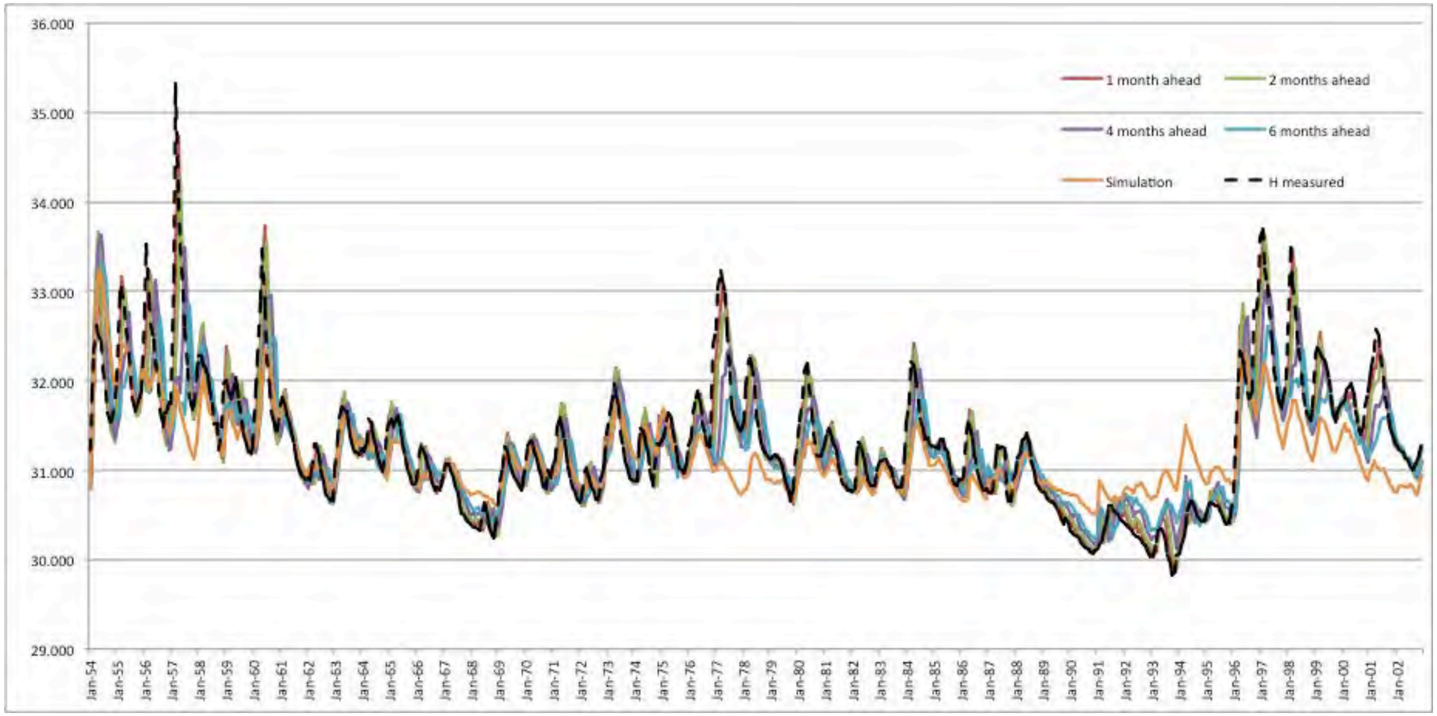
7th GIT Meeting – Bologna 13-15 June 2012

DATA-MINING IN ENGINEERING GEOLOGY



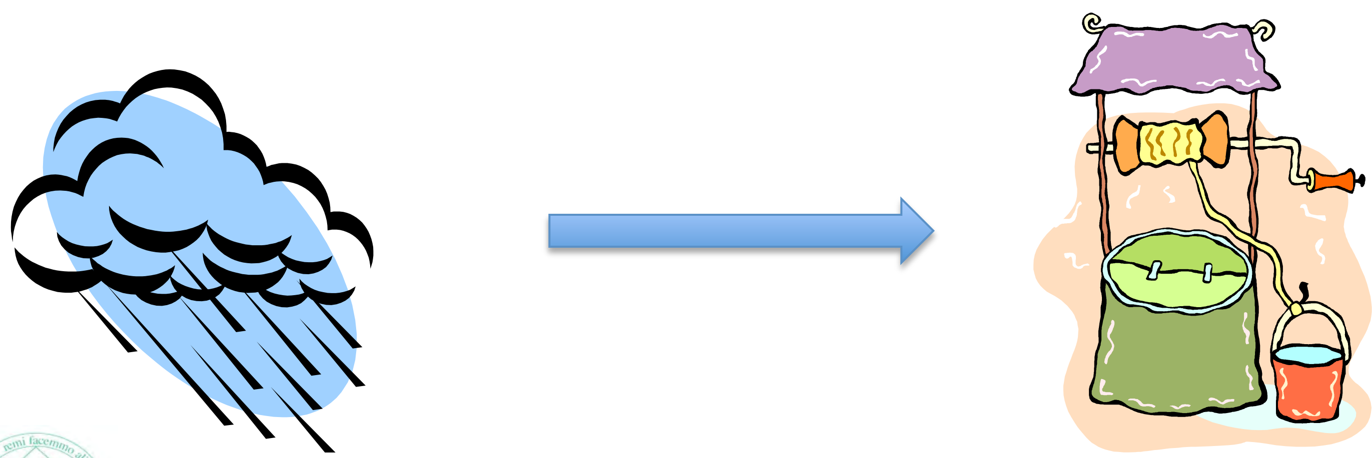
Angelo Doglioni, Vincenzo Simeone

Technical University of Bari

Department of Civil Engineering and Architecture

Introduction

- Data-mining applied to natural systems
- Dynamic response of groundwater to rainfall
- How can I do?
- Some specific applications



Data-Driven models

Measured data
A priori physical knowledge → **Model** → **Prediction Simulation Scientific knowledge**

- *The conceptualisation of an "environmental" system is commonly made by inferring models from observations and studying their properties (Ljung, 1999).*
- The functional form of relationships between variables and the numerical parameters in those functions are unknown and need to be estimated. (Statistical or Regressive or Data-Driven)

Data-mining: mathematical formulation

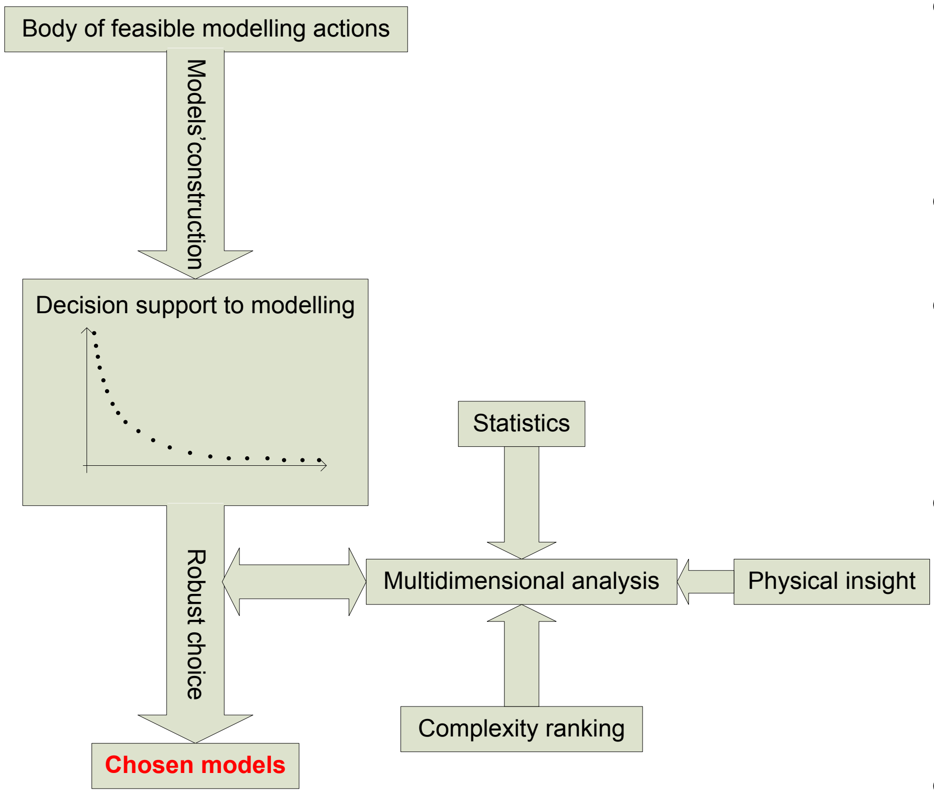
- From a modelling point of view, a physical system having an output value *y* dependent on a set of inputs **X** and parameters **q**, can be mathematically formalized as:

$$y = F(\mathbf{X}, \boldsymbol{\theta}) \quad \mathbf{X} \in \Re^m$$

- where *F* is a function in the space dimensionally equal to the number of inputs *m*.

- Data-driven techniques aim at reconstructing *F* from input/output data.

Multi-Objective Evolutionary Polynomial Regression (EPR-MOGA)



- Search for **structurally parsimonious formulas**
- **Fitness to data**
- Set of **non-dominated models**
- **Comparisons among non-dominated equations**
- **Analysis of similar input** across equations

EPR - MOGA

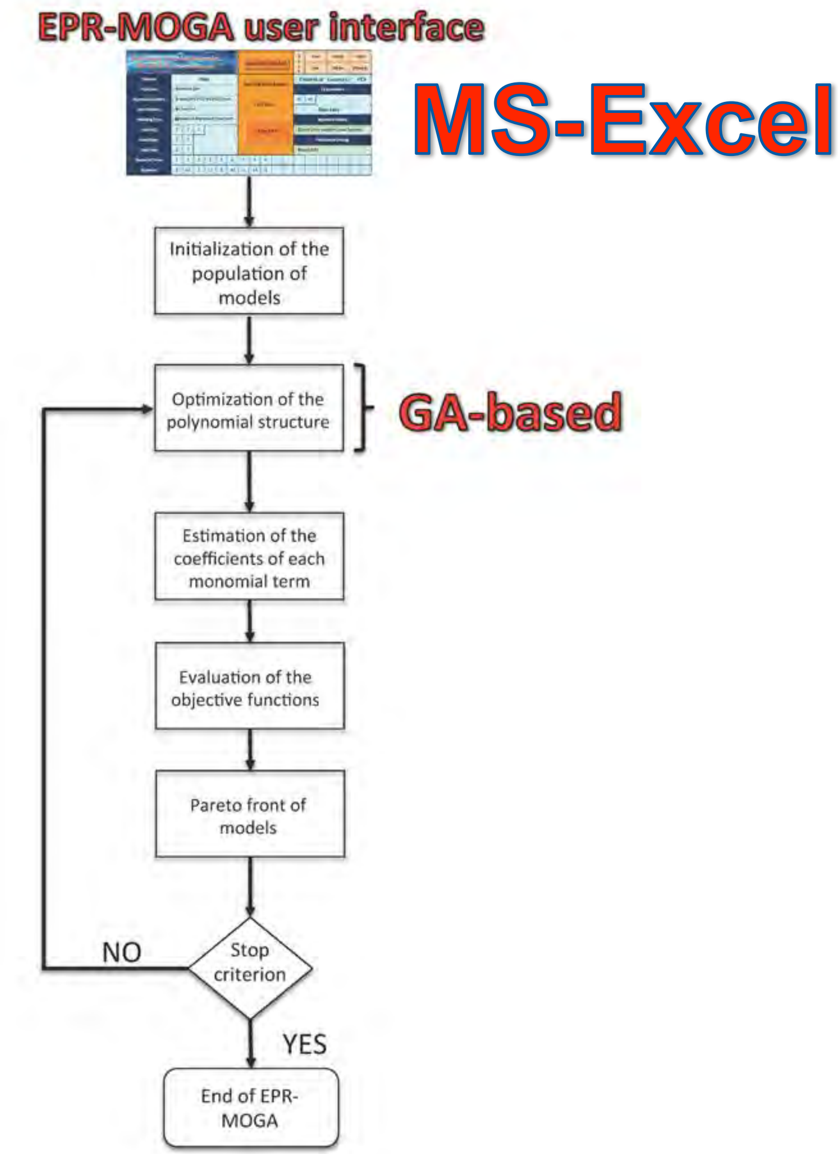
$$\hat{y} = \sum_{j=1}^{m} F\left(\mathbf{X}, f\left(\mathbf{X}\right), a_j\right) + a_o$$

$$\mathbf{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot f\left(\mathbf{X}_1 \cdot \ldots \cdot \mathbf{X}_k\right)$$

$$\mathbf{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot \left[\mathbf{X}_1 \cdot \ldots \cdot \mathbf{X}_k \cdot f\left(\mathbf{X}_1 \cdot \ldots \cdot \mathbf{X}_k\right)\right]$$

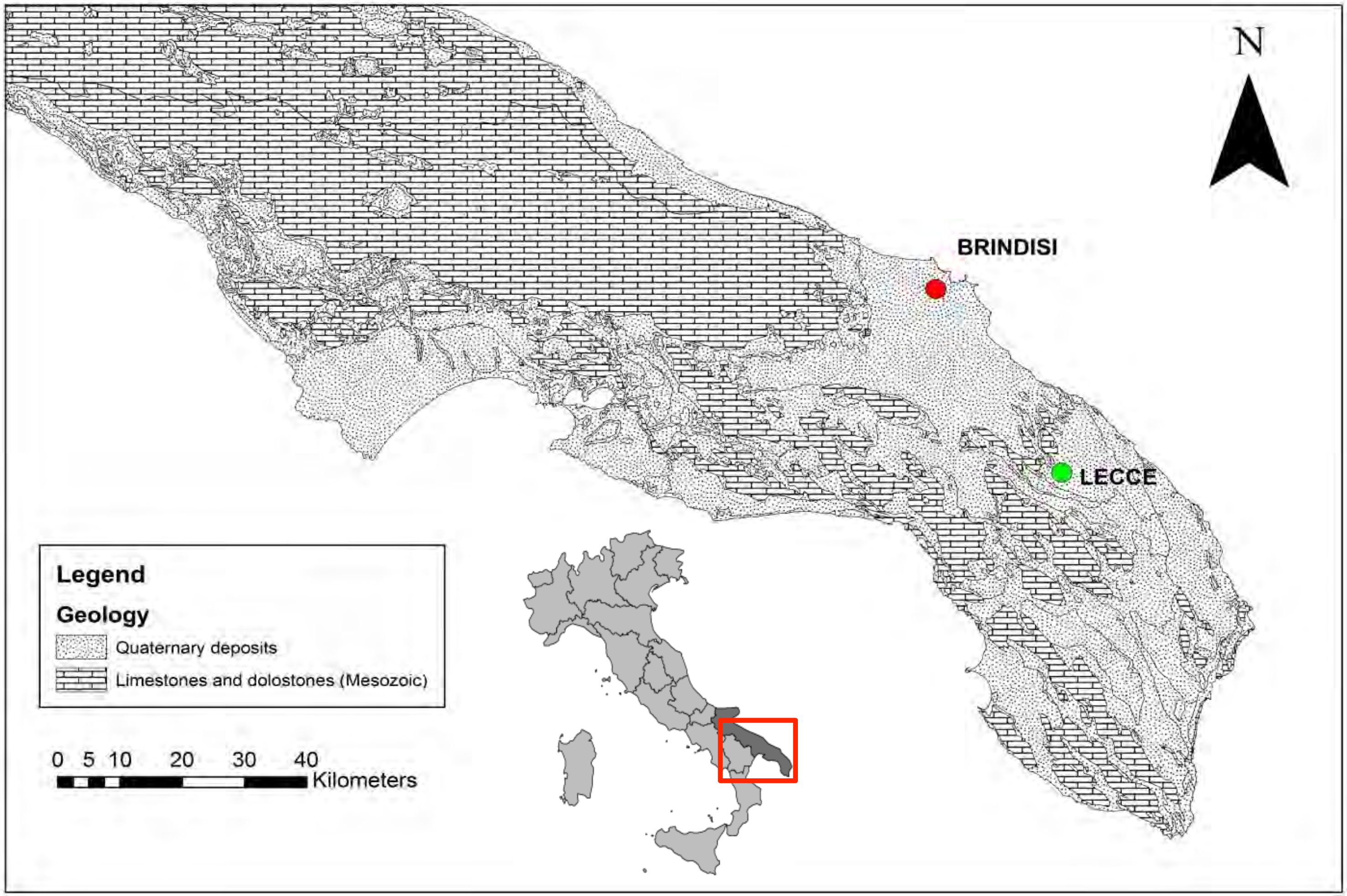
$$\mathbf{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot \left[\mathbf{X}_1 \cdot \ldots \cdot \mathbf{X}_k \cdot f\left(\mathbf{X}_1\right) \cdot \ldots \cdot f\left(\mathbf{X}_k\right)\right]$$

$$\mathbf{Y} = f\left(a_0 + \sum_{j=1}^{m} \left[a_j \cdot \mathbf{X}_1 \cdot \ldots \cdot \mathbf{X}_k\right]\right)$$



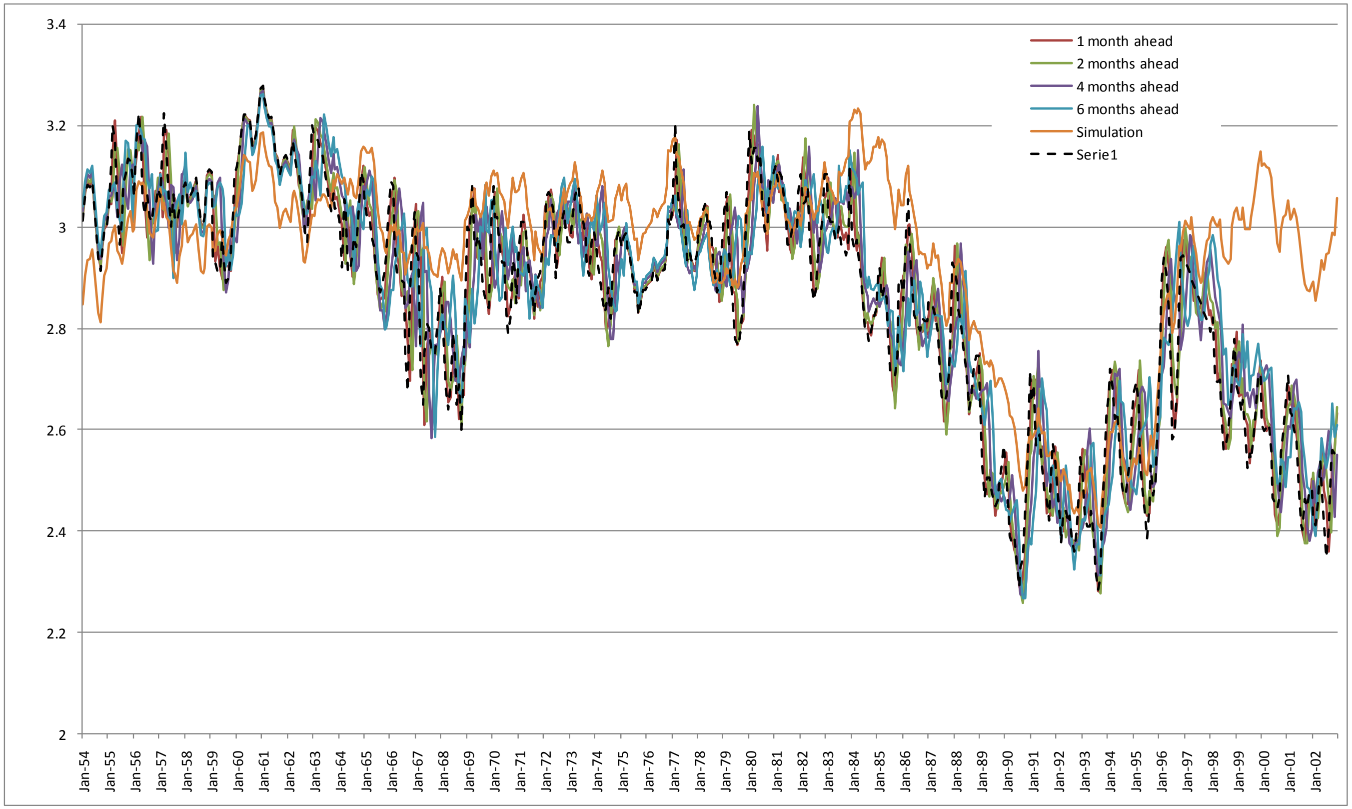
Some cases study 1/2

January 1953 – December 2002



Deep karst aquifer of Lecce

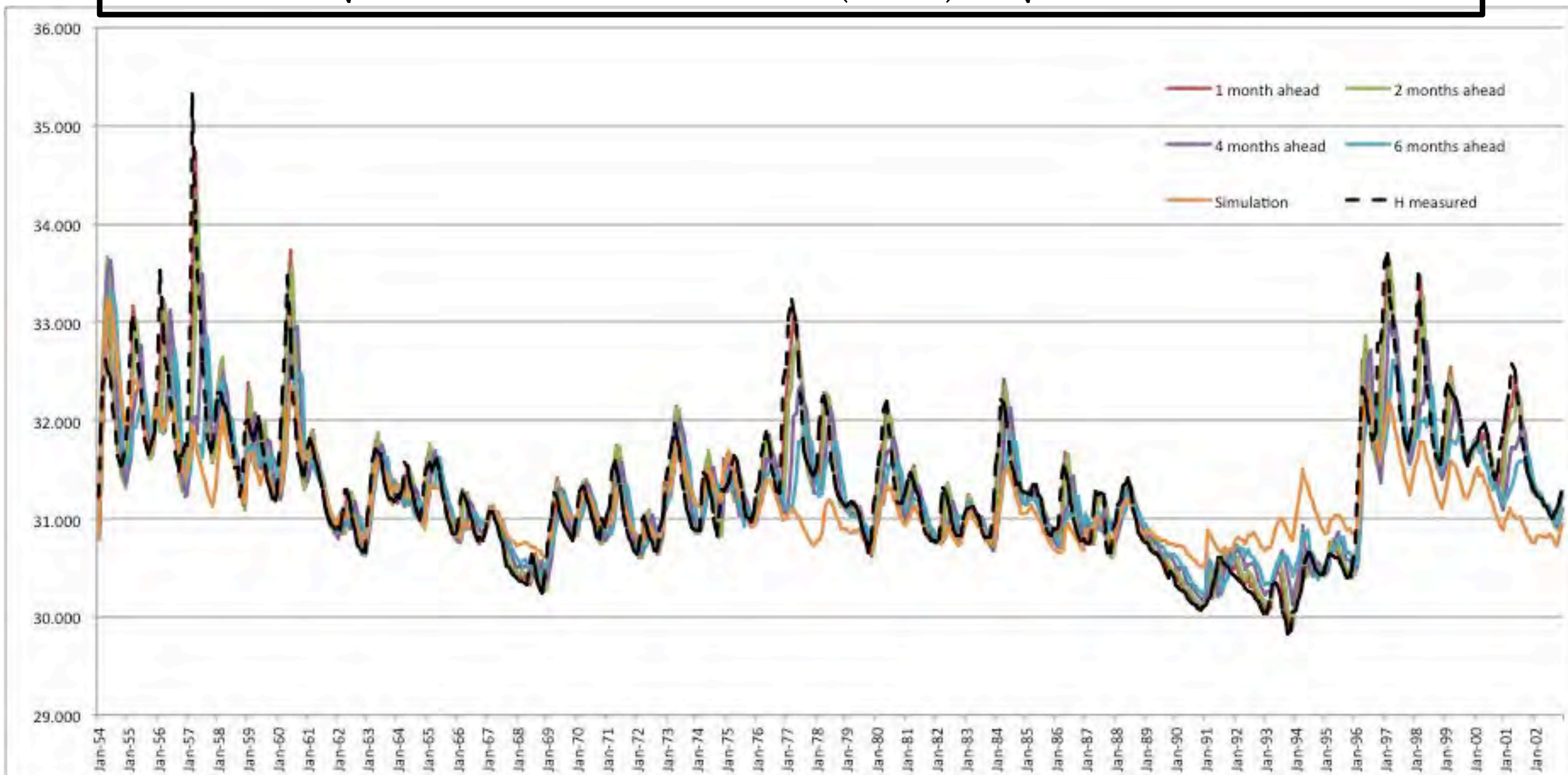
$$H_t = 0.0080995 \cdot P_t^{0.5} + 0.9819 \cdot H_{t-1}$$



		CoD 1-month	CoD 2-months	CoD 4-months	CoD 6-months	CoD simulation
Lecce	Training (53-77)	0.9495	0.8871	0.8025	0.7677	0.6075
	Test (78-2002)	0.9448	0.8677	0.7334	0.6753	0.3877

Shallow porous aquifer of Brindisi

$$H_t = 10.1 \cdot \sqrt{H_{t-1}} + 5.94 \cdot 10^{-6} \cdot P_{t-2} \cdot \left(H_{t-2}\right)^2 \cdot \sqrt{H_{t-1} \cdot P_{t-3} \cdot P_{t-4}} - 25.3$$

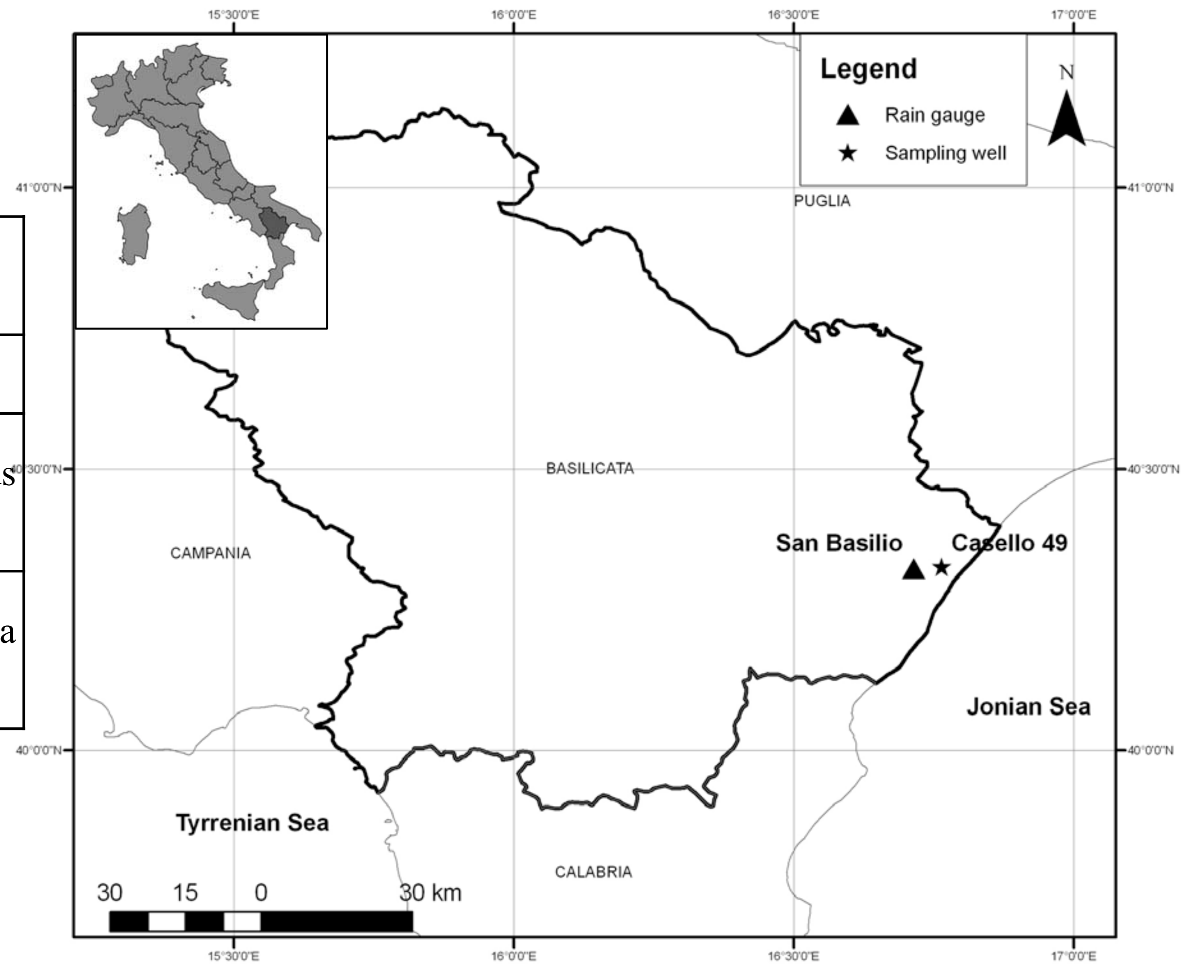


		CoD 1-month	CoD 2-months	CoD 4-months	CoD 6-months	CoD simulation
Brindisi	Training (53-77)	0.9493	0.8849	0.8043	0.7712	0.6055
	Test (78-2002)	0.9524	0.8882	0.7647	0.7073	0.4768

Some cases study 2/2

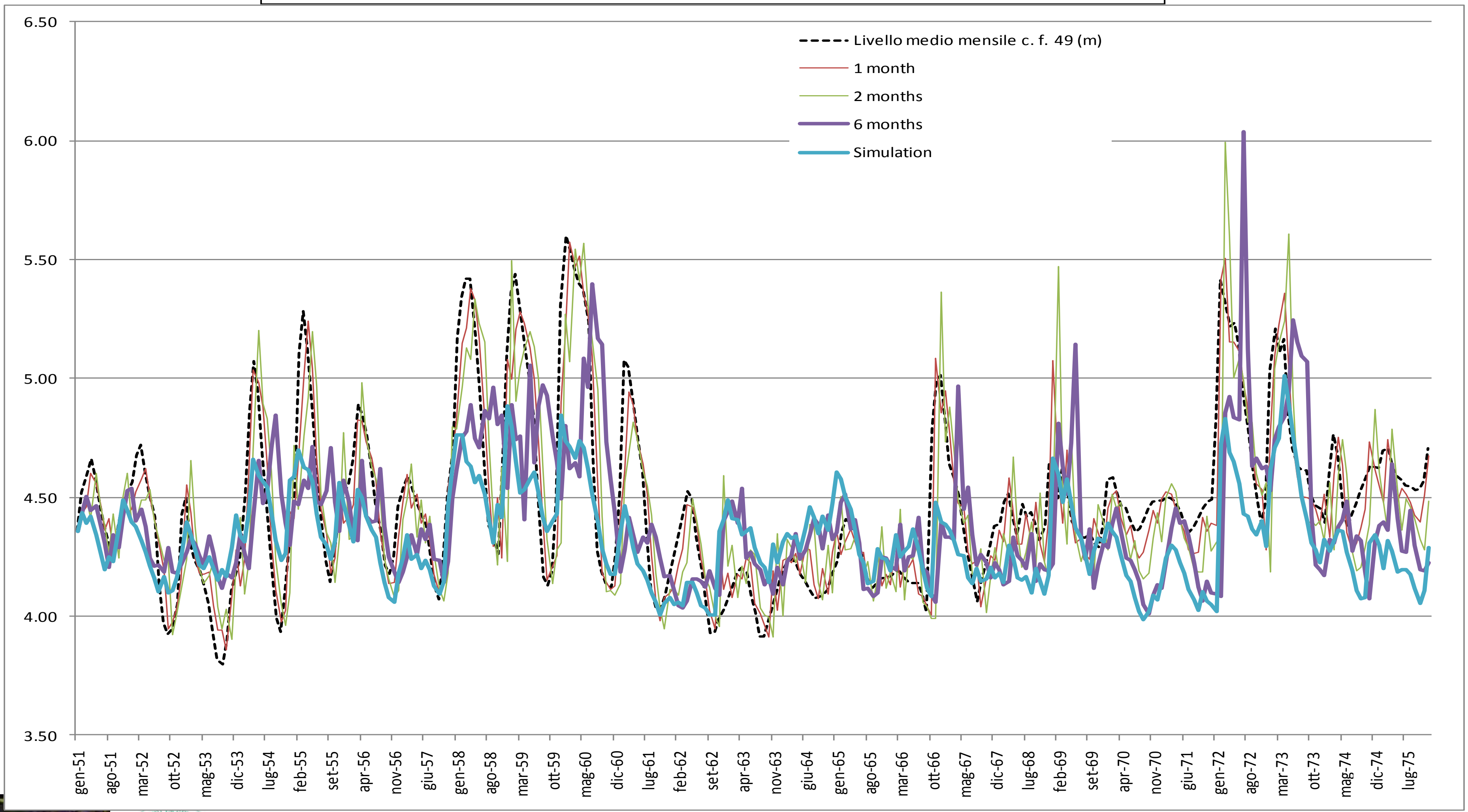
January 1951 – December 1975

Thickness of strata		
2.00		Ochraceous silty andy clay
7.50		Ochraceous and grey sands and coarse sands
10.00		Grey sands and silts in strata of small thickness



Shallow porous aquifer of Metaponto

$$H_t = 0.0021289 \cdot P_{t-1} + 0.096254 \cdot H_{t-1}^2 + 2.4205$$



Conclusions

- The data-mining approach is quite effective at modelling those scenarios where time-series of monitoring data are available
- EPR-MOGA returned simple and physically sound models for the investigated aquifers
- EPR-MOGA-returned models can be used for the management and planning of groundwater exploitation
- EPR-MOGA-returned models allow for simulation of scenarios of rainfall shortage in terms of impact on groundwater

Thank you for the attention!

Do you want to try?

a.doglioni@poliba.it

http://www.hydroinformatics.it