Supplementary material

Environmental exposure assessment in the German National Cohort (NAKO)

Kathrin Wolf*, Marco Dallavalle, Fiona Niedermayer, Gabriele Bolte, Tobia Lakes, Tamara Schikowski, Karin Halina Greiser, Lars Schwettmann, Ronny Westerman, Nikolaos Nikolaou, Jeroen Staab, Robert Wolff, Gunthard Stübs, Stefan Rach, Alexandra Schneider, Annette Peters**, Barbara Hoffmann**

* Corresponding author (Institute of Epidemiology, Helmholtz Zentrum München GmbH, German Research Center for Environmental Health, Neuherberg, Germany; kathrin.wolf@helmholtz-munich.de)

** Shared last authorship

Table of contents

Appendix A: Supplementary Methods

Supplementary Methods. Geocoding of participants residential addresses

Participant's addresses at the time of random sampling and baseline examination were gecoded using the "Georeferenced Address Data" provided by the German Federal Agency for Cartography and Geodesy (BKG). The "Georeferenced Address Data" ((c) GeoBasis-DE/ BKG, Deutsche Post Direkt GmbH, Statistisches Bundesamt, Wiesbaden (2021), https://www.bkg.bund.de) contains Germany-wide information on postal addresses and coordinates based on the "Official House Coordinates of Germany (HK-DE)" maintained by the "Central Office for House Coordinates and Building Polygons" (ZSHH).

Addresses were matched using a proprietary backtracking algorithm. Prior to matching the participant addresses, the algorithm was tested on a subsample against a commercial solution and achieved comparable or even higher accuracy. After normalizing both the participant addresses and the addresses from the BKG dataset, we aimed for a full match of location, postcode (ZIP), street, number, and house number suffix. Since we had access to multiple versions of the BKG dataset (2018, 2021), additional matching was performed against the other version of the dataset if the initial matching quality was below 1-X. Addresses that could not be initially matched (<0.2 %) were manually corrected and matched again. Each dataset was evaluated for its matching quality based on which fields were successfully matched and which fields did not align.

We introduced six matching quality levels (1-5, 9), with the quality ranging from 1 (unambiguous match) to 9 (no match found). Each level contains further subcategories that specify which fields were successfully matched. Of the original 205,414 participant addresses, 204,234 (≈99.4 %) were matched with a quality of 1. Table M1 presents the matching quality results.

Table M1. Matching quality of the geocoding: levels, number (N) and percentage (%) of addresses falling in the respective category and category description.

Level	Matches		Description
	N	%	
1-1	5	0.0024	match with location, zip and street (participant address contains no number)
1-11	8	0.0039	full match with location, street, and number (no house number suffix in participant address)
1-12	1	0.0005	full match with location, street, number and house number suffix
1-2	151,479	73.7433	full match with location, ZIP, street and number (no house number suffix in participant address)
1-3	20,062	9.7666	full match with location, ZIP, street, number and house number suffix
1-4	1	0.0005	full match with location (BKG-dataset-location contains participant-address-location), ZIP and street (no number in participant address)
1-5	29,168	14.1996	full match with location (BKG-dataset-location contains participant-address-location), ZIP, street and number (no house number suffix in participant address)
1-6	3,236	1.5754	full match with location (BKG-dataset-location contains participant-address-location), ZIP, street, number and house number suffix
1-8	218	0.1061	full match with ZIP, street and number (no house number suffix in participant address)
1-9	56	0.0273	full match with ZIP, street, number and house number suffix
2-1	119	0.0579	match with location, ZIP, street, number and house number suffix (house number suffix matched case-insensitive)
2-2	29	0.0141	match with location (BKG-dataset-location contains participant-address-location), ZIP, street, number and house number suffix (house number suffix matched case-insensitive)
2-3	2	0.0010	match with location, street, number and house number suffix (house number suffix matched case-insensitive)
3-1	5	0.0024	match with location, ZIP and street (no number in participant address and there is no BKG-address without a number, first entry in given street is assigned, usually number 1)
3-11	5	0.0024	match with location, street and number (house number suffix in participant address but there is no BKG-address with given number and house number suffix, entry with number and without suffix is assigned, 1a->1)
3-2	171	0.0832	match with location, ZIP, street and number (no house number suffix in participant address but there is no BKG-address with given number and without a suffix, entry with number and suffix is assigned, 1->1a)
3-3	476	0.2317	match with location, ZIP, street and number (house number suffix in participant address but there is no BKG-address with given number and suffix, entry with number and without suffix is assigned, 1a->1)
3-4	35	0.0170	match with location (BKG-dataset-location contains participant-address-location), ZIP, street and number (no house number suffix in participant address but there is no BKG-address with given number and without a suffix, entry with number and suffix is assigned, 1->1a)

3-5	131	0.0638	match with location (BKG-dataset-location contains participant-address-location), ZIP, street and number (house number suffix in participant address but there is no BKG-address with given number and suffix, entry with number and without suffix is assigned, 1a->1)
3-6	3	0.0015	match with ZIP and street (no number in participant address but there is no BKG-address without number, first entry in given street is assigned, usually number 1)
3-8	3	0.0015	match with ZIP, street and number (house number suffix in participant address but there is no BKG-address with given number and suffix, entry with number and without suffix is assigned, 1a->1)
4-1	76	0.0370	match with location, ZIP and street (number in participant address but there is no BKG-address with given number in this street, the number with the lowest numerical distance to participant number is used)
4-2	31	0.0151	match with location (BKG-dataset-location contains participant-address-location), ZIP and street (number in participant address but there is no BKG-address with given number in this street, the number with the lowest numerical distance to participant number is used)
4-3	4	0.0019	match with ZIP and street (number in participant address but there is no BKG-address with given number in this street, the number with the lowest numerical distance to participant number is used)
4-4	5	0.0024	full match with ZIP, street, number and house number suffix (no house number suffix given) (there is more than one entry in BKG that matches with these fields, manual selection performed)
4-5	1	0.0005	full match with ZIP, street, number and house number suffix (there is more than one entry in BKG that matches with these fields, manual selection is performed)
4-6	1	0.0005	match with location and street (number in participant address but there is no BKG-address with given number in this street, the number with the lowest numerical distance to participant number is used)
4-7	63	0.0307	full match with location, street and number (no house number suffix given) (there is more than exactly one entry in BKG that matches with these fields, manual selection performed)
4-8	5	0.0024	full match with location, street, number and house number suffix (there is more than exactly one entry in BKG that matches with these fields, manual selection performed)
5-1	2	0.0010	match with location and ZIP (no entry in BKG with the given street, first entry in BKG for given location and ZIP assigned)
5-2	6	0.0029	match with ZIP (no entry in BKG with the given street, first entry in BKG for given ZIP assigned)
5-3	3	0.0015	match with location (no entry in BKG with the given street, first entry in BKG for given location assigned)
9	4	0.0019	location and ZIP do not exist in BKG-Data (manual selection not possible with the given address details)

Appendix B: Supplementary Tables and Figures

Table S1. Number of participants per study centre indicator (where examination took place; rows) and study centre region indicator (based on address locations; columns).

Study region Study centre	Augs-burg	Berlin[1]	Bremen	Düssel-dorf	Essen	Frei-burg	Halle	Ham-burg	Han-nover	Kiel	Leipzig	Mann-heim	Mün-ster	Neubran-denburg	Regens-burg	Saar-brücken	None[2]
Augsburg	20491	2	0	0	0	0	0	0	1	0	0	0	1	0	0	0	103
Berlin[1]	0	30999	0	0	1	0	0	2	1	0	2	1	0	1	0	0	58
Bremen	0	1	10417	0	0	0	0	5	0	0	1	0	1	0	0	0	60
Düsseldorf	0	1	0	9010	0	0	0	0	0	0	0	0	1	0	0	0	80
Essen	0	0	0	4	10570	0	0	0	0	1	0	0	1	0	0	0	57
Freiburg	0	0	0	0	0	10018	0	0	0	0	0	0	0	0	0	0	40
Halle	0	1	0	0	0	0	10104	0	0	0	4	0	0	0	0	0	13
Hamburg	0	0	0	0	2	0	0	10026	0	0	0	0	0	0	0	0	31
Hannover	0	0	1	0	0	0	0	0	9968	0	0	0	0	0	0	0	47
Kiel	0	0	0	0	0	0	0	1	0	9478	1	0	0	0	0	0	8
Leipzig	0	2	0	0	0	0	1	0	0	0	10810	0	0	0	0	0	37
Mannheim	1	2	0	0	0	0	0	0	0	0	0	10189	0	0	0	0	78
Münster	0	0	0	1	1	0	0	0	0	1	0	0	9916	0	0	0	82
Neubran-denburg	0	3	0	0	0	0	0	2	0	0	0	0	0	21943	0	0	33
Regensburg	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9980	0	33
Saar-brücken	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9951	59
Sum	20493	31011	10418	9015	10574	10018	10105	10036	9970	9480	10818	10190	9920	21944	9980	9951	819
Sum other study centre[3]	2	12	1	5	4	0	1	10	2	2	8	1	4	1	0	0	

[1] All three Berlin centres combined.

[2] Participants who no longer lived in one of the study centre regions during baseline examination.

[3] Participants who lived in one of the study centre regions but travelled to another study centre for the baseline examination.

Figure S1: Spatial patterns of population density per 5 km x 5 km grid in Germany and Berlin (detailed map right, 100 m x 100 m grid), 2018. Areas circled in red indicate NAKO study regions.

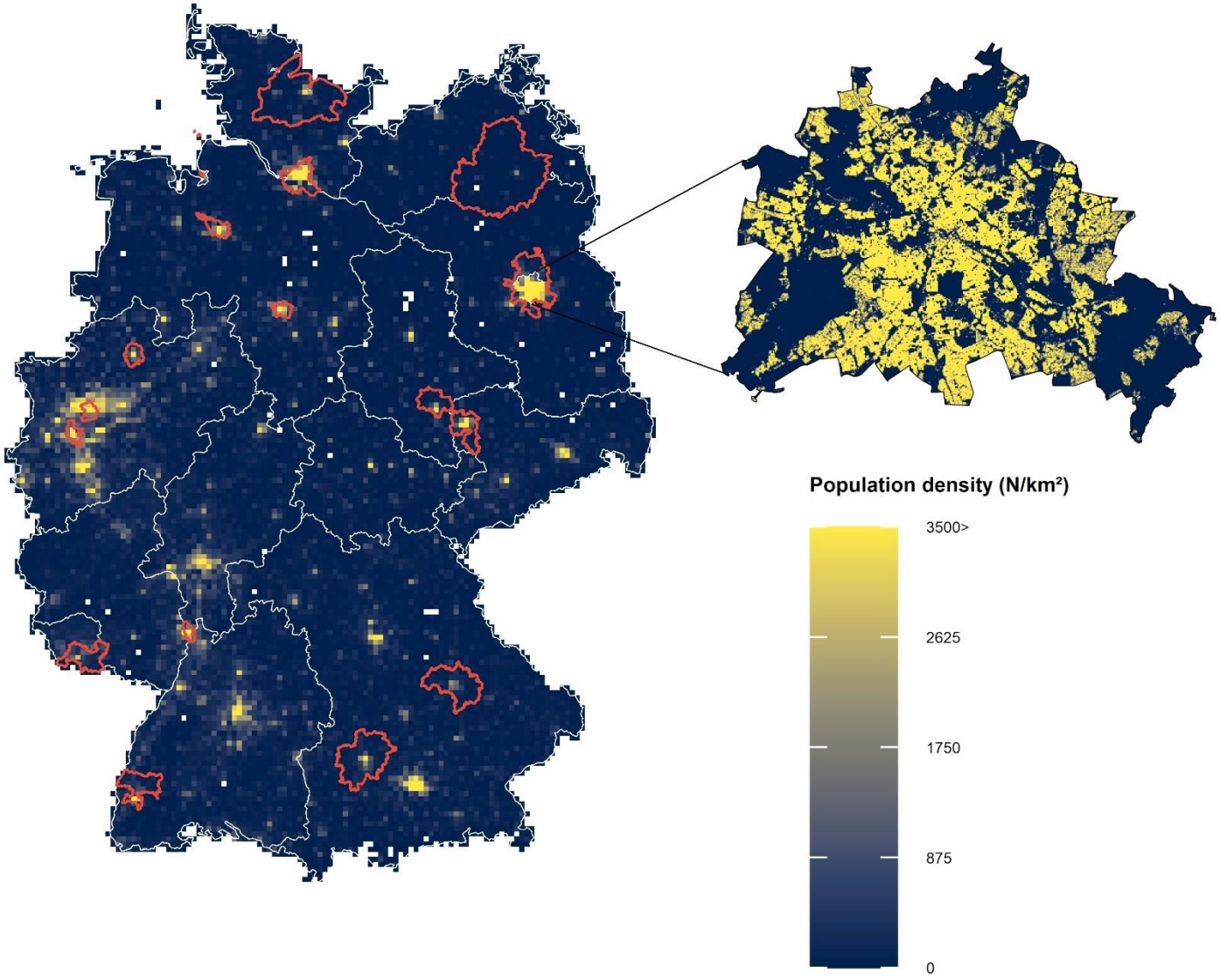

Figure S2: Spatial patterns of annual mean $PM_{2.5}$, $PM_{2.5}$ absorbance, NO_2 and warm season O_3 in Germany, 2010. Air pollution concentrations were predicted via hybrid landuse regression models within the ELAPSE project.

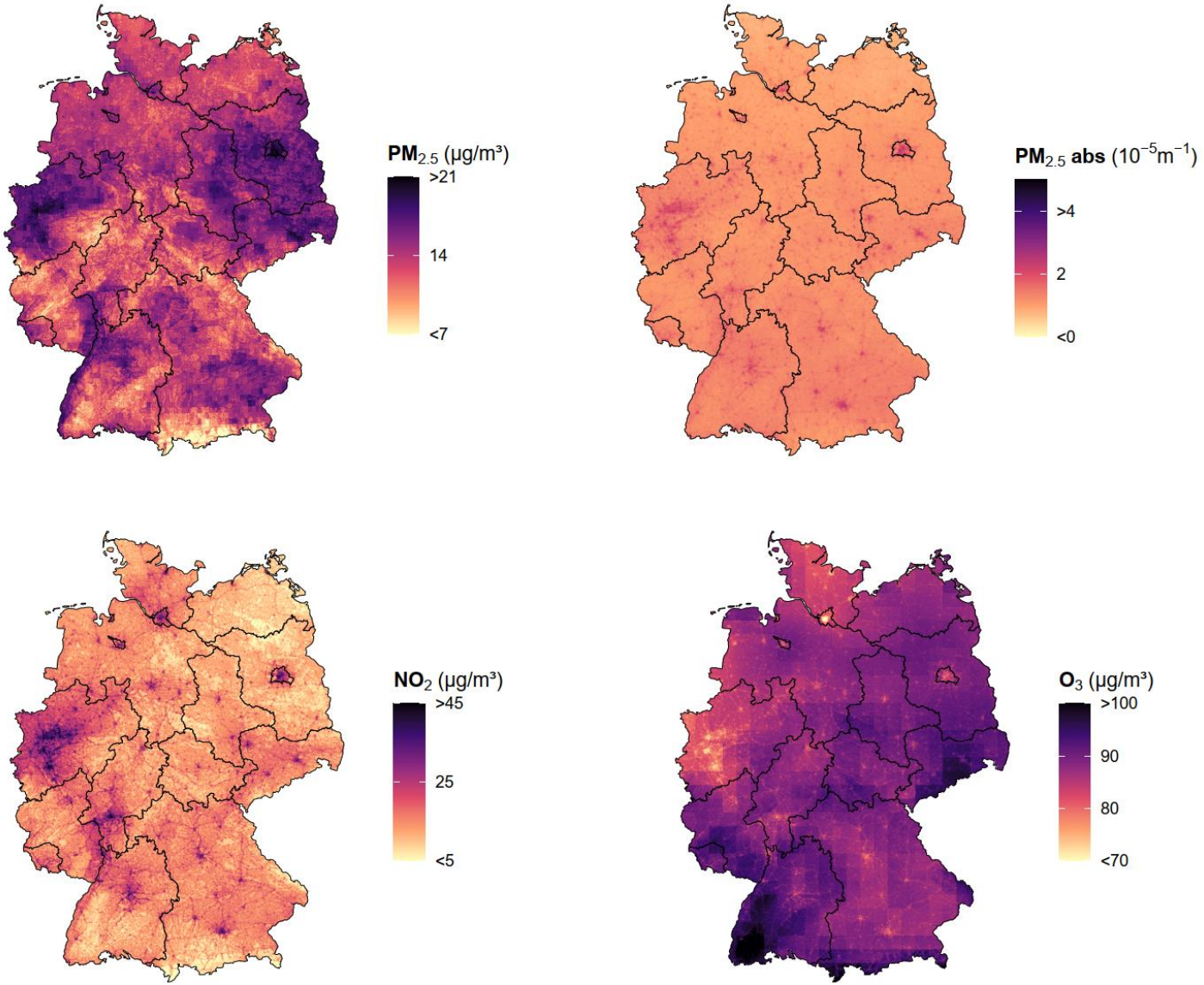

Figure S3: Spatial patterns of annual mean PM_{10}, $PM_{2.5}$, NO_2 and warm season O_3 in Germany, 2014-2019. Air pollution concentrations predicted via optimal interpolation combining air quality measurements with the chemical transport model REM-CALGRID by Federal Environment Agency (UBA).

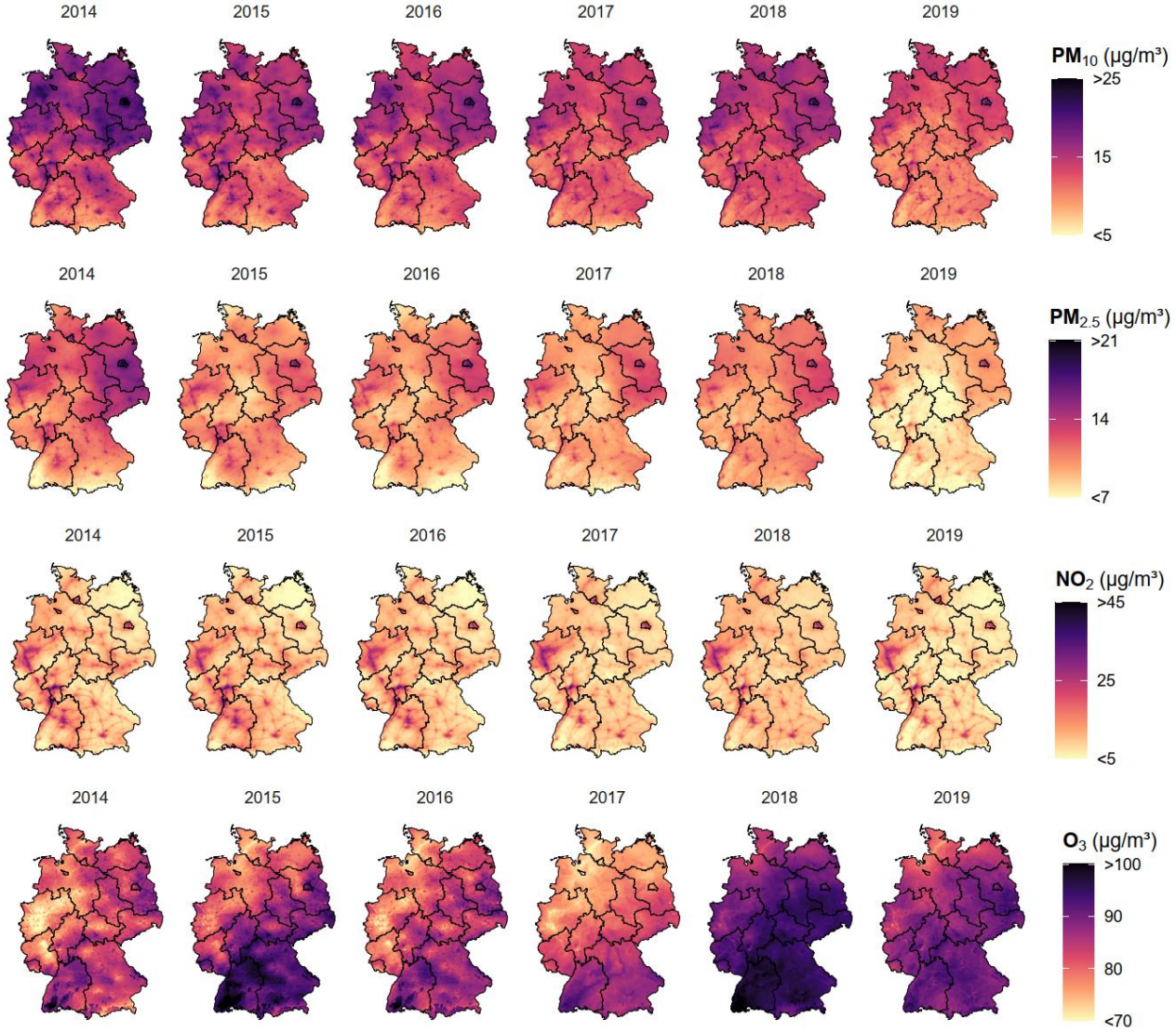

Figure S4: Percentage of NAKO participants above the World Health Organization (WHO) Air Quality Guidelines (AQG), European Union (EU) Ambient Air Quality Standards (AAQS) proposal for 2030 and current EU AAQS for $PM_{2.5}$ (top) and NO_2 (bottom).

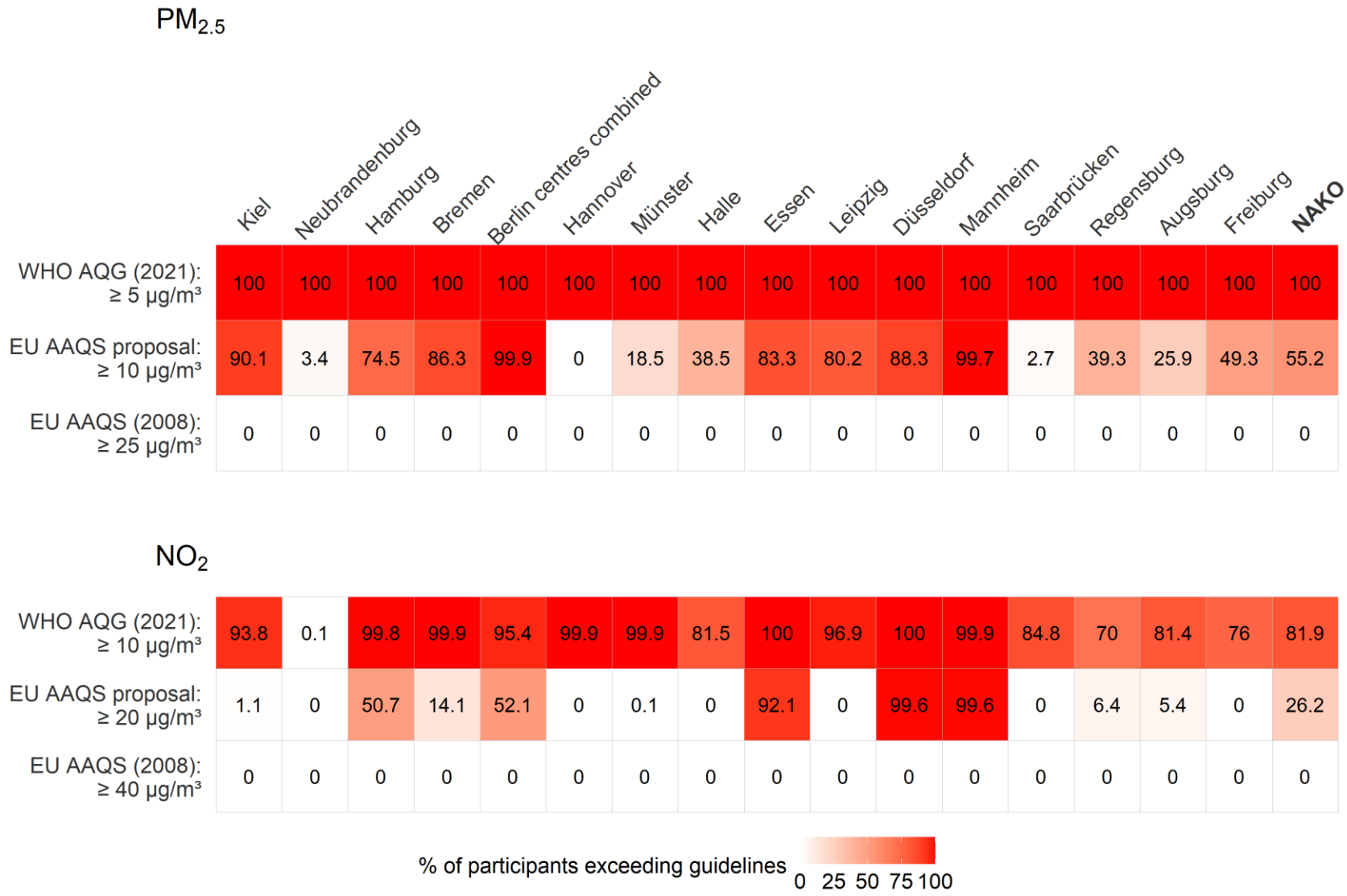

Figure S5: Annual 2010 mean levels of $PM_{2.5}$, $PM_{2.5}$ absorbance, NO_2, warm season O_3 concentrations at the residencies of NAKO participants stratified by degree of urbanisation. Concentrations were derived from the ELAPSE model.

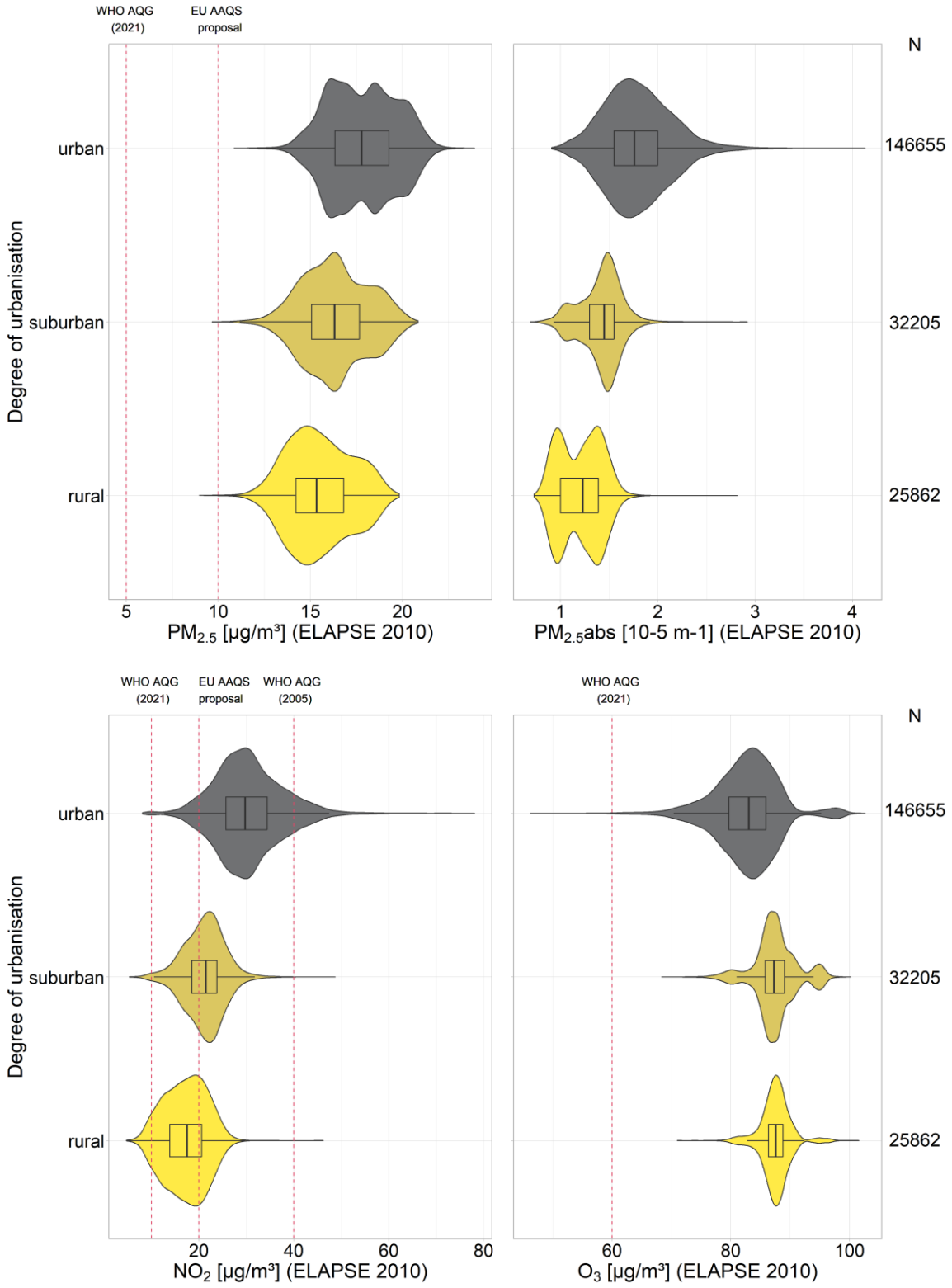

Figure S6: Annual 2014 and 2019 mean PM_{10}, $PM_{2.5}$, NO_2 and warm season O_3 at residencies of NAKO participants stratified by degree of urbanisation. Concentrations were derived from the Federal Environment Agency (UBA) model.

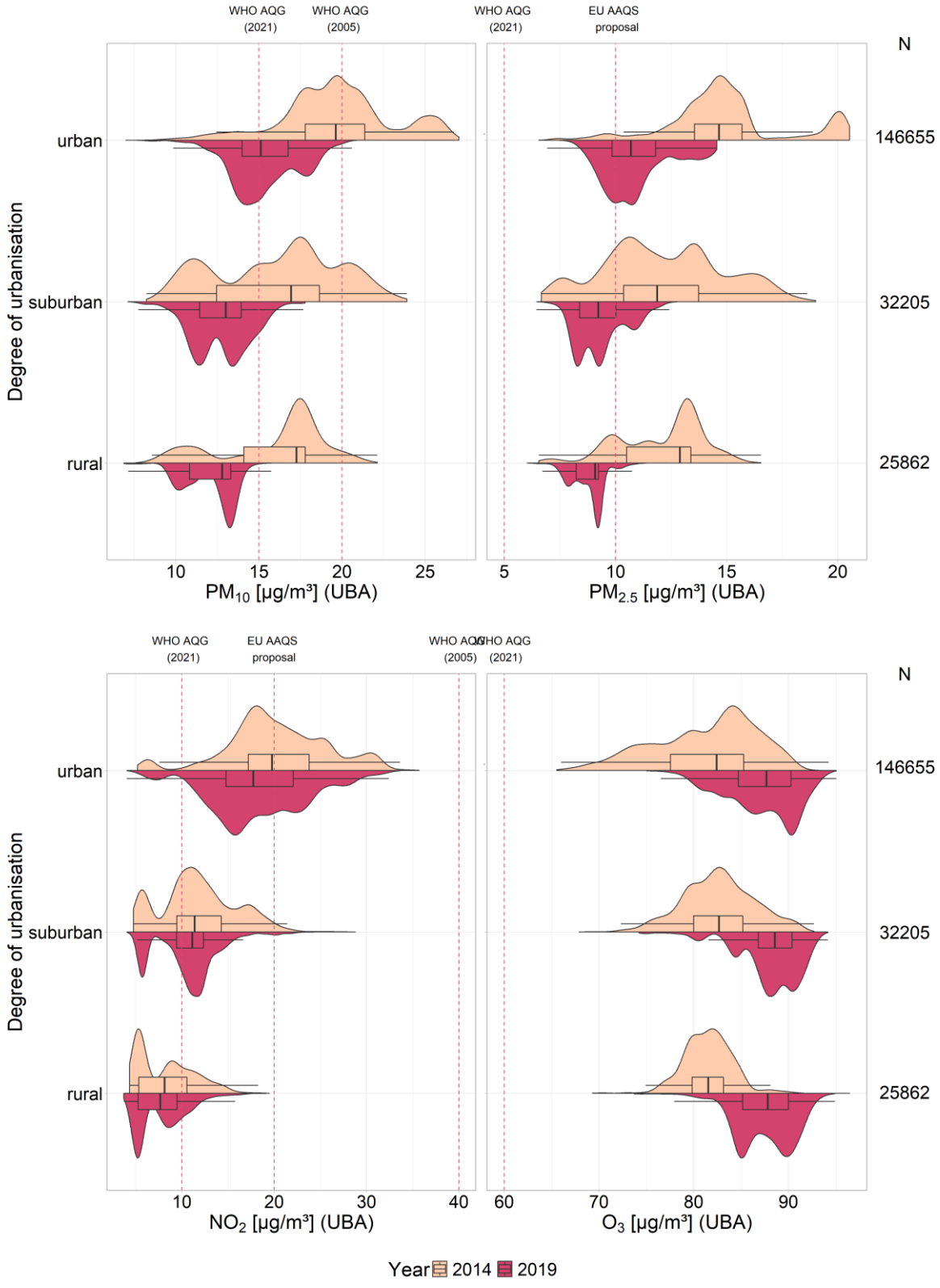

Figure S7a: Annual 2010 mean levels of $PM_{2.5}$ and $PM_{2.5}$ absorbance at residencies of NAKO participants stratified by study centre (y-axis ordered from North to South). Concentrations were derived from the ELAPSE model.

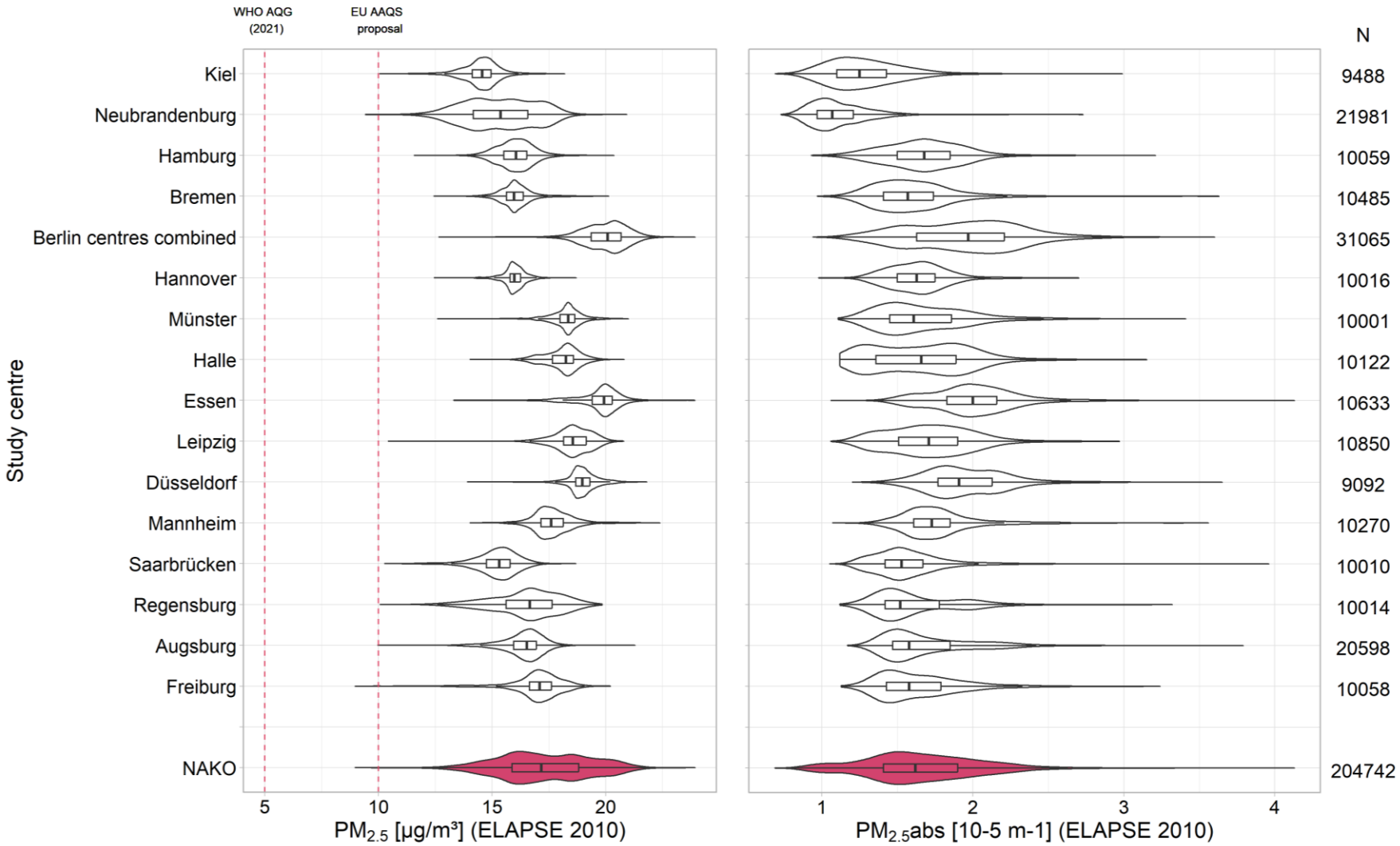

Figure S7b: Annual 2010 mean levels of NO_2 and warm season O_3 at residencies of NAKO participants stratified by study centre (y-axis ordered from North to South). Concentrations were derived from the ELAPSE model.

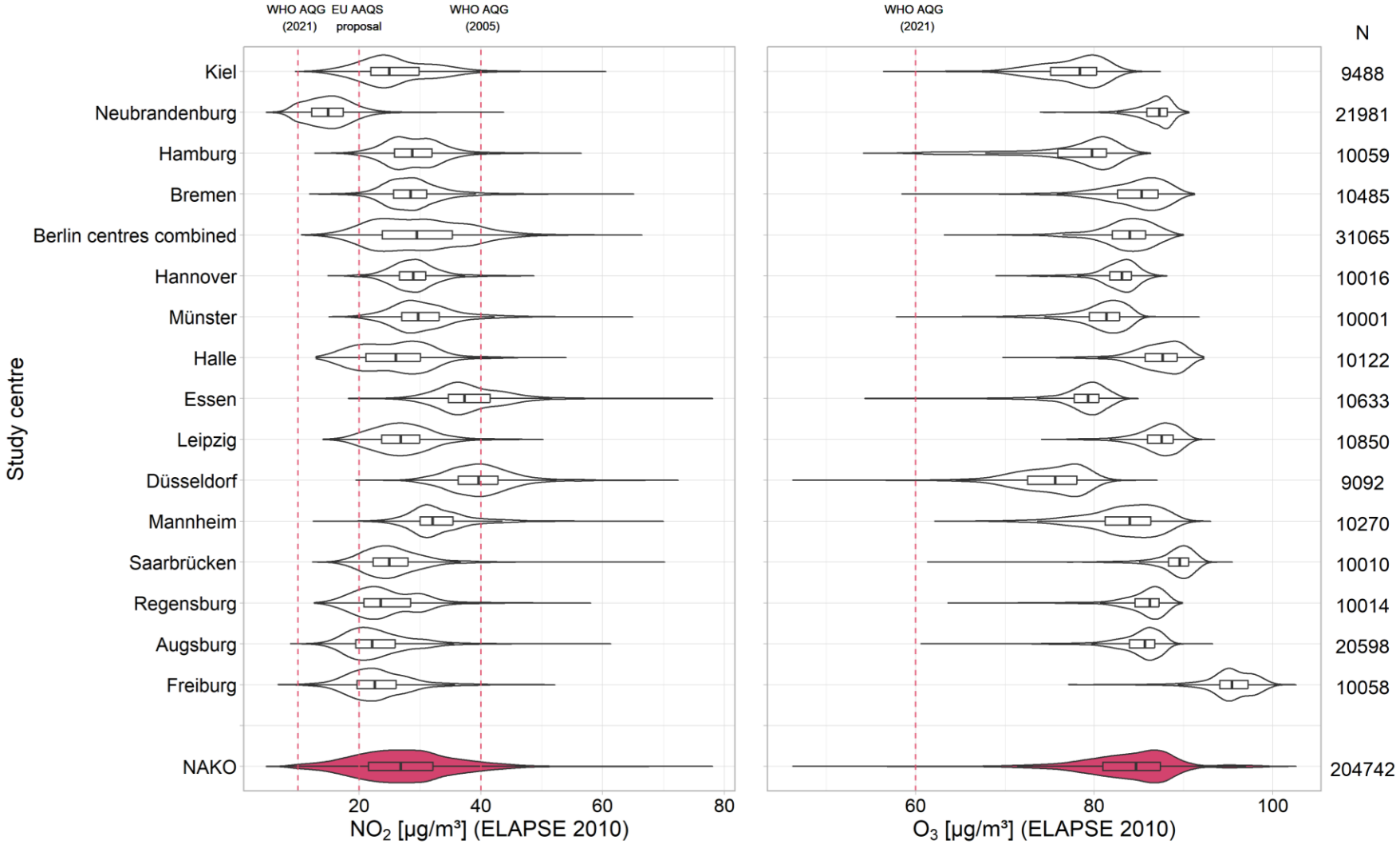

Figure S8a: Annual 2014 and 2019 mean PM_{10} and $PM_{2.5}$ at residencies of NAKO participants stratified by study centre (y-axis ordered from North to South). Concentrations were derived from the Federal Environment Agency (UBA) model.

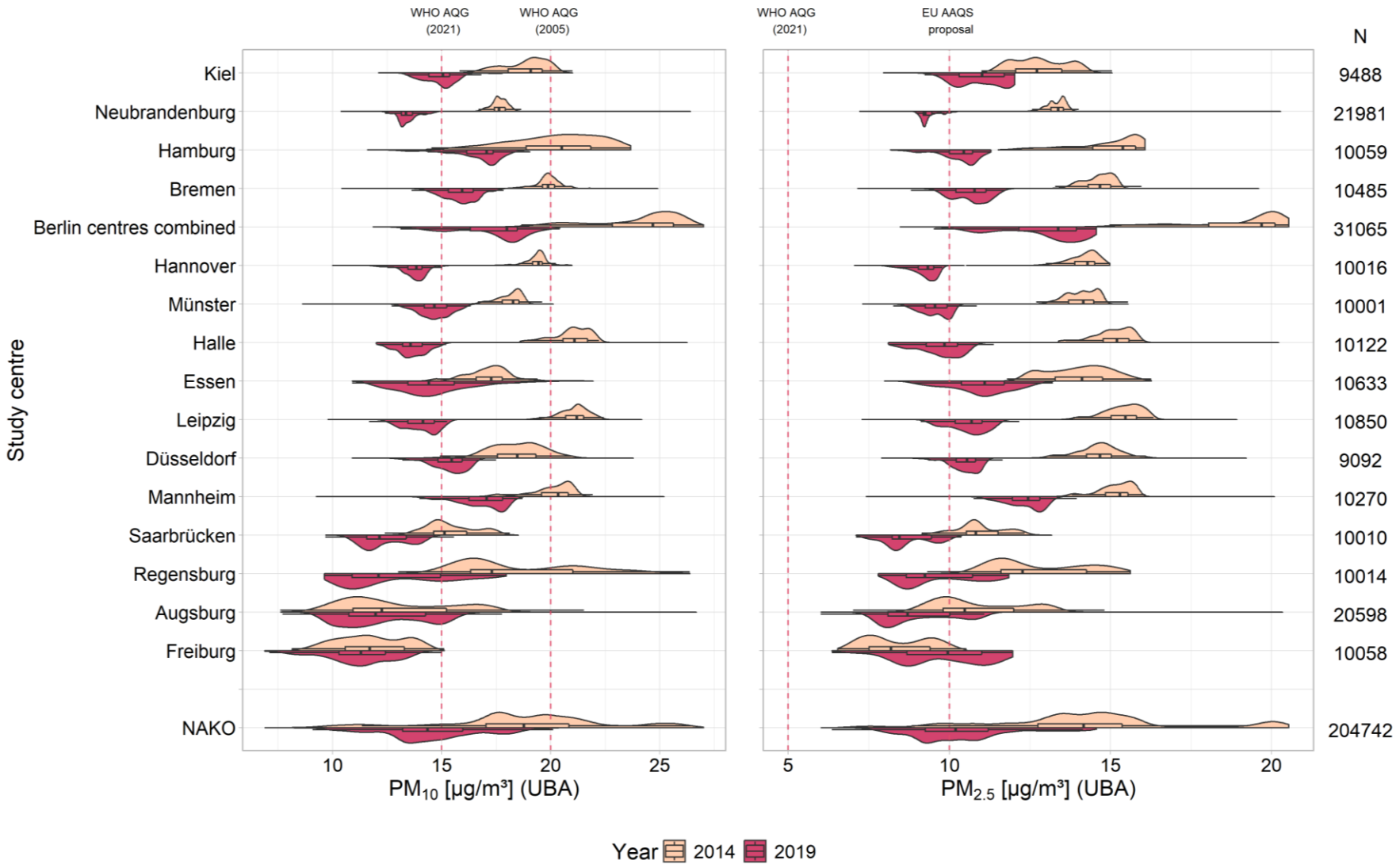

Figure S8b: Annual 2014 and 2019 mean NO_2 and warm season O_3 at residencies of NAKO participants stratified by study centre (y-axis ordered from North to South). Concentrations were derived from the Federal Environment Agency (UBA) model.

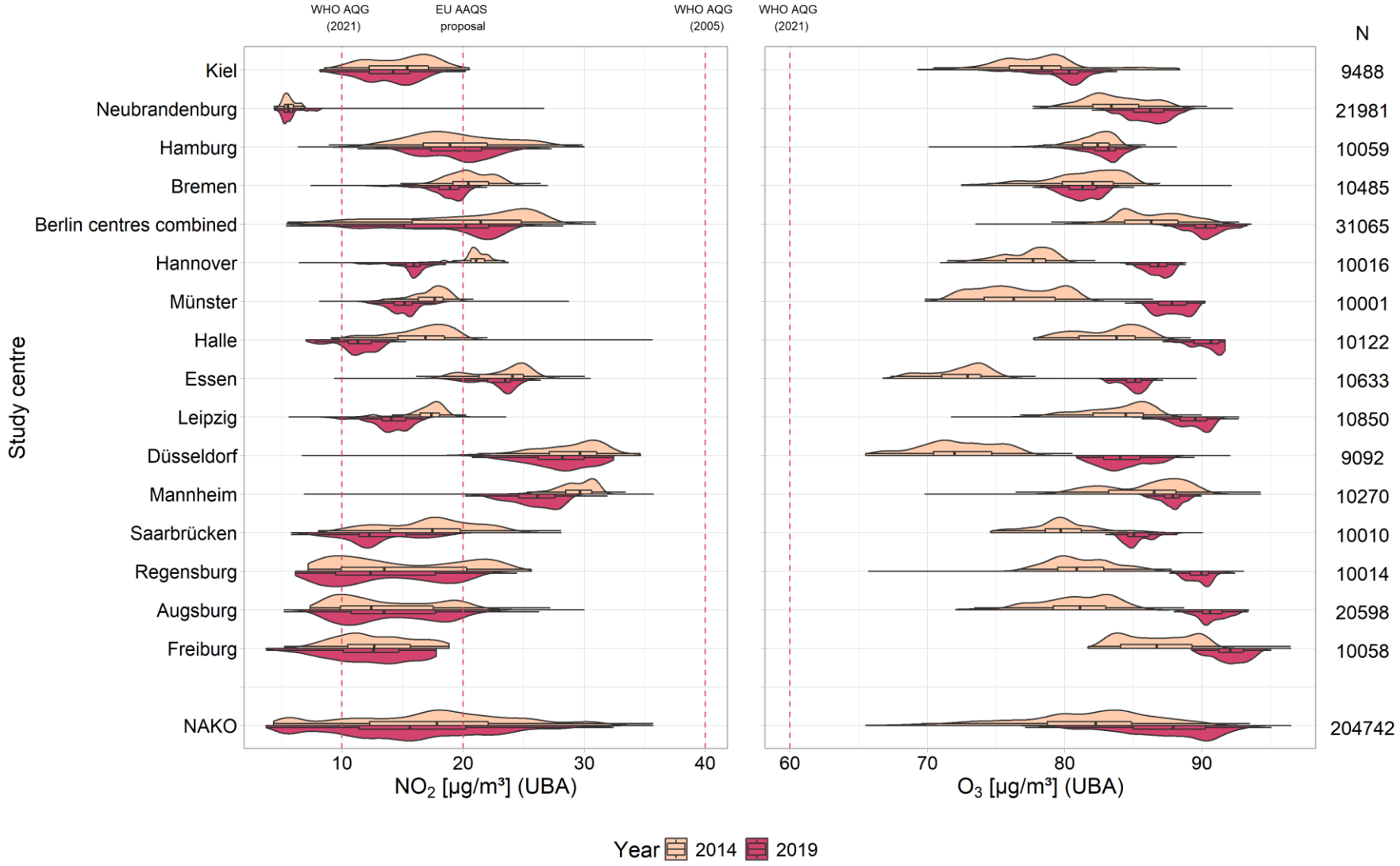

Figure S9: Spatial patterns of annual ambient traffic noise day-evening-night levels (Lden) in Germany and Berlin (detailed map right), 2017. Data was downloaded from the central data repository EIONET of the European Environment Agency, corrected for topological errors and harmonized to a German wide raster.

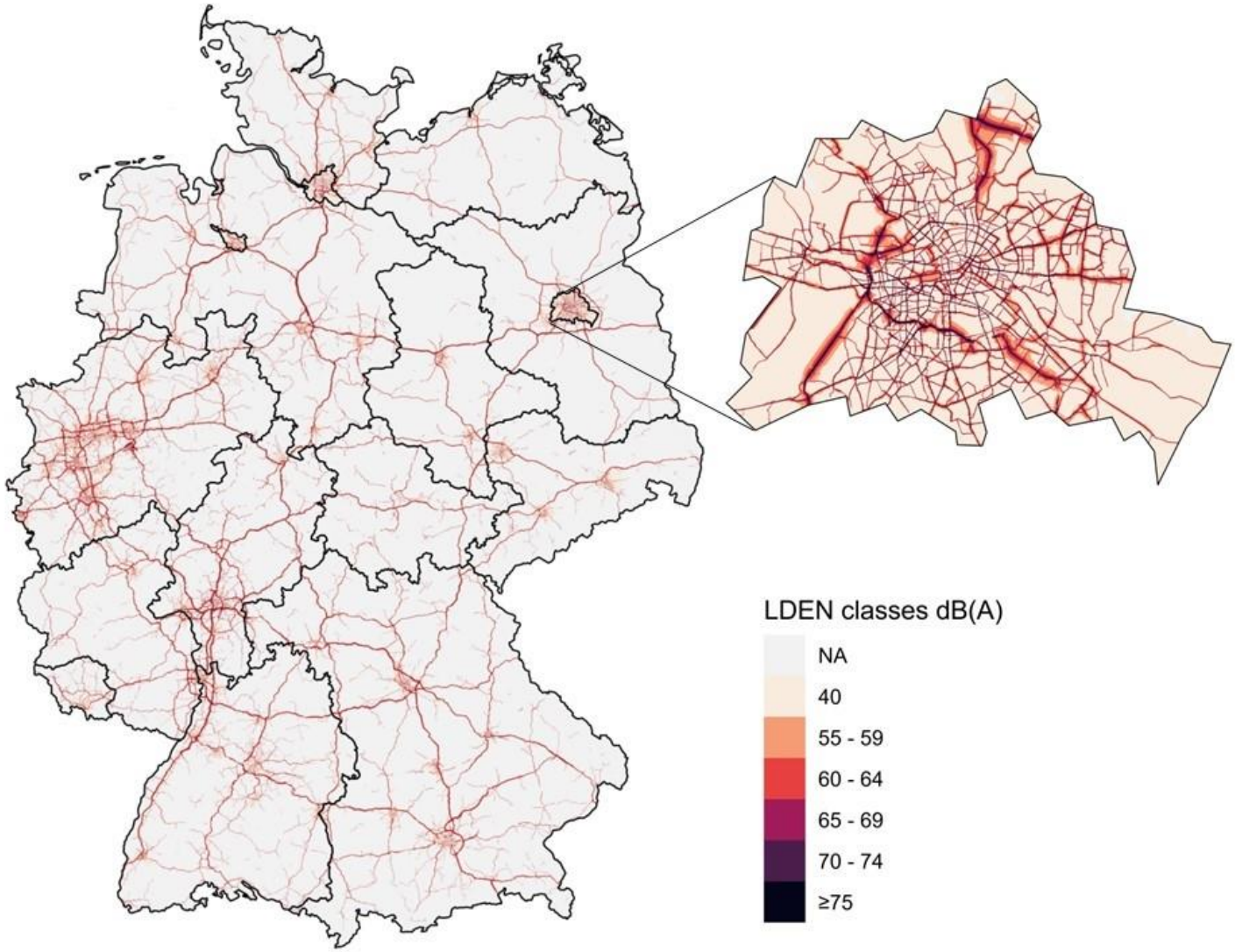

Figure S10: Annual 2017 ambient traffic noise day-evening-night levels (Lden) at the residencies of NAKO participants stratified by degree of urbanisation. Proportion of missings indicates participants living in regions not covered by the Environmental Noise Directive (END) mapping obligation, whereas 40dB(A) was assigned to participants living in END mapping regions without noise predictions.

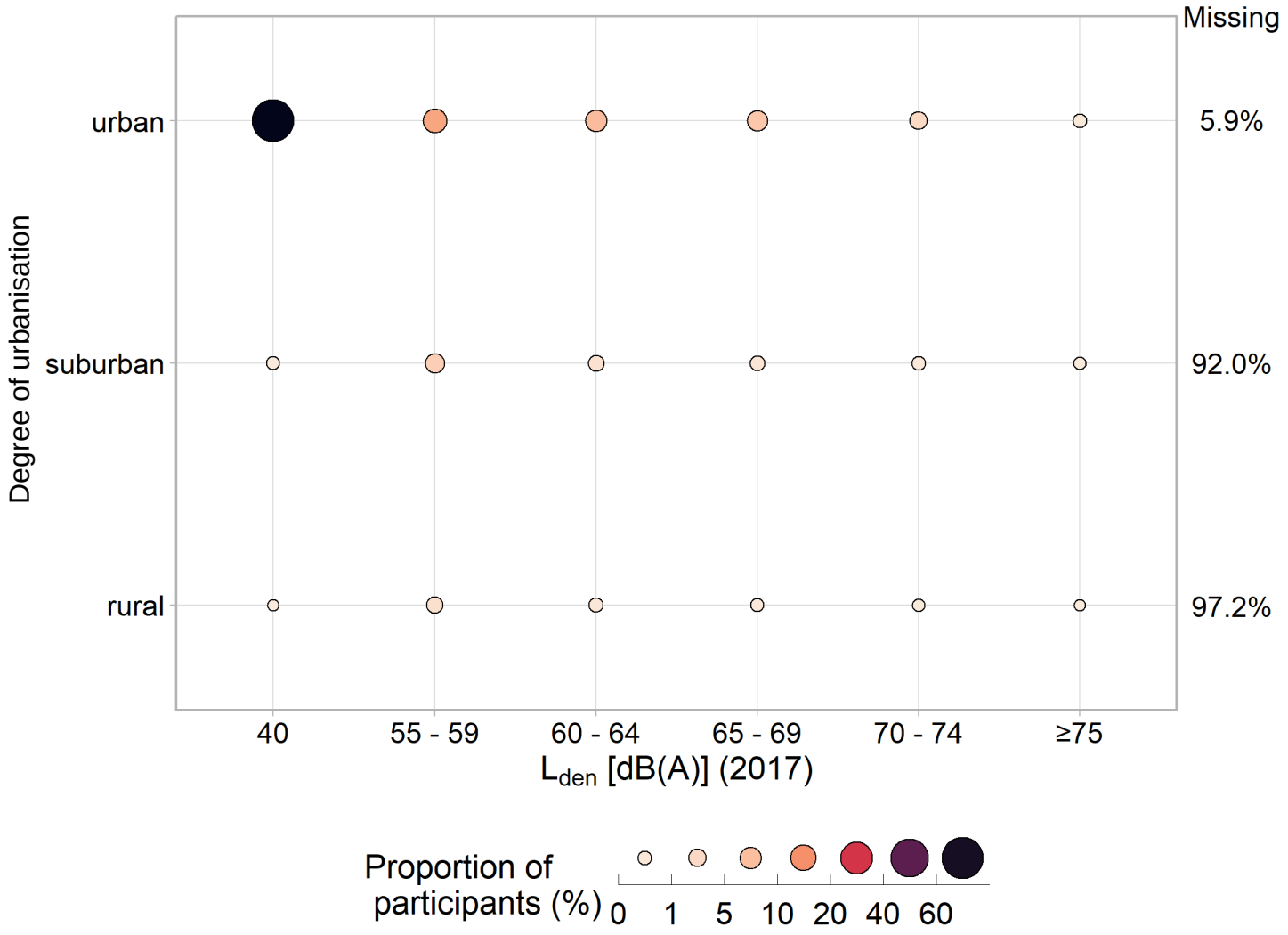

Figure S11: Annual 2017 ambient traffic noise day-evening-night levels (Lden) at the residencies of NAKO participants stratified by study centre (ordered from North to South) and pooled. Proportion of missings indicates participants living in regions not covered by the Environmental Noise Directive (END) mapping obligation, whereas 40dB(A) was assigned to participants living in END mapping regions without noise predictions.

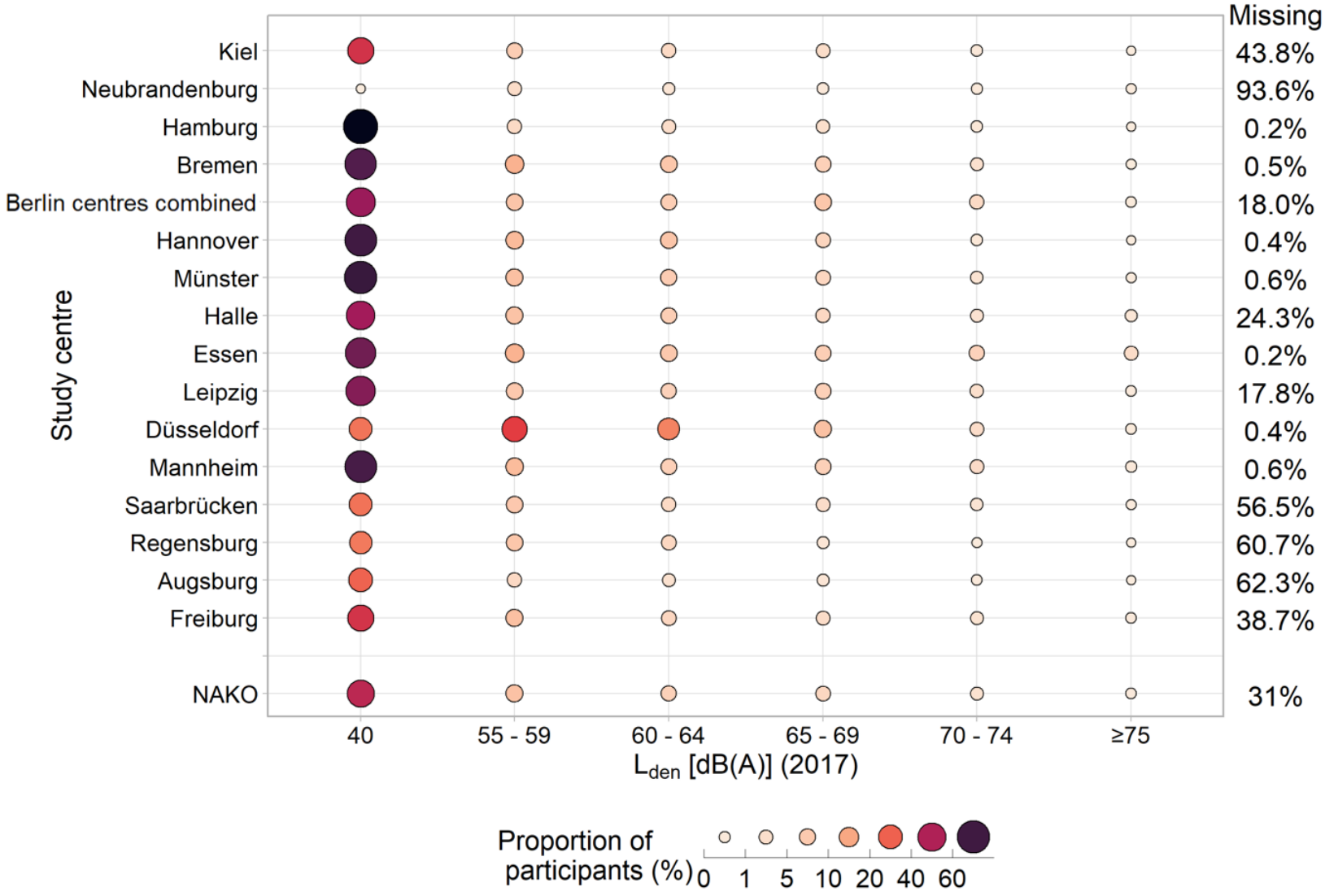

Figure S12: Spatial patterns of the predicted annual mean (T_{mean}), minimum (T_{min}), and maximum (T_{max}) air temperature in Germany, 2014-2019.

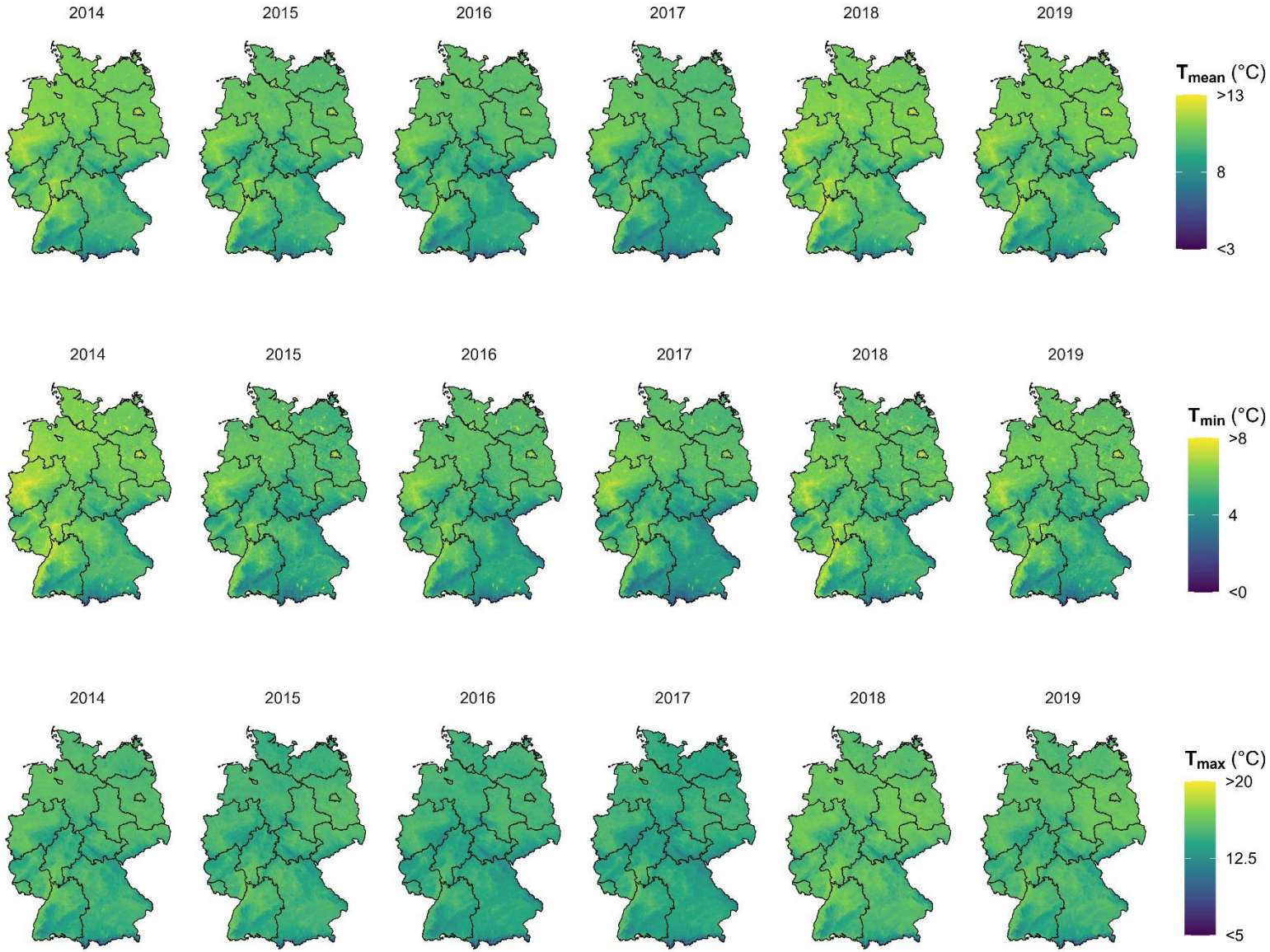

Figure S13: Annual mean (T_{mean}), minimum (T_{min}), maximum (T_{max}) air temperature and the diurnal temperature range (T_{range}) at the residencies of NAKO participants, 2014-2019 stratified by degree of urbanisation.

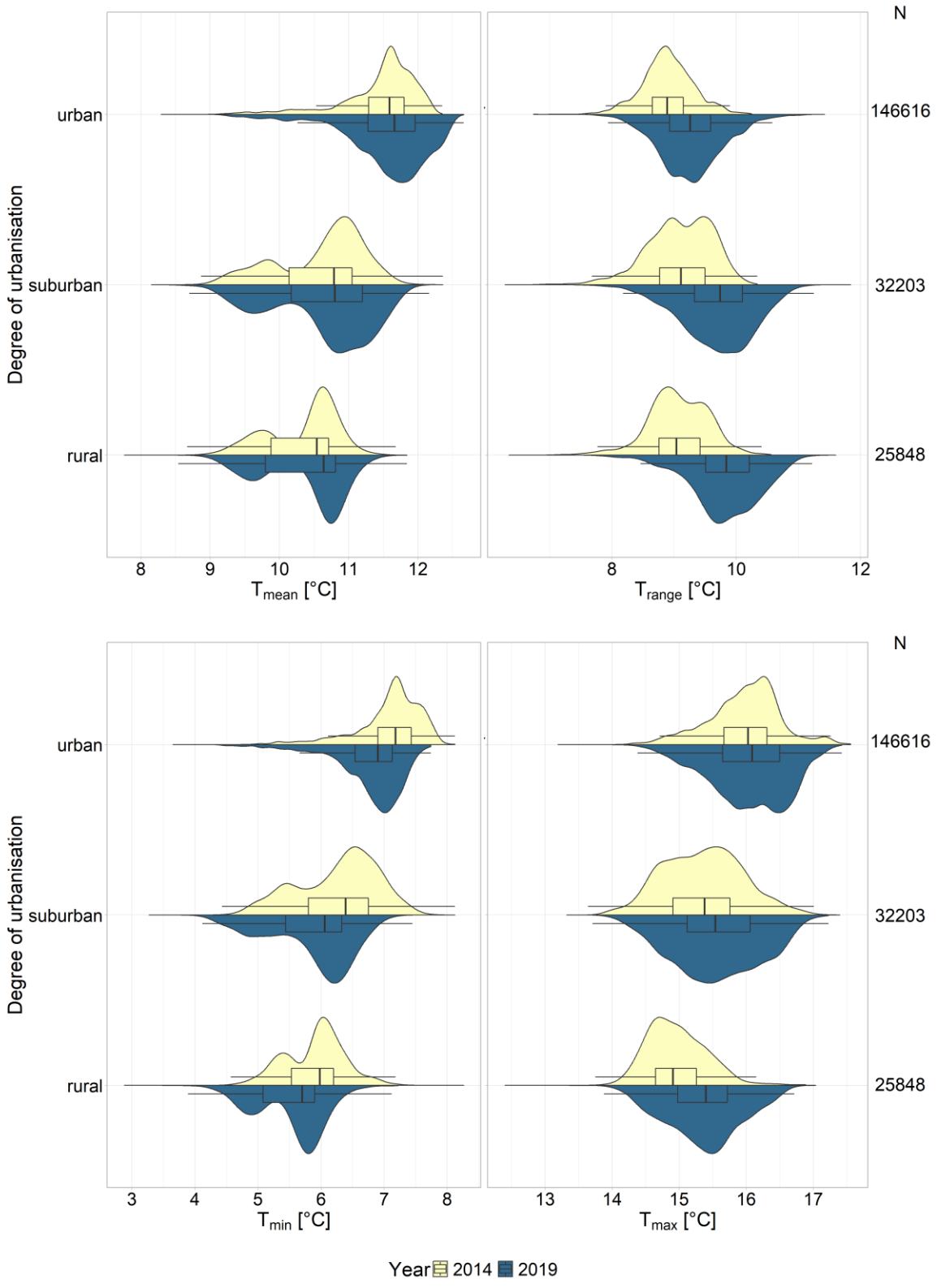

Figure S14a: Annual 2014 and 2019 mean air temperature (left) and diurnal air temperature range (right) at the residencies of NAKO participants stratified by study centre (y-axis ordered from North to South) and all NAKO participants combined (NAKO). N indicates number of participants.

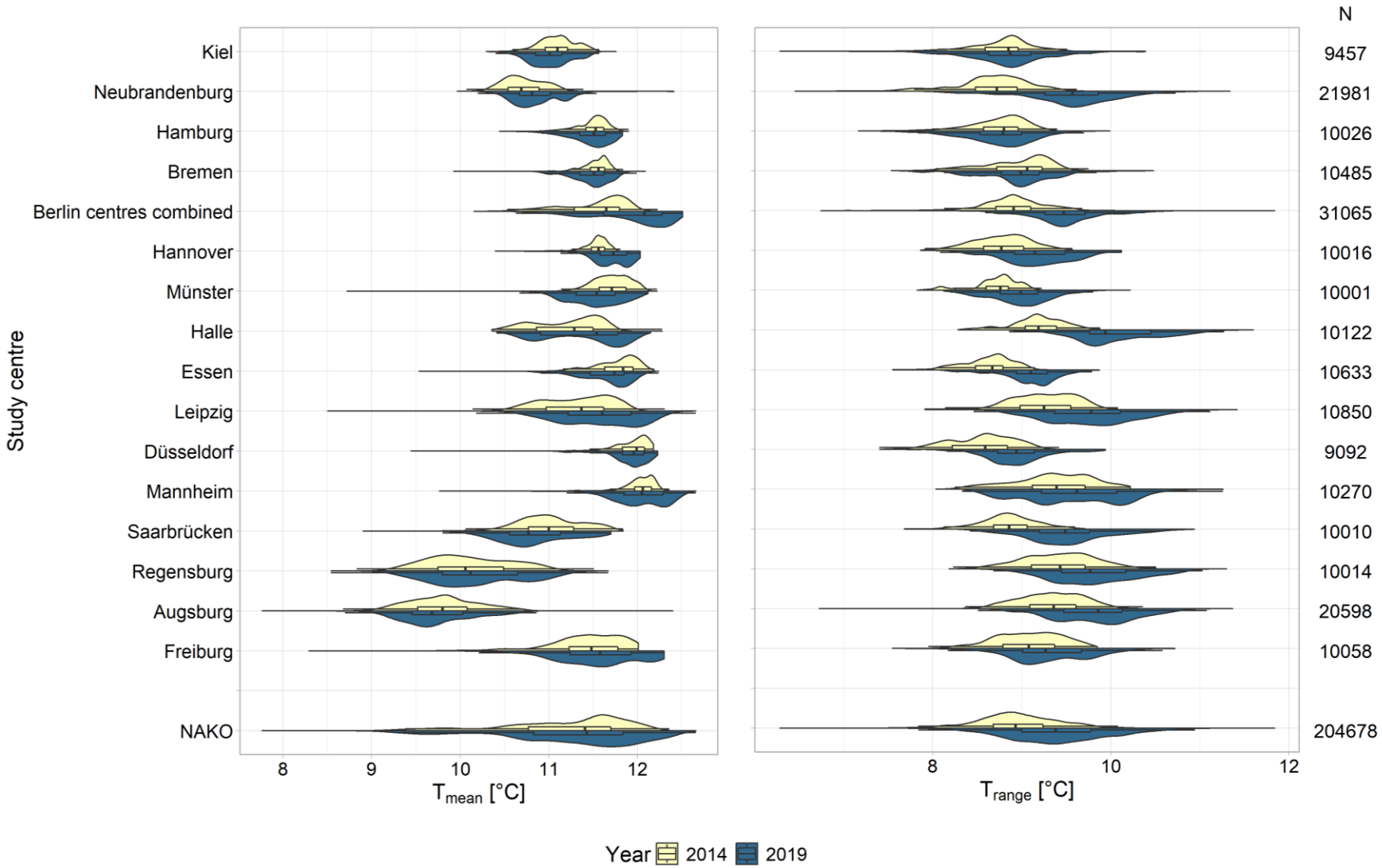

Figure S14b: Annual 2014 and 2019 minimum and maximum air temperature at the residencies of NAKO participants stratified by study centre (y-axis ordered from North to South; Berlin centres combined) and all NAKO participants combined (NAKO). N indicates number of participants.

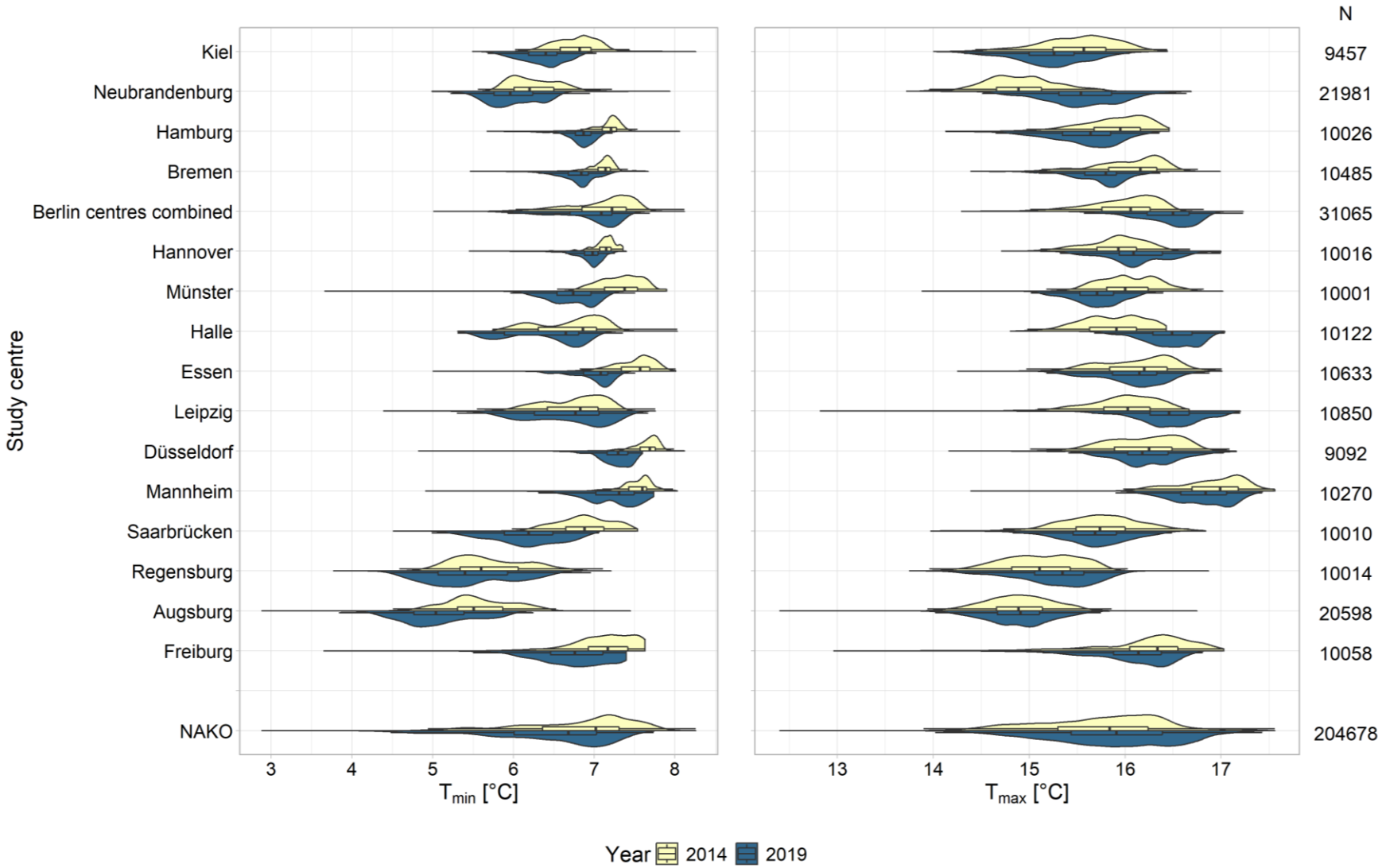

Figure S15: Spatial patterns of relative humidity (annual mean and standard deviation) in Germany, 2014-2019.

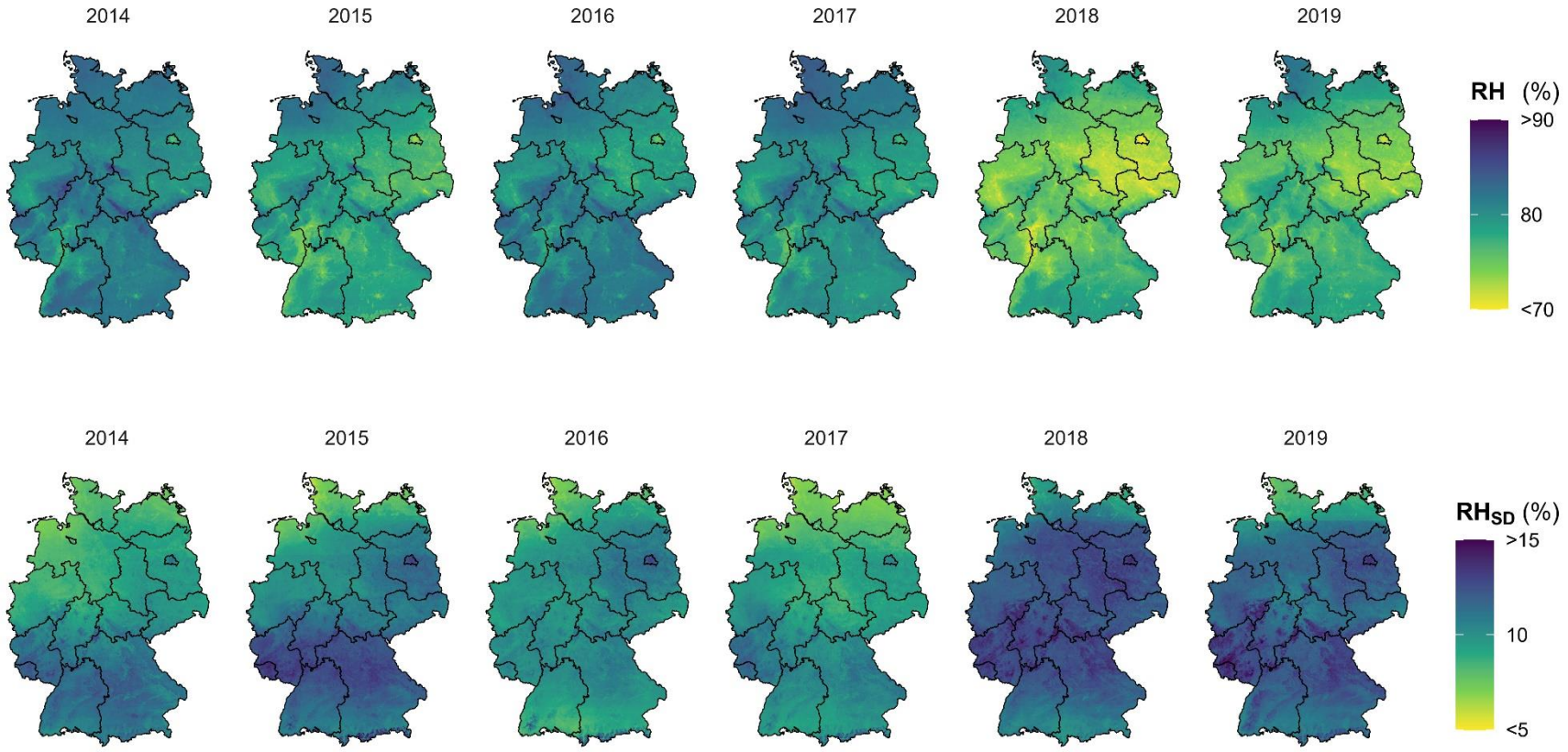

Figure S16: Relative humidity (annual mean and standard deviation) at the residencies of NAKO participants, 2014-2019 stratified by degree of urbanisation.

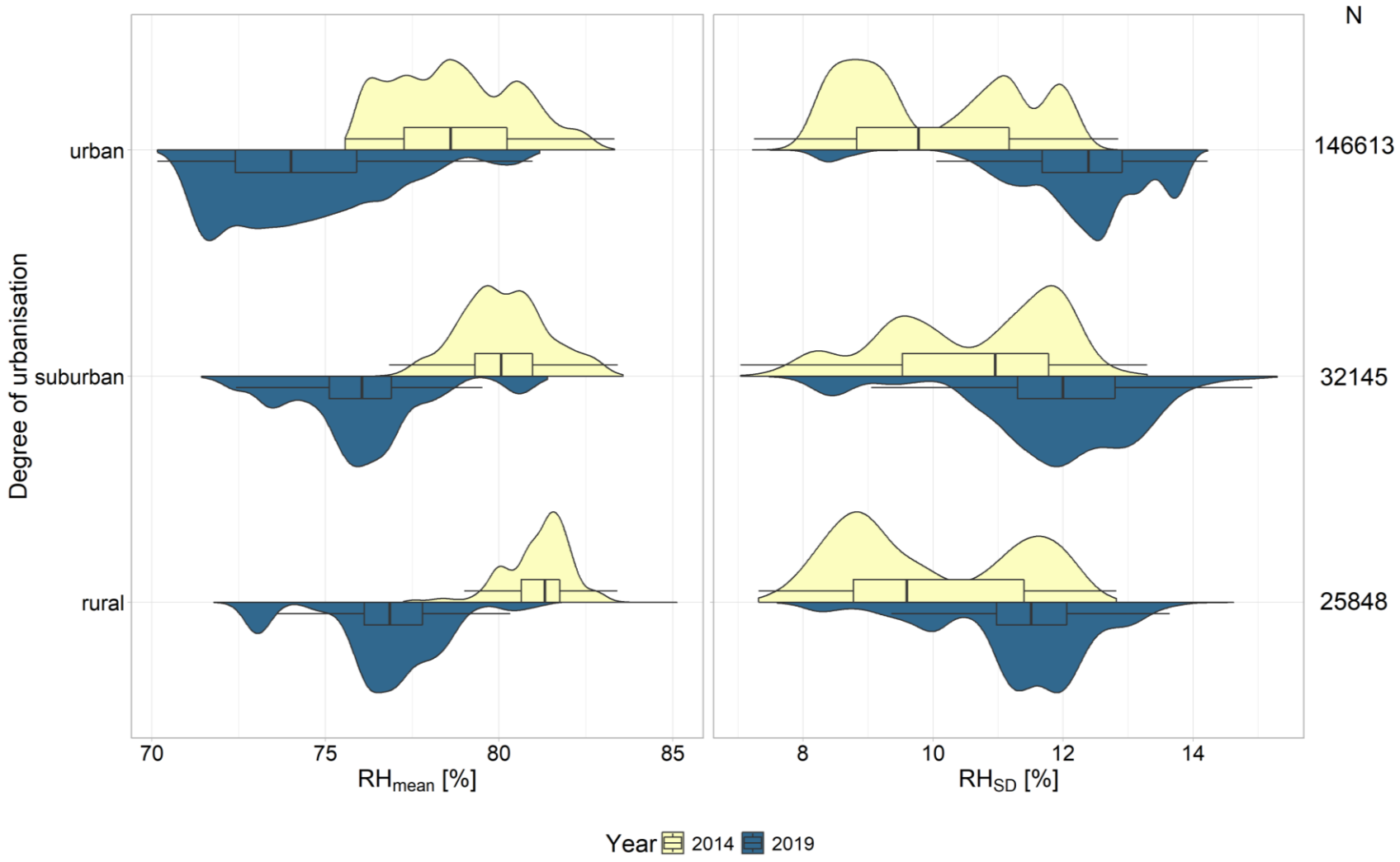

Figure S17: Annual 2014 and 2019 mean relative humidity (left) and its standard deviation (right) at the residencies of NAKO participants stratified by study centre (y-axis ordered from North to South) and all NAKO participants combined (NAKO).

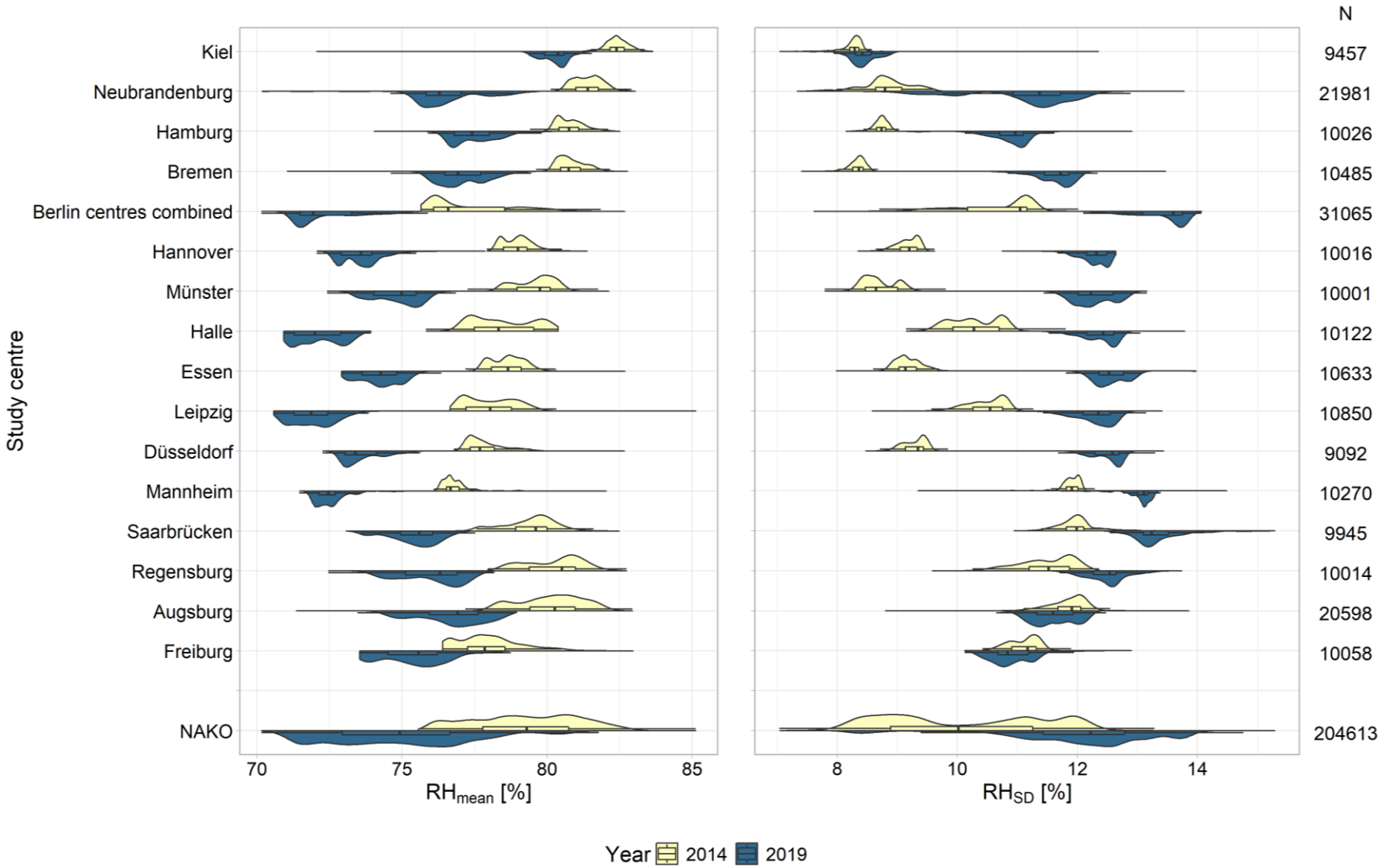

Figure S18: Spatial patterns of Normalized Difference Vegetation Index (NDVI, average 27.06.2015 - 29.09.2017) within 300 m (left) and 1 km (right) in Germany.

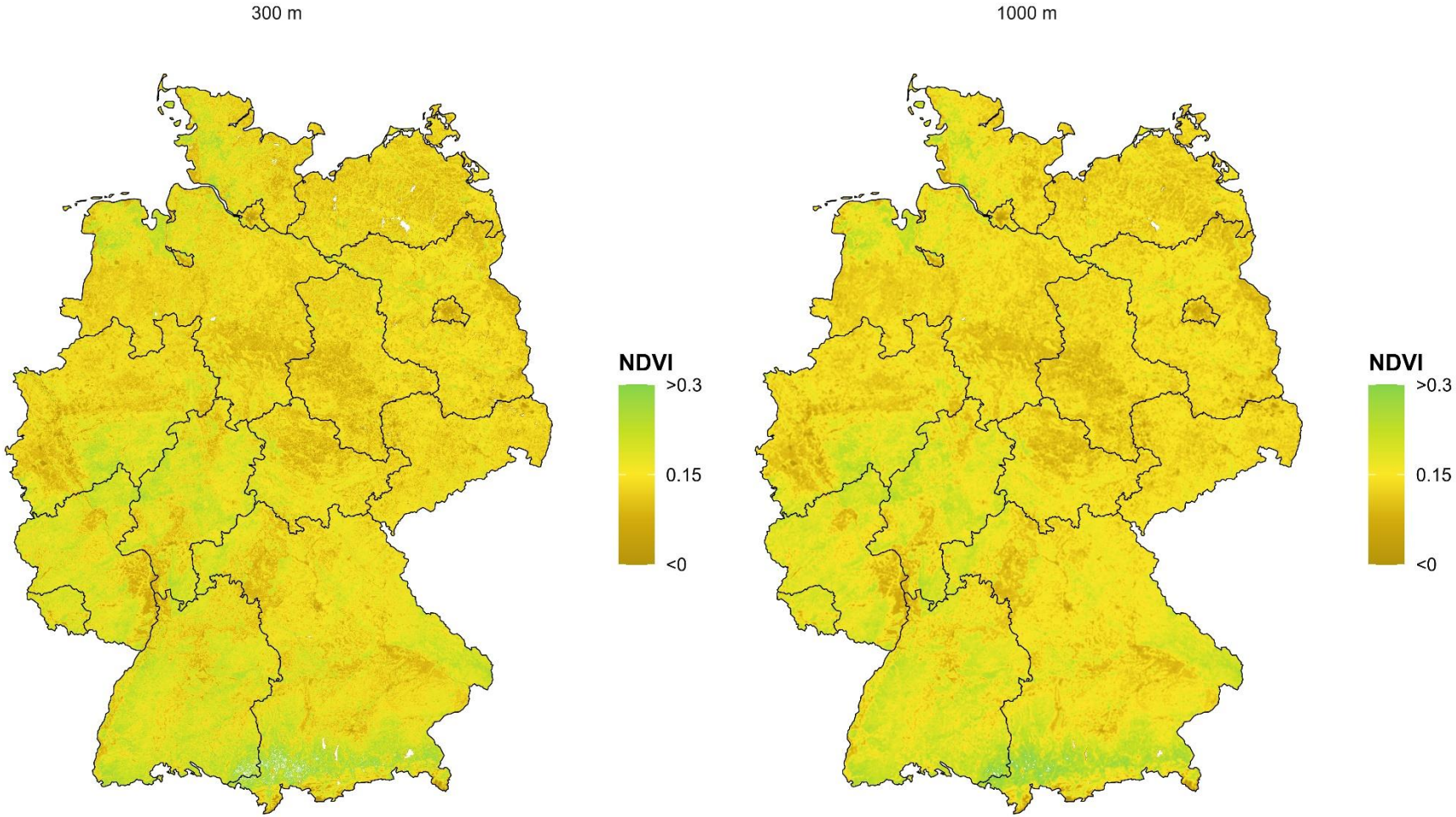

Figure S19: Spatial patterns of MODIS Normalized Difference Vegetation Index (NDVI, average of monthly values for vegetation period from March to October) in Germany.

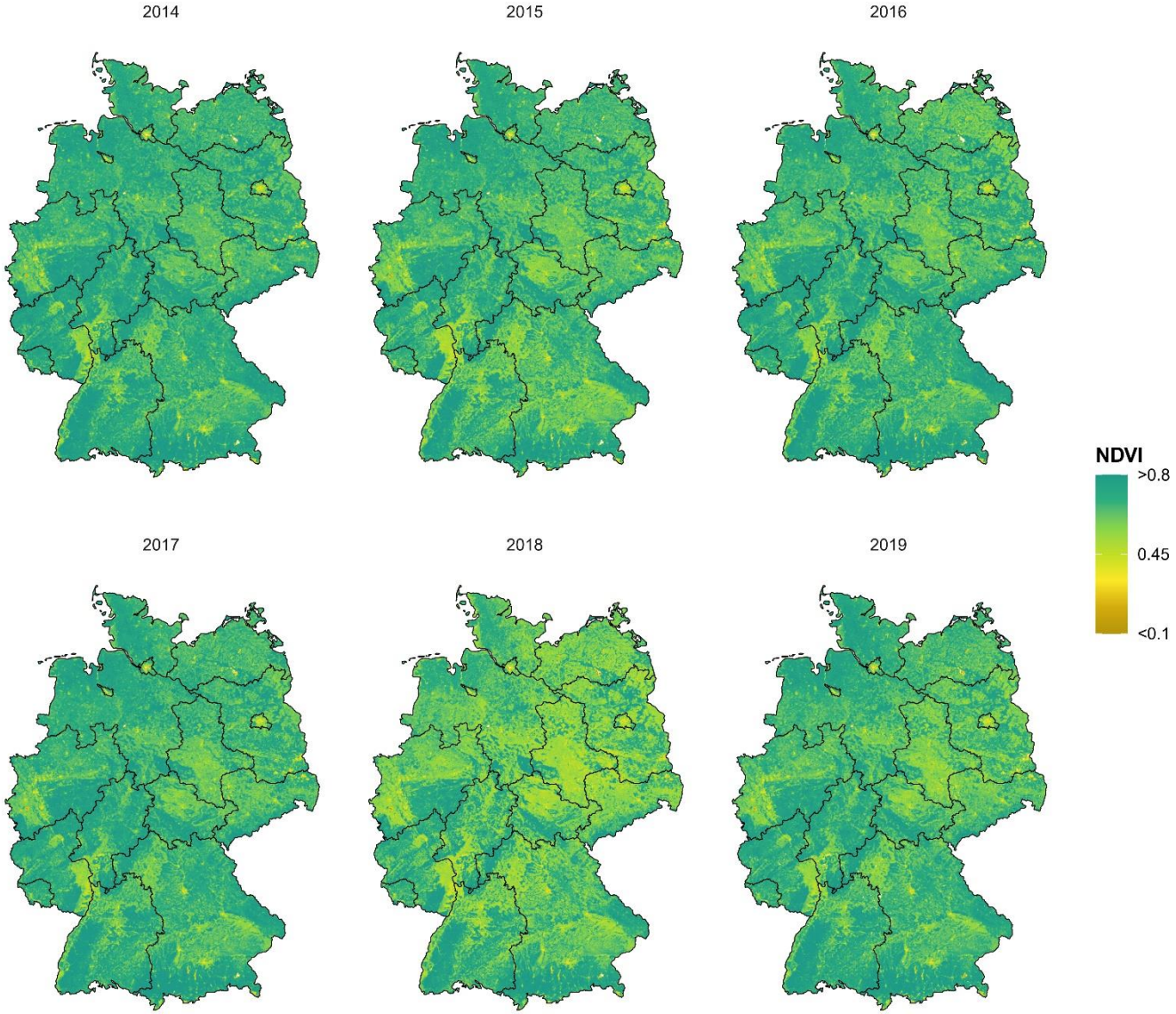

Figure S20: Normalized Difference Vegetation Index (NDVI from German Aerospace Center, average 27.06.2015 - 29.09.2017 within 1km (left) and average monthly March to October values for 2014 and 2019 (right)) at the residencies of NAKO participants stratified by degree of urbanisation.

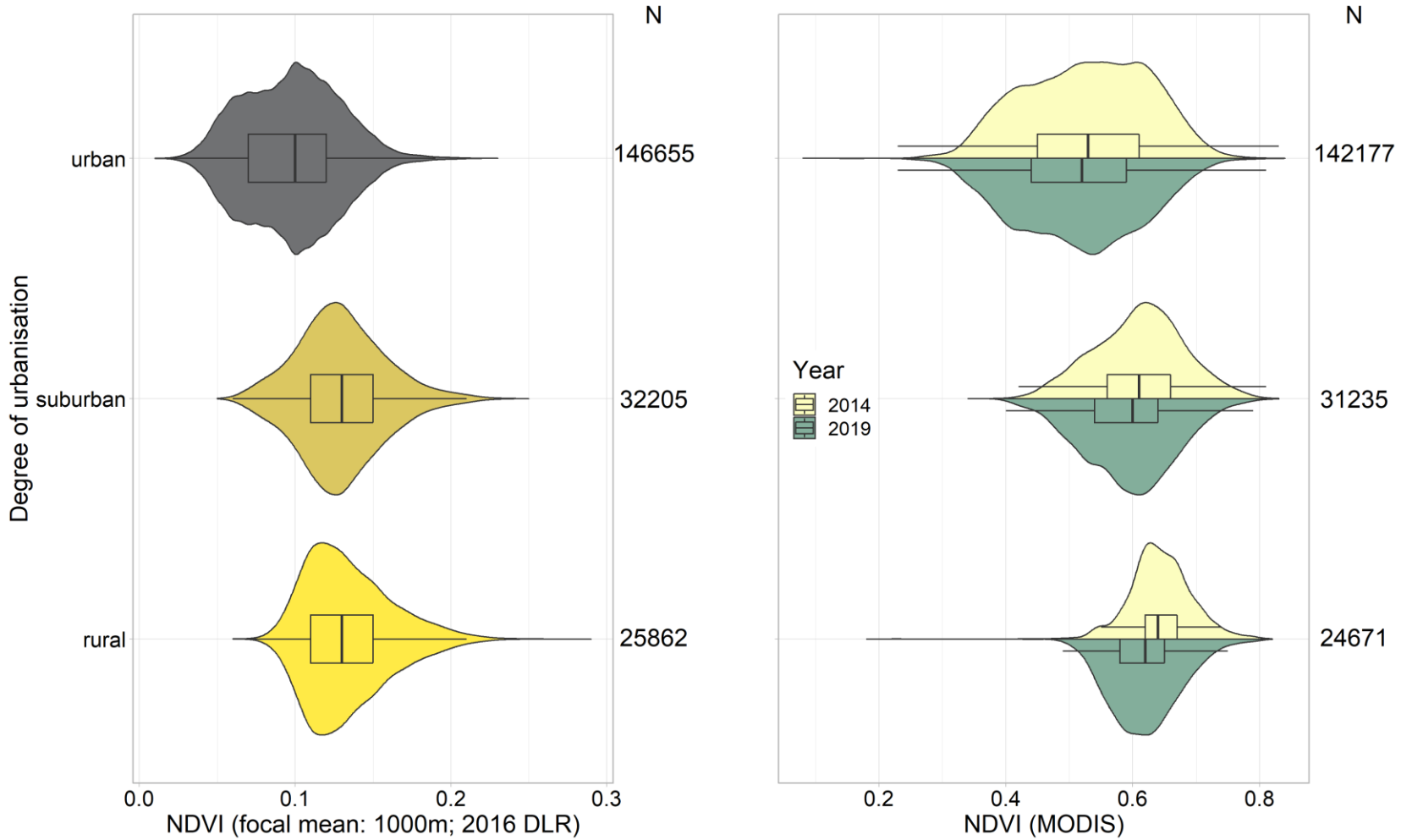

Figure S21: Normalized Difference Vegetation Index (NDVI from German Aerospace Center, average 27.06.2015 - 29.09.2017 within 1 km (left) and average monthly March to October values for 2014 and 2019 (right)) at the residencies of NAKO participants stratified by study centre (y-axis ordered from North to South) and all NAKO participants combined (NAKO).

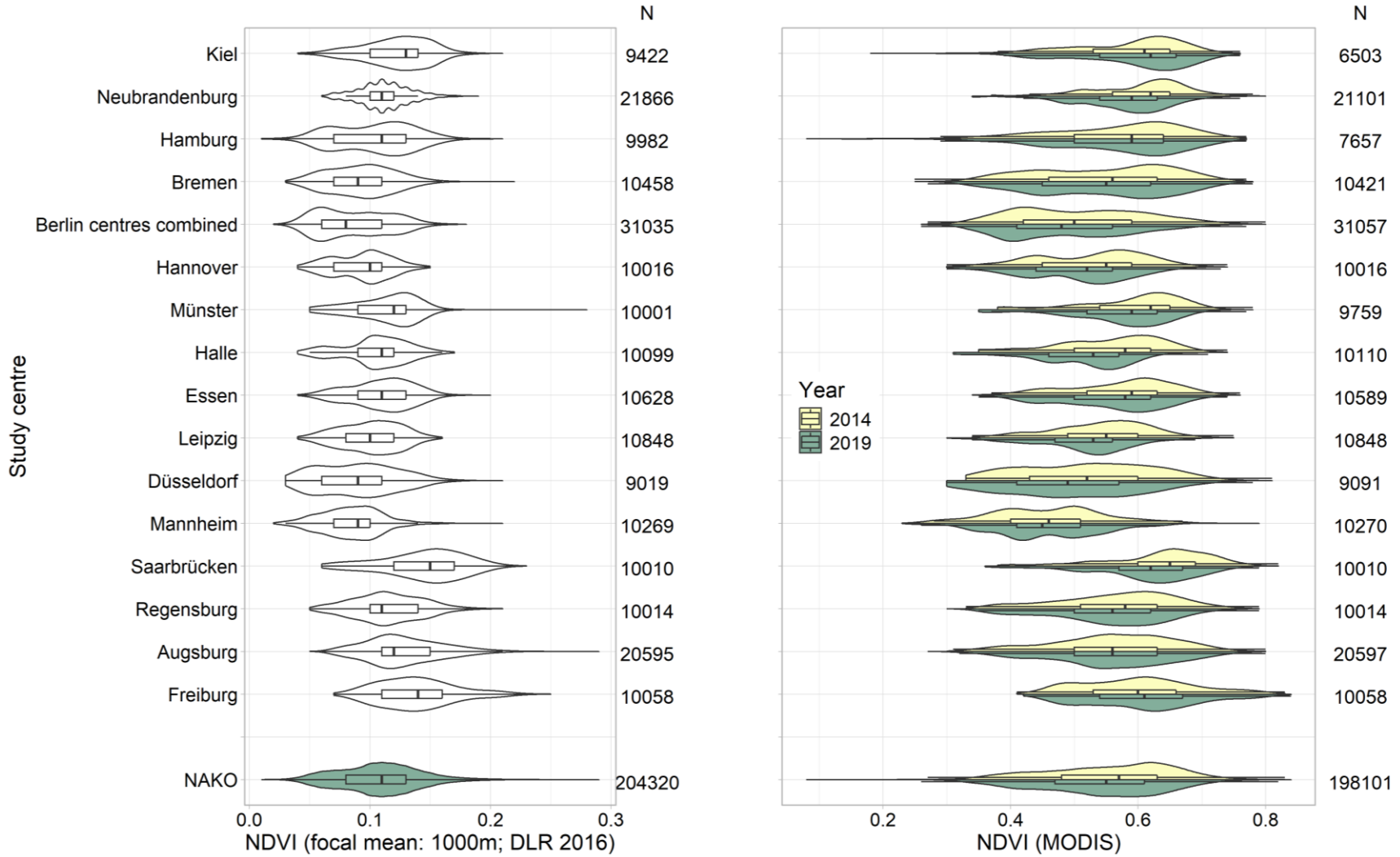

Figure S22: Spatial patterns of four land cover types (artificial land, open soil, vegetation, water; compiled by from German Aerospace Center based on Sentinel-2 imagery from 27.06.2015 - 29.09.2017) in Germany.