Modelling anomalies in the land surface phenology of Europe using inter-annual variations in climatic drivers

Maria Villarin-Claveria[1], Victor Rodriguez-Galiano[2,3]

1 Human Geography, University of Seville, Seville, Spain

2 Physical Geography, University of Seville, Seville, Spain

3 Geography and Environment, University of Southampton, Southampton, United Kingdom

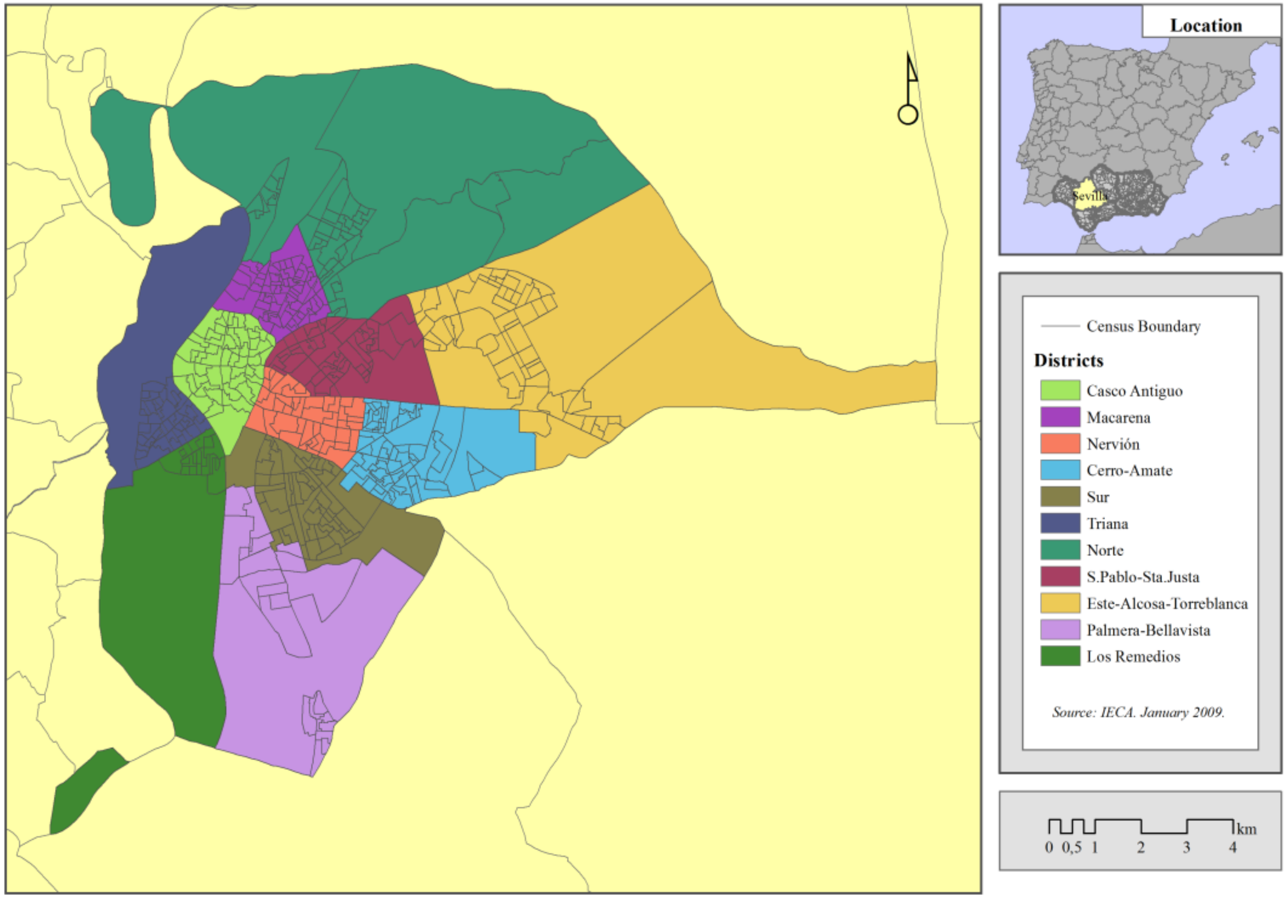
Objectives

I. To assess regression trees for modelling water demand.

II. To reveal new insights into the socio-demographic and building-urban drivers of water demand in Sevilla city at a census track level.

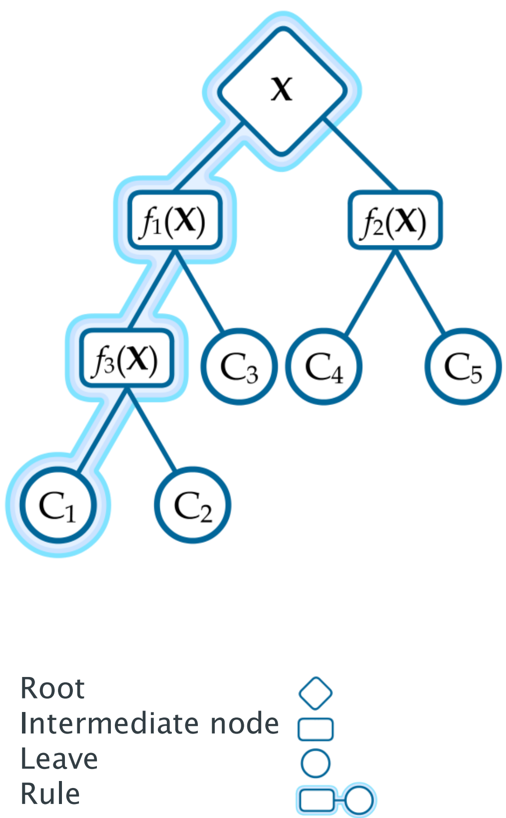
III. To compare the performance and graphical outputs of CART Regression trees and Random Forest.

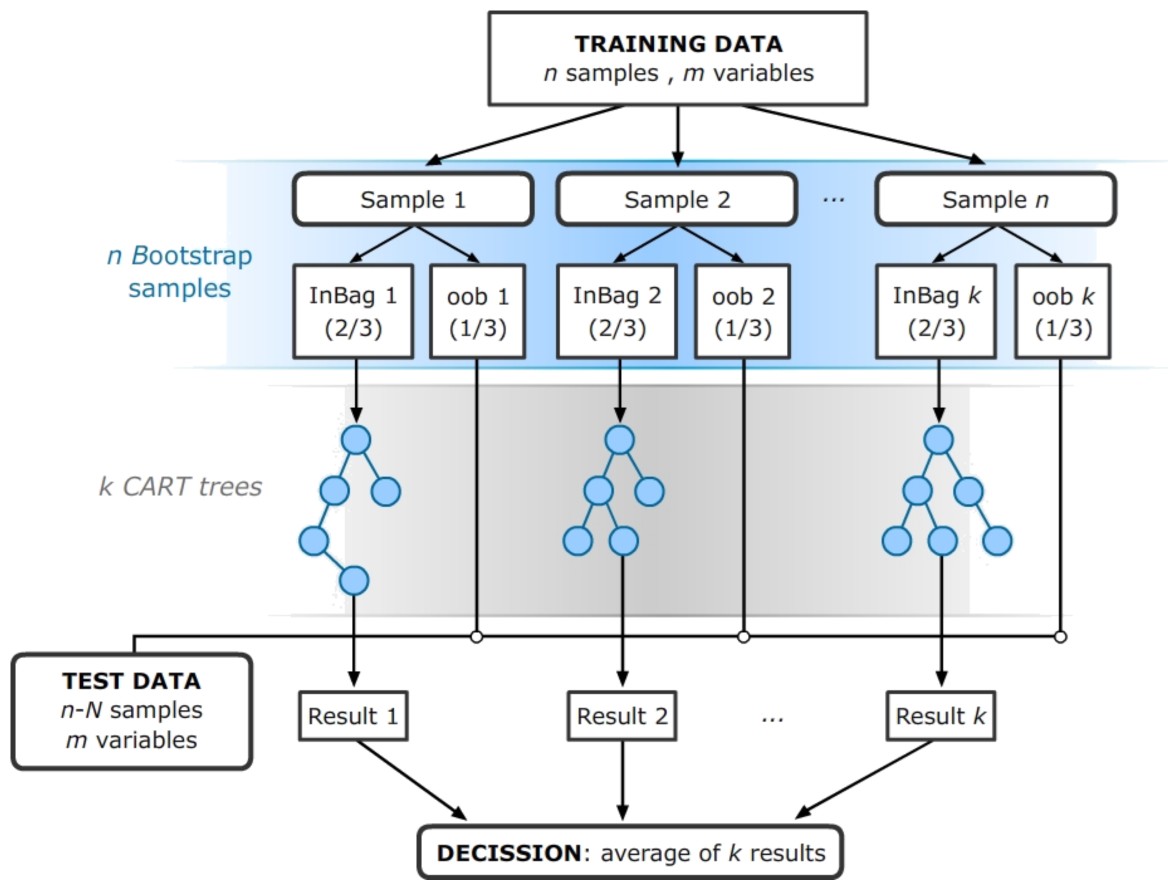
Study area



Census tract is a geographic region defined for the purpose of taking a census.
Census tracts represent the smallest territorial unit for which population data are available in many countries

Modelling methods: Regression trees and Random Forest





Why Regression Trees?

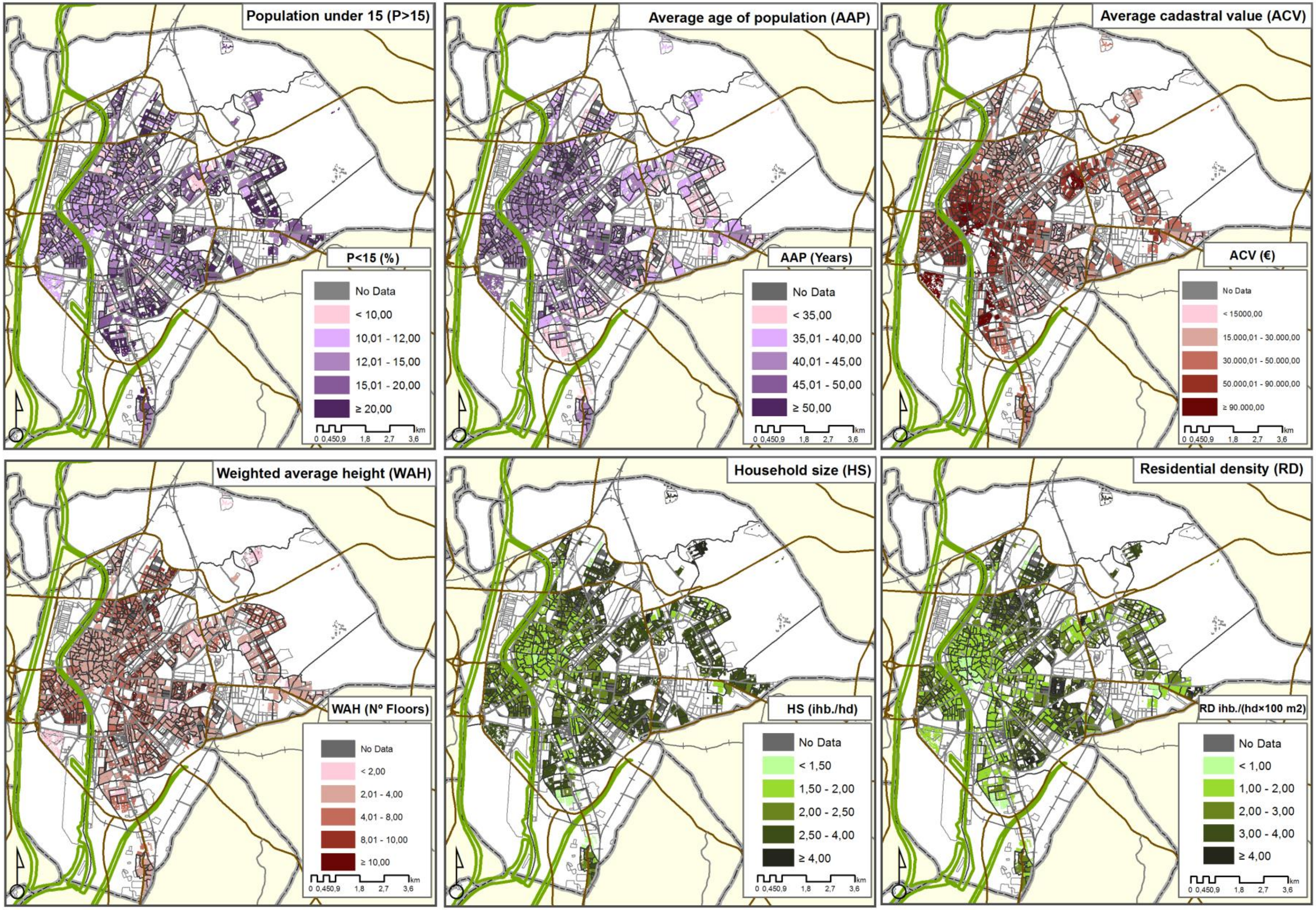
- Ability to learn complex patterns, considering nonlinear relationships between explanatory and dependent variables.
- Generalisation ability, hence applicable to incomplete or noisy databases.
- Integration of different types of data in the analysis due to the absence of assumptions about the data used (e.g. normality).
- Interpretability of results, since RT allows obtaining patterns for a better explanation of a given phenomenon, showing the most important variables and their threshold values.

Explanatory variables (I)

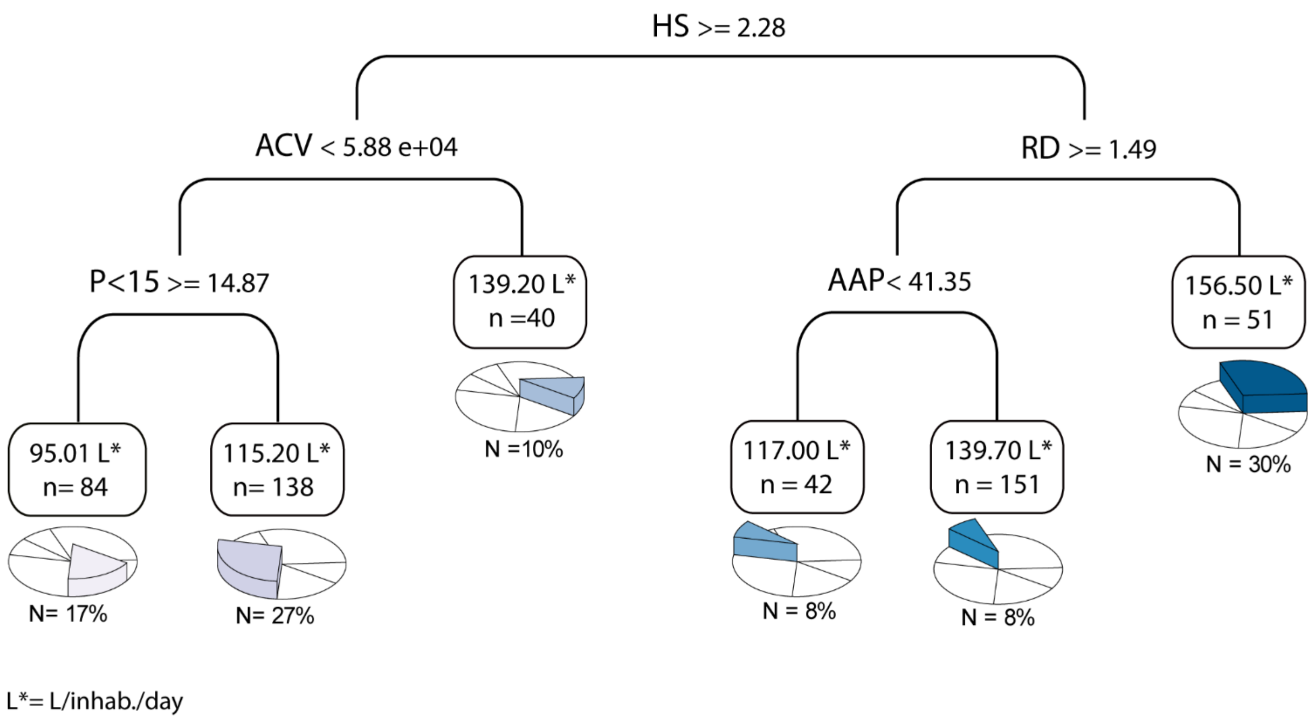
Explanatory variable	Units
Domestic water consumption (DC)	L/(inhab. day)
Population under 15 (P<15)	%
Population between 15 and 34 (P1534)	%
Population between 35 and 64 (P3564)	%
Population over 65 (P>65)	%
Youth index (YI)	%
Aging index (AI)	%
Foreigners (FRG)	%
Average age of population (AAP)	years
Average cadastral value (ACV)	€
Average built surface area (ABSA)	m^2
Weighted average height (WAH)	number of floors
Average gross density (AGD)	inhab./m^2
Average net density (AND)	inhab./m^2 (constructed)
Household size (HS)	inhab./household
Residential density (RD)	inhab./(household $\times$ 100m^2)

P<15	P1534	P3564	P>65	AAP	FRG	YI	AI	WAH	ACV	ABSA	AGD	AND	HS	RD
-0.398**	-0.207**	0.04	0.360**	0.453**	**0.93***	-0.287**	0.369**	0.132**	0.420**	**0.197****	0.02	-0.308**	-0.329**	-0.479**

Explanatory variables (II)



Regression Tree (RT) Model.

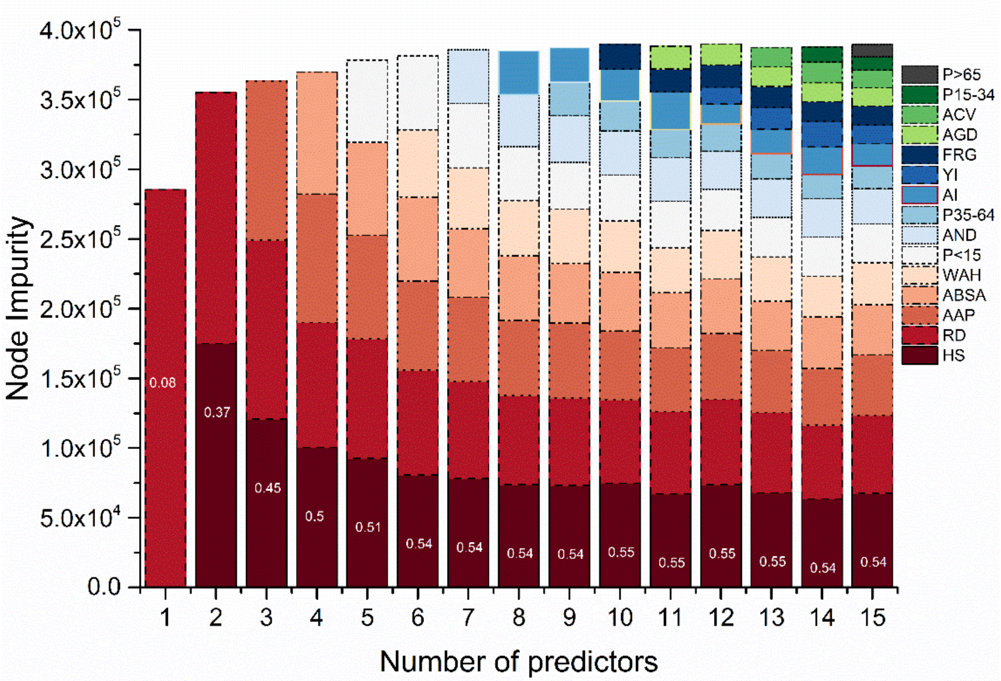


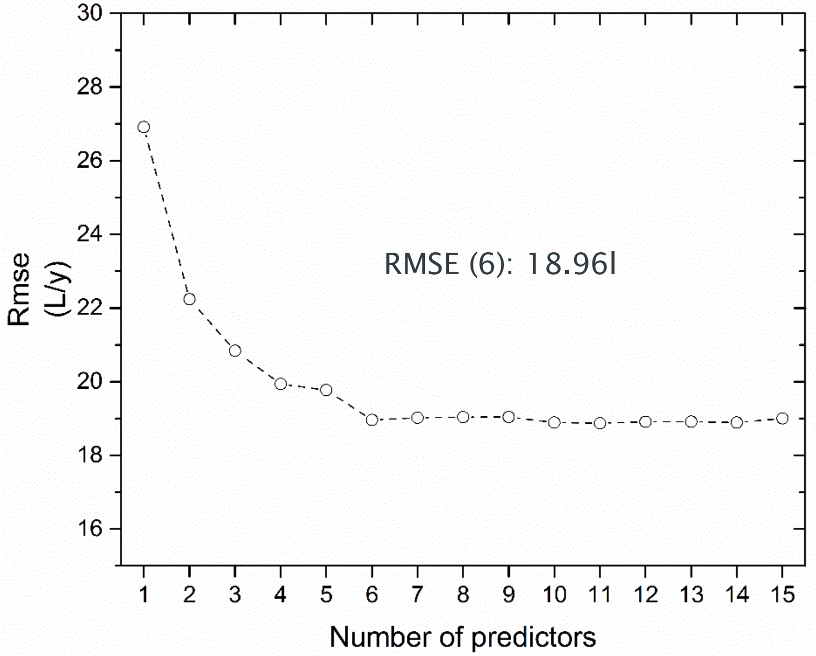
RMSE: 20.70l
R^2: 0.46

HS	RD	AAP	YI	P>65	ACV	P1534	P<15	ABSA	AI	AND
20	14	13	11	10	8	6	6	5	5	1

Household size (HS); Average cadastral value (ACV); Residential density (RD); Population under 15 (P<15); Average age of population (AAP)

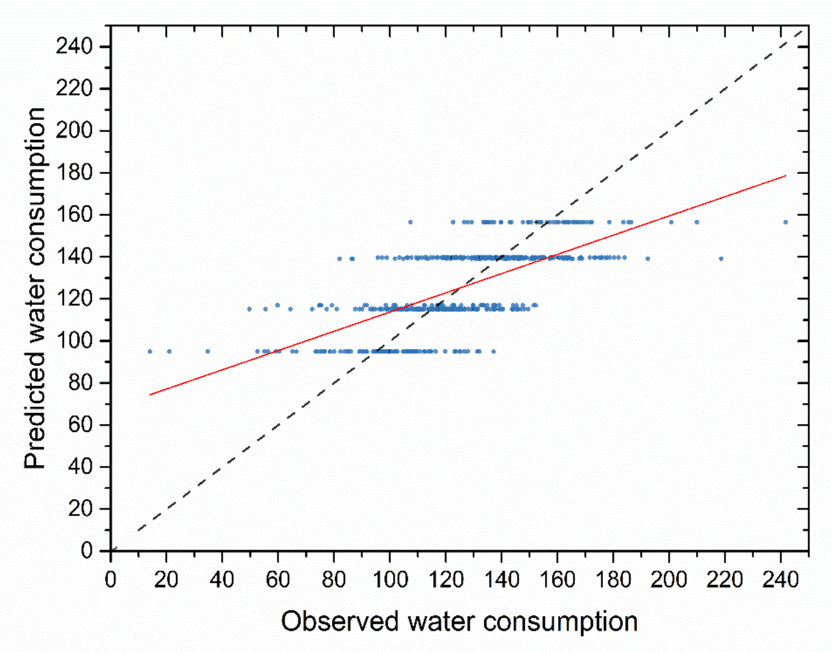
Random Forest(RF) Model.

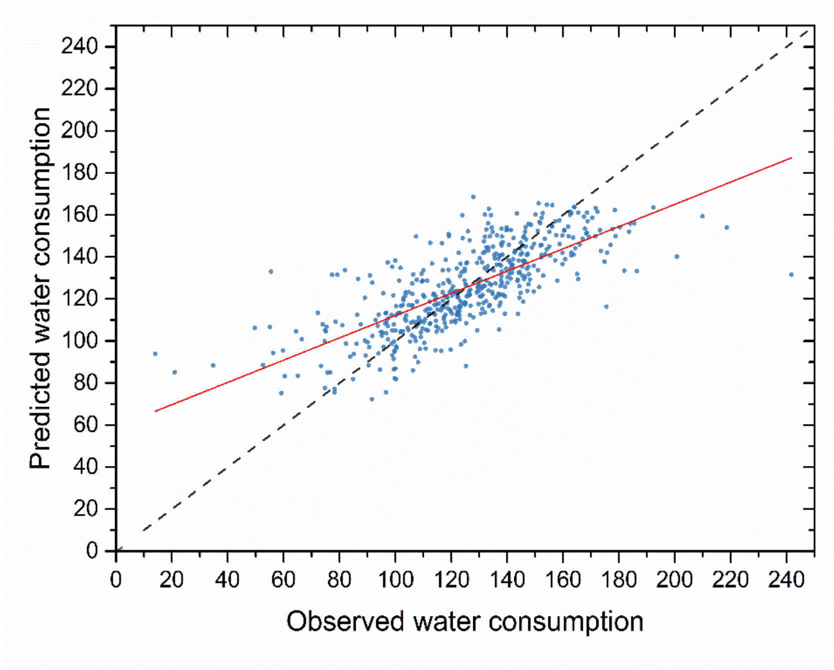




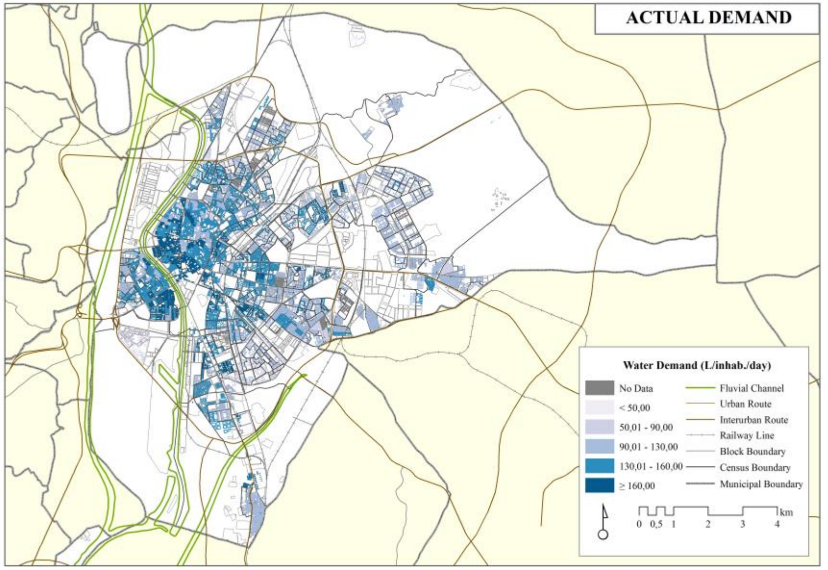
Household size (HS); Residential density (RD); Average age of population (AAP); Average built surface area (ABSA); Weighted average height (WAH); Population under 15 (P<15);

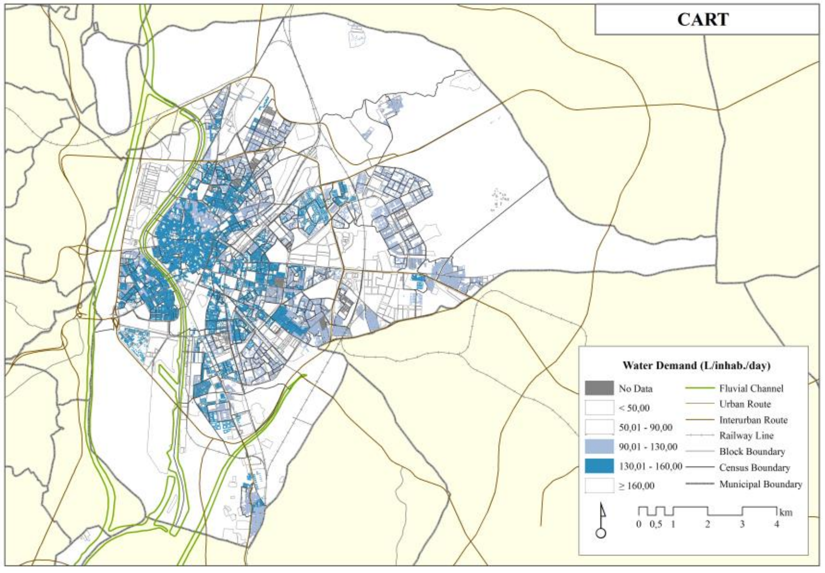
CART vs Random Forest (I)

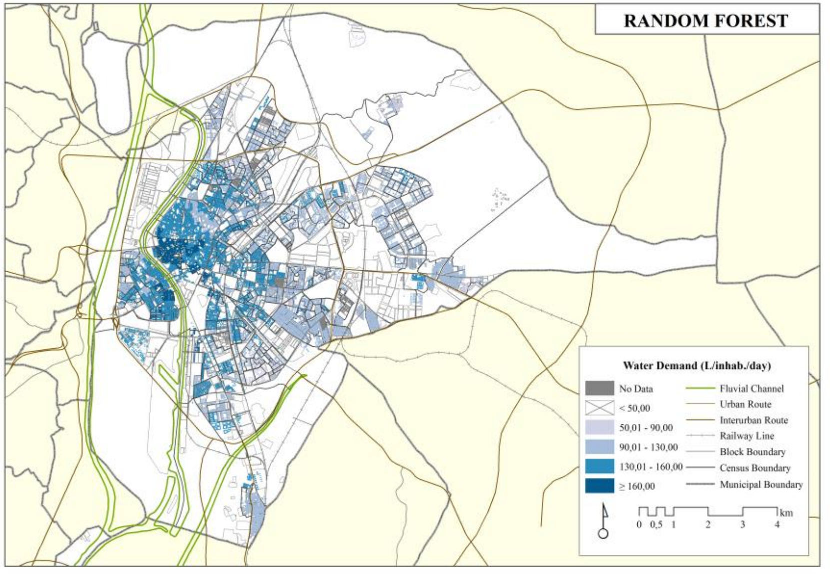




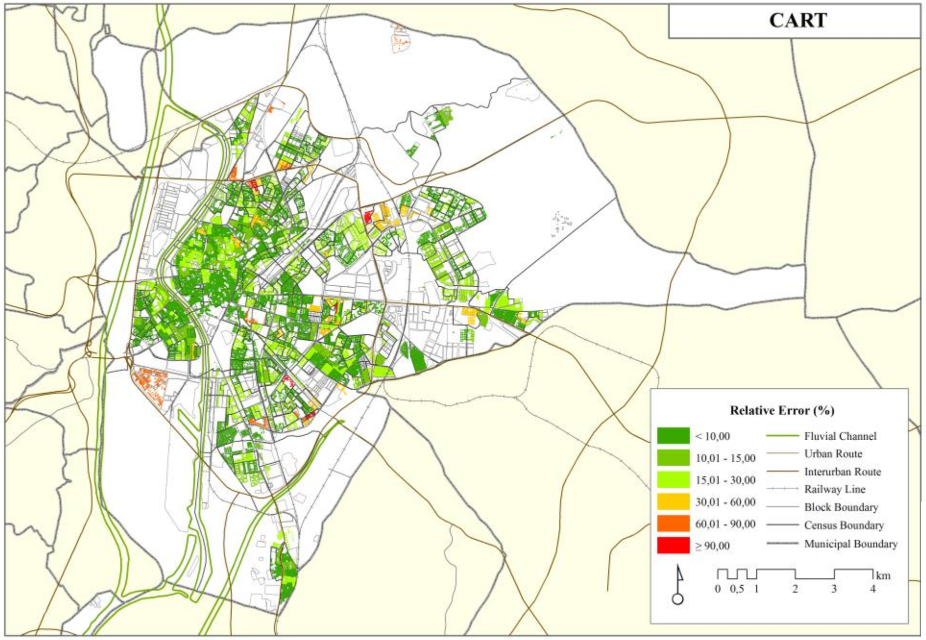
CART vs Random Forest (II)

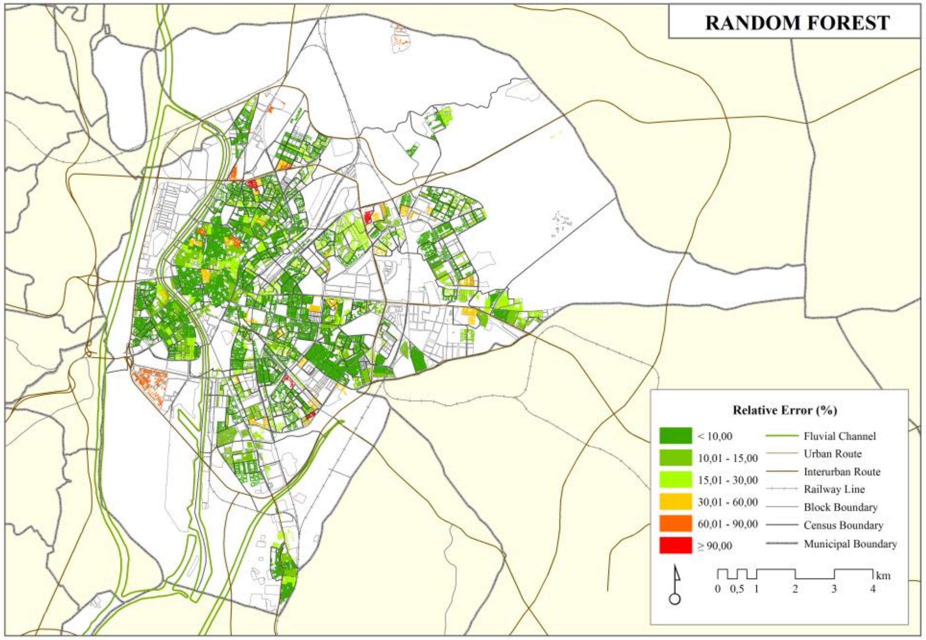






CART vs Random Forest (III)





Conclusions

I. Domestic water consumption in Sevilla city is driven mainly by the household size and residential density, and to a lower extent by the age of population, and built surface area.

II. The relations between water consumption and the socio-demographic and building-urban predictors are non-linear.

III. The regression tree models (both CART and RF), multivariate, spatially non-stationary and non-linear machine learning approach, outperformed multivariate linear models.

IV. RF showed a better performance than CART trees, explaining 54% of the variance with a RMSE value equal to 18.96 l.