

Microclimate analysis in Basse Santa Su, The Gambia

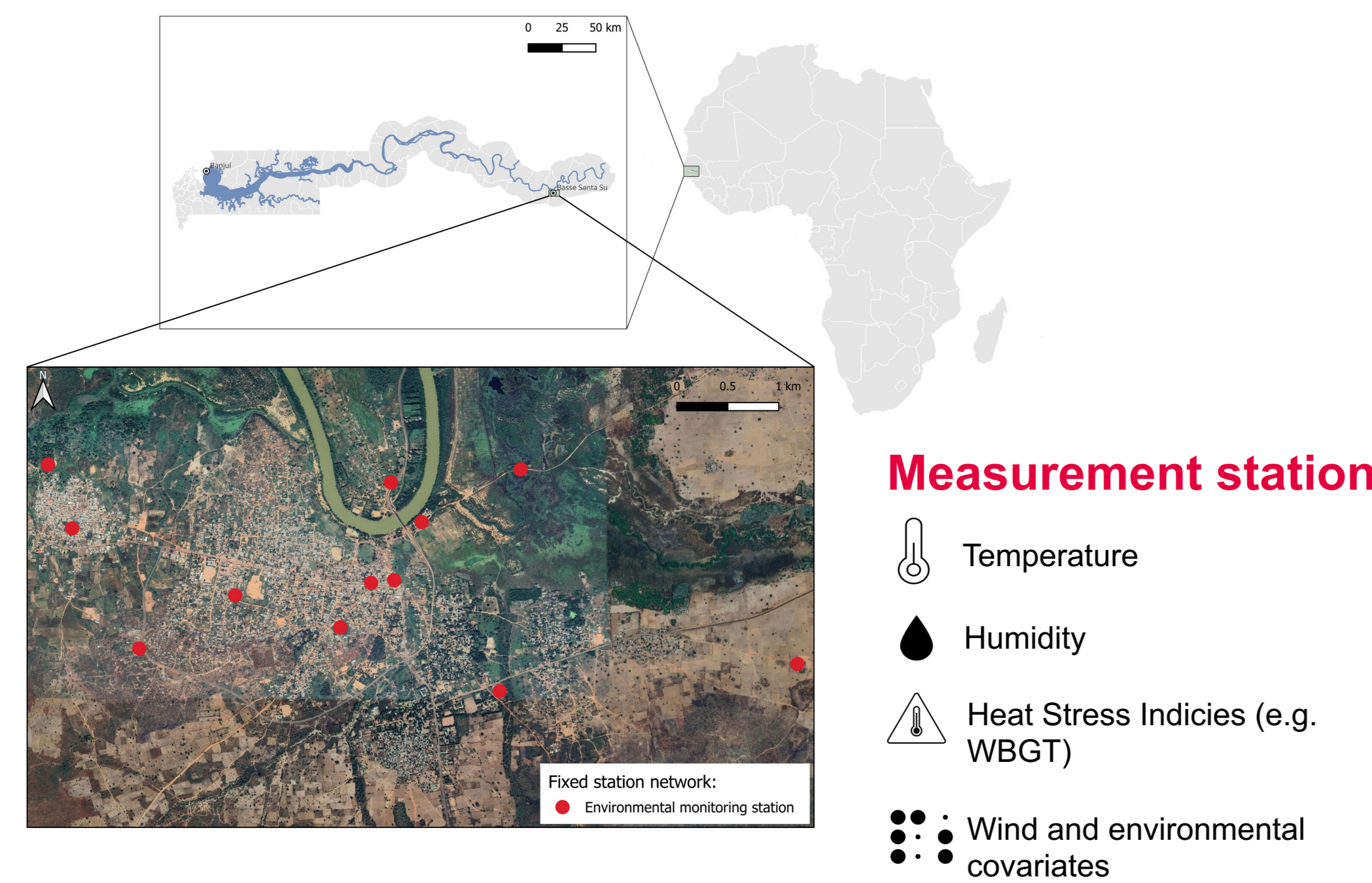
Elisabeth Tadiri (Elisabeth.Tadiri@unibe.ch)^{1,2}, Apolline Saucy^{1,2}, Ana Bonell^{3,4}, Moritz Burger^{2,5}, Moritz Gubler^{2,5,6}, Ana M. Vicedo-Cabrera^{1,2}

¹Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland, ²Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland, ³Medical Research Council Unit The Gambia at the London School of Hygiene & Tropical Medicine, Fajara, The Gambia, ⁴Centre on Climate Change and Planetary Health, London School of Hygiene & Tropical Medicine Faculty of Epidemiology and Population Health, London, England, United Kingdom, ⁵Institute of Geography, University of Bern, Bern, Switzerland, ⁶Institute for Research, Development and Evaluation, Bern University of Teacher Education, Bern, Switzerland

Motivation

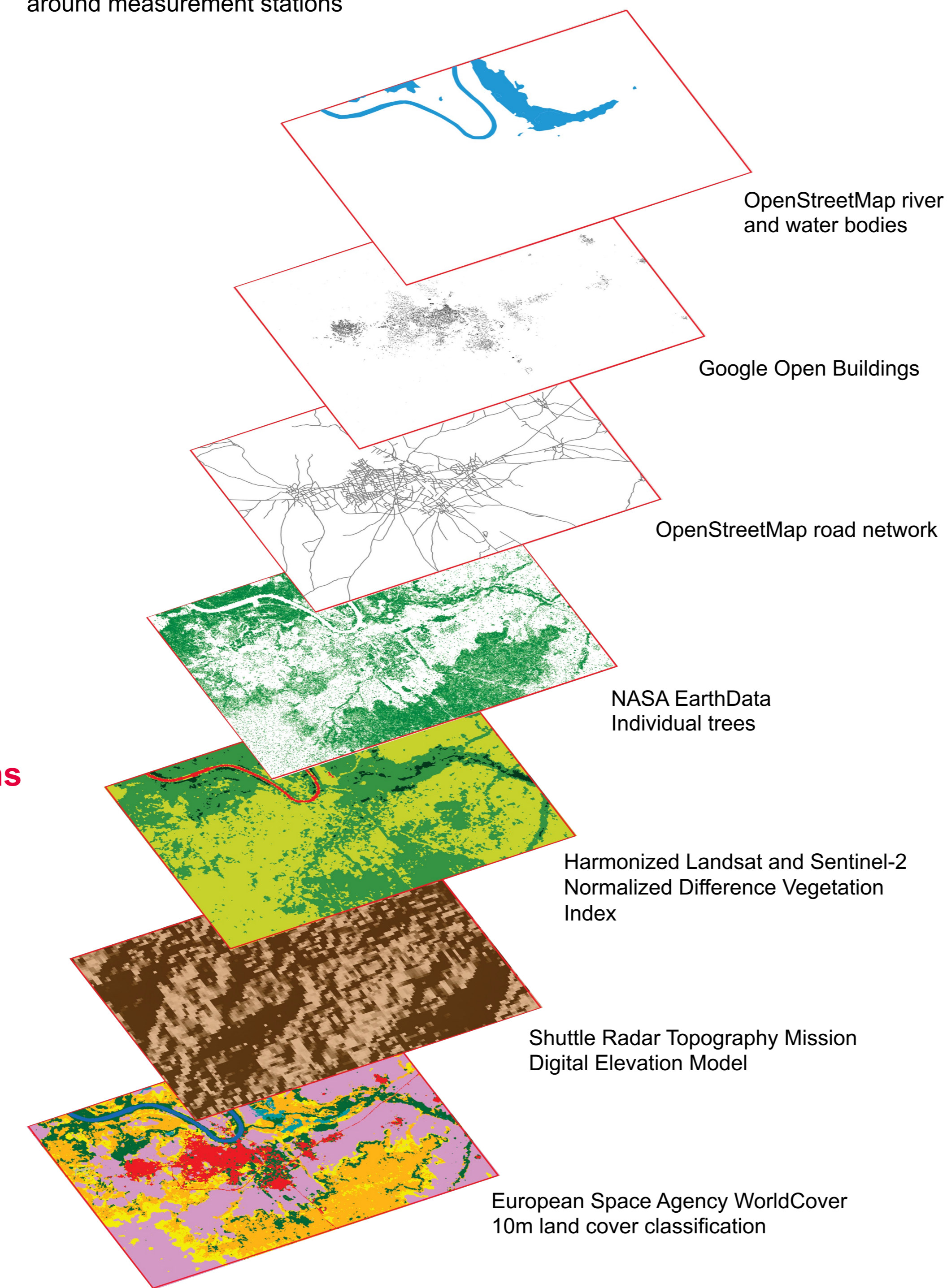
- Basse Santa Su experiences **extreme humid heat conditions**, but fine-scale, local meteorological data is lacking
- Bridge gap between coarse regional meteorological data, and local variable conditions to **identify localized urban heat stress** patterns and conditions potentially affecting human health

Data and Methods



Satellite-derived data

Predictors within 25m-500m buffers around measurement stations



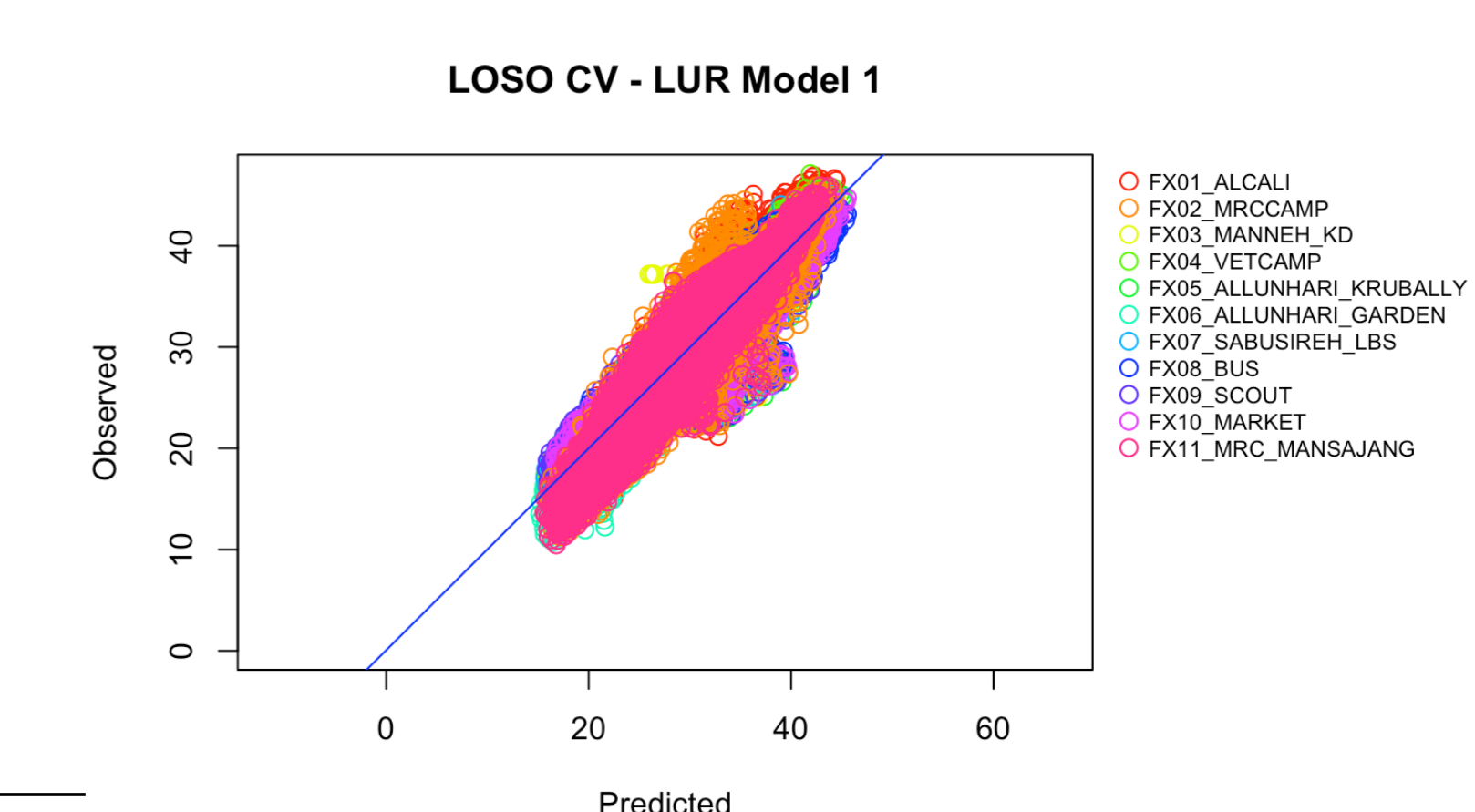
Model comparison

Base temperature model:

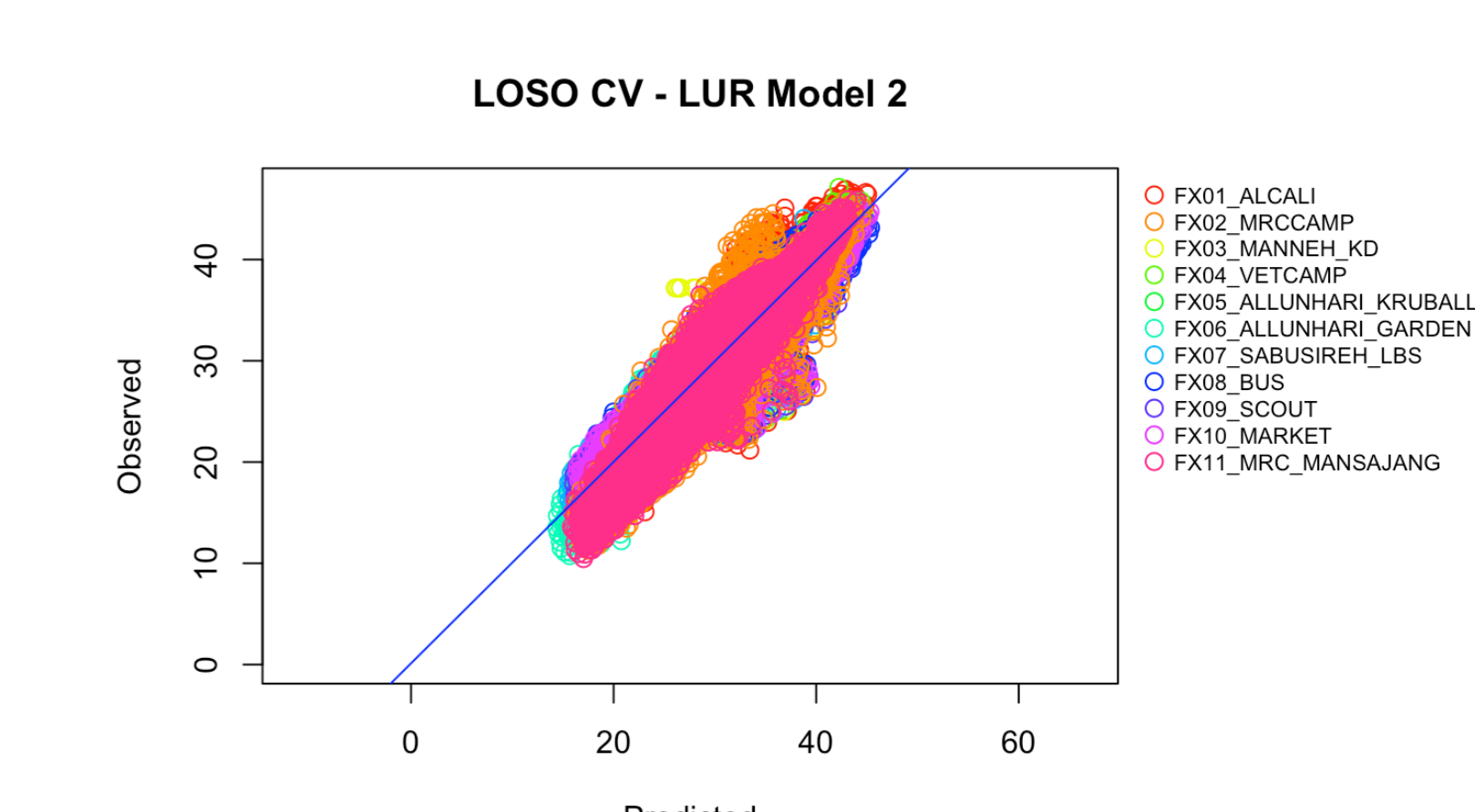
$$temp_hourly_i = \beta_0 + \beta_1 \cdot era5land_temp + S_1(hod) + S_2(doy) + \epsilon_i$$

Adjusted $R^2 = 0.882$, RMSE = 2.067

Variable	Coefficient	Standard Error	p value	Adj. VIF	Adj. R^2	RMSE	LOSO CV R^2	LOSO CV RMSE
1. LUR Model - Forward selection								
(Intercept)	7.122	0.098	<0.001		0.893	1.970	0.886	2.034
Era5-Land hourly 2m temperature	0.943	0.003	<0.001	2.213				
pbs(hour of day, df = 3)1	-5.277	0.059	<0.001					
pbs(hour of day, df = 3)2	-5.683	0.026	<0.001					
pbs(hour of day, df = 3)3	-7.239	0.058	<0.001					
pbs(day of year, df = 3)1	3.146	0.059	<0.001					
pbs(day of year, df = 3)2	-2.404	0.038	<0.001					
pbs(day of year, df = 3)3	1.175	0.069	<0.001					
Number of trees in 100m	-0.011	0.000	<0.001	1.559				
Mean NDVI in 25m	-1.775	0.086	<0.001	2.319				
Percent grassland cover in 50m	-0.038	0.002	<0.001	1.337				
Percent bare/sparse vegetation cover in 100m	0.417	0.021	<0.001	1.18				
Length of paved roads (m) in 150m	0.001	0.000	<0.001	1.394				
Percent built-up cover in 500m	0.004	0.000	<0.001	1.502				



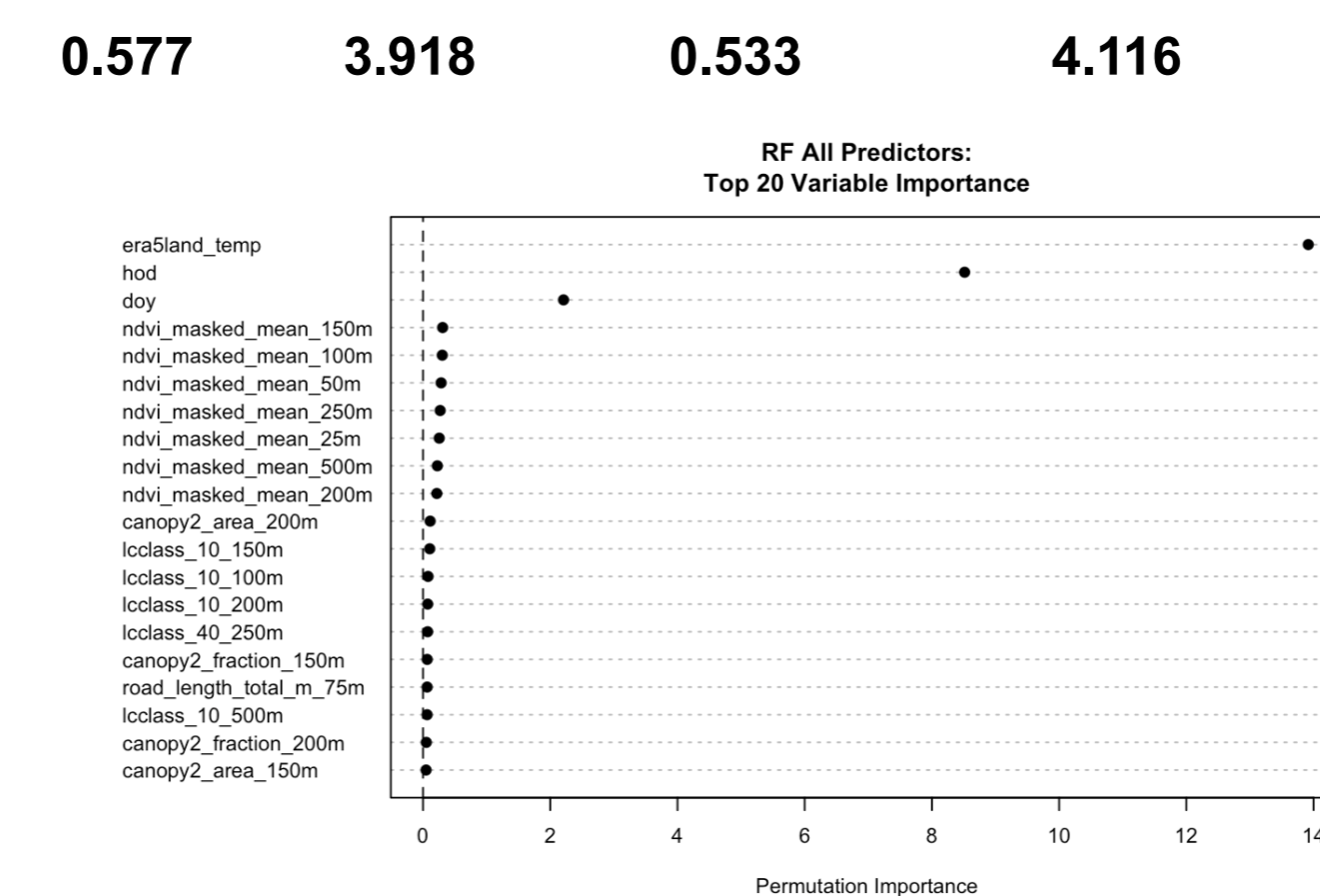
Variable	Coefficient	Standard Error	p value	Adj. VIF	Adj. R^2	RMSE	LOSO CV R^2	LOSO CV RMSE
2. LUR Model - Domain Selection								
(Intercept)	9.094	0.113	<0.001		0.894	1.966	0.886	2.033
Era5-Land hourly 2m temperature	0.943	0.003	<0.001	2.212				
pbs(hour of day, df = 3)1	-5.275	0.059	<0.001					
pbs(hour of day, df = 3)2	-5.683	0.026	<0.001					
pbs(hour of day, df = 3)3	-7.241	0.058	<0.001					
pbs(day of year, df = 3)1	3.144	0.059	<0.001					
pbs(day of year, df = 3)2	-2.416	0.038	<0.001					
pbs(day of year, df = 3)3	1.166	0.069	<0.001					
Number of trees in 100m	-0.024	0.000	<0.001	2.046				
Mean NDVI in 25m	-1.813	0.083	<0.001	2.247				
Distance from the river (m)	0.000	0.000	<0.001	1.719				
Percent permanent water bodies cover in 100m	-0.069	0.003	<0.001	1.236				
Length of paved roads (m) in 500m	0.000	0.000	<0.001	2.081				
Elevation (m)	-0.009	0.001	<0.001	1.574				
Percent building coverage in 50m	-0.679	0.078	<0.001	1.846				



	OOB R^2	OOB RMSE	LOSO CV R^2	LOSO CV RMSE
--	-----------	----------	---------------	--------------

3. RF Model - All Predictors

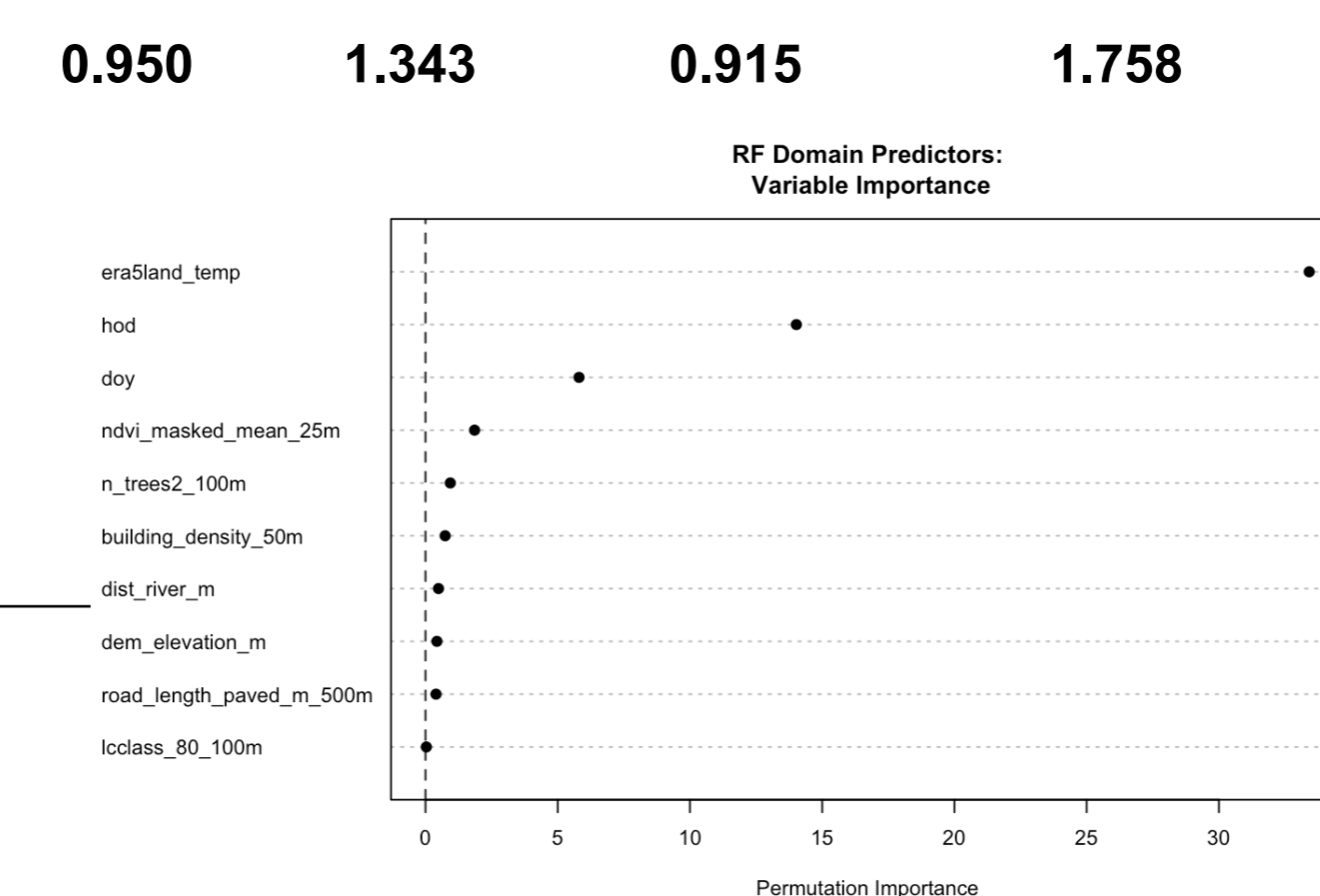
Type: Regression
Number of trees: 500
Sample size: 111947
Number of independent variables: 163
Mtry: 12
Target node size: 5
Variable importance mode: permutation
Spliitrule: variance



	OOB R^2	OOB RMSE	LOSO CV R^2	LOSO CV RMSE
--	-----------	----------	---------------	--------------

4. RF Model - Domain Predictors

Type: Regression
Number of trees: 500
Sample size: 111947
Number of independent variables: 10
Mtry: 3
Target node size: 5
Variable importance mode: permutation
Spliitrule: variance



Modelling approach

- **Multilinear land-use regression (LUR)** modelling of temperature to maximize the percentage of explained variability (R^2)
 1. Forward selection strategy from all predictors, selecting for expected direction of relationship and additional R^2 threshold of 0.0001
 2. Domain-restricted selection strategy – max R^2 from each domain
- **Random forest (RF)** regression modelling with 500 trees
 3. All predictors
 4. Domain-restricted predictors
- **Leave-one-station-out spatial cross-validation (LOSO CV)** to evaluate model performances

Next steps and questions

- High-resolution maps of heat stress across study area to identify **differential spatial distributions of heat stress**
- Investigate relationship between **individual heat stress exposure** and observed **physiological heat strain** in healthy adults
- Model and predict each **input** for heat stress index separately, then calculate heat stress index from predictions, **or** calculate **heat stress index**, then model and predict heat stress index?
- **Traditional LUR vs ML methods** – what is the added value of ML?