Time series regression: advancements in this new tool for epidemiological analyses

Part II: multi-city analysis

Antonio Gasparrini and Ben Armstrong

London School of Hygiene and Tropical Medicine

23 February 2012

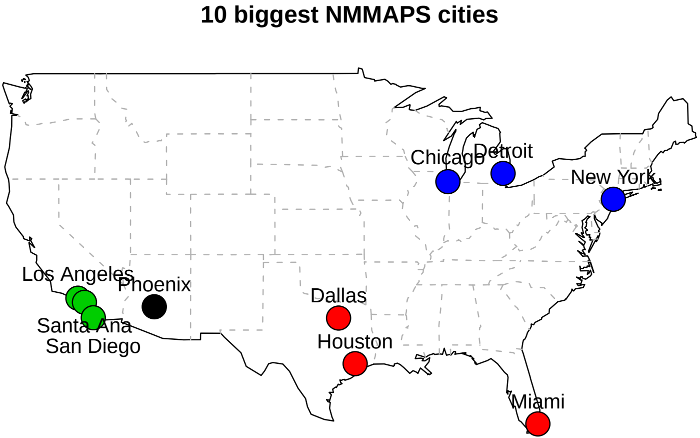
Multi-city analysis

Time series analysis on environmental stressors often involves data from multiple cities

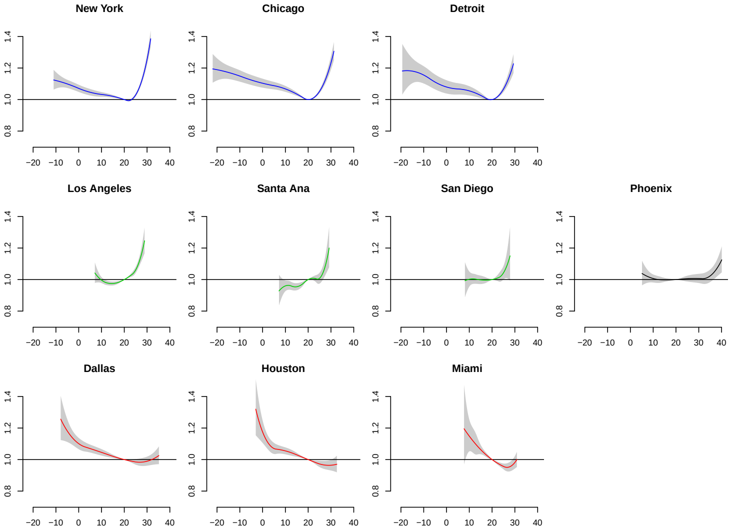
The reason: health effects usually change depending on **city-specific modifiers**, such as:

- climatic factors
- demographic factors
- socio-economic characteristics
- prevalence of air conditioning

An example


10 biggest NMMAPS cities
Chicago
Detroit
New York
Los Angeles
Phoenix
Dallas
Santa Ana
San Diego
Houston
Miami

Temperature and mortality



Multi-city analysis

The analytical framework is based on a **two-stage hierarchical design**:

1. A first-stage time series regression model to estimate the exposure-response while controlling for potential confounders **in each city**
2. A second-stage meta-analytical model to obtain a pooled estimate and investigate heterogeneity **across cities**

However, DLNMs hardly fit into this modelling approach

Simple approaches of pooling

Traditional meta-analytic techniques only works for pooling estimates of a **single parameter**

Adopting less sophisticated approaches, this can be achieved by:

- **Restricting**: seasonal analysis assuming a linear relationship
- **Simplifying**: linear-threshold parameterization
- **Averaging** over a predetermined lag period
- **Summarizing**: computing RR for specific absolute or relative temperatures

Limitations

These approaches, if generally appropriate, may not be suitable for investigating detailed associations. In particular:

- risk of **biases** due to wrong assumptions, or **limited info** on the true non-linear/delayed relationship
- **unbalance** between fairly complex first-stage models, compared to relatively simple second-stage meta-analytic procedures

What if we could **retain complexity** from the first-stage model, allowing the synthesis of more complex summary measures?

Multivariate meta-analysis

Traditionally, an extension of traditional meta-analysis to combine estimates of **multiple outcomes** from RCT

MV-meta may also be applied to combine the estimates of **multi-parameter associations** from different studies

In this case, the estimated coefficients $\hat{\boldsymbol{\theta}}_i$ of the function $s(x_t, \boldsymbol{\theta}_i)$, used in the first stage model to describe the association in each of the $i = 1, \ldots, m$ cities

Algebraic definition - I

Given the estimated $\hat{\boldsymbol{\theta}}_i$ and associated (co)variance matrix $\mathbf{S}_i$:

Within-study model

$$\hat{\boldsymbol{\theta}}_i \sim \mathrm{N}_k(\boldsymbol{\theta}_i\,,\,\mathbf{S}_i)$$

Between-study model

$$\boldsymbol{\theta}_i \sim \mathrm{N}_k(\boldsymbol{\theta}\,,\,\boldsymbol{\Psi})$$

with $\boldsymbol{\Psi}$ as the between-study (co)variance matrix

Algebraic definition - II

Marginally:

Multivariate meta-analysis

$$\hat{\boldsymbol{\theta}}_i \sim \mathrm{N}_k(\boldsymbol{\theta}\,,\,\mathbf{S}_i + \boldsymbol{\Psi})$$

Multivariate meta-regression

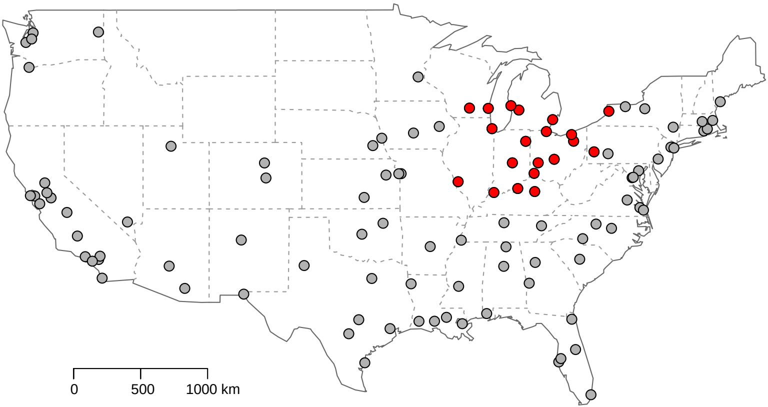
$$\hat{\boldsymbol{\theta}}_i \sim \mathrm{N}_k(\mathbf{X}_i\boldsymbol{\beta}\,,\,\mathbf{S}_i + \boldsymbol{\Psi})$$

with $\mathbf{X}_i$ as a design matrix obtained by city-level predictors $\mathbf{x}_i = [x_{1i}, x_{2i}, \ldots, x_{pi}]^\mathsf{T}$

$\boldsymbol{\theta}$ (or $\boldsymbol{\beta}$) and components of $\boldsymbol{\Psi}$ need to be estimated

An application

20 NMMAPS cities in the Industrial Mid-West region



Two-stage analysis

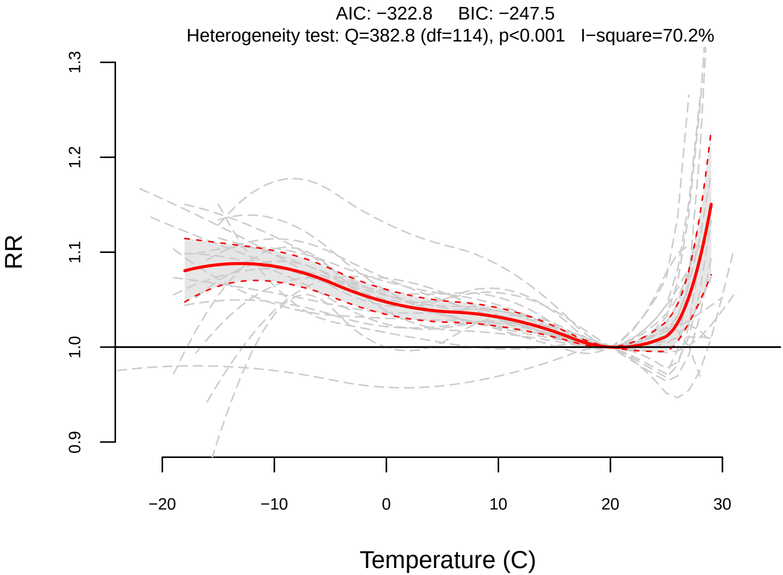
Investigating the association between temperature and mortality

First-stage time series regression model with a **quadratic B-spline**, with:

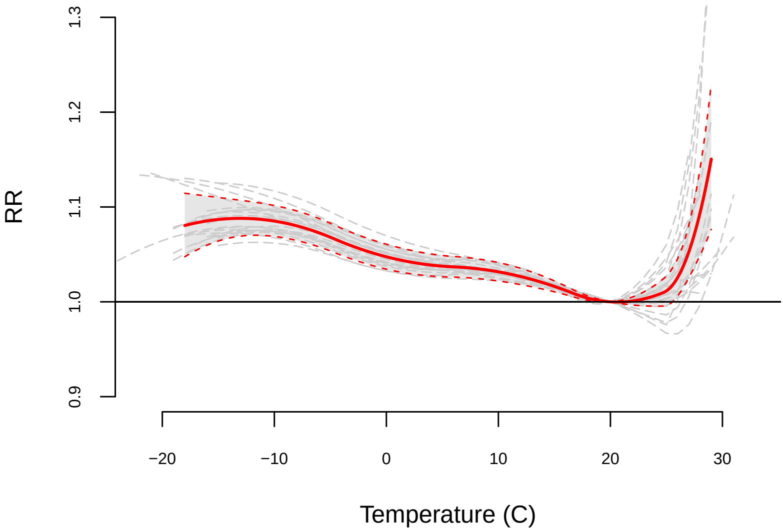
- Temperature averaged over lag 0-3
- 6 *df*
- 4 internal knots and 2 boundary knots placed at the same temperatures
- controlled for seasonality and day of the week

Then a second-stage multivariate meta-analysis and meta-regression

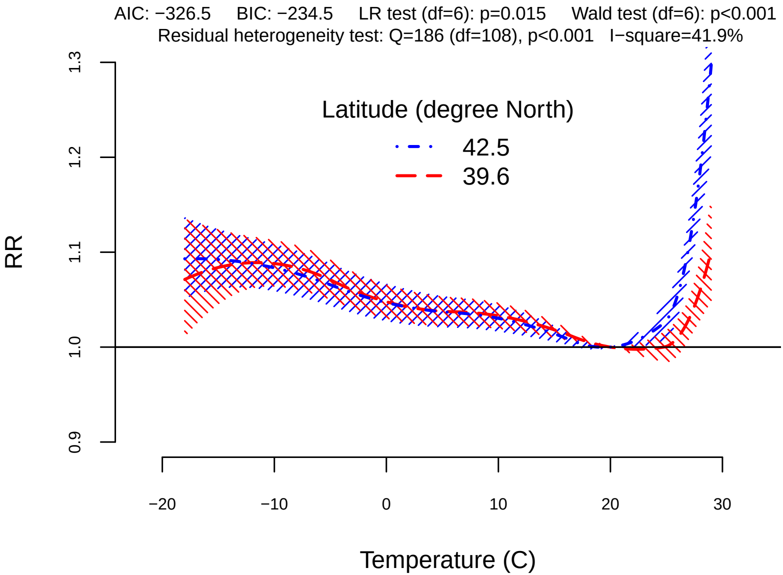
Pooled relationship



Best linear unbiased prediction



Multivariate meta-regression



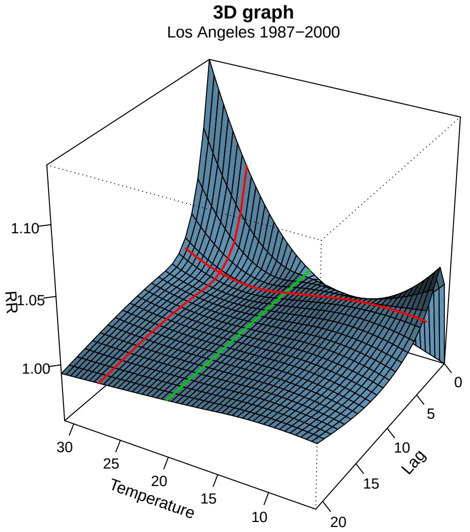
DLNMs: a reminder

They are specified through a **cross-basis**, a tensor product between the basis matrices $\mathbf{Z}$ and $\mathbf{C}$ for predictor and lag, with dimensions v_x and v_ℓ, giving:

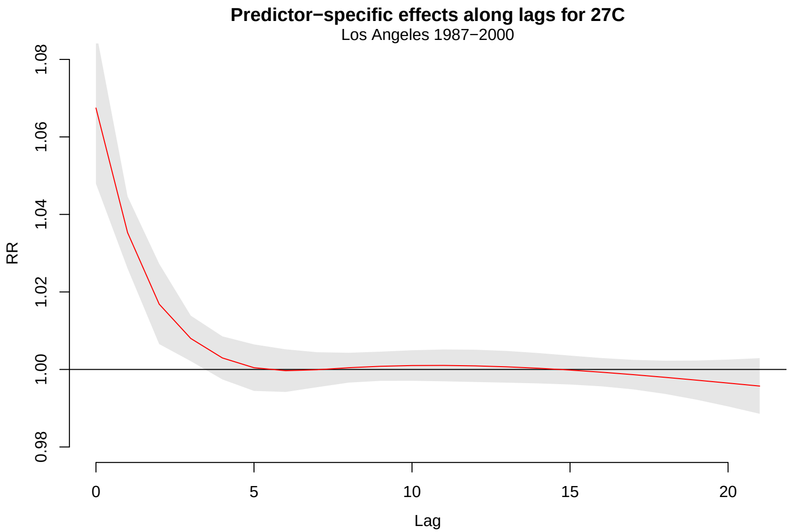
$$s(x_t;\boldsymbol{\theta}) = \sum_{j=1}^{v_x}\sum_{k=1}^{v_\ell} \mathbf{r}_{tj\cdot}^{\mathsf{T}}\mathbf{c}_{\cdot k}\theta_{jk} = \mathbf{w}_{t\cdot}^{\mathsf{T}}\boldsymbol{\theta} \tag{1}$$

The cross-basis matrix $\mathbf{W}$ has dimension $v_x \times v_\ell$: for complex models, this **dimensionality is not compatible** with multivariate meta-analysis

Bi-dimensional relationship

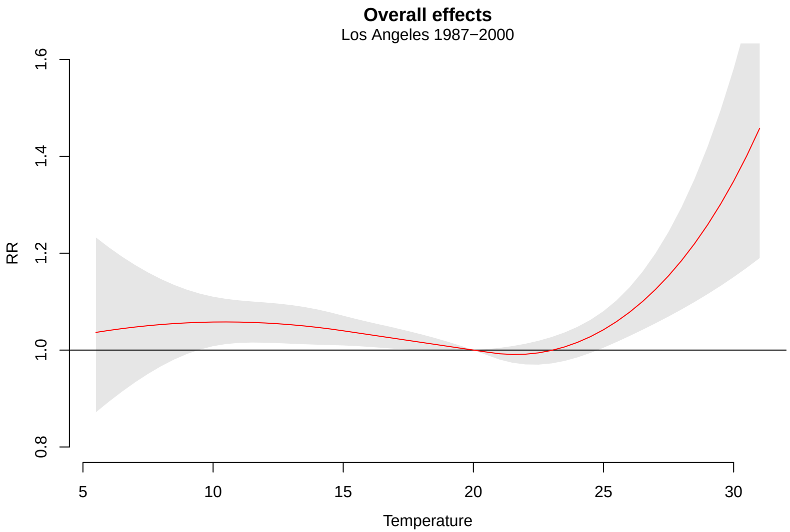


Summary - I



Predictor−specific effects along lags for 27C
Los Angeles 1987−2000
RR
1.08
1.06
1.04
1.02
1.00
0.98
0
5
10
15
20
Lag

Summary - II



Overall effects
Los Angeles 1987−2000
RR
0.8
1.0
1.2
1.4
1.6
5
10
15
20
25
30
Temperature

Reducing DLNMs - I

For a fitted DLNM with estimated parameters $\hat{\boldsymbol{\theta}}$, **summaries of the fit** may be re-expressed in terms of reduced set of parameters $\hat{\boldsymbol{\eta}}$ of original one-dimensional bases $\mathbf{Z}$ or $\mathbf{C}$.

These reduced parameters are computed through a **transformation matrix** $\mathbf{M}$, by:

$$\hat{\boldsymbol{\eta}} = \mathbf{M}\hat{\boldsymbol{\theta}}$$

$$\mathrm{V}(\hat{\boldsymbol{\eta}}) = \mathbf{M}\mathrm{V}(\hat{\boldsymbol{\theta}})\mathbf{M}^\mathsf{T}$$

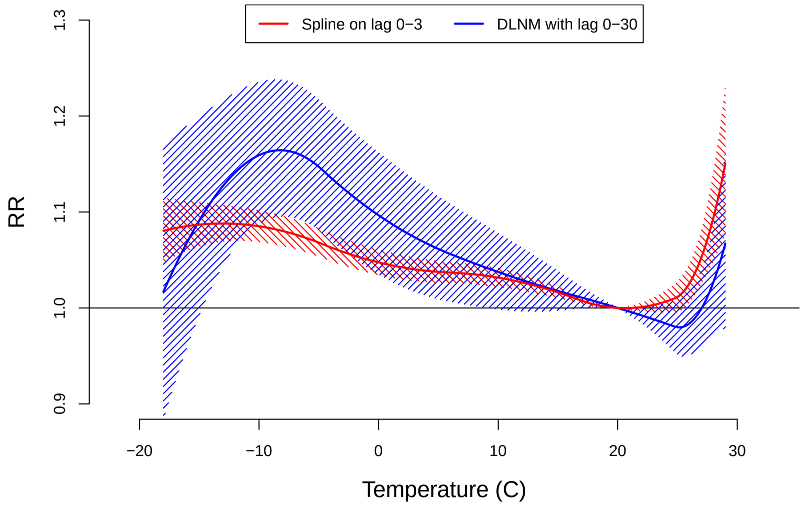
Reducing DLNMs - II

The computation of the matrix $\mathbf{M}$ is dependent on the type of summary:

$$\mathbf{M} = \begin{cases} \mathbf{I}_{(v_\ell)} \otimes \mathbf{z}_{[x_0]}^{\mathrm{T}} & \text{for predictor-specific effects at } x_0 \\ \mathbf{c}_{[\ell_0]}^{\mathrm{T}} \otimes \mathbf{I}_{(v_x)} & \text{for lag-specific effects at } \ell_0 \\ \mathbf{1}_{(L+1)}^{\mathrm{T}} \mathbf{C} \otimes \mathbf{I}_{(v_x)} & \text{for overall effects,} \end{cases}$$

Overall effects are computed as $\mathbf{Z}\hat{\boldsymbol{\eta}}$, predictor-specific effects as $\mathbf{C}\hat{\boldsymbol{\eta}}$, with a **reduced dimensions** v_x and v_ℓ, respectively, more compatible with MV-meta models

Comparison



Spline on lag 0−3
DLNM with lag 0−30
RR
1.3
1.2
1.1
1.0
0.9
−20
−10
0
10
20
30
Temperature (C)

DLNMs and MV-meta

Distributed lag non-linear models and **multivariate meta-analysis** represent useful statistical tools for time series analysis of environmental factors

The methodologies are implemented in the two R packages dlnm and mvmeta, both available on CRAN

Methodologies potentially applicable **beyond time-series analysis**

Next publications

Multivariate meta-analysis for non-linear and other multi-parameter associations. Submitted to *Statistics in Medicine* (with R code)

Reducing and meta-analyzing distributed lag non-linear models. To be submitted soon (with R code)

A general statistical framework for exposure-time-response relationships based on distributed lag models. To be submitted soon (with R code)

R packages

http://cran.r-project.org/web/packages/dlnm/index.html

http://cran.r-project.org/web/packages/mvmeta/index.html

Further questions

antonio.gasparrini@lshtm.ac.uk