GLOBAL GEO HEALTH DATA CENTER Scientific Framework Software architecture: web app + services

Introduction

Our environment has a considerable impact on health. For instance, air pollution increases the risk for cardiovascular disease, a warm and humid | climate supports the spread of vectors causing | malaria, and green space may improve mental health. Understanding these health impacts is a | massive challenge as it requires quantifying the exposure to these environmental variables for each individual in a population. The Global and Geo Health Data Centre (GGHDC) is taking this challenge by providing a web service that enriches population data with information on personal exposures to the environment. We combine high performance geocomputation and spatial data analysis to calculate personal exposures for individuals using their location data, either directly measured using mobile devices or by agent-based simulation modelling. Exposures are calculated from archived national and global environmental ¦ information (up to 5 m spatial and 1 h temporal resolution) or data generated on the fly using environmental models running as microservices.

GGHDC Team info@globalgeohealthdatacenter.com

Human Geography & Planning

University Medical Centre Utrecht

- Physical Geography (PCRaster)
- ITS

Institute for Risk Assessment

SURFsara

eam: Rick Grobbee, Martin Dijst, Bert Brunekreef, Derek Karssenberg, Ilonca Vaartjes, Folkert-Jan de Groot, Kor de Jong, Oliver Schmitz, Ivan Soenario, Maciek Strak, Harm de Raaff, Leendert van Bree, Peter Hessels, Dick Ettema

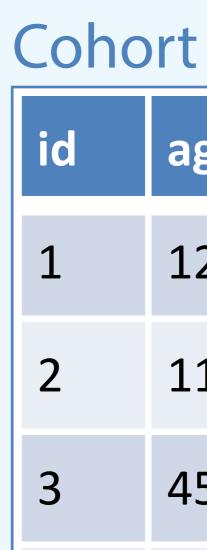
Contributions: Michiel Geijer, Gerard Hoek, Anna-Maria Ntarladima, Maartje Poelman, Mei-Po Kwan, Monique Simons, Carlijn Kamphuis, Marco Helbich, Amit Birenboim, Maarten Zeylmans van Emmichoven

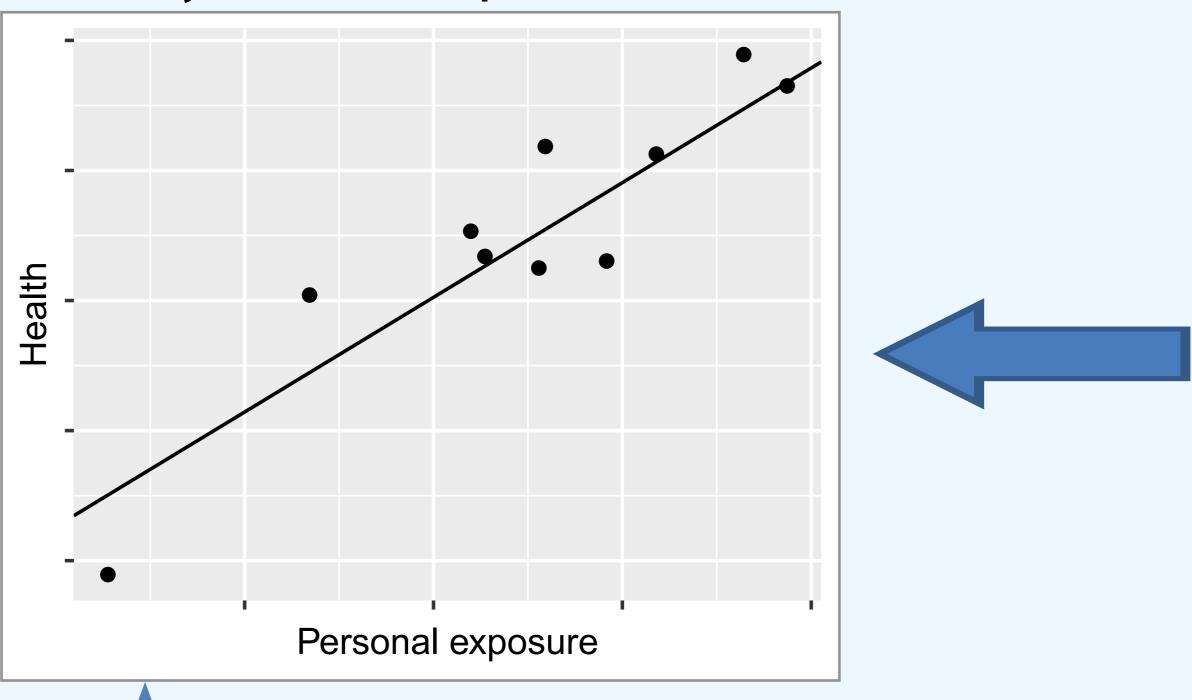
Funded by Utrecht University



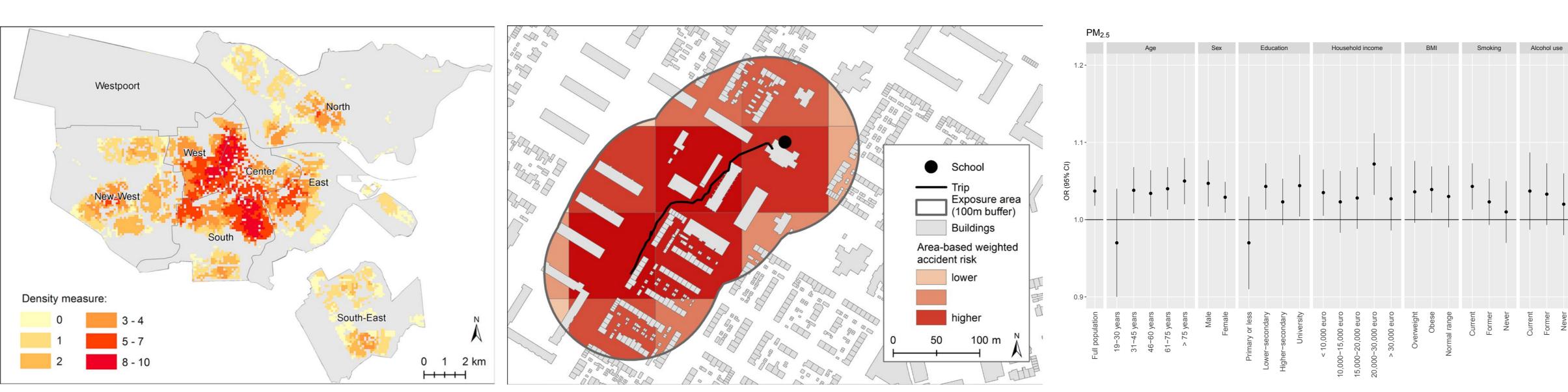


Utrecht University





Examples health impact



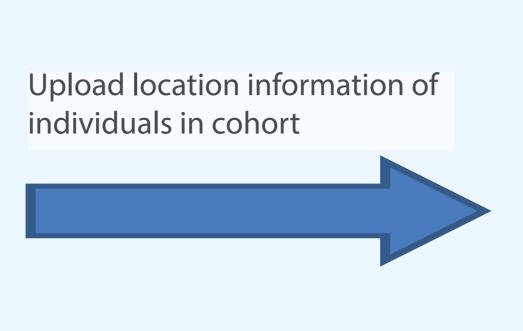
geography.

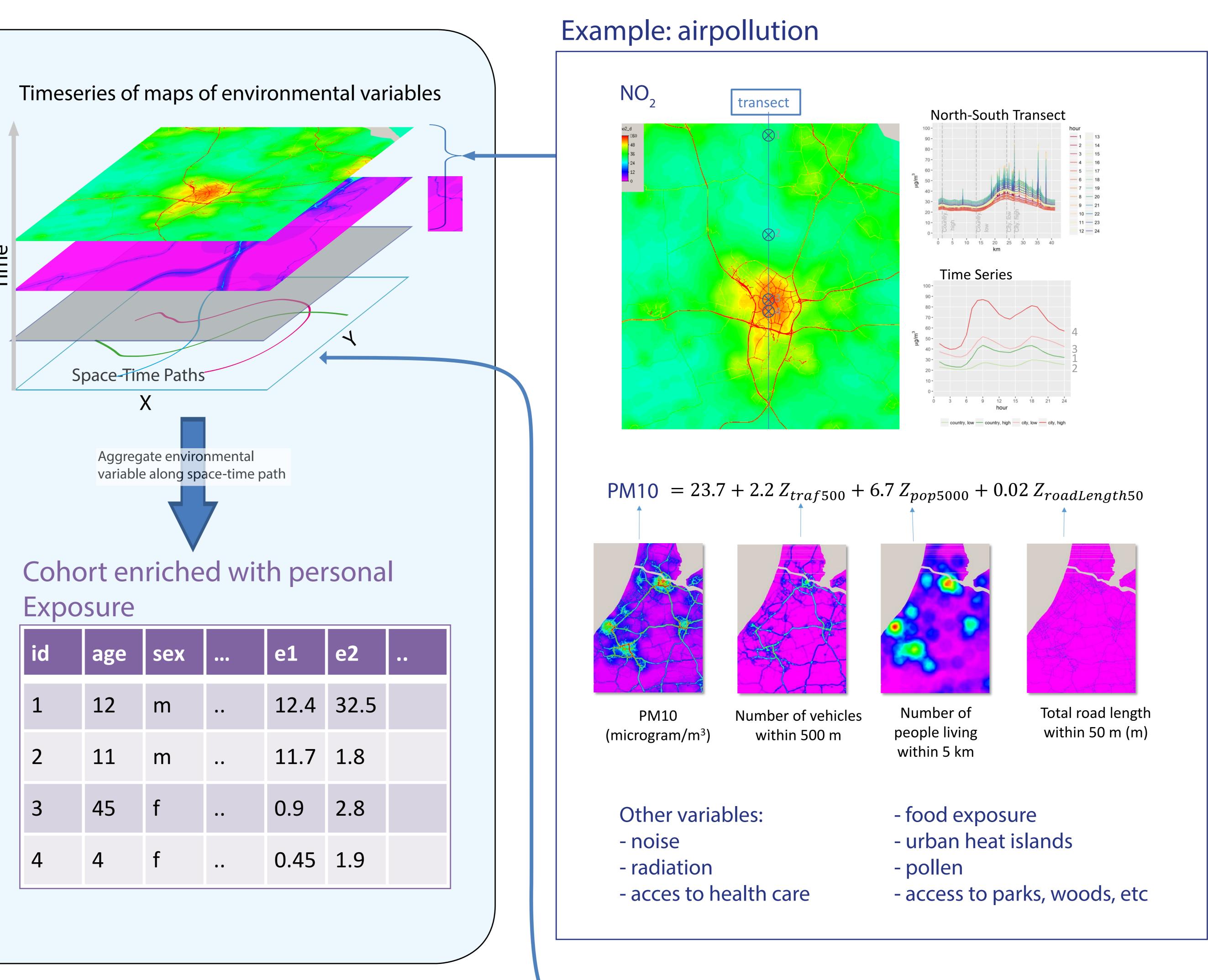
I I

I I

I I

12 m 11 m 45 ·





Identify health impacts

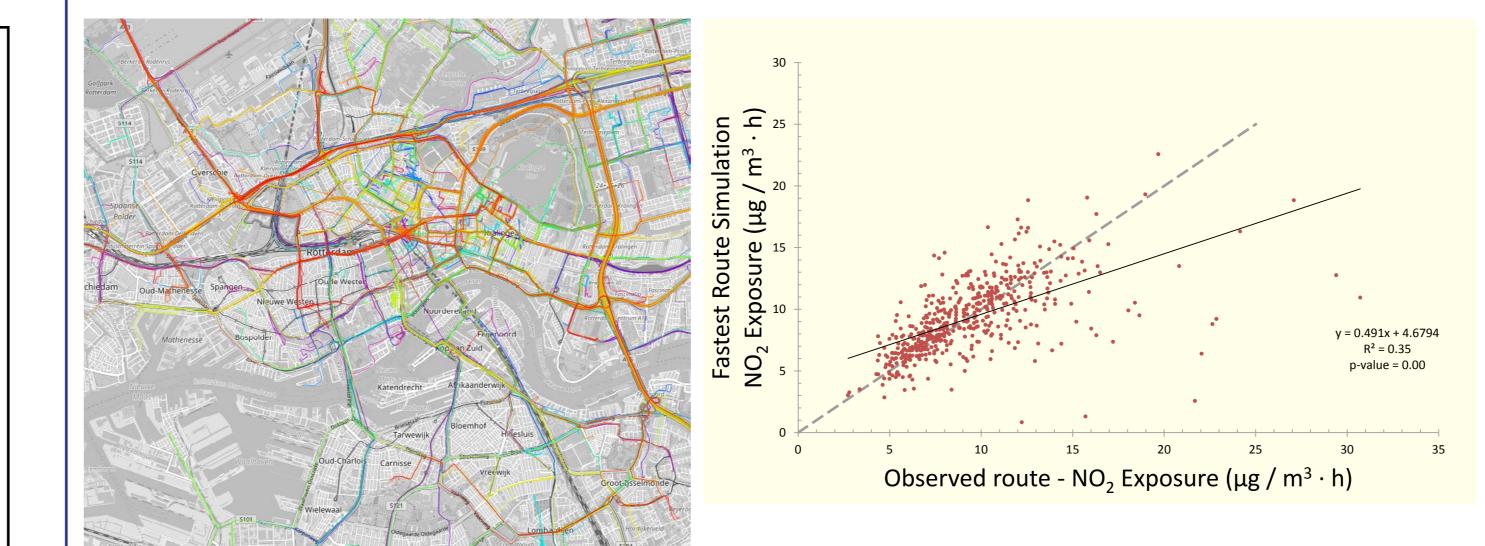
id	age	sex	•••	e1	e2	••
1	12	m	••	12.4	32.5	
2	11	m	••	11.7	1.8	
3	45	f	••	0.9	2.8	
4	4	f	••	0.45	1.9	

Density of supermarkets within 1000 meters per cell, used to calculate accessibility to healthy food. M.Helbich et al 2017, Applied

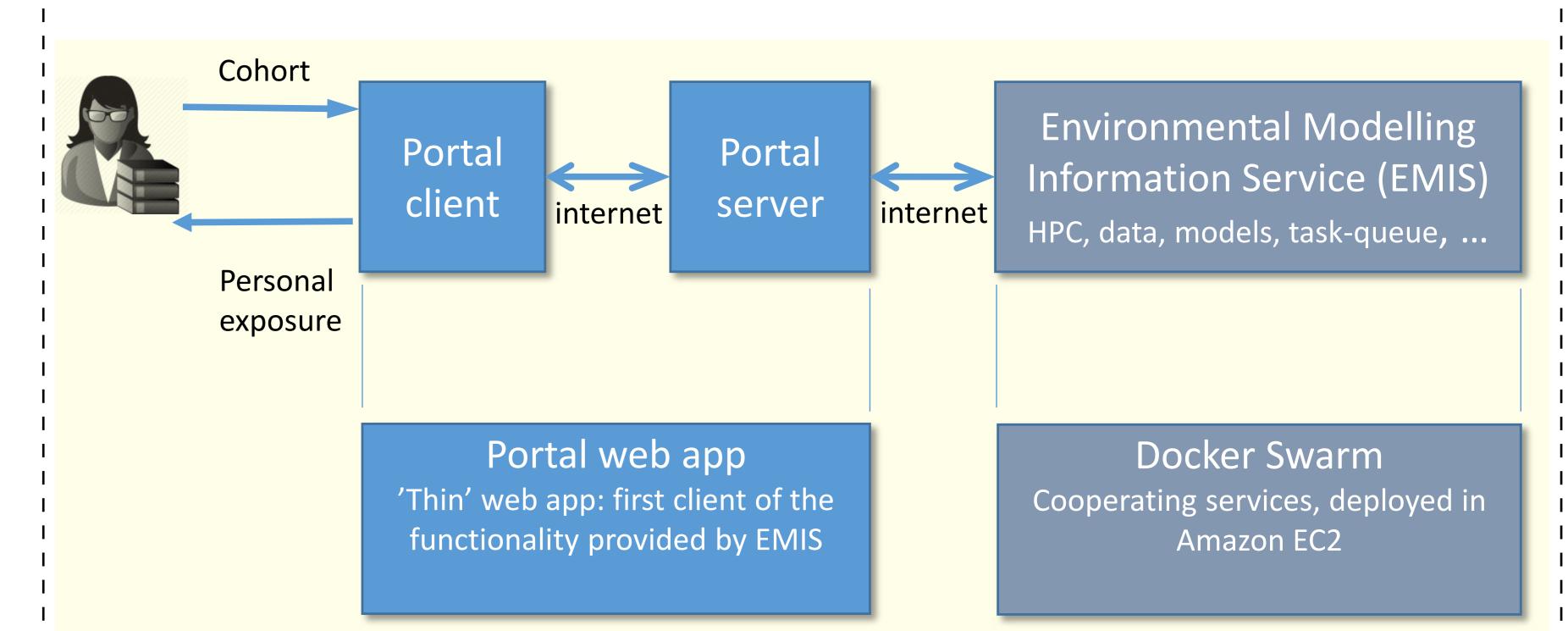
Accident risk exposure along a walking trip (brighter cells receive a lower weight). M. Helbich et al. 2016, Health & Place.

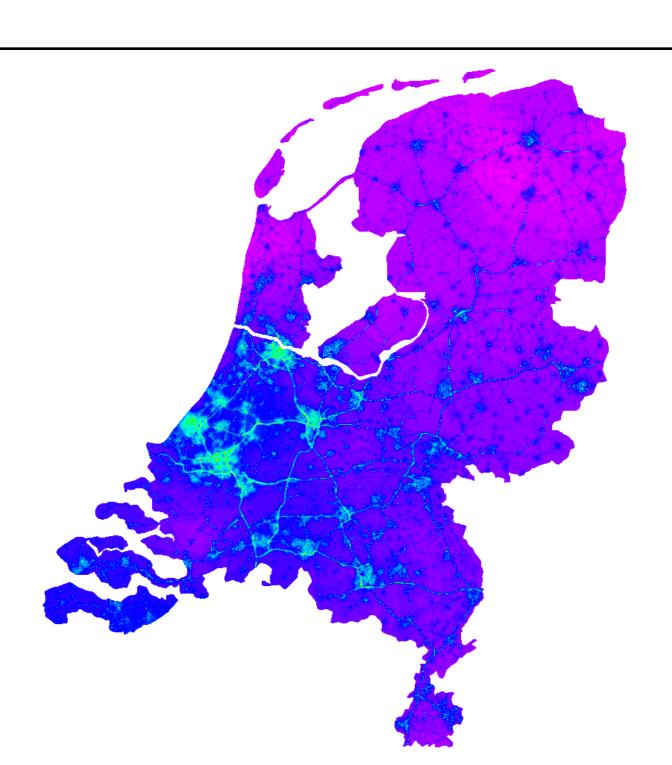
Assessment of linearity of the association between air pollution and diabetes (simulated data). M. Strak et al. 2017 (unpublished).

• Example: modelled space-time path

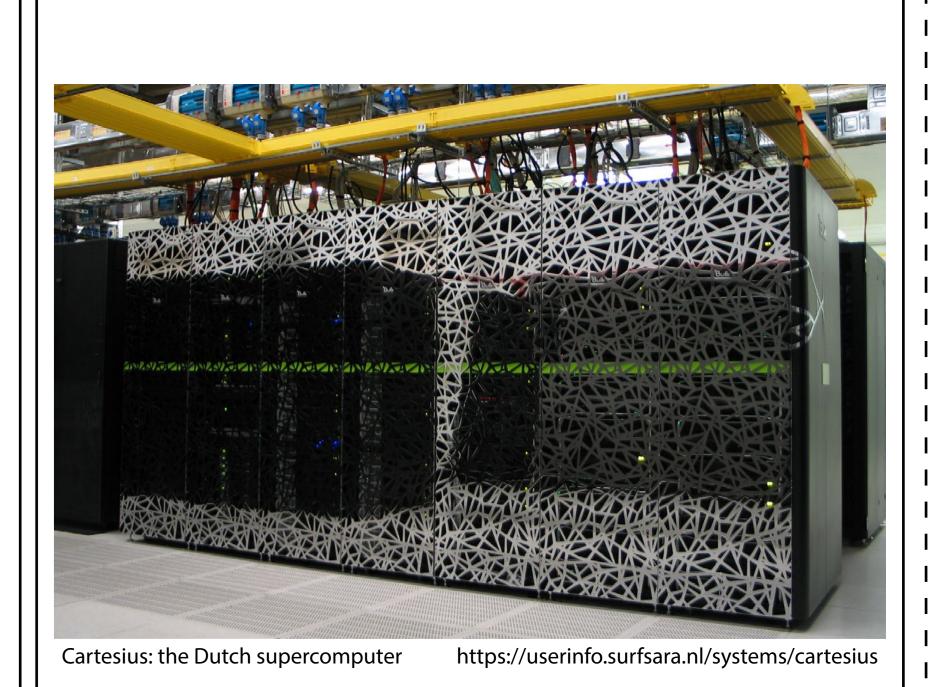


Left: space-time paths calculated with google routing API. Right: exposure along modeled route vs exposure along observed route. (Calculations: M. Geijer, data: D. Ettema).





Environmental datasets: - 6 air pollution models, 25 predictors - 5m grid and hourly time step nationwide - 1.5 billion raster cells, 15Gb per timestep - 125TB per model per modelled year



Geocomputation on HPC facilities: - parallel algorithms - distributed algorithms - parallel I/O

Modelling software stack

User applications Agent- and field-based models

API: Python binding

Core: modelling support (C++) LUE data model, parallel algorithms, meworks for temporal modelling, err propagation, agent based modelling, .

> **3rd party libraries** Boost, (parallel) HDF5, HPX,

- *Target user* is a domain expert, not a programmer
- *Target domain* undefined: generic building blocks
- Target platform is any platform, including HPC clusters