

# Percorsi orientamento PNRR

Università degli Studi di Milano-Bicocca  
Liceo Scientifico Lussana (Bergamo)  
Febbraio 2025

*Paolo Maranzano*

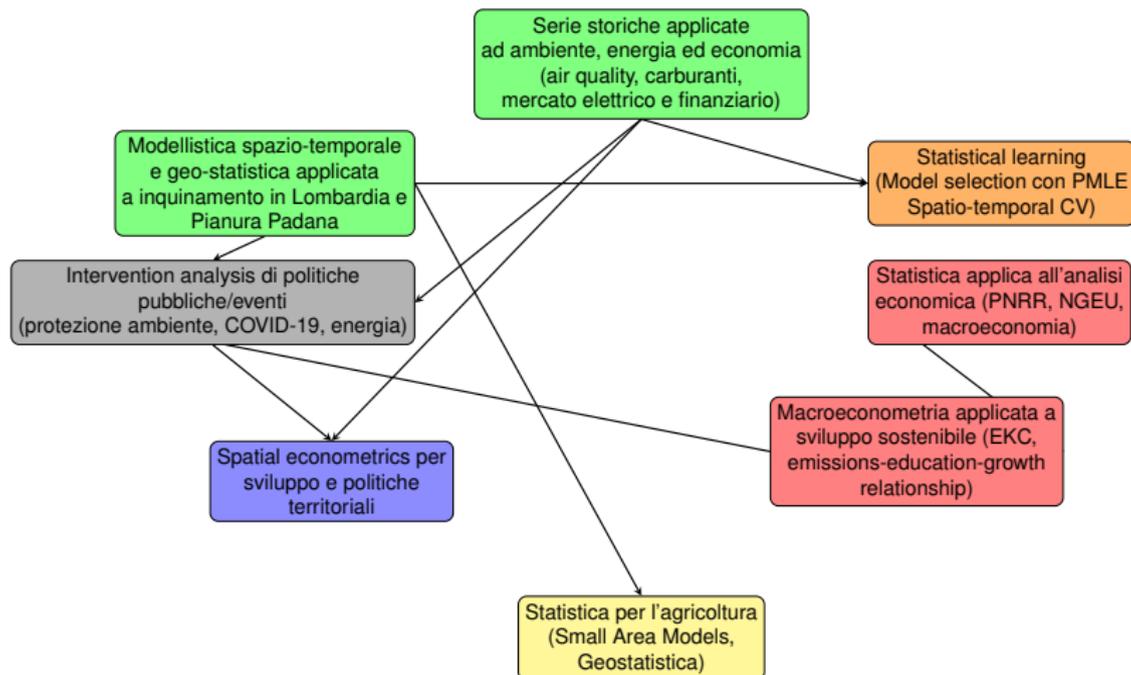
**Introduzione alle serie storiche e alla geostatistica spazio-temporale**

18-26 Febbraio 2025

# Paolone...

- Gen. 2022 - oggi: ricercatore in Statistica Economica (SECS-S/03) @ Università degli Studi di Milano-Bicocca (DEMS)
- 2021: Dottorato di ricerca in Statistics @ UniMiB. Tesi: *Essays on air quality statistics: time series methods for evaluating environmental protection policies in Lombardy, Italy.*
- 2017: LM in Scienze Statistiche ed Economiche (UniMiB)
- 2015: LT in Economia in 2015 (UniBg)
- Progetto **Agrimonia**: Agriculture Impact on Italian Air Quality (<https://agrimonia.net/>)
- Progetto **SCARFACE**: Sequestering CARbon through Forests, AgriCulture, and land use
- **Mail**: [paolo.maranzano@unimib.it](mailto:paolo.maranzano@unimib.it)
- **Pagina personale**: <https://www.paolomaranzano.net/>

# Attività e interessi di ricerca



# Definizione di una serie storica

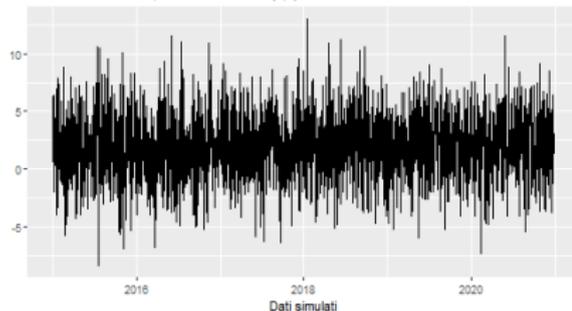
Molti fenomeni di natura ambientale, sociale, economica, industriale, etc. **evolvono/cambiano nel corso del tempo:**

- Prezzi della benzina che cambia giorno per giorno
- Le concentrazioni atmosferiche di un inquinante cambia ora per ora (o al minuto)
- I parametri meteorologici (es. temperatura o velocità del vento) cambiano di minuto in minuto
- Gli indici sociali (es. grado di povertà) economici di un paese (es. PIL) che cambia trimestralmente o annualmente
- Numero di pezzi difettosi in una linea di produzione industriale può cambiare di ora in ora

# Definizione di una serie storica

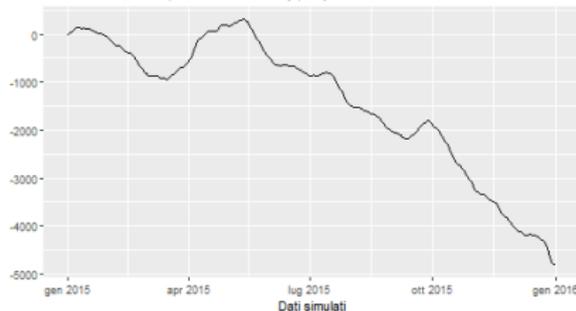
## White Noise

White noise  $n = 365 \times 6$ , con distribuzione  $N(2,9)$



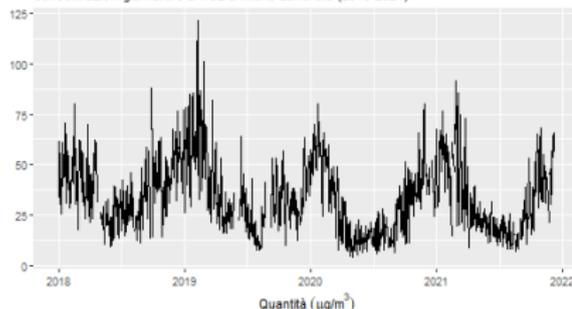
## Random Walk

Random Walk  $n = 365$ , con distribuzione  $N(0,100)$



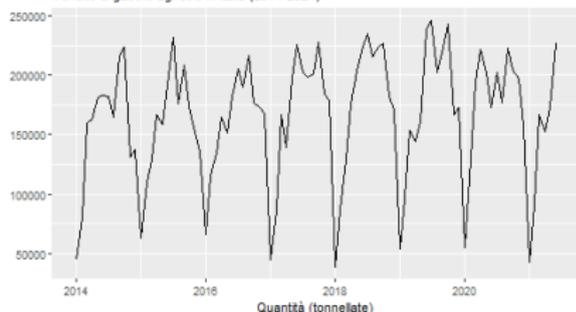
## Dati reali

Concentrazioni giornaliere di NO<sub>2</sub> a Milano Lambrate (2018-2021)

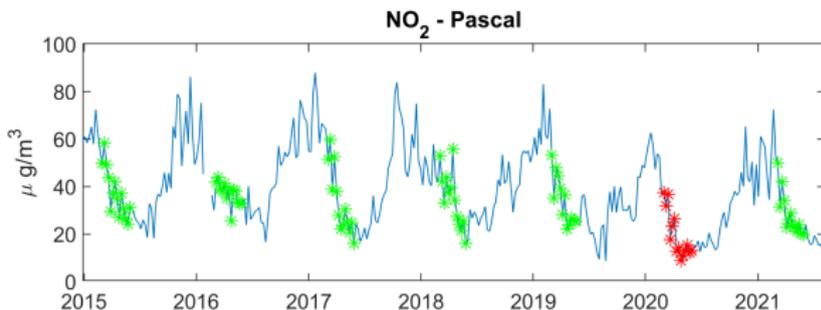
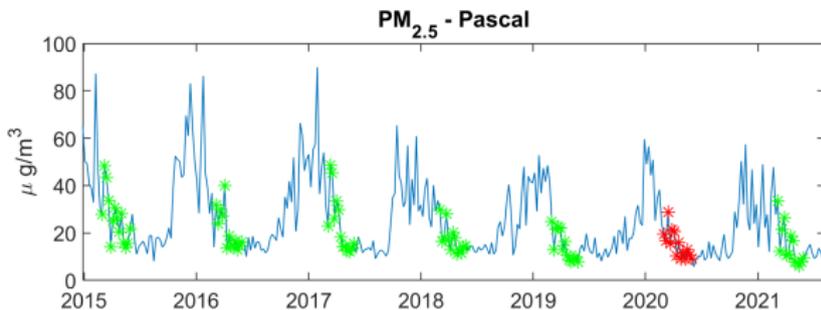


## Dati reali

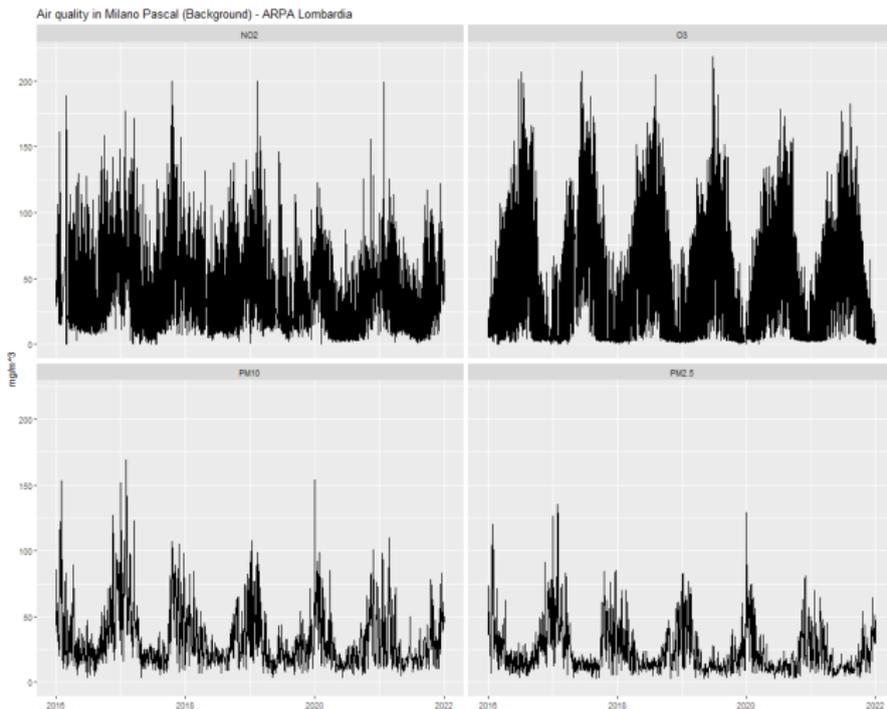
Vendite di gasolio agricolo in Italia (2014-2021)



# Definizione di una serie storica



# Definizione di una serie storica



# Definizione di una serie storica

Cosa si intende per serie storica? Una serie storica univariata è una sequenza di osservazioni (o misurazioni) ordinate temporalmente. Dato un insieme di  $n$  indici ordinati rispetto al tempo (gli istanti di tempo: ore, giorni, mesi, anni)

$$\{t_1, t_2, \dots, t_n\}$$

indichiamo la serie storica  $Y$  come i valori associati ad ogni indice:

$$\{y_{t_1}, y_{t_2}, \dots, y_{t_n}\}$$

- Gli istanti di tempo possono essere sia equispaziati (la distanza tra un istante e l'altro è costante) che non-equidistanti;
- Il tempo che intercorre tra una misura e la successiva è detta **frequenza di campionamento** della serie storica (es. mensile se ogni valore raccolto rappresenta un mese diverso)

# Analisi di una serie storica

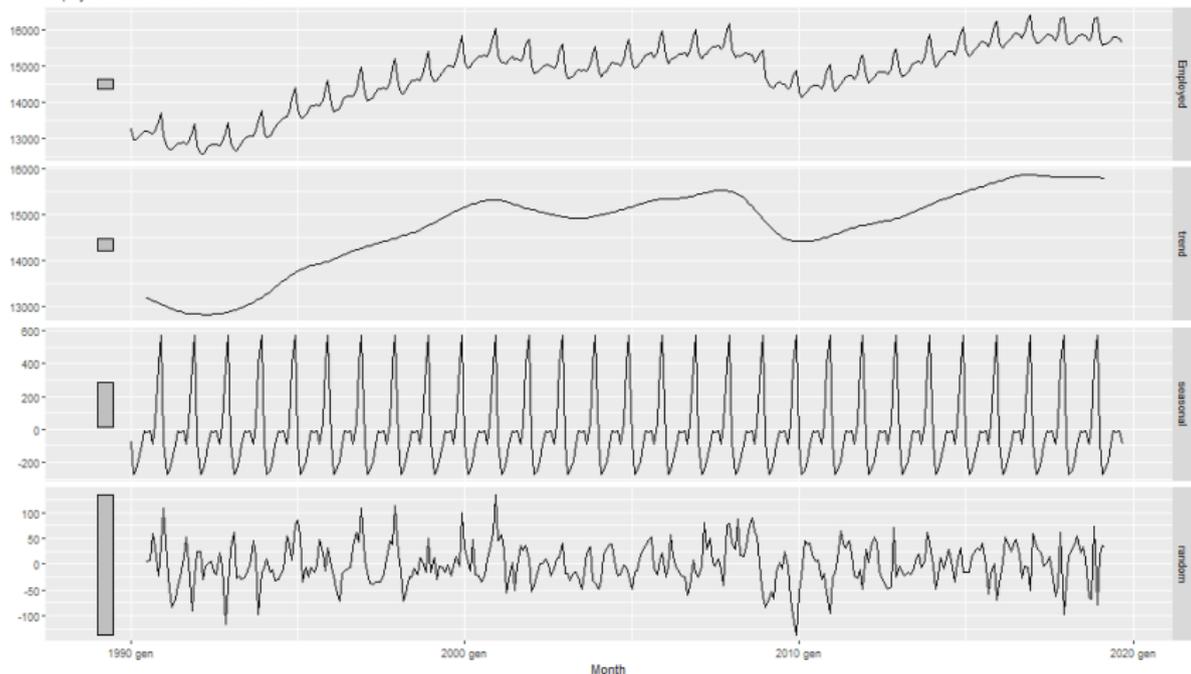
Come possiamo analizzare una serie storica?

- Analisi strutturale e **decomposizione**: identificare e isolare le componenti (strutture/pattern regolari) che caratterizzano la serie attraverso modelli statistici e tecniche di decomposizione
- **Previsione** di valori nel presente o **a brevissimo termine**: *real-time forecasting* o *nowcasting*
- **Previsione di valori futuri** basandoci sul passato: *forecasting*
- **Ricostruzione** di valori passati (es. imputazione di valori mancanti): *backcasting*
- Analisi d'**impatto**: quantificare l'impatto/effetto di politiche (*policy impact evaluation*) o eventi (*event studies*) nel corso del tempo
- **Clustering** di serie storiche: identificare similitudini nel comportamento e nella dinamica delle serie storiche per poi raggrupparle in gruppi omogenei

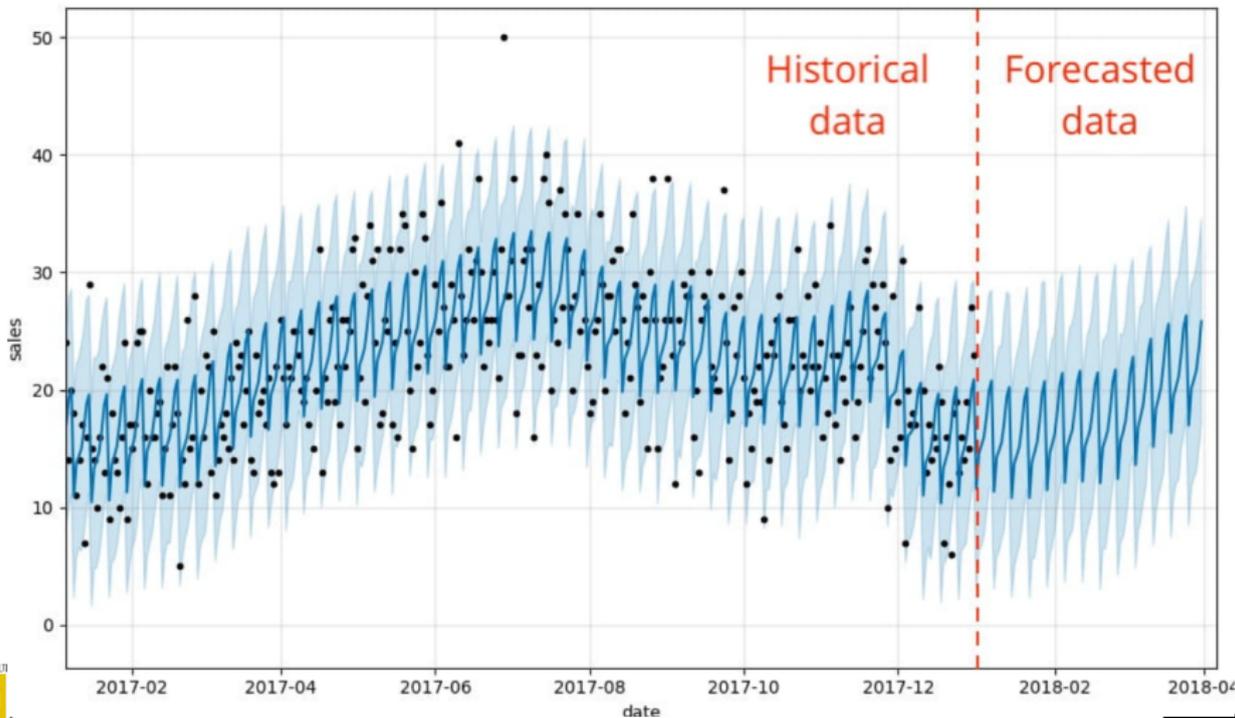
# Analisi di una serie storica: decomposizione

Classical additive decomposition of total US retail employment

Employed = trend + seasonal + random

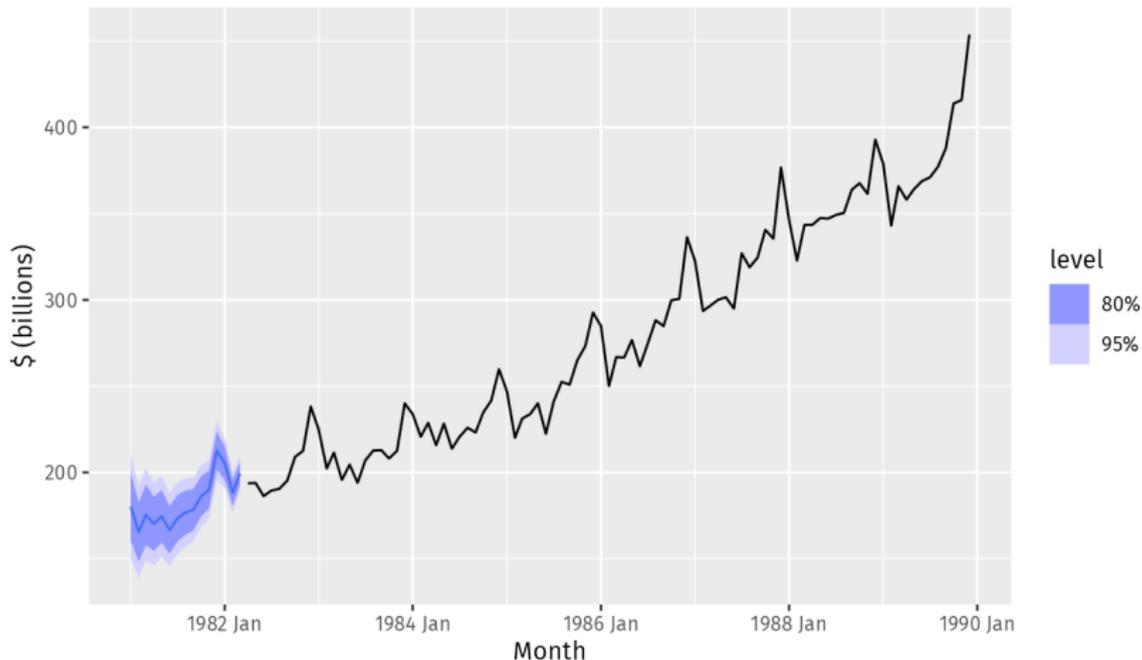


# Analisi di una serie storica: forecasting



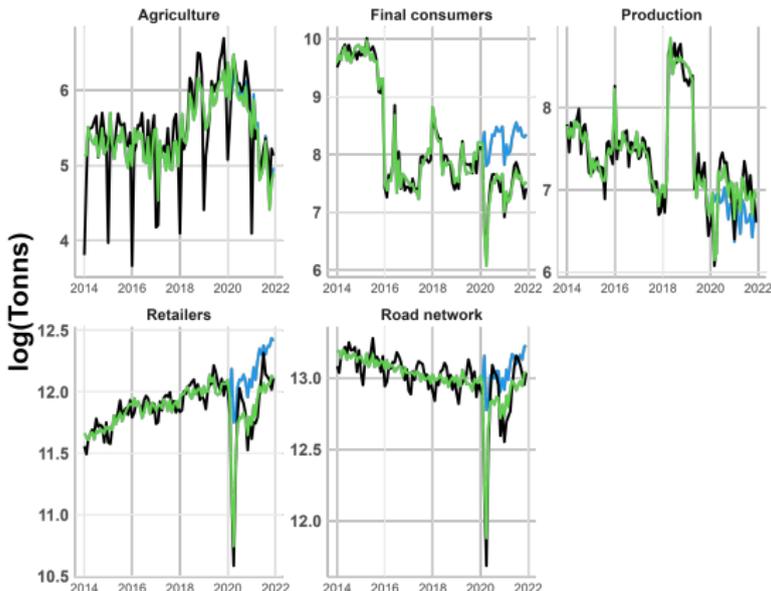
# Analisi di una serie storica: backcasting

## Backcasts of Australian food expenditure



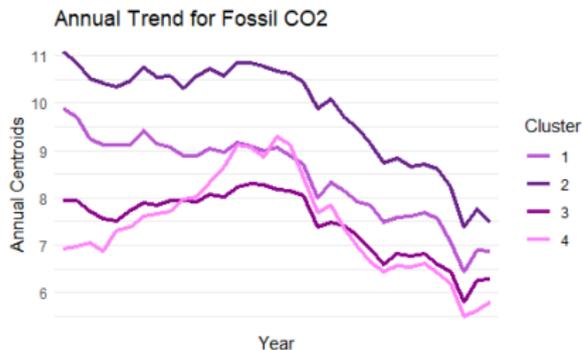
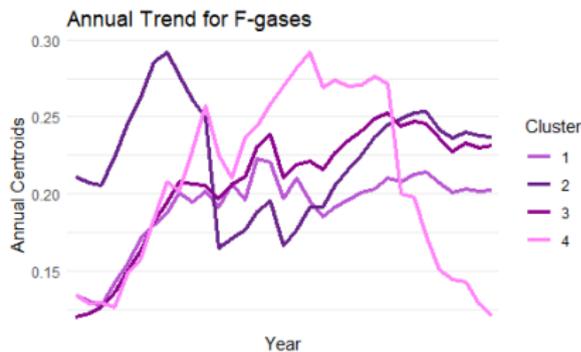
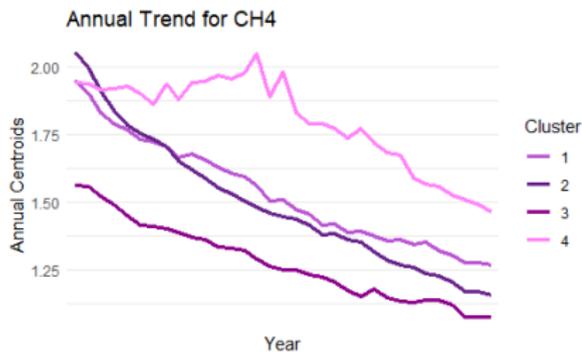
# Analisi di una serie storica: impatto

## Gasoline



■ COVID + Seasonal adj. ■ Original ■ Seasonal adj.

# Analisi di una serie storica: clustering



# Vocabolario delle serie storiche

Per descrivere una serie storica utilizzeremo alcuni termini specifici, tipo:

- **Componenti strutturali:** Trend, Stagionalità, Ciclo, Break strutturale
- **Casualità**
- **Stazionarietà, memoria e persistenza**
- **Passato, presente e futuro**

# Analisi esplorativa grafica

Il primo passo per una corretta analisi delle serie storiche è rappresentarla graficamente.

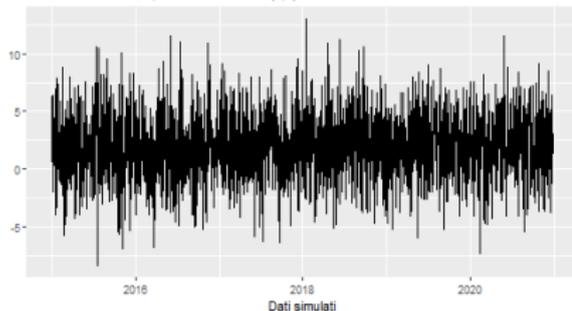
Per quanto possa sembrare banale, dal grafico possiamo intuire tantissime caratteristiche della serie:

- Presenza di componenti strutturali (trend, ciclo, stagionalità, breaks)
- Grado di persistenza (memoria della serie) e stazionarietà
- Presenza di valori anomali
- Altri pattern particolari, tipo eteroschedasticità

# Rappresentazione della serie storica

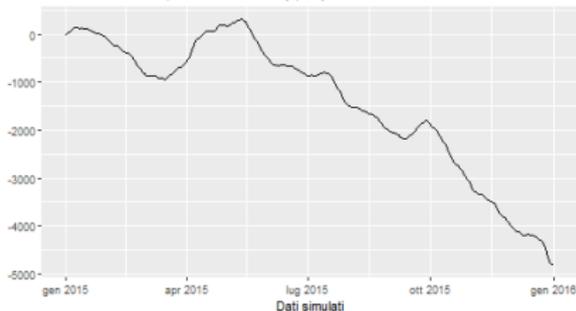
## White Noise

White noise  $n = 365 \times 6$ , con distribuzione  $N(2,9)$



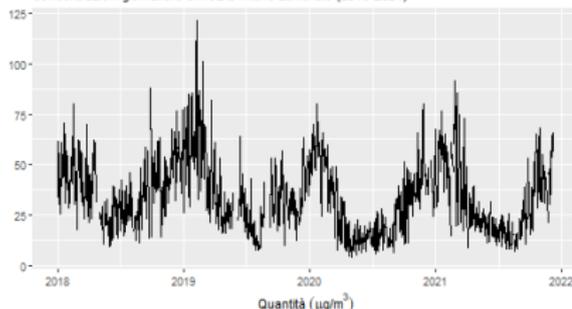
## Random Walk

Random Walk  $n = 365$ , con distribuzione  $N(0,100)$



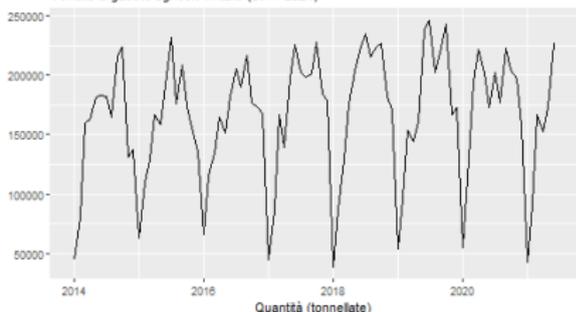
## Dati reali

Concentrazioni giornaliere di NO<sub>2</sub> a Milano Lambrate (2018-2021)



## Dati reali

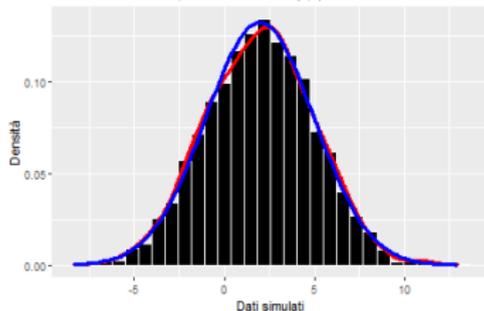
Vendite di gasolio agricolo in Italia (2014-2021)



# Distribuzione: Istogramma

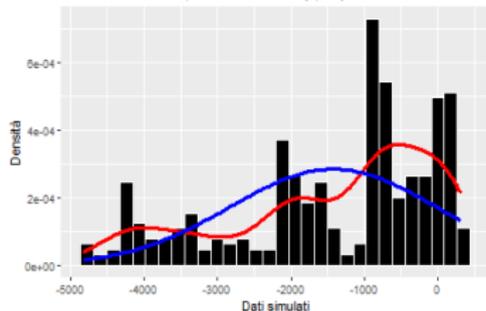
Istogramma: normalità

White noise  $n = 365 \times 6$ , con distribuzione  $N(2,9)$



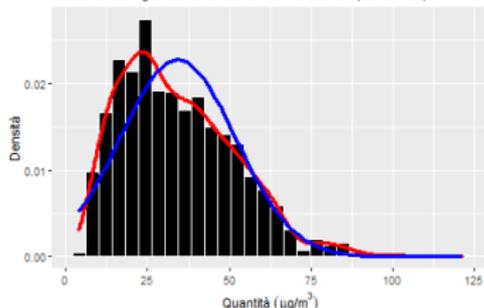
Istogramma: normalità

Random Walk  $n = 365$ , con distribuzione  $N(0,100)$



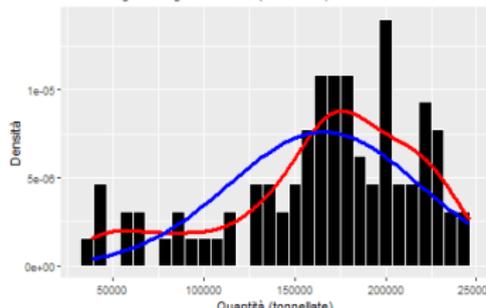
Istogramma: asimmetria positiva

Concentrazioni giornaliere di NO<sub>2</sub> a Milano Lambrate (2018-2021)

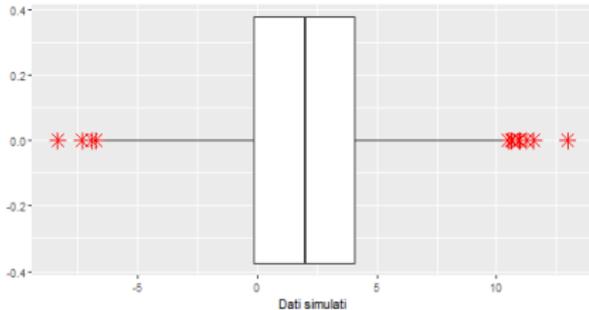
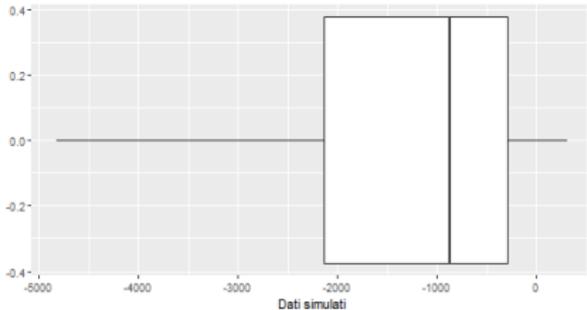
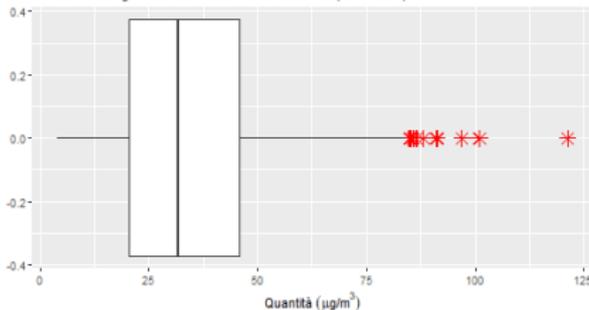


Istogramma: asimmetria negativa

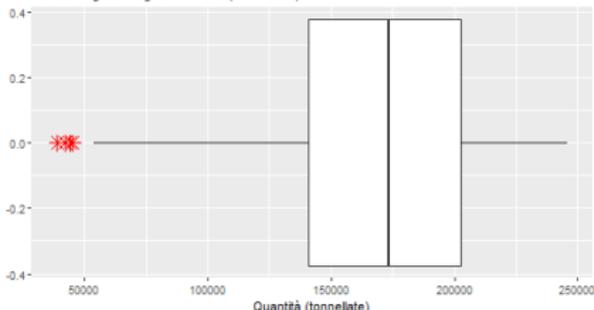
Vendite di gasolio agricolo in Italia (2014-2021)



# Distribuzione: Box-plot

**Boxplot: normalità**White noise  $n = 365 \times 6$ , con distribuzione  $N(2,9)$ **Boxplot: normalità**Random Walk  $n = 365$ , con distribuzione  $N(0,100)$ **Boxplot: asimmetria positiva**Concentrazioni giornaliere di NO<sub>2</sub> a Milano Lambrate (2018-2021)**Boxplot: asimmetria negativa**

Vendite di gasolio agricolo in Italia (2014-2021)



# Componenti delle serie storiche

Molte serie temporali sono caratterizzate dalla **sovrapposizione** di varie **componenti strutturali** che ne determinano l'andamento:

- Trend
- Stagionalità
- Ciclo
- Break strutturale

Quando si sceglie un metodo di previsione, bisognerà prima identificare i modelli di serie temporali nei dati, e poi scegliere un metodo che sia in grado di catturare correttamente i modelli.

# Trend

## Trend

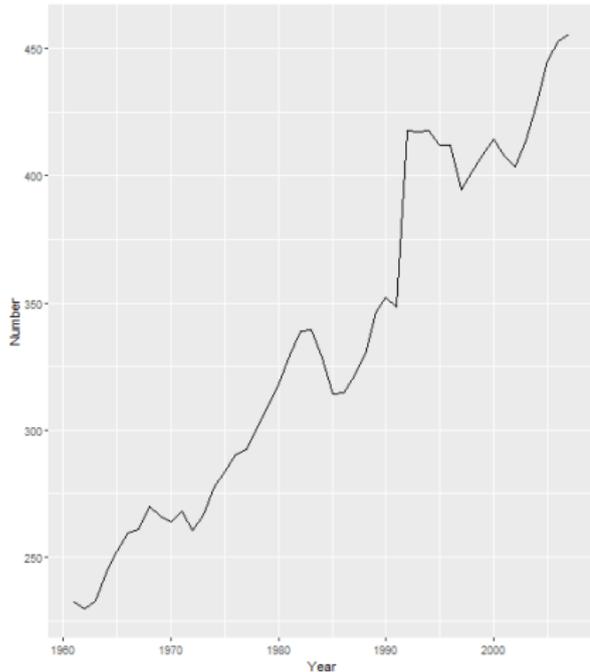
Una tendenza (trend) esiste quando c'è un aumento o una diminuzione a lungo termine nei dati.

- Non deve essere necessariamente lineare: posso avere trend deterministici polinomiali o logaritmici
- A volte ci si riferisce a una tendenza come a un "cambiamento di direzione", quando potrebbe passare da una tendenza all'aumento a una tendenza alla diminuzione (qui è più appropriato di parlare di **livello medio** della serie)

# Trend

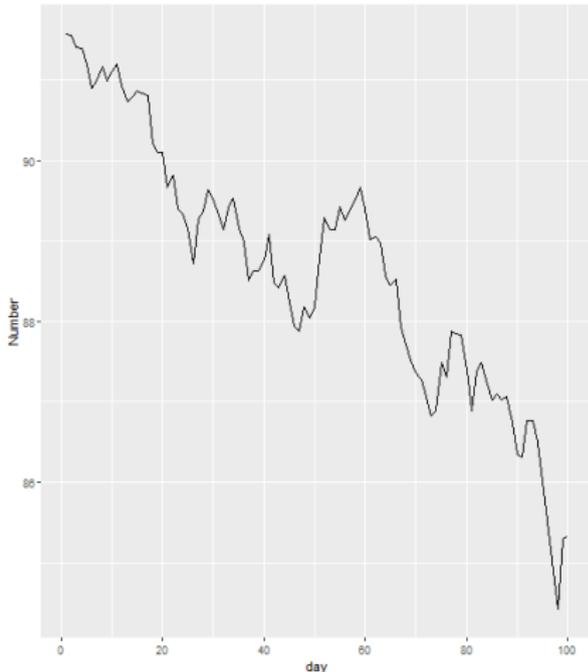
### Trend crescente

Annual sheep livestock numbers in Asia.



### Trend decrescente

US treasury bill contracts



# Ciclicità e stagionalità

## Stagionalità

La stagionalità è un fenomeno che si manifesta quando una serie temporale è influenzata da fattori stagionali come il periodo dell'anno o il giorno della settimana. La stagionalità ha sempre una frequenza **fissa e nota**. Le vendite mensili dei farmaci antidiabetici di cui sopra mostrano una stagionalità che è indotta in parte dal cambiamento del costo dei farmaci alla fine dell'anno solare.

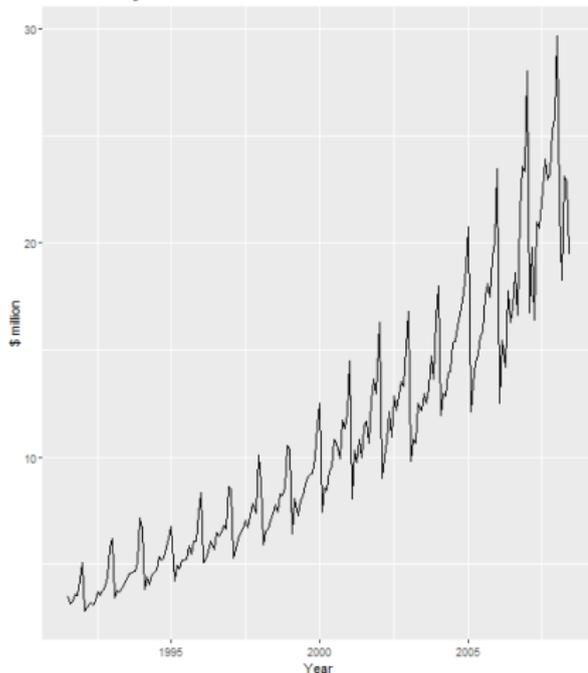
## Ciclicità

Un ciclo si verifica quando i dati mostrano aumenti e diminuzioni che **non** hanno una frequenza fissa. Queste fluttuazioni sono di solito dovute a condizioni economiche, e sono spesso collegate al *ciclo economico*. La durata di queste fluttuazioni è solitamente di almeno 2 anni.

# Stagionalità

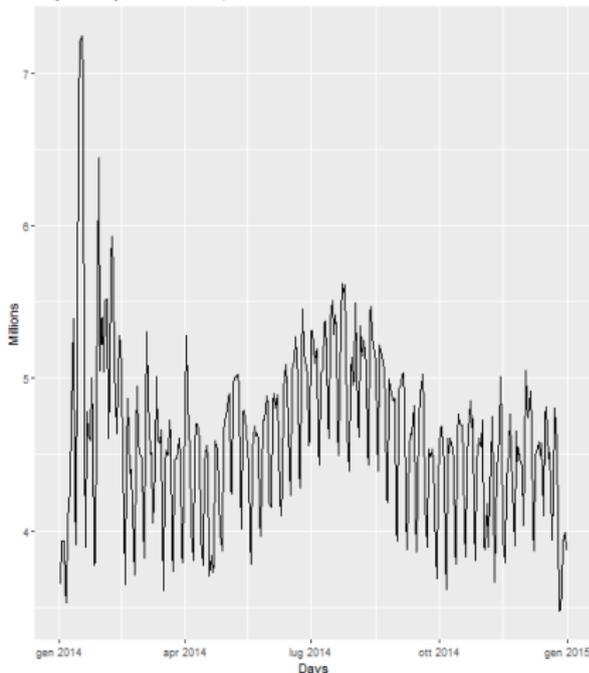
### Stagionalità moltiplicativa

Antidiabetic drug sales



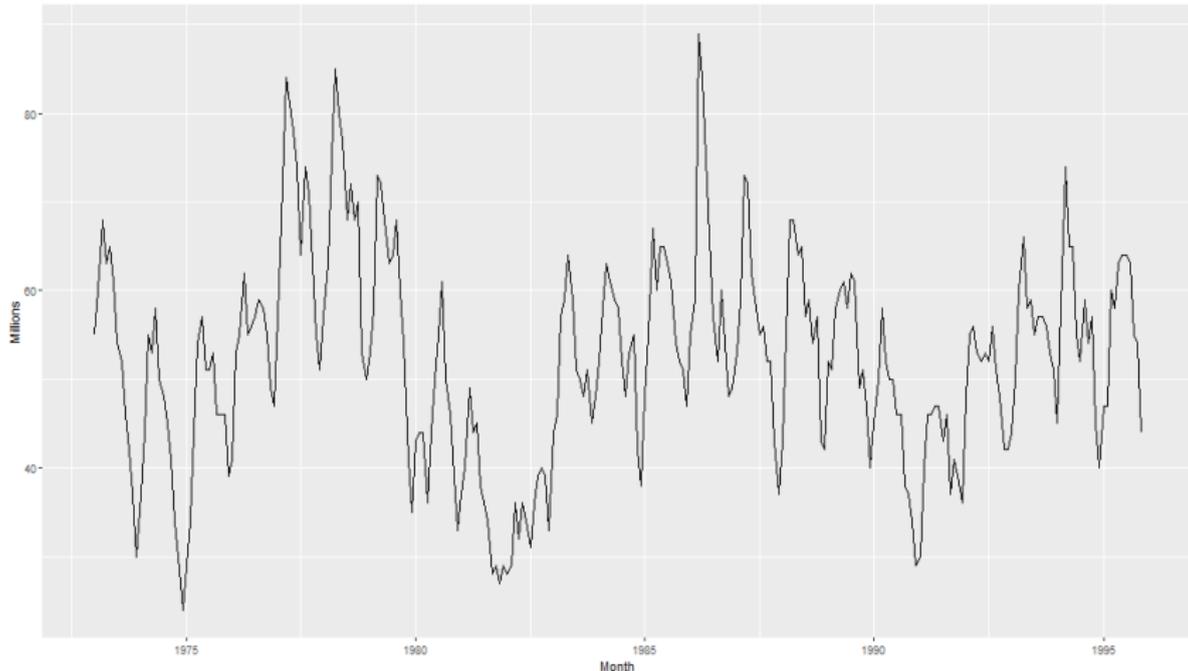
### Stagionalità additiva

Daily electricity demand: Victoria, Australia



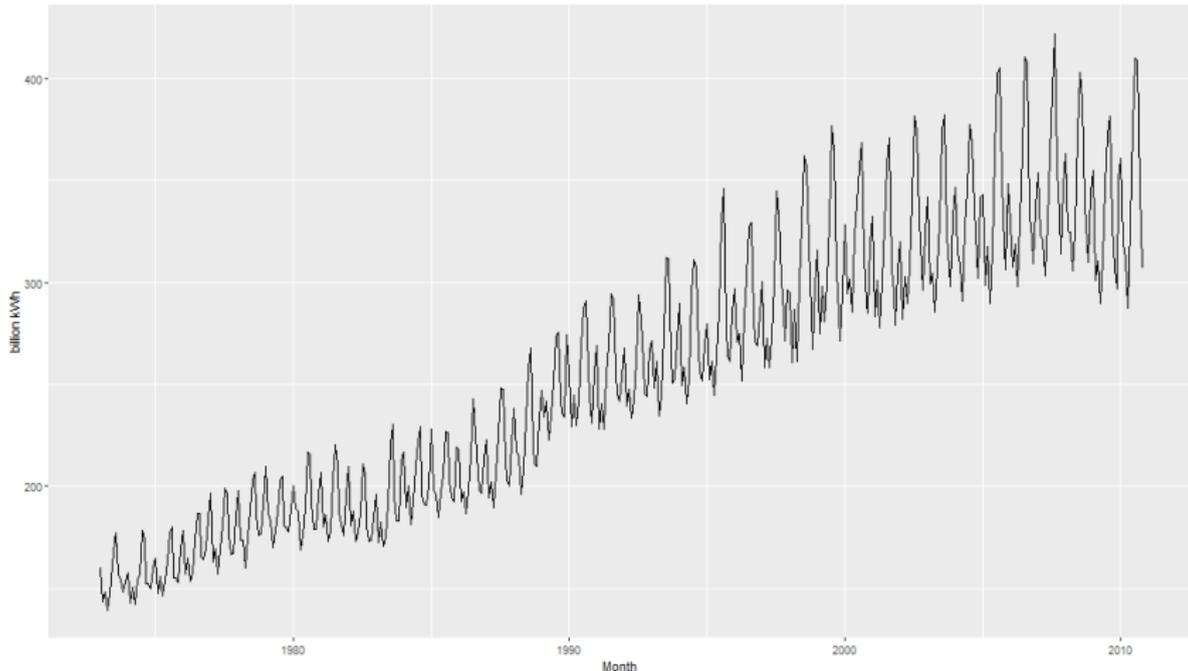
# Stagionalità additiva + Ciclo

Stagionalità additiva + ciclo  
Sales of new one-family houses in US



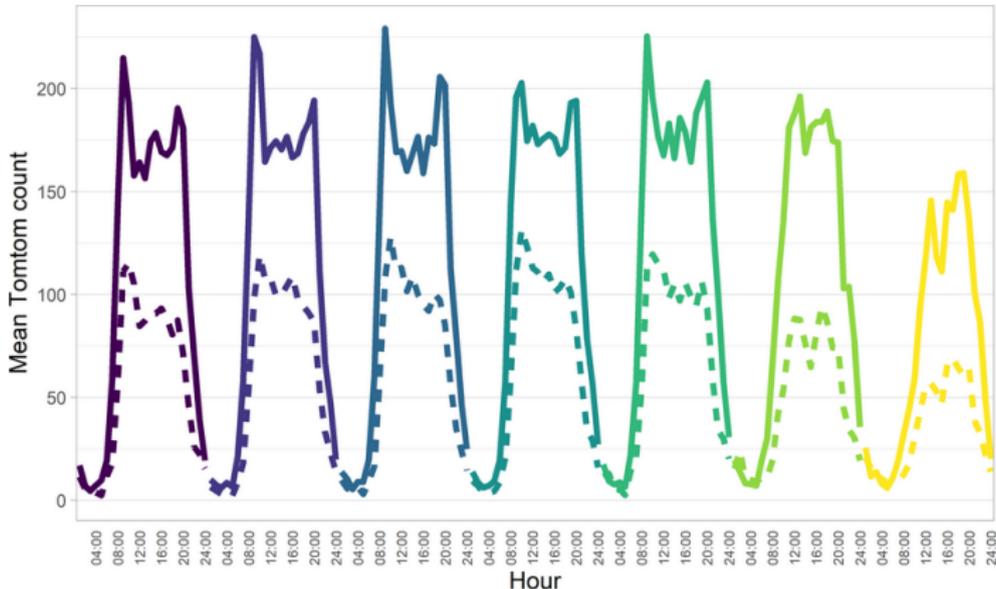
# Stagionalità moltiplicativa + Trend

Stagionalità moltiplicativa + trend  
Electricity net generation measured



# Analisi di una serie storica: decomposizione

Plot



Weekday

Monday

Tuesday

Wednesday

Thursday

Friday

Saturday

Sunday

Centralina

marche

verziere

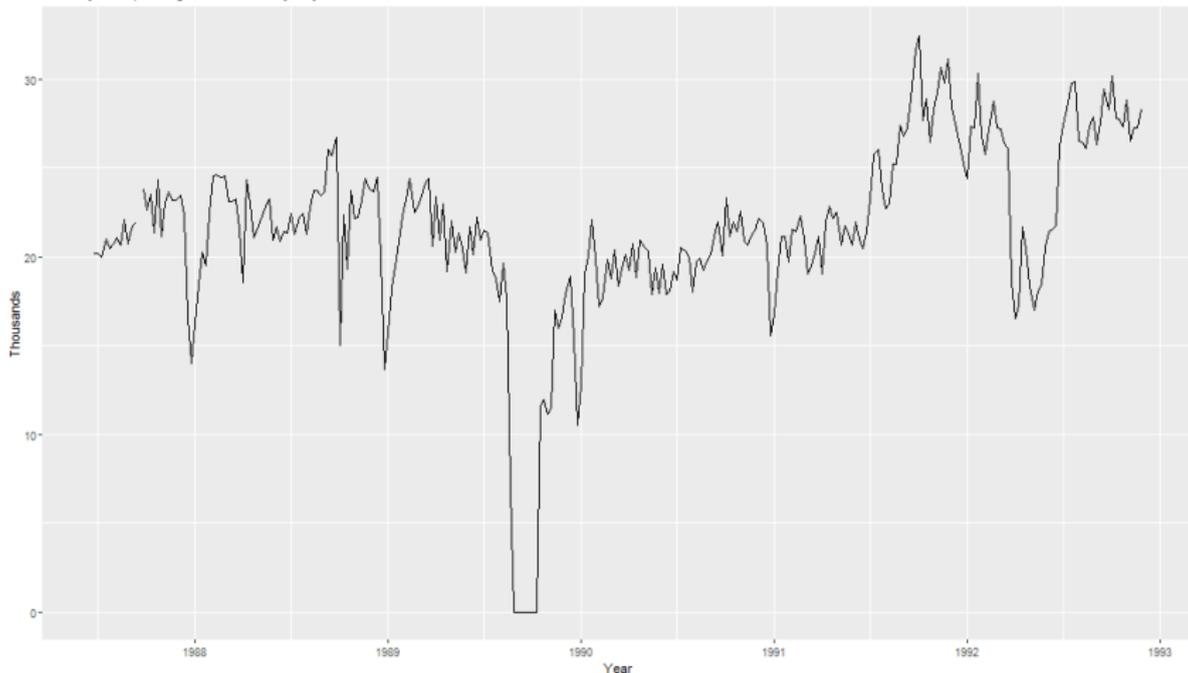
# Break strutturali

Un break strutturale si verifica quando c'è una variazione repentina (variazione verticale nel grafico) verso l'alto o verso il basso in una serie storica. Ci possono essere breaks temporanei, ma anche permanenti; repentini o adattivi (lenti).

- Crisi petrolifera anni '70
- Crisi finanziaria-reale (2008) o crisi dei debiti sovrani (2011-2013)
- Pandemia di COVID-19
- Guerra in Ucraina e Inflazione del 2022-2023
- Introduzione di una politica pubblica (es. zona a traffico limitato oppure obbligo di cintura di sicurezza)

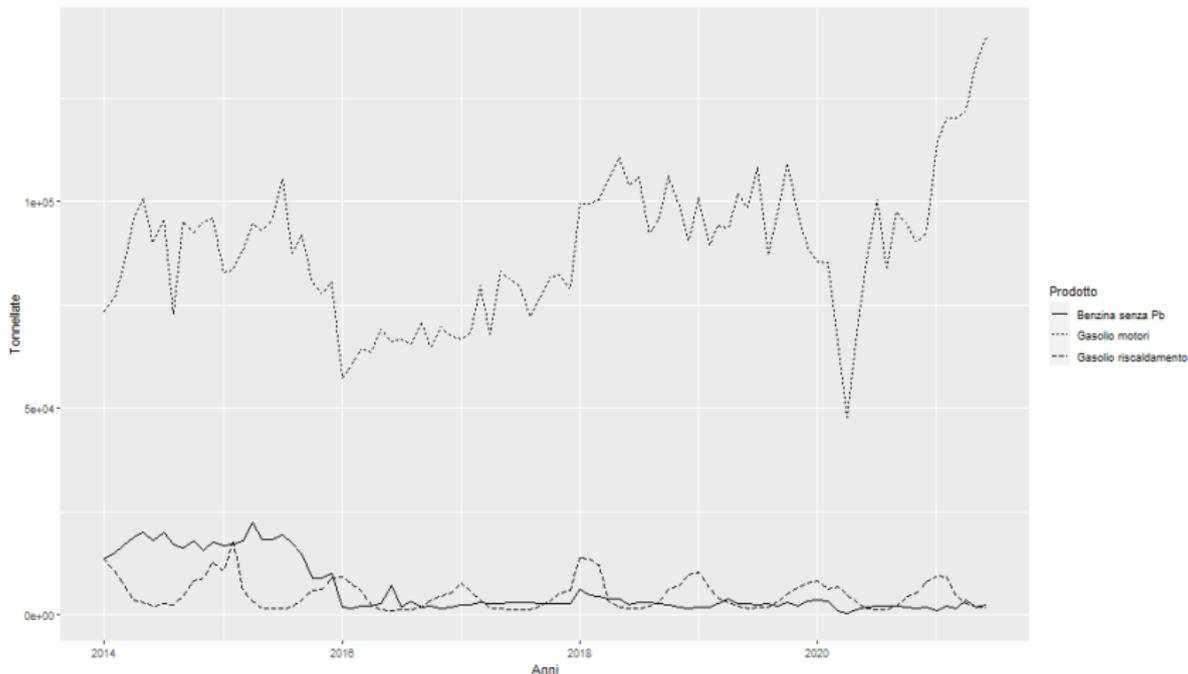
# Break strutturale: esempi

Break strutturali / Policy intervention  
Economy class passengers: Melbourne-Sydney



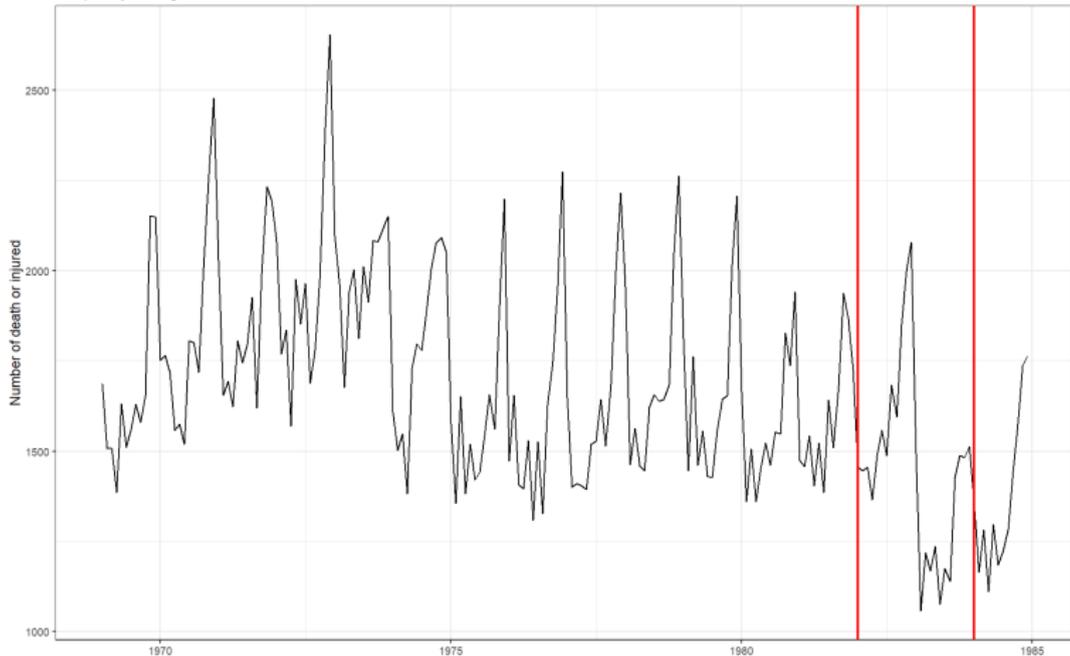
# Break strutturale: esempi

Vendita carburanti ai consumatori finali  
Pandemia da COVID-19



# Break strutturale: esempi

Monthly totals of car drivers in UK killed or seriously injured Jan 1969 to Dec 1984.  
Compulsory wearing of seat belts was introduced on 31 Jan 1983.



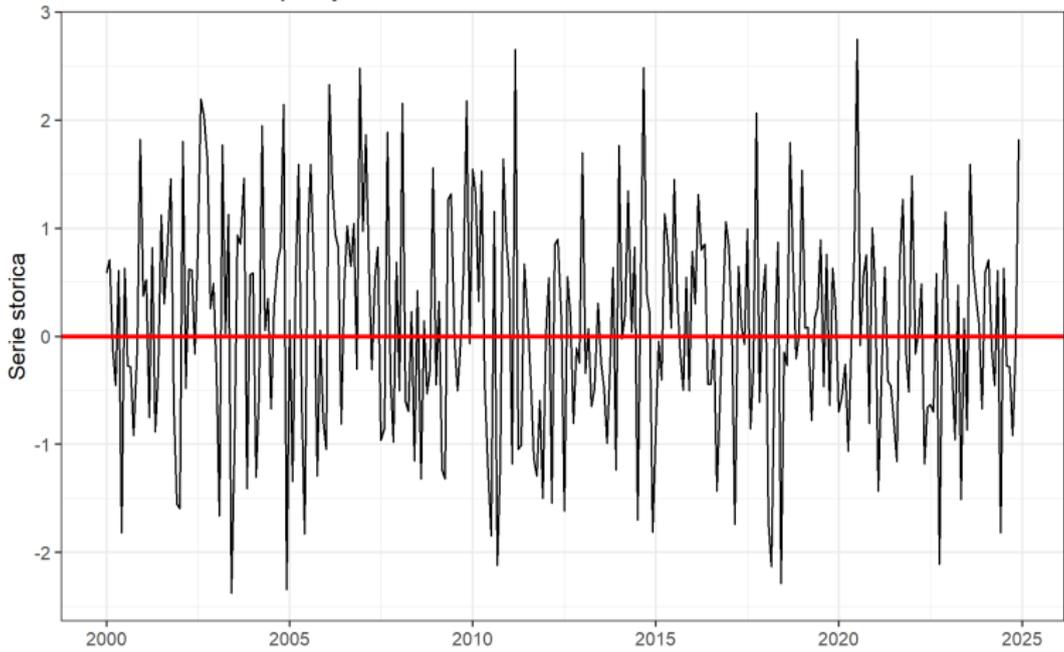
# Casualità

La serie storica può essere caratterizzata da un **andamento completamente casuale** in cui non si riescono ad individuare pattern specifici e non sono prevedibili nel futuro. Se la serie possiede alcune specifiche condizioni parleremo di **White Noise** o **Rumore bianco**:

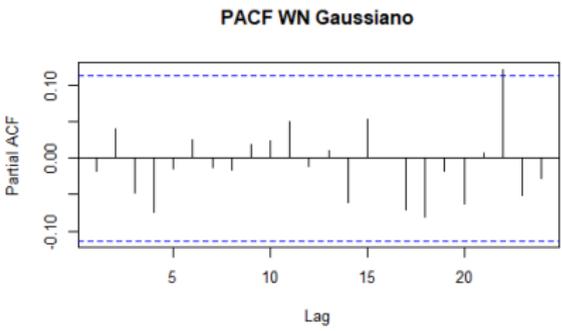
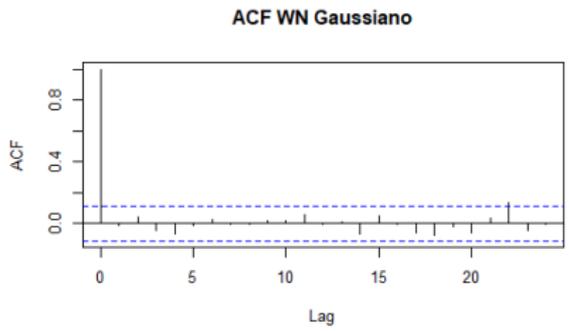
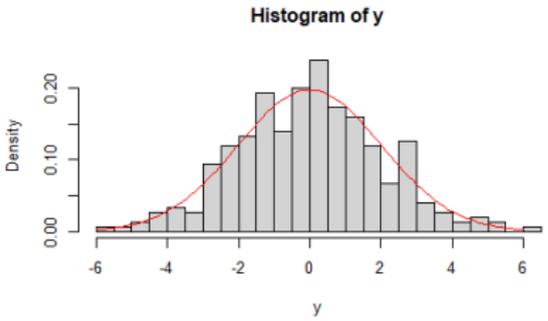
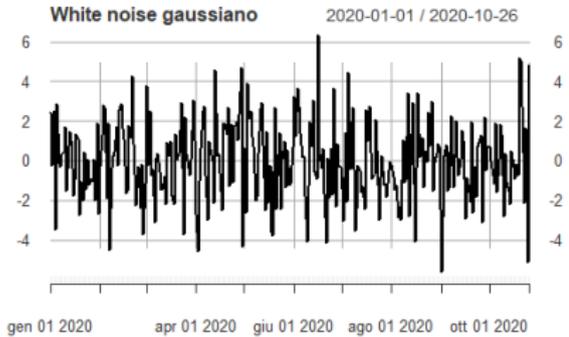
- Ogni osservazione è dovuta esclusivamente al caso, non vi sono pattern che l'hanno generata
- Osservazioni incorrelate nel tempo (non autocorrelate): ogni osservazione è indipendente dalle altre
- Il futuro è *st statisticamente imprevedibile*: si tratta di variazioni casuali non prevedibili attorno alla media e con una variabilità costante, in cui ogni valore è indipendente dagli altri

# Casualità

White Noise: completely random time series



# Casualità



# Stazionarietà

Una serie storica è stazionaria quando le sue proprietà statistiche (es. la forma delle componenti e la sua variabilità) sono stabili nel tempo, cioè mantiene nel tempo un certo equilibrio statistico.

Perché ci interessa la stazionarietà? Perché se una serie storica è "stabile" nel tempo allora potrò usare valori passati per studiare le proprietà statistiche in qualunque istante di tempo e prevedere (in maniera statisticamente valida) i valori futuri!

# Stazionarietà

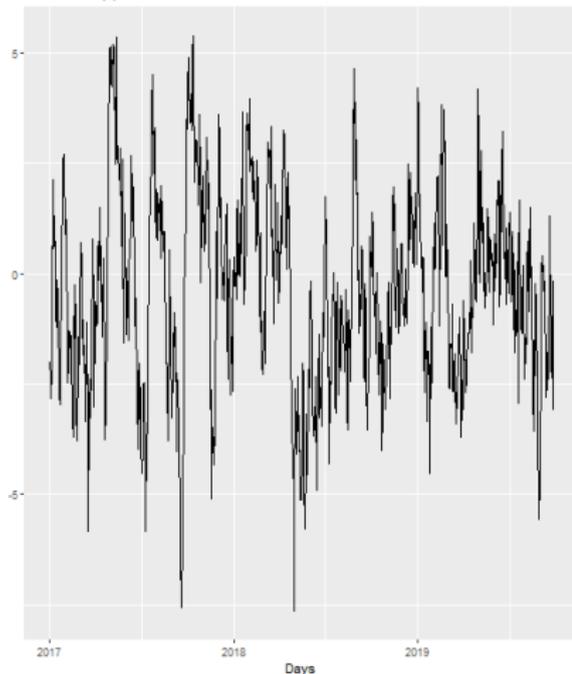
Il concetto di stazionarietà di una serie storica è intimamente connesso al concetto di **memoria o persistenza della serie storica** (grado di dipendenza tra un istante  $t$  e la storia passata e futura): tanto più la memoria tanto è elevata e tanto più i valori presenti e futuri saranno influenzati dal passato.

- Una serie è molto persistente quando anche il passato "abbastanza" remoto influenza il presente
- ... "abbastanza" significa che il passato molto molto lontano non può essere influente (**memoria finita**)
- Quando il passato non è influente sul presente e sul passato si dice che la serie storica non ha memoria

# Stazionarietà

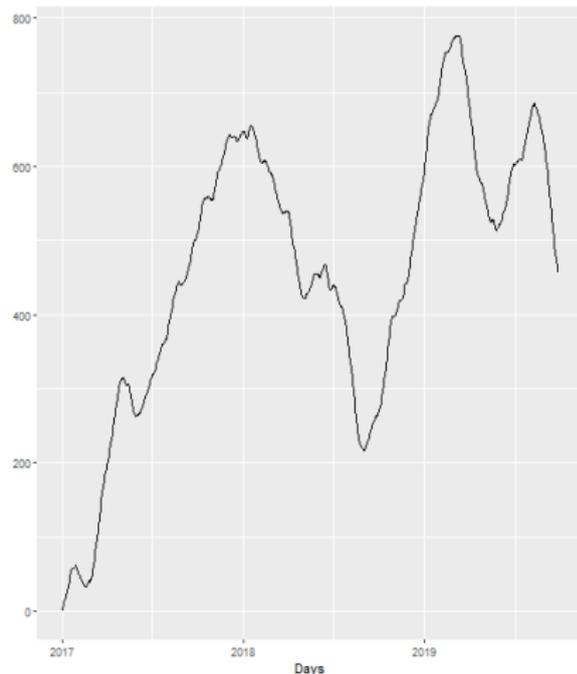
Serie stazionaria

Processo AR(1)



Serie non stazionaria

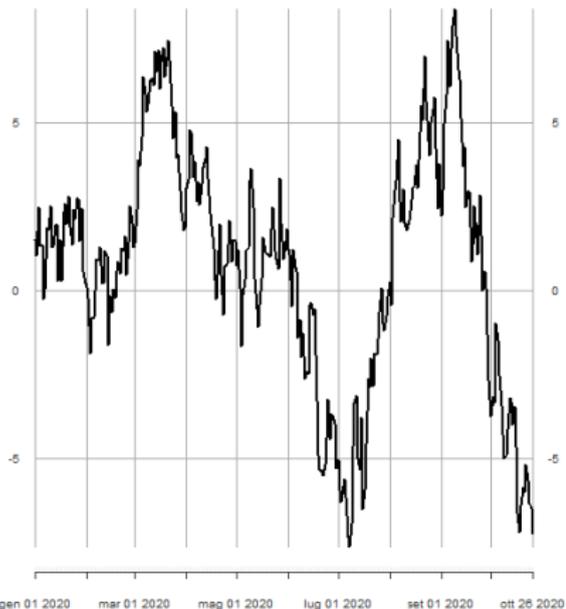
Random walk



# Stazionarietà

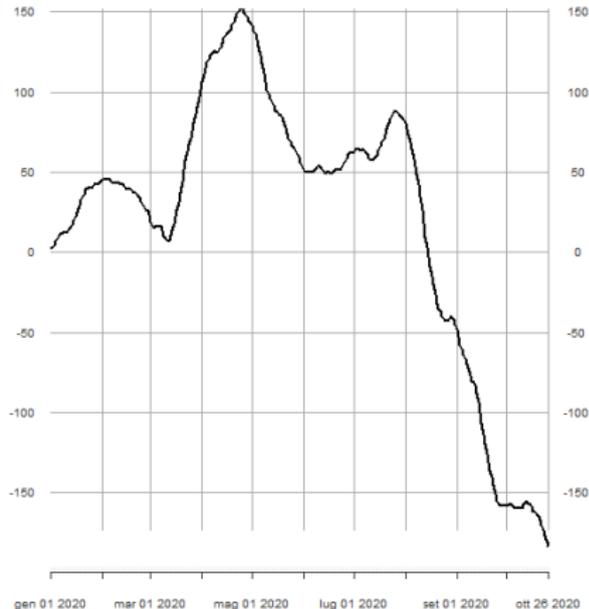
TS stazionaria

2020-01-01 / 2020-10-26



TS non stazionaria

2020-01-01 / 2020-10-26



# Persistenza e autocorrelazione

Una volta intuite le caratteristiche più evidenti della serie passiamo all'analisi della distribuzione.

Classici strumenti della statistica descrittiva basati sulla **correlazione lineare** tra i valori osservati e i valori ritardati della serie storica:

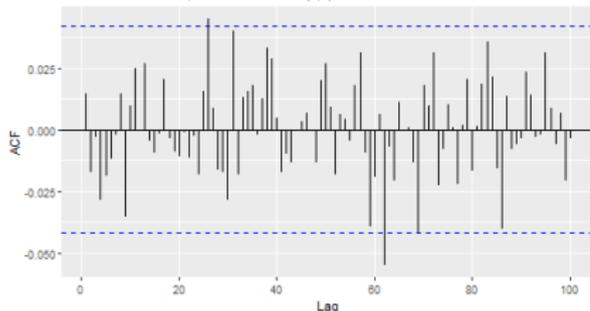
- Autocorrelogramma (ACF) e Autocorrelogramma parziale (PACF)
- Correlazione tra i ritardi (lagplot)

Ricordiamo che due fenomeni **correlano linearmente** quando **evolvono/variano/cambiano contemporaneamente**, cioè quando **al variare di uno varia l'altro (covariazione)**.

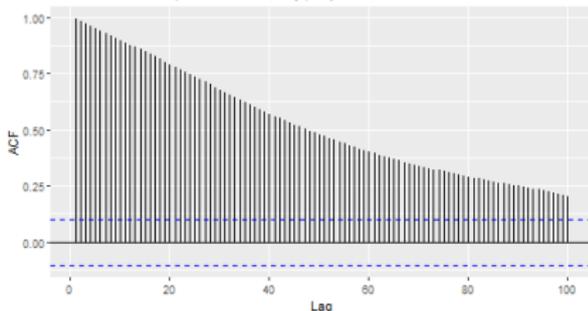
**Attenzione!!** Non stiamo dicendo che vi sia un legame causale, cioè che X cambia in seguito ad un cambiamento in Y (o viceversa)... questa è **causalità**.

# Persistenza: Autocorrelogramma

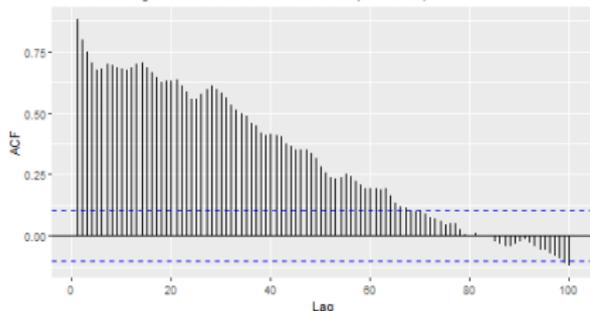
ACF: Serie con persistenza nulla

White noise  $n = 365^6$ , con distribuzione  $N(2,9)$ 

ACF: Serie con persistenza nulla

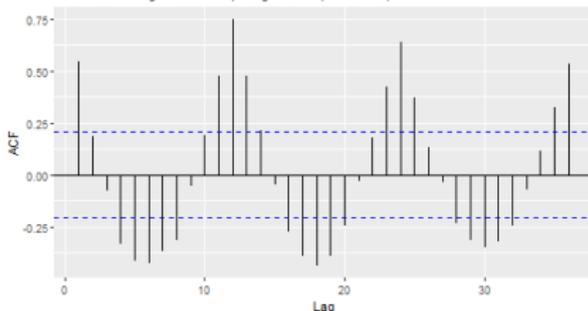
Random Walk  $n = 365$ , con distribuzione  $N(0,100)$ 

ACF: Serie con elevata persistenza

Concentrazioni giornaliere di NO<sub>2</sub> a Milano Lambrate (2018-2021)

ACF: Serie stagionale con media persistenza

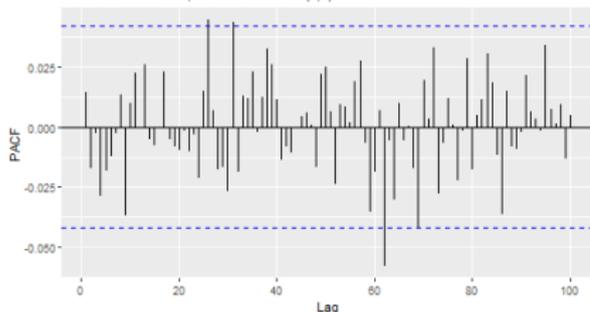
Vendita mensile gasolio motori per agricoltura (2014-2021)



# Persistenza: Autocorrelogramma parziale

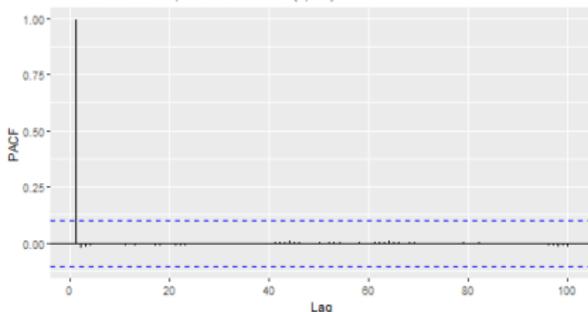
PACF: Serie con persistenza parziale nulla

White noise  $n = 365^6$ , con distribuzione  $N(2,9)$



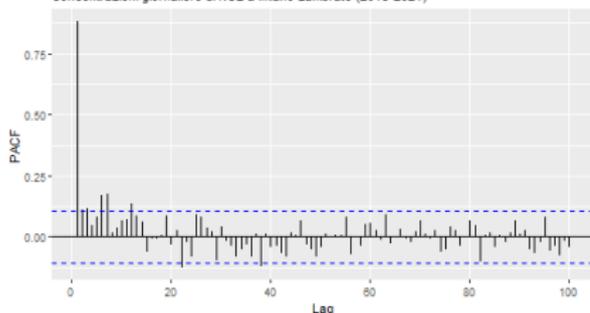
PACF: Serie con persistenza parziale nulla

Random Walk  $n = 365$ , con distribuzione  $N(0,100)$



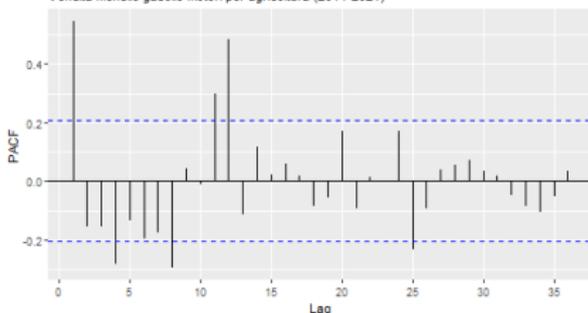
PACF: Serie con forte persistenza parziale di ordine 1

Concentrazioni giornaliere di  $NO_2$  a Milano Lambrate (2018-2021)



PACF: Serie stagionale con buona persistenza parziale

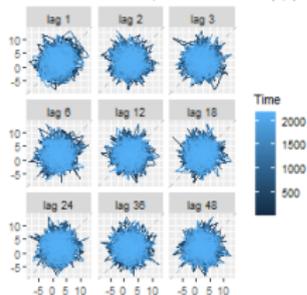
Vendita mensile gasolio motori per agricoltura (2014-2021)



# Persistenza: Lag plots

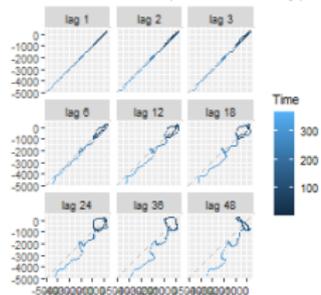
Lag-plots: no autocorrelazione

White noise  $n = 365 \times 6$ , con distribuzione  $N(2,9)$



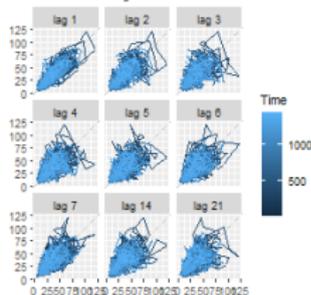
Lag-plots: no autocorrelazione

Random Walk  $n = 365$ , con distribuzione  $N(0,100)$



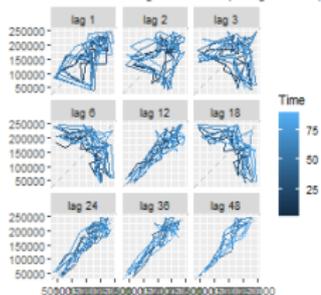
Lag plots: Serie con forte persistenza parziale di ordine 1

Concentrazioni giornaliere di NO<sub>2</sub> a Milano Lambrate (2018-2021)



Lag plots: Serie stagionale con buona persistenza

Vendita mensile gasolio motori per agricoltura (2014-2021)



# Dati spazio-temporali

Finora ci siamo concentrati su fenomeni ambientali, economici, sociali, etc. che **evolvono nel tempo** (serie storiche) oppure di fenomeni che **evolvono nello spazio** (analisi di dati spaziali).

Ovviamente, possiamo pensare di "mettere insieme" la dimensione spaziale e quella temporale per analizzare **dati geostatistici dinamici o spazio-temporali**.

Potenziali esempi riguardano:

- Misurazioni di parametri meteorologici o inquinanti in tante locazioni spaziali per un certo numero di giorni/mesi/anni (**serie storiche georeferenziate**);
- Quantità mensili di carburante vendute nelle province italiane per un certo numero di anni;
- Misurare un certo fenomeno economico nelle province europee per un certo numero di anni;
- Misurare il grado di sostenibilità ambientale di alcune aziende in funzione della loro localizzazione spaziale.

# Dati spazio-temporali

Il **dato georeferenziato o geostatistico** contiene informazioni riguardo la **geolocalizzazione**, le **forme spaziali** (es. poligoni o punti) e i **valori/misurazioni** del fenomeno da studiare. Il dato può essere **statico** (**cross-sezionale**) o **dinamico** (**spazio-temporale**).

I dati georeferenziati possono essere divisi in almeno due categorie:

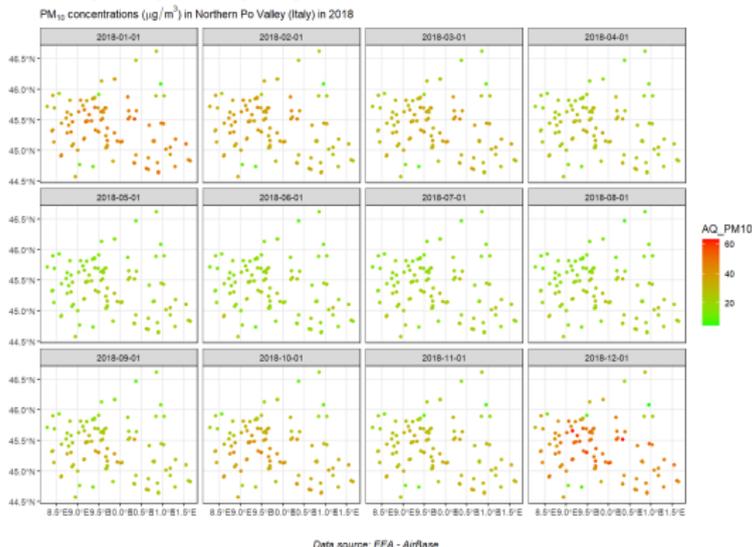
- 1 **Dati geostatistici**, in cui il fenomeno di interesse è misurato per più istanti di tempo presso una serie di coordinate spaziali;
- 2 **Dati areali/areali/reticolari**, in cui il fenomeno di interesse muta nel tempo presso un insieme finito di poligoni (es. pixel da satellite, dati comunali o regionali)

Ad ogni tipologia di dati sono associate metodologie analitiche differenti.

# Dati spatio-temporali: geostatistica dinamica

Rilevazione mensile di particolato atmosferico (PM<sub>10</sub>) in 84 stazioni di monitoraggio ambientale gestite da ARPA Lombardia nel 2018.  
Data source: ARPA Lombardia

Example of spatio-temporal geostatistical data

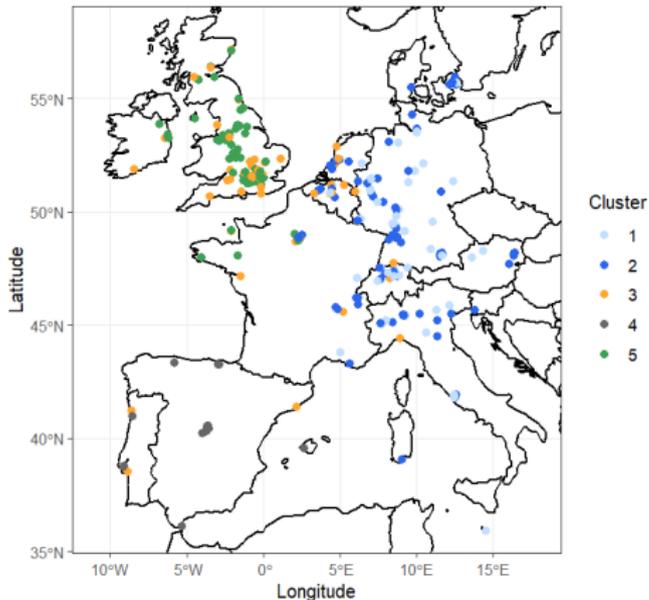


# Dati spazio-temporali: geostatistica dinamica

Raggruppamento di aziende europee in base al grado (crescente) di sostenibilità ambientale ESG tra il 2013 e 2023: mappa dei gruppi

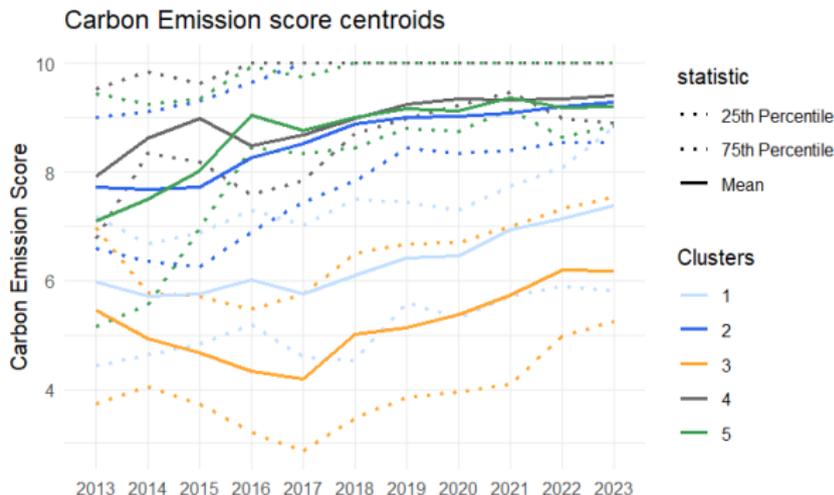
Data source: ARPA Lombardia

Map of spatiotemporal clusters in 2013-2023



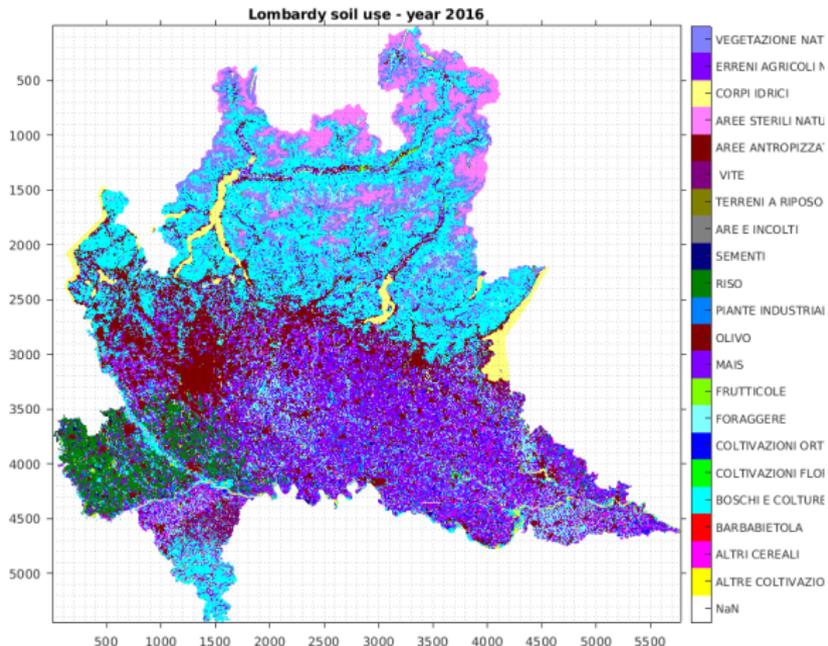
# Dati spazio-temporali: geostatistica dinamica

Raggruppamento di aziende europee in base al grado (crescente) di sostenibilità ambientale ESG tra il 2013 e 2023: valori medi nel tempo  
Data source: ARPA Lombardia



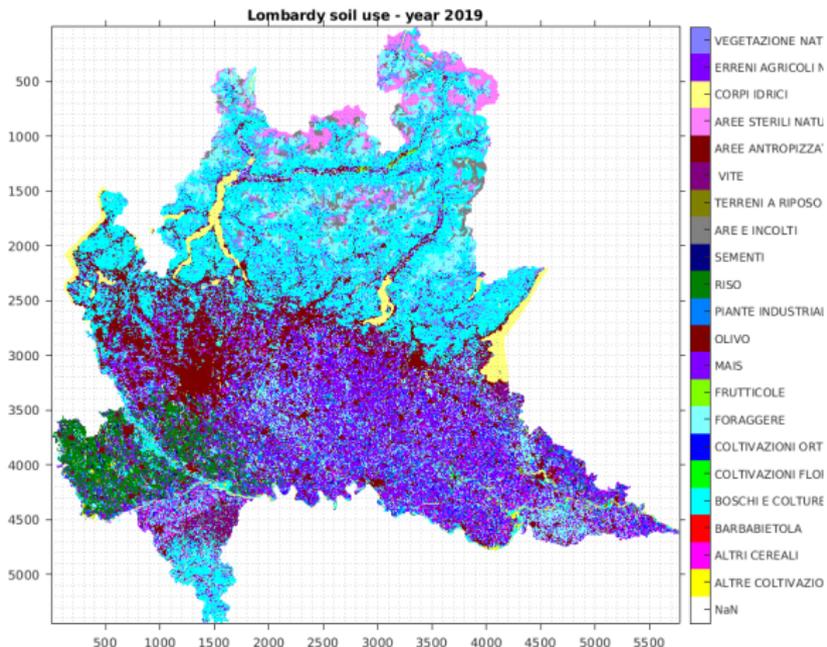
# Dati spazio-temporali: dati areali dinamici

Dati satellitari su griglia regolare o pixel sull'utilizzo del suolo in Lombardia nel 2016 (Fonte: Copernicus Corine Land Cover)



# Dati spazio-temporali: dati areali dinamici

Dati satellitari su griglia regolare o pixel sull'utilizzo del suolo in Lombardia nel 2019 (Fonte: Copernicus Corine Land Cover)

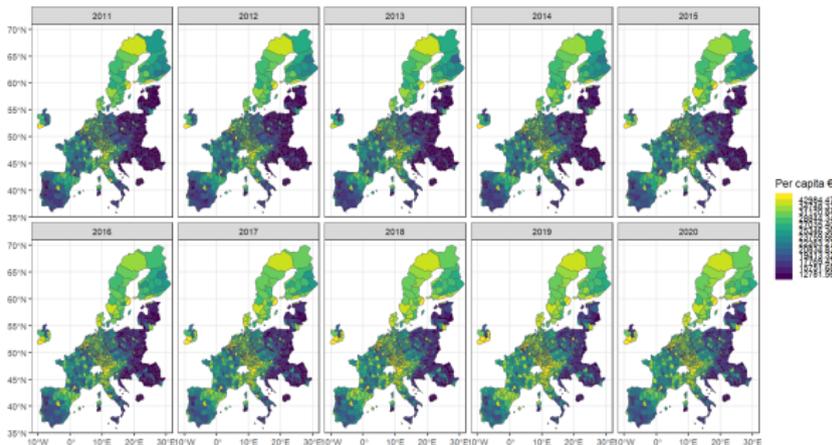


# Dati spazio-temporali: dati areali dinamici

Esempio di dato spazio-temporale areale su poligono: PIL pro capite (€ per abitante) misurate in 1088 province europee (NUTS-3) dal 2011 al 2020 (Fonte: Eurostat)

## Example of spatio-temporal areal data

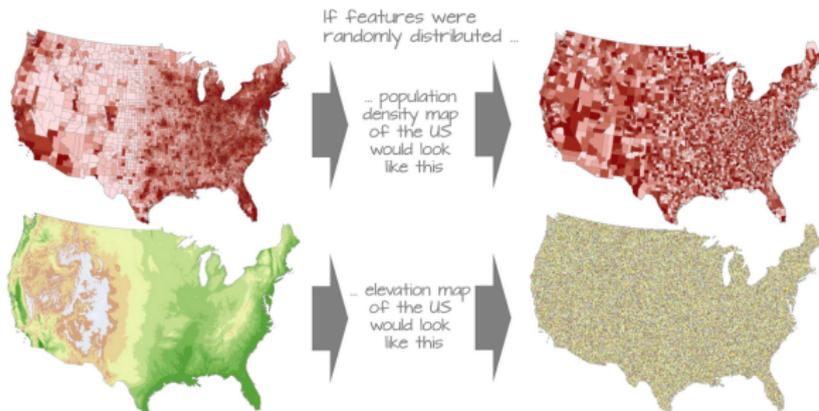
Per capita GDP (€ PPS) in the European Union provinces (NUTS-3 classification)



Data source: Eurostat

# Perchè i dati possono mostrare **regolarità nello spazio?**

**Prima legge della Geografia** o **dipendenza/autocorrelazione spaziale**: “tutto è correlato a tutto il resto, ma le cose vicine sono più correlate di quelle lontane” (Tobler, 1970). Wsempio: le quantità di inquinanti misurate in centraline geograficamente vicine saranno simili tra di loro.



# Perchè i dati possono mostrare **regolarità nello spazio?**

**Seconda legge della Geografia** or **eterogeneità spaziale**: le relazioni statistiche tra le variabili possono cambiare in un certo spazio in base alla geografia (Goodchild, 2004). Esempio: la relazione tra temperatura e inquinante può essere più forte in pianura che in montagna

Geographically Weighted Regression 441

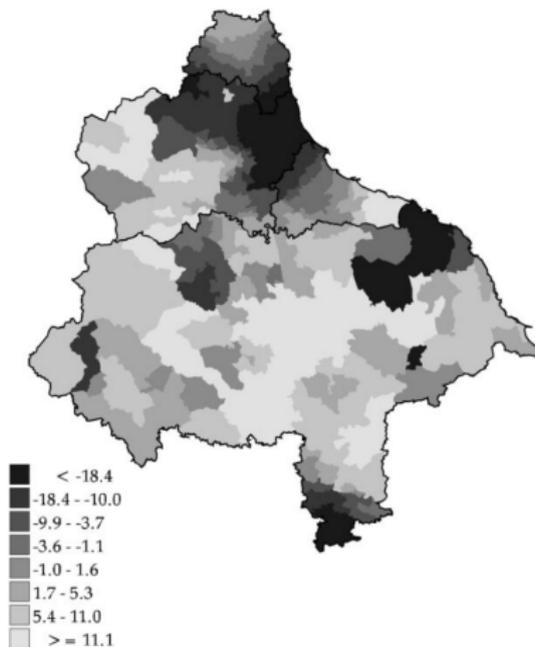


Fig. 2. Geographical variation of the density coefficient

# Perchè i dati possono mostrare **regolarità nello spazio?**

**Third Law of Geography** or **geographic similarity**: "Più le configurazioni geografiche (cioè la combinazione e la disposizione delle condizioni geografiche rilevanti intorno a un luogo) di due punti/aree sono vicine, più simili sono i valori misurati della variabile di interesse presso questi due punti/aree." (Zhu & Turner, 2022).

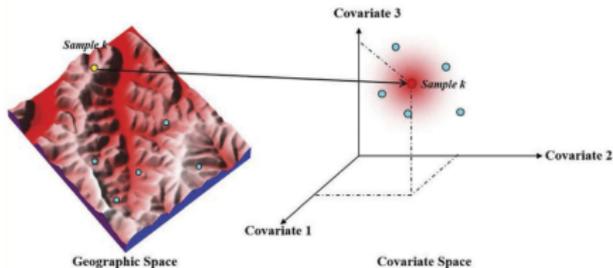
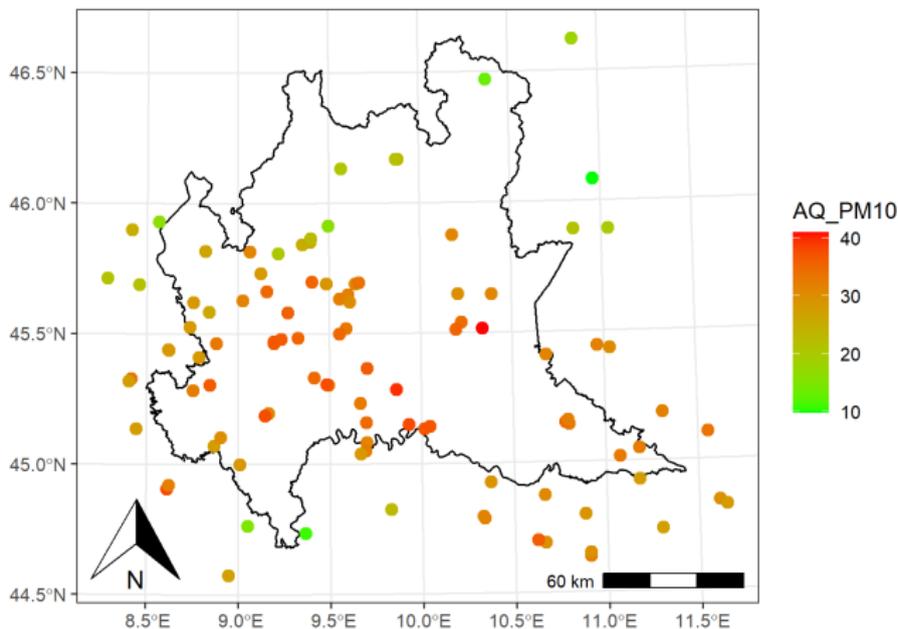


Figure 4. The representiveness of a single sample to a location of prediction can be quantified by the similarity between the geographic configurations of these two points.

# Proviamo a lavorare un pò sui dati della qualità dell'aria in Lombardia

Average PM<sub>10</sub> ( $\mu\text{g}/\text{m}^3$ ) concentrations in the extended Lombardy region



Grazie  
dell'attenzione e  
buon lavoro!