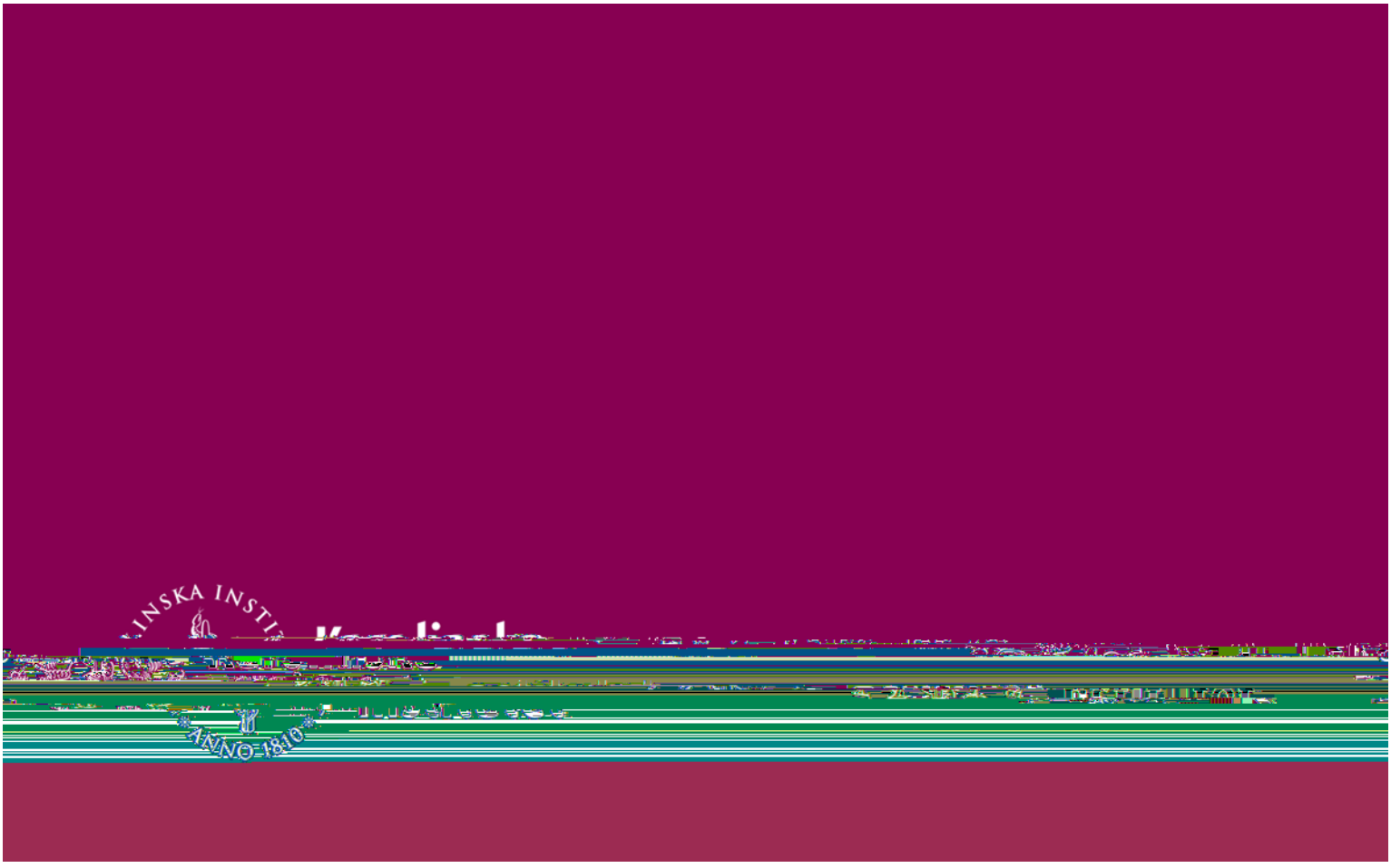
IMM Ambient Exposome database Application Procedure

Institute of Environmental Medicine (IMM)



The *IMM ambient exposome*

Temperature	Daily mean temperature (°C)	Spatiotemporal models	Models developed by IMM including monitoring sites, satellite data, land cover, meteorology, population, among others

In the data request form, the applicant will need to indicate which data from the IMM exposome database that are requested, and specify the following aspects:

The applicant should carefully check the area coverage of each model. For example, the air pollution dispersion model data are only available for Stockholm and Uppsala County, whereas the spatiotemporal model data are available for the whole of Sweden. If the model covers larger areas, it goes at the expense of lower spatial resolution.

Multiple time periods of exposures are available, and these vary depending on the exposure factor of interest. Carefully check the time period of availability in the descriptions below.

Depending on the aim of study, the applicant should differentiate short-term exposure studies (such as daily exposure levels) and long-term studies (annual exposure levels).

3.1.3. Satellite-based spatiotemporal models

Spatio-temporal model	
Model	*This multi-stage methodology is based on a machine learning method ς random forests ς to estimate daily mean concentrations of 5 air pollutants. This model uses data from multiple sources, including satellite retrievals, dispersion models, land cover, meteorology, population, road traffic, among others. A methodology was developed to calibrate such data on measurements of air pollutants from existing monitoring networks. The calibration model was evaluated through cross-validation procedures, by holding out individual monitors, training the model on the others, and checking the performance of the model fit on the left-out sites. Then, the model was applied to all places and days with no ob nBT/Fe502 0 0 1 182.58 582.67 Tm0.502 g0.502 G[*

Model	*Data on spatial greenness as a normalized difference vegetation index (NDVI), with 25 x 25m resolution, were obtained from satellite images (Landsat 5, Landsat 7 and Landsat 8) for the period 1 May-30 September 1988ς2019. To avoid underestimating NDVI values due to random cloud contamination, the annual value of each 25 x 25m pixel was replaced with a 5-year average value (2 years before, the current year, 2 years after). Because NDVI for water is represented by negative values, we replaced all water surfaces (lakes, streams, oceans) with "data-free" (no-data) surfaces.* *References:* *Chander G et al. Remote Sensing of Environment 2009, 113:5, https://doi.org/10.1016/j.rse.2009.01.007* *Persson A et al. Environment International 2018, 121:1, doi:10.1016/j.envint.2018.10.009*
Indicators	*Address based NDVI values (0-1) are calculated within corresponding buffers, for example circles of 100m, 250m, 500m or 1000m around coordinates.* *Coverage/proportion of water i.e. blue space could be calculated for each corresponding buffer.*
Source Temporal resolu	*Processed satellite images (Landsat 5, Landsat 7 and Landsat 8)*

3.4.1. Monitoring stations

Data on monitoring stations across Stockholm are available through SLB-analys (in Swedish Stockholms Luft- och Bulleranalys). We can provide guidance on how to obtain the data, the data request will need to go through SLB-analys. For more information check their website: https://www.slb.nu/slbanalys/

3.4.2. Satellite-based spatiotemporal models

Spatio-temporal hybrid model	
Model	*This multi*

Responsible contact	*exposome-imm@ki.se*

questionnaire data/self-reported, address data is not available or others). The applicant also needs to provide information about the and used. The coordinate system used in by the IMM Exposome Group is Sweref 99 TM (EPSG 3006).

However, we can give support and contacts how to geocode and to perform an accurate manual cleaning, which is needed to provide sufficient quality of the geocoding process. Please contact us for further guidance.

1) The Researcher transfers a file containing personal identification numbers (personnummer) of the study participants to the Swedish Tax Agency (Skatteverket) and unique study id numbers.
2) The Tax Agency extracts information from the Population Registry (Folkbokföringsregistret) for every individual and for selected years on for example the following variables (as agreed on beforehand):
    a. Street address (Gatuadress)
    b. Zip code (Postnummer)
    c. Town (Postort)
    d. County code (Länskod)
    e. Municipality code (Kommunkod)
    f. Registration date (Folkbokföringsdatum), the date the person moved to the address
3) The Tax Agency removes the personal identification numbers and only the study id numbers are kept as identifiers.
4) The Tax Agency transfers the file back to the Researcher.
5) The Researcher transfers the file to the company Metria. Addresses are matched with address coordinates from the Property Registry (Fastighetsregistret). The variables added are X and Y SWEREF coordinates and an indicator of the quality of the match, called rank.
6) The file is transferred back to the Researcher. The quality of the match, indicated by the rank variable, ranges from 0-1, depending on the precision of the rank. Below 0.8 is considered low rank, and these coordinates may be imprecise or inaccurate. The Researcher needs to decide how to handle these (exclude, correct manually, etc).

2. Access to data

The Recipient is responsible for ensuring that Exposome data within the framework of the Agreement are only used in accordance with paragraph 1 and that IMM must be involved in all cooperation where the transferred data are utilized. Access to Exposome data is given based on a standardized application procedure and approval after evaluation in a designated review group.

Data request - IMM Exposome database template

The applicant should read the document IMM Ambient Exposome database Application Procedure before filling in the data request form. This clarifies which exposures can be provided by IMM and what data are needed to estimate the exposures. The data request form needs to be completed by the principal investigator of the project. The IMM Ambient Exposome Database Review Committee meets approximately every 3 months to discuss the feasibility of each project (check the dates for Review Committee meetings: https://ki.se/en/imm/the-imm-ambient-exposome-database). A formal answer will be sent to the applicant regarding the use of IMM Exposome data.

(The PI must have a PhD, if not, please provide a collaborator with a PhD)

Work institution:	
PI Surname:	PI First name:
E-mail:	Telephone:
[redacted]ss:	
[redacted]	

(give a short overview of the aims of the study, the background, and the methodology):

Title of the project:

[redacted]

☐	No, but data will be available in the near future Please, specify:

What is the source of the addresses available in your study?	
☐	Registry data
☐	Questionnaire data/self-reported
☐	Address data are not available
☐	Others Please, specify:

Are the addresses geocoded (e.g., information on the exact longitude and latitude)? *(The IMM Exposome group does <u>not</u> provide service to geocode the addresses. However, we can give*

Remember that linkage of geographical data to health data where individuals can be identified are typically subject to requirement of ethical approval. The applicant PI is responsible for obtaining any ethical approval.

Karolinska Institutet
171 77 Stockholm
Telefon 08-524 800 00
ki.se