How do environmental factors correlate across different spatial measurement methods?

The Problem

Environmental factors in exposure measurement are everywhere around us. Examples include green space, air pollution, crime, and grocery stores. When investigating the effect of multiple environmental factor exposures to a health outcome in a study, it is possible that some may correlate, producing multicollinearity in statistical models.

The Aim

This study investigates to what extent the relationships between exposures are impacted by how they are quantified. This is done by assessing the exposure of 84 environmental factors around Dutch residential addresses using three exposure models with varying exposure size (100, 250, 500, and 1000 meters) and assessing the difference in exposure outcomes. 10 percent of all residential addresses in The Netherlands were used. Findings can help researchers in epidemiology choose the best model and exposure size for the most accurate results for their research.

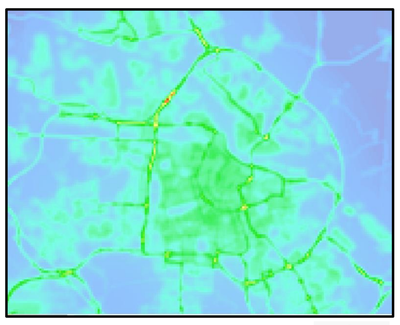
Thematic categorization

All 84 environmental factors are organized into categories and subgroups:

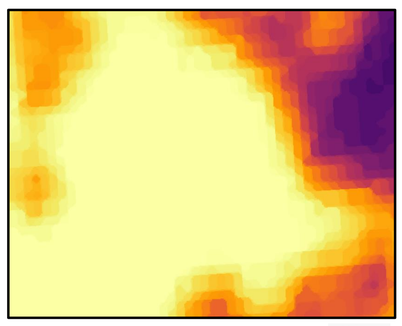
Category 1: Physico-chemical environment	Category 2: Built environment	Category 3: Social environment	Category 4: food environment
1)Air pollution	1)Crops	1) Social security	1) Food facilities
2)Temperatures quarterly	2)Green space and blue space	2) Education level	
3)Urban exposures	3)Grey space	3) Income	
	4)Roads and railroads	4) Livability	
		5) Crime	
		6) Schools and hospitals	

Spatial data types

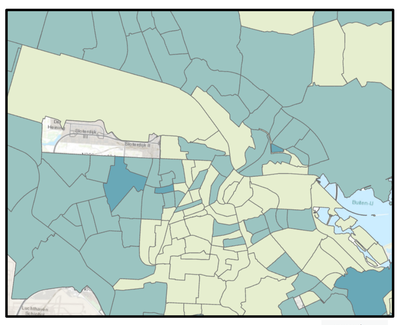
Environmental factors can come in many different spatial data forms, making it difficult to compare them and their exposures. Below are examples of different kinds of spatial data, centered around Amsterdam (though the study area is the entire Netherlands).



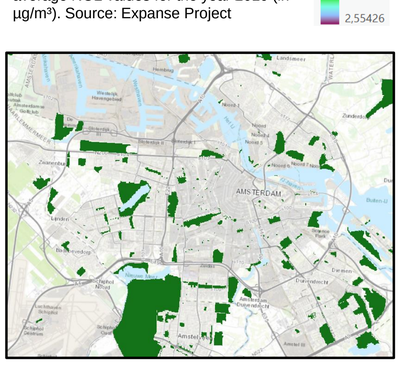
a) Field
average NO2 values for the year 2019 (in µg/m³). Source: Expanse Project



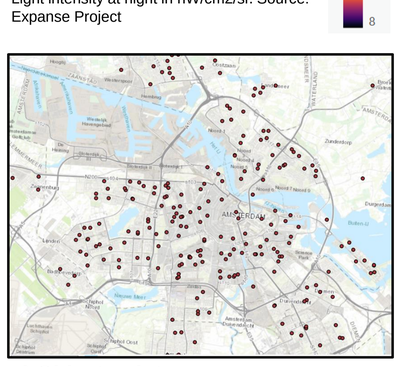
b) Event
Light intensity at night in nW/cm2/sr. Source: Expanse Project



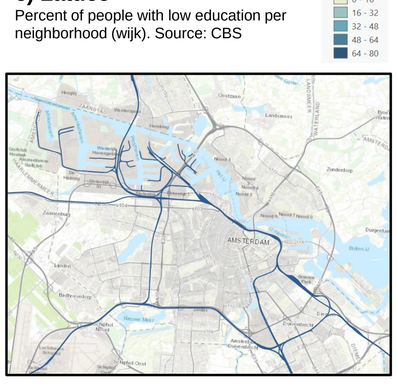
c) Lattice
Percent of people with low education per neighborhood (wijk). Source: CBS



d) Coverage
Parks (in green). Source: OpenStreetMap

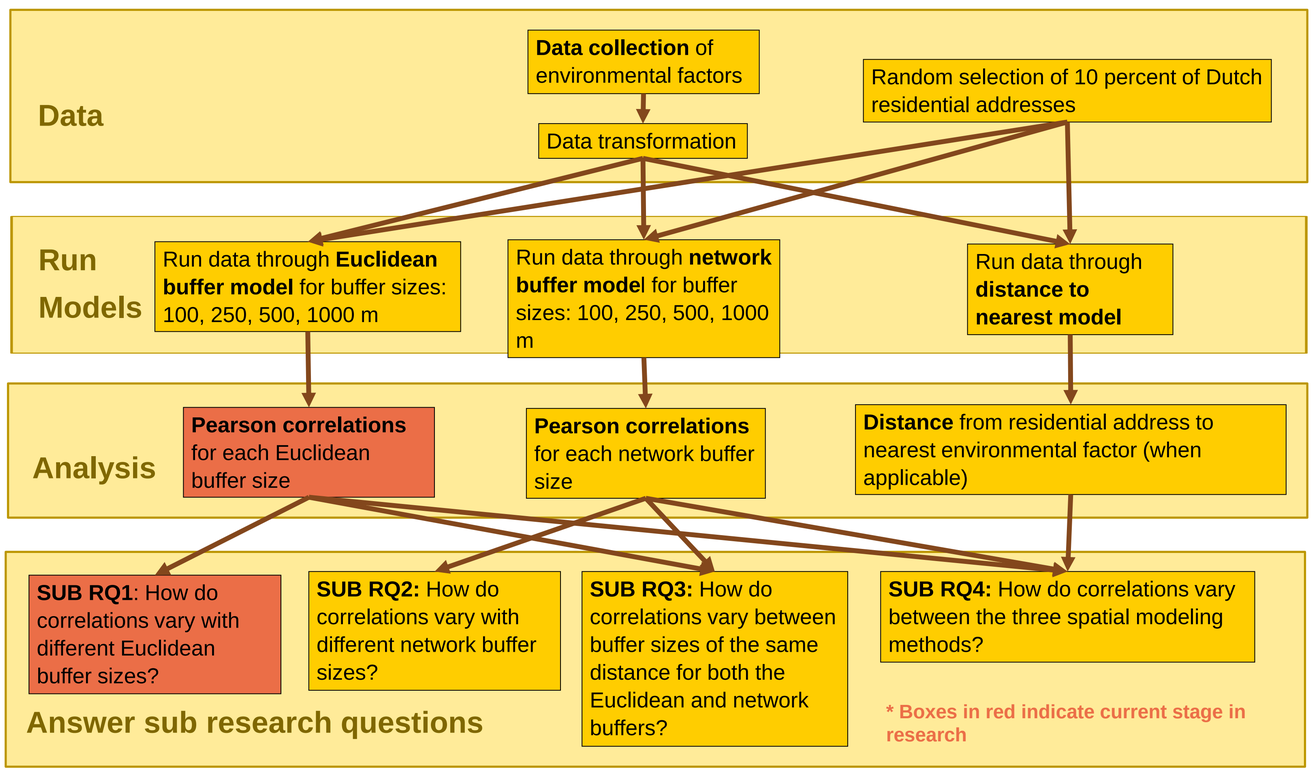


e) Object point
Primary schools. Source: DUO



f) Network
Main railroads (in blue). Source: OpenStreetMap

Summary of Methods



Preliminary Results

This table is a sample of the Pearson correlations (PC) of the environmental factor exposures found for Euclidean buffer sizes 100, 250, 500, and 1000 meters with the center as the residential address. The color scale shows the degree of change between the Pearson correlation values for each buffer size.

PC values in green indicate a positive correlation. PC values in red indicate a negative correlation. PC values in yellow indicate no correlation. The deeper the green/red, the stronger the positive/negative correlation.

variable1	variable2	100m_PCvalue	250m_PCvalue	500m_PCvalue	1000m_PCvalue
PM2.5	PM10	0.57	0.63	0.68	0.6
NO2	Ozone	-0.72	-0.74	-0.77	-0.8
PM2.5	UFP	0.16	0.18	0.18	0.6
PM2.5	BC	0.29	0.3	0.3	0.88
PM2.5	pm25_ni_slr	-0.01	-0.02	-0.06	-0.02
NO2	pm25_si_slr	0.68	0.02	0.01	0.04
pm25_cu_slr	pmc_fe_slr	0	0.98	0.99	1
pm25_cu_slr	temp_may_2019	0.56	0	-0.01	-0.01
NO2	AllSound	0.51	0.6	0.67	0.73
pm25_cu_slr	NightLightExpanse	0.59	0.01	0.01	0.02
NO2	cereals	-0.29	-0.41	-0.4	-0.59
PM2.5	fodder	-0.03	-0.06	-0.06	-0.15
cereals	grains	0.3	0.43	0.43	0.7
NO2	MSAVI_ME300	-0.76	-0.77	-0.79	-0.8
PM2.5	MSAVI_ME300	-0.19	-0.2	-0.22	-0.8
pm25_si_slr	NDVI_ME300	-0.56	0	0.03	0.08
NO2	Trees	-0.28	-0.33	-0.29	-0.15
NO2	Parks	0.02	0.07	0.15	0.26
pm25_si_slr	urbanity	-0.67	-0.01	0	0
PM2.5	popdensity	0.14	0.14	0.15	0.68
pmc_si_slr	Imp300m	0.55	0.01	0	0.01
urbanity	Imp300m	-0.56	-0.61	-0.7	-0.81
root_crops	SS85	-0.02	-0.03	-0.03	0.01
SS84	EduLow	0.54	0.53	0.5	0.46
EduLow	AvIncmPIncRec	-0.63	-0.63	-0.63	-0.63
temp_feb_2019	LowestIncmPcnt	,-0.13	-0.14	-0.15	-0.17
EduLow	Lvblty	-0.65	-0.65	-0.64	-0.63
NO2	SoCoh	-0.67	-0.7	-0.73	-0.76
PM2.5	SoCoh	-0.21	-0.22	-0.22	-0.76
Ozone	SoCoh	0.75	0.75	0.75	0.75
UFP	SoCoh	-0.55	-0.6	-0.56	-0.66
BC	SoCoh	-0.72	-0.74	-0.76	-0.78
temp_feb_2019	Facilities	0.69	0.69	0.7	0.71
pm25_cu_slr	T_dest_vio	0.43	0	0	0.02
NsnceIncrty	T_dest_vio	-0.57	-0.58	-0.59	-0.62
artificial_land	T_theft	0.15	0.24	0.26	0.49
Facilities	T_theft	0.52	0.53	0.54	0.56
Facilities	primaryschools	-0.08	-0.01	0.17	0.36
pm25_v_slr	MBO	0	0	0.01	0
MBO	col_and_uni	0.57	0.7	0.71	0.68
NO2	social_drinking	0	0.03	0.09	0.16
MSAVI_ME300	fastfood	0.02	-0.07	-0.15	-0.24
popdensity	fastfood	-0.03	0.14	0.28	0.4
primaryschools	superandminimarket	0.56	0.44	0.27	0.27

Preliminary Findings

These first results for sub research question one show that results vary widely. While some correlations remain consistent across different buffer sizes, other correlations contain a PC value higher or lower than other buffer sizes. This may be attributed to the notion that as buffer size increases, correlation increases. The spatial type of the environmental factor data may also influence the correlation. Other reasons unaccounted for may play a role as well. However, these findings indicate that buffer size can make a difference in magnitude of correlation.

How environmental factors correlate across different spatial measurement methods
Csilla Vámos, Tian Tian*, Simon Scheider*, Roel Vermeulen**
**Utrecht University*

exposomeNL

Utrecht University

Csilla Vámos
c.k.vamos@uu.nl