. F., Fouilloux, A., Brioude, J., Philipp, A., Seibert, P., and Stohl, A. (2019) The Lagrangian particle dispersion model FLEXPART version 10.4. *Geoscientific Model Development*, *12*, 4955-4997, <https://doi.org/10.5194/gmd-12-4955-2019>.
- Seibert, P., and Frank, A. (2004). Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode. *Atmospheric Chemistry and Physics*, *4*, 51-63.
- Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L. D., Fagerli, H., Flechard, C. R.,

- Hayman, G. D., Gauss, M., Jonson, J. E., Jenkin, M. E., Nyíri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, Á., and Wind, P. (2012). The EMEP MSC-W chemical transport model – technical description. *Atmospheric Chemistry and Physics*, *12*, 7825–7865, <https://doi.org/10.5194/acp-12-7825-2012>.
- Soares, J., Makar, P. A., Aklilu, Y., and Akingunola, A. (2018a). *Hierarchical clustering network analysis of ambient air monitoring in Alberta - Phases 1 and 2* (Oil Sands Monitoring Program Technical Report Series No. 5.0). <https://open.alberta.ca/publications/9781460141199>
- Soares, J., Makar, P. A., Aklilu, Y., and Akingunola, A. (2018b). The use of hierarchical clustering for the design of optimized monitoring networks. *Atmospheric Chemistry and Physics*, *18*(9), 6543–6566. <https://doi.org/10.5194/acp-18-6543-2018>.
- Strizhkina I., Ilyin, I., Rozovskaya, O., and Travnikov, O. (2021). *Heavy metals and POPs: Pollution assessment of toxic substances on regional and global scales. Part I. Supplementary materials for heavy metals* (MSC-E Data Report 1/2021). Moscow: Meteorological Synthesizing Centre – East.
- Travnikov, O., Angot, H., Artaxo, P., Bencardino, M., Bieser, J., D'Amore, F., Dastoor, A., De Simone, F., Diéguez, M. D. C., Dommergue, A., Ebinghaus, R., Feng, X. B., Gencarelli, C. N., Hedgecock, I. M., Magand, O., Martin, L., Matthias, V., Mashyanov, N., Pirrone, N., Ramachandran, R., Read, K. A., Ryjkov, A., Selin, N. E., Sena, F., Song, S., Sprovieri, F., Wip, D., Wängberg, I., and Yang, X. (2017). Multi-model study of mercury dispersion in the atmosphere: atmospheric processes and model evaluation. *Atmospheric Chemistry and Physics*, *17*, 5271–5295, <https://doi.org/10.5194/acp-17-5271-2017>.

Annex

A.1 Modelled regions (clusters): 25 regions representing the areas of representativeness of a single station based on $(1-R) \times \text{EuD}$ metrics

A.1.1 SO_x

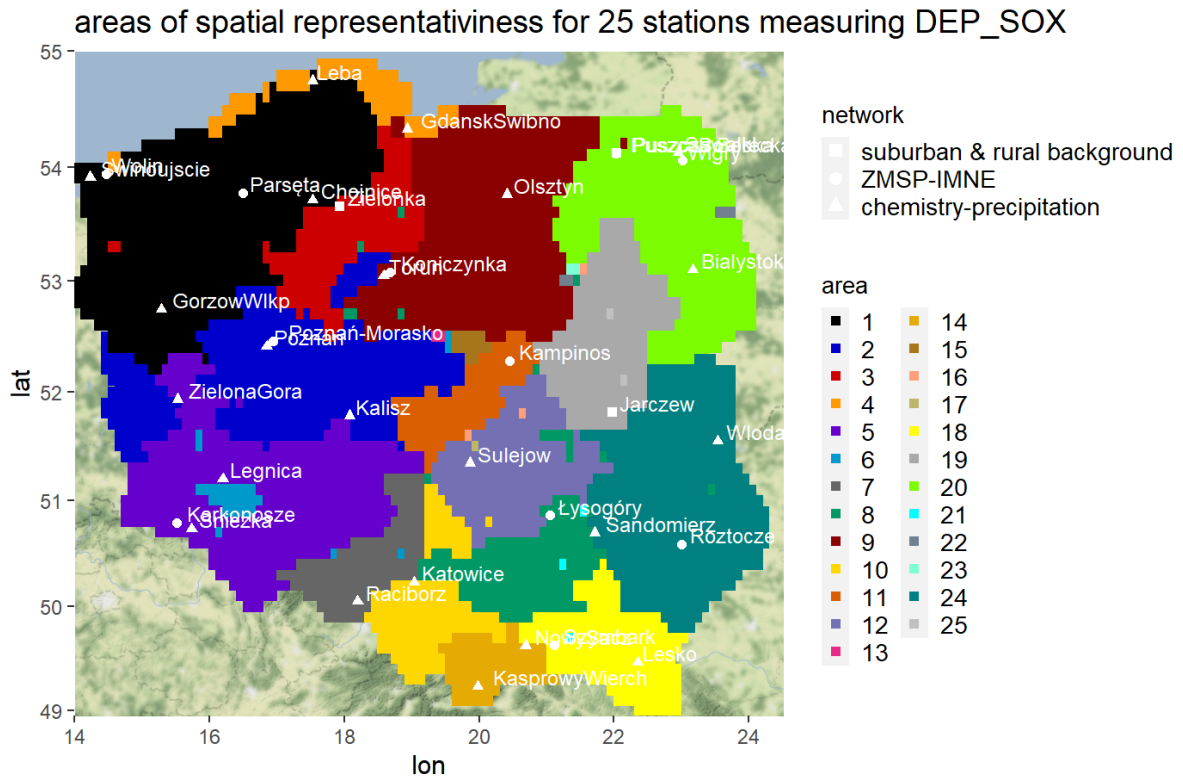


Figure 26: Areas of spatial representativeness for 25 stations measuring SO_x deposition.

Table 14: Cluster information based on the results presented in Figure 5 and Figure 27 for the monitoring stations listed in Table 2.

25 regions	name
1	Chojnice
1	GorzowWlkp
1	Parseta
1	Wolin
2	Poznan
2	Torun
2	ZielonaGora
2	Poznan-Morasko
3	Zielonka
4	Leba
5	Kalisz
5	Sniezka
5	Swinoujscie
5	Karkonosze
6	Legnica
7	Raciborz
8	Sandomierz
8	Lysogóry
9	GdanskSwibno
9	Olsztyn
9	Koniczynka
10	Katowice
11	Kampinos
12	Sulejow
14	KasprowyWierch
14	NowySacz
18	Lesko
18	Szymbark
19	Jarczew
20	Bialystok
20	Suwalki
20	PuszczaBorecka
20	Puszcza Borecka
20	Wigry
24	Wlodawa
24	Roztocze

A.1.2 OXN

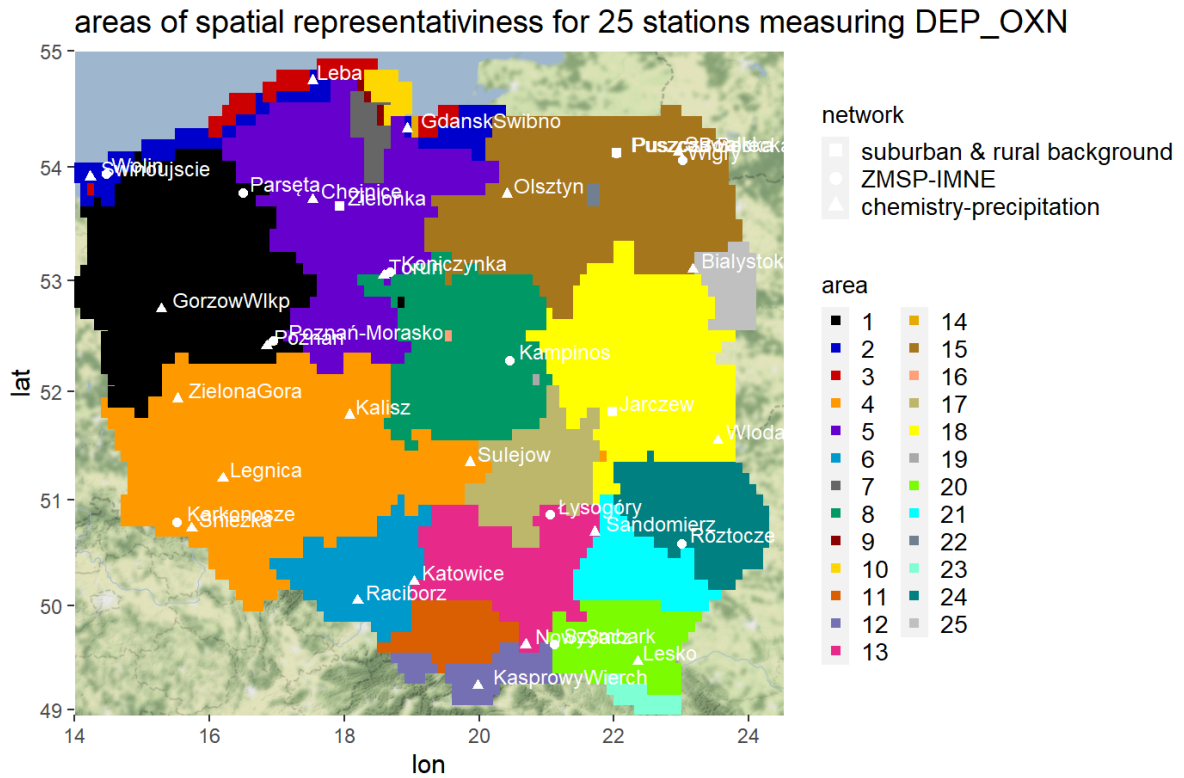


Figure 27: Areas of spatial representativeness for 25 stations measuring OXN deposition.

Table 15: Cluster information based on the results presented in Figure 5 and Figure 27 for the monitoring stations listed in Table 2.

25 regions	name
1	GorzowWlkp
1	Poznan
1	Parseta
1	Poznan-Morasko
2	Wolin
3	Leba
4	Kalisz
4	Legnica
4	Sniezka
4	Sulejow
4	Swinoujscie
4	ZielonaGora
4	Karkonosze
5	Chojnice
5	GdanskSwibno
5	Zielonka
5	Koniczynka
6	Katowice
6	Raciborz
8	Torun
8	Kampinos
12	KasprowyWierch
12	NowySacz
12	Szymbark
13	Sandomierz
13	Lysogóry
15	Bialystok
15	Olsztyn
15	Suwalki
15	PuszczaBorecka
15	Puszcza Borecka
15	Wigry
18	Wlodawa
18	Jarczew
20	Lesko
24	Roztocze

A.1.3 RDN

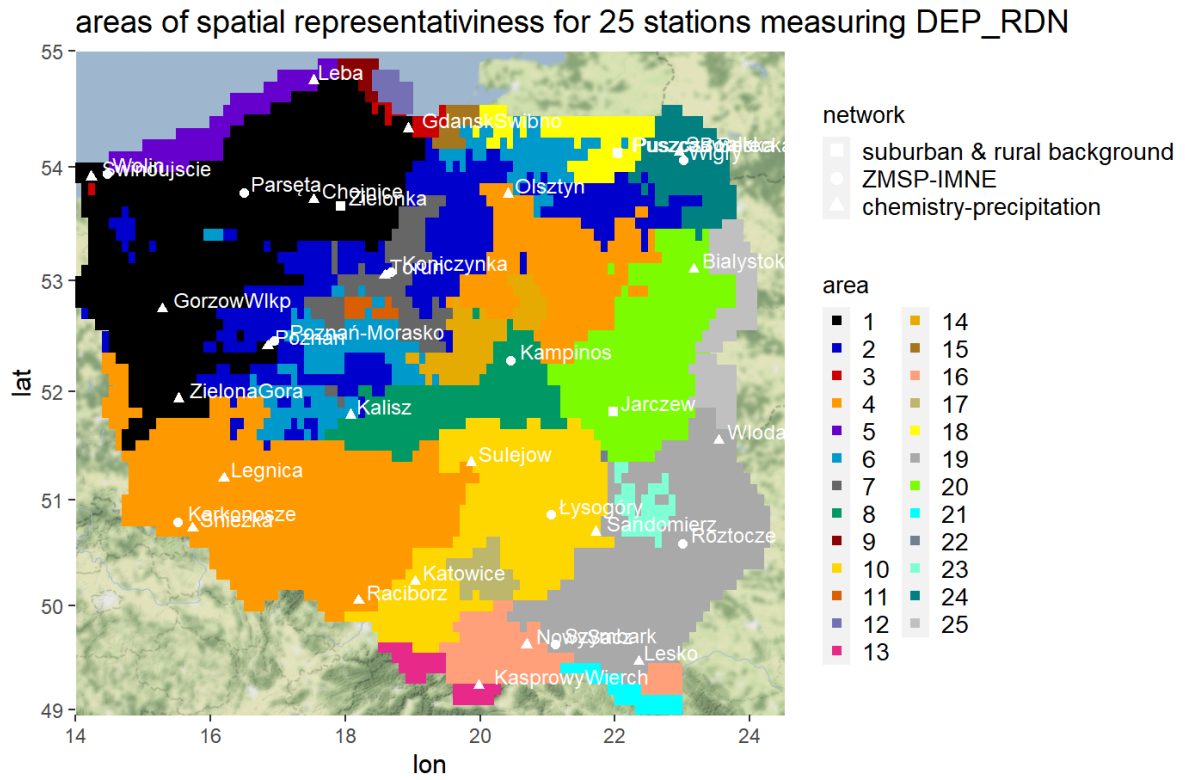


Figure 28: Areas of spatial representivness for 25 stations measuring RDN deposition.

Table 16: Cluster information based on the results presented in Figure 7 and Figure 28 for the monitoring stations listed in Table 2.

25 regions	name
1	GorzowWlkp
1	Legnica
1	Poznan
1	Sniezka
1	Swinoujscie
1	Torun
1	ZielonaGora
1	PuszczaBorecka
1	Karkonosze
1	Parseta
1	Kampinos
1	Lysogóry
1	Puszcza Borecka
1	Poznan-Morasko
1	Roztocze
1	Wigry
1	Koniczynka
1	Szymbark
2	Wolin
4	Chojnice
4	GdanskSwibno
4	Leba
4	Raciborz
4	Jarczew
5	Zielonka
6	Kalisz
8	Katowice
10	NowySacz
14	Sulejow
16	Olsztyn
17	KasprowyWierch
19	Lesko
19	Suwalki
20	Sandomierz
20	Wlodawa
24	Bialystok

A.1.4 Cd

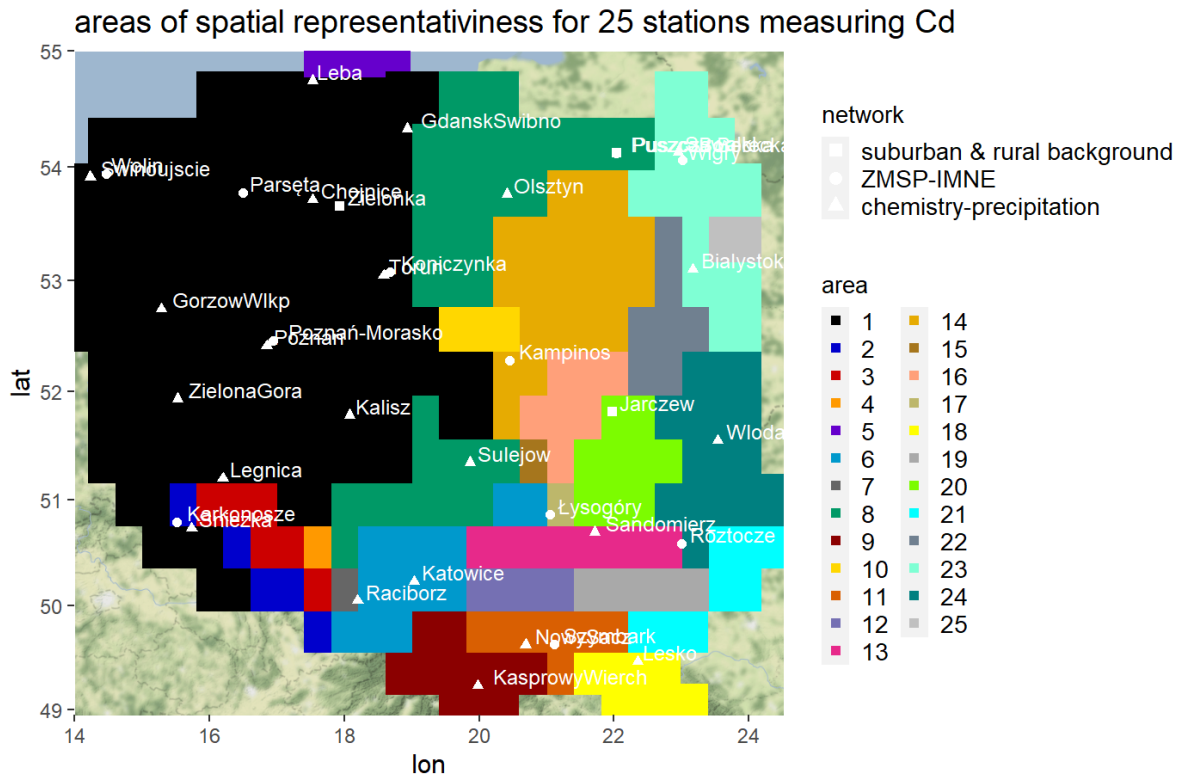


Figure 29: Areas of spatial representativeness for 25 stations measuring Cd deposition.

Table 17: Cluster information based on the results presented in Figure 9 and Figure 29 for the monitoring stations listed in Table 2.

25 regions	name
1	Chojnice
1	GdanskSwibno
1	GorzowWlkp
1	Kalisz
1	Leba
1	Legnica
1	Poznan
1	Sniezka
1	Swinoujscie
1	Torun
1	ZielonaGora
1	Zielonka
1	Parseta
1	Wolin
1	Poznan-Morasko
1	Koniczynka
2	Karkonosze
6	Katowice
6	Lysogóry
7	Raciborz
8	Olsztyn
8	Sulejow
8	PuszczaBorecka
8	Puszcza Borecka
9	KasprowyWierch
11	NowySacz
11	Szymbark
13	Sandomierz
13	Roztocze
14	Kampinos
16	Jarczew
18	Lesko
22	Bialystok
23	Suwalki
23	Wigry
24	Wlodawa

A.1.5 Hg

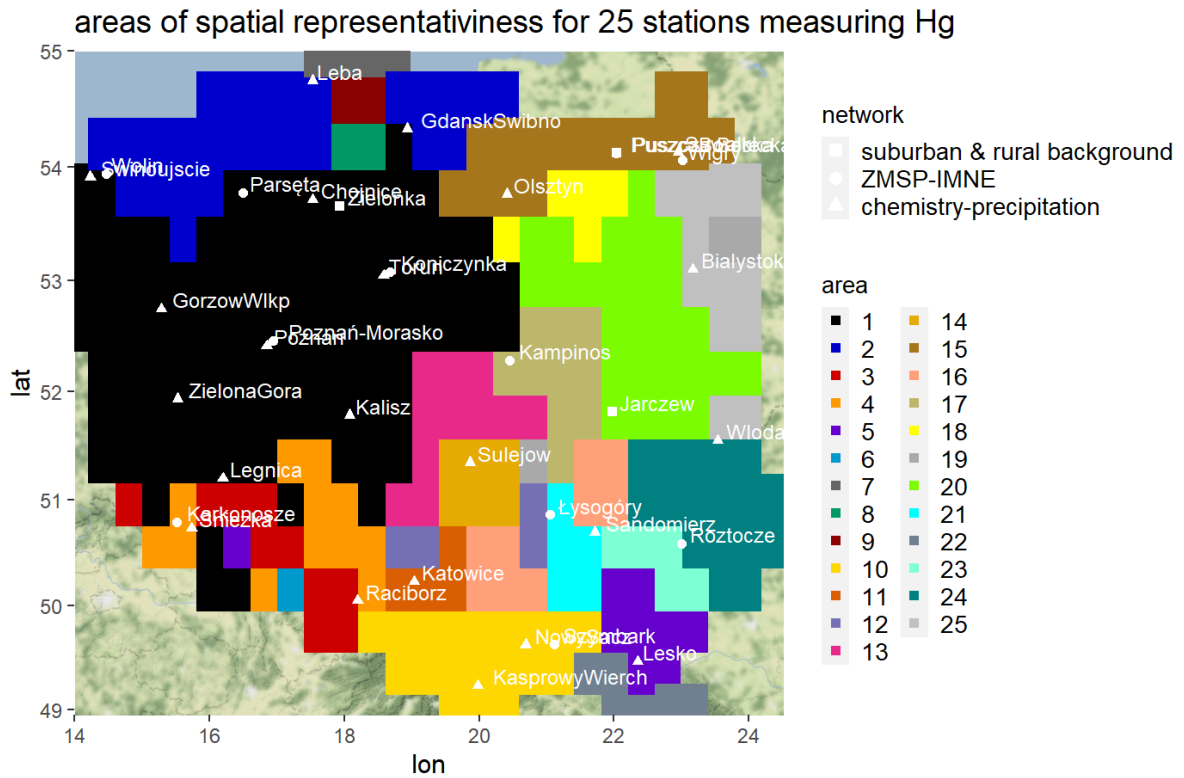


Figure 30: Areas of spatial representativeness for 25 stations measuring Hg deposition.

Table 18: Cluster information based on the results presented in Figure 11 and Figure 30 for the monitoring stations listed in Table 2.

25 regions	name
1	Chojnice
1	GdanskSwibno
1	GorzowWlkp
1	Kalisz
1	Legnica
1	Poznan
1	Swinoujscie
1	Torun
1	ZielonaGora
1	Zielonka
1	Parseta
1	Wolin
1	Poznan-Morasko
1	Koniczynka
2	Leba
3	Raciborz
4	Sniezka
4	Karkonosze
10	KasprowyWierch
10	NowySacz
10	Szymbark
11	Katowice
12	Lysogóry
14	Sulejow
15	Olsztyn
15	Suwalki
15	PuszczaBorecka
15	Puszcza Borecka
15	Wigry
17	Jarczew
17	Kampinos
20	Bialystok
21	Sandomierz
22	Lesko
23	Roztocze
24	Wlodawa

A.1.6 Pb

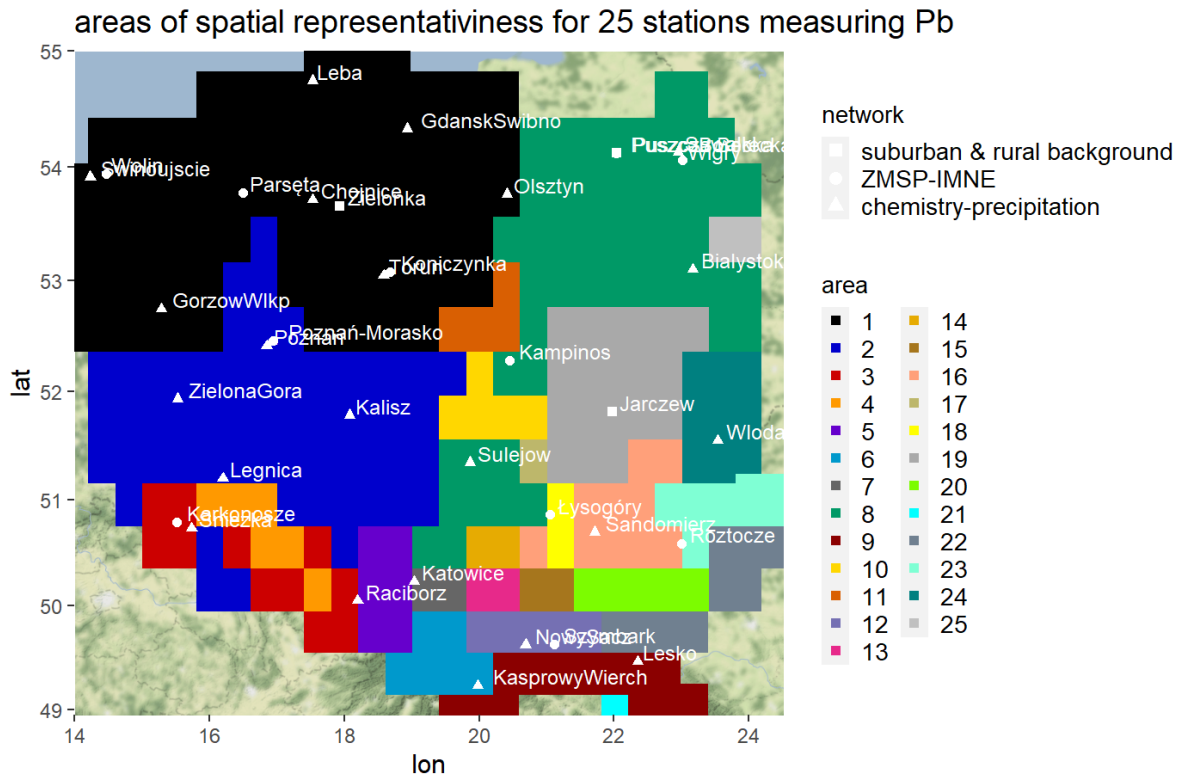


Figure 31: Areas of spatial representativeness for 25 stations measuring Pb deposition.

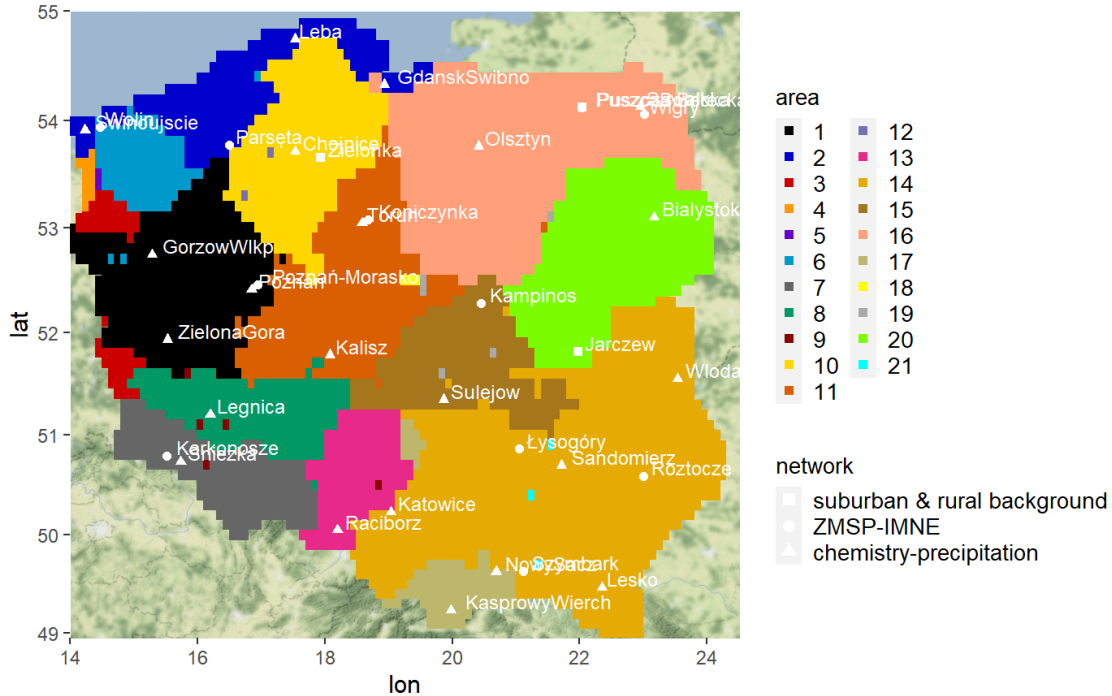
Table 19: Cluster information based on the results presented in Figure 13 and Figure 31 for the monitoring stations listed in Table 2.

25 regions	name
1	Chojnice
1	GdanskSwibno
1	GorzowWlkp
1	Leba
1	Olsztyn
1	Torun
1	Zielonka
1	Parseta
1	Wolin
1	Koniczynka
2	Kalisz
2	Legnica
2	Poznan
2	Swinoujscie
2	ZielonaGora
2	Poznan-Morasko
3	Raciborz
3	Sniezka
3	Karkonosze
5	Katowice
8	Bialystok
8	Sulejow
8	Suwalki
8	PuszczaBorecka
8	Kampinos
8	Lysogóry
8	Puszcza Borecka
8	Wigry
9	KasprowyWierch
9	Lesko
12	NowySacz
12	Szymbark
16	Sandomierz
16	Roztocze
19	Jarczew
24	Wlodawa

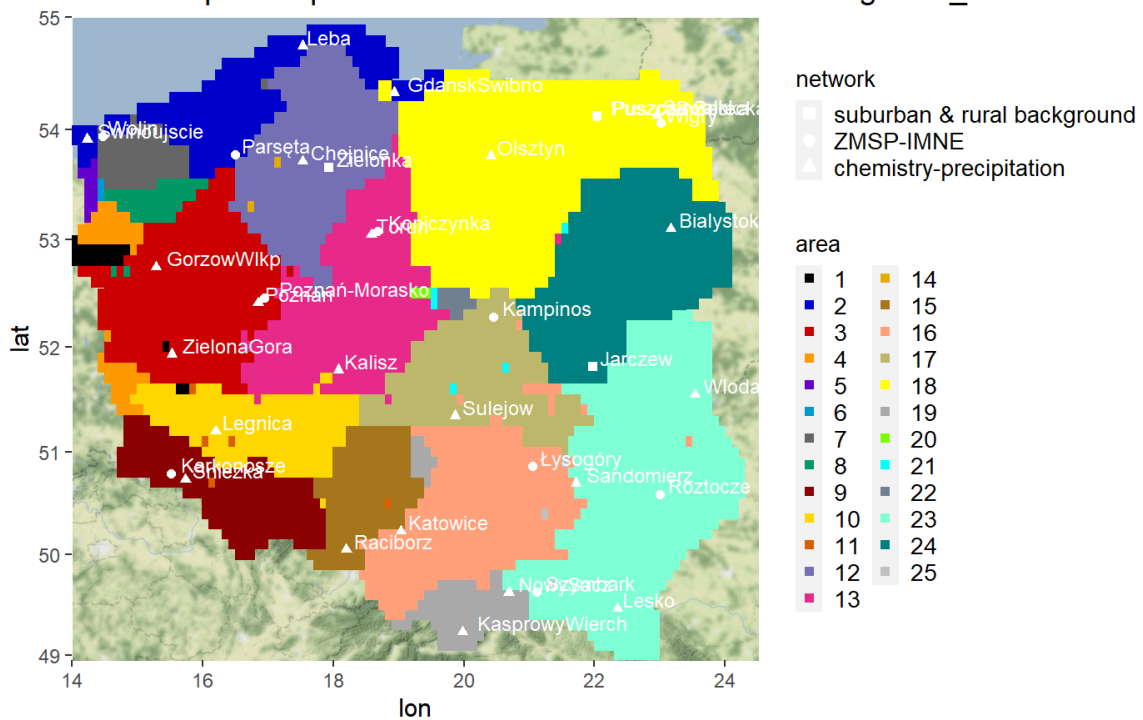
A.2 Modelled regions (clusters): 21 and 25 regions representing the areas of representativeness of a single station based on 1-R (temporal variation)

A2.1 SO_x

areas of spatial representativeness for 21 stations measuring DEP_SOX

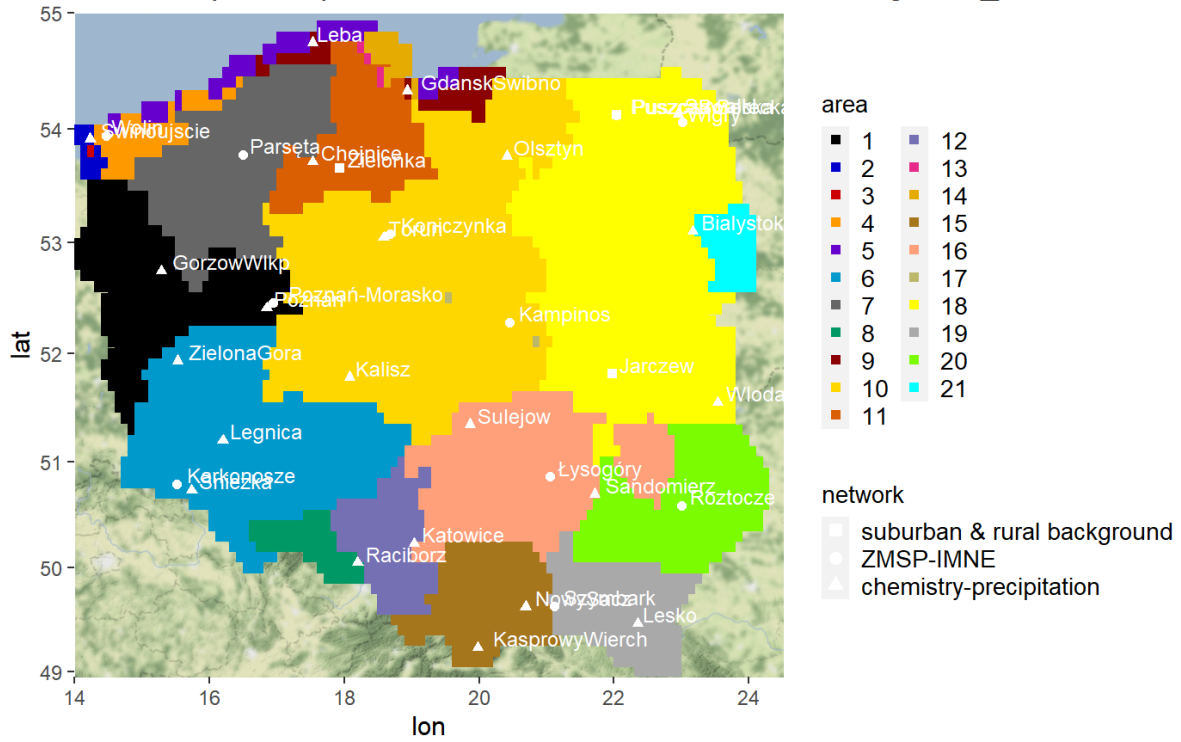


areas of spatial representativeness for 25 stations measuring DEP_SOX

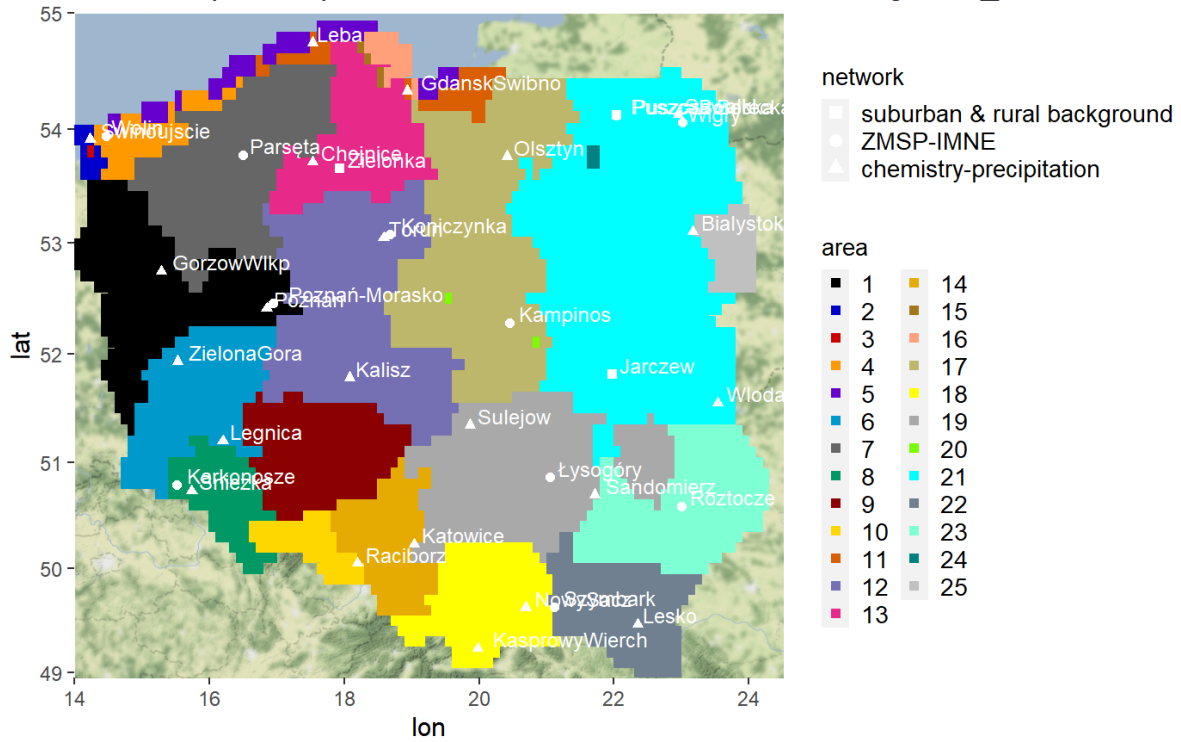


A.2.2 OXN

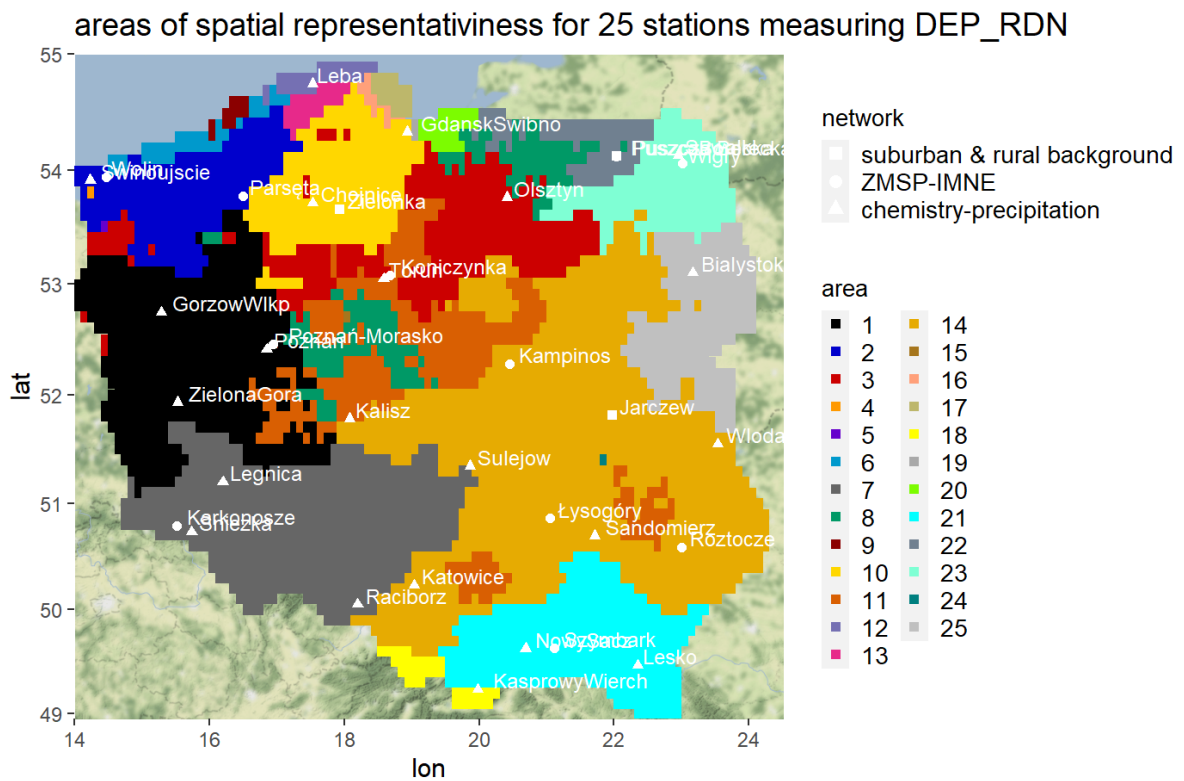
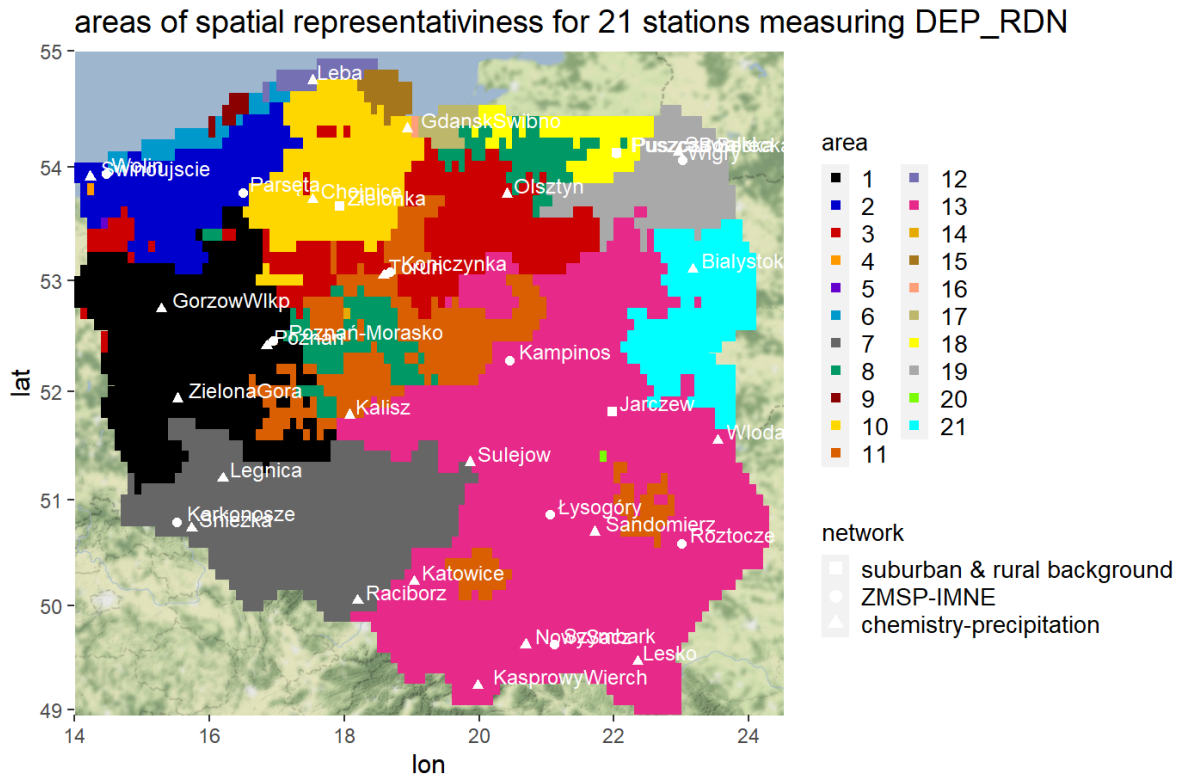
areas of spatial representativeness for 21 stations measuring DEP_OXN



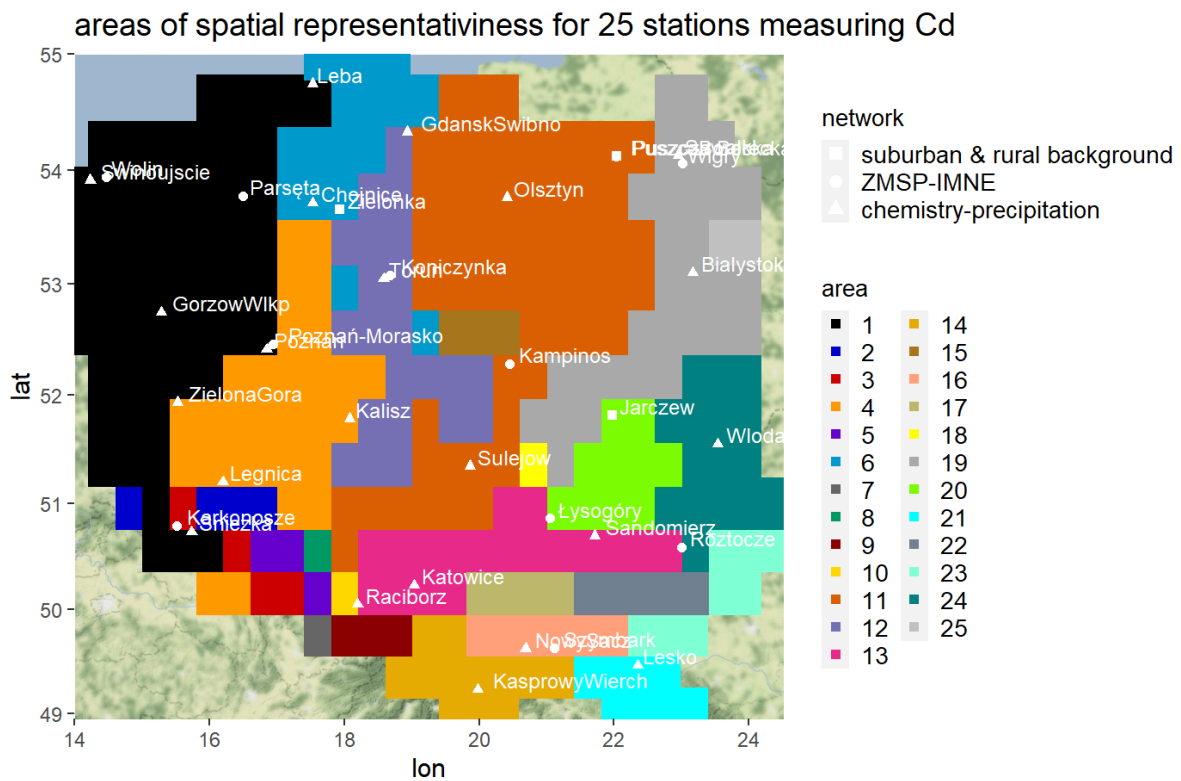
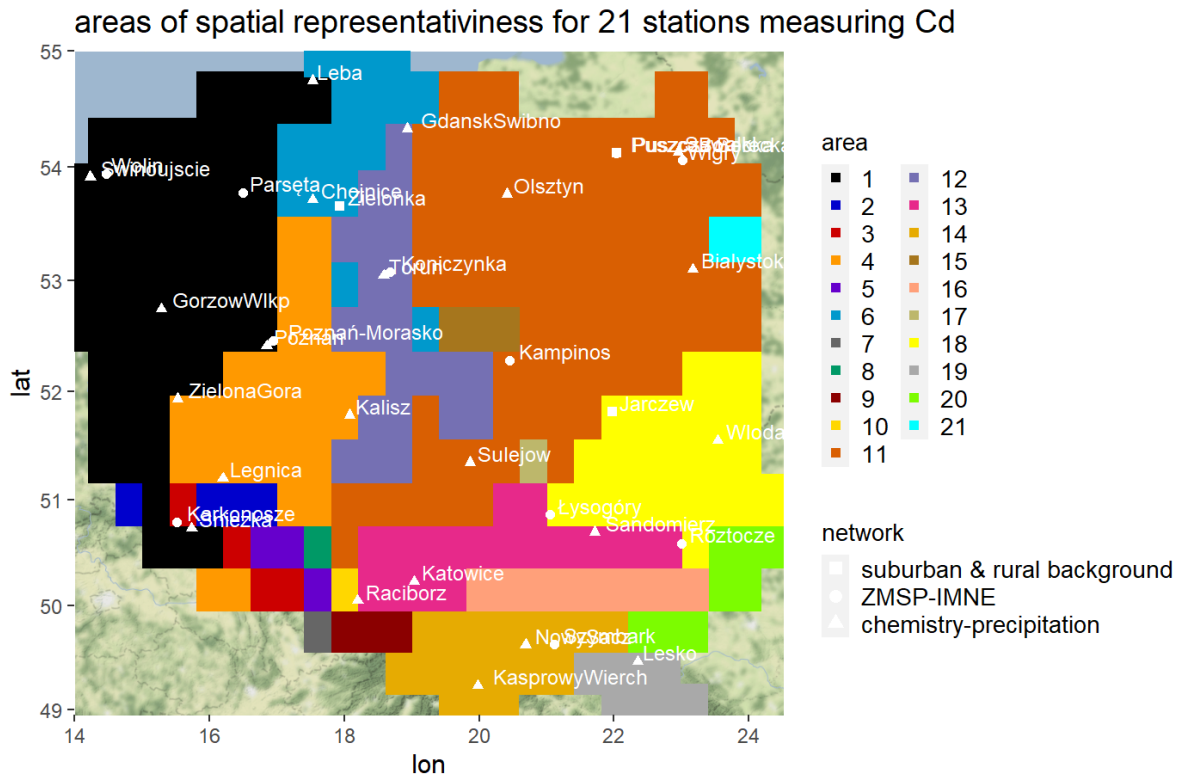
areas of spatial representativeness for 25 stations measuring DEP_OXN



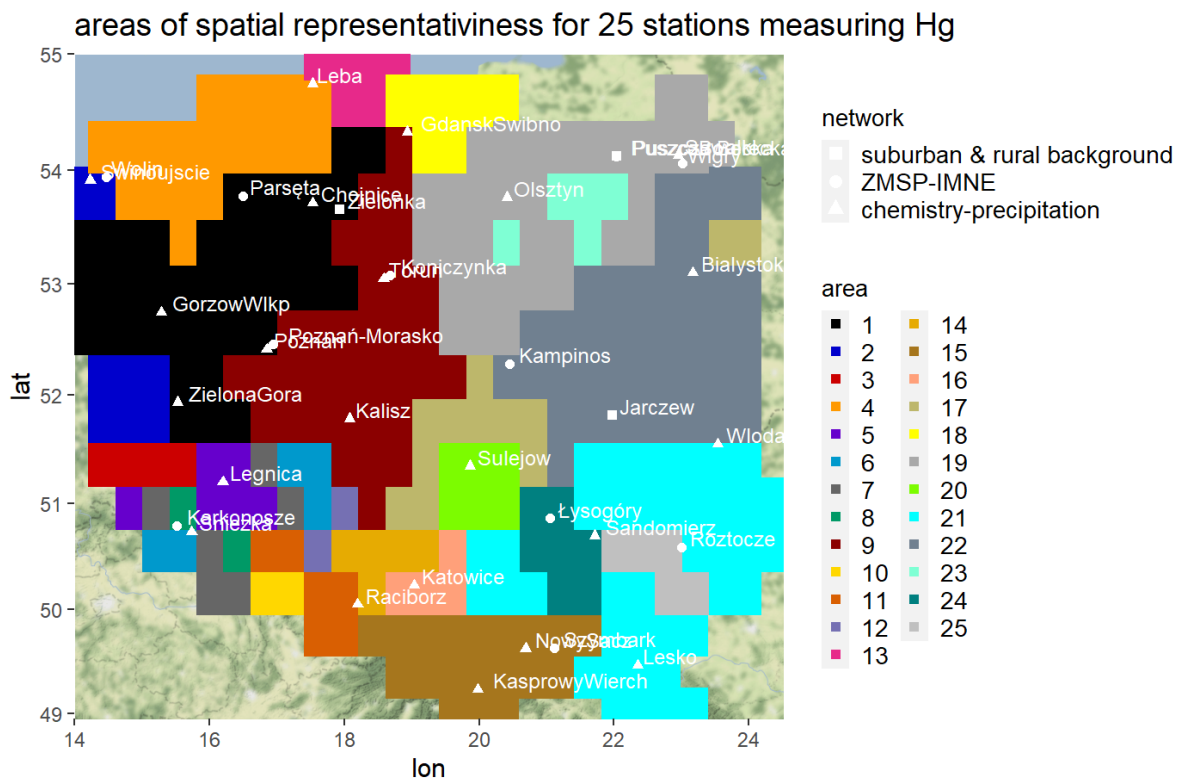
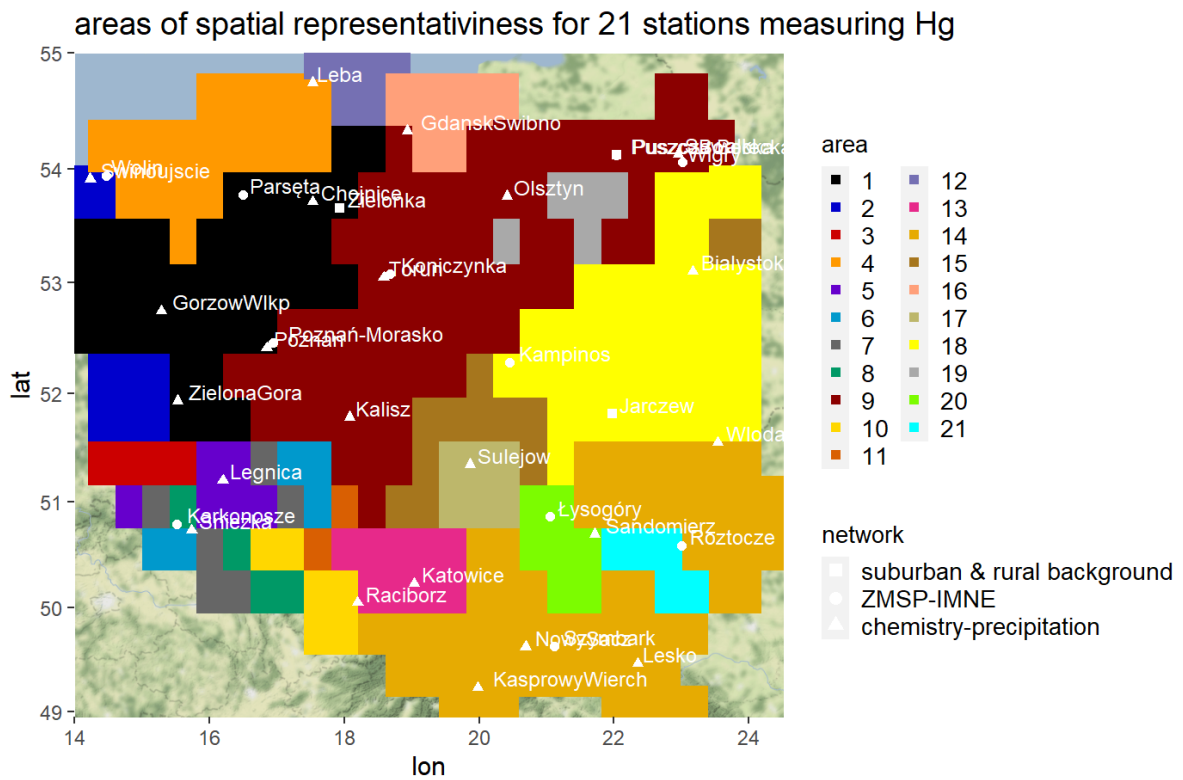
A.2.3 RDN



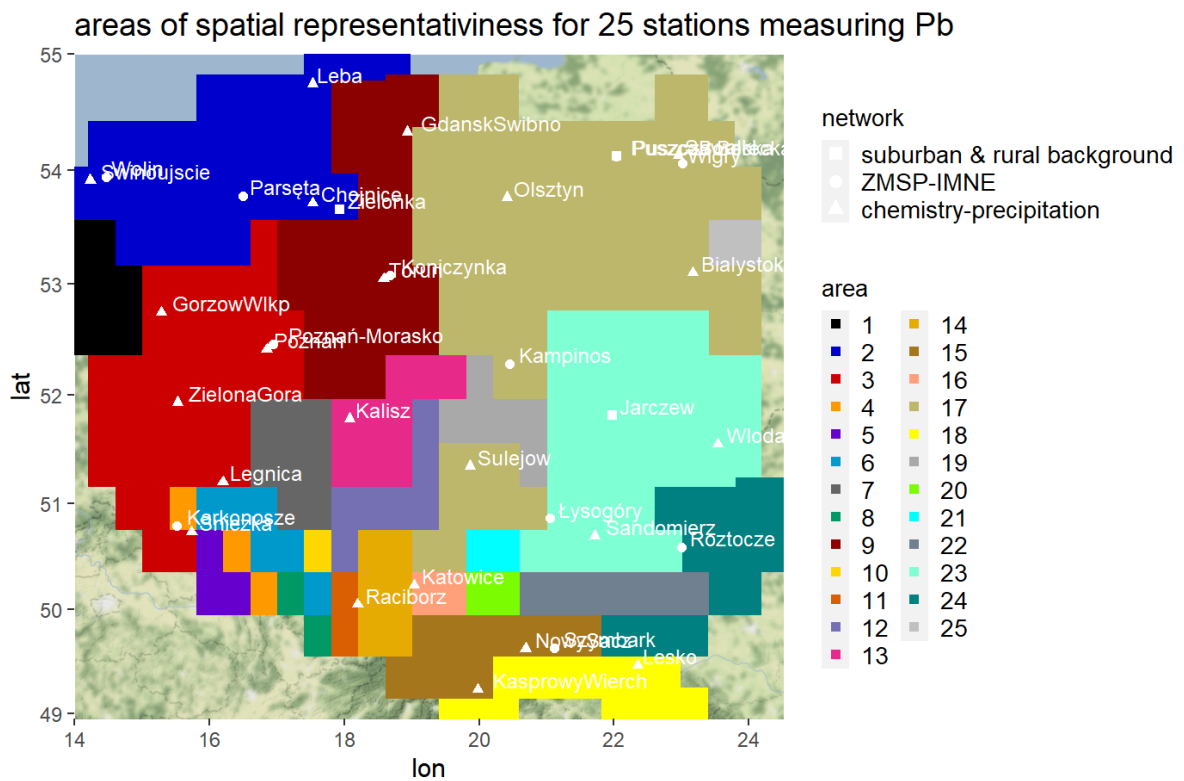
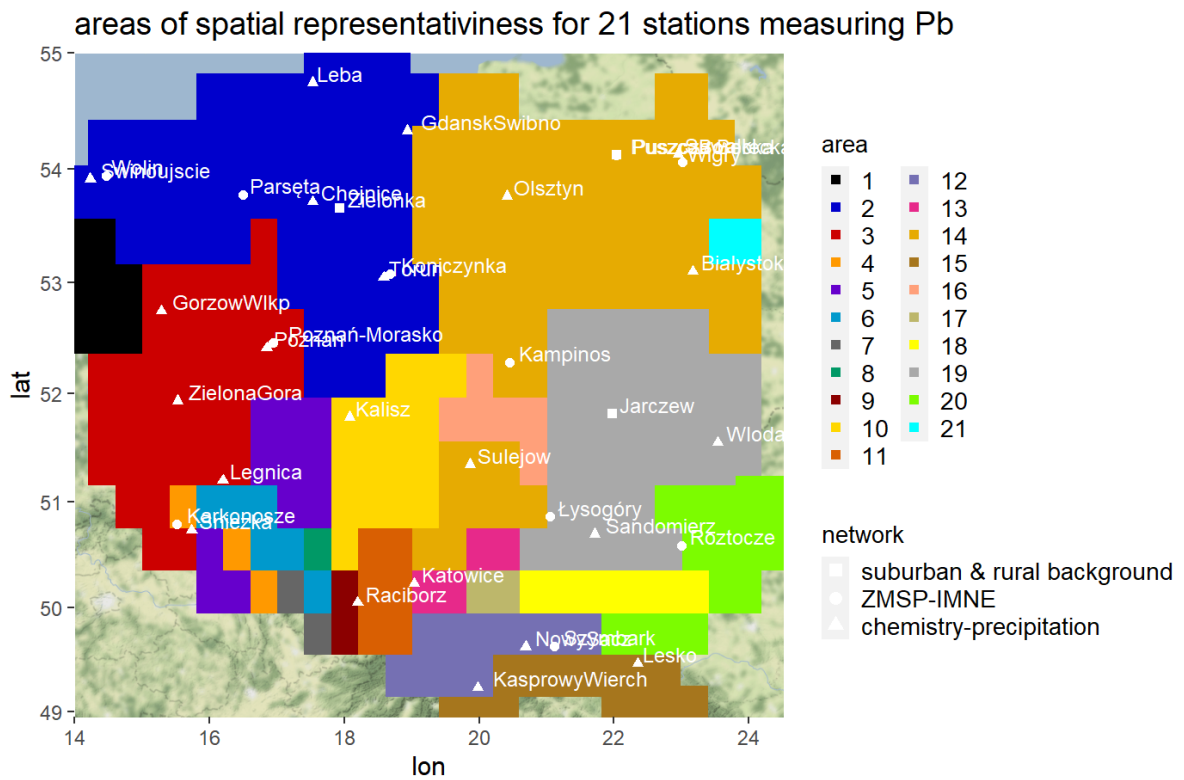
A.2.4 Cd



A.2.5 Hg



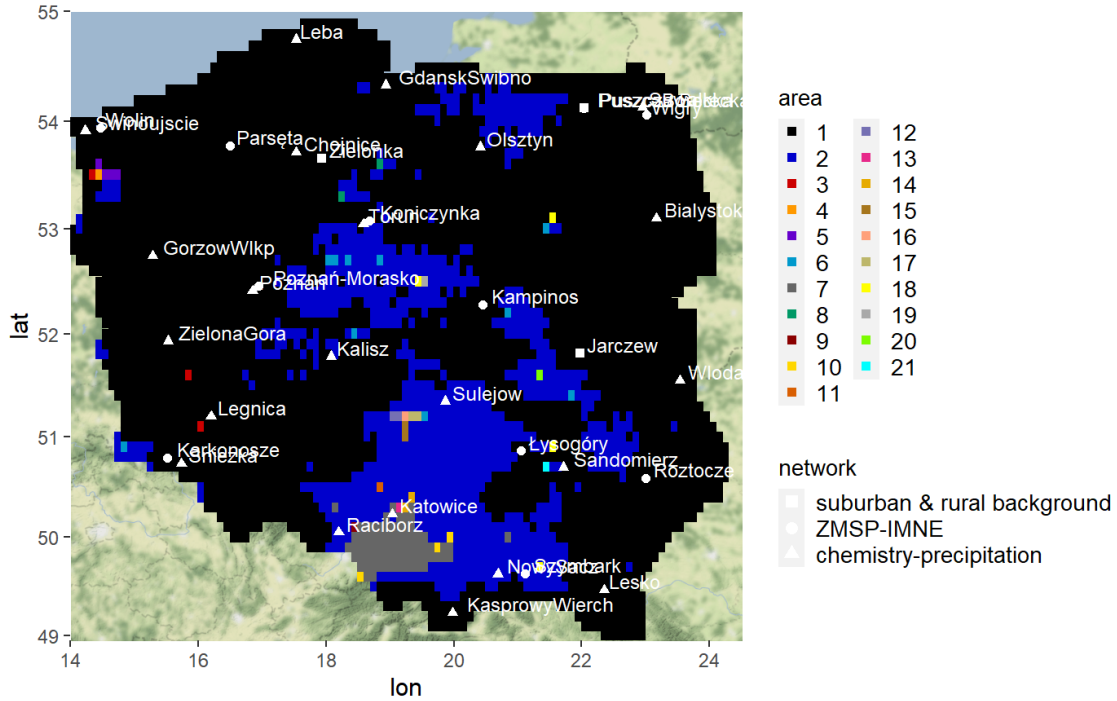
A.2.6 Pb



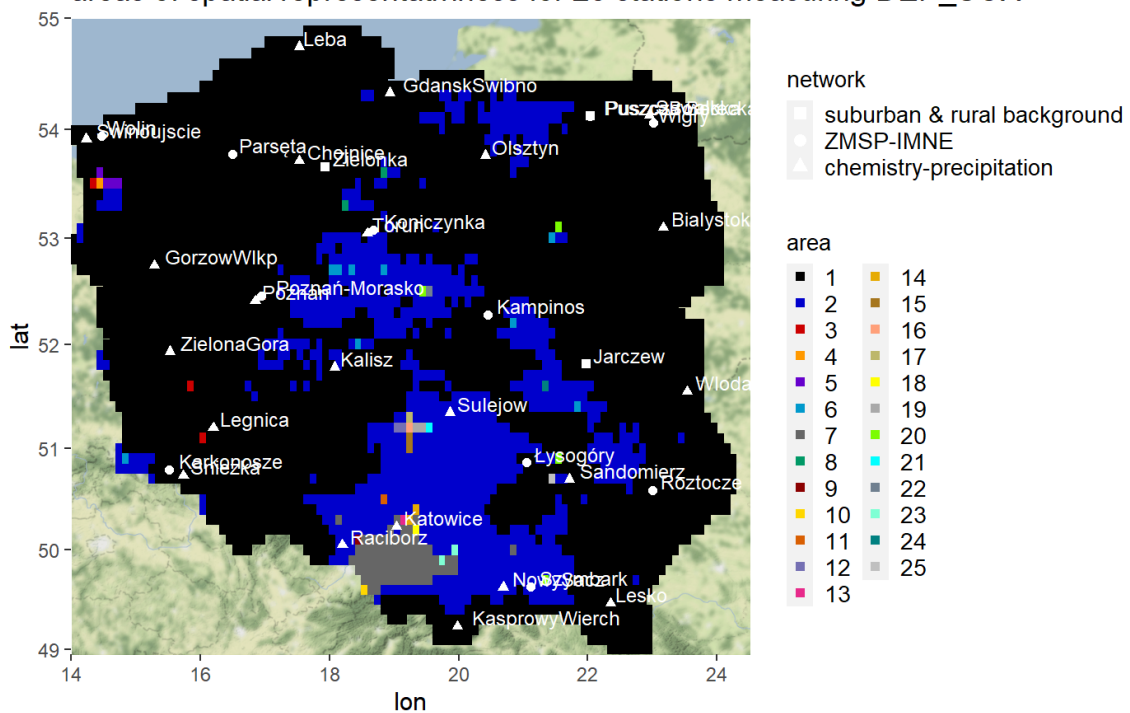
A.3 Modelled regions (clusters): 21 and 25 regions representing the areas of representativeness of a single station based on EuD (magnitude variation)

A.3.1 SO_x

areas of spatial representativeness for 21 stations measuring DEP_SO_x

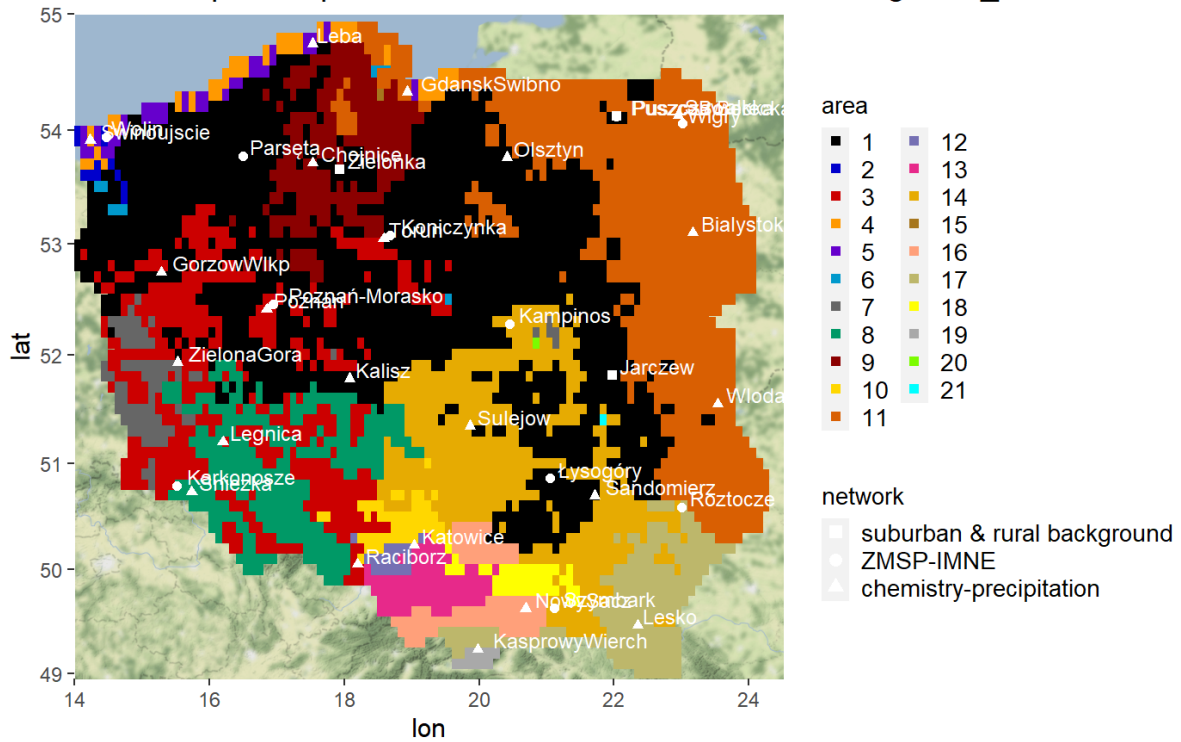


areas of spatial representativeness for 25 stations measuring DEP_SO_x

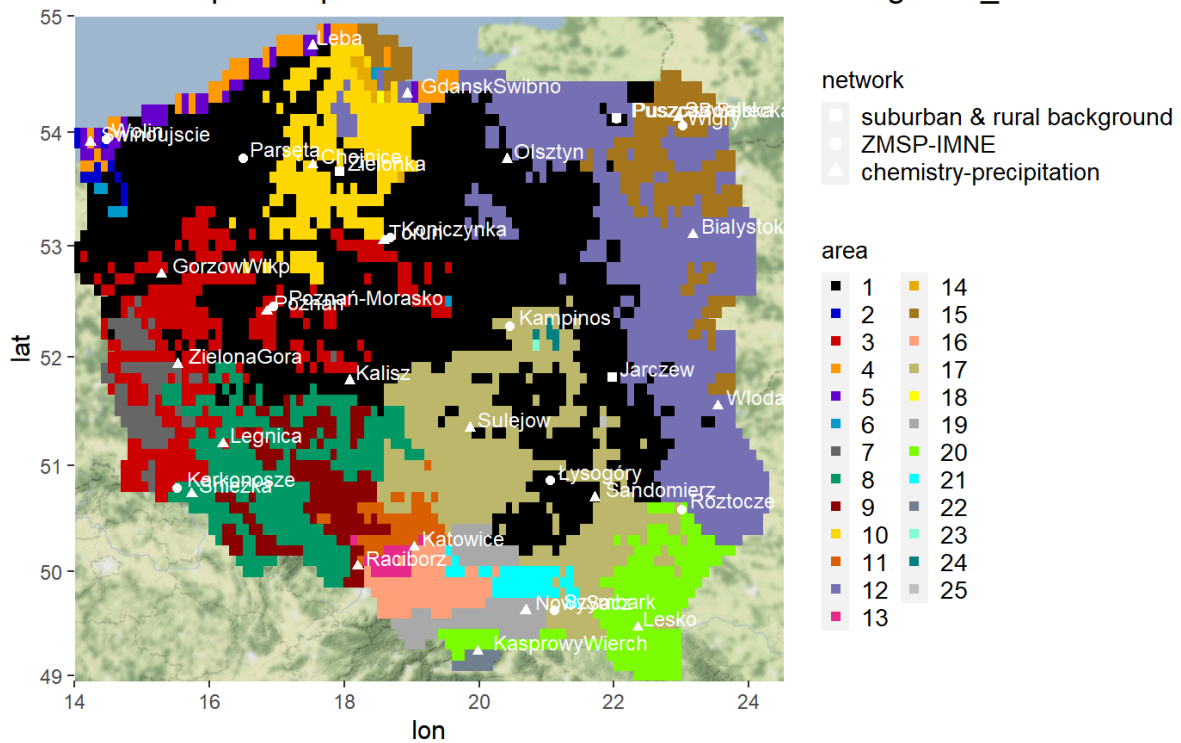


A.3.2 OXN

areas of spatial representativeness for 21 stations measuring DEP_OXN

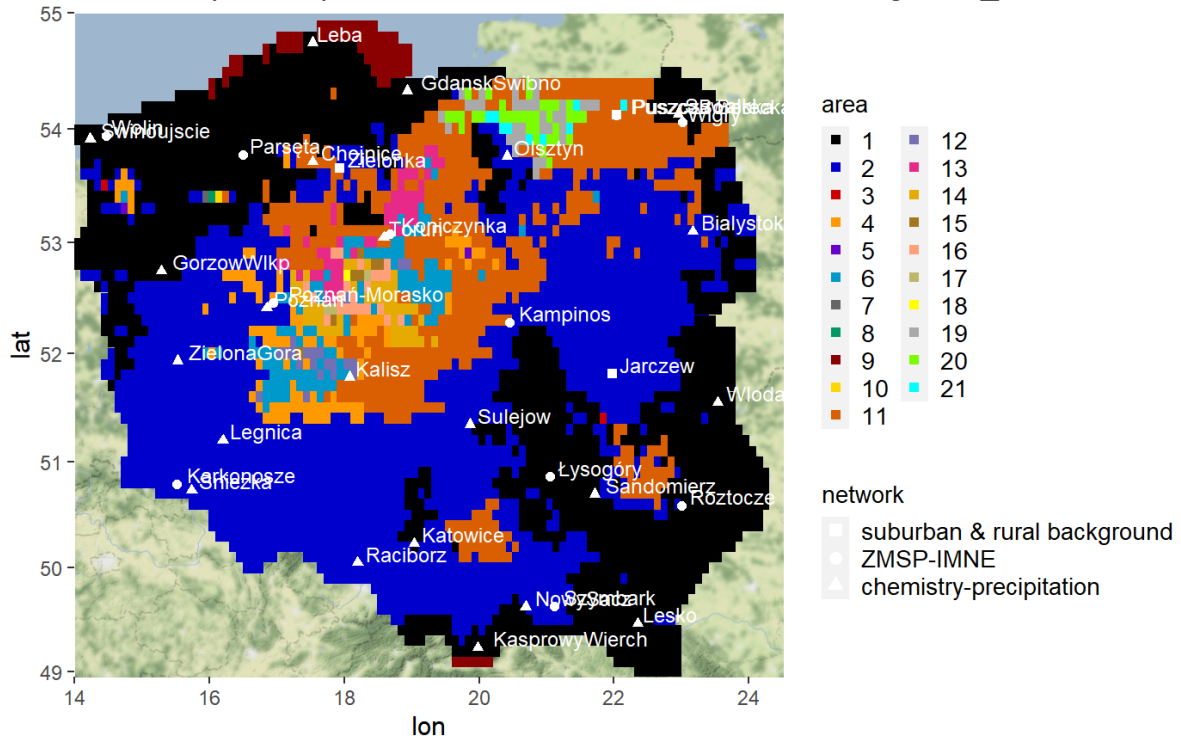


areas of spatial representativeness for 25 stations measuring DEP_OXN

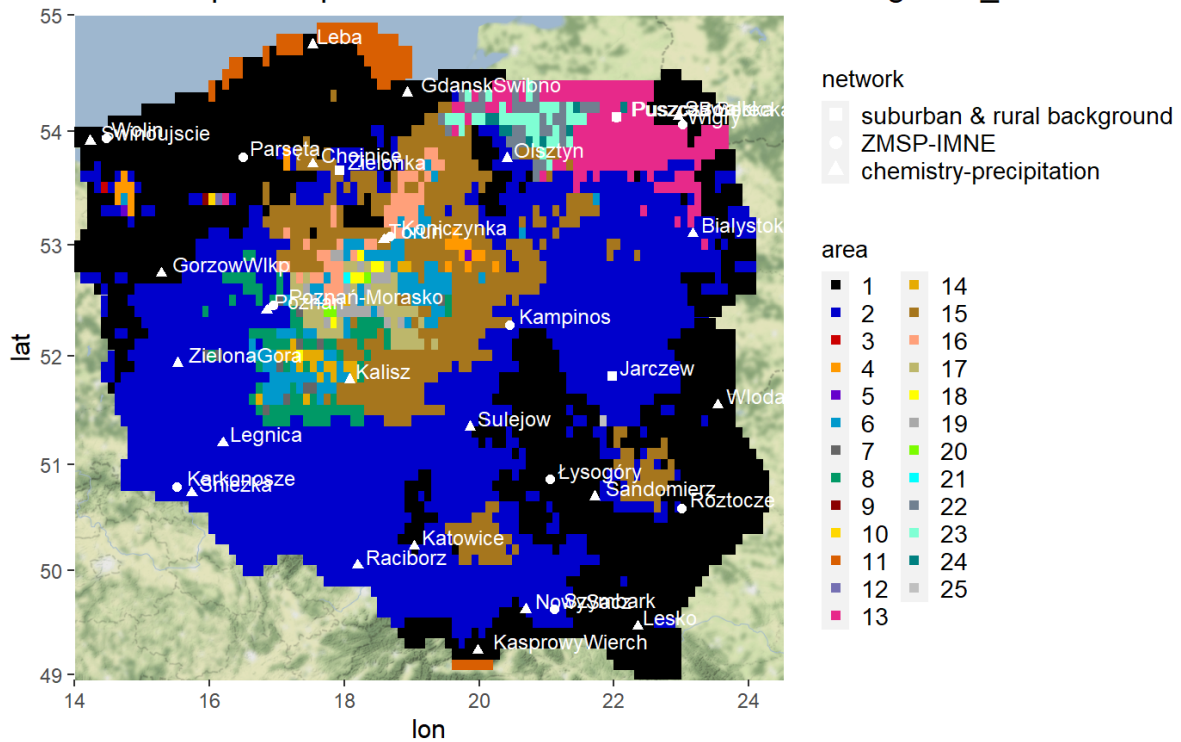


A.3.3 RDN

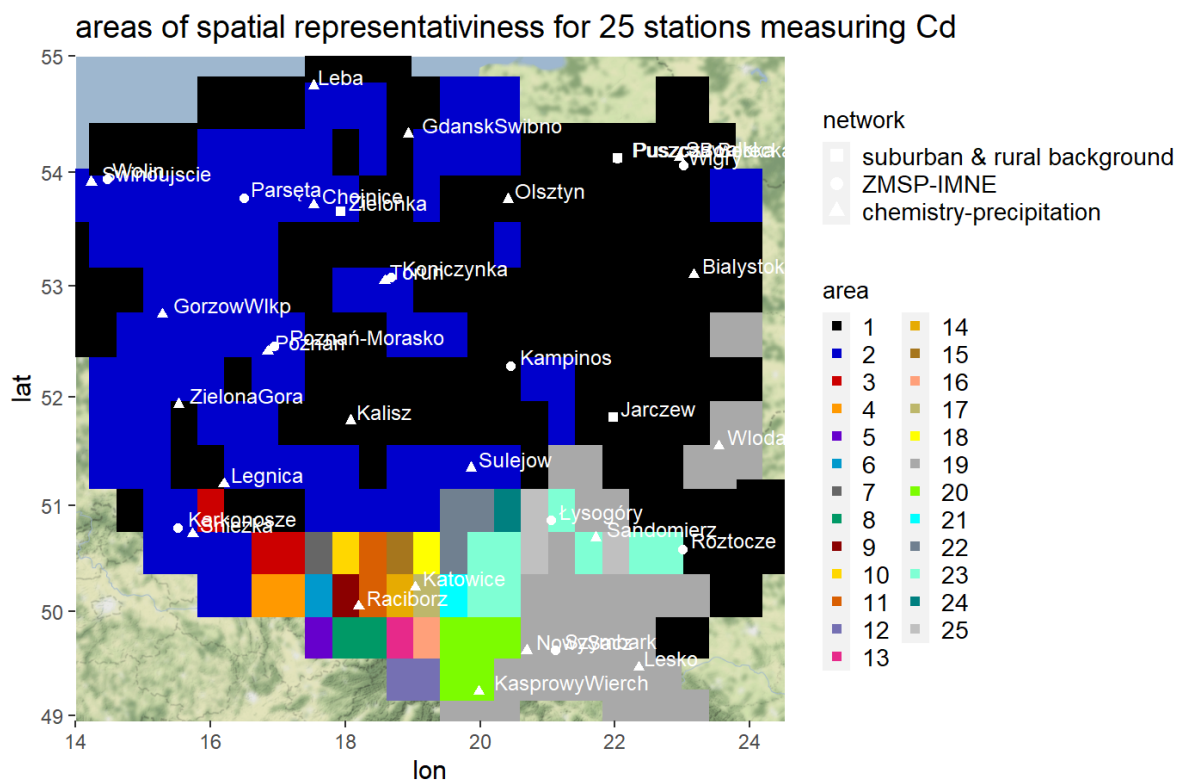
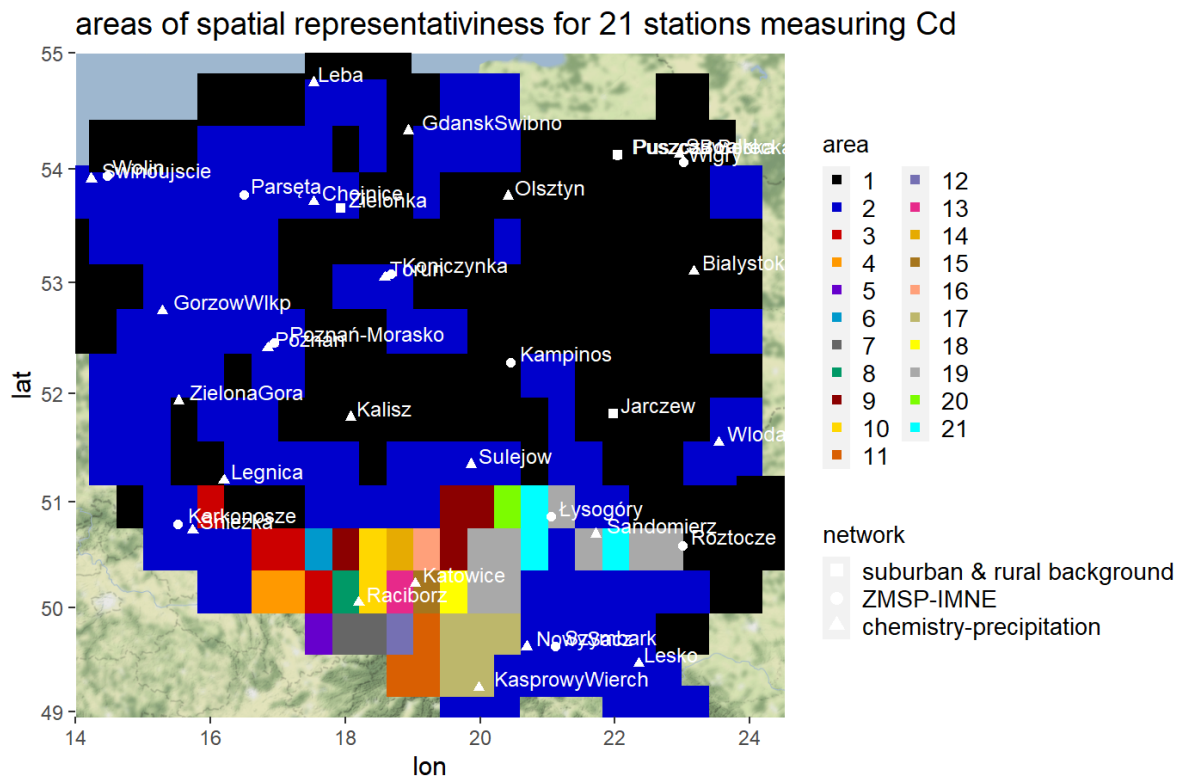
areas of spatial representativeness for 21 stations measuring DEP_RDN



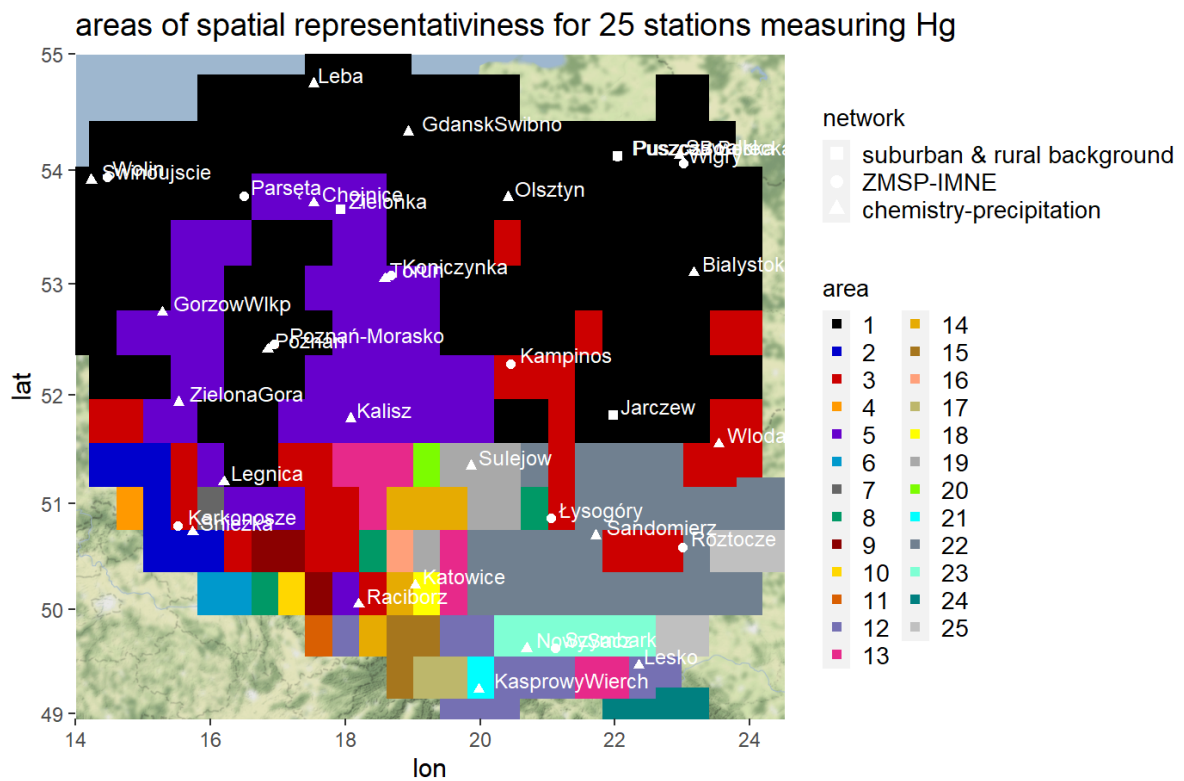
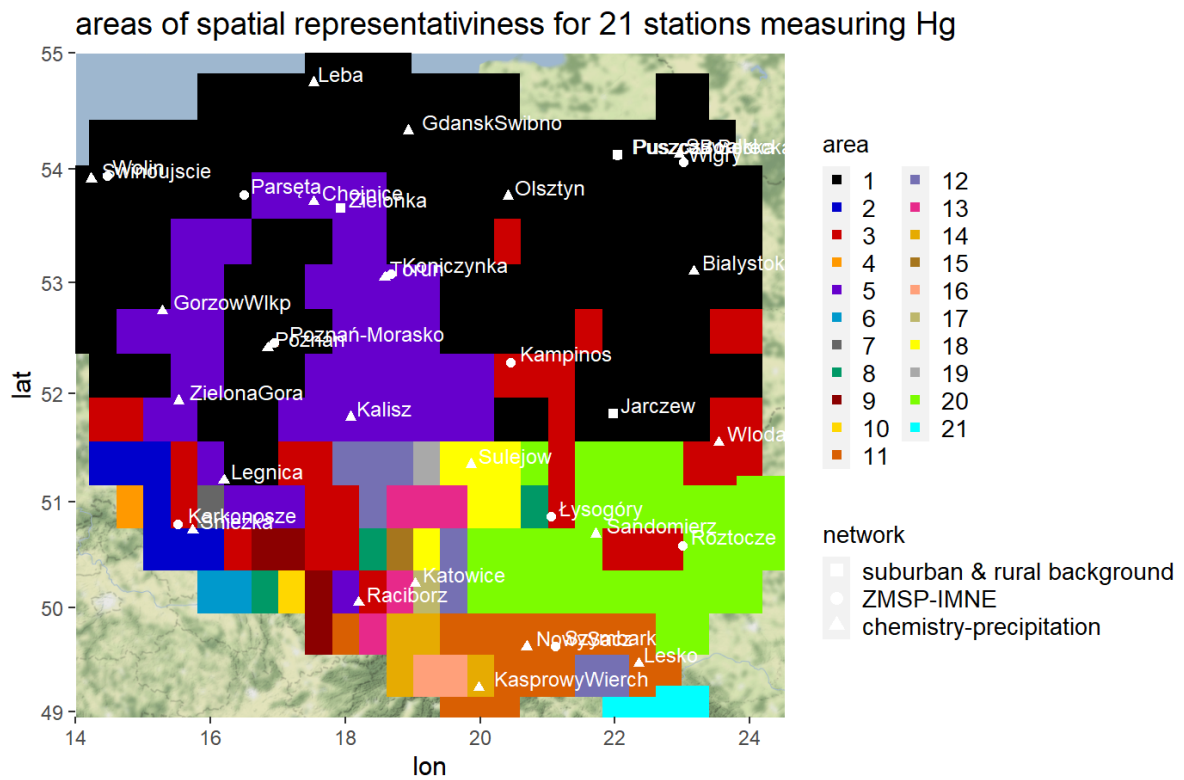
areas of spatial representativeness for 25 stations measuring DEP_RDN



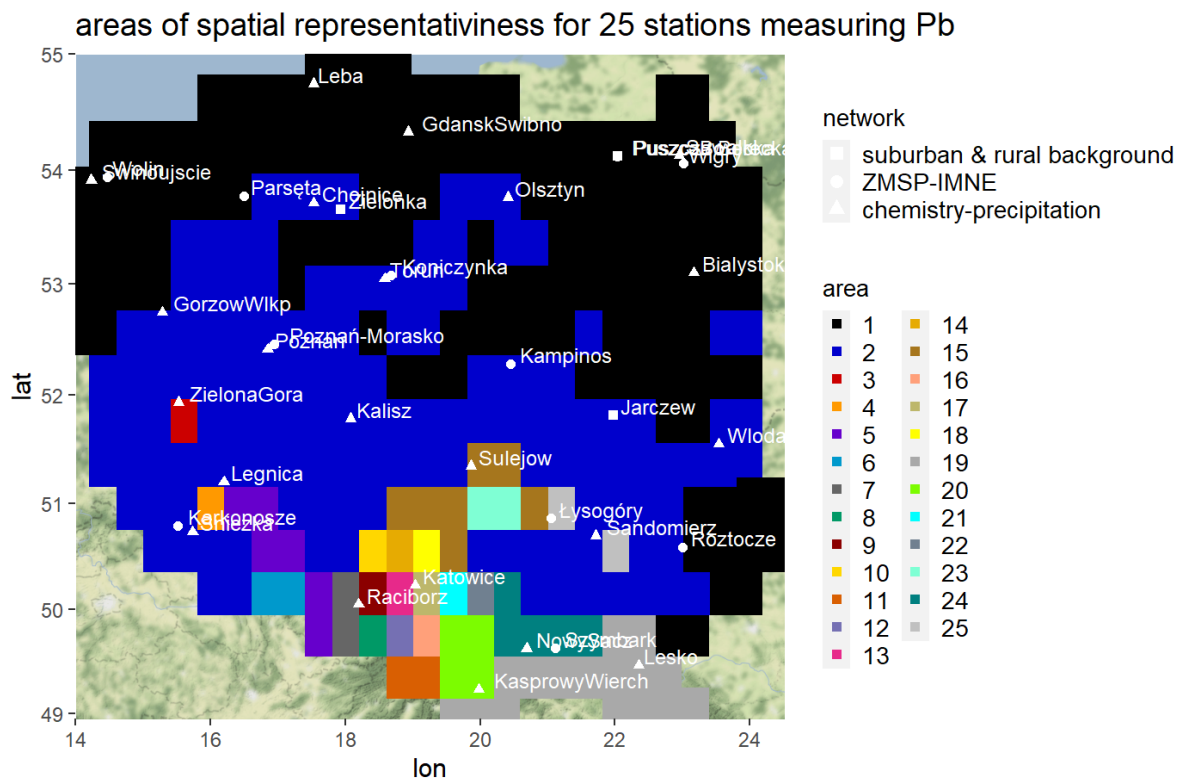
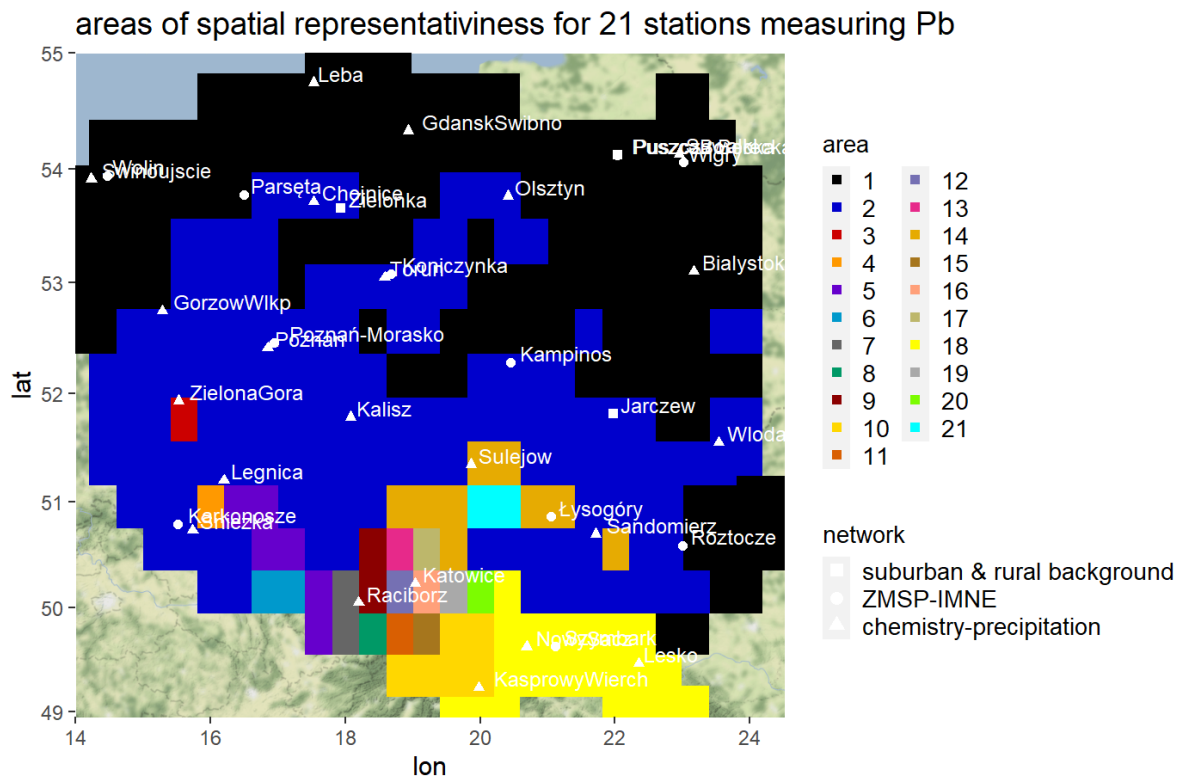
A.3.4 Cd



A.3.5 Hg

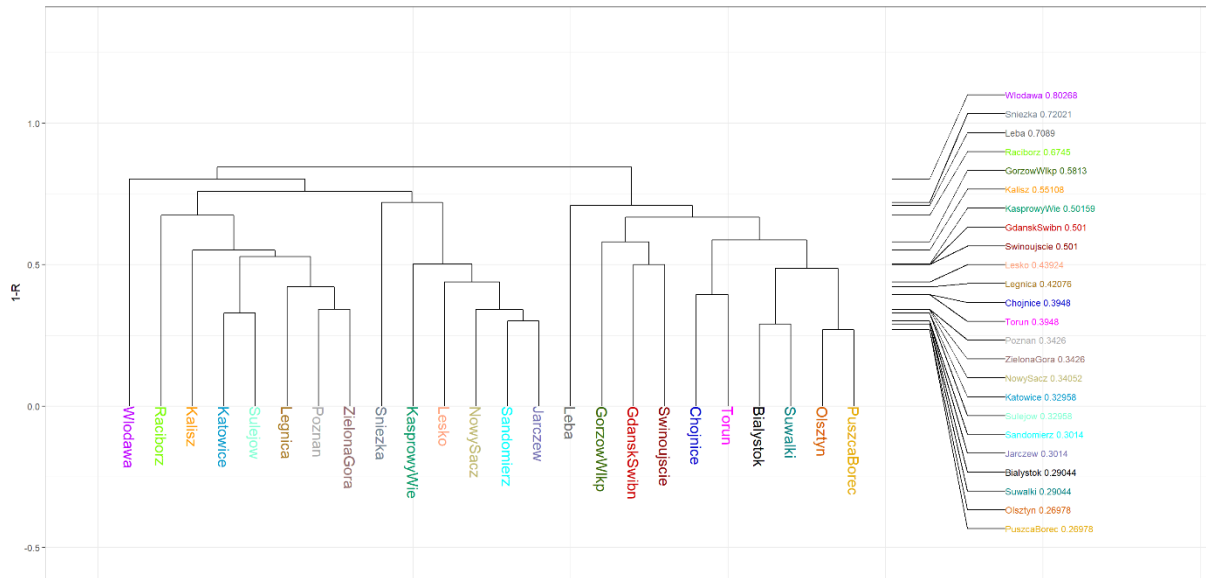


A.3.6 Pb

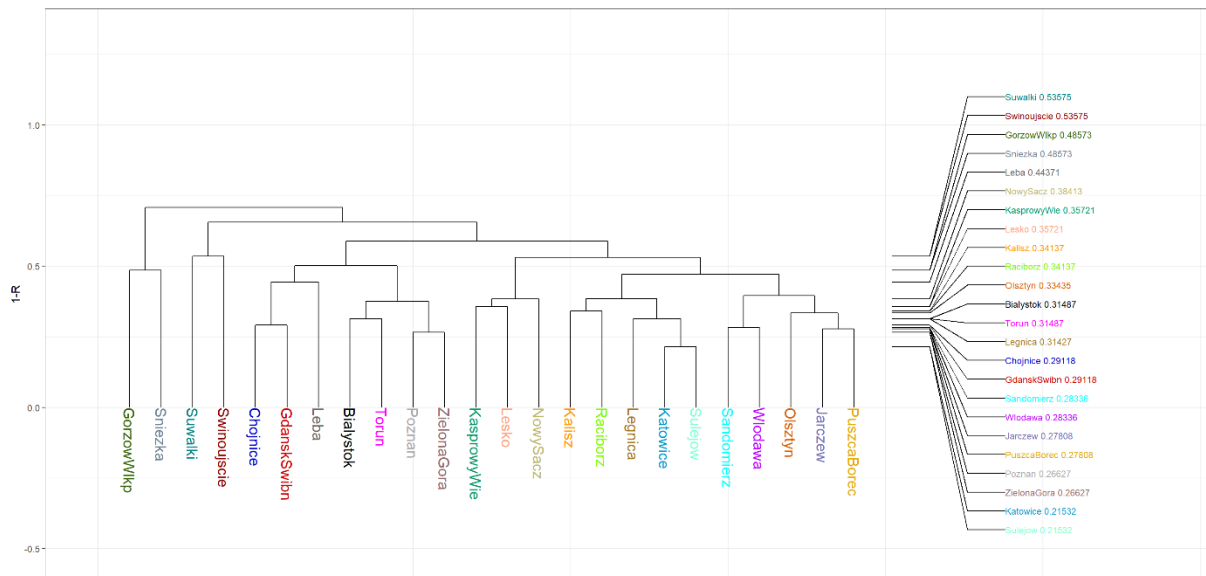


A.4 Station dissimilarity analysis based on the temporal variation (1-R)

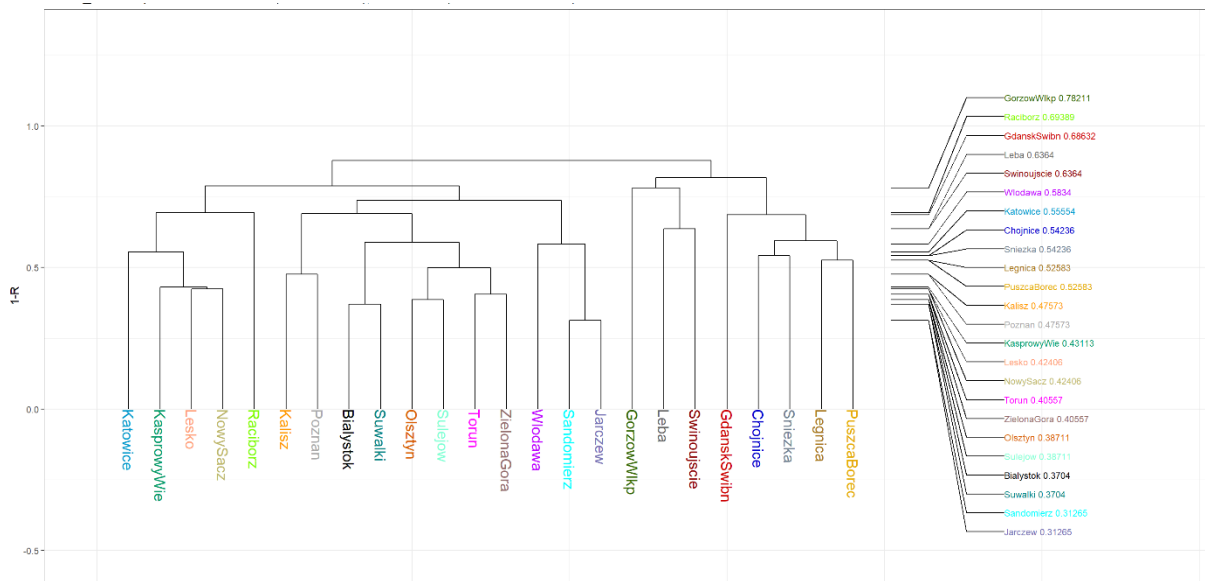
A4.1 SO4



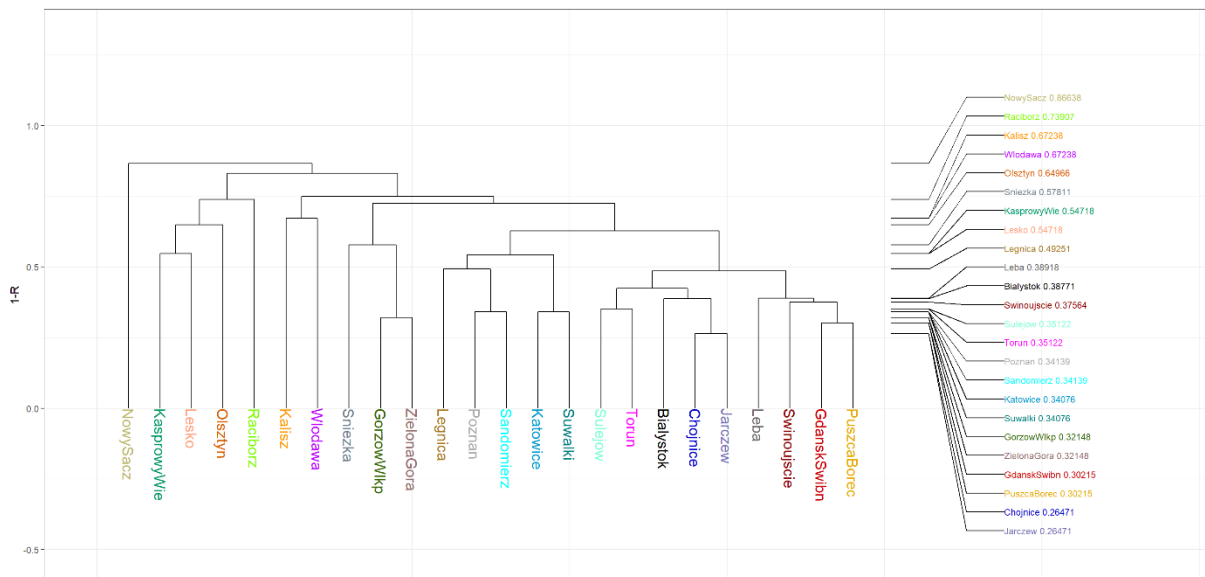
A4.2 NH4



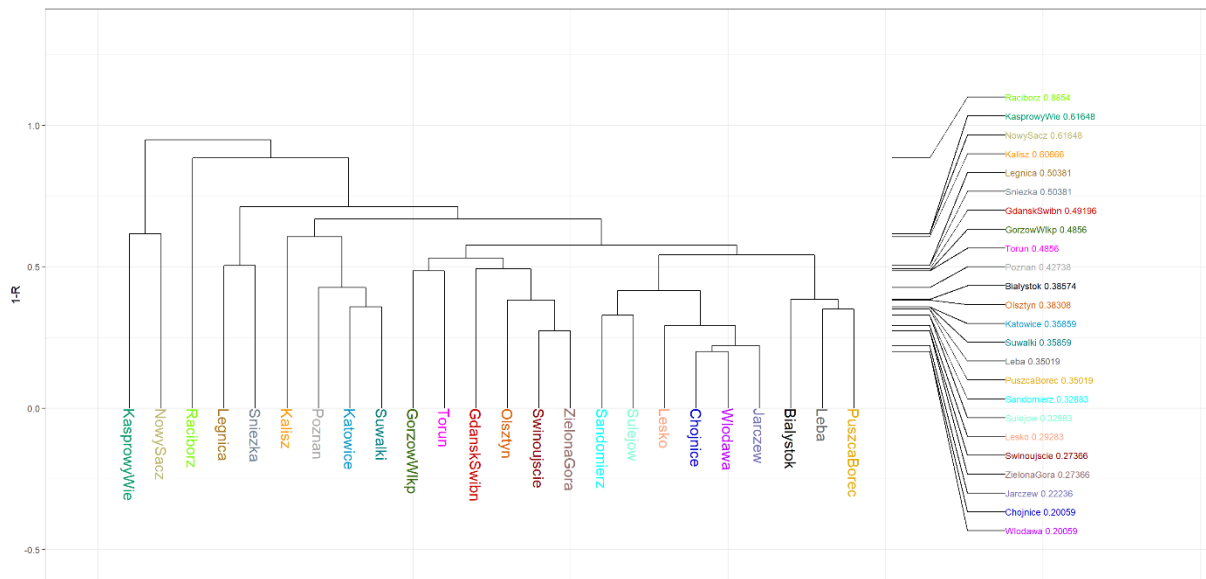
A4.3 Ca



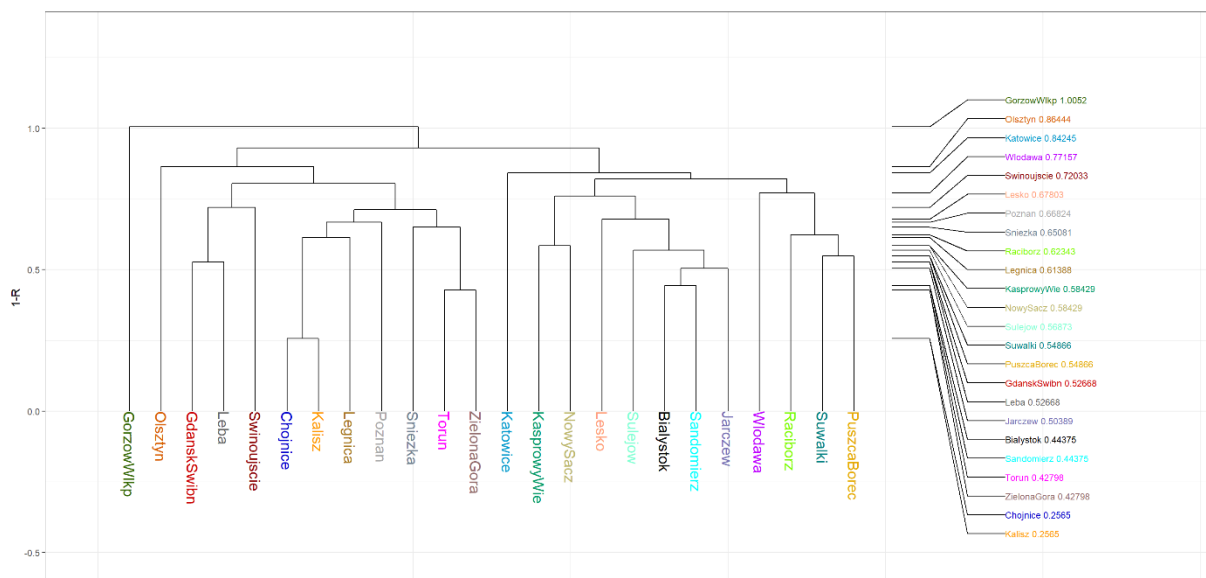
A.4.4 Cl



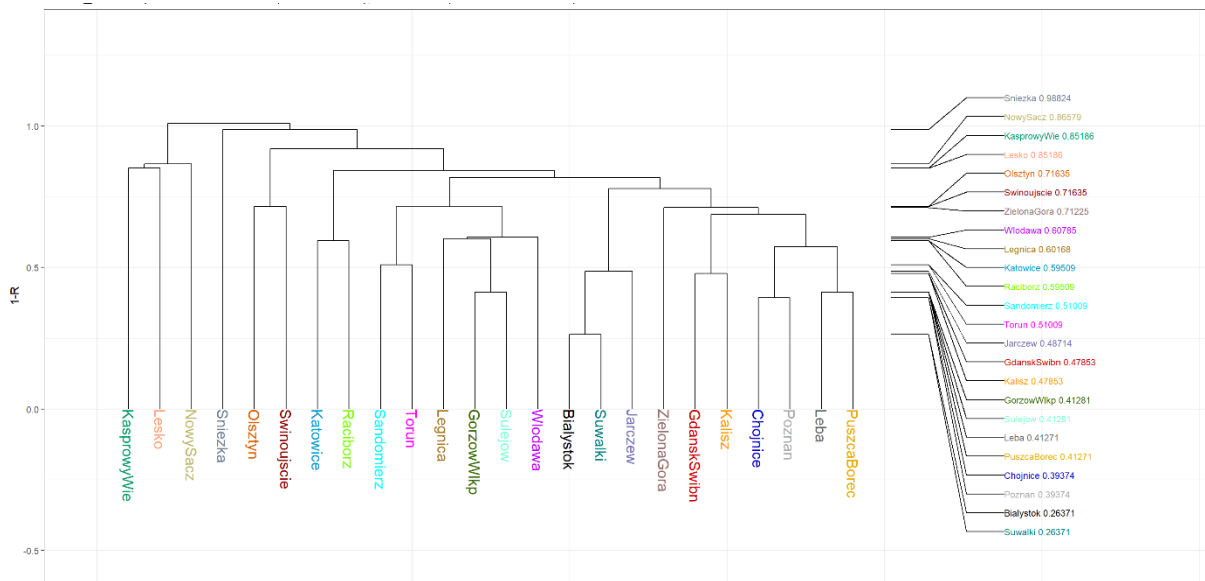
A.4.5 Na



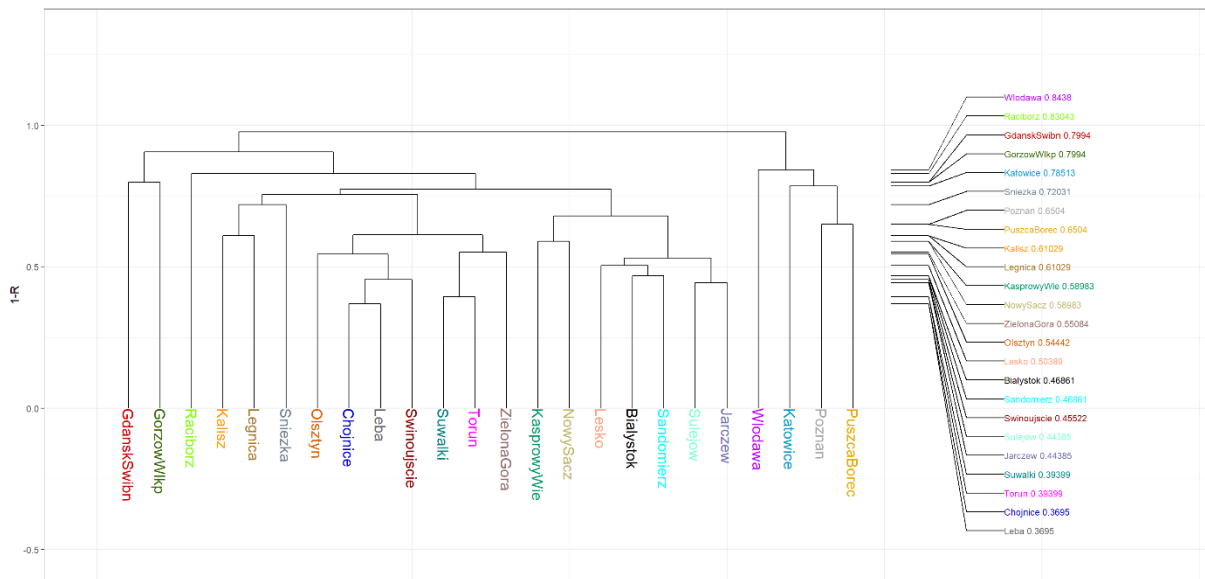
A.4.6 K



A.4.7 H+

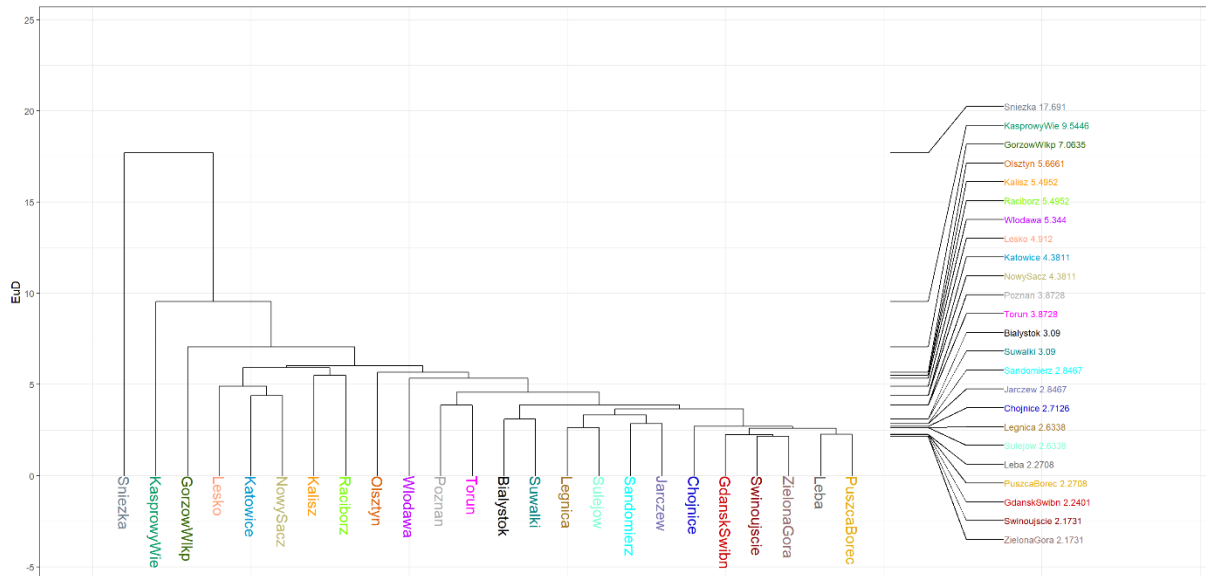


A.4.8 Mg

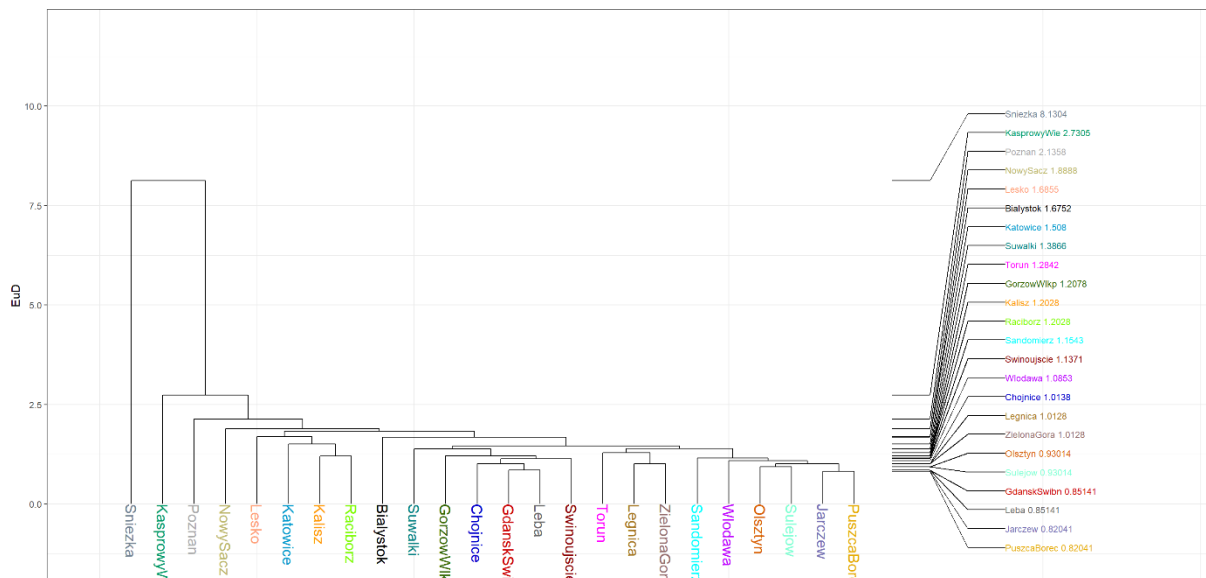


A.5 Station dissimilarity analysis based on the magnitude of deposition levels (EuD)

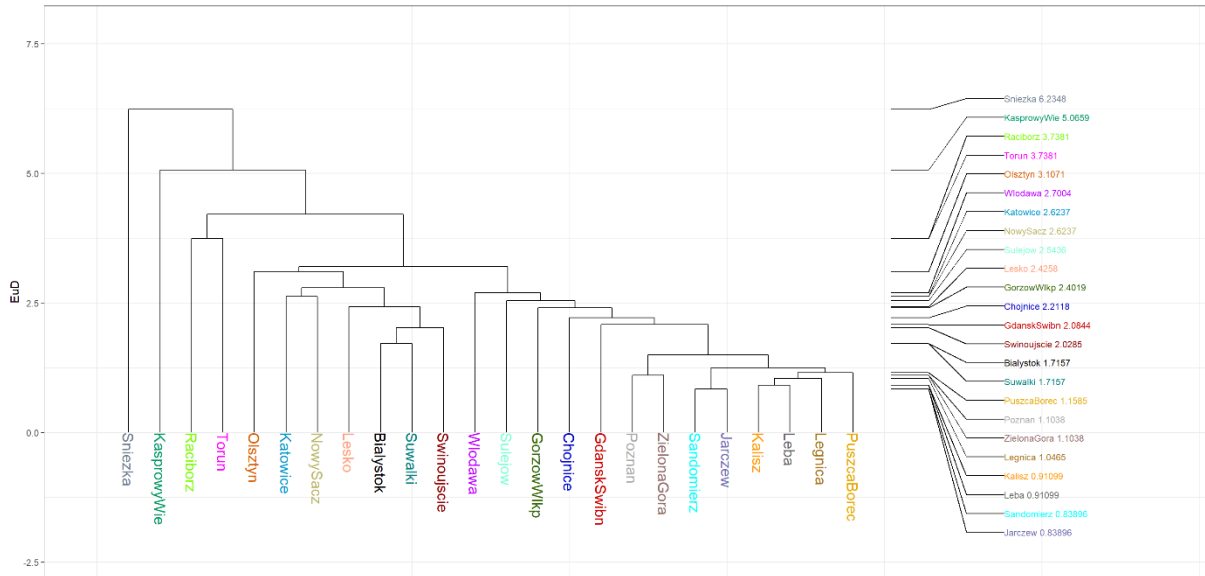
A5.1 SO4



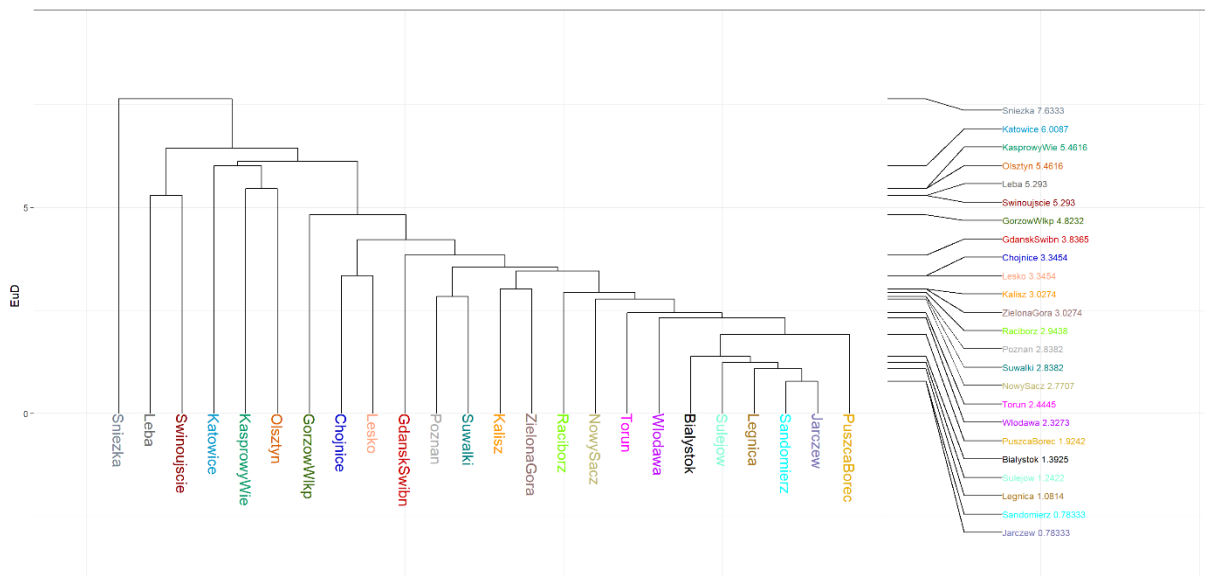
A5.2 NH4



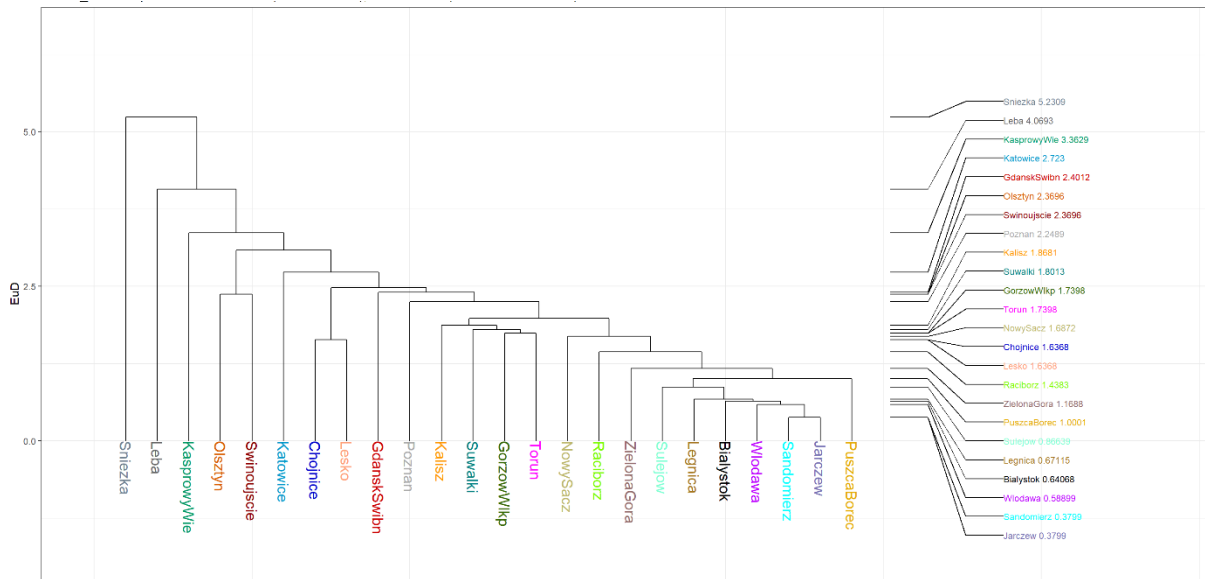
A5.3 Ca



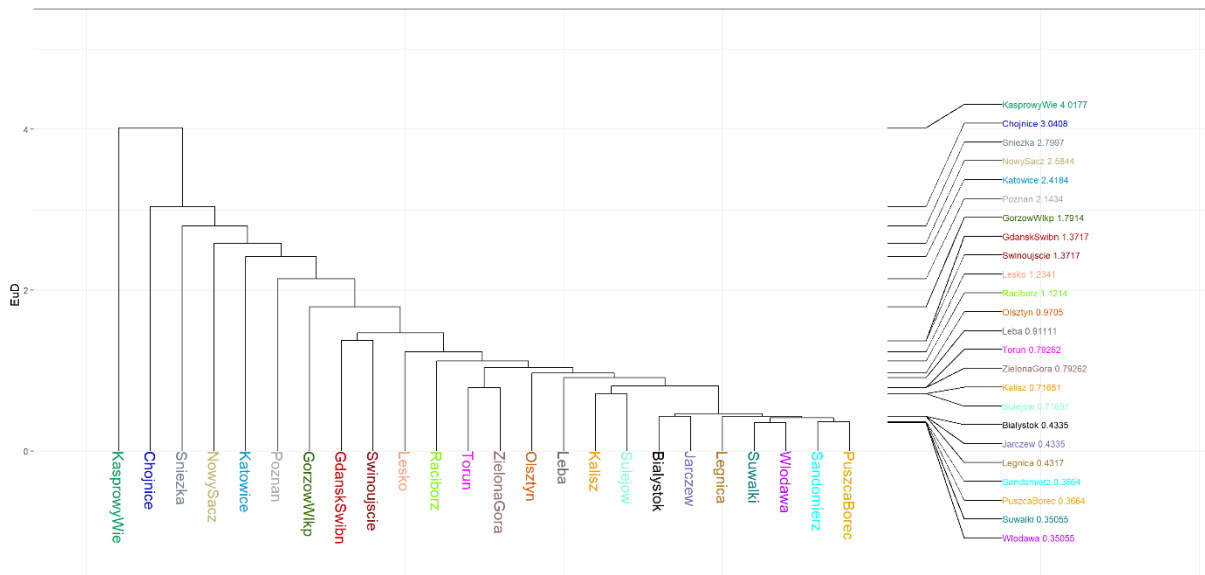
A5.4 CI



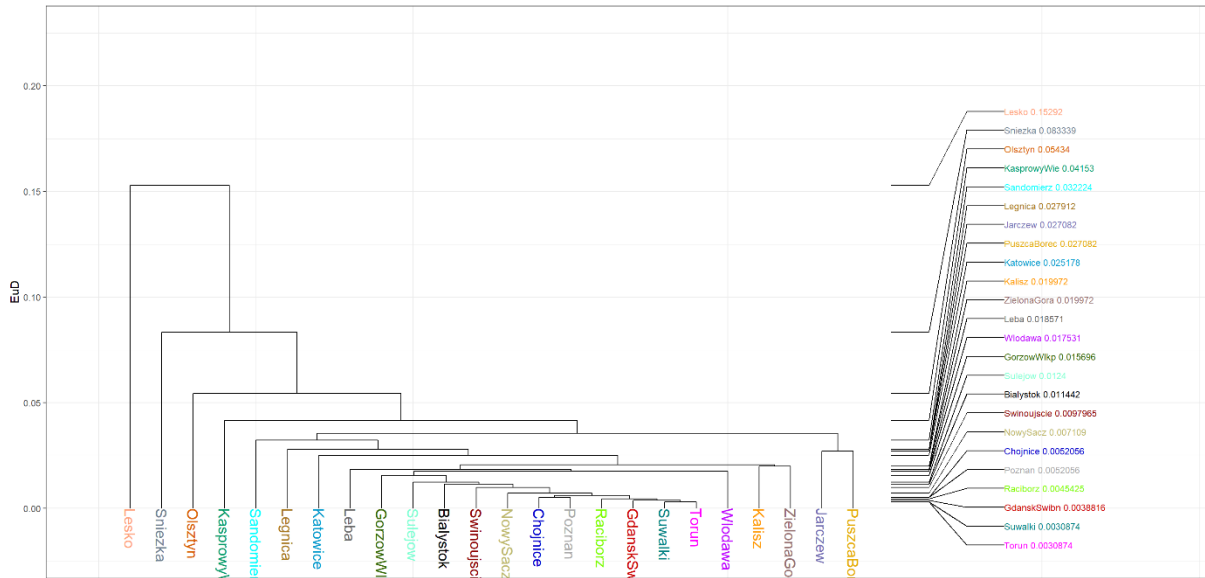
A.5.5 Na



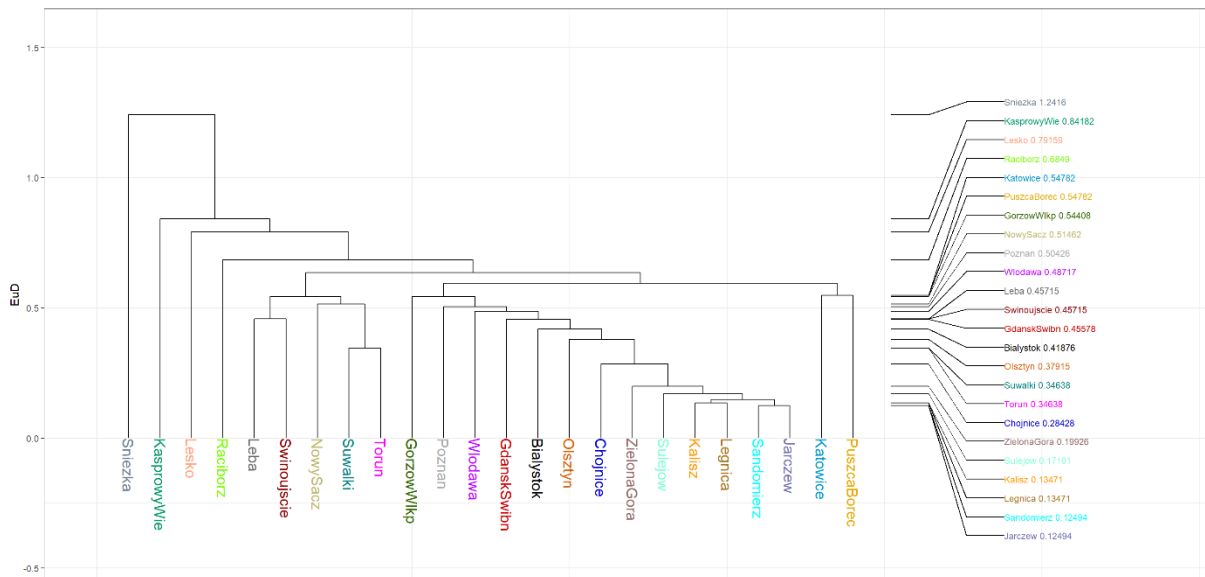
A.5.6 K



A.5.7 H+



A.5.8 Mg



NILU

The climate and environmental research institute NILU is an independent, non-profit research institution established in 1969. Through its research NILU increases the understanding of atmospheric composition, climate change, air quality, environmental contaminants, health effects, sustainable systems, circular economy, and digitalisation. Based on its research, NILU markets integrated services and products within analysing, monitoring and consulting. NILU is concerned with increasing public awareness about climate change and environmental pollution.

NILU's values: Integrity - Competence - Benefit to society

NILU's vision: Create sustainable development through internationally leading climate and environmental research

NILU

P.O. Box 100, NO-2027 KJELLER, Norway

E-mail: nilu@nilu.no

<http://www.nilu.no>

Enterprise no.: 941 705 561

ISBN: 978-82-425-3140-7

ISSN: 2464-3327

