TÉCNICO LISBOA

CERENA
Centro de Recursos Naturais e Ambiente

Ciência espacial de dados para auxiliar na gestão da pandemia COVID-19

Maria João Pereira, Manuel Ribeiro, Amílcar Soares, Ana F. Duarte, Leonardo Azevedo

leonardo.azevedo@tecnico.ulisboa.pt

@leoap

encontro CIÊNCIA '22
ENCONTRO COM A CIÊNCIA E TECNOLOGIA EM PORTUGAL
16-18 maio

Data and risk maps in epidemiology

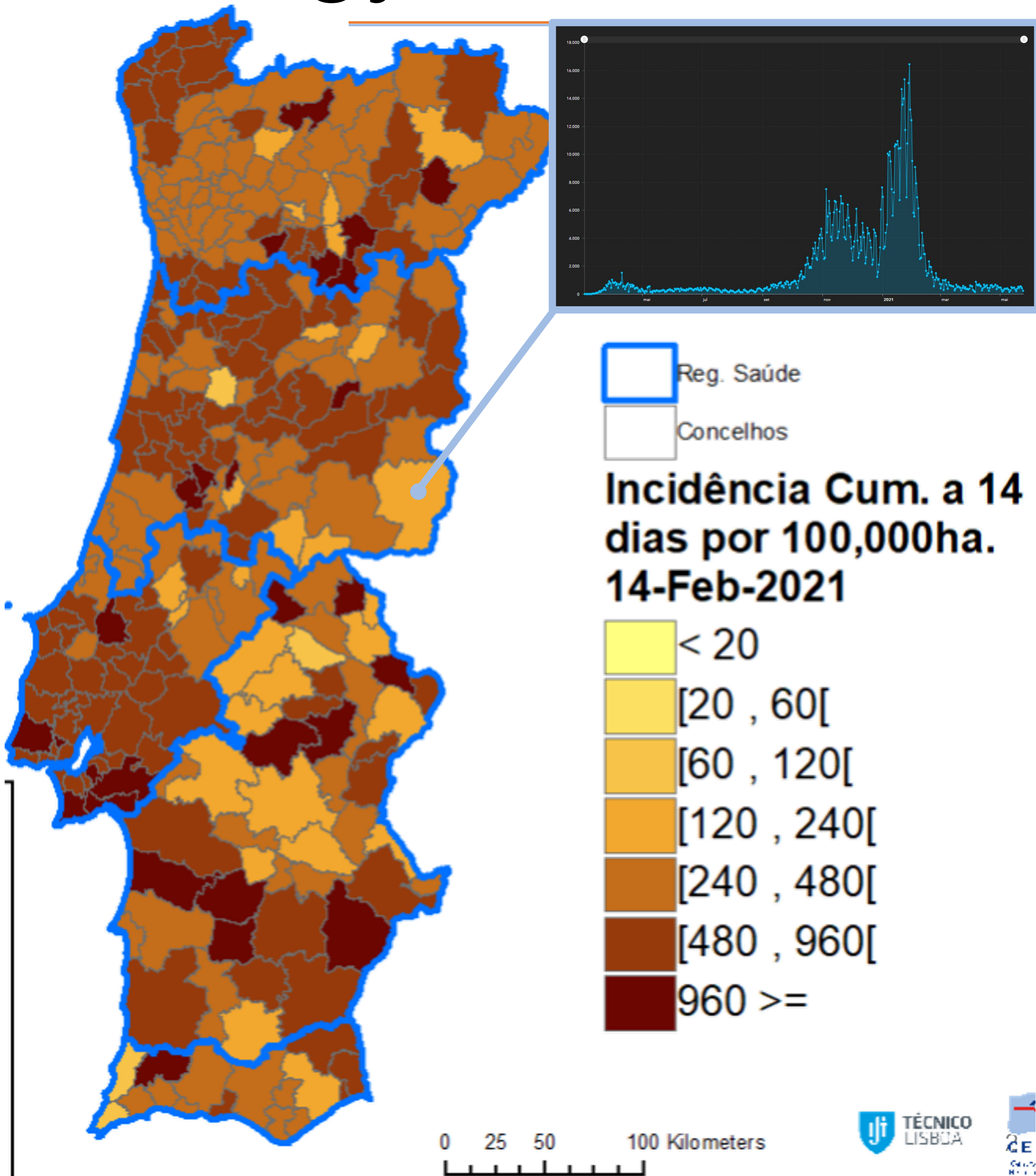
- Data viz tools to monitor and control disease spread, preventing and mitigation measures effectiveness and allocation of resources

- Most natural and anthropogenic phenomena are spatially structured;

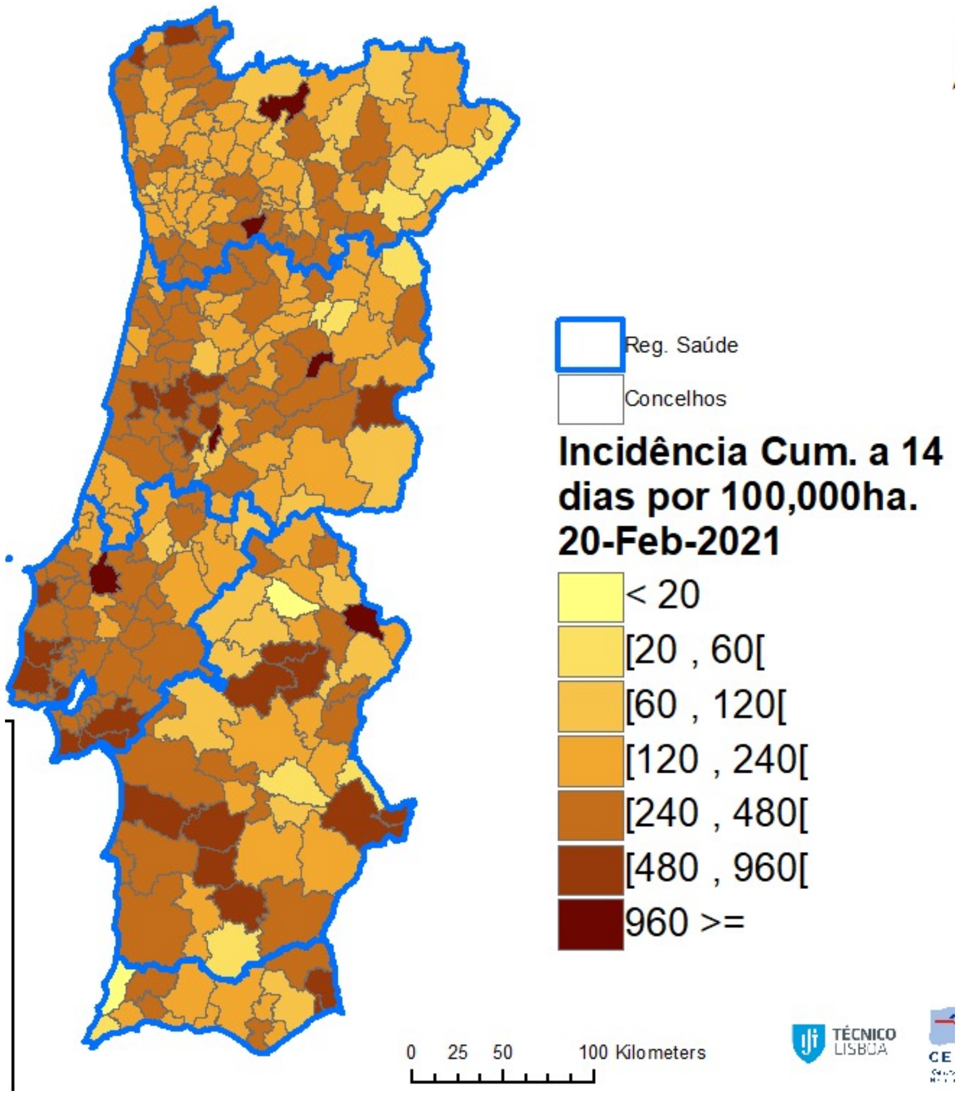
- SARS-CoV-2 epidemic has simultaneous time and space dynamics;

- **Incidence** = the **rate of new cases** of a disease occurring in a **specific population** over a **given period of time**

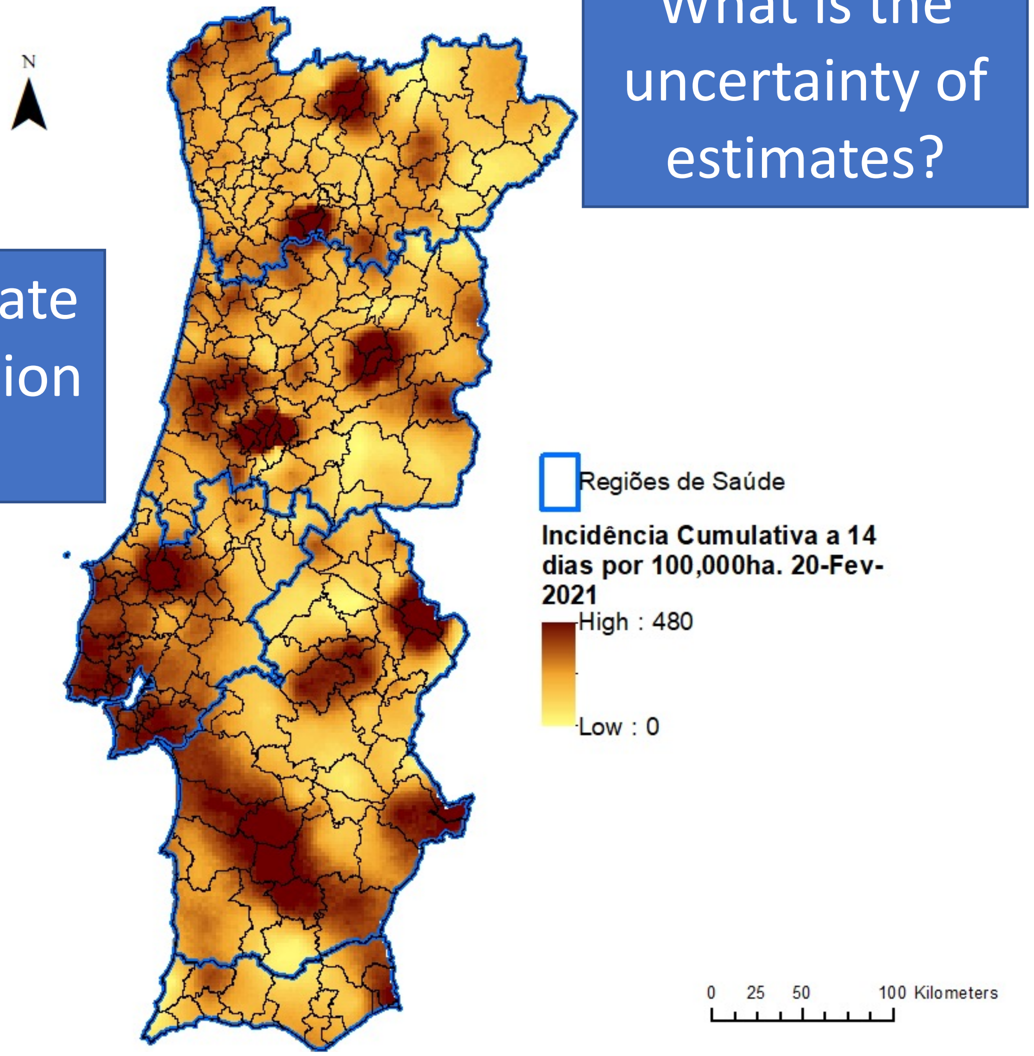
- **Incidence** as a measure of **risk** presented in **discontinuous maps**



Risk is not discontinuous!



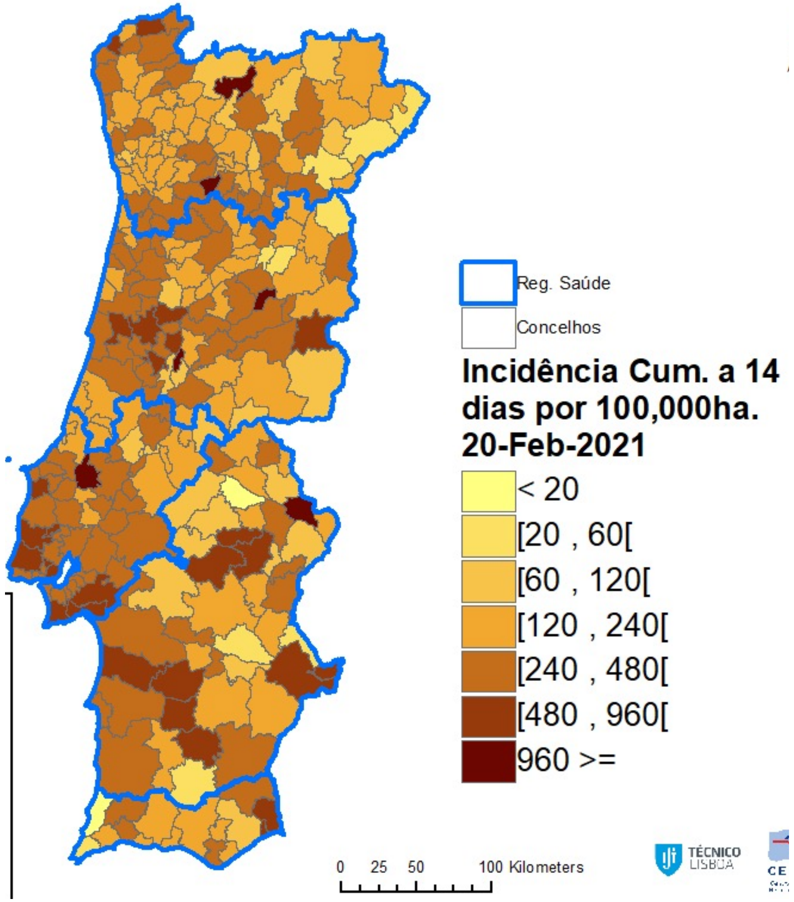
How to generate a high-resolution map?

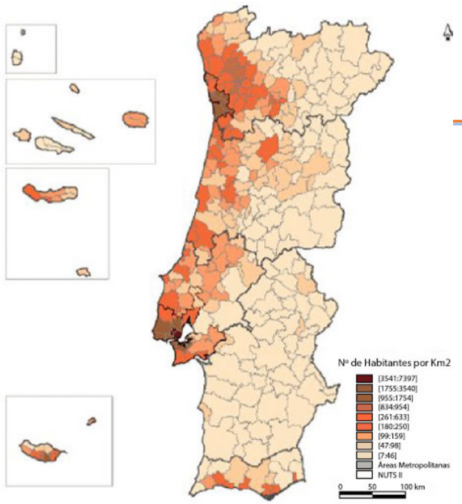


What is the uncertainty of estimates?

Challenges

$$\textbf{Incidence} = \frac{new\ infections\ per\ municipality\ in\ 2\ weeks\ period}{municipal\ resident\ population}$$

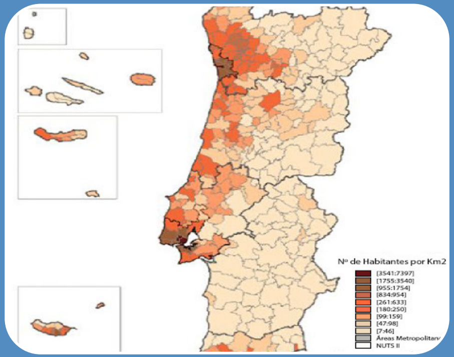




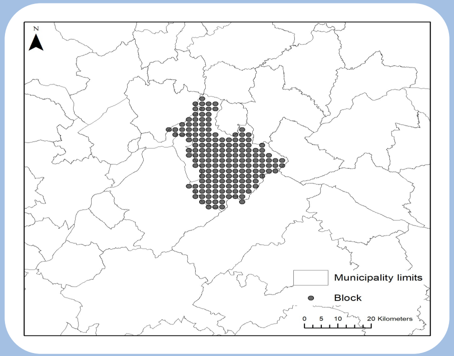
data uncertainty depends on the resident population of the municipality

Each municipality data has a different **support:** "block" data or "areal" data

Methology



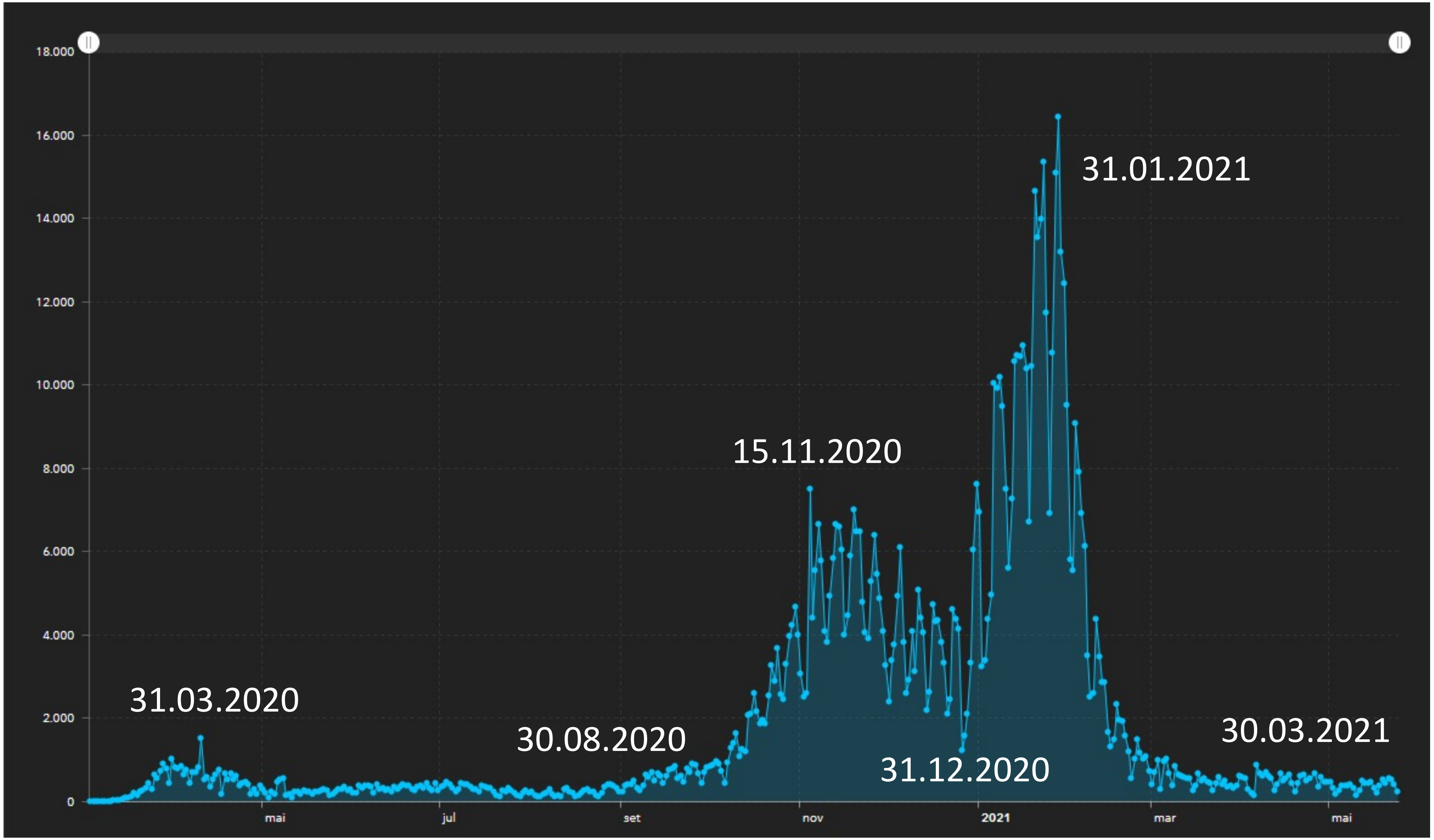
Poisson model for rare diseases proposed by Waller and Gotway [2004] and extended into a geostatistical framework by Goovaerts [2005] and Oliveira et al. [2013].



Direct block sequential simulation (block-DSS) by Liu and Journel [2009].

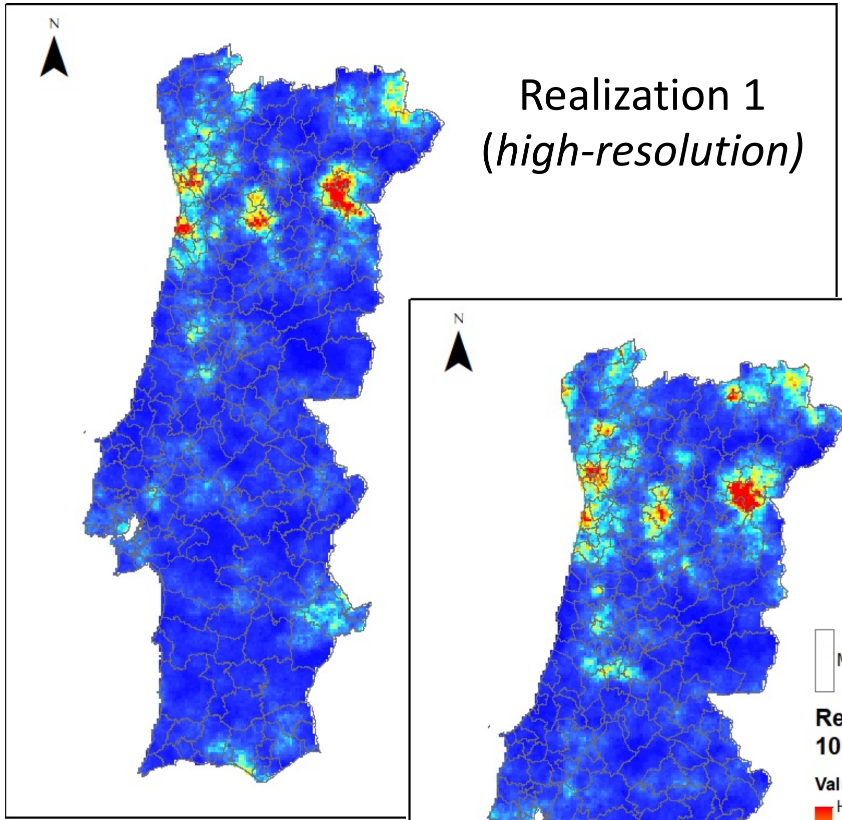
Azevedo, L., Pereira, M.J., Ribeiro, M.C. *et al.* Geostatistical COVID-19 infection risk maps for Portugal. *Int J Health Geogr* **19,** 25 (2020). https://doi.org/10.1186/s12942-020-00221-5

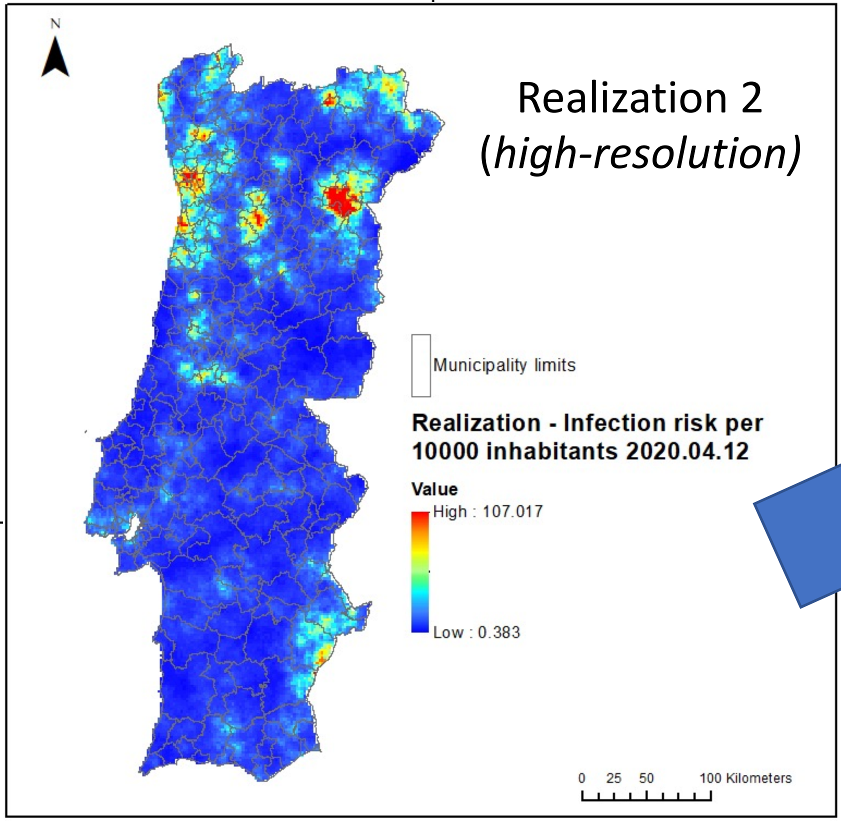
Daily number of new cases



... 07.03.2022

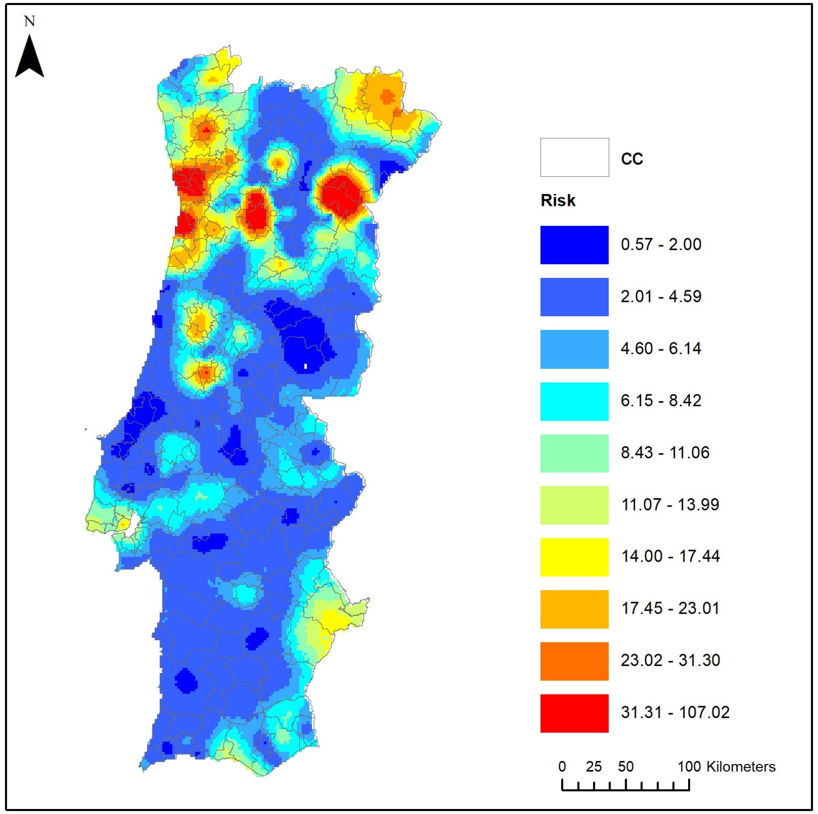
Daily updated Infection Risk Maps



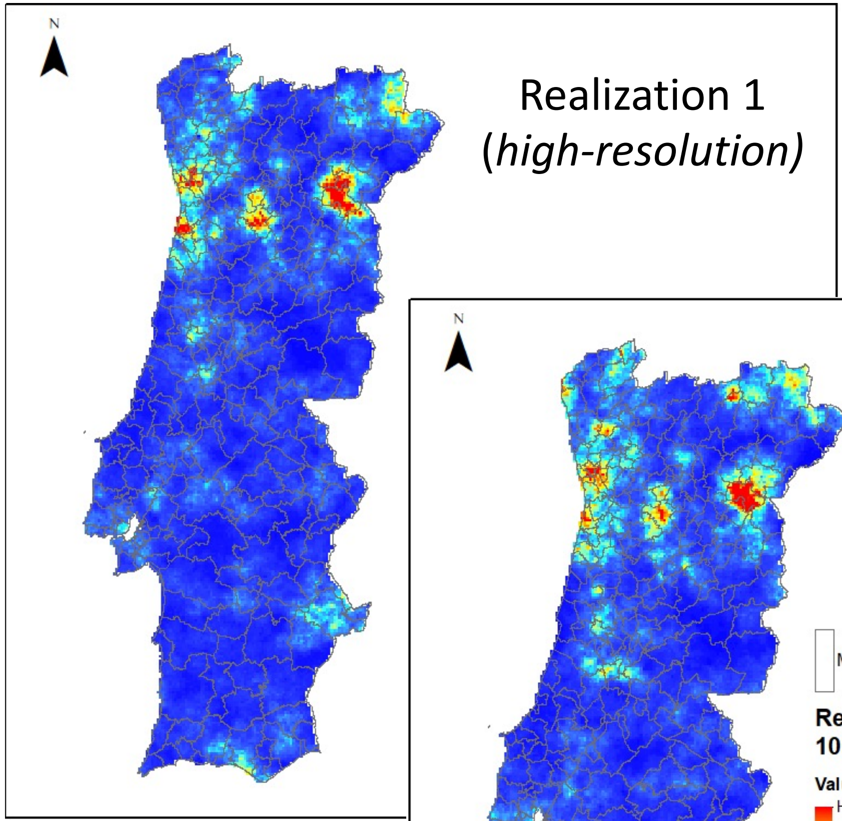


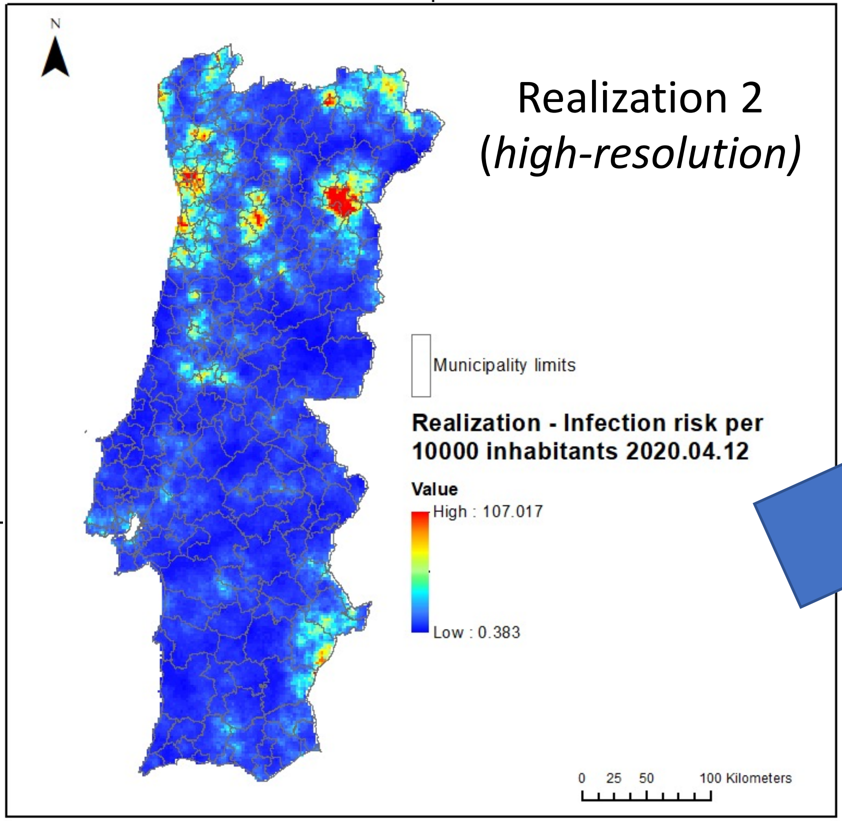
.... N_s *realizations*

Average or median risk map



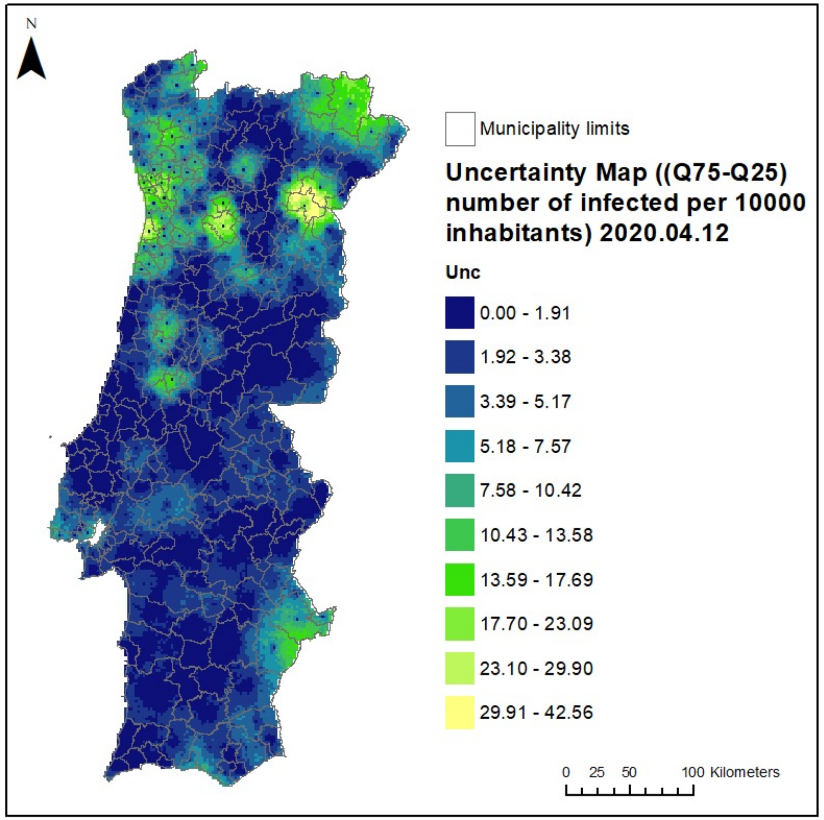
Daily updated Infection Risk Maps



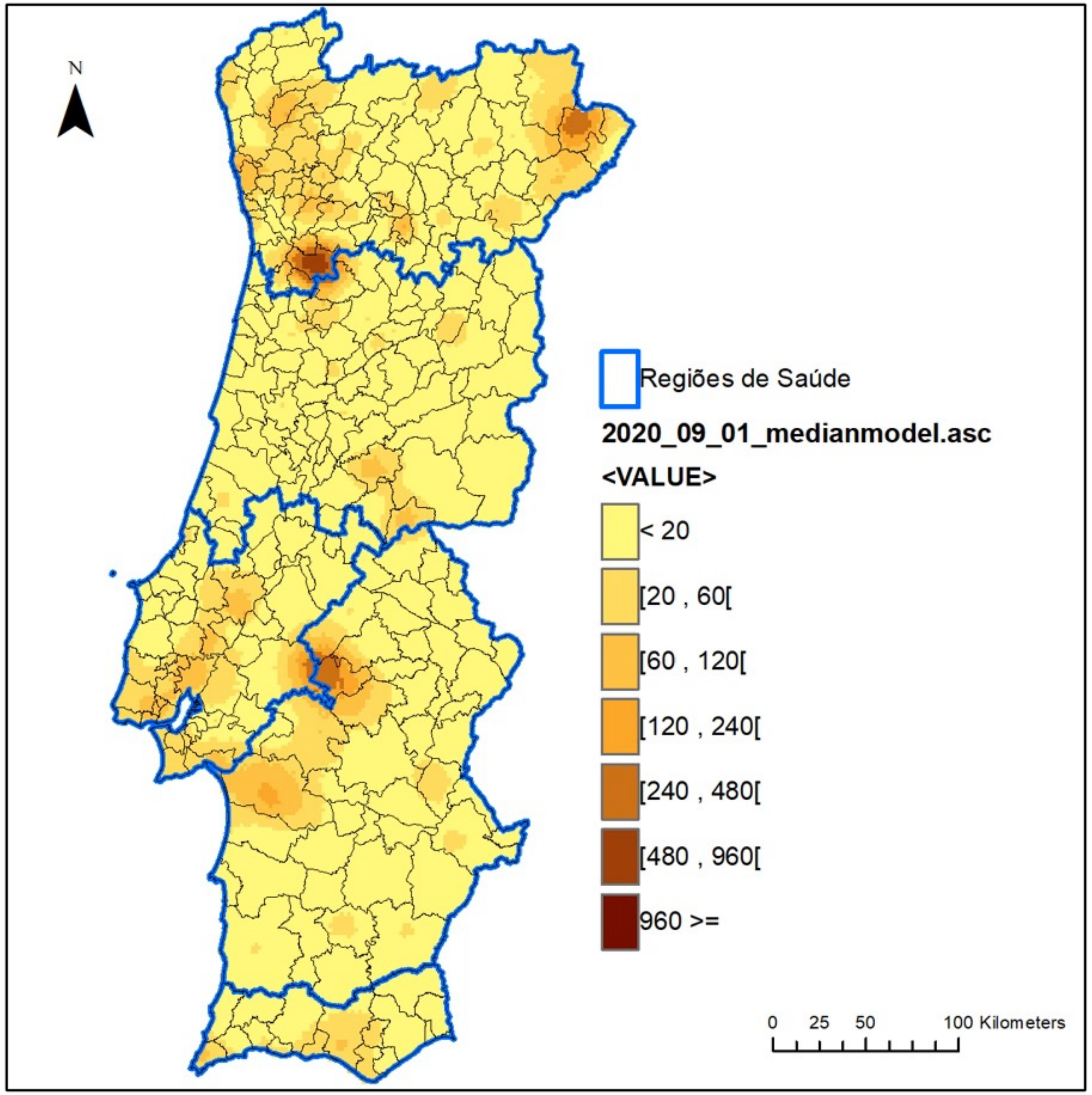


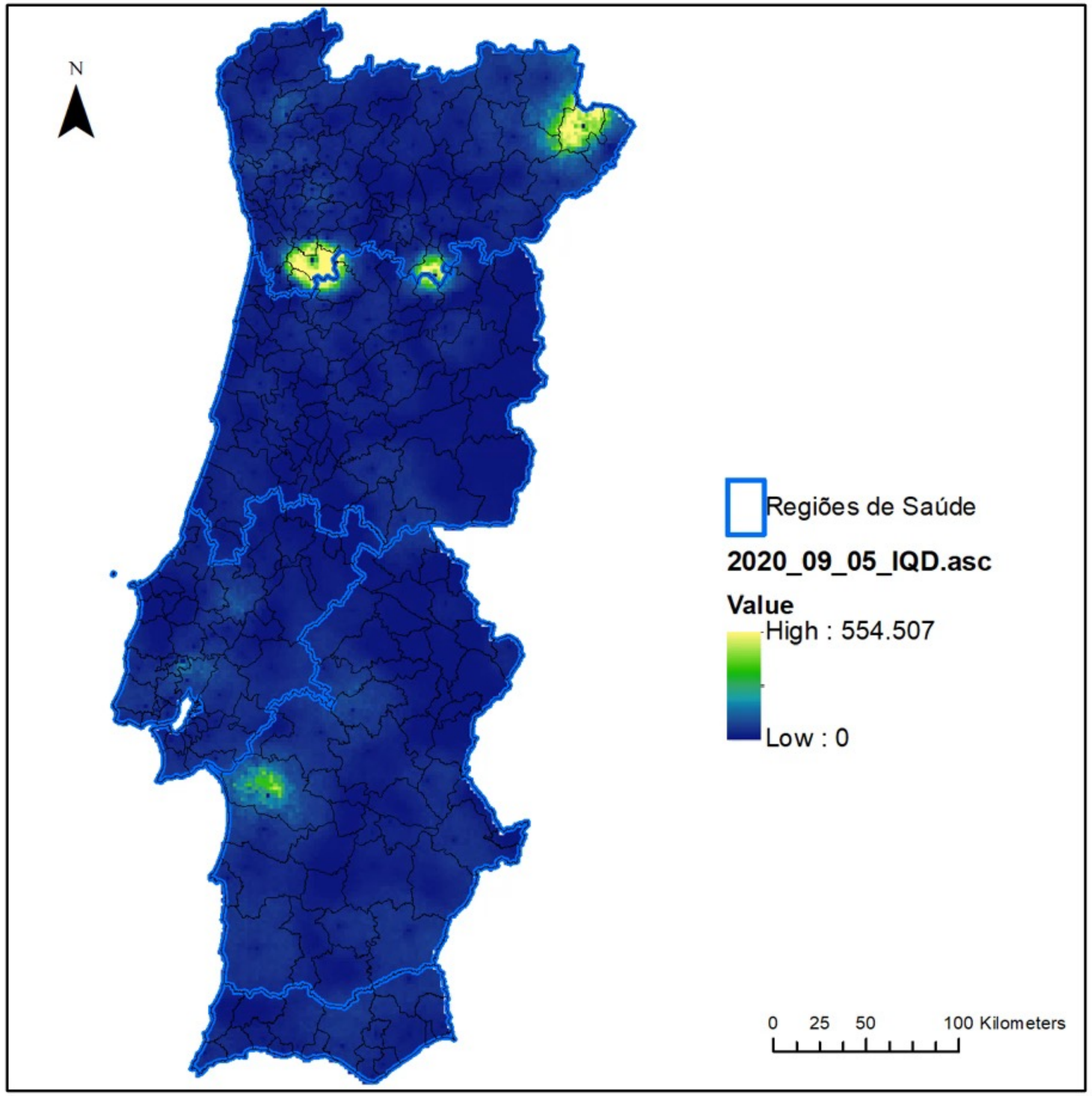
.... N_s *realizations*

Spatial uncertainty map



Second and third epidemic wave

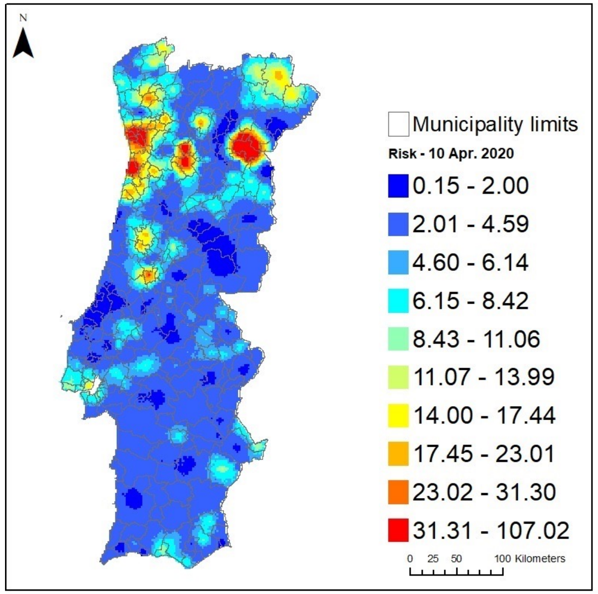




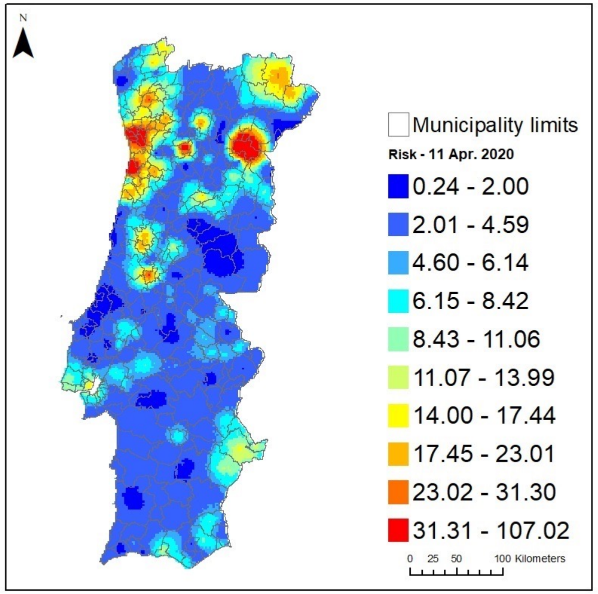
Temporal trend of COVID-19 infection risk

TÉCNICO LISBOA

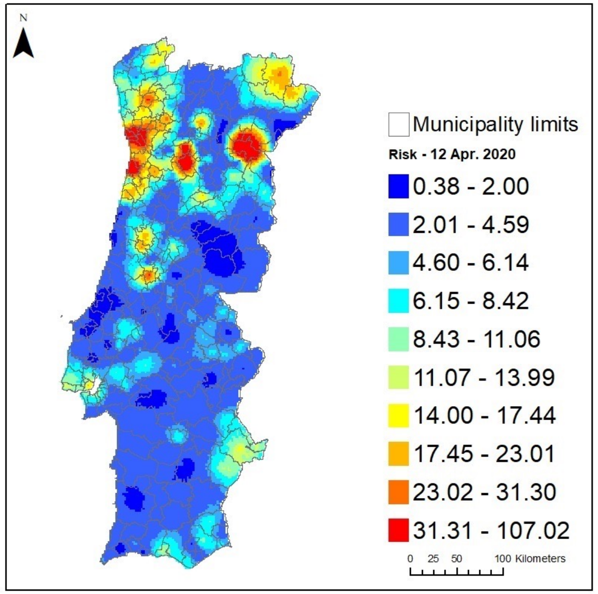
CERENA
Centro de Recursos Naturais e Ambiente



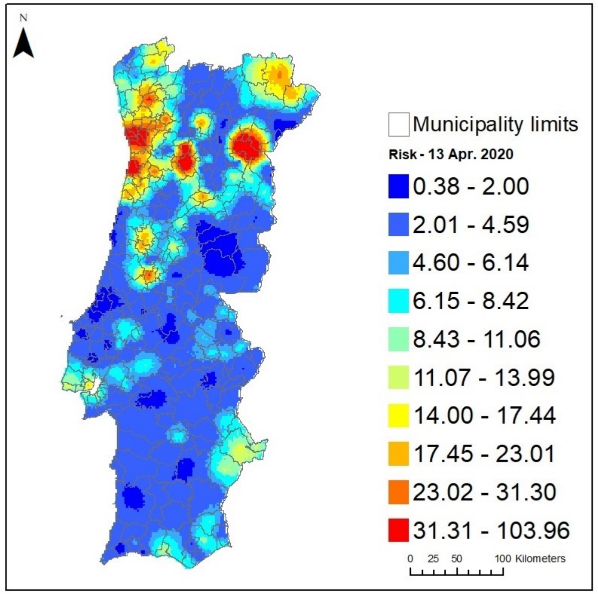
a)



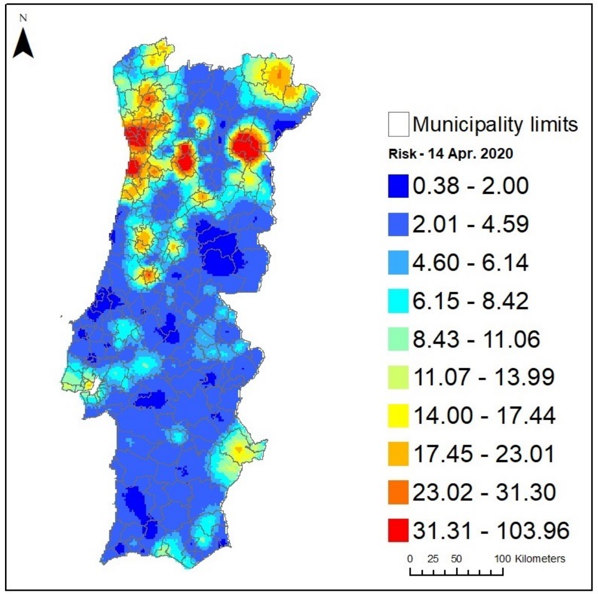
b)



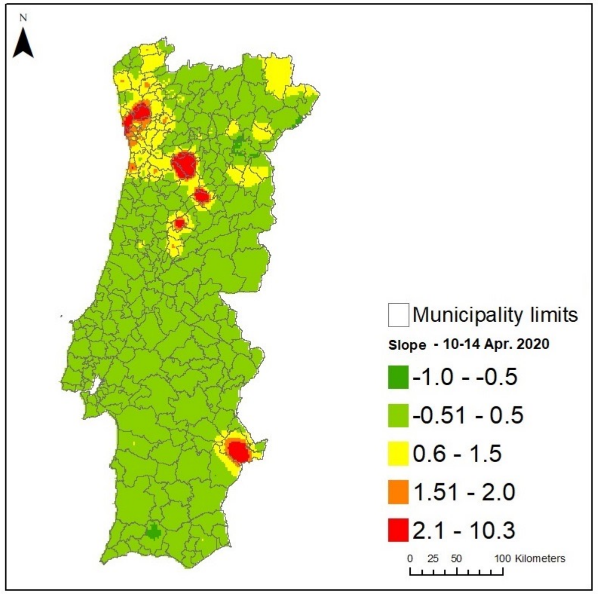
c)



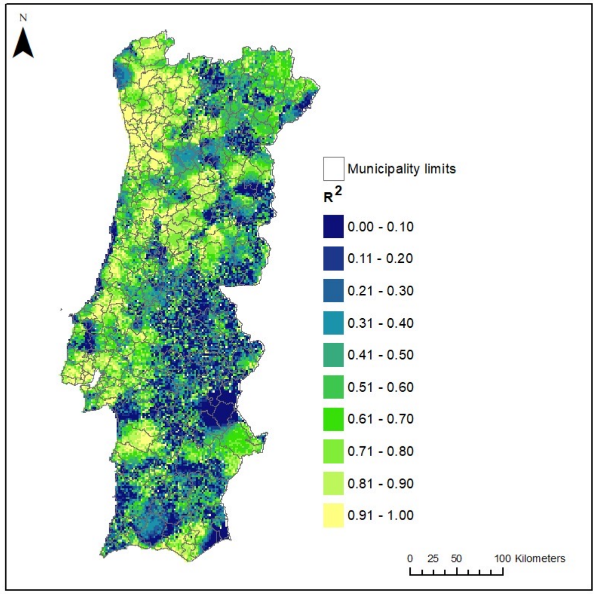
d)



e)

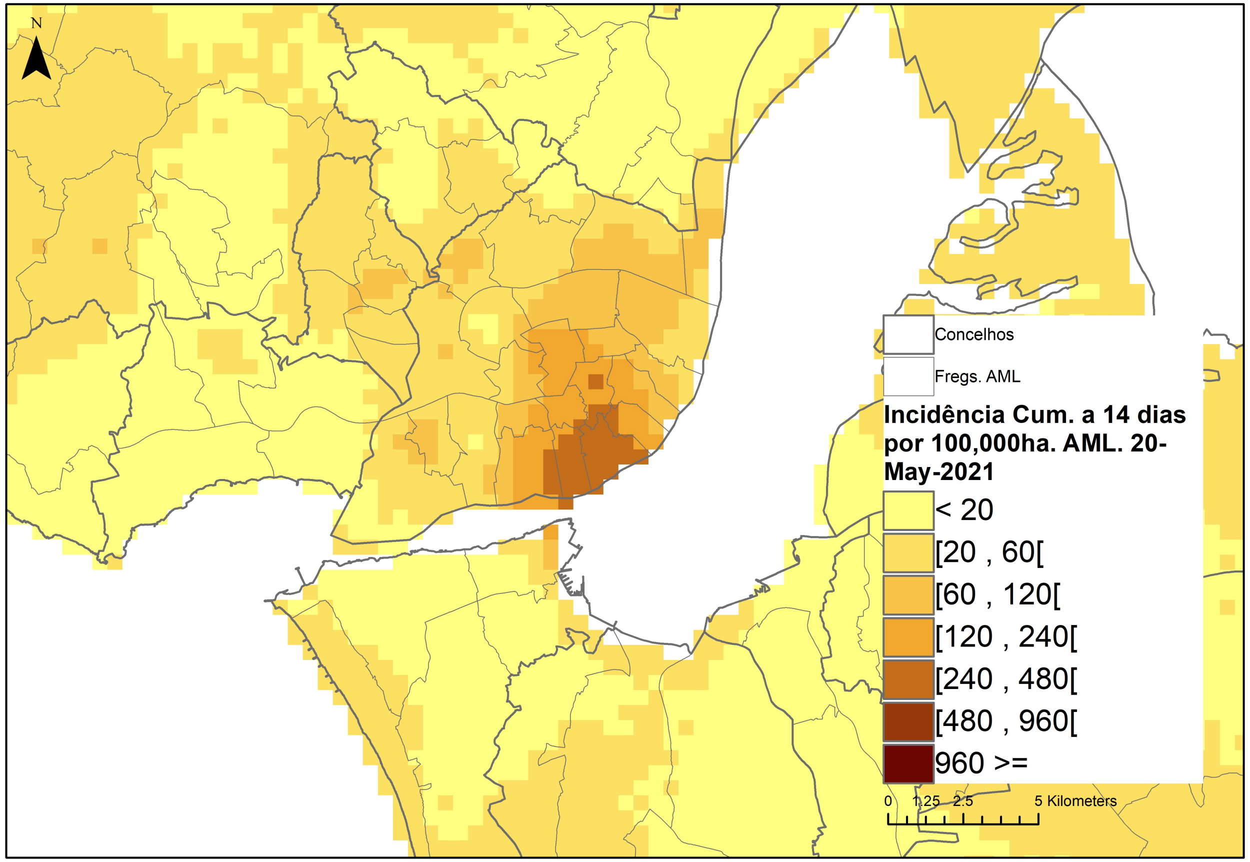


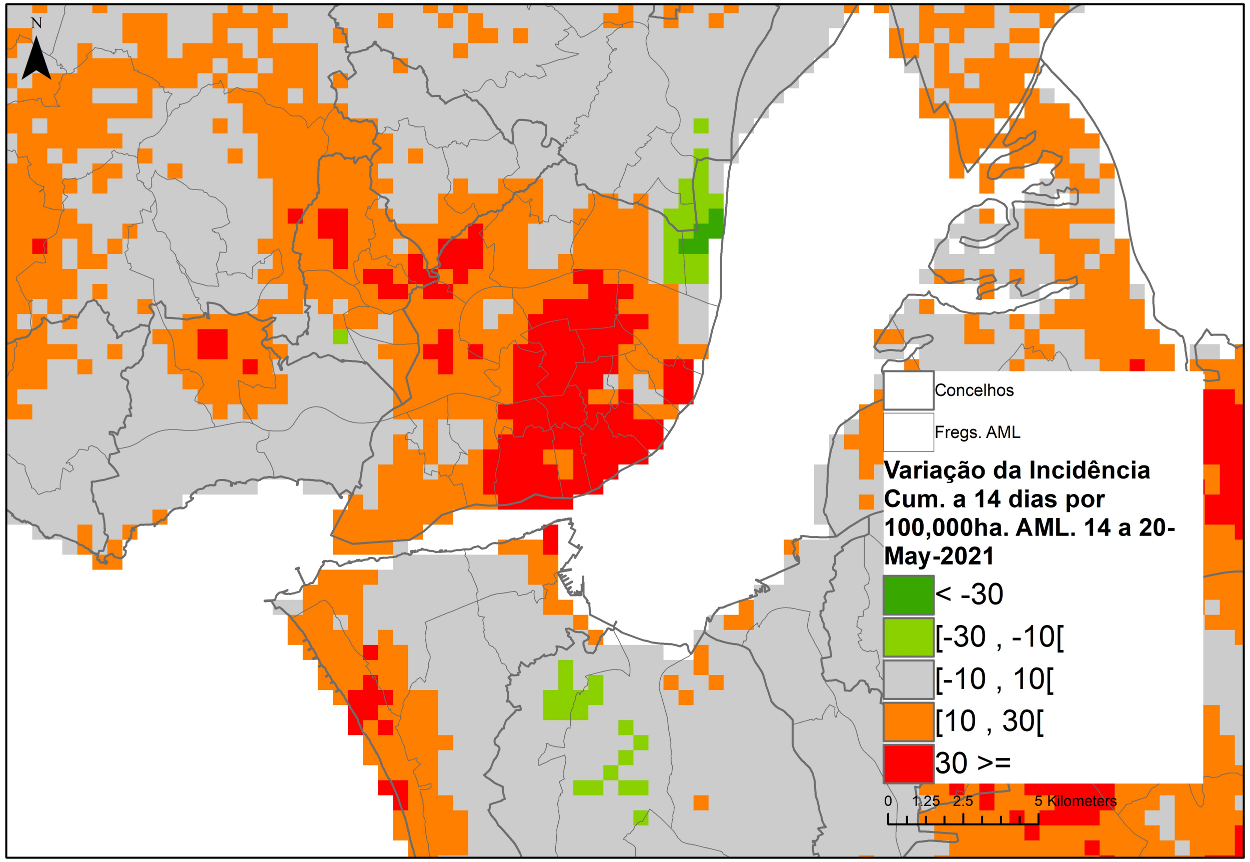
f)



g)

Lisbon metropolitan area @ parish level





Ongoing projects

- **SMOCK-** Spatial Modelling for mapping COVID-19 risk
  - Funded by: Research4COVID, 6 months, started 01.07.2020

- **SCOPE-**Spatial Data Science Services for COVID-19 Pandemic
  - Funded by: AI4COVID - Data Science and Artificial Intelligence in the Public Administration to strengthen the fight against COVID-19 and future pandemics 3 years, started 01.03.2021

Partners

TÉCNICO LISBOA

CERENA
Centro de Recursos Naturais e Ambiente

CEG-IST
CENTRO DE ESTUDOS DE GESTÃO DO INSTITUTO SUPERIOR TÉCNICO | TÉCNICO LISBOA

Interactive Technologies Institute
LARSyS

Instituto Nacional de Saúde Doutor Ricardo Jorge

inesc id lisboa

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

DGS desde 1899
Direção-Geral da Saúde

IPST
Instituto Português do Sangue e da Transplantação, IP

TÉCNICO LISBOA

CERENA
Centro de Recursos Naturais e Ambiente

Ciência espacial de dados para auxiliar na gestão da pandemia COVID-19

Maria João Pereira, Manuel Ribeiro, Amílcar Soares, Ana F. Duarte, Leonardo Azevedo

leonardo.azevedo@tecnico.ulisboa.pt

@leoap

encontro CIÊNCIA '22
ENCONTRO COM A CIÊNCIA E TECNOLOGIA EM PORTUGAL
16-18 maio

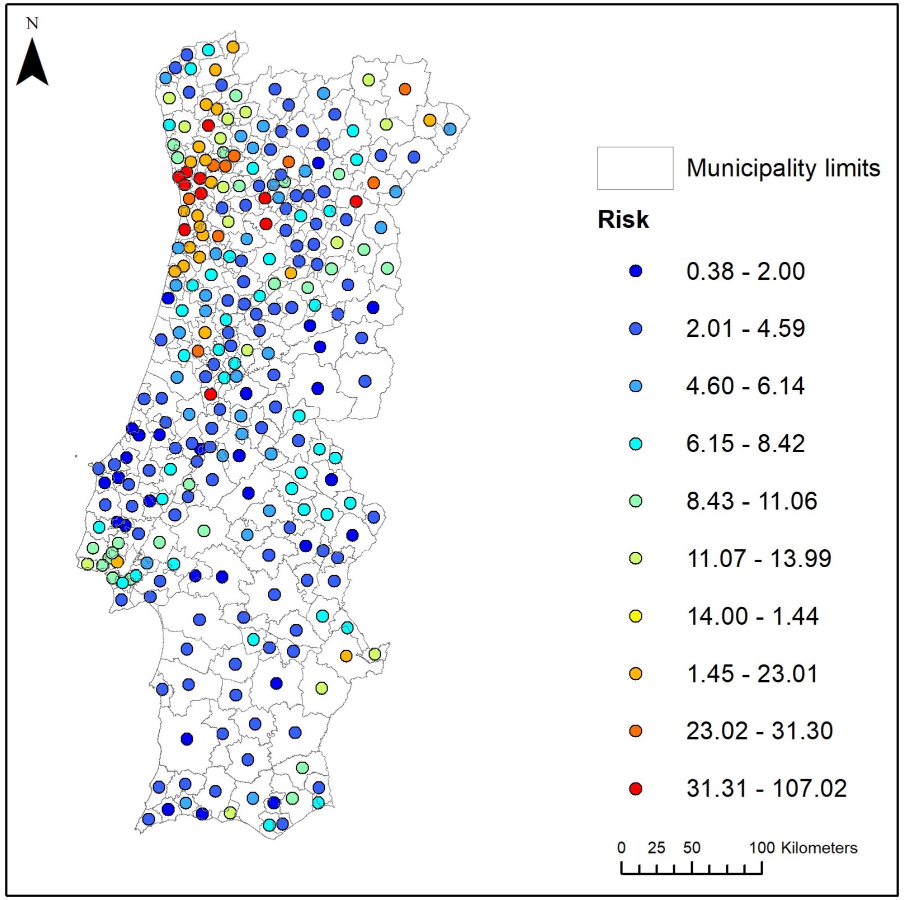
Incidence

$c(\boldsymbol{u}_\alpha)$ - the number of new infections recorded in each municipality on a period of consecutive 14 days, referenced by its geometric centroid $\boldsymbol{u}_\alpha$

$n(\boldsymbol{u}_\alpha)$ - the size of the population at risk (i.e., resident population of a given municipality) at each location $\boldsymbol{u}_\alpha$

$z(\boldsymbol{u}_\alpha)$ - incidence

$$z(\boldsymbol{u}_\alpha) = \frac{c(\boldsymbol{u}_\alpha)}{n(\boldsymbol{u}_\alpha)}$$



Poisson model

$c(\boldsymbol{u}_\alpha)$ is a realization of a Poisson random variable $C(\boldsymbol{u}_\alpha)$, with distribution parameter that is the “expected number of counts per unit of time”.

This parameter is the product of the population size, $n(\boldsymbol{u}_\alpha)$, and the local risk, $R(\boldsymbol{u}_\alpha)$, with expected mean m.

The expectation of risk at any location is equal to the expectation of the incidence

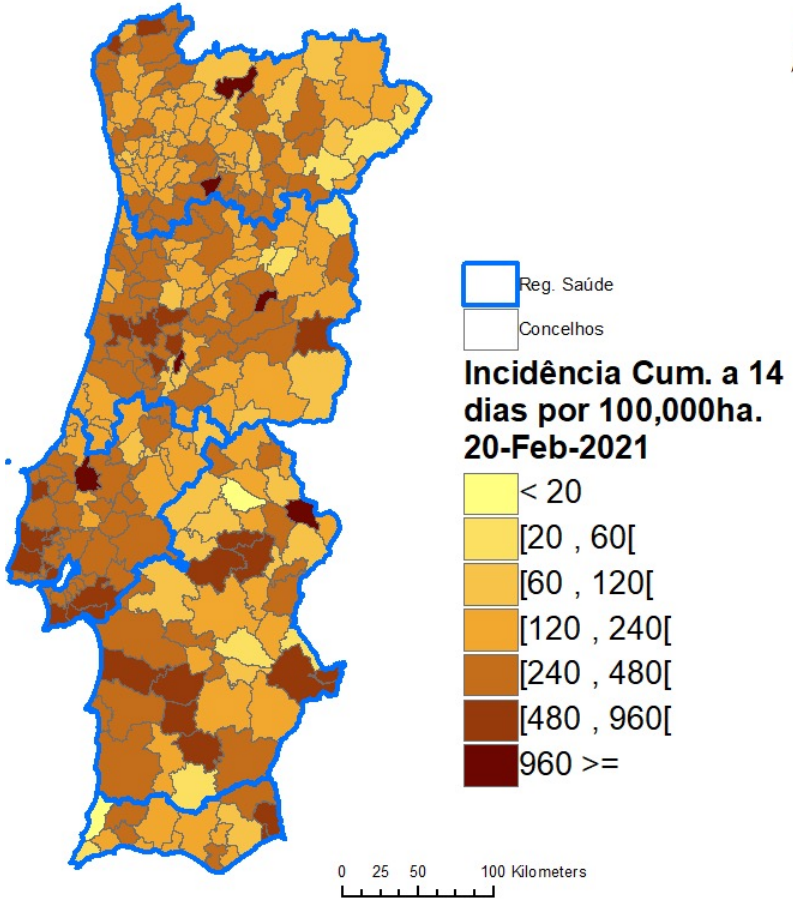
$$E[Z(\boldsymbol{u}_\alpha)] = E[R(\boldsymbol{u}_\alpha)] = m$$

and the risk variance is equal to the incidence variance plus an error term related to the size of the population,

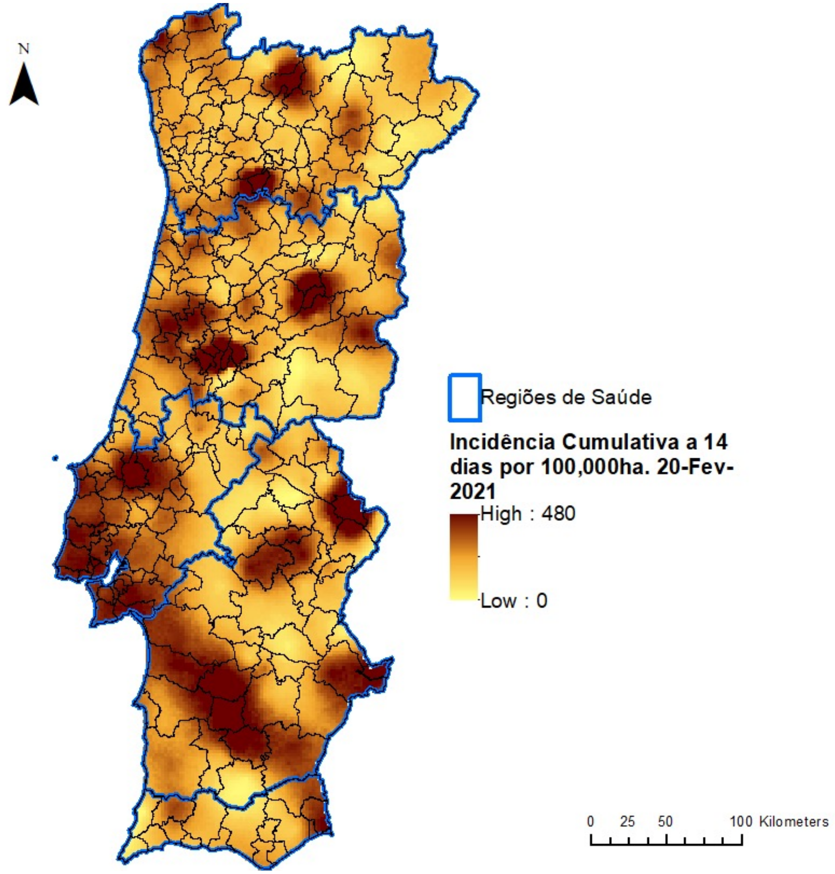
$$Var[Z(\boldsymbol{u}_\alpha)] = Var[R(\boldsymbol{u}_\alpha)] + E[R(\boldsymbol{u}_\alpha)/n(\boldsymbol{u}_\alpha)] = \sigma_R^2 + m/n(\boldsymbol{u}_\alpha)$$

Change of support

Block support



Point support



Direct Sequential Simulation (Soares, 2001)

1. Define a random path that visits each node, x, of the simulation grid;
2. For each node, x, search the conditioning data;
3. Calculate the local covariance values, build and solve kriging system to obtain the local mean and variance kriging estimate at location x;
4. Draw a value from the global probability distribution function centered at the local mean and bounded by the local variance obtained in previous step;
5. Add the simulated value to the data set and repeat all steps until all grid nodes are simulated for one realization;

Repeat until a given pre-defined number of realizations are generated

Block-DSS (Liu and Journel, 2009)

Define a random path that visits each node, x, of the simulation grid;

For each node, x, search the conditioning data;

Calculate the local covariance values, build and solve kriging system to obtain the local mean and variance kriging estimate at location x;

Draw a value from the global probability distribution function centered at the local mean and bounded by the local variance obtained in iii);

Add the simulated value to the data set and repeat all steps until all grid nodes are simulated for one realization;

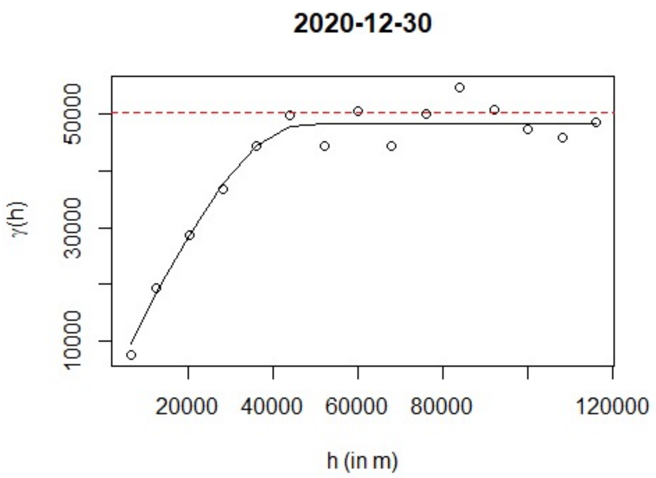
Repeat until a given pre-defined number of realizations are generated

Semivariogram of COVID-19 incidence

- Weighted experimental semivariogram

$$\gamma(\boldsymbol{h}) = \frac{1}{2\sum_{\alpha=1}^{N(\boldsymbol{h})} w(\boldsymbol{h})} \sum_{\alpha=1}^{N(\boldsymbol{h})} \{w(\boldsymbol{h})\,[z(\boldsymbol{v}_\alpha) - z(\boldsymbol{v}_\alpha + \boldsymbol{h})]^2\}$$

$$w(\boldsymbol{h}) = \frac{n(\boldsymbol{v}_\alpha)\, n(\boldsymbol{v}_\alpha + h)}{n(\boldsymbol{v}_\alpha) + n(\boldsymbol{v}_\alpha + h)}$$



Block kriging

block data, B($\boldsymbol{v}_\alpha$), are defined as the spatial linear average of point values, $Z(\mathbf{x}')$, within the block volume, $\boldsymbol{v}_\alpha$:

$$B(\boldsymbol{v}_\alpha) = \frac{1}{|\boldsymbol{v}_\alpha|} \int_{\boldsymbol{v}_\alpha} L_\alpha\big(Z(\boldsymbol{x}')\big) d\boldsymbol{x}' \qquad \forall\, \alpha$$

$$Z^*_{SK}(\mathrm{x}_{\mathbf{u}}) - m = \sum\nolimits_\alpha \lambda_\alpha\,(x_u).\,[z(x_\alpha) - m] + \sum\nolimits_\beta \lambda_\beta\,(x_u).\,[B_v(x_\beta) - m]$$

Kriging system $\begin{bmatrix} \lambda_\alpha \\ \lambda_\beta \end{bmatrix} = \begin{bmatrix} C(.,.), C(.,v) \\ C(v,.), C(v,v) \end{bmatrix}^{-1} \cdot \begin{bmatrix} C(.,xu) \\ C(v,xu) \end{bmatrix}$

Integrating data uncertainty derived from population size

Assuming block errors, r, are homoscedastic and not cross-correlated, with zero mean and known variance, then the covariance between two errors located at $\boldsymbol{v}_\alpha$ and $\boldsymbol{v}_\beta$ is:

$$C\left[r(\boldsymbol{v}_\alpha), r\left(\boldsymbol{v}_\beta\right)\right] = \begin{cases} \sigma_R^2(\boldsymbol{v}_\alpha) & if\ \boldsymbol{v}_\alpha = \boldsymbol{v}_\beta \\ 0\ if\ \boldsymbol{v}_\alpha \neq \boldsymbol{v}_\beta \end{cases}$$

If errors are independent of the signal, uncorrelated and with known variance, then:

$$C(v, v) = \begin{cases} C(\boldsymbol{v}_\alpha, \boldsymbol{v}_\beta) + \sigma_R^2(\boldsymbol{v}_\alpha) & if\ \boldsymbol{v}_\alpha = \boldsymbol{v}_\beta \\ C\left(\boldsymbol{v}_\alpha, \boldsymbol{v}_\beta\right) if\ \boldsymbol{v}_\alpha \neq \boldsymbol{v}_\beta \end{cases}$$

Spatial Data Analysis