Temporal and spatial statistical methods to remove external effects on groundwater levels with R

Torino R net, 29th November 2012

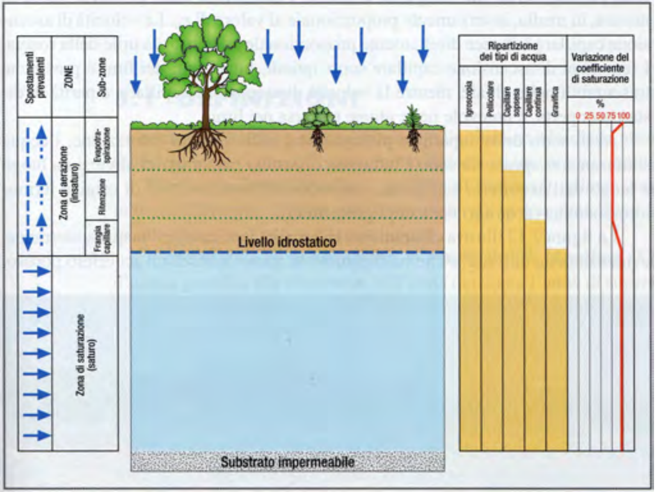
Daniele Imparato[1], Andrea Carena[2] and Mauro Gasparini[2]

[1] Dep. of Economics, Università dell'Insubria, Varese
[2] Dep. of Mathematical Sciences, Politecnico di Torino

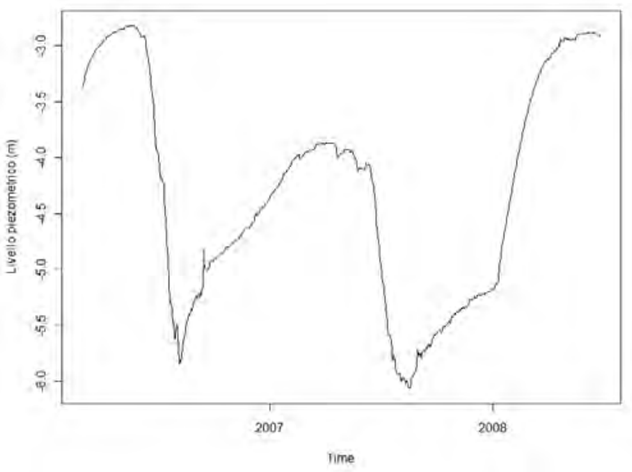
Problem overview

- Joint collaboration with Regione Piemonte - Direzione Ambiente - Settore Tutela Ambientale delle Acque
- **Initial goal**: monitoring the statistical trend of groundwater levels

Geological description of groundwater



An example: groundwater time series near Lobbi



The regulatory framework

Parte B - Stato quantitativo

Nella Tabella 4 è riportata la definizione di buono stato quantitativo delle acque sotterranee

Tabella 4- Definizione di buono stato quantitativo

Elementi	Stato buono
Livello delle acque sotterranee	Il livello/portata di acque sotterranee nel corpo sotterraneo è tale che la media annua dell'estrazione a lungo termine non esaurisca le risorse idriche sotterranee disponibili. Di conseguenza, il livello delle acque sotterranee non subisce alterazioni antropiche tali da: -impedire il conseguimento degli obiettivi ecologici specificati per le acque superficiali connesse; -comportare un deterioramento significativo della qualità di tali acque; -recare danni significativi agli ecosistemi terrestri direttamente dipendenti dal corpo idrico sotterraneo. Inoltre, alterazioni della direzione di flusso risultanti da variazioni del livello possono verificarsi, su base temporanea o permanente, in un'area delimitata nello spazio; tali inversioni non causano tuttavia l'intrusione di acqua salata o di altro tipo né imprimono alla direzione di flusso alcuna tendenza antropica duratura e chiaramente identificabile che possa determinare siffatte intrusioni. Un importante elemento da prendere in considerazione al fine della valutazione dello stato quantitativo è inoltre, specialmente per i complessi idrogeologici alluvionali, l'andamento nel tempo del livello piezometrico. Qualora tale andamento, evidenziato ad esempio con il metodo della regressione lineare, sia positivo o stazionario, lo stato quantitativo del copro idrico è definito buono. Ai fini dell'ottenimento di un risultato omogeneo è bene che l'intervallo temporale ed il numero di misure scelte per la valutazione del trend siano confrontabili tra le diverse aree. E' evidente che un intervallo di osservazione lungo permetterà di ottenere dei risultati meno influenzati da variazioni naturali (tipo anni particolarmente siccitosi).

La media annua dell'estrazione a lungo termine di acque sotterranee è da ritenersi tale da non esaurire le risorse idriche qualora non si delineino diminuzioni significative, ovvero trend negativi significativi, delle medesime risorse.
Ai fini della valutazione della conformità a dette condizioni, è necessario, nell'ambito della revisione dei piani di gestione e dei piani di tutela da pubblicare nel 2015, acquisire le informazioni utili a valutare il bilancio idrico.

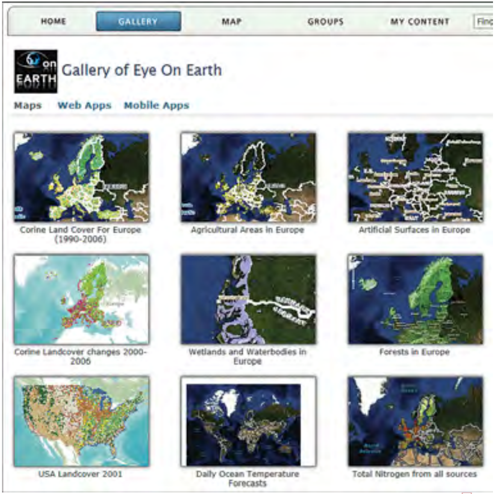
Final aim of the project

- To model the neighboring effects of rain precipitation and waterways on groundwater levels
- To reconstruct a groundwater virgin level as an indicator of the state of health of groundwater itself

The area of interest

Geographic Information System (GIS)

- A Geographic Information System (GIS) is a system to visualize, question, analyze geographic data.



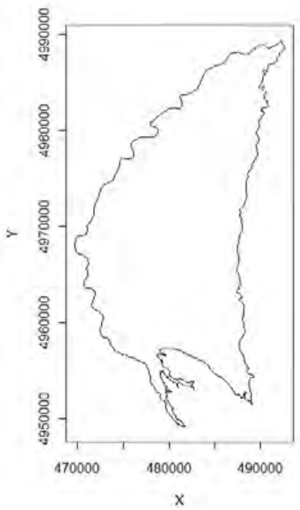
Shapefile

- A shapefile is a popular geospatial vector data format produced by GIS
- It contains general information on the geography (points, lines, polylines, area, perimeter, etc.)
- It consists of the following files:
  - .shp - file of geometries (main file);
  - .shx - index file;
  - .dbf - database file

GIS and R

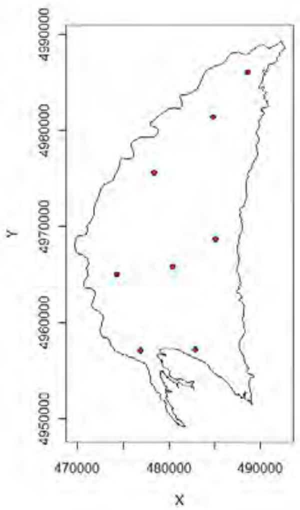
- **R-GIS project**: an incubator for developing spatial packages `http://r-forge.r-project.org/projects/r-gis/`
- In practical applications, it is often necessary to import a shapefile to R.
- the package **shapefiles** includes functions to read and write shapefiles (`read.shapefile` and `write.shapefile`)
- It converts easily dataframes to shapefiles and viceversa - see e.g. `convert.to.shapefile` and `convert.to.simple`

Our data



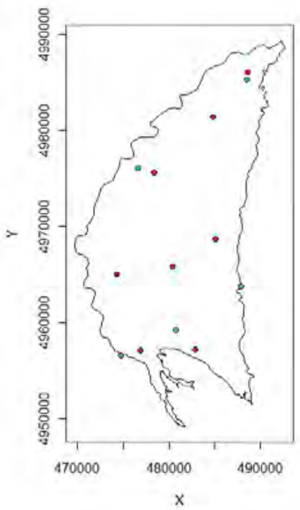
- Shapefile

Our data



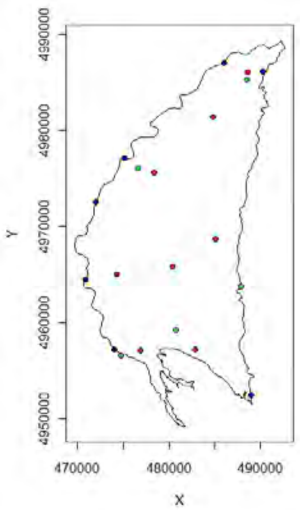
- Shapefile
- 8 piezometers

Our data



- Shapefile
- 8 piezometers
- 5 rain gauges

Our data

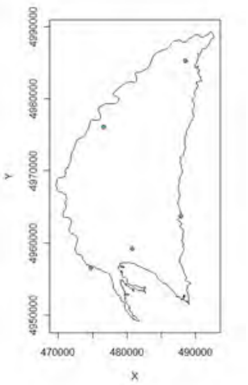


- Shapefile
- 8 piezometers
- 5 rain gauges
- 7 hydrometers

Steps of the statistical analysis

- Estimating the cumulative measure of rain precipitation on the area considered
- Capturing the contribution of rain on the piezometric level and removing such effect
- Capturing the contribution of near waterways on the piezometric level and removing such effect
- Removing jointly both contributions to get a *virgin* level of groundwater

Rain contribution



- Rain precipitation is known only on 5 observation points on a daily scale: this contribution is negligible for the geology of groundwater.
- A cumulative datum for rain precipitation is necessary
- The rain surface on the area is estimated through a **kriging** approach [Diggle and Ribeiro (2007)]

Spatial model

Let $Z_i : i = 1, \ldots, 5$ denote the rain precipitation measured by the i-th pluviometer: we assume

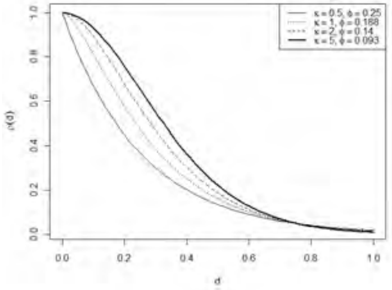
$$\sqrt{Z_i} = S(\mathbf{x}_i) + N_i$$

where $N_i \sim \mathcal{N}(0, \tau^2)$ is a white noise and $S(\mathbf{x})$ is a stationary Gaussian process with constant (unknown) mean $\mu(\mathbf{x}) = \beta$ and variance σ^2.

Spatial correlation

We assume a Matérn-type spatial correlation:

$$\rho(d) = (d/\phi)^k \, B_k(d/\phi) \left\{2^{k-1}\Gamma(k)\right\}^{-1}$$



Model parameters

The parameters are estimated through a Maximum Likelihood approach.

- Correlation parameters:
  - shape parameter k
  - scale parameter ϕ
- variances:
  - variance of the process σ^2
  - variance of error (*nugget*) $\tau^2 = \mathrm{Var}[Z_i | S(\mathbf{x}_i)]$

Kriging

Definition

For a stationary Gaussian model, the **kriging** predictor $\hat{T}(\mathbf{x})$ of $T = T(S(\mathbf{x}))$ minimizes the mean squared error.

$$\hat{T}(\mathbf{x}) = \mu(\mathbf{x}) + \sum_{i=1}^{n} w_i(\mathbf{x})[z_i - \mu(\mathbf{x})]$$

- Simple kriging

$$\begin{cases} \text{min} & E[\{\hat{T}(\mathbf{x}) - T(\mathbf{x})\}^2] \\ \text{sub} & \hat{T}(\mathbf{x}) = \{1 - \sum_i w_i(\mathbf{x})\}\mu(\mathbf{x}) + \sum_i w_i(\mathbf{x}) z_i \end{cases}$$

Kriging

Definition

For a stationary Gaussian model, the **kriging** predictor $\hat{T}(\mathbf{x})$ of $T = T(S(\mathbf{x}))$ minimizes the mean squared error.

$$\hat{T}(\mathbf{x}) = \mu(\mathbf{x}) + \sum_{i=1}^{n} w_i(\mathbf{x})[z_i - \mu(\mathbf{x})]$$

- Simple kriging
- Ordinary kriging

$$\begin{cases} \min & E[\{\hat{T}(\mathbf{x}) - T(\mathbf{x})\}^2] \\ \text{sub} & \hat{T}(\mathbf{x}) = \{1 - \sum_i w_i(\mathbf{x})\}\mu(\mathbf{x}) + \sum_i w_i(\mathbf{x}) z_i \\ \text{sub} & \sum_i w_i(\mathbf{x}) = 1 \end{cases}$$

Kriging

Definition

For a stationary Gaussian model, the **kriging** predictor $\hat{T}(\mathbf{x})$ of $T = T(S(\mathbf{x}))$ minimizes the mean squared error.

$$\hat{T}(\mathbf{x}) = \mu(\mathbf{x}) + \sum_{i=1}^{n} w_i(\mathbf{x})[z_i - \mu(\mathbf{x})]$$

- Simple kriging
- Ordinary kriging

$$\begin{cases} \min & E[\{\hat{T}(\mathbf{x}) - T(\mathbf{x})\}^2] \\ \text{sub} & \hat{T}(\mathbf{x}) = \{1 - \sum_i w_i(\mathbf{x})\}\mu(\mathbf{x}) + \sum_i w_i(\mathbf{x}) z_i \\ \text{sub} & \sum_i w_i(\mathbf{x}) = 1 \end{cases}$$

R implementation

The package **geoR**

R implementation

The package **geoR**

1. **likfit**
   - input: structure of spatial correlation
   - output: parameter estimation via ML

R implementation

The package **geoR**

1. **likfit**
    - input: structure of spatial correlation
    - output: parameter estimation via ML

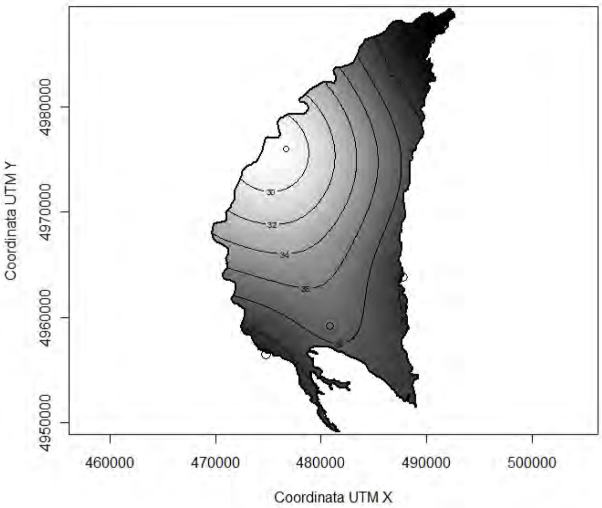
2. **krige.control**
    - creates a list defining the model components and the type of kriging to be performed.

R implementation

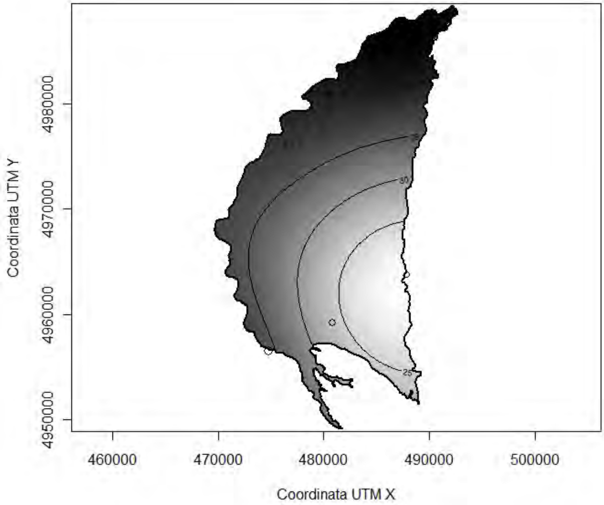
The package **geoR**

1. **likfit**
   - input: structure of spatial correlation
   - output: parameter estimation via ML
2. **krige.control**
   - creates a list defining the model components and the type of kriging to be performed.
3. **krige.conv**
   - kriging prediction over a grid of points
   - it can read the borders of an area from a shapefile

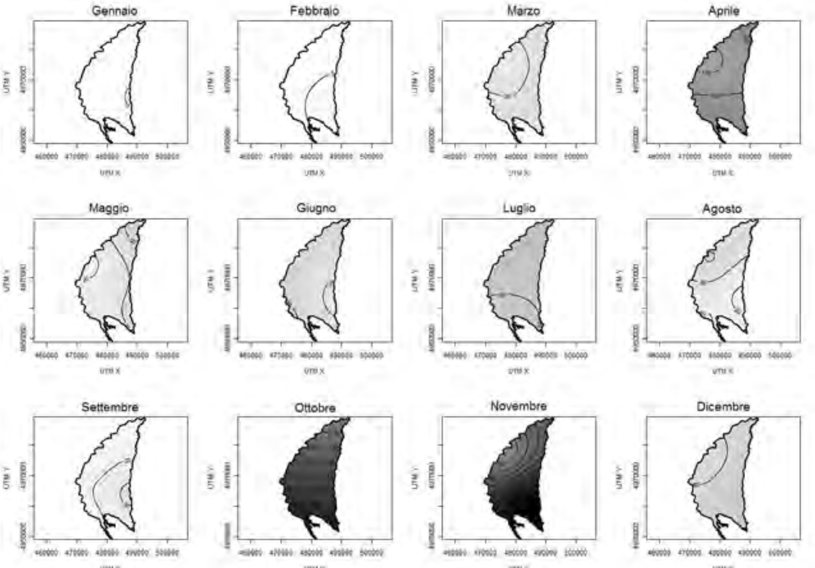
Examples: rain surface prediction on the area in March 2000



Examples: rain surface prediction on the area in September 2007



Examples: monthly rain surface prediction on the area in 2000



Contributions of rain and waterways on groundwater

- Choice of the optimal time scaling
  - Monthly scale
  - 2 week-shifted daily scale
  - weekly scale
  - **daily scale**
- **Transfer Function Model** implementation [Box and Jenkins (1970)]

Transfer Function Model

Definition

A **transfer function model** between two stationary processes Y_t and X_t is defined through

$$Y_t = \sum_{i=0}^{\infty} \nu_i X_{t-i} + N_t$$

Model identification and estimation

- If $\{X_t\}$ is a white noise, the coefficients ν_i are estimated through

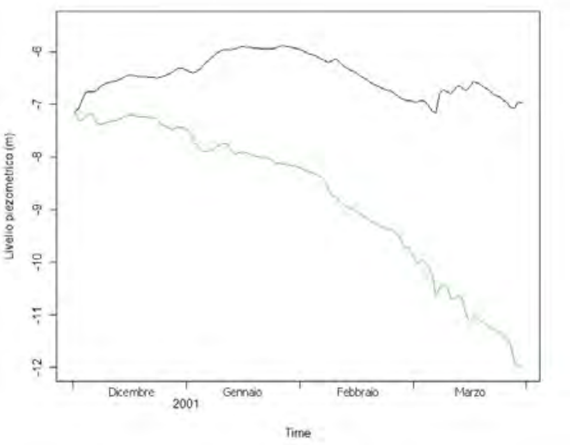
$$\nu_l = \frac{\gamma_{xy}(l)}{\gamma_x(0)}$$

- The significant lags are identified through the correlogram $\gamma_{xy}(l)$
- If $\{X_t\}$ is a general stationary process, a pre-whitening of both series is needed using Box-Jenkins modelization.

Removing rain contribution I

Piezometric level near Tortona

$$Y_t = 0.013\,X_t + 0.008\,X_{t-1} + \varepsilon_t$$



Removing rain contribution II

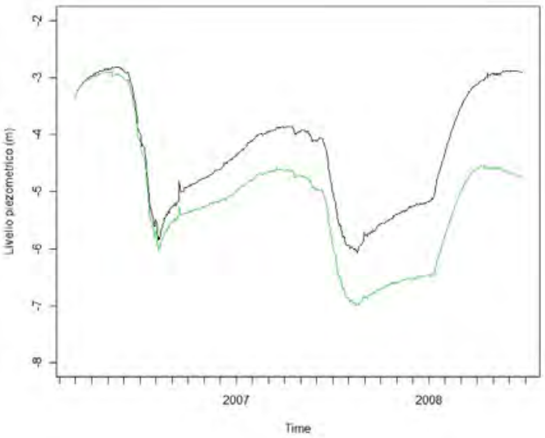
Piezometric level at Lobbi

$$Y_t = 0.00218\, X_t - 0.00066\, X_{t-1} + \eta_t$$

Removing rain contribution II

Piezometric level at Lobbi

$$Y_t = 0.00218\, X_t - 0.00066\, X_{t-1} + \eta_t$$



Removing waterway contribution

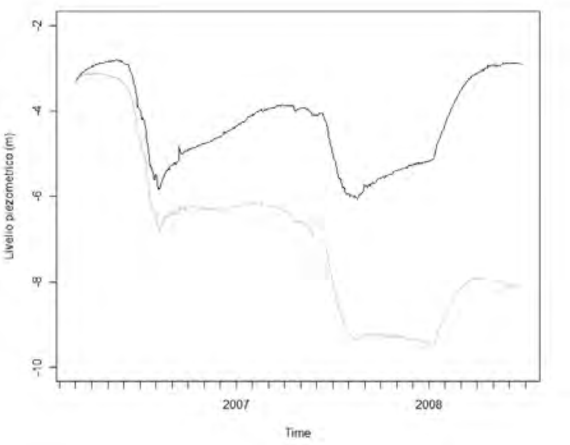
Piezometric level at Lobbi, near Tanaro river (Montecastello)

$$Y_t = 0.0402\, W_t - 0.0419\, W_{t-1} + \epsilon_t$$

Removing waterway contribution

Piezometric level at Lobbi, near Tanaro river (Montecastello)

$$Y_t = 0.0402\, W_t - 0.0419\, W_{t-1} + \epsilon_t$$



Removing joint effects

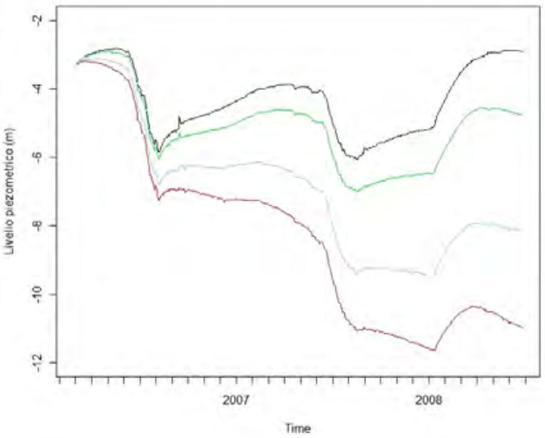
Piezometric level near Lobbi

$$Y_t = 0.04021\, W_t - 0.04186\, W_{t-1} + 0.00082\, X_t + \xi_t$$

Removing joint effects

Piezometric level near Lobbi

$$Y_t = 0.04021\, W_t - 0.04186\, W_{t-1} + 0.00082\, X_t + \xi_t$$



Conclusions and future developements

- Transfer function models are a valid alternative to a deterministic modelization through PDE
- Interesting similarity seems to hold between the estimated groundwater virgin levels and the so-called "exhaustion curves" of water springs
- Alarm thresholds for such predictions may be discussed, aimed at building semi-automatic procedures for controlling locally the health of the groundwater
- Improved control over private irrigation

Bibliography (1/2)

Ailliot P., Thompson C., Thomson P.,
Space-time modelling of precipitation by using a hidden Markov model and censored Gaussian distributions.
Journal Of The Royal Statistical Society Series C, 58, p. 405-426, 2009.

Battaglia F.,
Metodi di previsione statistica,
Springer-Verlag Italia, Milano, 2007.

Bove A., Destefanis E., De Luca D. A., Masciocco L., Ossella L., Tonussi M.,
Studio idrogeologico finalizzato alla caratterizzazione dell'acquifero superficiale nel territorio di pianura della provincia di Alessandria.
Dipartimento di Scienze della Terra, Universit di Torino, 2005.

Civita M.,
Idrogeologia applicata e ambientale,
CEA, Casa Editrice Ambrosiana, 2005.

Di Molfetta A.,
Ingegneria degli acquiferi,
Politeko Edizioni, seconda edizione, Torino, 2003.

Diggle P.J., Ribeiro Jr.P.J.,
Model-based Geostatistics,
Springer Series in Statistics, 2007.

Bibliography (2/2)

Haining R.,
Spatial Data Analysis, Theory and Practice, Cambridge University Press, 2005.

Karnieli A.,
Application of Kriging Technique to Areal Precipitation Mapping in Arizona.
GeoJournal, 22.4, 391-398, 1990.

Imparato D., Carena A., Gasparini M.
Temporal and spatial statistical methods to remove external effects on groundwater levels. In *Advanced Statistical Methods for the Analysis of Large Data-Sets* Di Ciaccio A., Coli M. and Ibanez, J.M.A. Eds., Springer, 241-251, 2012.

Ravines R.R., Schmidt A.M., Migon H.S., Renn C.D.,
A joint model for rainfall-runoff: the case of Rio Grande Basin.
Journal of Hydrology, 353, 189-200, 2008.

Shumway R.H., Stoffer D.S.,
Time Series Analysis and Its Applications,
Springer Texts in Statistics, second edition, 2006.

Yi M.-J., Lee K.-K.,
Transfer function-noise modelling of irregularly observed groundwater heads using precipitation data.
Journal of Hydrology, 288, 272-287, 2004.