Combining KDE and DBSCAN clustering to understand road traffic accidents: The case of Setúbal, Portugal.

Pedro Nogueira, Marcelo Silva, Paulo Infante, Paulo Rebelo Manuel, Leonor Rego, Anabela Afonso and Gonçalo Jacinto
marcelogs@uevora.pt

MOPREVIS

UNIVERSIDADE DE ÉVORA

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

GNR GUARDA NACIONAL REPUBLICANA

Road Traffic Accidents

- What
- Why
- Where

Methods

- Kernel Density Estimation
- Density Based Spatial Clustering
- DBSCAN based on KDE

Conclusions

- What is observed
- How can it be useful

Road traffic accidents

Contextualization

Why study RTA?

20 June 2022

WHO Global status report on road safety

Key facts

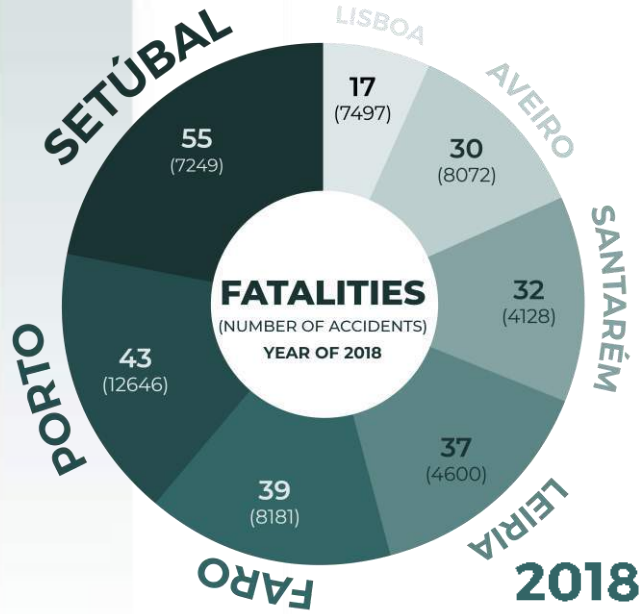
- The United Nations General Assembly has set an ambitious target of halving the global number of deaths and injuries from road traffic crashes by 2030 (A/RES /74/299).
- Road traffic injuries are the leading cause of death for children and young adults aged 5-29 years.
- Approximately 1.3 million people die each year as a result of road traffic crashes.
- More than half of all road traffic deaths are among vulnerable road users: pedestrians, cyclists, and motorcyclists.
- 93% of the world's fatalities on the roads occur in low- and middle-income countries, even though these countries have approximately 60% of the world's vehicles.
- Road traffic crashes cost most countries 3% of their gross domestic product.

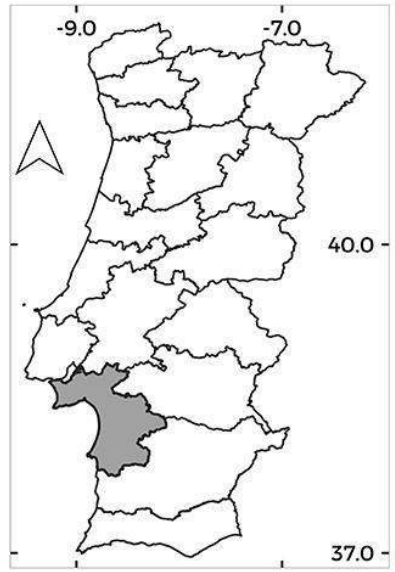
Geographical setting

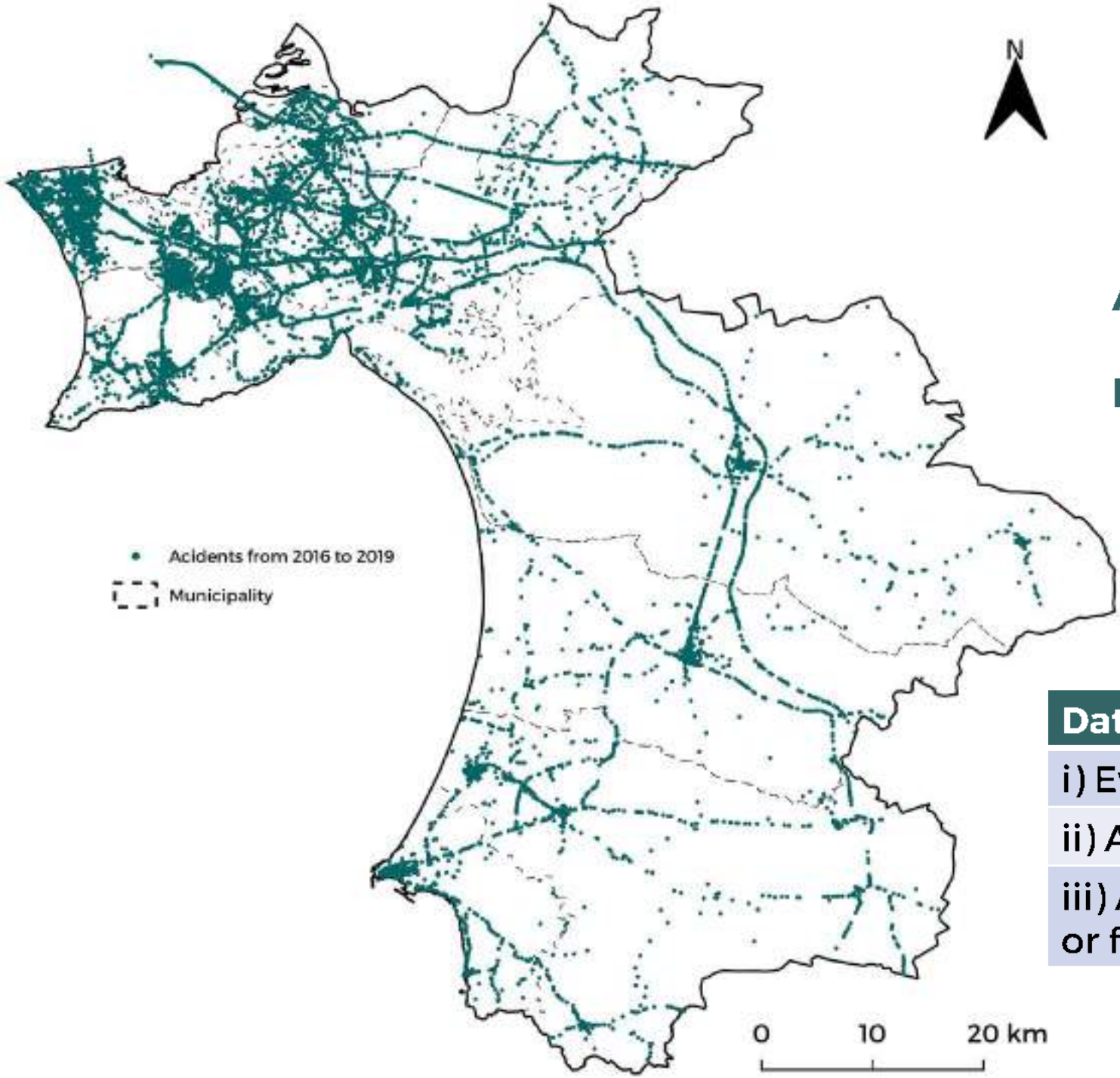
Portugal

Setúbal district

2 main zones: Urban and Rural







Assimetric distribution

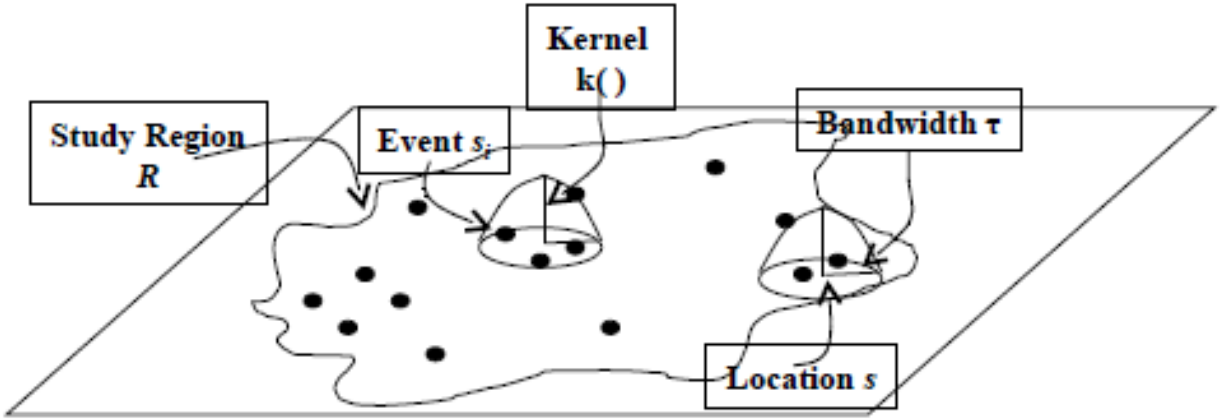
Follow road network

Dataset	Number
i) Every accident	27847
ii) Accidents with victims	5970
iii) Accidents with serious or fatal injuries	570

RTA distribution

Kernel Density Estimation

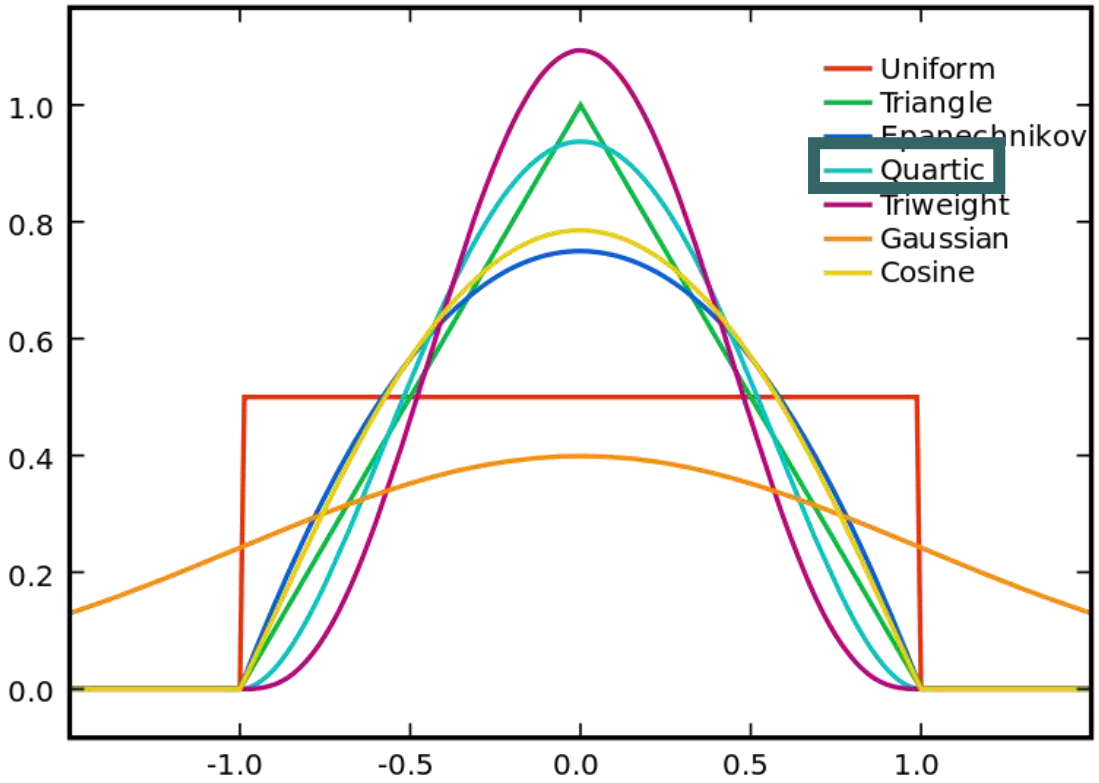
Kernel Density Estimation



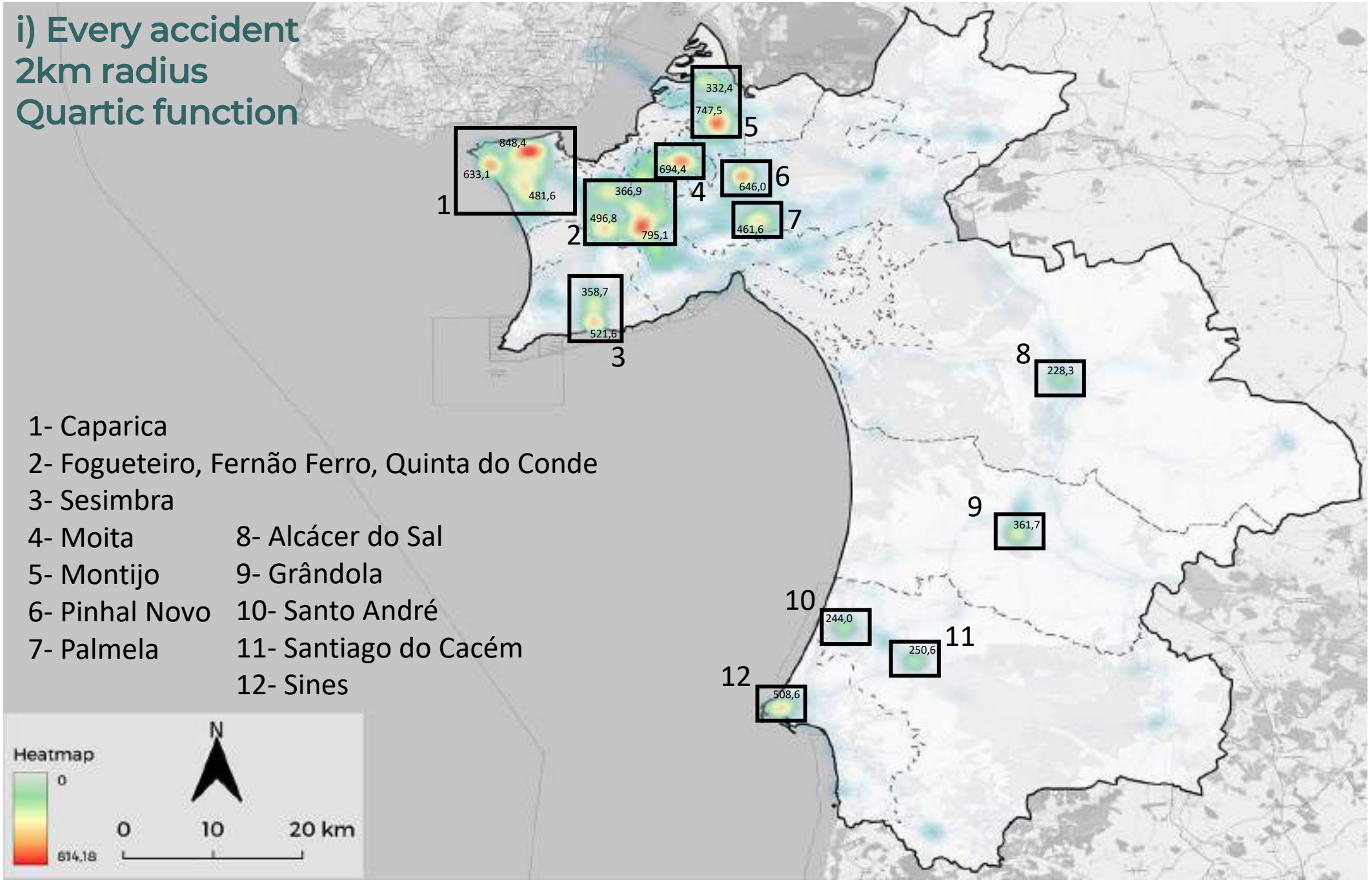
Occurs within study region;

Defines kernels based on events with given bandwidth;

Density is related to number of events in the location.

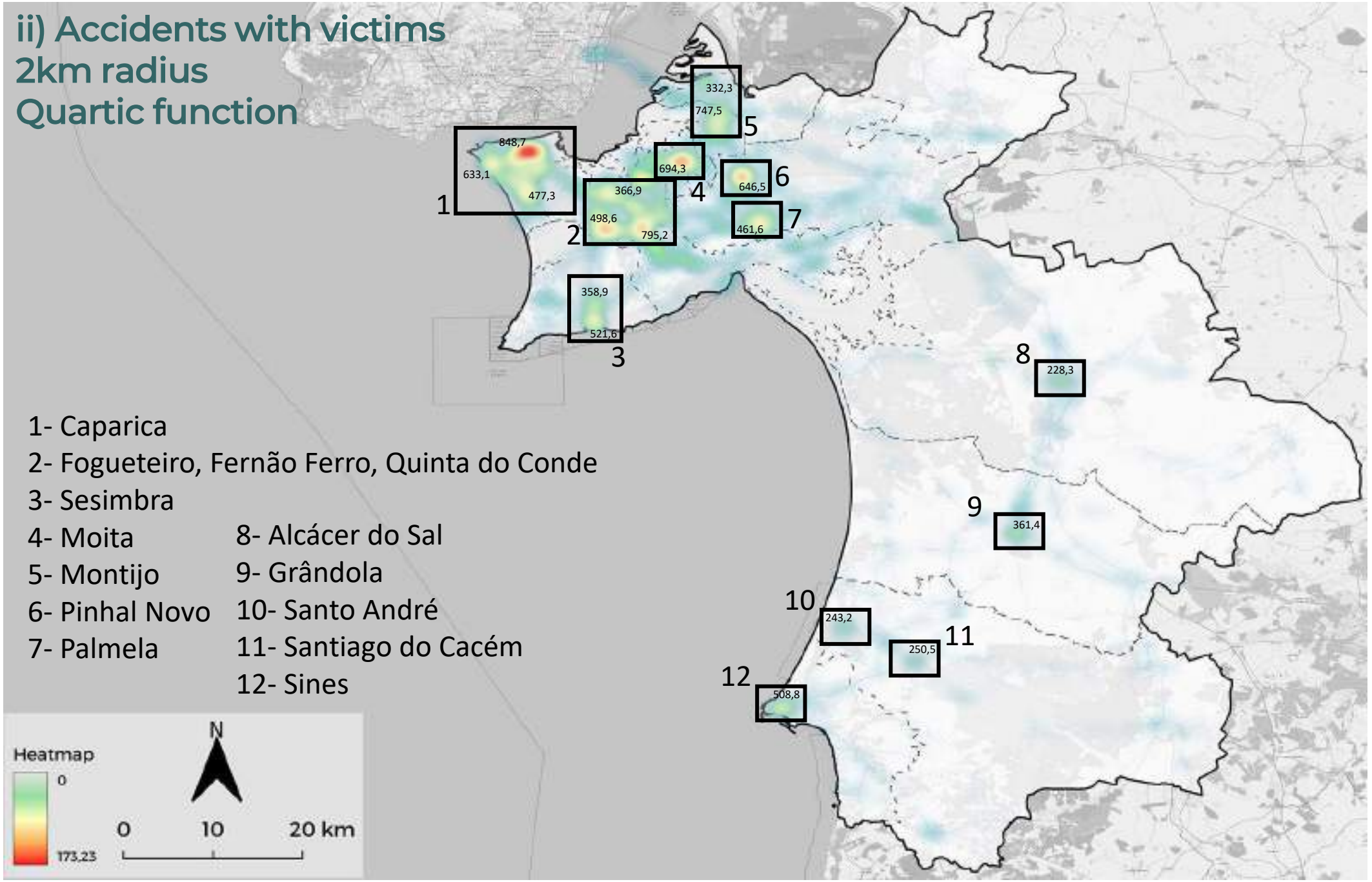


i) Every accident
2km radius
Quartic function



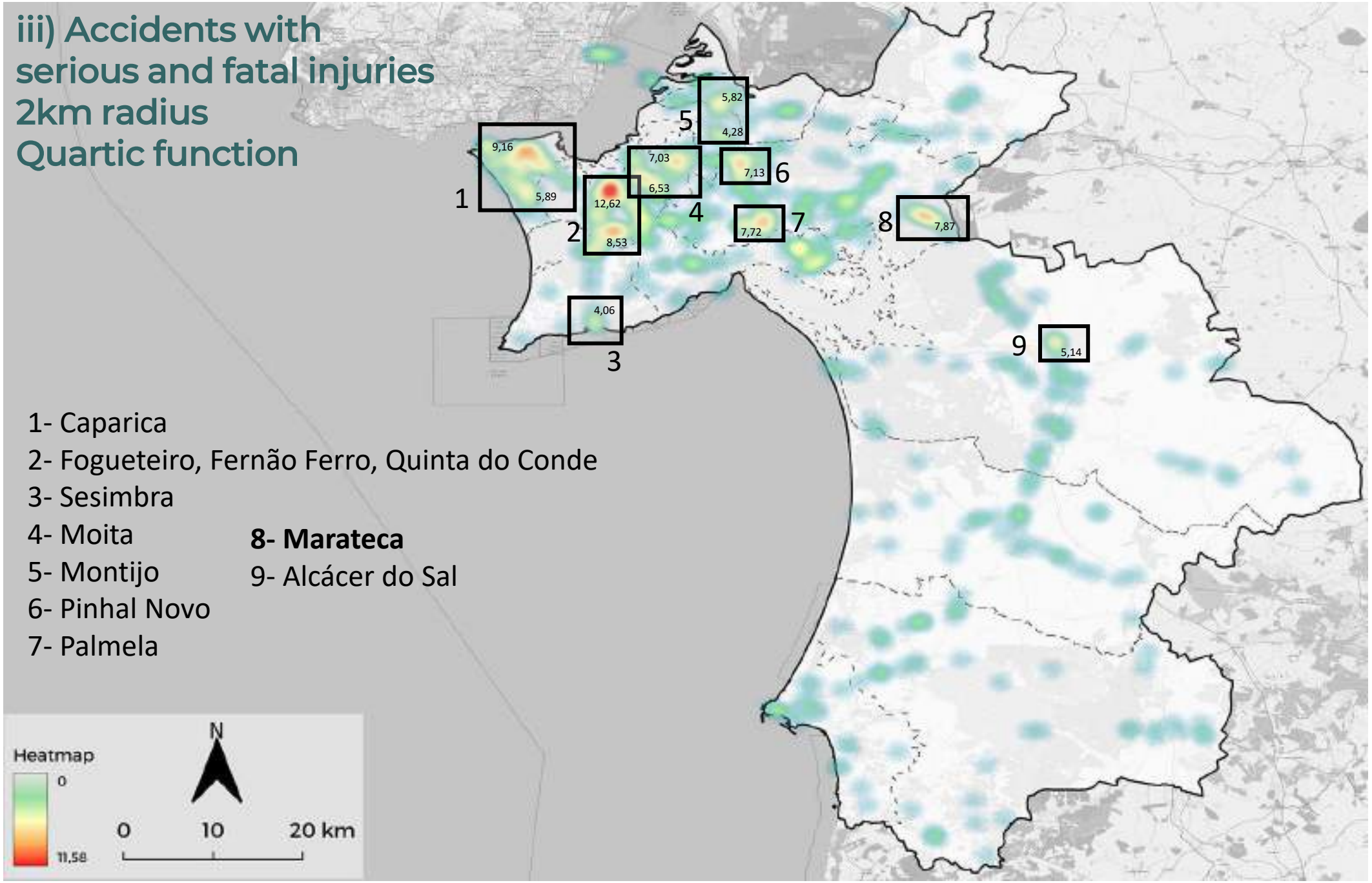
1- Caparica
2- Fogueteiro, Fernão Ferro, Quinta do Conde
3- Sesimbra
4- Moita
5- Montijo
6- Pinhal Novo
7- Palmela
8- Alcácer do Sal
9- Grândola
10- Santo André
11- Santiago do Cacém
12- Sines

ii) Accidents with victims
2km radius
Quartic function



1- Caparica
2- Fogueteiro, Fernão Ferro, Quinta do Conde
3- Sesimbra
4- Moita
5- Montijo
6- Pinhal Novo
7- Palmela
8- Alcácer do Sal
9- Grândola
10- Santo André
11- Santiago do Cacém
12- Sines

iii) Accidents with serious and fatal injuries 2km radius Quartic function



1- Caparica
2- Fogueteiro, Fernão Ferro, Quinta do Conde
3- Sesimbra
4- Moita
5- Montijo
6- Pinhal Novo
7- Palmela

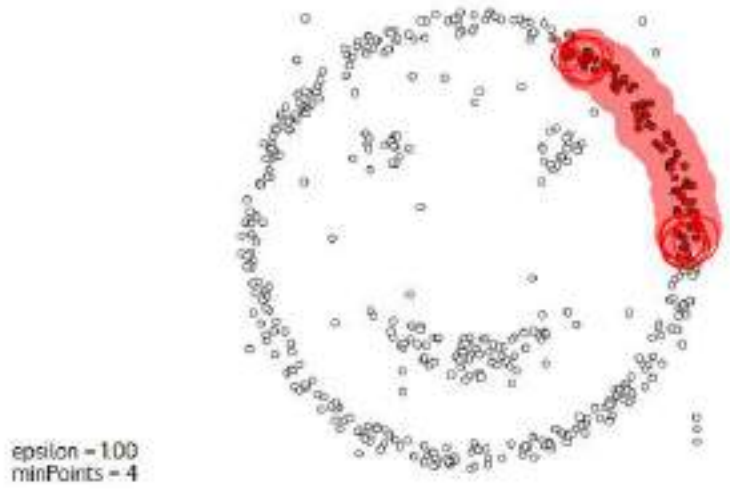
8- Marateca

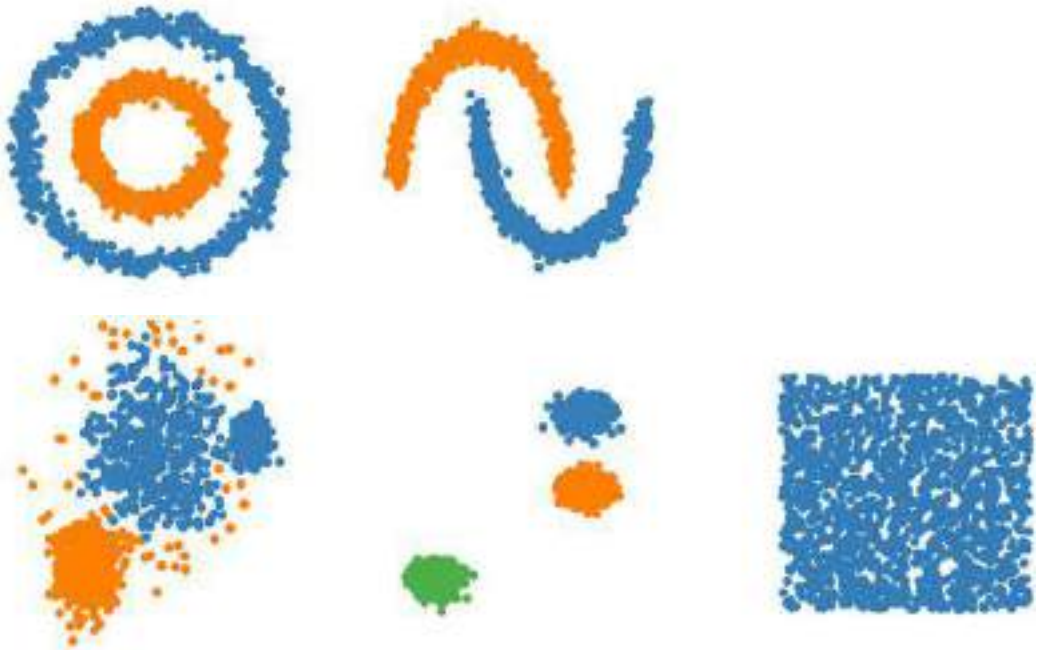
9- Alcácer do Sal

RTA distribution

Density Based Clustering

Density Based Spatial Clustering of Applications with Noise



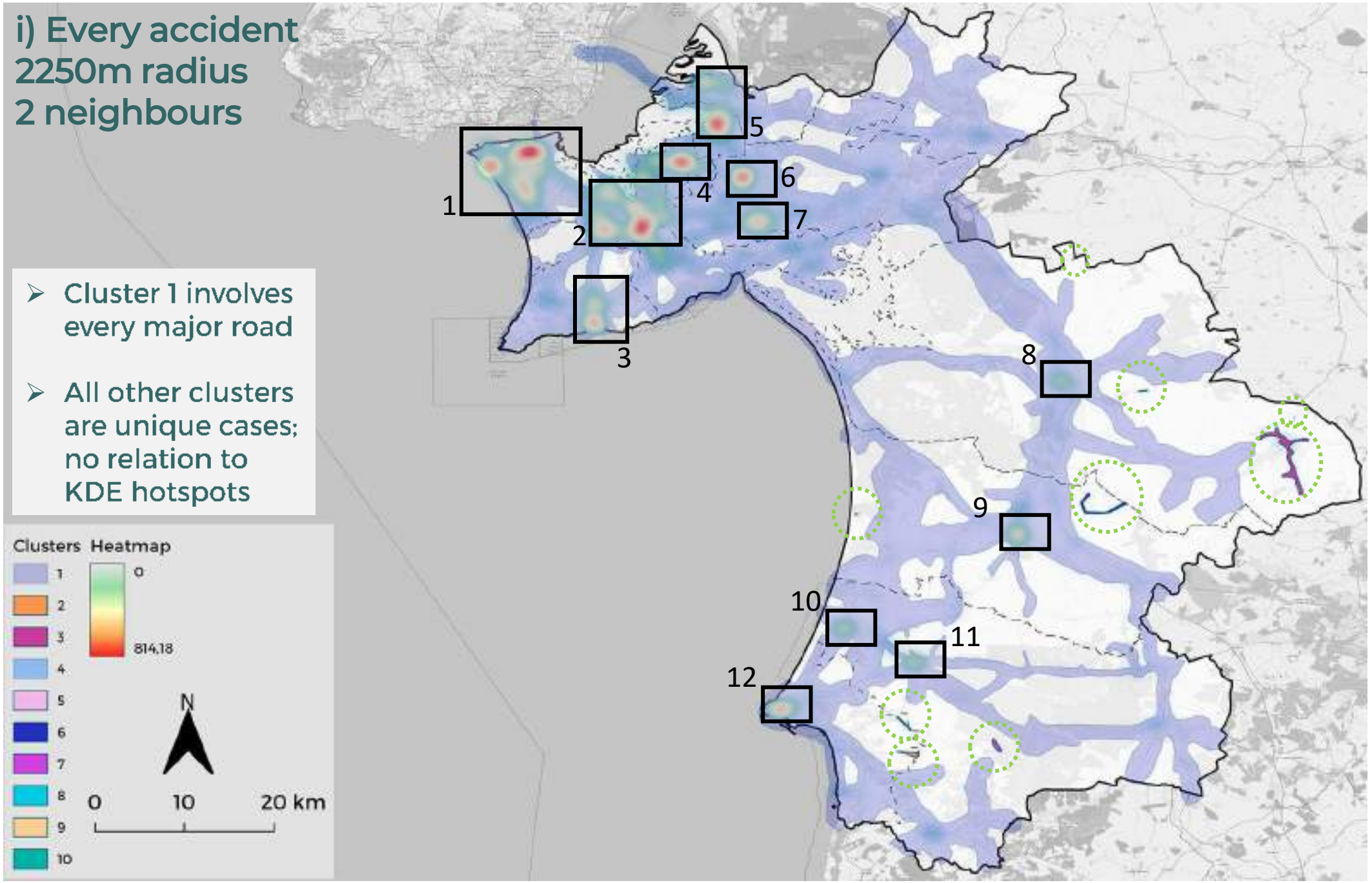


Determines clusters based on point relation, not just proximity

Searches for neighbours within a defined distance

Has to have a minimum number of neighbours to belong to the cluster

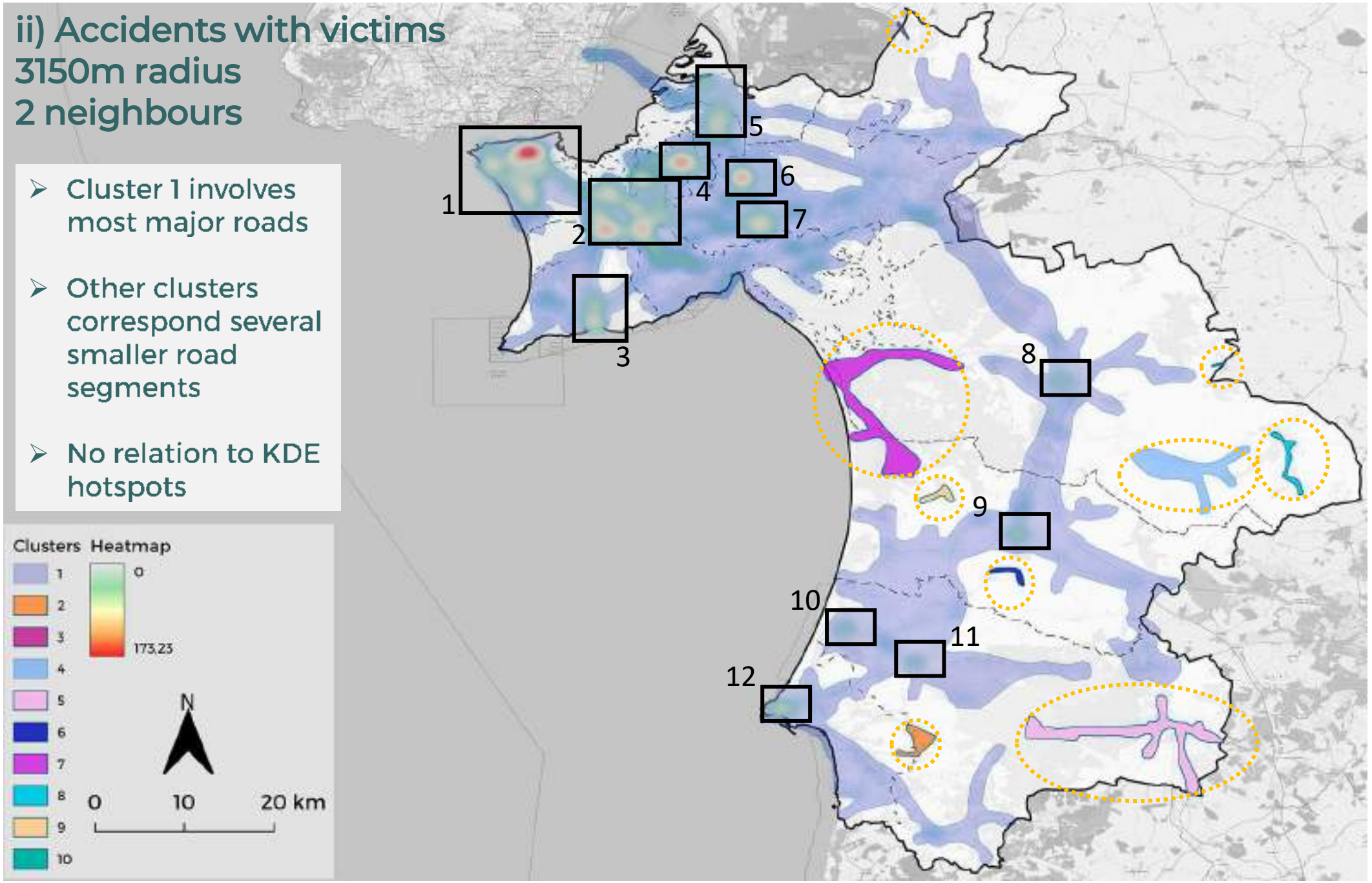
i) Every accident 2250m radius 2 neighbours



- Cluster 1 involves every major road
- All other clusters are unique cases; no relation to KDE hotspots

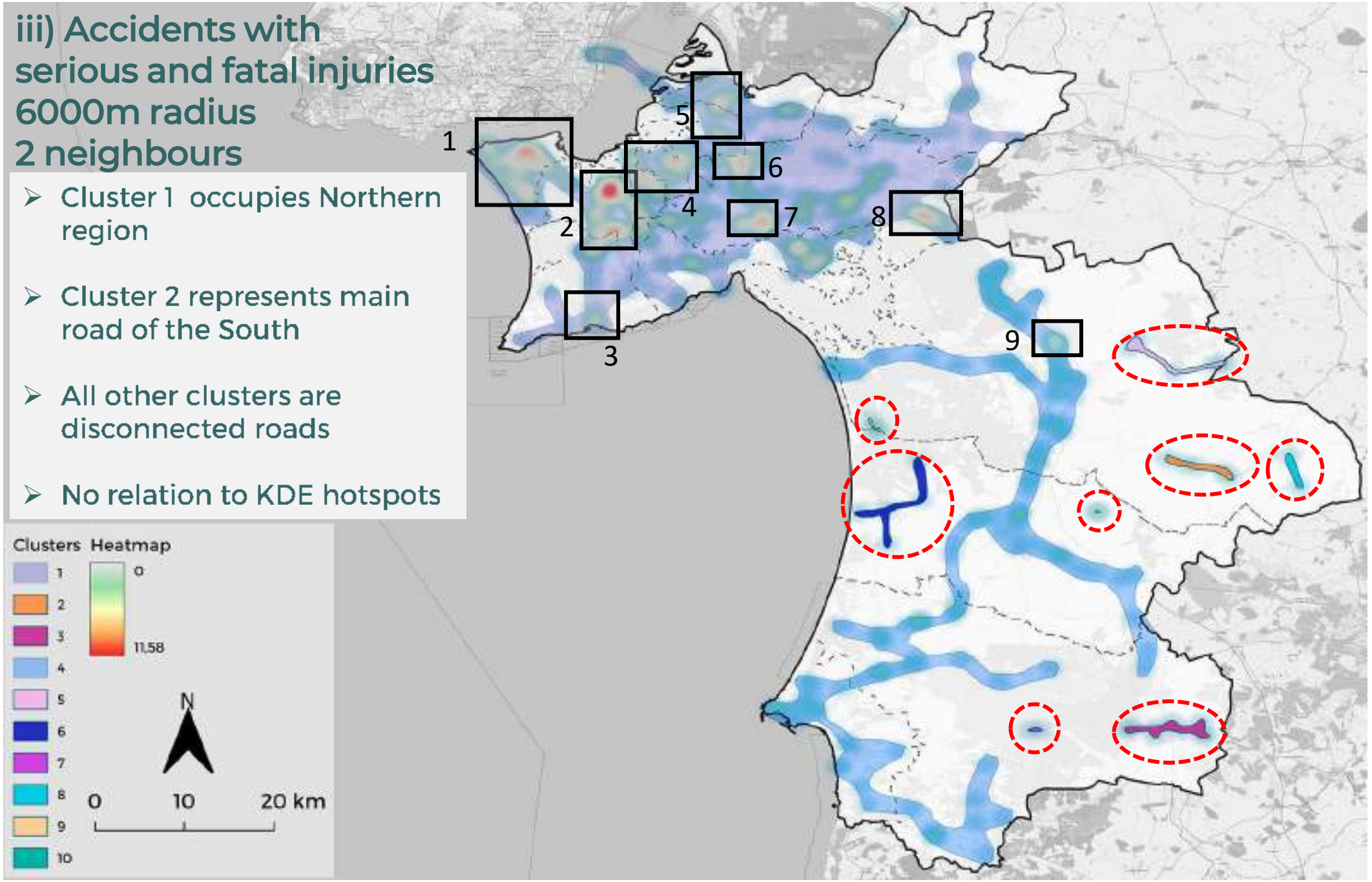
ii) Accidents with victims
3150m radius
2 neighbours

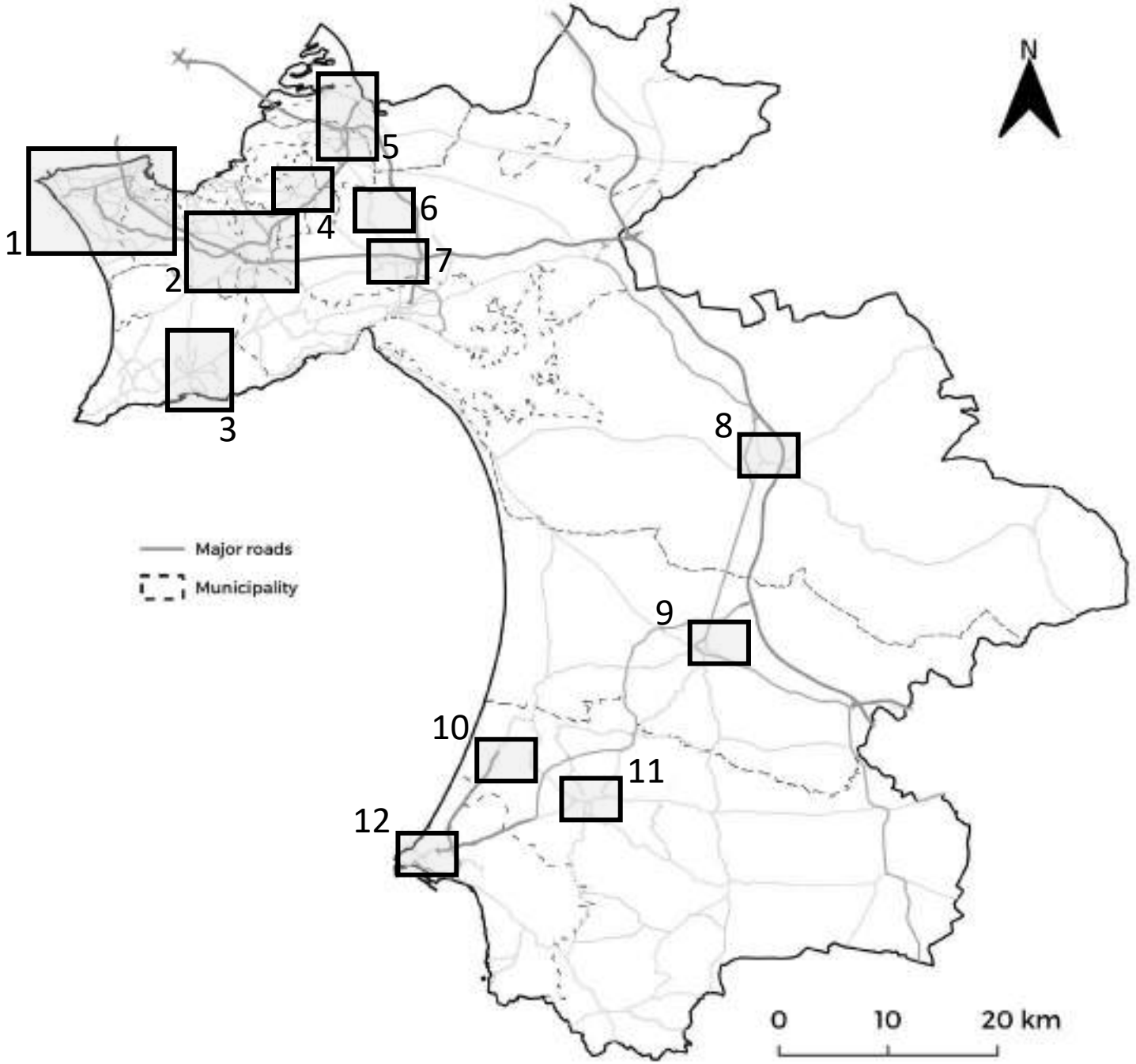
- Cluster 1 involves most major roads
- Other clusters correspond several smaller road segments
- No relation to KDE hotspots

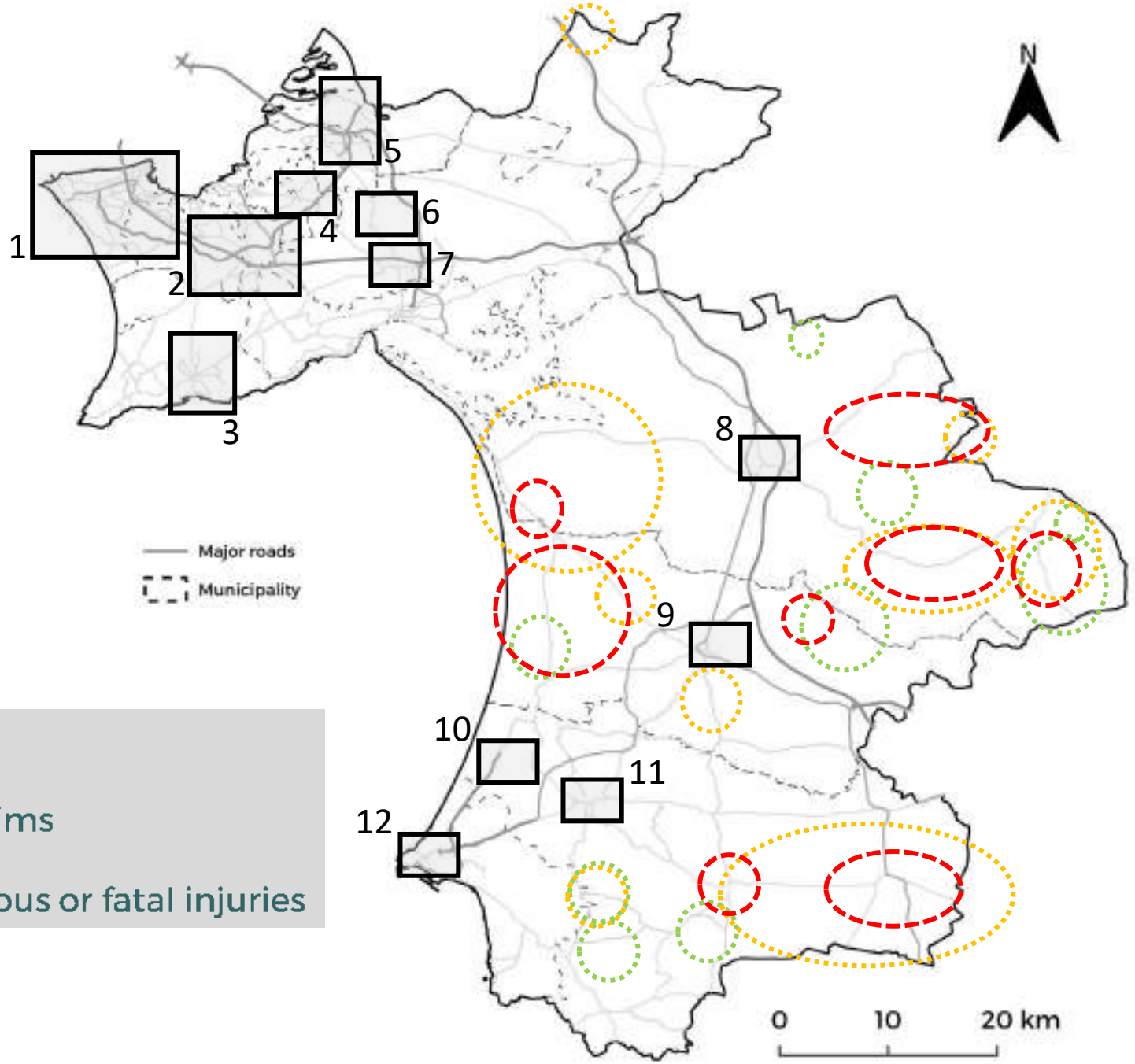


iii) Accidents with serious and fatal injuries 6000m radius 2 neighbours

- Cluster 1 occupies Northern region
- Cluster 2 represents main road of the South
- All other clusters are disconnected roads
- No relation to KDE hotspots



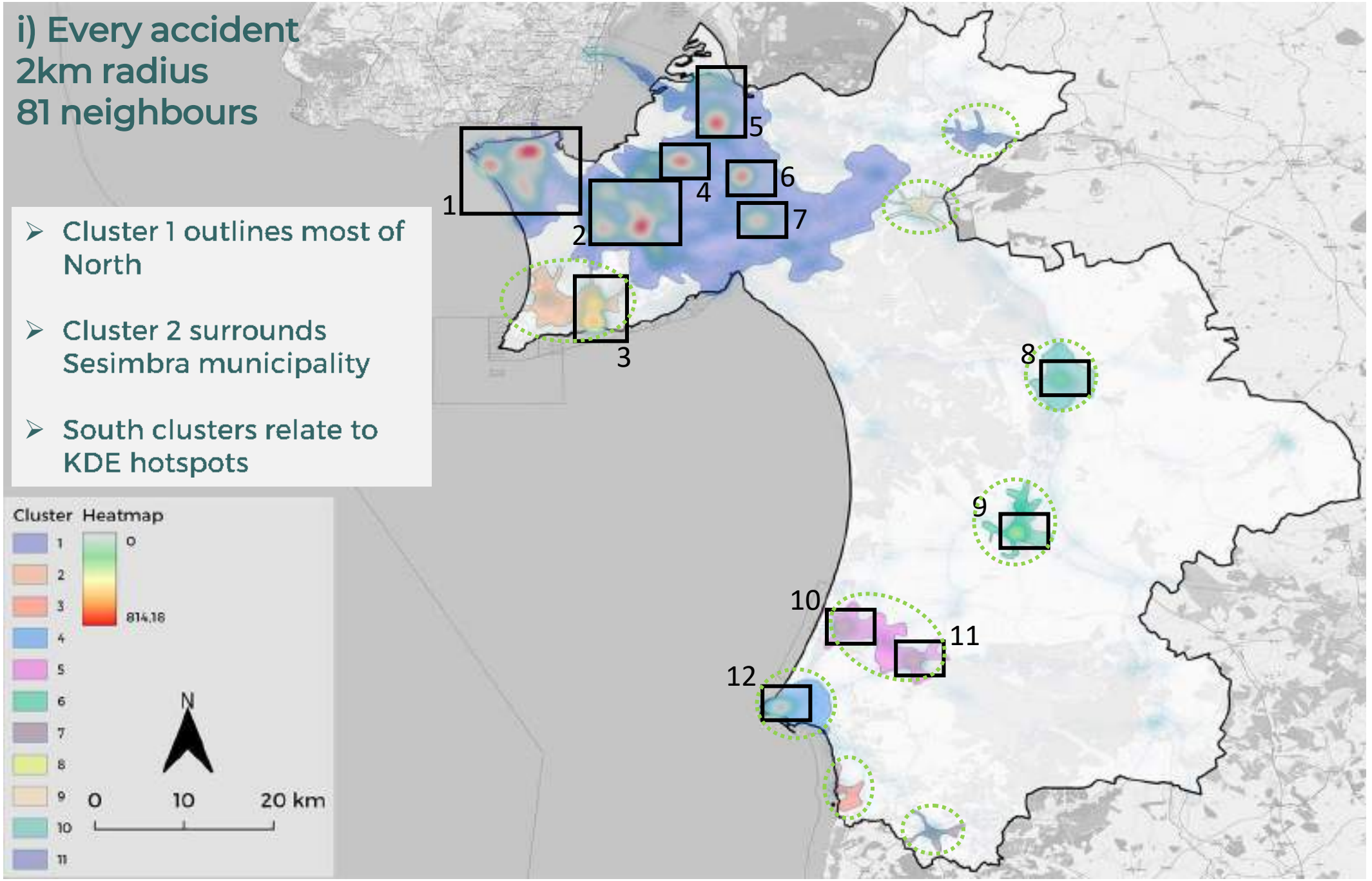

N
1
2
3
4
5
6
7
8
9
10
11
12
Major roads
Municipality
0
10
20 km



- Clusters for every accident
- Clusters for accidents with victims
- Clusters for accidents with serious or fatal injuries

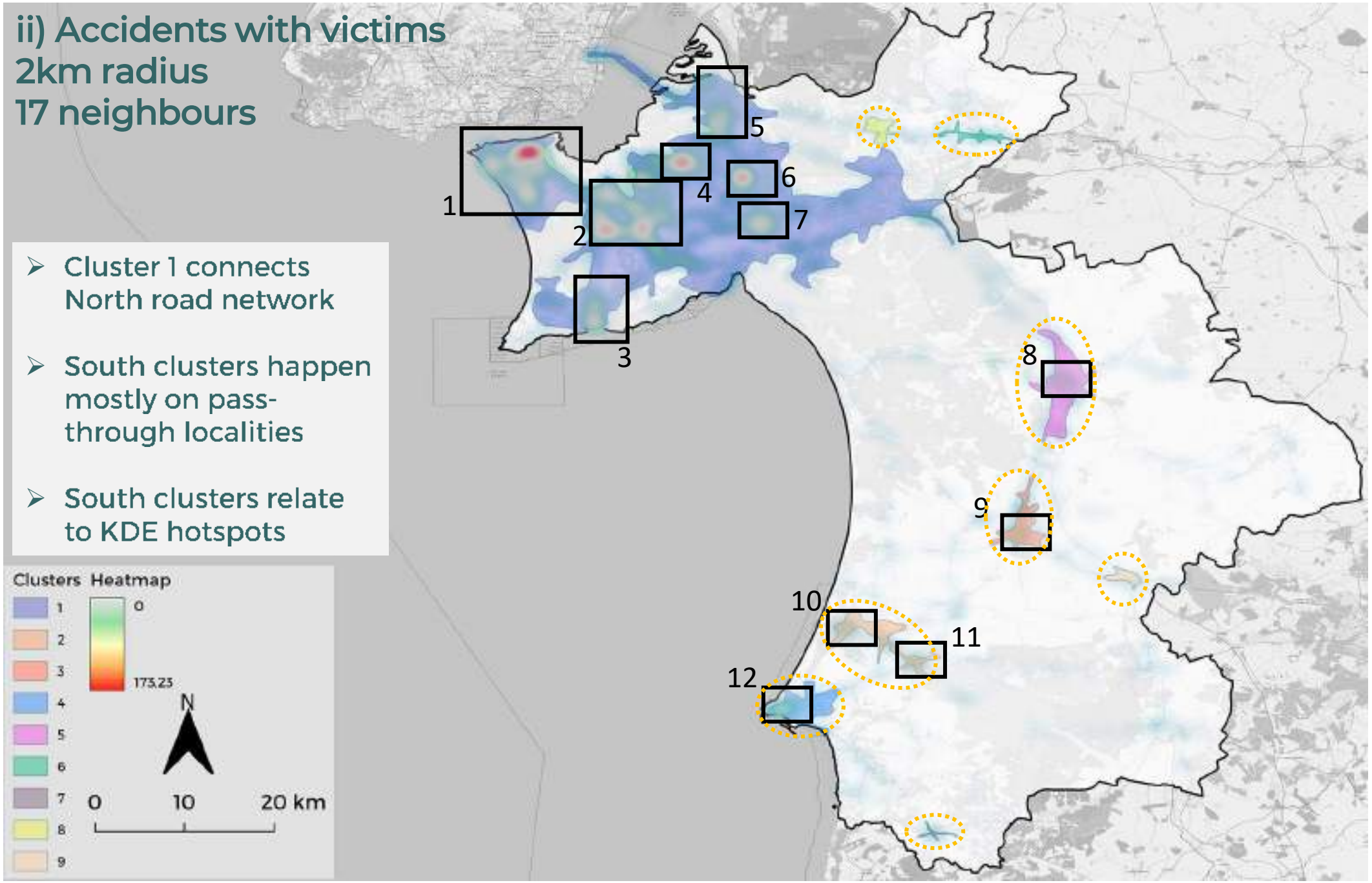
i) Every accident
2km radius
81 neighbours

- Cluster 1 outlines most of North
- Cluster 2 surrounds Sesimbra municipality
- South clusters relate to KDE hotspots



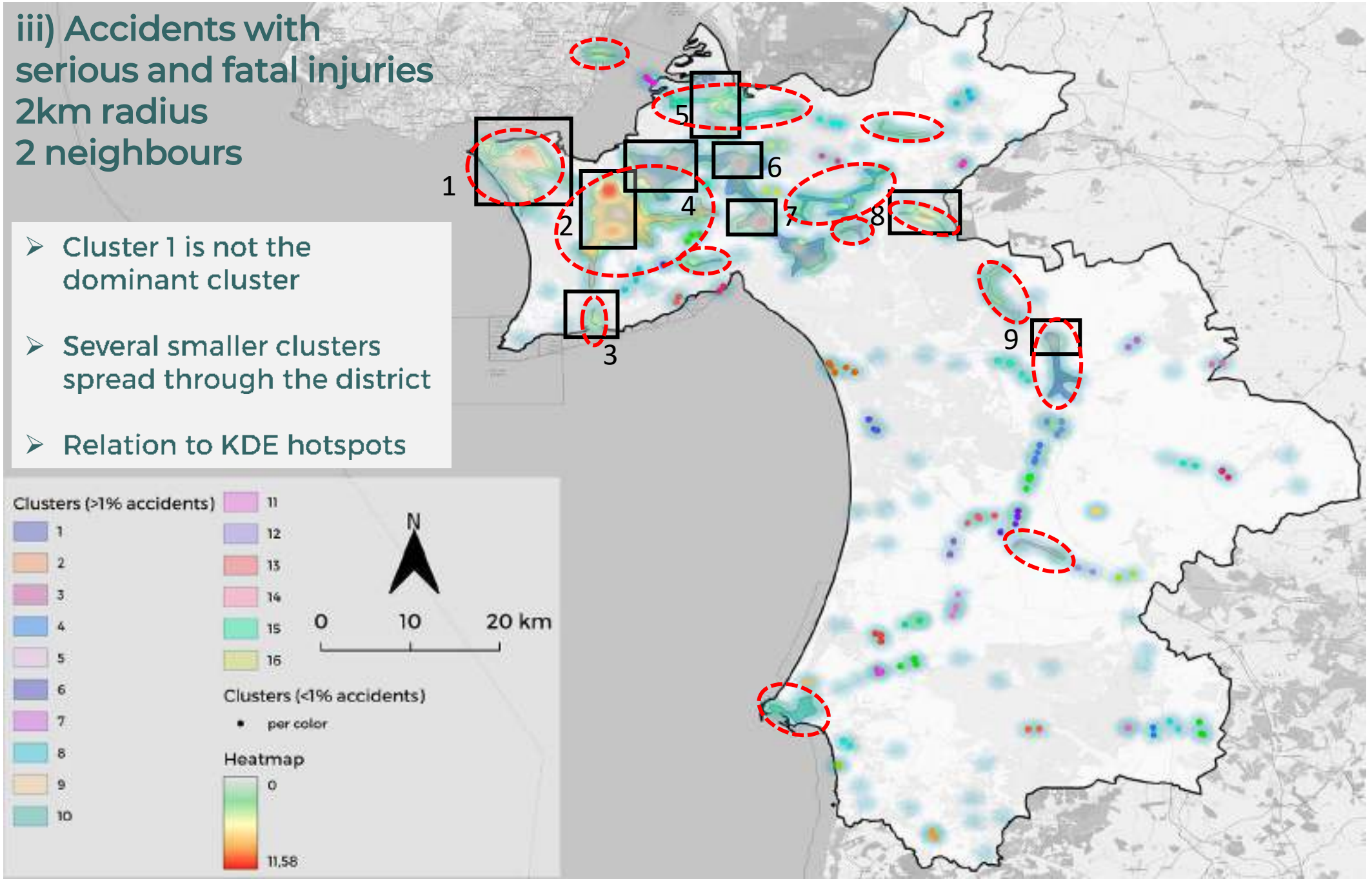
ii) Accidents with victims
2km radius
17 neighbours

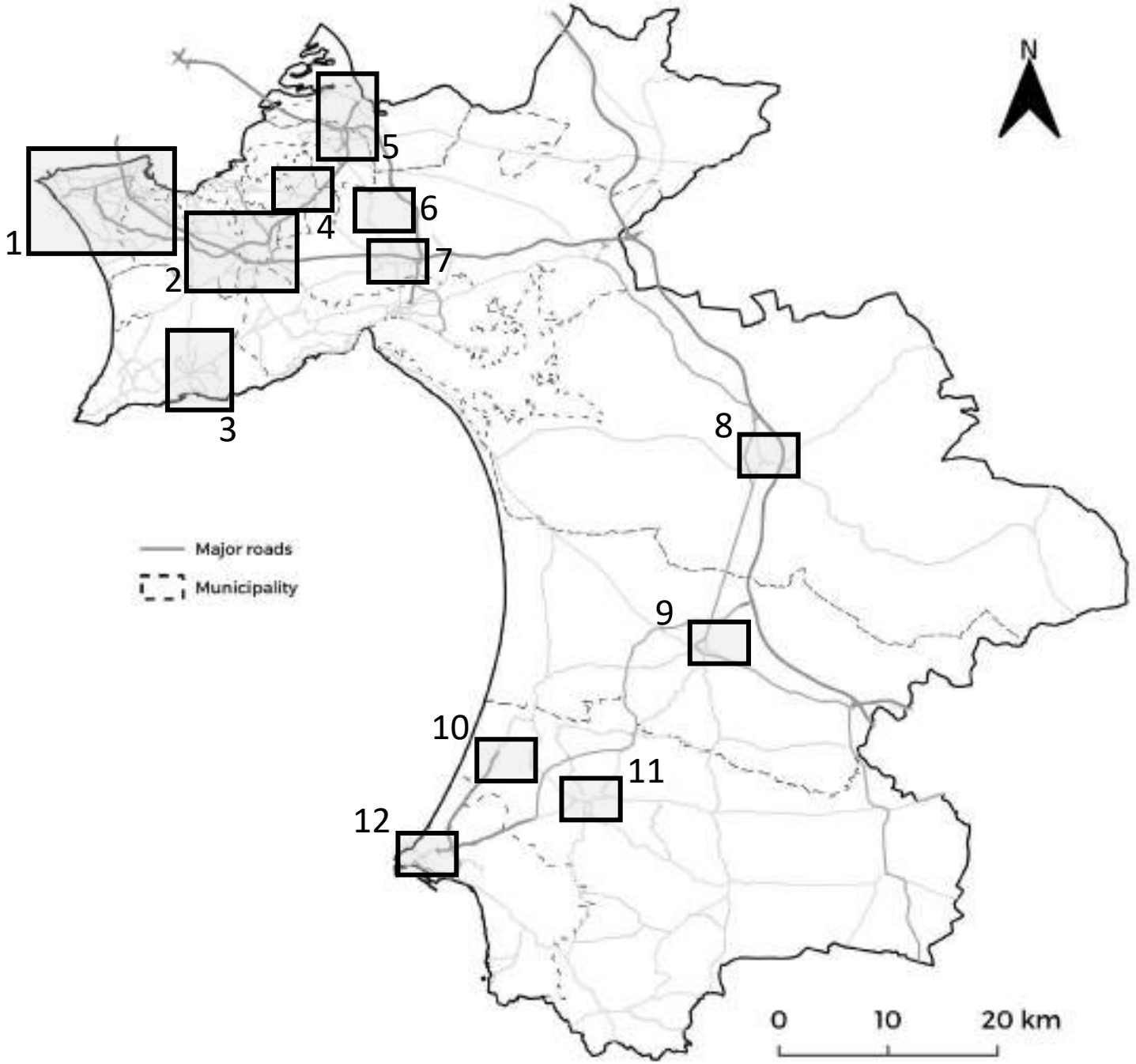
- Cluster 1 connects North road network
- South clusters happen mostly on pass-through localities
- South clusters relate to KDE hotspots

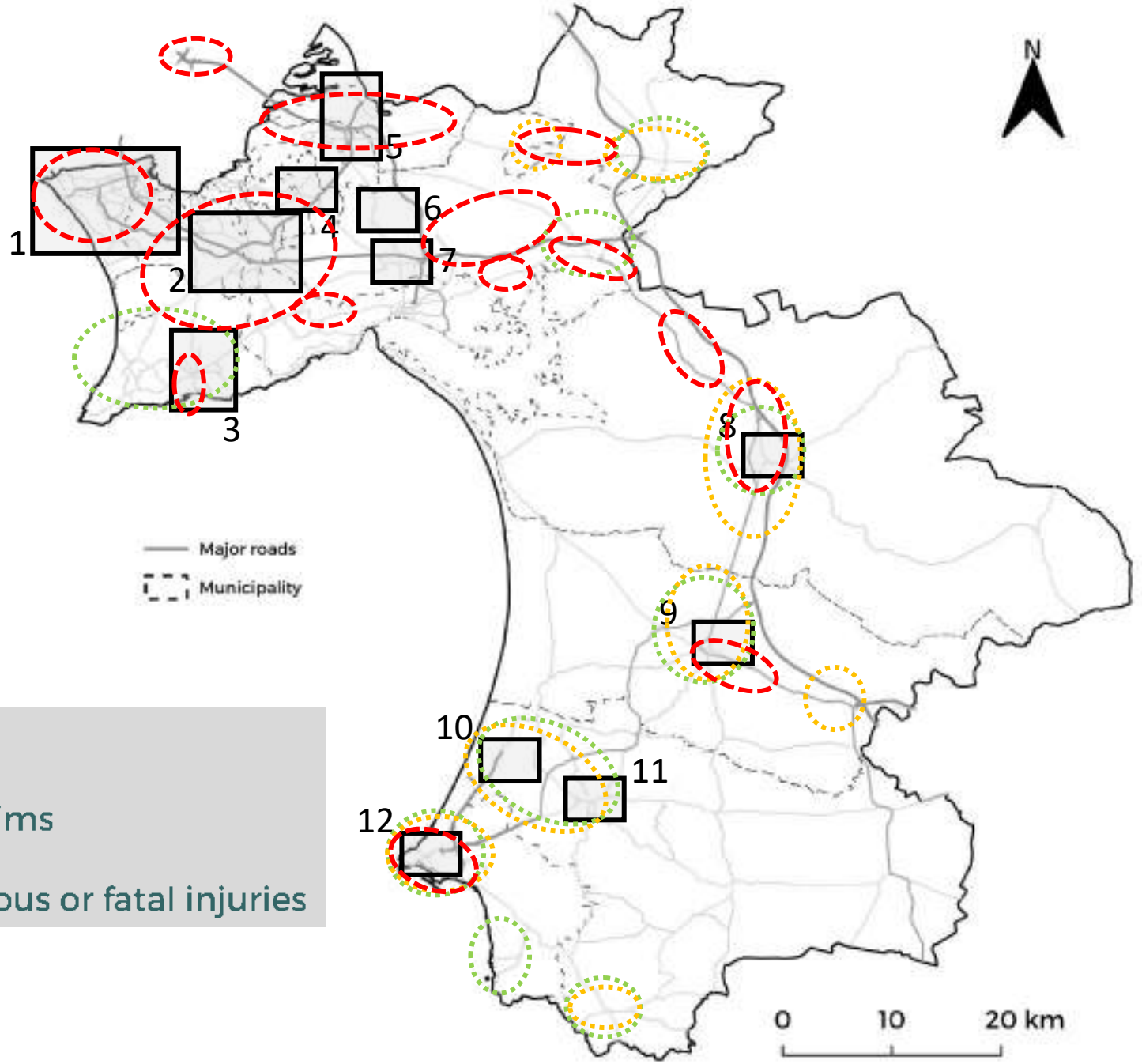


iii) Accidents with serious and fatal injuries 2km radius 2 neighbours

- Cluster 1 is not the dominant cluster
- Several smaller clusters spread through the district
- Relation to KDE hotspots




N
1
2
3
4
5
6
7
8
9
10
11
12
Major roads
Municipality
0
10
20 km



- Clusters for every accident
- Clusters for accidents with victims
- Clusters for accidents with serious or fatal injuries

Concluding remarks

- KDE mostly highlights localities and important road network intersections

DBSCAN adjusted for 10 clusters	VS.	DBSCAN adjusted for 10% of max KDE
Large, dominating cluster		Smaller clusters, essentially outlining intersections
Dataset i) and ii) share similarity between cluster locations		Dataset i) and ii) are similar, but ii) lost two clusters
Dataset iii) separates urban from rural; rural has several smaller clusters		Dataset iii) breaks the urban area in several sectors
No relation to KDE hotspots		Relation to KDE hotspots
Overlap between dataset ii) and iii) clusters		Better overlap between all dataset clusters

The combination of KDE and DBSCAN is greater than the sum of their individual parts

Concluding remarks – Pratical uses

- Law enforcement can use the maps to know where to deploy a patrol to monitor traffic.
  - Smaller clusters mean less uncertainty on location
  - Available units are deployed more efficiently

- Municipalities can use the maps to locate roads with problems within their jurisdiction, and correct them.

- By analysing time periods, DBSCAN-KDE maps highlight areas more accident-prone, helping to plan traffic control operations

Aknowledgements:
This research was funded by the Portuguese funding agency, FCT-Fundação para a Ciência e a Tecnologia, grant number FCT DSAIPA/DS/0090/2018, "MOPREVIS—Modelação e Predição de Acidentes de Viação no Distrito de Setúbal", within the scope of the National Initiative on Digital Skills e.2030, Portugal INCoDe.2030

We appreciate your time and attention.

MOPREVIS

UNIVERSIDADE DE ÉVORA

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

GNR
GUARDA NACIONAL REPUBLICANA

Further Reading

Daniel Santos, Vitor Nogueira, José Saias, Paulo Quaresma, Paulo Infante, Gonçalo Jacinto, Anabela Afonso, Pedro Nogueira, Marcelo Silva, Rosalina Costa, Patrícia Gois, and Paulo Manuel, 2022, Machine Learning Approach to Identify Factors that Influence Accident Severity, IFCS22

Gatrell, A. C., Bailey, T. C., Diggle, P. J., & Rowlingson, B. S. (1996). Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology. Transactions of the Institute of British Geographers, 21(1), 256–274. https://doi.org/10.2307/622936

Paulo Infante; Gonçalo Jacinto; Anabela Afonso; Leonor Rego; Vitor Nogueira; Paulo Quaresma; José Saias; Daniel Santos; Pedro Nogueira; Marcelo Silva; Rosalina Pisco Costa; Patrícia Gois; Paulo Rebelo Manuel, 2022, Comparison of Statistical and Machine-Learning Models on Road Traffic Accident Severity Classification, Computers 2022, Volume 11, Issue 5, 80

Paulo Infante; Gonçalo Jacinto; Anabela Afonso; Leonor Rego; Vitor Nogueira; Paulo Quaresma; José Saias; Daniel Santos; Pedro Nogueira; Marcelo Silva; Rosalina Pisco Costa; Patrícia Gois; Paulo Rebelo Manuel, 2022, Some Determinants for Road Accidents Severity in the District of Setúbal, Recent Developments in Statistics and Data Science, Springer Proceedings in Mathematics and Statistics, in press

Paulo Infante, Gonçalo Jacinto, Anabela Afonso, Leonor Rego, Vitor Nogueira, Paulo Quaresma, José Saias, Daniel Santos, Pedro Nogueira, Marcelo Silva, Rosalina Pisco Costa, Patrícia Gois, and Paulo Rebelo Manuel, 2022, Some Factors That Influence the Nature of Road Traffic Accidents, IFCS22

World Health Organization. (2018). Global status report on road safety 2018, Geneva