Incidentalità stradale in Lombardia

Corso di Formazione: Elementi di geo-statistica e microsimulazione spaziale

25 Giugno 2024

Andrea Gilardi, Francesca Ieva, Laura Sangalli, Piercesare Secchi

PoliS Lombardia

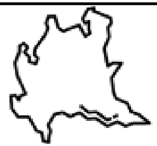
Regione Lombardia

Introduction

Goals:

- Introduce a series of **statistical methods** to describe and analyse road accident data.
- **Explore** the spatio-temporal car crashes **hotspots** in Milan and on Via Olimpica, years 2018 - 2021.
- Pinpoint areas and periods of **high incidence**, examining the **safety patterns’** variations along the years.

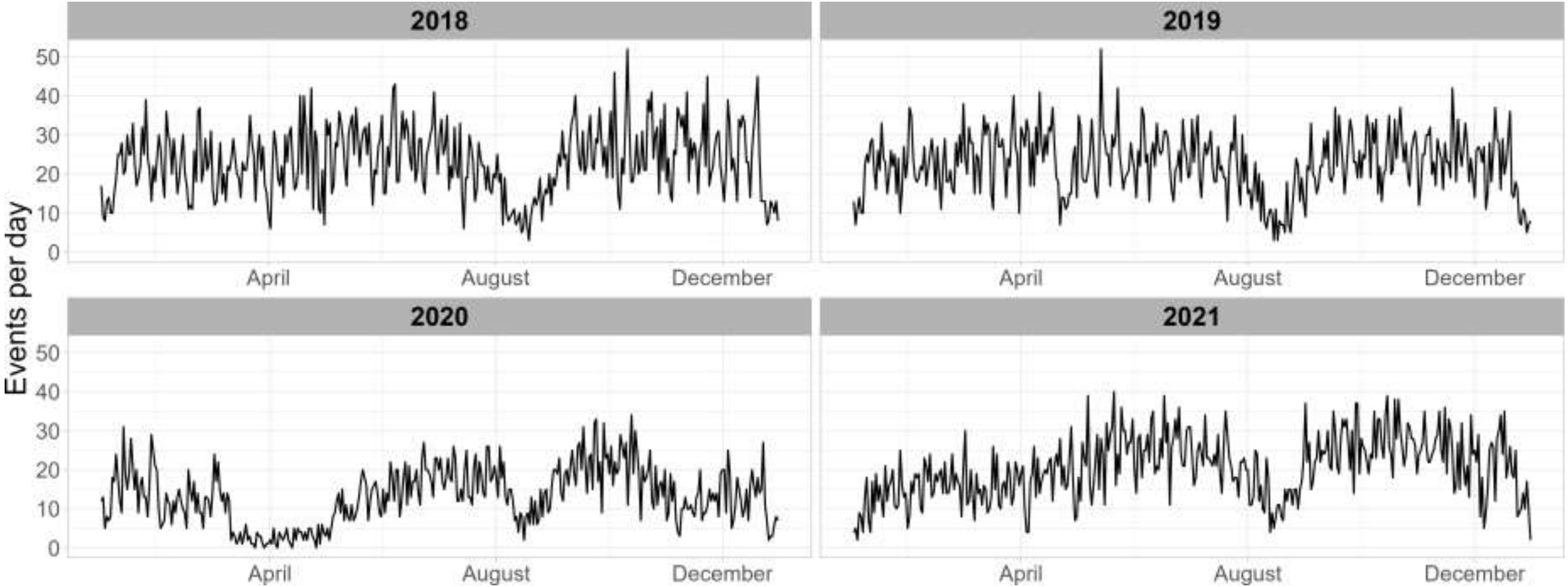
Data sources

	Spatial domain	Temporal domain	Sources
Car crashes		2018 - 2021	PoliS Lombardia
Ambulance interventions		2018 - 2020	AREU
Open Street Map		Continuous updates	OpenStreetMap

Spatio-temporal analysis - Milan

Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.
POLITECNICO DI MILANO

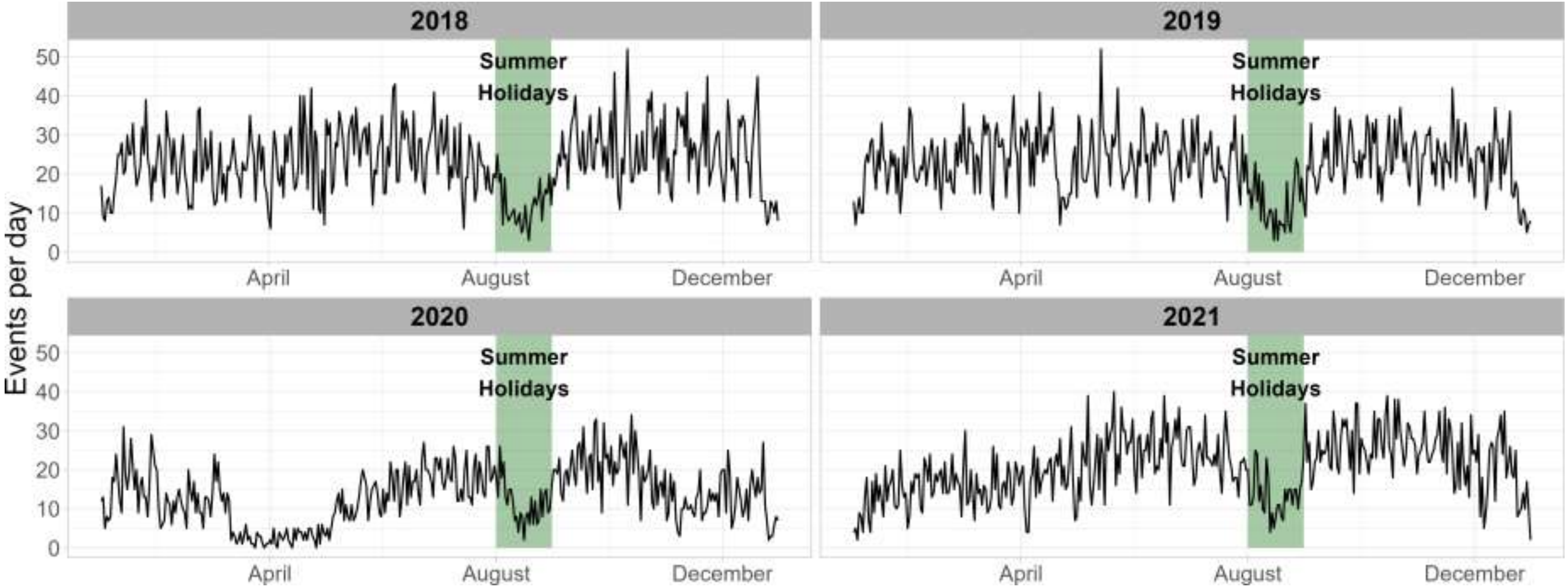
Temporal analysis - Milan



Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

POLITECNICO DI MILANO

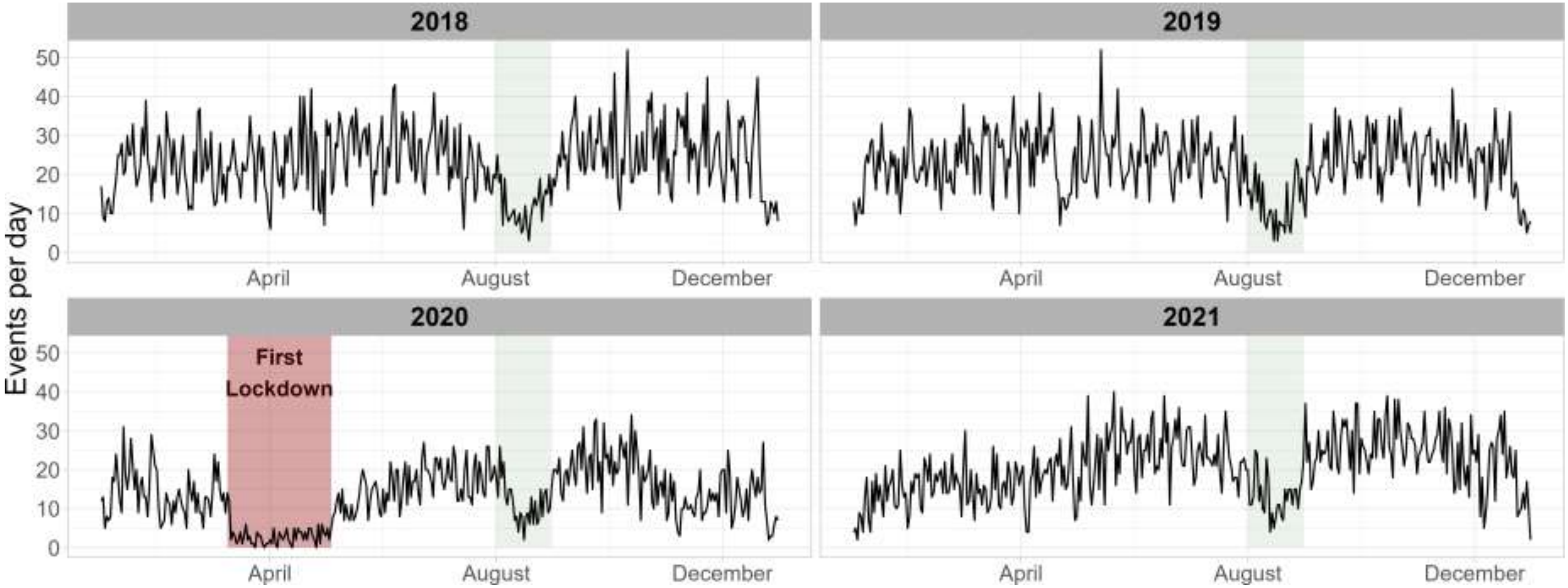
Temporal analysis - Milan



Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

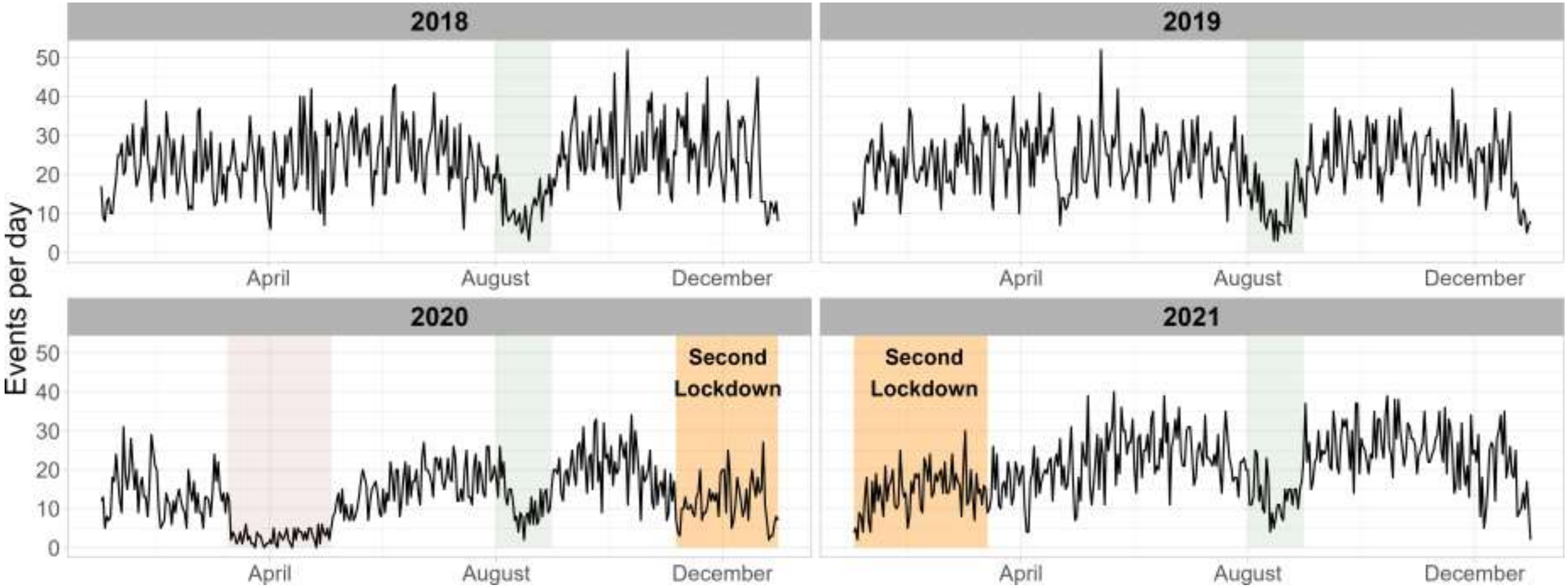
POLITECNICO DI MILANO

Temporal analysis - Milan



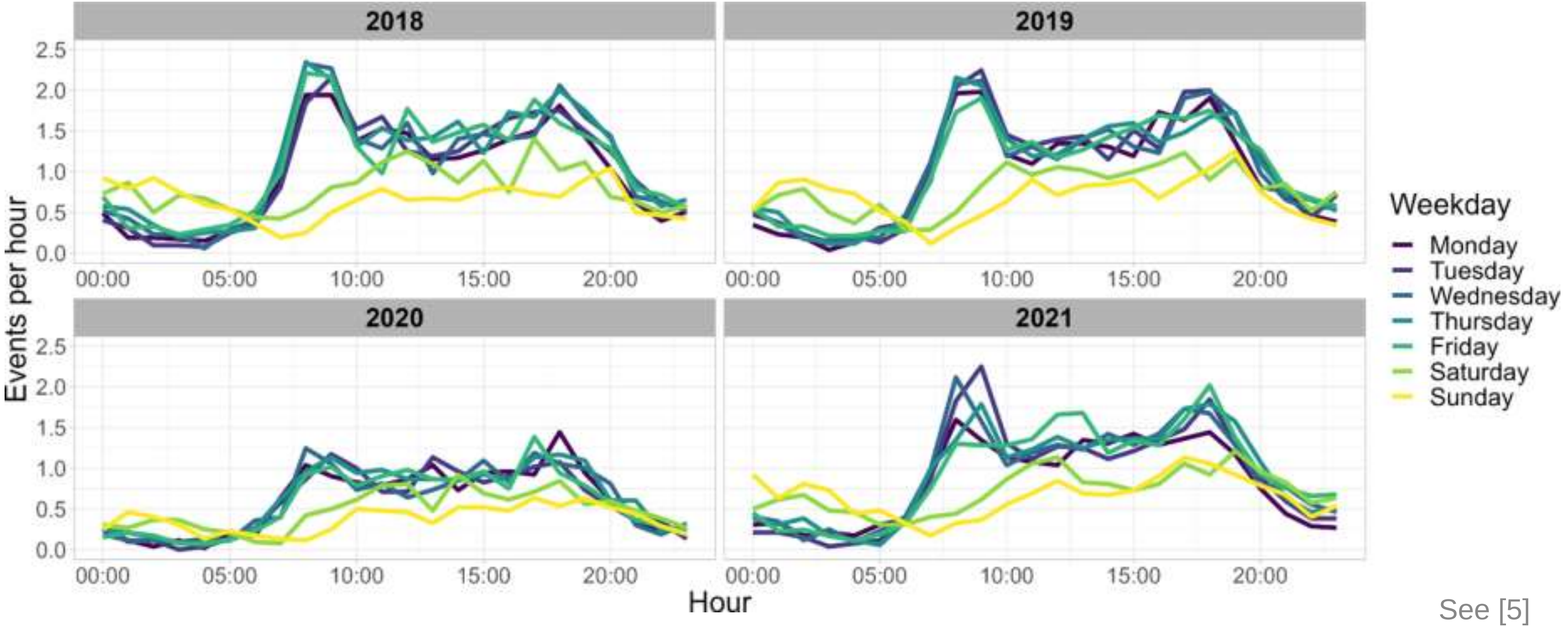
Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Temporal analysis - Milan



Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

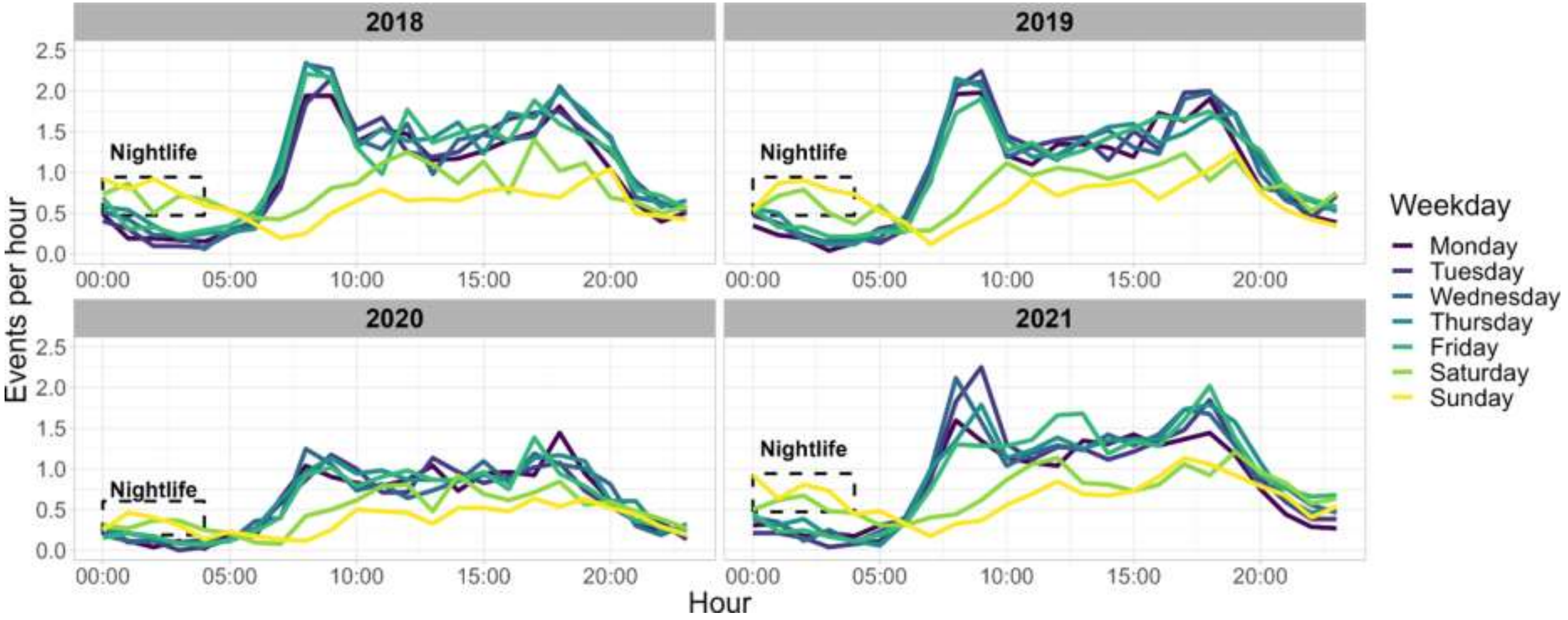
Temporal analysis - Milan



See [5]

Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

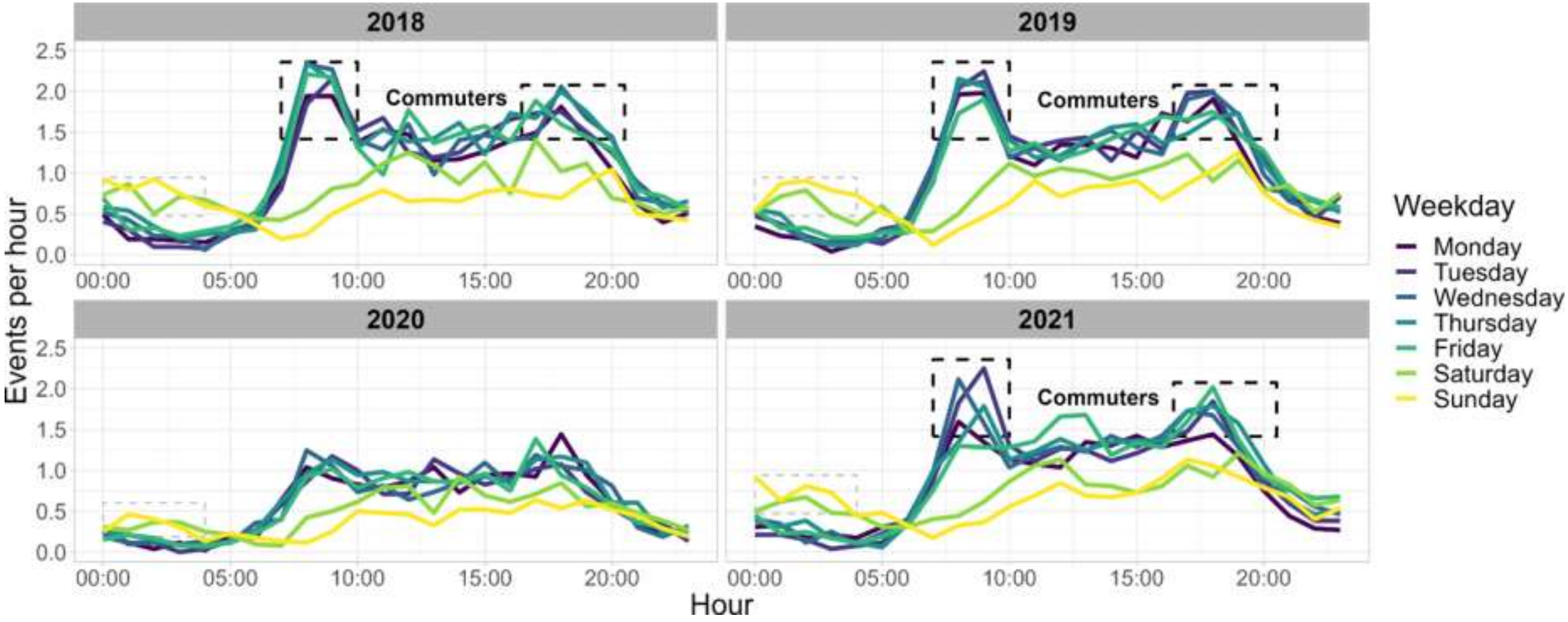
Temporal analysis - Milan



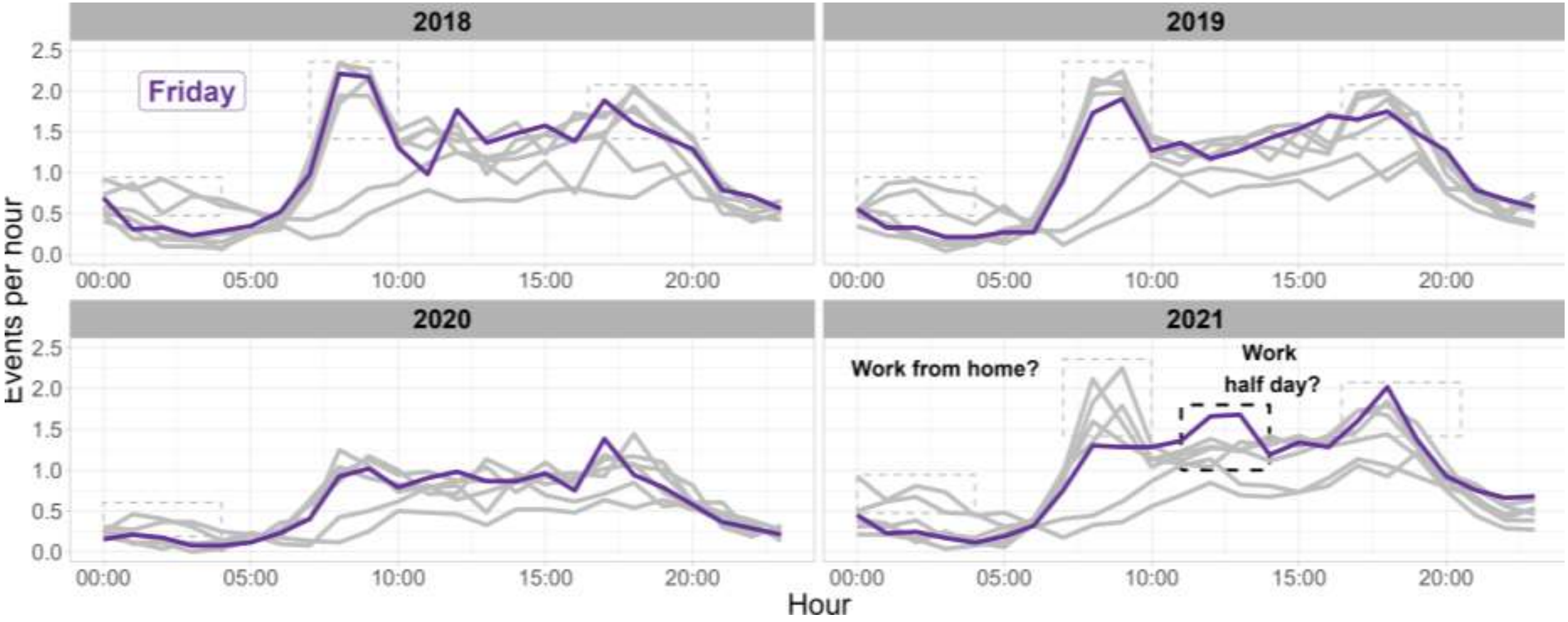
Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

POLITECNICO DI MILANO

Temporal analysis - Milan



Temporal analysis - Milan



Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Take-home messages

1. Weekly seasonal effects due to **summer holidays** in August, which are also present in 2020 and 2021

2. **Drastic reductions** due to COVID-19 lockdowns

3. Hourly seasonal components linked to **nightlife** and **commuters** are clearly evident in 2018, 2019, and 2021.

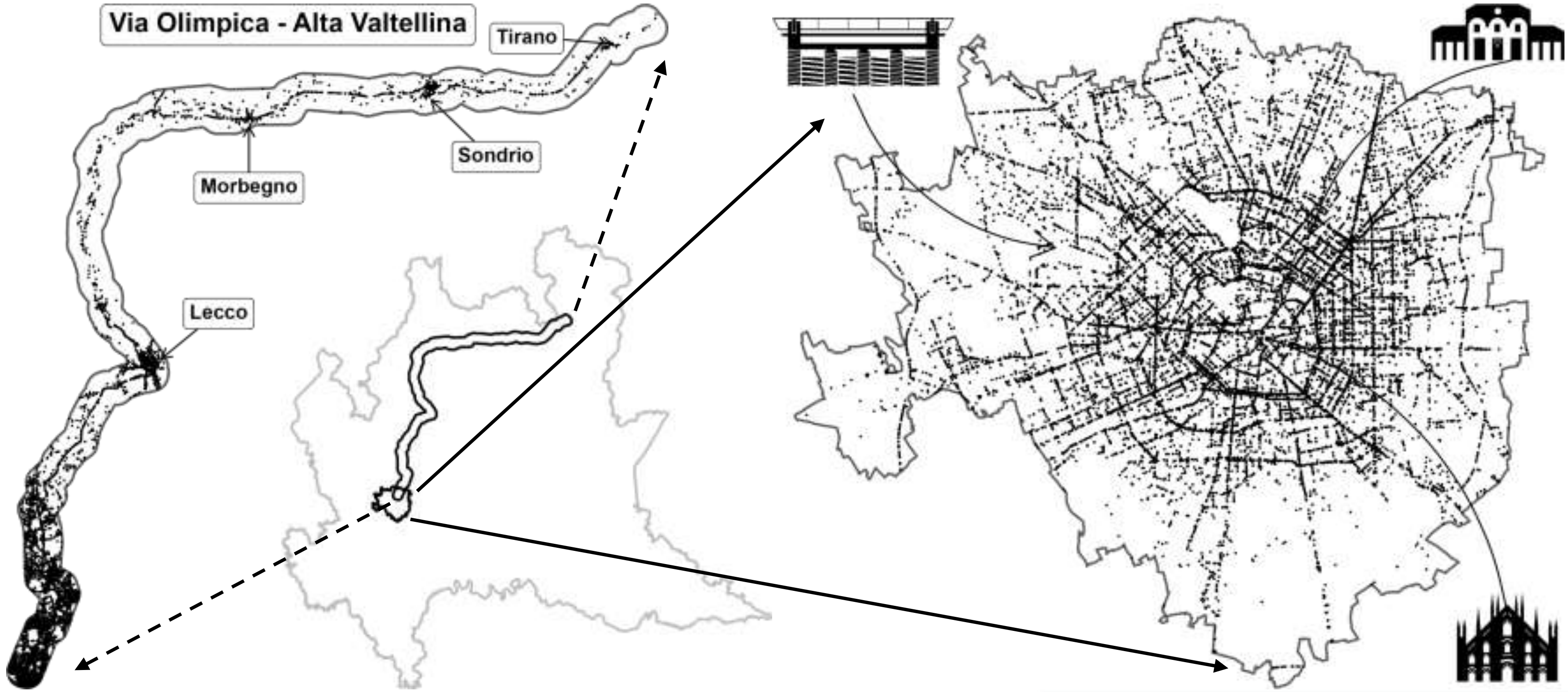
Take-home messages

4. These hourly patterns are **almost absent** in 2020

5. Nevertheless, the behaviour on **friday** during 2021 is substantially different from 2018 and 2019.

Work-from-home policies adopted by large firms near Milan?

Spatial representation



Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Spatial Statistics - Theory

Road accidents represent a classic example of **spatial point patterns**: sets of **points** place inside an **observation window (W)**

Typical statistical analyses involve:

1. Detection of spatial trends/**hotspots**;
2. Identification of **clustering behaviours**
3. Exploring the relationship with **external covariates** (e.g. the road type)

Spatial Statistics - Theory

- Given a spatial point pattern $X = \{x_1, \ldots, x_n\}$, we can define its **intensity** function $\lambda(s), s \in W$ as

Intensity function at location $s \in W$

Expected number of events in proximity of s

$$\lambda(s) = \lim_{|d(s)| \to 0} \frac{\mathbb{E}[N(d(s))]}{|d(s)|}$$

Area of a region surrounding s

Spatial Statistics - Theory

- $\boldsymbol{\lambda}(\boldsymbol{s})$ represents the **expected number of events** per unit area that we may observe in **proximity** of $\boldsymbol{s}$

- When $\lambda(s)$ is **constant** for all $s \in W$, we say that the pattern is **homogeneous,** otherwise it's **inhomogeneous**

- We can define the **density** of $\boldsymbol{X}$ as a **normalised** intensity:

$$f(s) = \frac{\lambda(s)}{n}$$

- The usual (non-parametric) **estimator** of the intensity function for a (planar) point pattern is:

The estimate of $\lambda(s)$

$$\hat{\lambda}(s) = \frac{1}{e(s)} \sum_{i=1}^{n} K(s, x_i, h)$$

Spatial Statistics - Theory

- The usual (non-parametric) **estimator** of the intensity function for a (planar) point pattern is:

$K(\cdot)$ is a kernel function (i.e. an arbitrary density function)

The estimate of $\lambda(s)$

$$\hat{\lambda}(s) = \frac{1}{e(s)} \sum_{i=1}^{n} K(s, x_i, h)$$

- The usual (non-parametric) **estimator** of the intensity function for a (planar) point pattern is:

$K(\cdot)$ **is a kernel function (i.e. an arbitrary density function)**

The estimate of $\lambda(s)$

$$\hat{\lambda}(\boldsymbol{s}) = \frac{1}{\boldsymbol{e}(\boldsymbol{s})} \sum_{i=1}^{n} \boldsymbol{K}(s, x_i, \boldsymbol{h})$$

$\boldsymbol{h}$ **is a smoothing parameter known as *bandwidth***

Spatial Statistics - Theory

- The usual (non-parametric) **estimator** of the intensity function for a (planar) point pattern is:

The estimate of $\lambda(s)$

$K(\cdot)$ is a kernel function (i.e. an arbitrary density function)

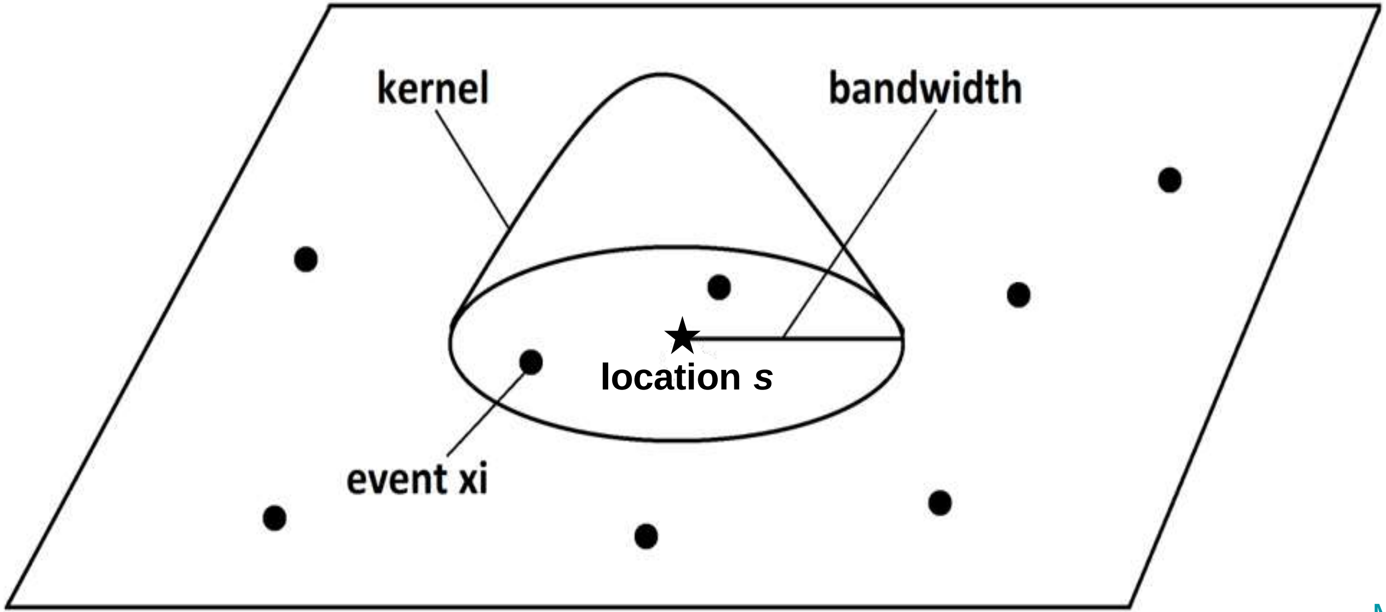
$$\hat{\lambda}(s) = \frac{1}{e(s)} \sum_{i=1}^{n} K(s, x_i, h)$$

An edge-correction term

h is a smoothing parameter known as *bandwidth*

GRINS FOUNDATION

Graphical representation of the estimation process



Courtesy of Moraga (2023)

Spatial Statistics - Theory

- Nevertheless, car crashes represent a classic example of **point patterns on a linear network (L)**

- A proper spatial analysis must always consider the **spatial domain** of the data…

GRINS FOUNDATION

- Therefore, we need to **correct** the usual estimator:

The network-corrected estimate of $\lambda(s)$

The Kernel function (same as before)

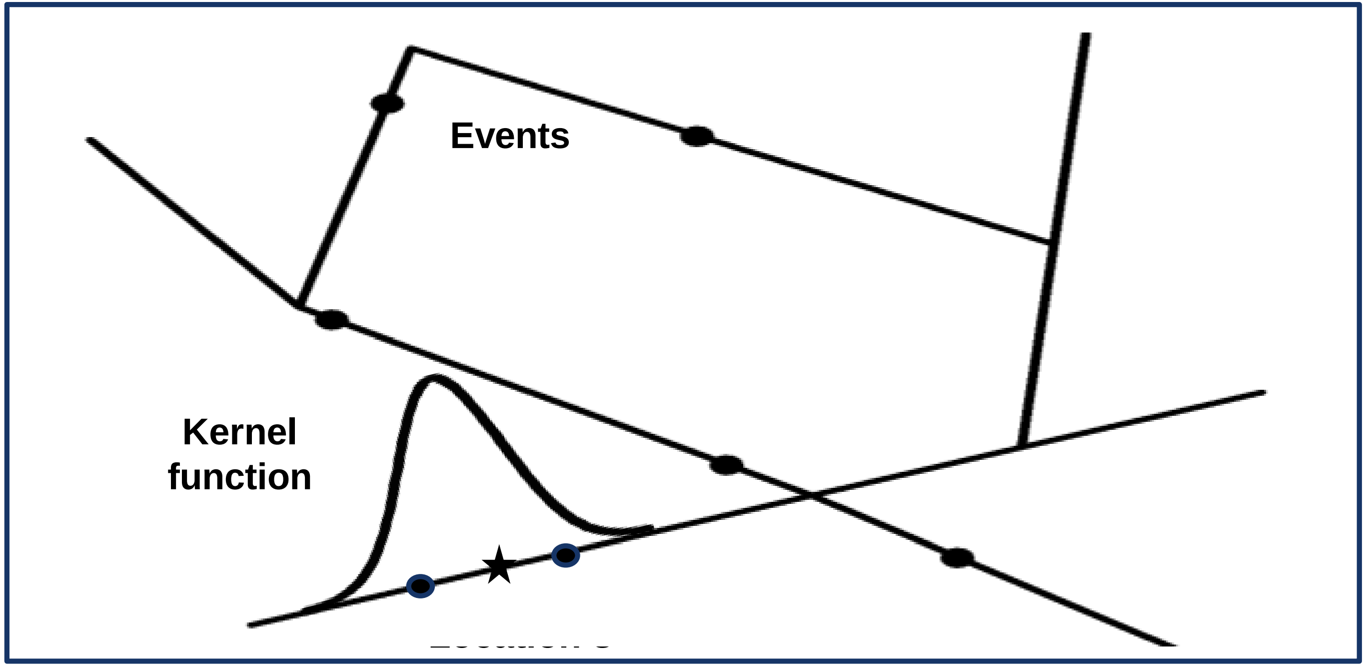
$$\widehat{\lambda_L}(s) = \frac{1}{e(s)} \sum_{i=1}^{n} \frac{K(s, x_i, h)}{c_L(s)}$$

A linear-network correction term

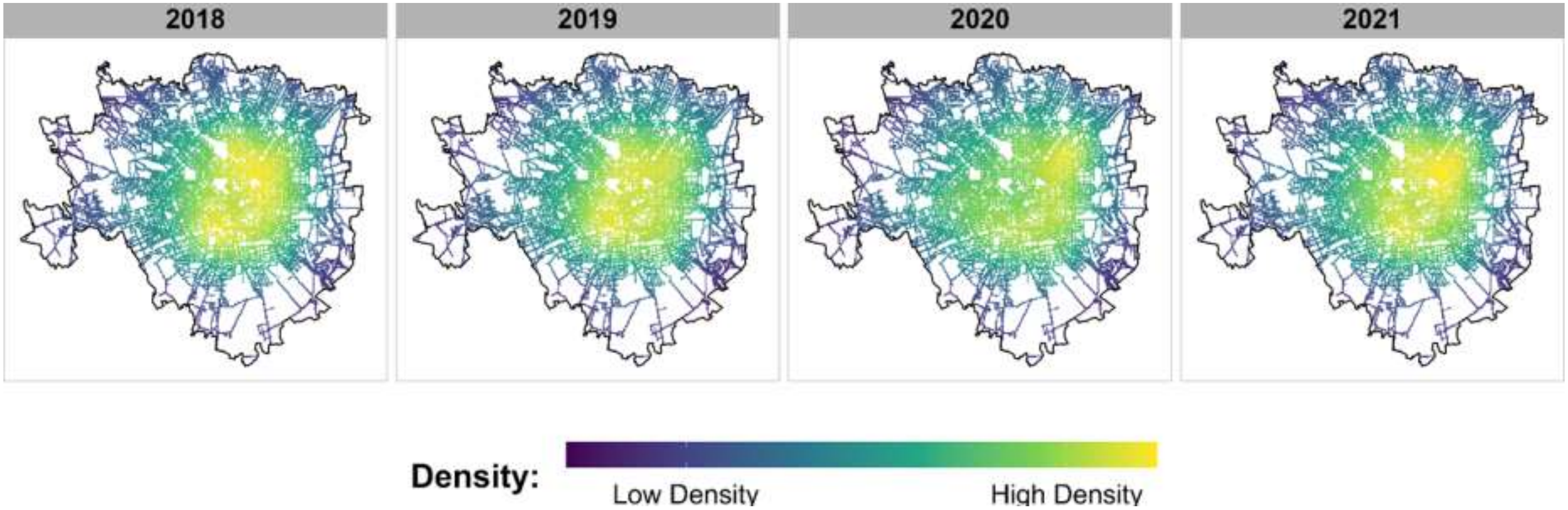
Spatial Statistics - Theory

GRINS FOUNDATION

Graphical representation of the estimation process

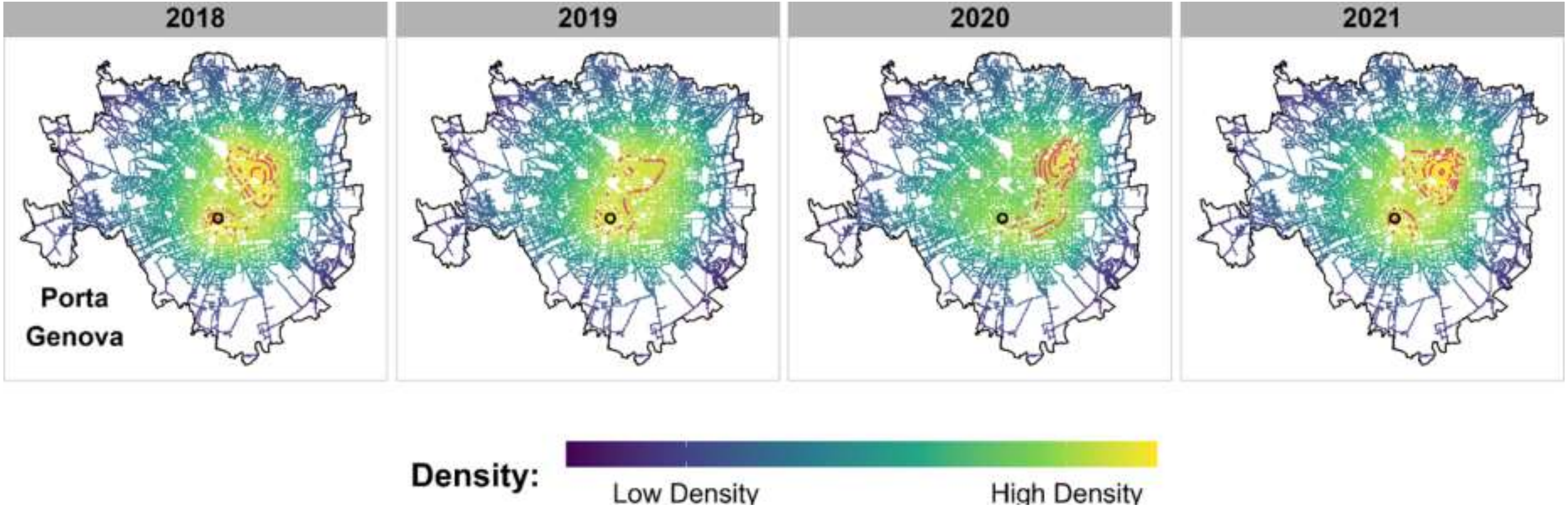


Smoothed density of car crashes



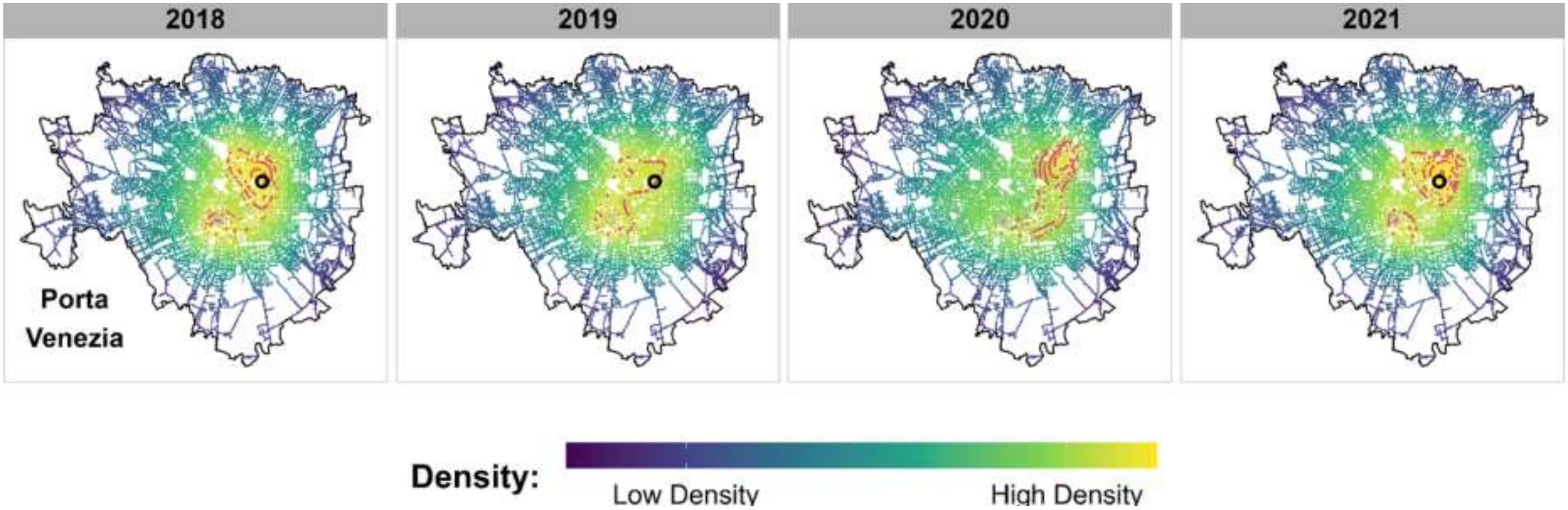
Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Smoothed density of car crashes

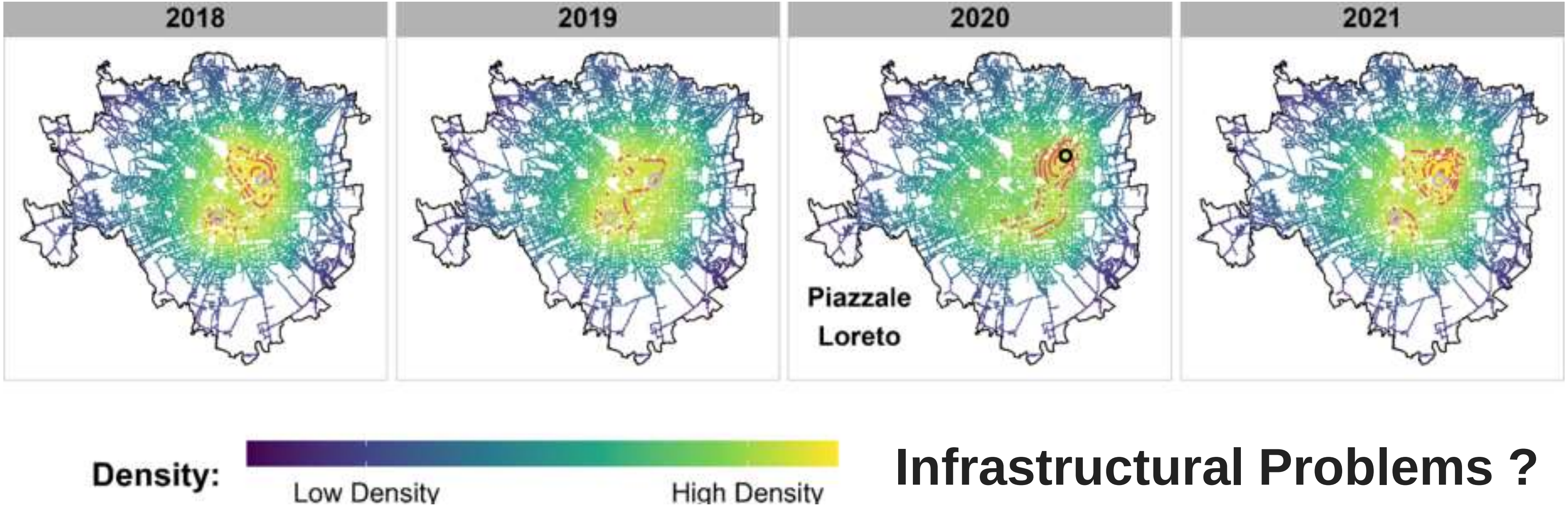


Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Smoothed density of car crashes



Smoothed density of car crashes



Infrastructural Problems ?

Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Spatial Statistics - Theory

- The **spatial relative risk,** $\rho(s)$, is a statistical technique to compare two point-patterns observed at **different times**.

- It can be mathematically defined as:

$$\widehat{\rho_t}(s) = \frac{\hat{f}_t(s)}{\hat{f}_t(s) + \hat{f}_{2018}(s)}$$

Spatial density during year t

Estimate of the spatial relative risk

Spatial density during a reference year (2018)

$$\widehat{\rho_t}(\boldsymbol{s}) = \frac{\hat{f}_t(\boldsymbol{s})}{\hat{f}_t(\boldsymbol{s}) + \hat{f}_{2018}(\boldsymbol{s})}$$

Interpretation:

$$\hat{\rho}_t(\boldsymbol{s}) \simeq 0 \Rightarrow \hat{f}_t(\boldsymbol{s}) \ll \hat{f}_{2018}(\boldsymbol{s})$$

The car crashes' density during year ***t*** at location **s** is **much smaller** than during 2018

$$\widehat{\rho_t}(\boldsymbol{s}) = \frac{\hat{f}_t(\boldsymbol{s})}{\hat{f}_t(\boldsymbol{s}) + \hat{f}_{2018}(\boldsymbol{s})}$$

Interpretation:

$$\hat{\rho}_t(\boldsymbol{s}) \simeq 0.5 \Rightarrow \hat{f}_t(\boldsymbol{s}) \simeq \hat{f}_{\mathbf{2018}}(\boldsymbol{s})$$

The car crashes' density during year ***t*** at location **s** is **almost the same as** during 2018

Spatial Statistics - Theory

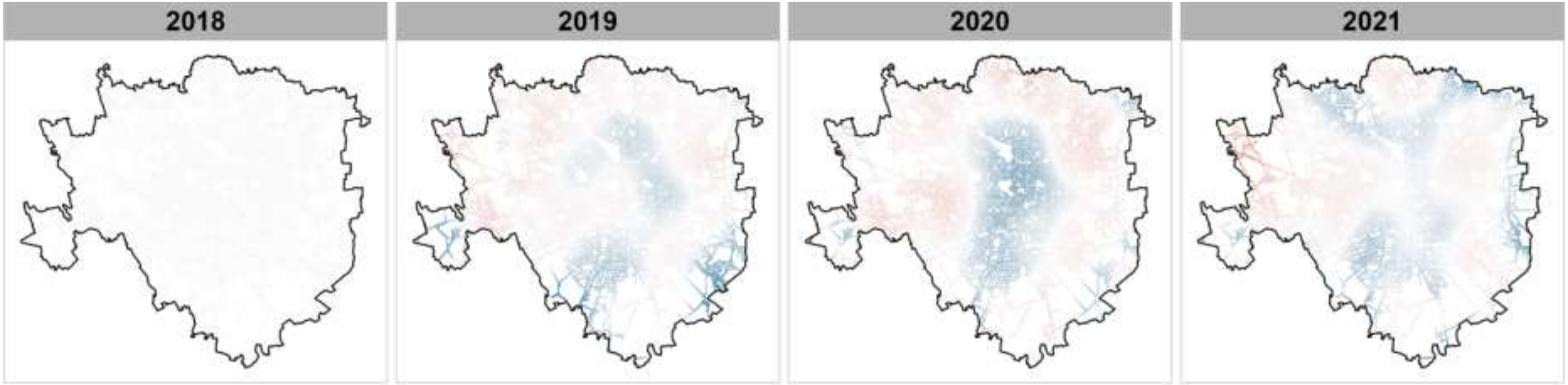
$$\widehat{\rho_t}(s) = \frac{\hat{f}_t(s)}{\hat{f}_t(s) + \hat{f}_{2018}(s)}$$

Interpretation:

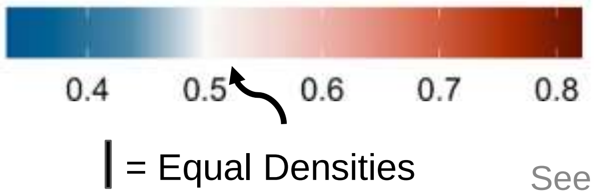
$$\hat{\rho}_t(s) \simeq 1 \Rightarrow \hat{f}_t(s) \gg \hat{f}_{2018}(s)$$

The car crashes' density during year ***t*** at location **s** is **much larger** than during 2018

Ratio of spatial densities



$$\widehat{\rho_t}(s) = \frac{\widehat{f}_t(s)}{\widehat{f}_t(s) + \widehat{f}_{2018}(s)}$$



See [1, 2, 4]

Take-home messages

1. During 2018, 2019, and 2021, the hotspots are located in proximity of **Porta Genova** and **Porta Venezia**.

2. We notice a **peak** of events near **Piazzale Loreto** during 2020.

3. The density-ratio maps showcase a reduction of car crashes density near the city centre in 2020, possibly due to the **lower traffic volumes**.

Spatial Statistics - Theory

- Let $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a point pattern and let $\boldsymbol{\lambda(s)}$ denote its intensity function

- We can also relate $\boldsymbol{\lambda(s)}$ to one or more fixed variables (e.g the road type). For example:

$$\boldsymbol{\lambda(s)} = \boldsymbol{\lambda(s;\theta)} = \mathbf{exp}(\boldsymbol{\theta_0 Z_0} + \boldsymbol{\theta_1 Z_1} + \cdots + \boldsymbol{\theta_p Z_p})$$

- Let $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a point pattern and let $\boldsymbol{\lambda}(\boldsymbol{s})$ denote its intensity function

- We can also relate $\boldsymbol{\lambda}(\boldsymbol{s})$ to one or more fixed variables (e.g the road type). For example:

$$\boldsymbol{\lambda}(\boldsymbol{s}) = \boldsymbol{\lambda}(\boldsymbol{s};\boldsymbol{\theta}) = \mathbf{exp}(\boldsymbol{\theta}_0\boldsymbol{Z}_0 + \boldsymbol{\theta}_1\boldsymbol{Z}_1 + \cdots + \boldsymbol{\theta}_p\boldsymbol{Z}_p)$$

Unknown coefficients (to be estimated)

- Let $\boldsymbol{X} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ be a point pattern and let $\boldsymbol{\lambda}(\boldsymbol{s})$ denote its intensity function

- We can also relate $\boldsymbol{\lambda}(\boldsymbol{s})$ to one or more fixed variables (e.g the road type). For example:

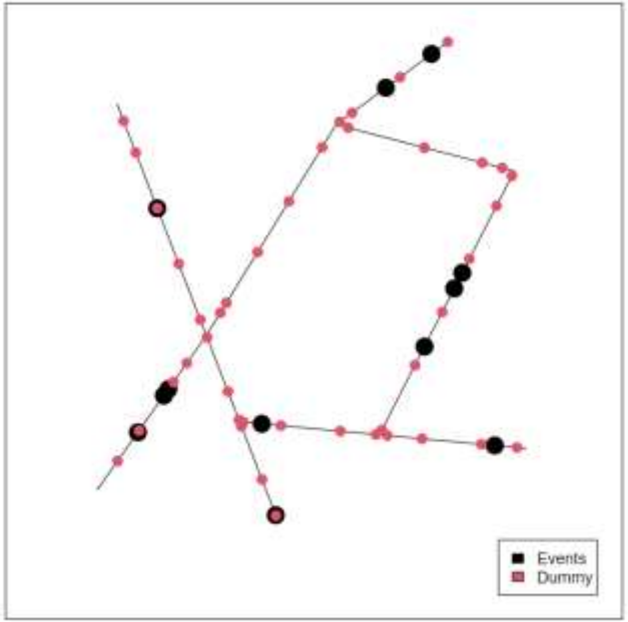
$$\boldsymbol{\lambda}(\boldsymbol{s}) = \boldsymbol{\lambda}(\boldsymbol{s};\boldsymbol{\theta}) = \exp(\boldsymbol{\theta}_0 \boldsymbol{Z}_0 + \boldsymbol{\theta}_1 \boldsymbol{Z}_1 + \dots + \boldsymbol{\theta}_p \boldsymbol{Z}_p)$$

Unknown coefficients (to be estimated)

Known spatial functions

Spatial Statistics - Theory

- The coefficients are estimated employing a techniques known as Berman Turner device:
  - Simulate a set of **dummy points** and *superimpose* them to the events
  - Compute a **weight** for each point after **splitting** the network into smaller parts
  - Calculate the Z_p value in **each segment** and sum the weights
  - Estimate $\theta_1, \ldots, \theta_p$ using standard **GLM** techniques



Deep dive on car crashes and road types

GRINS FOUNDATION

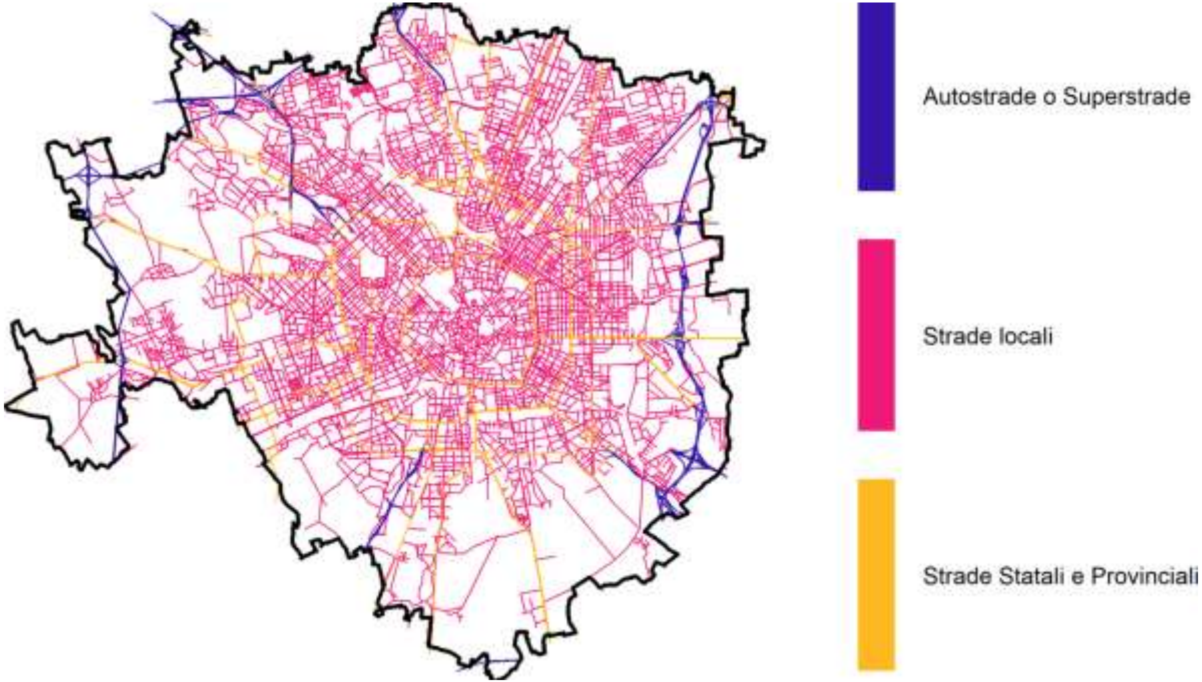
- According to a recent report shared by PoliS: *"La categoria di strada più a rischio incidentalità, lesività, e mortalità è quella delle* ***strade urbane*** *(di Regione Lombardia) dove si concentra il 76,9% di incidenti, il 49,9% di morti e il 73,7% di feriti."*

CATEGORIA STRADA	Incidenti	Morti	Feriti	Indice mortalità	Var.% incidenti 2020-2021	Var.% morti 2020-2021	Var.% feriti 2020-2021
Strade urbane [a]	19.860	178	24.807	0,9	29,3%	20,3%	29,2%
Autostrade	1.474	32	2.275	2,2	51,0%	6,7%	53,4%
Altre strade [b]	4.504	147	6.590	3,3	24,0%	5,8%	25,3%
TOTALE	**25.838**	**357**	**33.672**	**1,4**	**29,4%**	**12,6%**	**29,8%**

(a) Sono incluse nella categoria "Strade urbane" anche le Provinciali, Statali e Regionali entro l'abitato.
(b) Sono incluse nella categoria "Altre strade", le strade Statali, Regionali e Provinciali fuori dall'abitato e Comunali extraurbane.

Deep dive on car crashes and road types

We used two types of **parametric Poisson regression model** to further explore these findings.



Deep dive on car crashes and road types

Ratio of car accident rates **(Reference year: 2018)**

2019	2020	2021
0.97 (0.95; 1.00)	**0.57 (0.55; 0.60)**	0.88 (0.86; 0.91)

Interpretation: On average, the number of car crashes per metre observed in 2020 is **57%** of that observed in 2018. The 95% confidence interval is (0.55; 0.60).

Deep dive on car crashes and road types

***Ratio of car accident rates* (Reference category: Autostrade)**

	2018	2019	2020	2021
Strade Locali	1.32 (1.19; 1.46)	1.32 (1.18; 1.46)	**1.58 (1.37; 1.83)**	1.44 (1.28; 1.62)
Strade Statali e Provinciali	3.83 (3.44; 4.26)	3.70 (3.32; 4.12)	**4.26 (3.66; 4.96)**	4.08 (3.62; 4.59)

Interpretation: On average, the number of car crashes per metre observed in 2020 on *Strade Statali e Provinciali* is **426%** of that observed in *Autostrade e Superstrade* during the same year. The 95% confidence interval is (3.66; 4.96).

Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Take-home messages

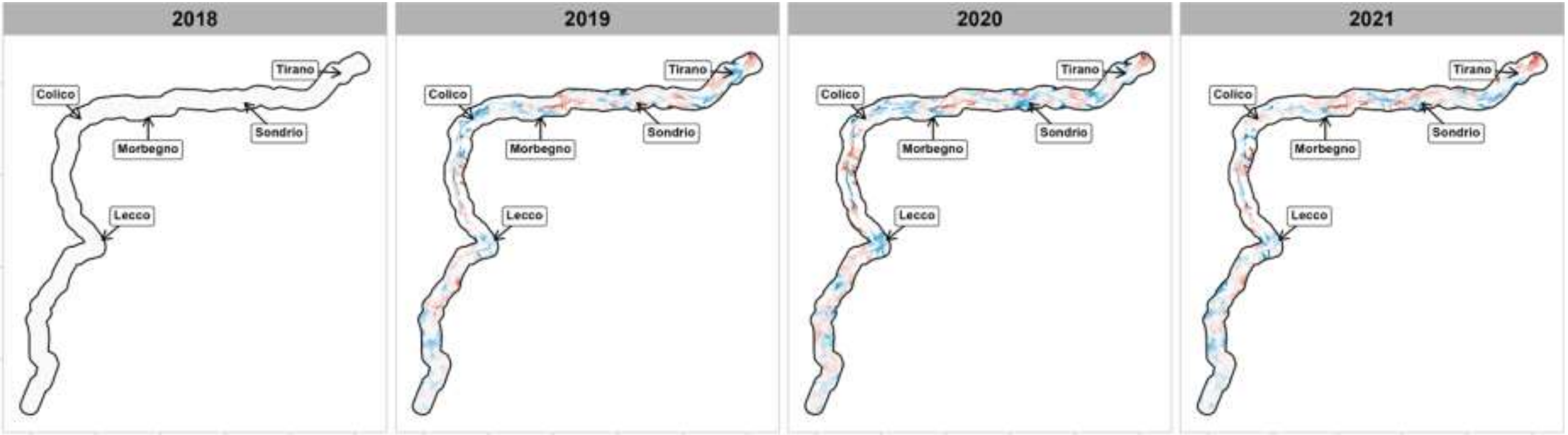
1. The statistical models let us **reproduce** and **replicate** the findings detailed in the report shared by **PoliS**.

2. We can notice that, on average, *Strade Statali* is much **more dangerous** than *Autostrade or Superstrade*.

3. These effects do not seem to be stationary along the years.

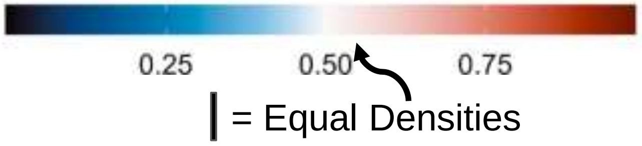
Spatio temporal analysis - Via Olimpica

Ratio of spatial densities



$$\widehat{\rho_t}(s) = \frac{\hat{f}_t(s)}{\hat{f}_t(s) + \hat{f}_{2018}(s)}$$

$\widehat{\rho_t}(s)$:



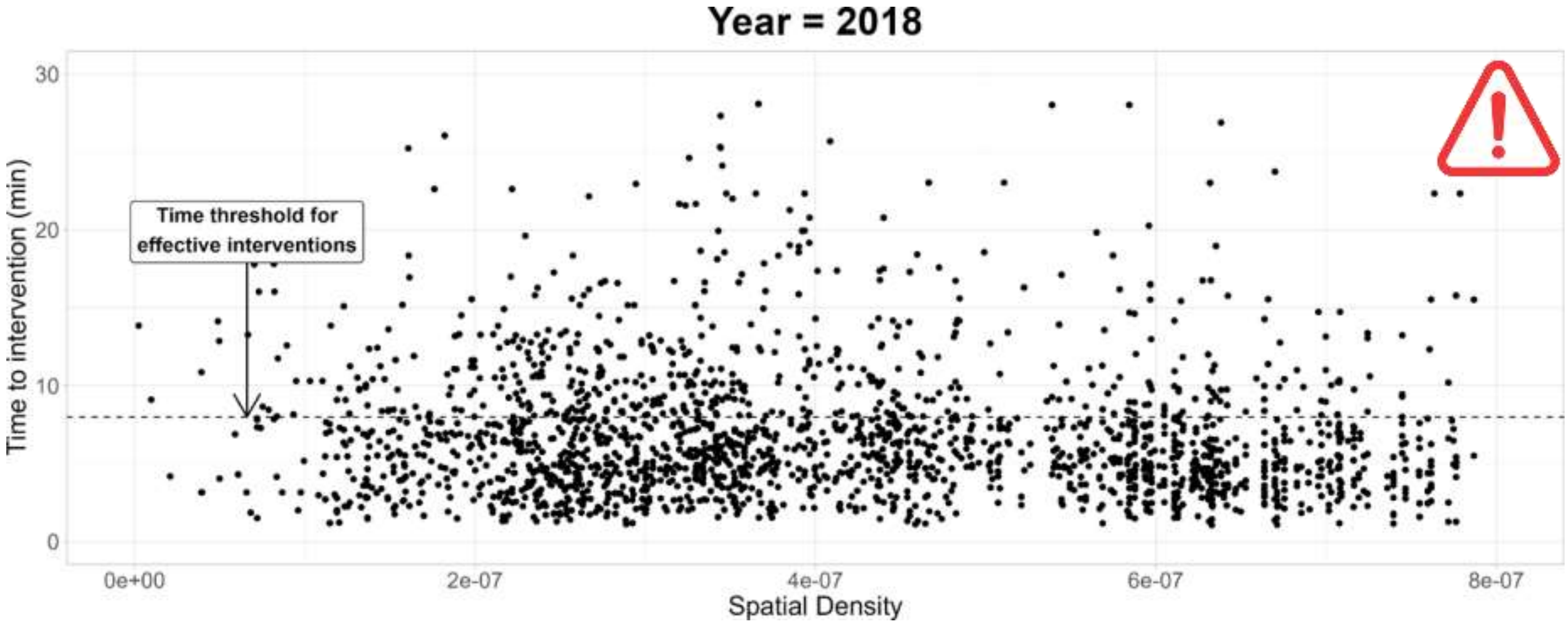
Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Data fusion

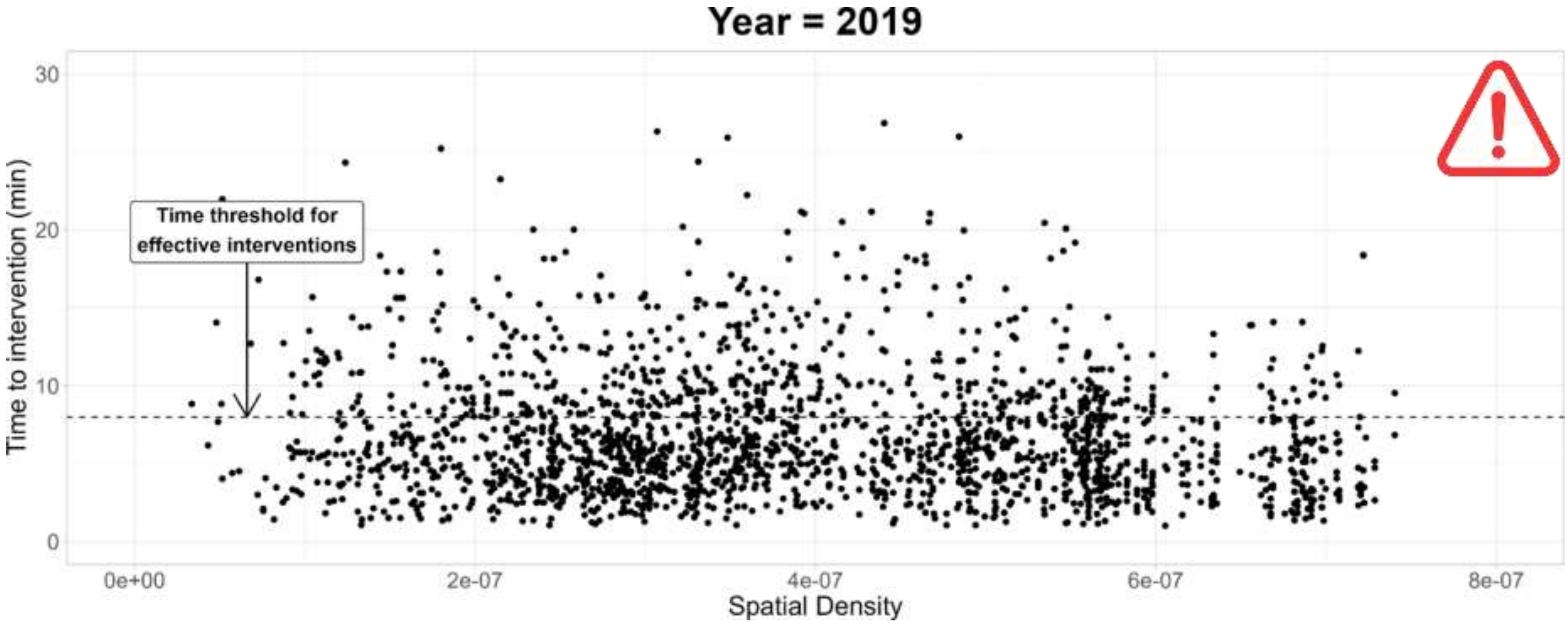
- First, we **merged** the car crashes data with the database of ambulance interventions using a **spatial matching operation**.

- Then, we derived an index that represents the **time** required for the first ambulance to **arrive at the location of the car crash**.

Spatial densities and time to intervention

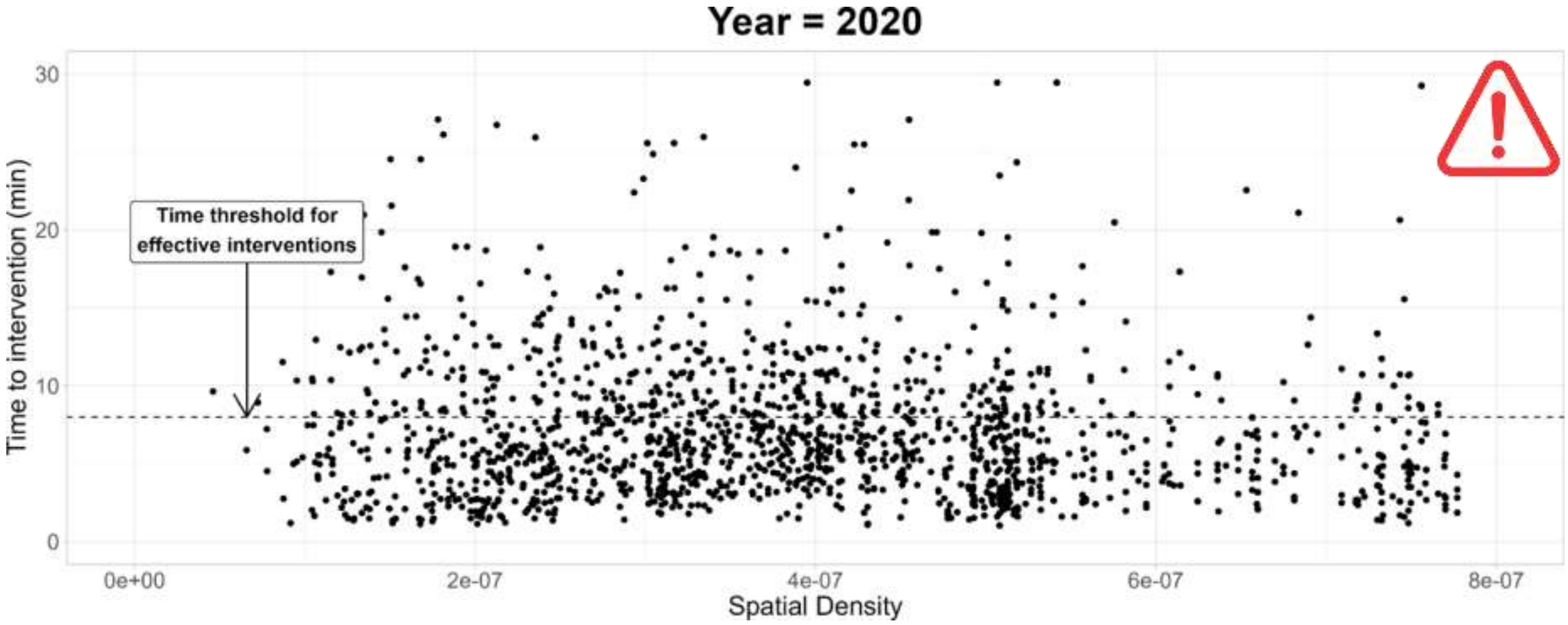


Spatial densities and time to intervention



Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Spatial densities and time to intervention



Take-home messages

1. The spatial ratios appear quite **stable** in the **first half of Via Olimpica**, from Milan to Colico.

2. There are slightly **more pronounced** differences in **Alta Valtellina**, especially near Morbegno or Sondrio.

3. During 2018, several crashes occurred in **proximity of an hotspot** and required **more than 8 minutes** to get an ambulance intervention.

Take-home messages

4. The situation is **remarkably better** in 2019!

5. Unfortunately, the **performances** of the ambulance intervention system **get worse** during 2020.

6. Integrating **PoliS** and **AREU** data let us identify risky areas where we could **optimise** the ambulance response times through **reallocation** algorithms.

Further developments

• Estimation of **road traffic** to measure the exposure [7].	
• Development of marked point process models for **multivariate estimation** of safety indices.	

Acknowledgments

This study was carried out within the GRINS: Growing Resilient, INclusive and Sustainable project, which is part of the Next Generation EU program (National Recovery and Resilience Plan).

References

Seminal and review papers:

[1] Kelsall, J. E., & Diggle, P. J. (1995). Non-parametric estimation of spatial variation in relative risk. *Statistics in medicine*, *14*(21-22), 2335-2342.

[2] McSwiggan, G., Baddeley, A., & Nair, G. (2020). Estimation of relative risk for events on a linear network. *Statistics and Computing*, *30*, 469-484.

[3] Rakshit, S., Davies, T., Moradi, M. M., McSwiggan, G., Nair, G., Mateu, J., & Baddeley, A. (2019). Fast kernel smoothing of point patterns on a large network using two-dimensional convolution. *International Statistical Review*, *87*(3), 531-556.

State of the art:

[4] Baddeley, Adrian, Gopalan Nair, Suman Rakshit, Greg McSwiggan, and Tilman M. Davies. "Analysing point patterns on networks—A review." *Spatial Statistics* 42 (2021): 100435.

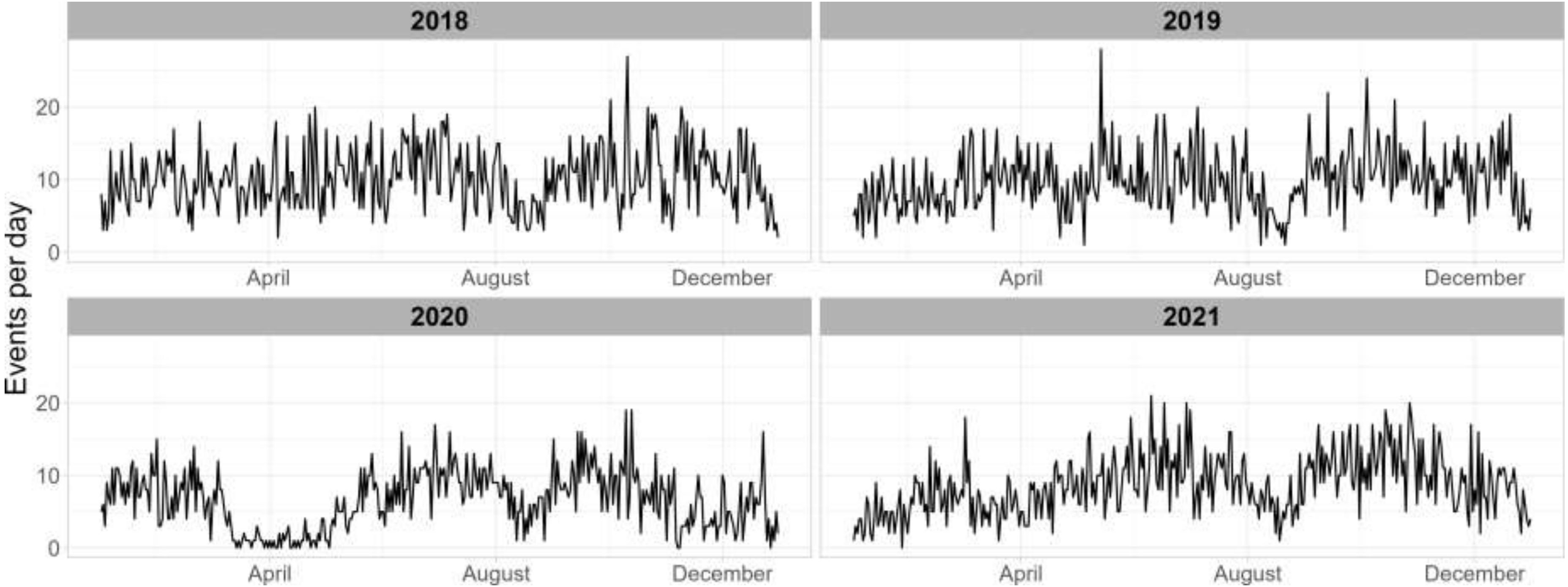
[5] Gilardi, A., Borgoni, R., Mateu, J. (2023+) A non-separable first-order spatio-temporal intensity for events on linear networks: an application to ambulance interventions. Accepted for publication in *Annals of Applied Statistics*.

[6] Ferraccioli, F., Arnone, E., Finos, L., Ramsay, J. O., & Sangalli, L. M. (2021). Nonparametric density estimation over complicated domains. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *83*(2), 346-368.

[7] Gilardi, A., Borgoni, R., Presicce, L., & Mateu, J. (2023). Measurement error models for spatial network lattice data: Analysis of car crashes in Leeds. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *186*(3), 313-334.

Appendix: Temporal analysis along Via Olimpica

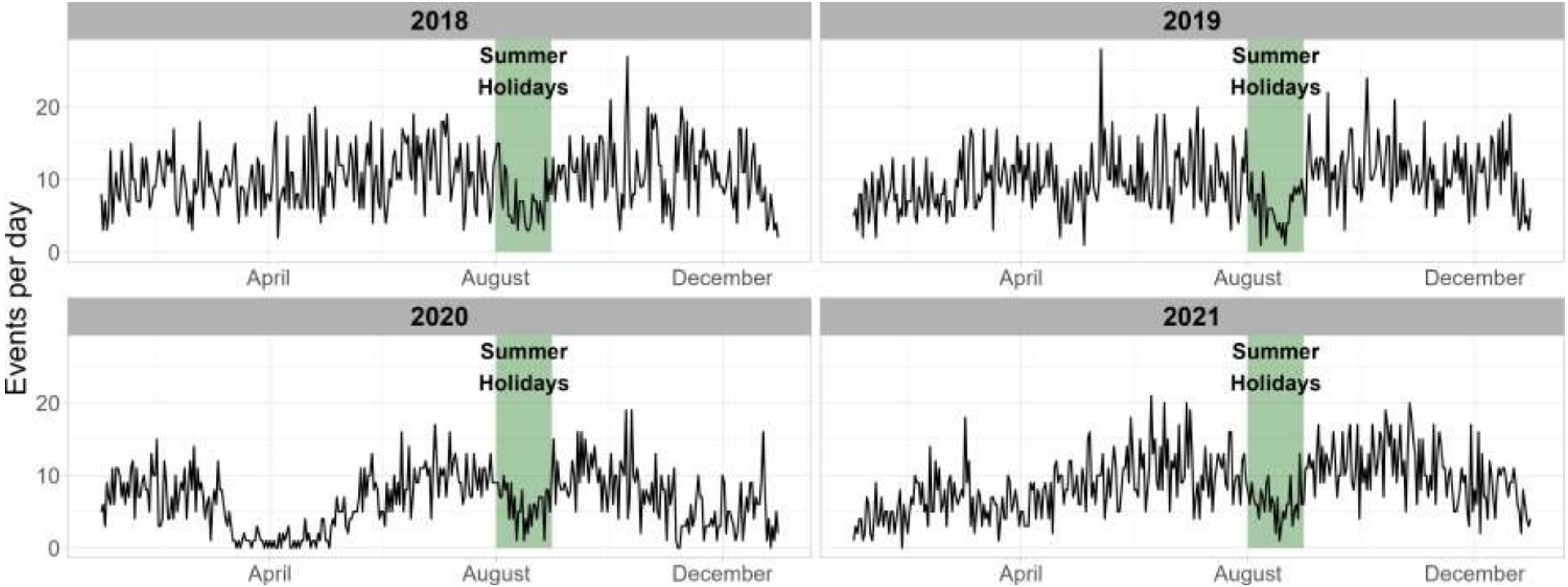
Temporal analysis - Via Olimpica



Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

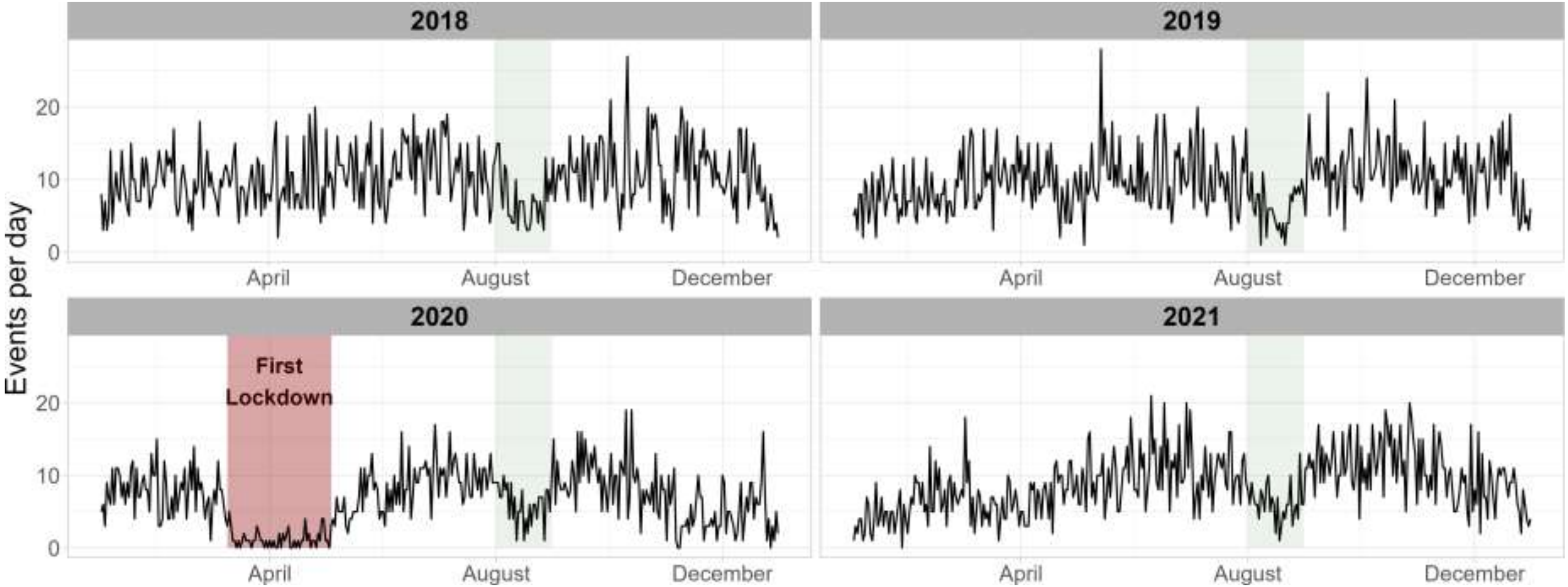
POLITECNICO DI MILANO

Temporal analysis - Via Olimpica



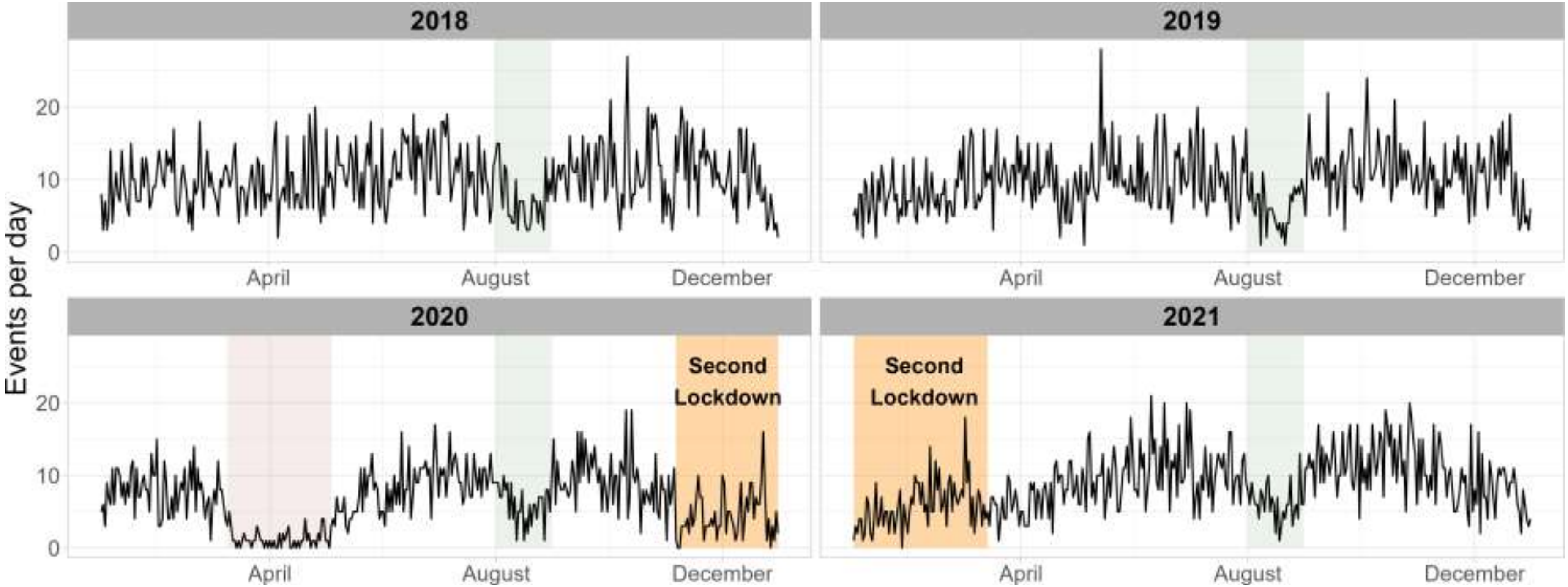
Temporal analysis - Via Olimpica

GRINS FOUNDATION



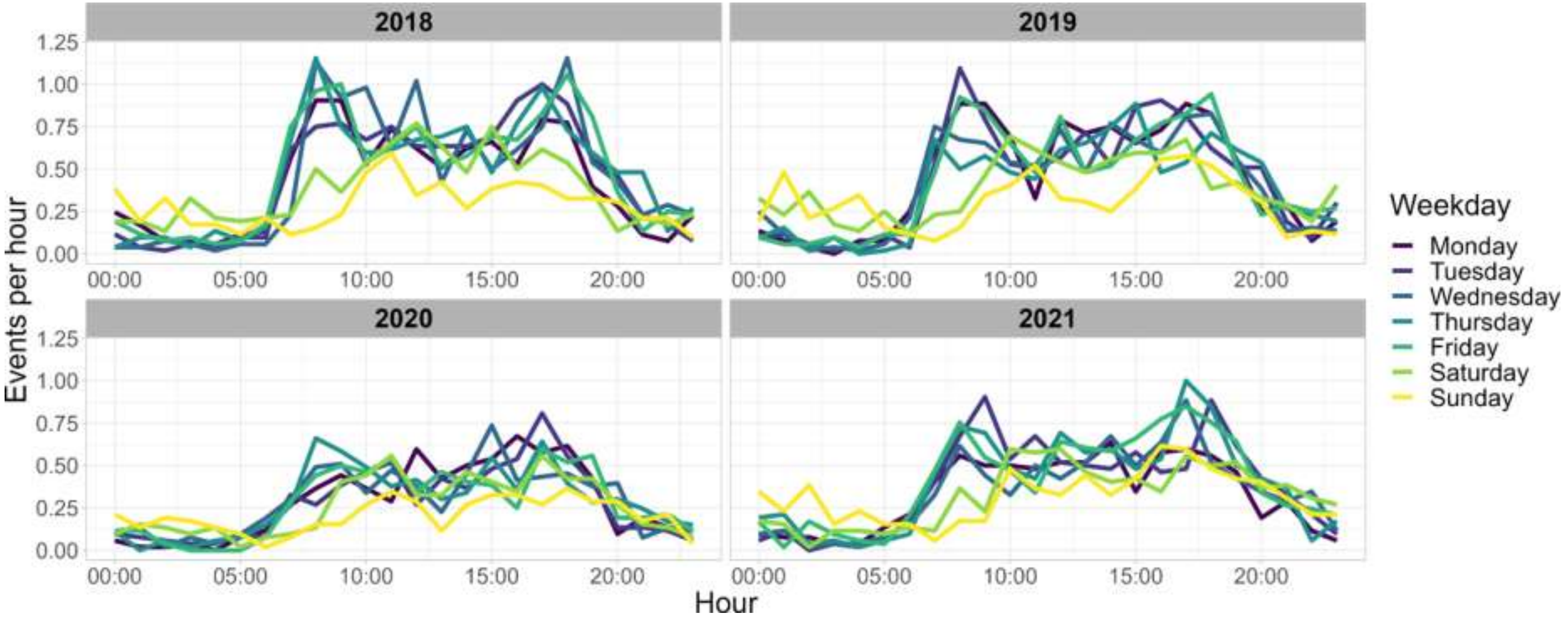
Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Temporal analysis - Via Olimpica



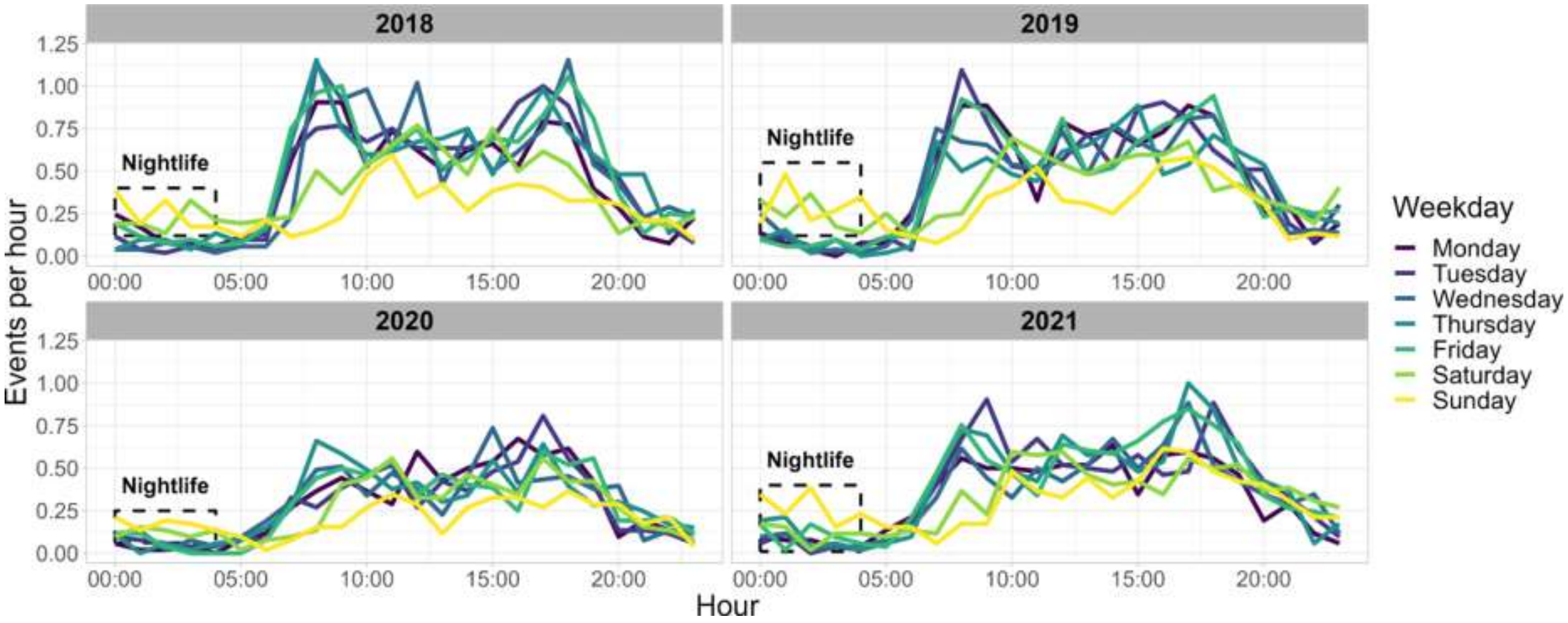
Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Temporal analysis - Via Olimpica



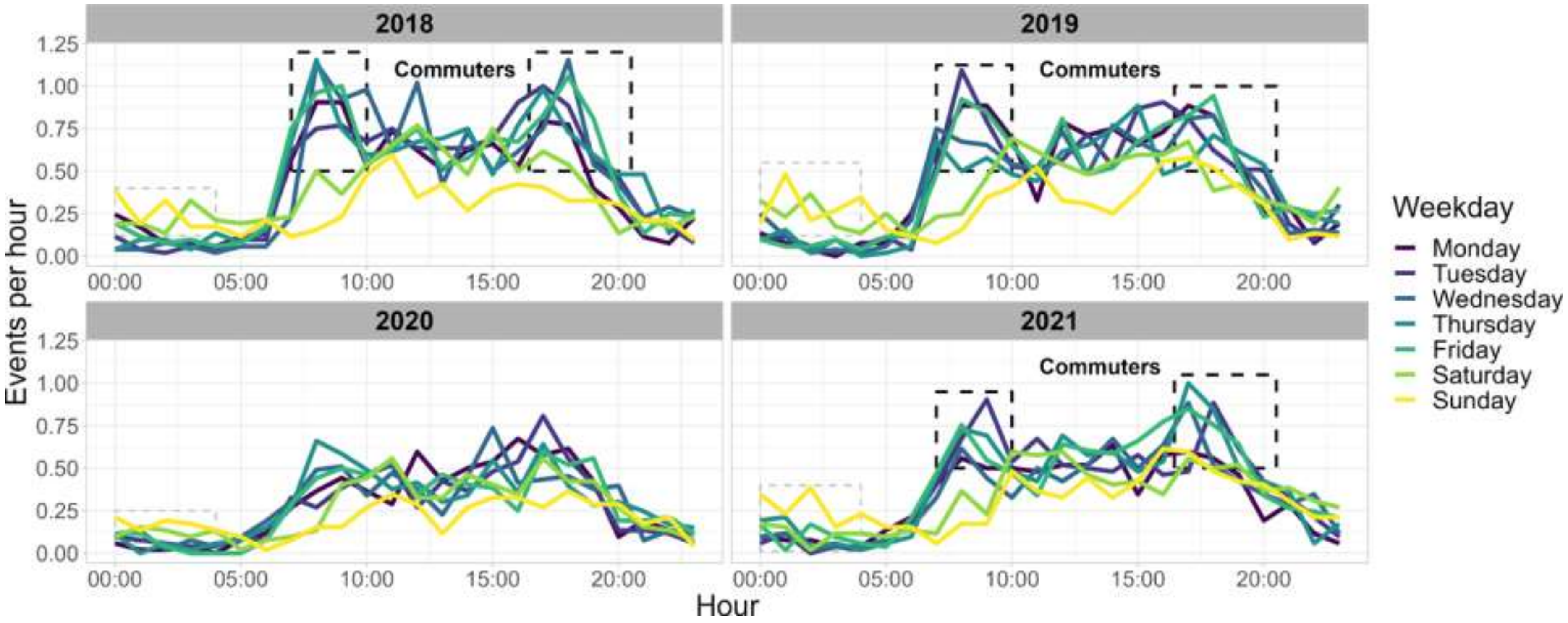
Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Temporal analysis - Via Olimpica



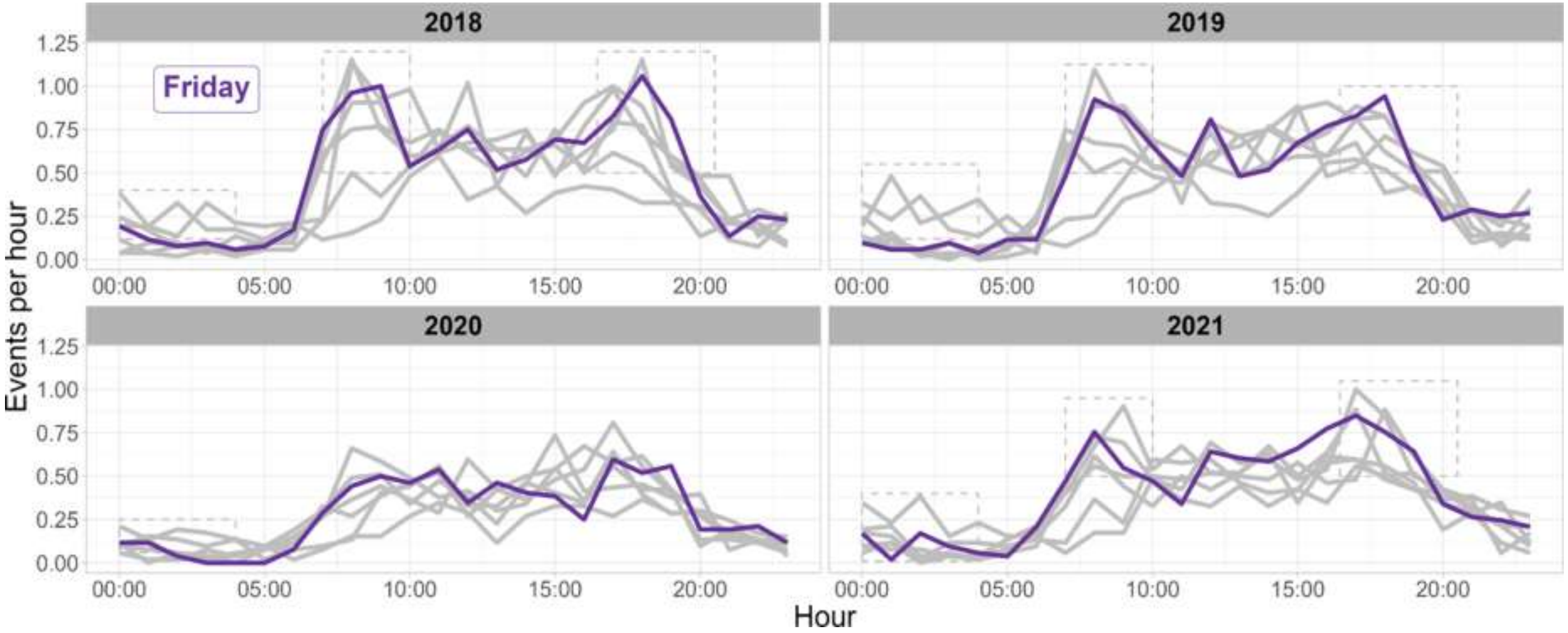
Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Temporal analysis - Via Olimpica



Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

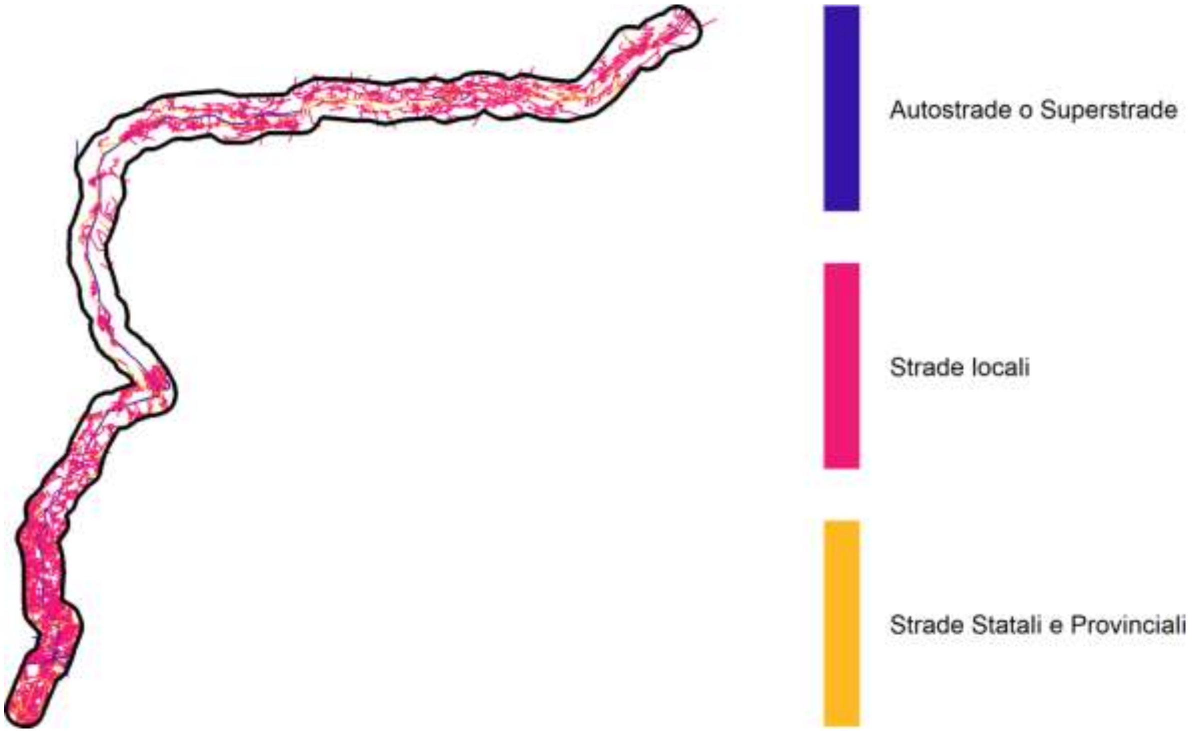
Temporal analysis - Via Olimpica



Gilardi, A., Ieva, F., Sangalli, L., Secchi, P.

Deep dive on car crashes and road types

The main road (**SS 36**) is classified as **Autostrada o Superstrada!**



Deep dive on car crashes and road types

***Ratio of car accident rates* (Reference year: 2018)**

2019	2020	2021
0.97 (0.95; 1.00)	**0.65 (0.62; 0.68)**	0.86 (0.82; 0.90)

Interpretation: On average, the number of car crashes per metre observed in 2020 is **65%** of that observed in 2018.

Deep dive on car crashes and road types

GRINS FOUNDATION

Ratio of car accident rates **(Reference category: Autostrade)**

	2018	2019	2020	2021
Strade Locali	0.55 (0.50; 0.61)	0.51 (0.45; 0.56)	**0.68 (0.59; 0.77)**	0.56 (0.50; 0.62)
Strade Statali e Provinciali	1.40 (1.26; 1.56)	1.16 (1.03; 1.29)	**1.40 (1.21; 1.62)**	1.25 (1.11; 1.42)

Interpretation: On average, the number of car crashes per metre observed in 2020 on *Strade Statali e Provinciali* is **140%** of that observed in *Autostrade e Superstrade* during the same year.