Time series analysis in environmental epidemiology

Doz. Dr. Hanns Moshammer
Environmental Health / ZPH

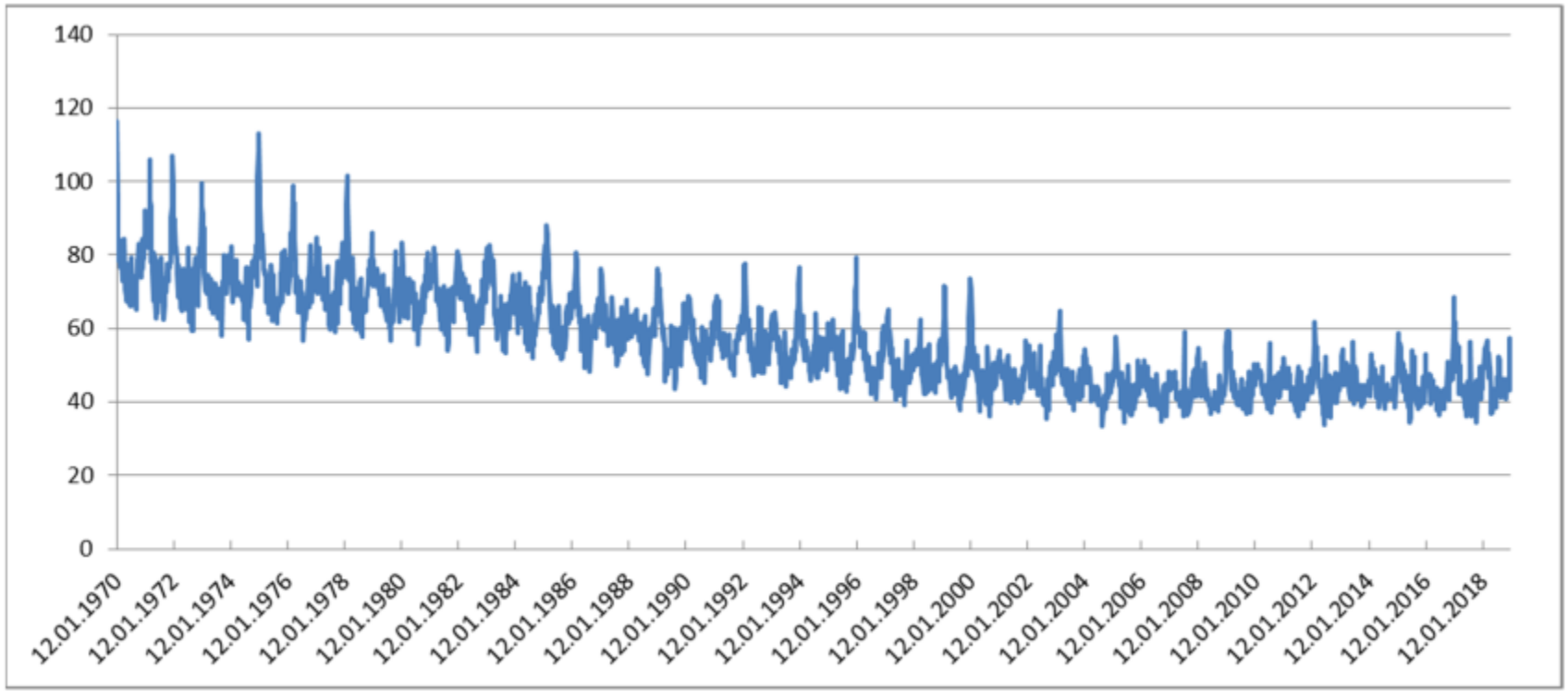
2021/11/10

Time series

An introduction

Time series

- Many variables vary temporally
  - Example: Number of daily deaths in Vienna (weekly averages)
- Underlying causes?



Time series

Strengths

- Easy access to data
- Large data-base
- No spatial confounding
- No confounding by individual traits
  - E.g. smoking, age, occupation

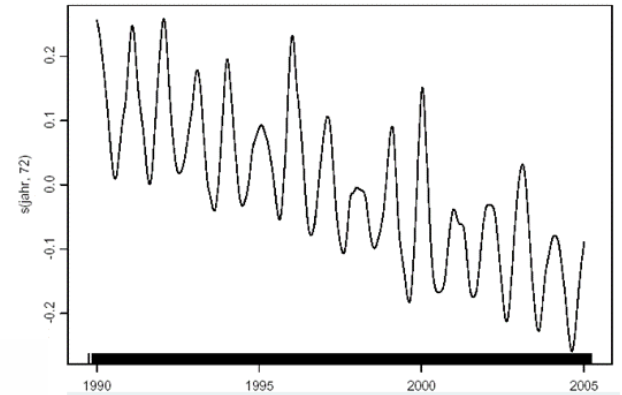
Weaknesses

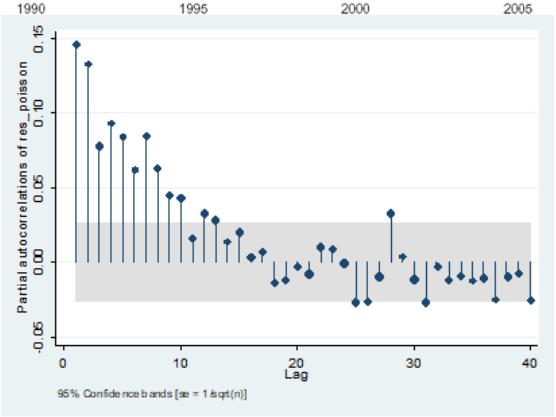
- No data on individual persons
- Confounding by temporal factors
  - Short- and long-term
- Unclear lag-structure
  - Immediate or retarded effects?
- Autocorrelation

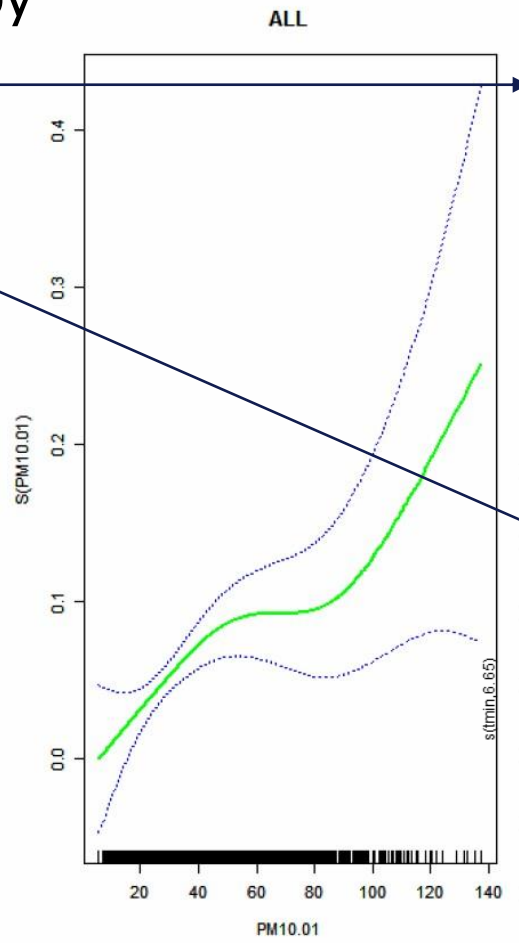
Time series

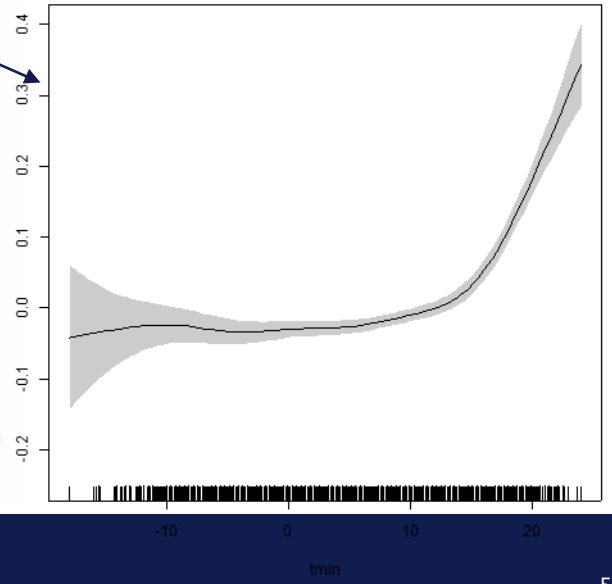
- Daily number of cases (here: deaths in Vienna)
- Control for seasonal and long-term trend by
  - Minimizing partial autocorrelation
- Control for meteorological factors
- Control for day of week, holidays

- Pollution effect
  - (same and previous day average in PM10)









Time series

What to consider

What you must consider I

- Duration of the effect (lag-structure)
  - Not only for pollutants, but also for meteorological factors:
  - Do you expect an immediate effect,
  - Or will an exposure today lead to more cases also on consecutive days?
  - Exposure and outcome of the same day?
  - Exposure and outcome on different days? (e.g. 1 day, 2 days, 3 days lag)
  - Exposure averaged over several days (e.g. 2 days, 1 week,…)
  - Weighted average (“distributed lag model”)

What you must consider II

- Long-term and seasonal variation in outcome
  - Also pollutants display a seasonal pattern
  - Confounding or over-adjustment?
    - Is seasonal pattern of outcome due to seasonal variation in pollution?
    - Or are seasonal patterns of outcome and pollution both affected by a third unmeasured factor?
  - Seasonal patterns not linear:
    - Approximation by sine-cosine function
    - Or non-parametric models (e.g. natural splines). Degrees of freedom?
      - How closely shall you smooth-out all “wriggles”?

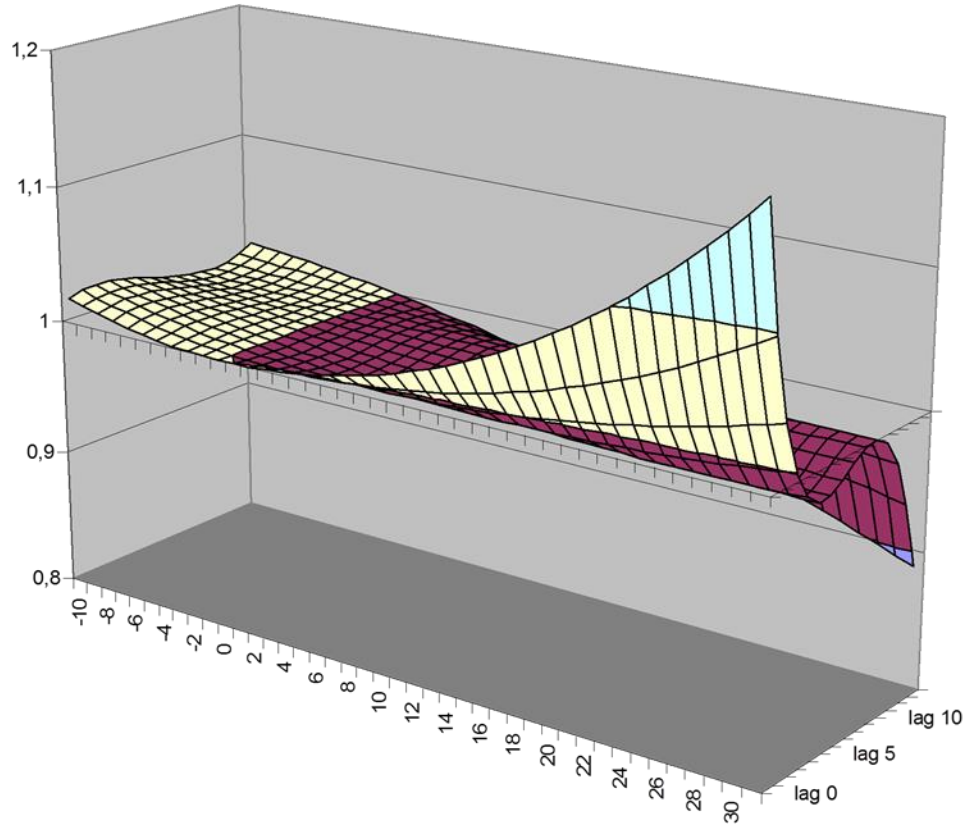
What you must consider III

- Meteorological factors
  - Temperature, humidity, air pressure
  - Subject to seasonal variation
  - Temperature is by itself a strong predictor of health and correlated with pollution
    - Possibly a strong confounder
  - Clearly a non-linear factor (both high and low temperatures are adverse!)
    - How to model temperature-outcome associations?
    - Same day or temperature averaged over several days?

Temperature effect

- Daily deaths in Vienna by temperature:
  - Hot temperatures:
  - Strong immediate effect
  - (and some “harvesting”)

  - Cold temperatures:
  - Weak effects over 14 days (and more)



Time series

Conclusion

Conclusion I

- Time series analysis is an elegant and straight-forward methodology
  - To demonstrate acute / immediate / short-term effects
  - By using existing data sets enabling study on large populations
  - Allowing the detection even of small effects

    Air pollution, recurring interventions, meteorological phenomena, etc.

  - Widely used in many settings around the world under different conditions
  - Several health end-points,
    - but number of deaths & number of hospital admissions most often used
    - Reliable and well-defined daily (or weekly) data

      Births, ER visits, ambulance calls, sick-leave and school absence, etc.

Conclusion II

- Time series analysis can still be tricky
  - Confounding by temporally co-varying factors (meteorological factors re air pollution)
  - How to control for seasonal variation?
  - How to chose the optimal lag structure?

  - How to communicate the results of time series analyses?
    - Often rather weak effects (PM air pollution RR per 10 µg/m³ around 1.01)
    - For example regarding deaths: some might only be brought forward by a few days
      - ("harvesting") – shall we count these deaths as well?
      - Individual risk versus population-wide risk

Thank you for your interest!

Doz. Dr. Hanns Moshammer

hanns.moshammer@meduniwien.ac.at

Umwelthygiene / ZPH

2021/11/10