Fourth European Conference on Ecological Modelling
Fourth International Workshop on Environmental Applications of Machine Learning
ECEM/EAML 2004
Bled, Slovenia, September 27 - October 1, 2004

Fourth International Workshop on Environmental Applications of Machine Learning
27-29 September 2004
Bled, Slovenia

A new Multi-Objective strategy to support model selection for environmental modelling

O. Giustolisi (*), A. Doglioni (*), D.A. Savic (**) and F. di Pierro (**)

(*) *Civil and Environmental Engineering Department, Engineering Faculty of Taranto, Technical University of Bari, Italy.*

(**) *Centre for Water Systems, Department of Engineering, School of Engineering, Computer Science and Mathematics, University of Exeter, United Kingdom.*

1. Outline

PURPOSES

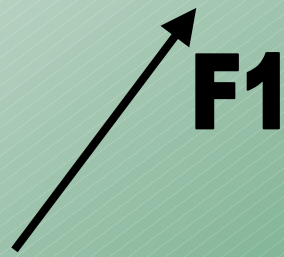
- **Data augmentation**
- **On-line forecasting**
- **Off-line forecasting**
- **Scientific knowledge discovery**

EPR

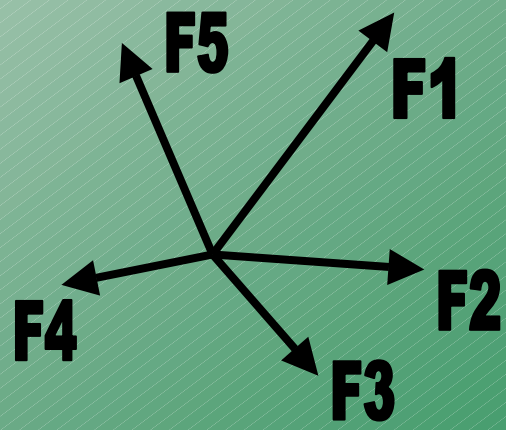
Evolutionary Computing

2. Multidimensional Modelling

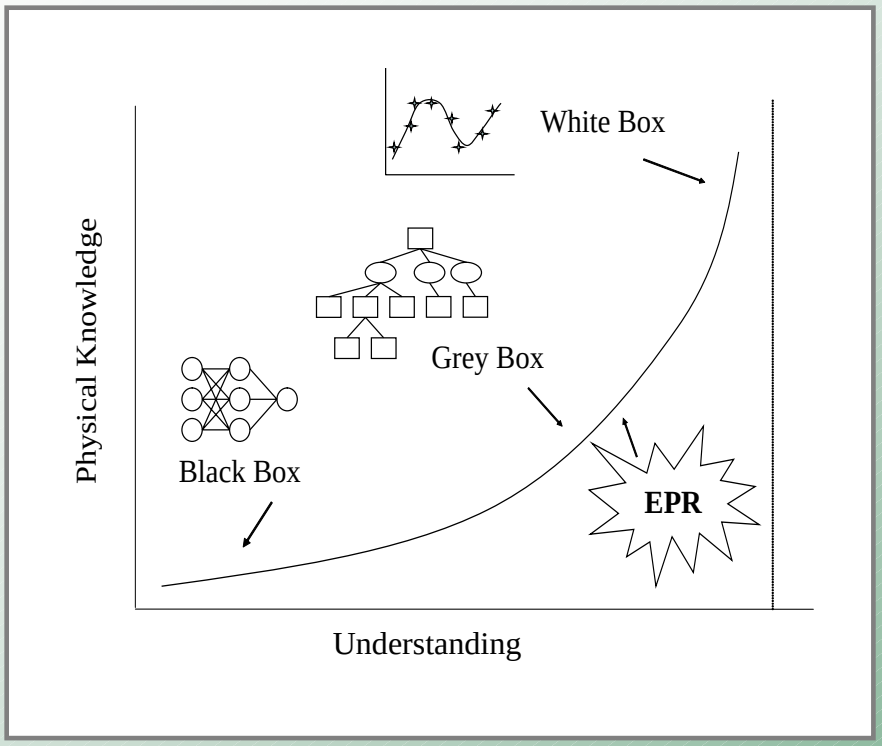
- Old strategies:
  - Single-Objective approach.
    - One objective function is optimized.
    - One-dimensional scenario.



- Novel strategies:
  - Multi-Objective approach.
    - Multipurpose modelling.
    - Multidimensional scenario: fitness vs. complexity.



3.1 Evolutionary Polynomial Regression



$$y = \sum_{j=1}^{m} F\left(\mathbf{X}, f\left(\mathbf{X}\right), a_j\right) + a_0$$

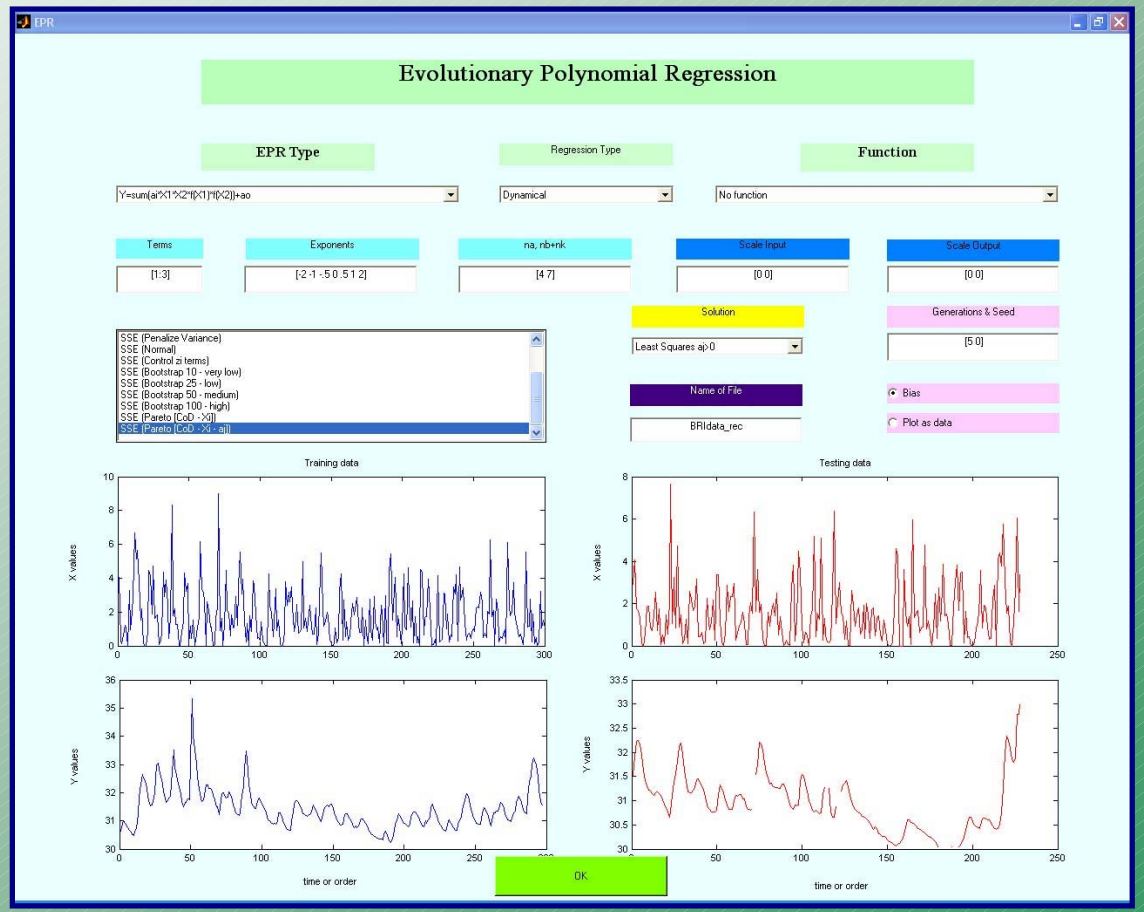
Hybrid Paradigm

GA + LS

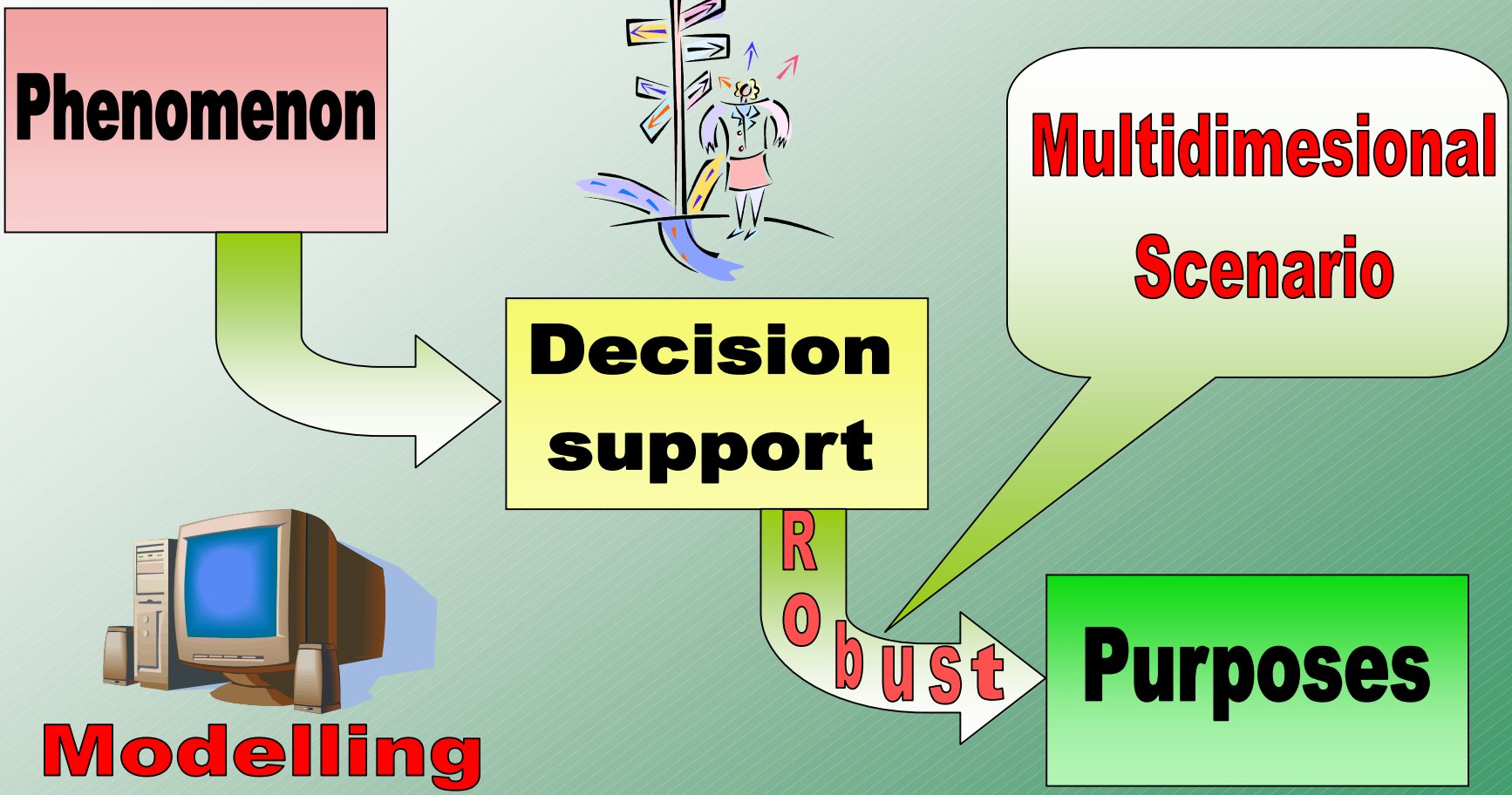
$$W_t = \left| C_{adim\_EPR} = \sqrt{\frac{8}{f}} = 0.4746\frac{ds}{hs}\cdot\ln\frac{2R}{ds} + \left(29.6138\ln\frac{hs}{ds} + 3.9319\frac{R\cdot S}{ds\cdot\frac{\mathbf{R}_w}{10^5}} + 50.55 \right) \right| \cdot 59914$$

3.2 MO-EPR: features

- Global method
- Parsimonious
- Non – overfitting
- Multiple models
- Interactive
- Multi-Objective
- LS $a_j > 0$

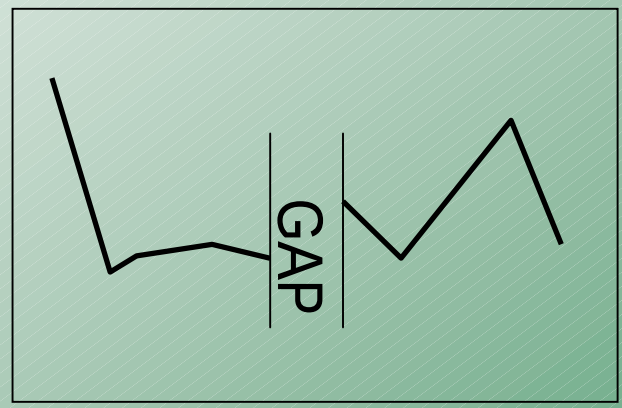


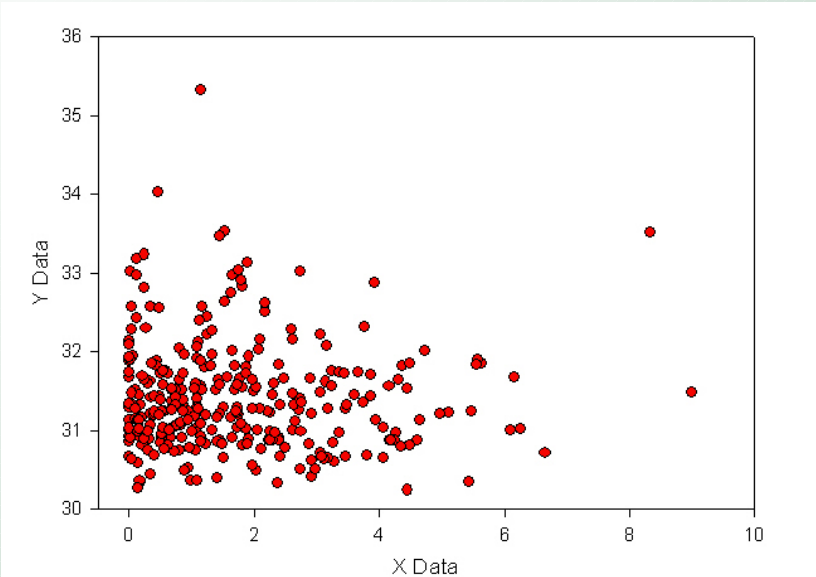
4. Environmental phenomena



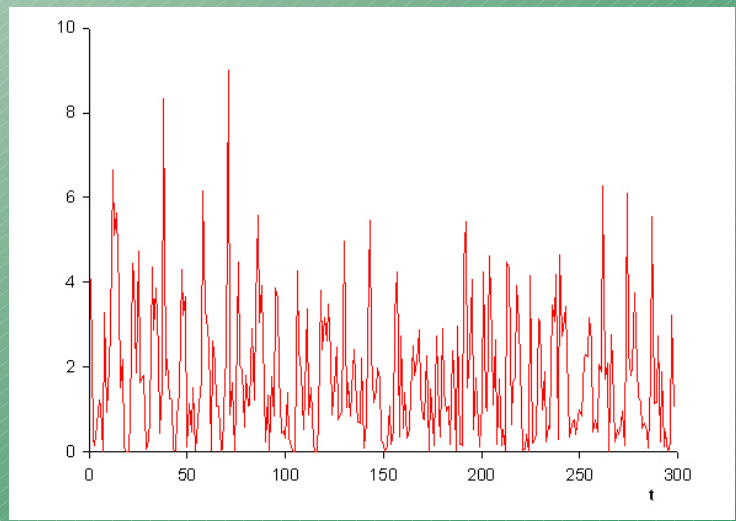
4.2 Environmental phenomena

Missing data



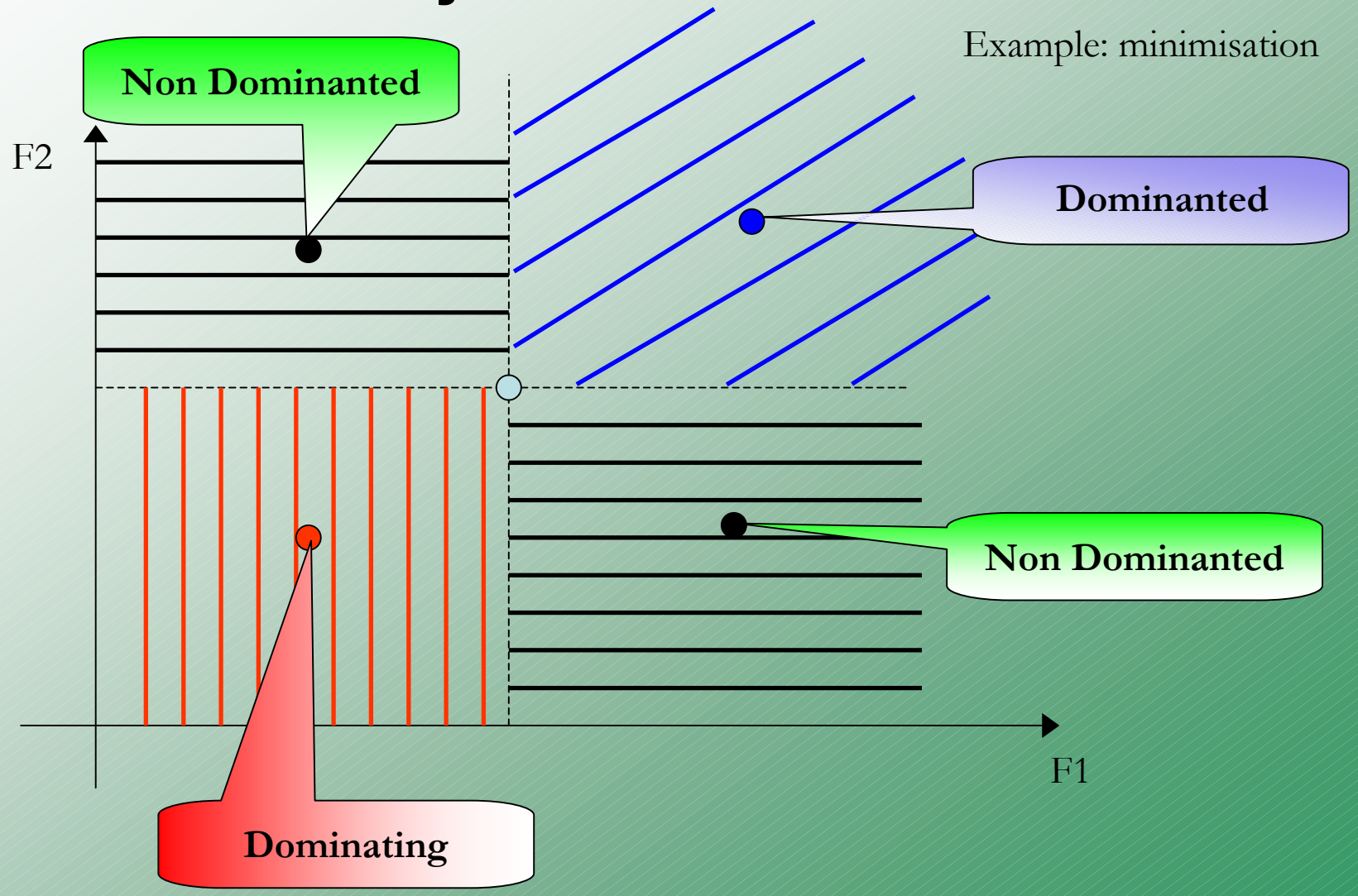


Non-linear dynamics



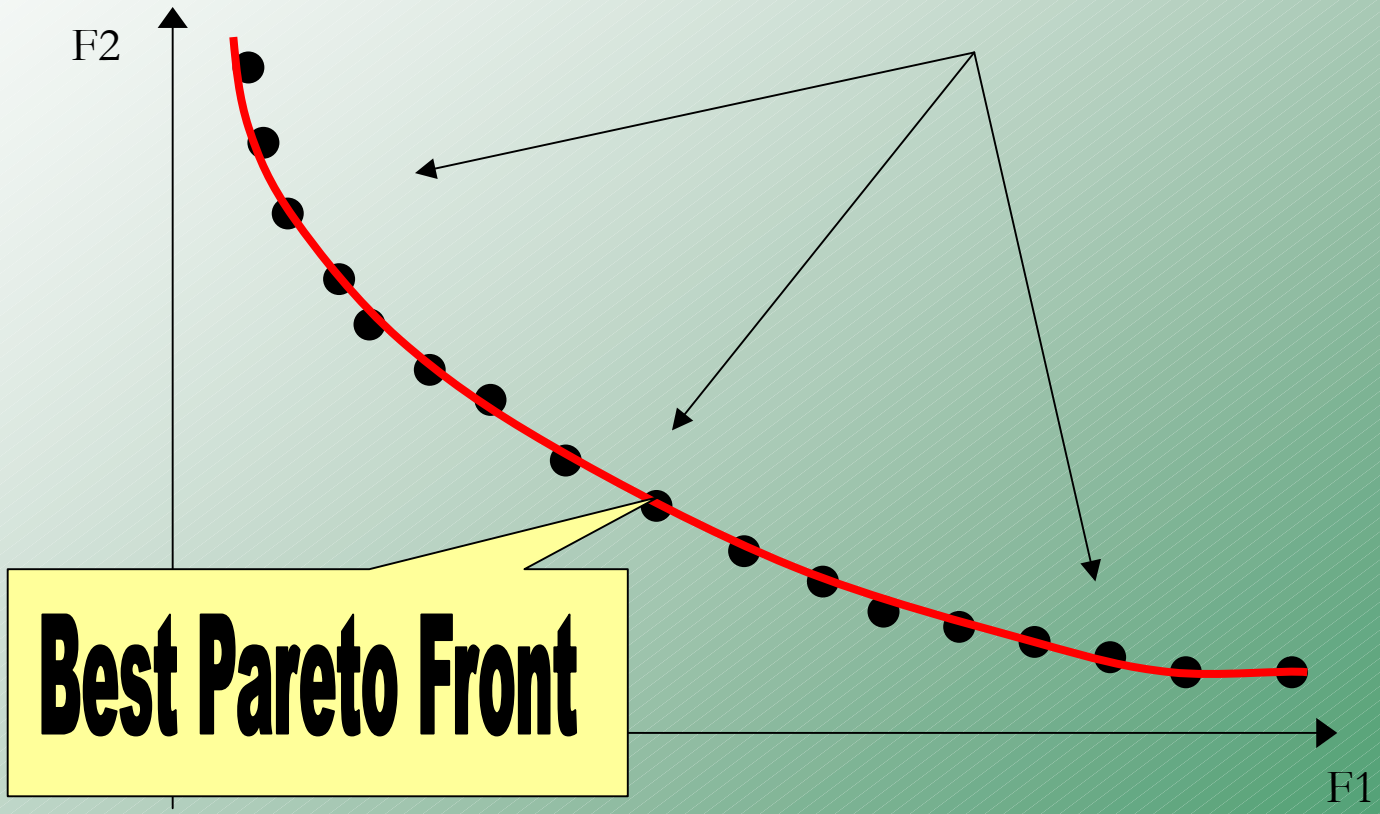
Non Gaussian noise

5.1 Multi-Objective search

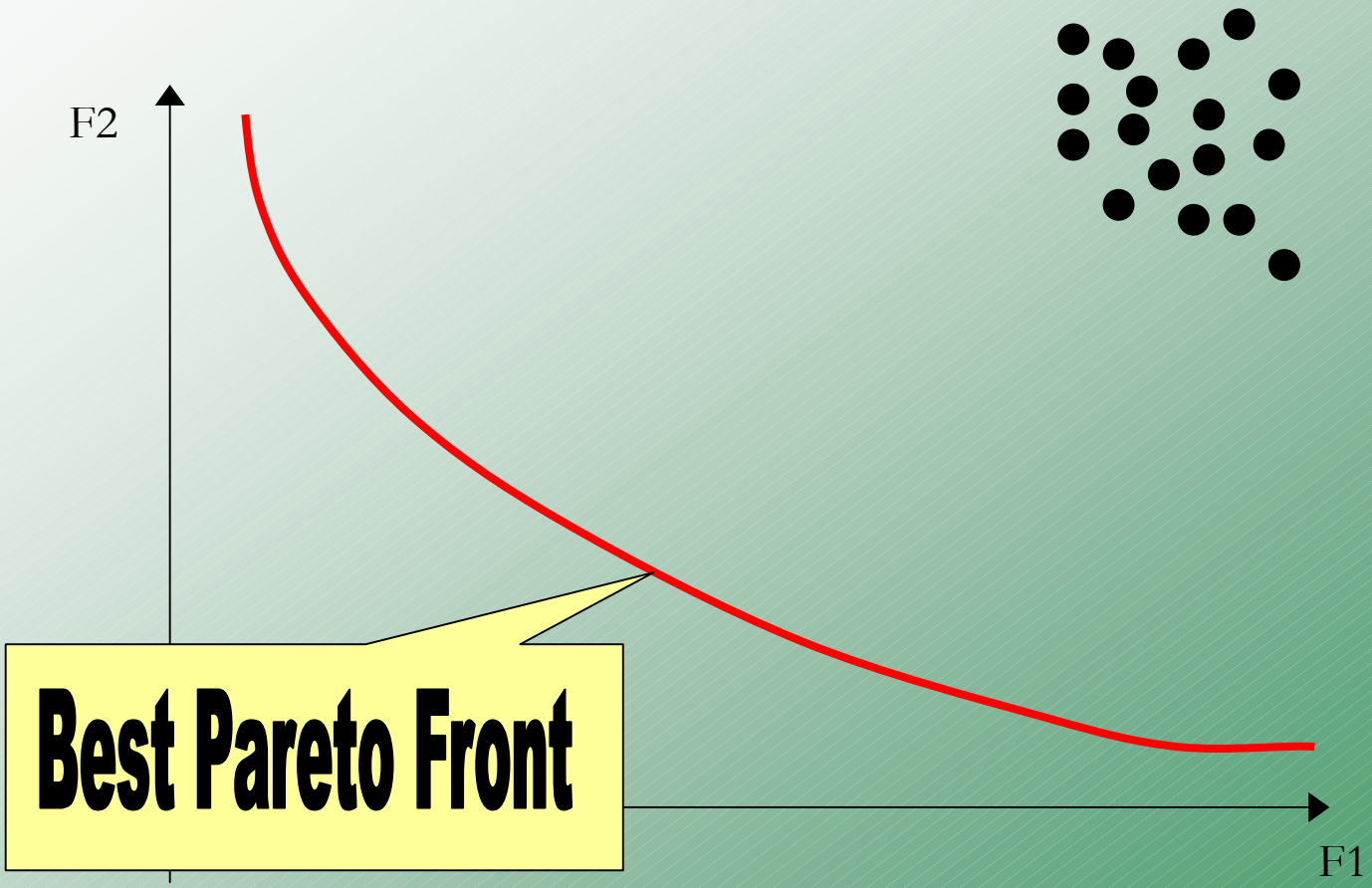


5.2 Multi-Objective search

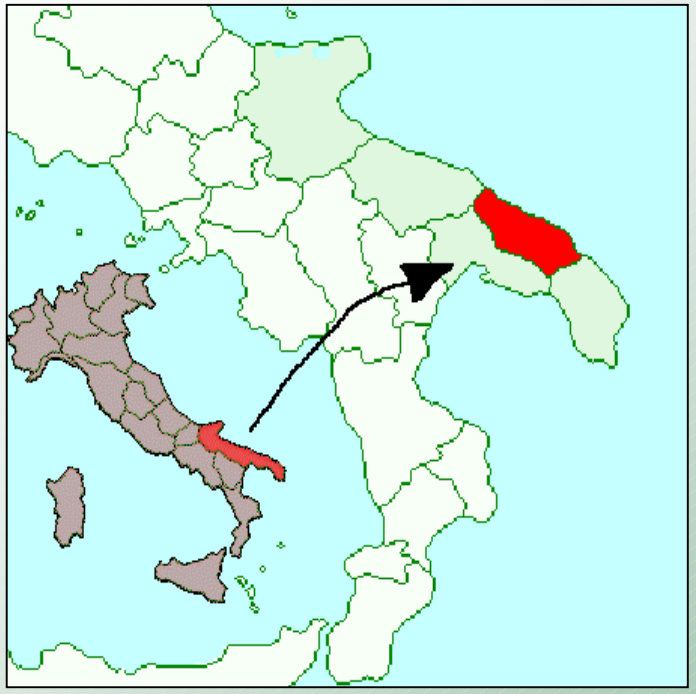
5.2 Multi-Objective search

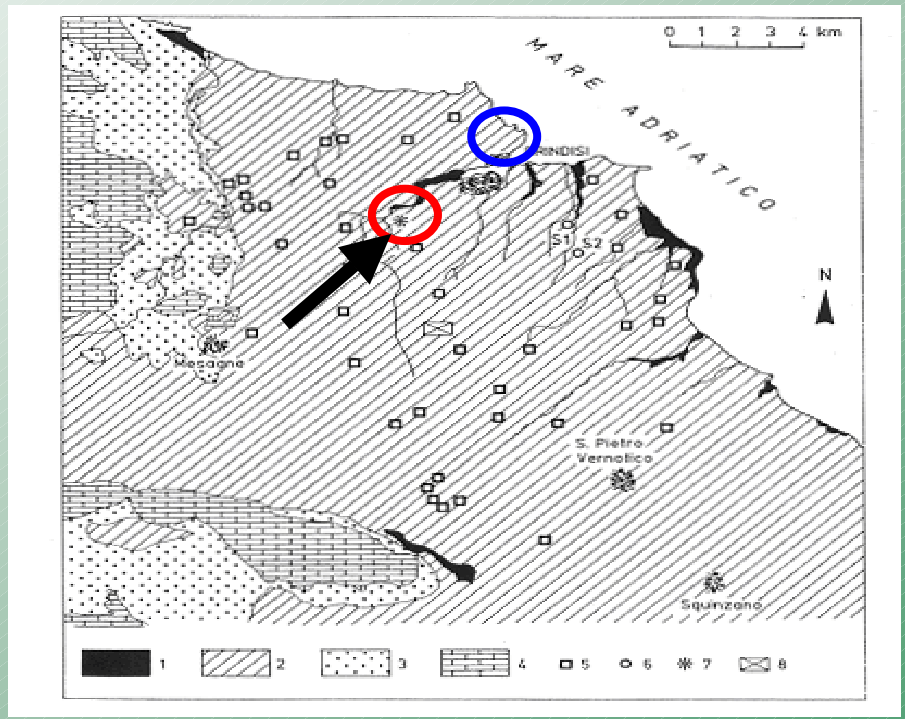


5.2 Multi-Objective search



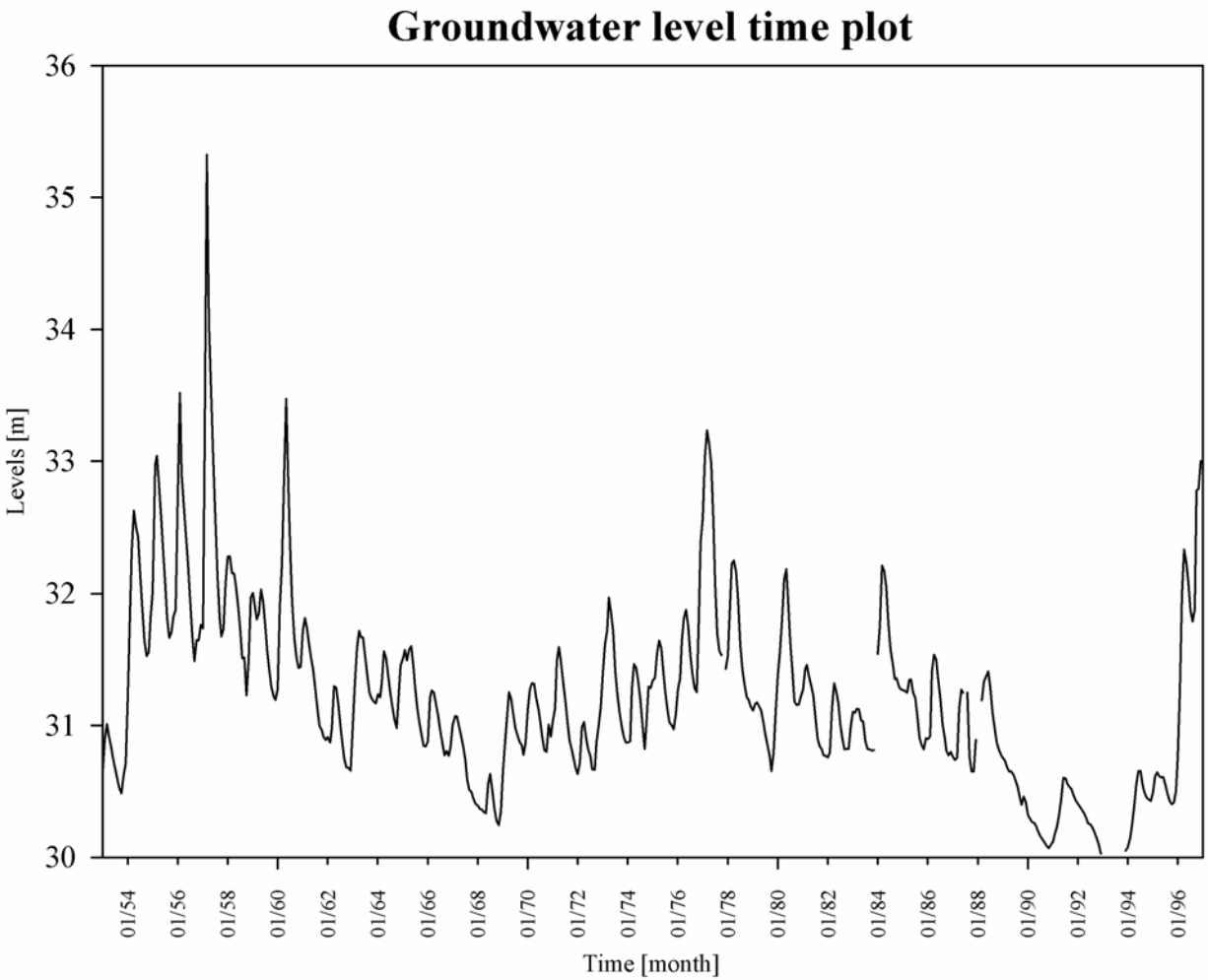
6.1 Case study





- Rainfall vs. Groundwater levels
- Brindisi - ITALY
- Monthly sampled data

6.2 GW time series

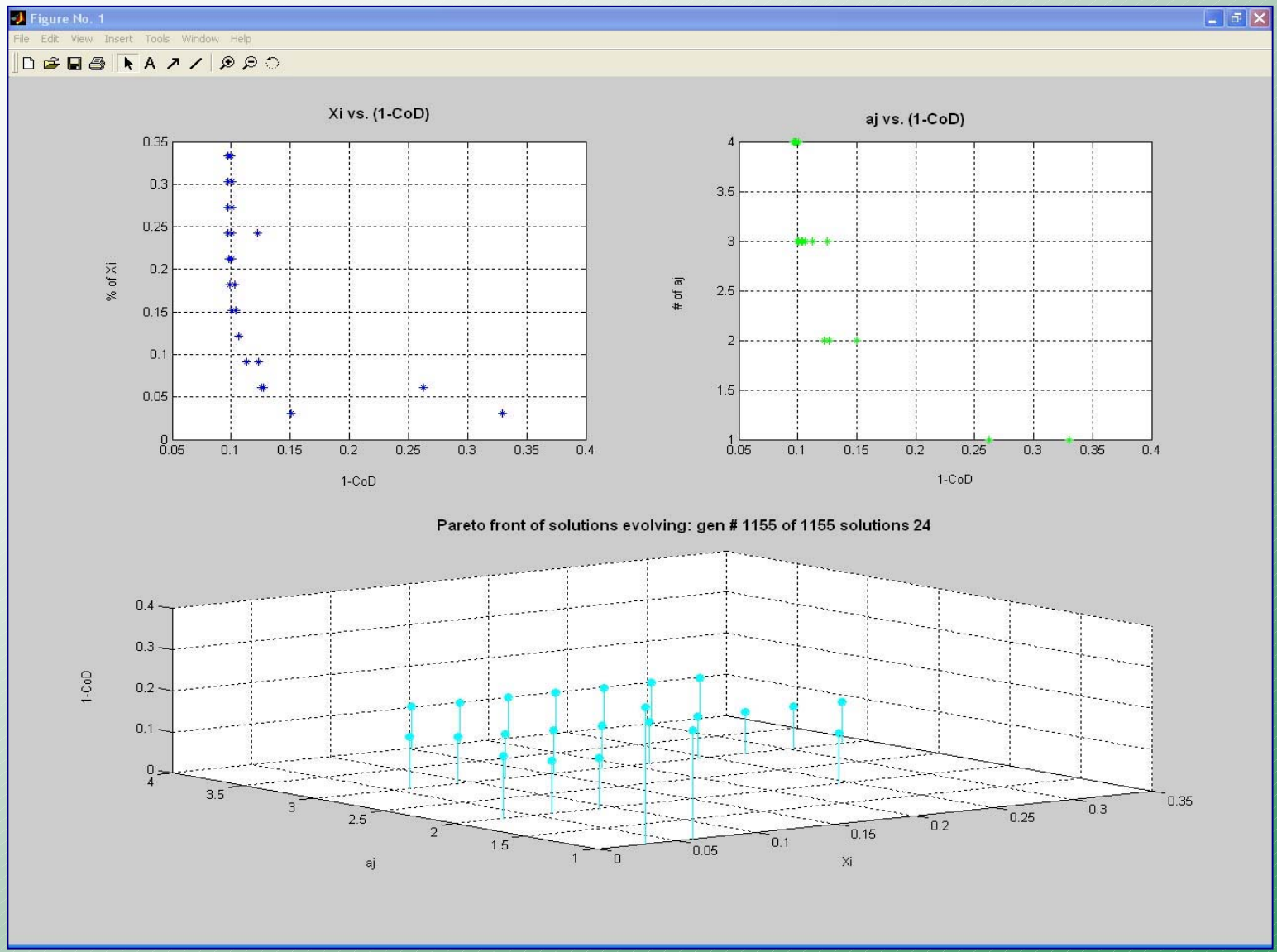


6.3 Case study: assumptions

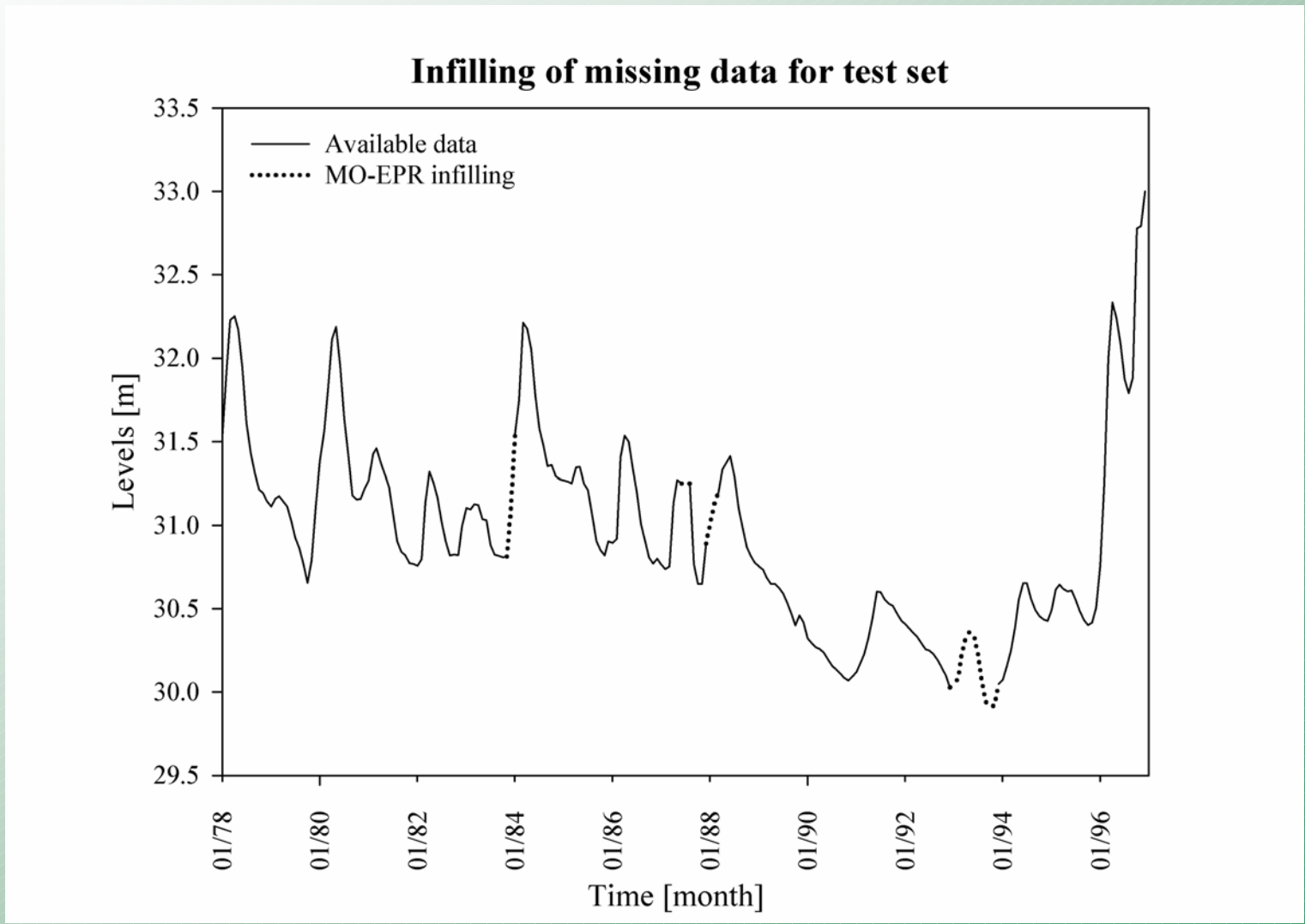
- Polynomially structured
- $m_{max} = 4$
- Training set
- 3 objective functions:
  - Fitness
  - # of monomial BB (complexity)
  - # of time each input is involved (complexity)
- Model evaluation on unseen data

$$\hat{y} = \sum_{j=1}^{m} F\left(\mathbf{X}, f\left(\mathbf{X}\right), a_j\right) + a_0$$

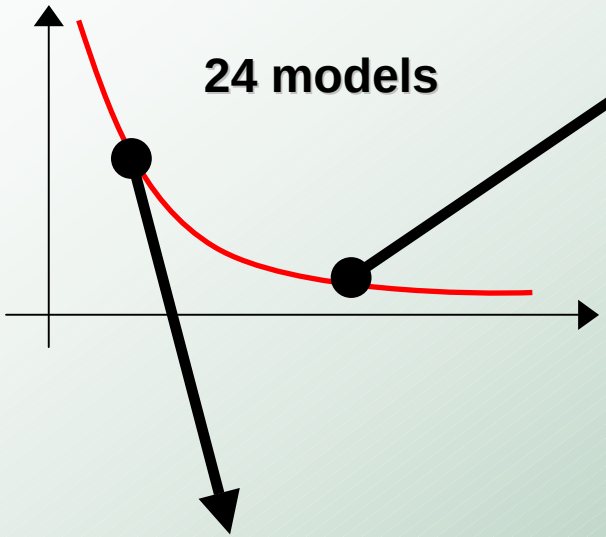
6.4 Pareto Front of solutions



6.5 Purpose: infilling of data



6.6 MO-EPR results – Test set



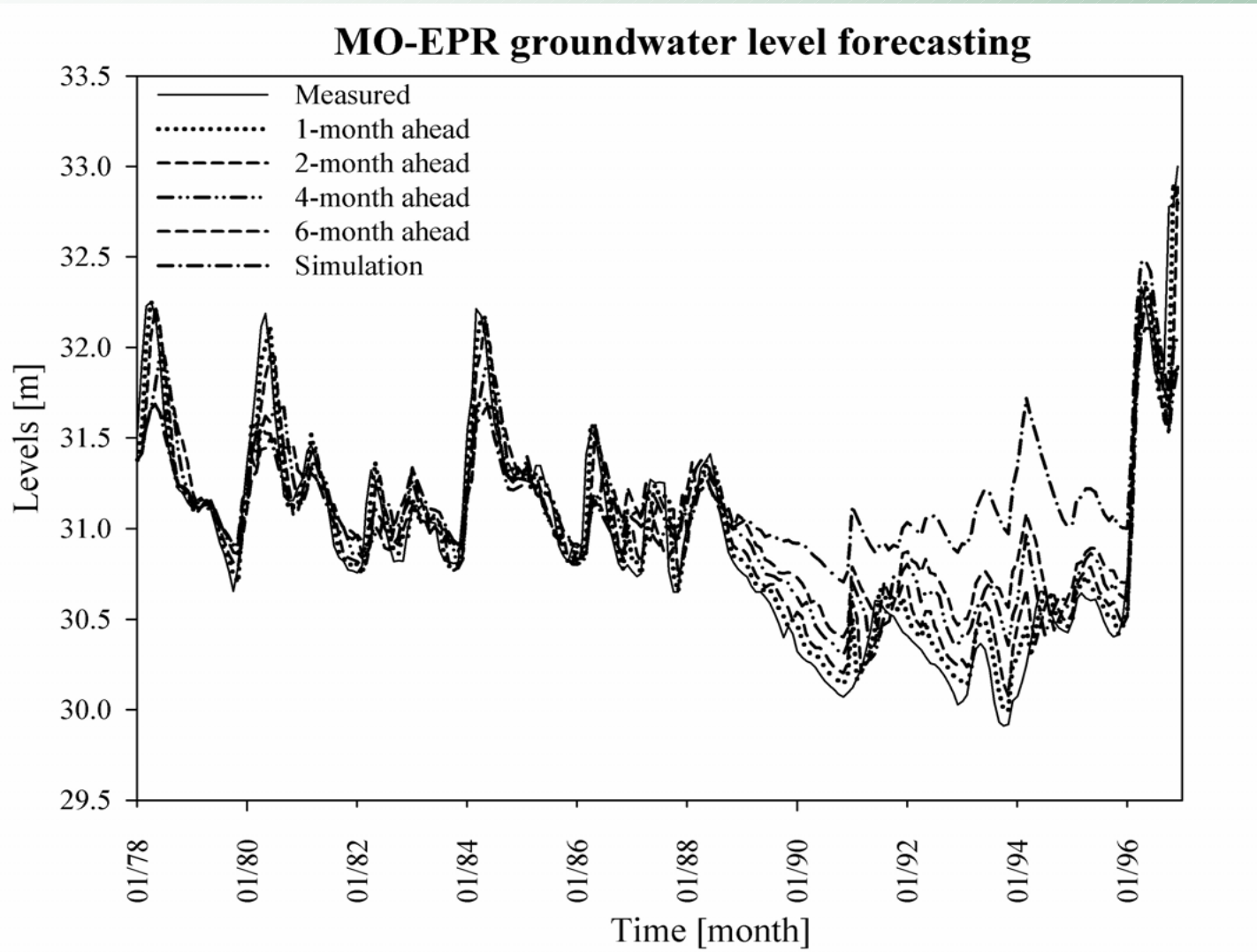
$$H_t = 0.92794 \cdot H_{t-1} + 0.070023 \cdot P_{t-1} + 2.1437$$

	CoD
2 months	0.878
4 months	0.726
6 months	0.646

$$H_t = 0.10417 \cdot H_{t-1}^{2} \cdot H_{t-2}^{-0.5} + 0.024556 \cdot P_{t-1} \cdot P_{t-2}^{0.5} \cdot P_{t-3}^{0.5} + 0.0054314 \cdot P_t^{2} + 12.9706$$

	CoD
1 months	0.956
Off-line prediction	0.242

6.7 MO-EPR results - Test set



7. Conclusions

- Multidimensional modelling:
  - Not only statistically based choice
  - "Pareto" surface based choice
- Wide range of solution/purposes:
  - Infilling
  - Prediction
  - Etc.
- Proposal: MO-EPR (Not unique!)

7. Conclusions

http://www.poliba.it/taranto/software/hydroinformatics/index.htm

Thank you!