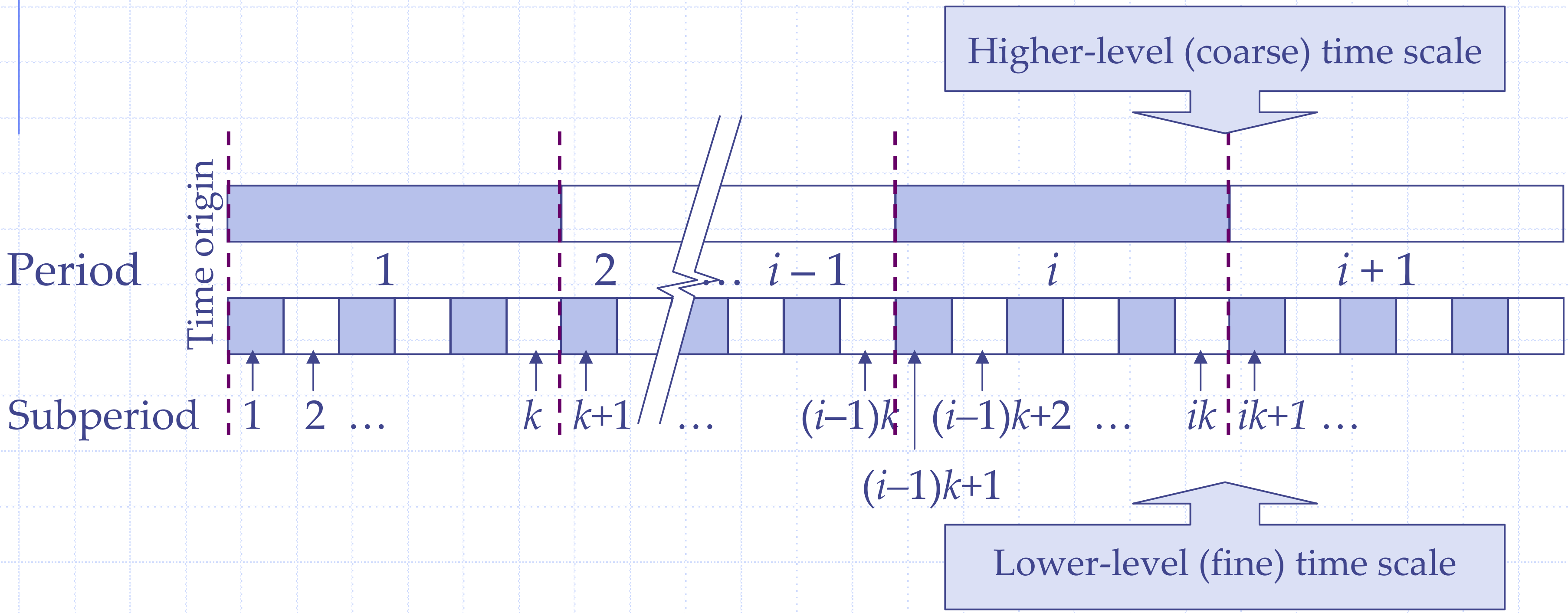
Workshop on Statistical and Mathematical Methods for Hydrological Analysis
Rome, May 2003
Universitá degli Studi di Roma "La Sapienza"
CNR - IRPI Istituto di Ricerca per la Protezione Idrogeologica GNDCI
Gruppo Nazionale per la Difesa dalle Catastrofi Idrogeologiche
Universitá degli Studi di Napoli Federico II

Rainfall Disaggregation Methods: Theory and Applications

Demetris Koutsoyiannis
Department of Water Resources, School of Civil Engineering,
National Technical University, Athens, Greece

Definition of disaggregation

1. Generation of synthetic data (typically using stochastic methods)
2. Involvement of two time scales (higher- and lower-level)

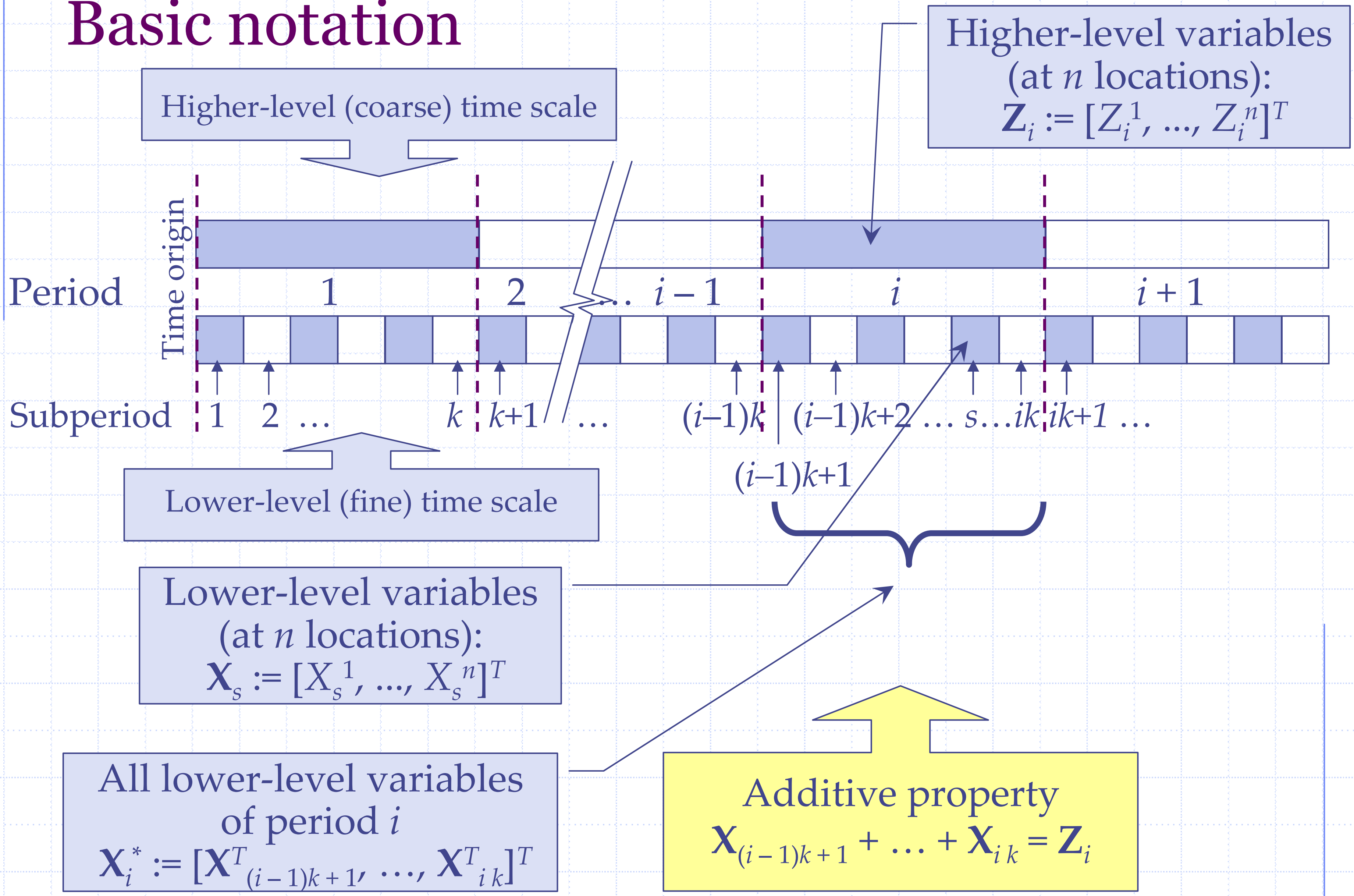


3. Use of different models for the two time scales (with emphasis on the different characteristics appearing at each scale)
4. Requirement that the lower-level synthetic series is consistent with the higher-level one

The utility of rainfall disaggregation

- **Enhancement of data records:** Disaggregation of widely available daily rainfall measurements into hourly records (often unavailable and frequently required by hydrological models)
- **Flood studies:** Synthesis of one or more detailed storm hyetograph (more severe than the observed ones) with known total characteristics (duration, depth)
- **Simulation studies:** Study of a hydrological system using multiple (rather than the single observed) sequences of rainfall series
- **Climate change studies:** Use of output from General Circulation Models (forecasts for different climate change scenarios), generally provided at a coarse time-scale (e.g. monthly) to hydrological applications that require a finer time scale

Basic notation



General purpose stochastic disaggregation: The Schaake-Valencia model (1972)

- The lower-level time series are generated by a “hybrid” model involving both time scales simultaneously
- The model has a simple mathematical expression

$$\mathbf{X}_i^* = \mathbf{a}\,\mathbf{Z}_i + \mathbf{b}\,\mathbf{V}_i$$

  where

  $\mathbf{V}_i$: vector of kn independent identically distributed random variates

  $\mathbf{a}$: matrix of parameters with size $kn \times n$

  $\mathbf{b}$: matrix of parameters with size $kn \times kn$

- The parameters depend on variance and covariance properties among higher- and lower-level variables, which are estimated from historical records
- The additive property is automatically preserved if parameters are estimated from historical records
- Due to the Central Limit Theorem $\mathbf{X}_i$ tend to have normal distributions

General purpose stochastic disaggregation: Weaknesses and remediation

Weakness	**Remediation**	**Comments**
Independence of consecutive lower-layer variables belonging to consecutive periods	Different model structures, simultaneously involving lower-layer variables of the earlier period	1. Use of even larger parameter sets 2. Only partial remediation
Inability to perform with non-Gaussian distribution	Use of nonlinear transformations of variables	Violation of the additive property
	Attempt to preserve the skewness	Infeasibility to preserve large skewness
Excessive number of parameters	Different model types: ▪ Staged disaggregation models ▪ Condensed disaggregation models ▪ Dynamic disaggregation models	Better performance compared to the original model types

Coupling stochastic models of different time scales: A more recent disaggregation approach

- Do not combine both time scales in a single "hybrid" model
- Instead, use totally independent models for each time scale
- Run the lower-level model independently of the higher-level one
- For each period do many repetitions and choose the generated lower-level series that is in closer agreement with the higher-level one
- Apply an appropriate transformation (adjustment) to the finally chosen lower-level series to make it fully consistent with the higher-level one

The single variate coupling form: The notion of accurate adjusting procedures

- In each period, use the lower-level model to generate a sequence of $\tilde{X}_s$ that add up to the quantity $\tilde{Z} := \sum_s \tilde{X}_s$, which is different from the known Z
- Adjust $\tilde{X}_s$ to derive the sequence of X_s that add up to Z
- The adjusting procedures, i.e. the transformations $X_s = f(\tilde{X}_s, \tilde{Z}, Z)$, should be such that the distribution function of X_s is identical to that of $\tilde{X}_s$

The two most useful adjusting procedures

1. Proportional adjusting procedure

$$X_s = \tilde{X}_s \, (Z / \tilde{Z})$$

- It preserves exactly the complete distribution functions if variables X_s are independent with two-parameter gamma distribution and common scale parameter
- It gives good approximations for gamma distributed X_s

2. Linear adjusting procedure

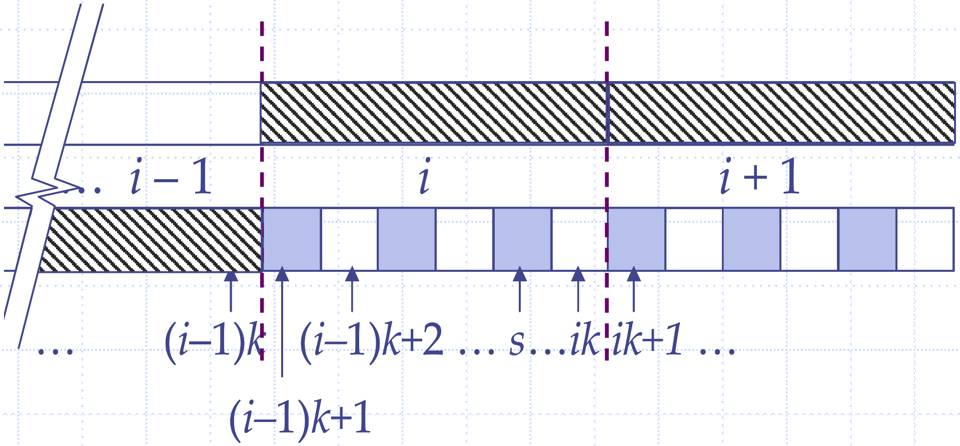
$$X_s = \tilde{X}_s + \lambda_s (Z - \tilde{Z})$$

where λ_s are unique coefficients depending of covariances of X_s and Z

- It preserves exactly the complete distribution functions if variables X_s are normally distributed
- It is accurate for the preservation of means, variances and covariances for any distribution of variables X_s

The general linear coupling transformation for the multivariate case

- It is a multivariate extension to many dimensions of the single-variate linear adjusting procedure
- It preserves exactly the complete distribution functions if variables X_s are normally distributed
- It preserves exactly means, variances and covariances for any distribution of variables X_s
- In addition, it enables linking with previous subperiods and next periods (with already generated amounts) so as to preserve correlations of lower-level variables belonging to consecutive periods



The general linear coupling transformation for the multivariate case

Mathematical formulation

$$\mathbf{X}_i^* = \tilde{\mathbf{X}}_i^* + \mathbf{h}\,(\mathbf{Z}_i^* - \tilde{\mathbf{Z}}_i^*)$$

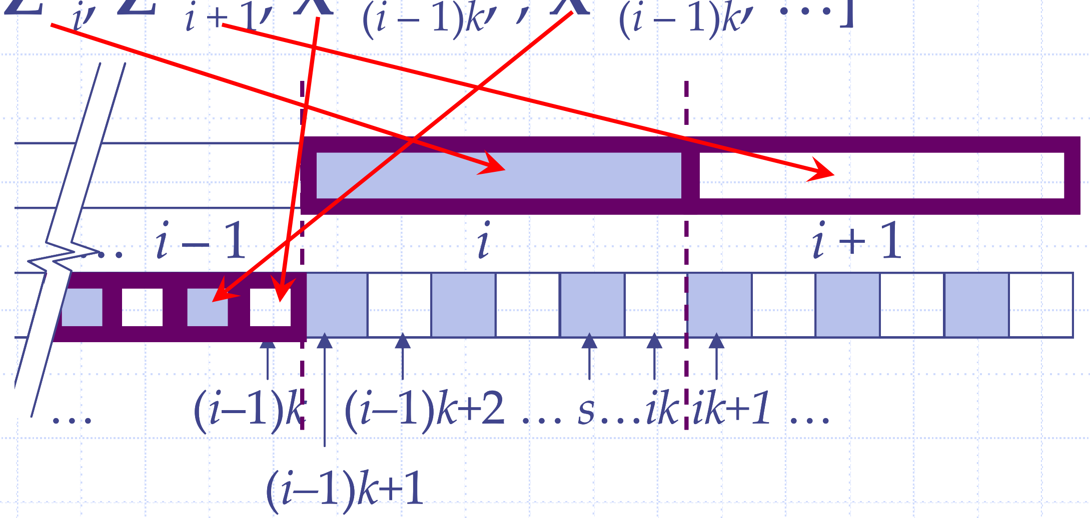
where

$$\mathbf{h} = \mathrm{Cov}[\mathbf{X}_i^*, \mathbf{Z}_i^*]\,\{\mathrm{Cov}[\mathbf{Z}_i^*, \mathbf{Z}_i^*]\}^{-1}$$

$$\mathbf{X}_i^* := [\mathbf{X}^T_{(i-1)k+1}, \ldots, \mathbf{X}^T_{ik}]^T,$$

$$\mathbf{Z}_i^* := [\mathbf{Z}^T_i, \mathbf{Z}^T_{i+1}, \mathbf{X}^T_{(i-1)k}, \mathbf{X}^T_{(i-1)k}, \ldots]^T$$

Important note:
h is determined from properties of the lower-level model only



Rainfall disaggregation - Peculiarities

- General purpose models have been used for rainfall disaggregation but for time scales not finer than monthly
- For finer time scales (e.g. daily, hourly, sub-hourly), which are of greater interest, the general purpose models were regarded as inappropriate, because of:
  - the intermittency of the rainfall process
  - the highly skewed, J-shaped distribution of rainfall depth
  - the negative values that linear models may produce

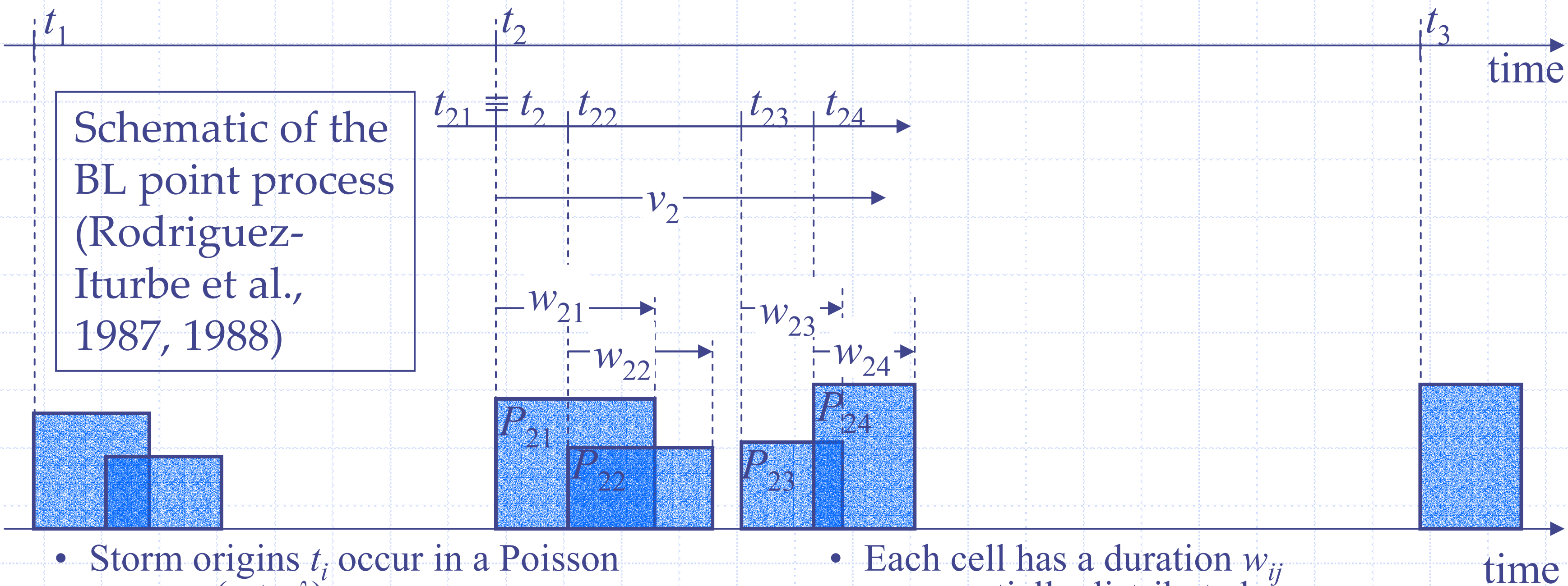
Special types of rainfall disaggregation models

- Urn models: filling of "boxes", representing small time intervals, with "pulses" of small rainfall depth increments
- Non-dimensionalised models: standardisation of the rainfall process either in terms of time or depth or both and use of certain assumptions for the standardised process (e.g. Markovian structure, gamma distribution)
- Models implementing non stochastic techniques such as
  - multifractal techniques
  - chaotic techniques
  - artificial neural networks
- All special type models are single variate
- Recently, a bi-variate model was developed and applied to the Tiber catchment (Kottegoda et al., 2003)

Implementation of general-purpose disaggregation models for rainfall disaggregation

- The disaggregation approach based on the coupling of models of different timescales can be directly implemented in rainfall disaggregation
- In single variate setting: A point process model, like the Bartlett-Lewis (BL) model, can be used as the lower-level model ⇨ **Hyetos**
- In multivariate setting there are two possibilities for the lower-level model
  - Use of a multivariate (space-time) extension of a point process model
  - Combination of a detailed single variate model and a simplified multivariate model ⇨ **MuDRain** The detailed single variate model can be replaced by observed time series if applicable

Hyetos: A single variate fine time scale rainfall disaggregation model based on the BL process



Schematic of the BL point process (Rodriguez-Iturbe et al., 1987, 1988)

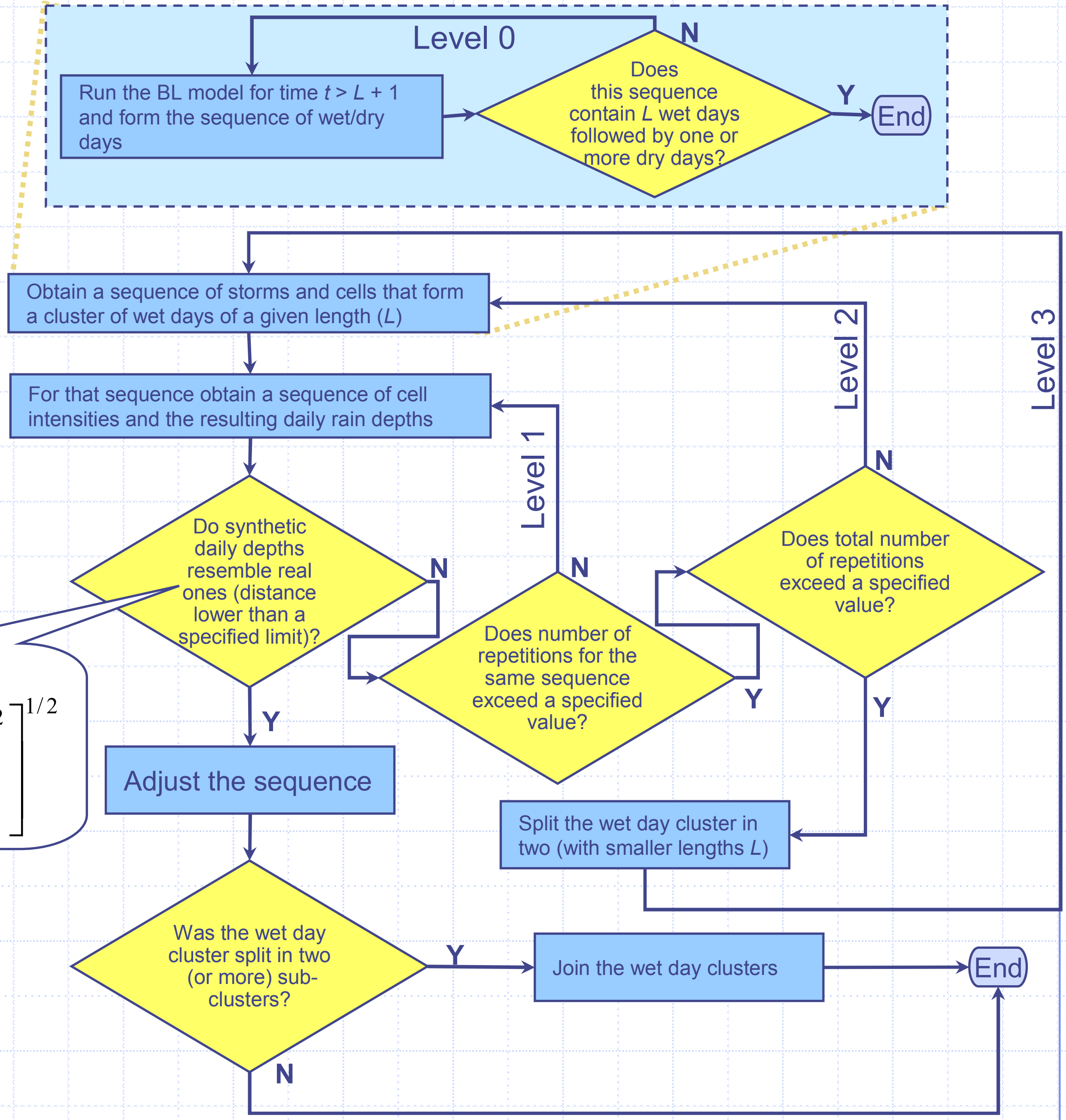
- Storm origins t_i occur in a Poisson process (rate λ)
- Cell origins t_{ij} arrive in a Poisson process (rate β)
- Cell arrivals terminate after a time v_i exponentially distributed (parameter γ)
- Each cell has a duration w_{ij} exponentially distributed (parameter η)
- Each cell has a uniform intensity P_{ij} with a specified distribution

Hyetos = BL + repetition + proportional adjusting procedure

Hyetos: Assumptions and procedures

- Different clusters of rain days (separated by at least one dry day) may be assumed independent
- This allows different treatment of each cluster of rain days, which reduces computational time rapidly as the BL model runs separately for each cluster
- Several runs are performed for each cluster, until the departure of daily sum from the given daily rainfall becomes lower than an acceptable limit
- In case of a very long cluster of wet days, it is practically impossible to generate a sequence of hourly depths with low departure of daily sum from the given daily rainfall; so the cluster is subdivided into sub-clusters, each treated independently of the others
- Further processing consists of application of the proportional adjusting procedure to achieve full consistency with the given sequence of daily depths.

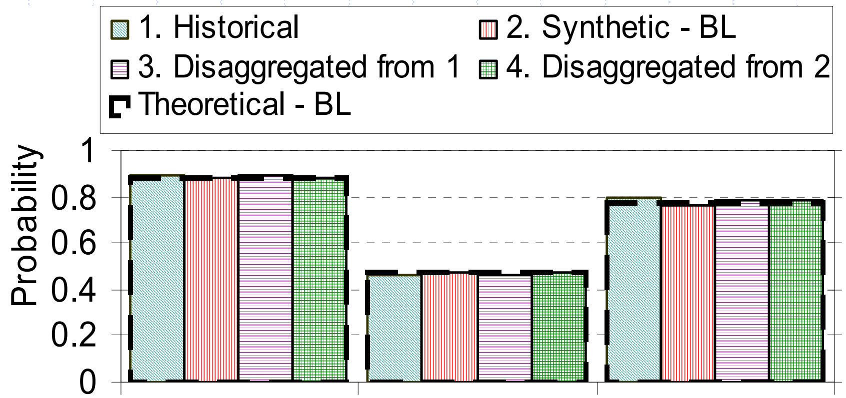
Hyetos: Repetition scheme



$$d = \left[\sum_{i=1}^{L} \ln \left(\frac{Z_i + c}{\tilde{Z}_i + c} \right)^2 \right]^{1/2}$$

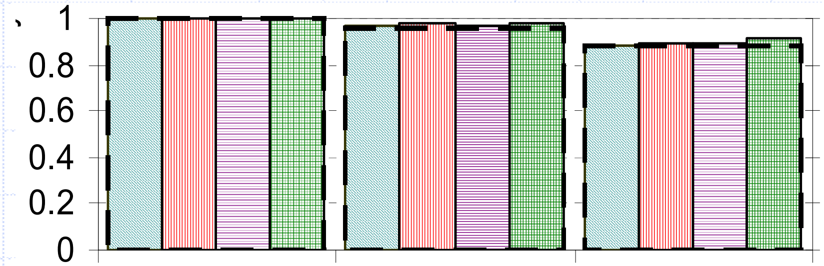
Hyetos: Case studies and model performance
1. Preservation of dry/wet probabilities

1. Heathrow Airport (England)
Wet throughout the year

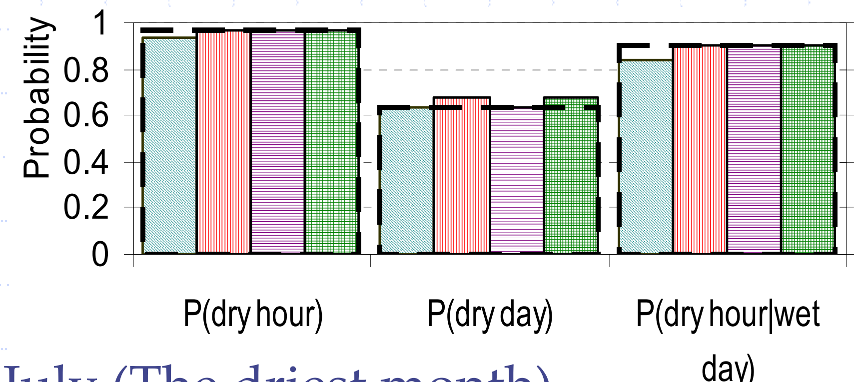


January (The wettest month)

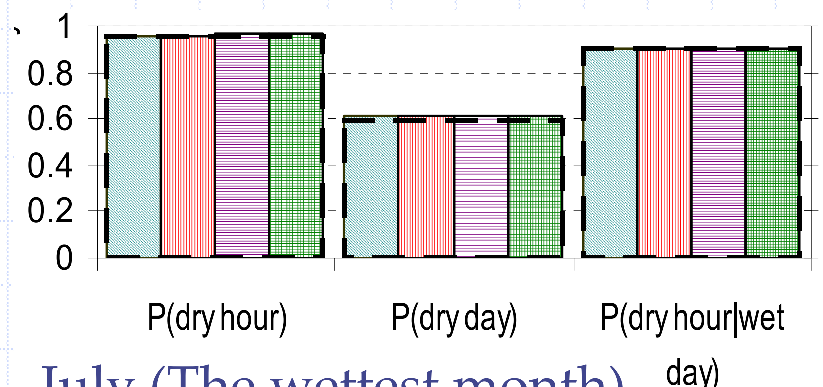
Walnut Gulch, Gauge 13 (USA)
Semiarid with a wet season



May (The driest month)



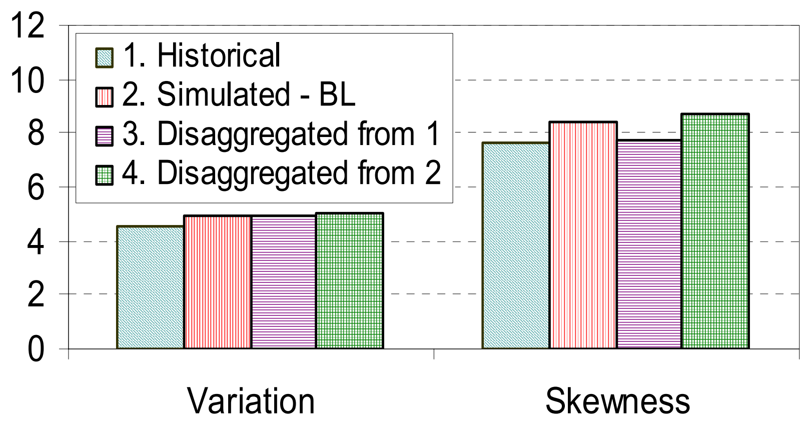
July (The driest month)



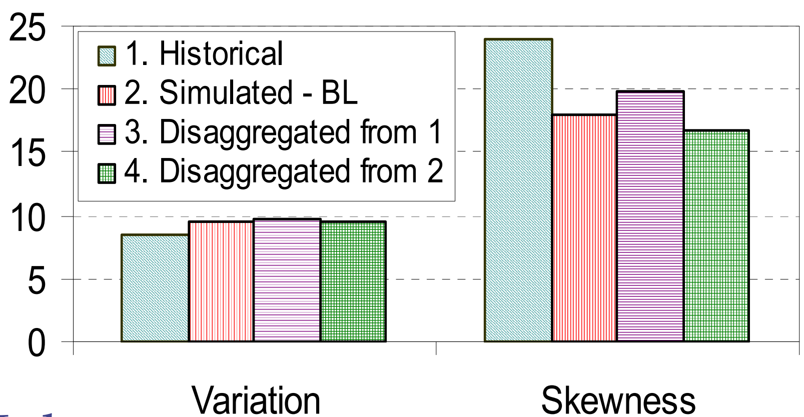
July (The wettest month)

2. Preservation of marginal moments

Heathrow Airport

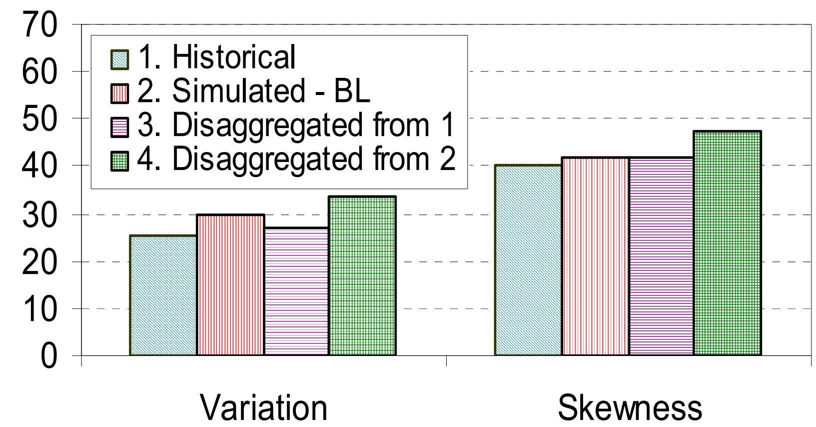


January

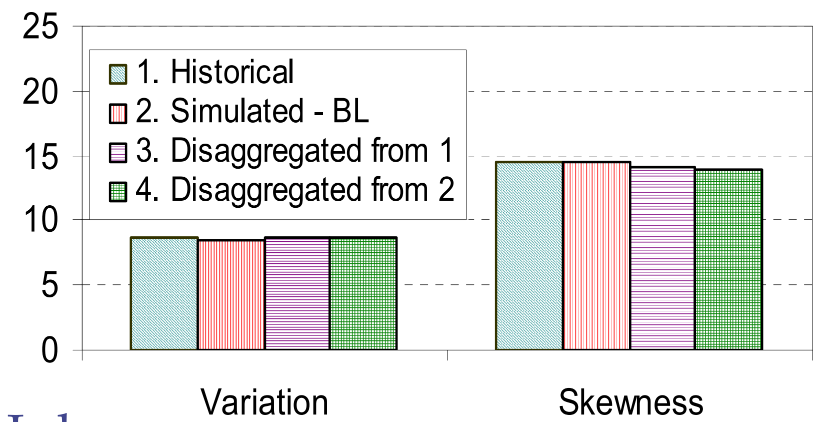


July

Walnut Gulch (Gauge 13)



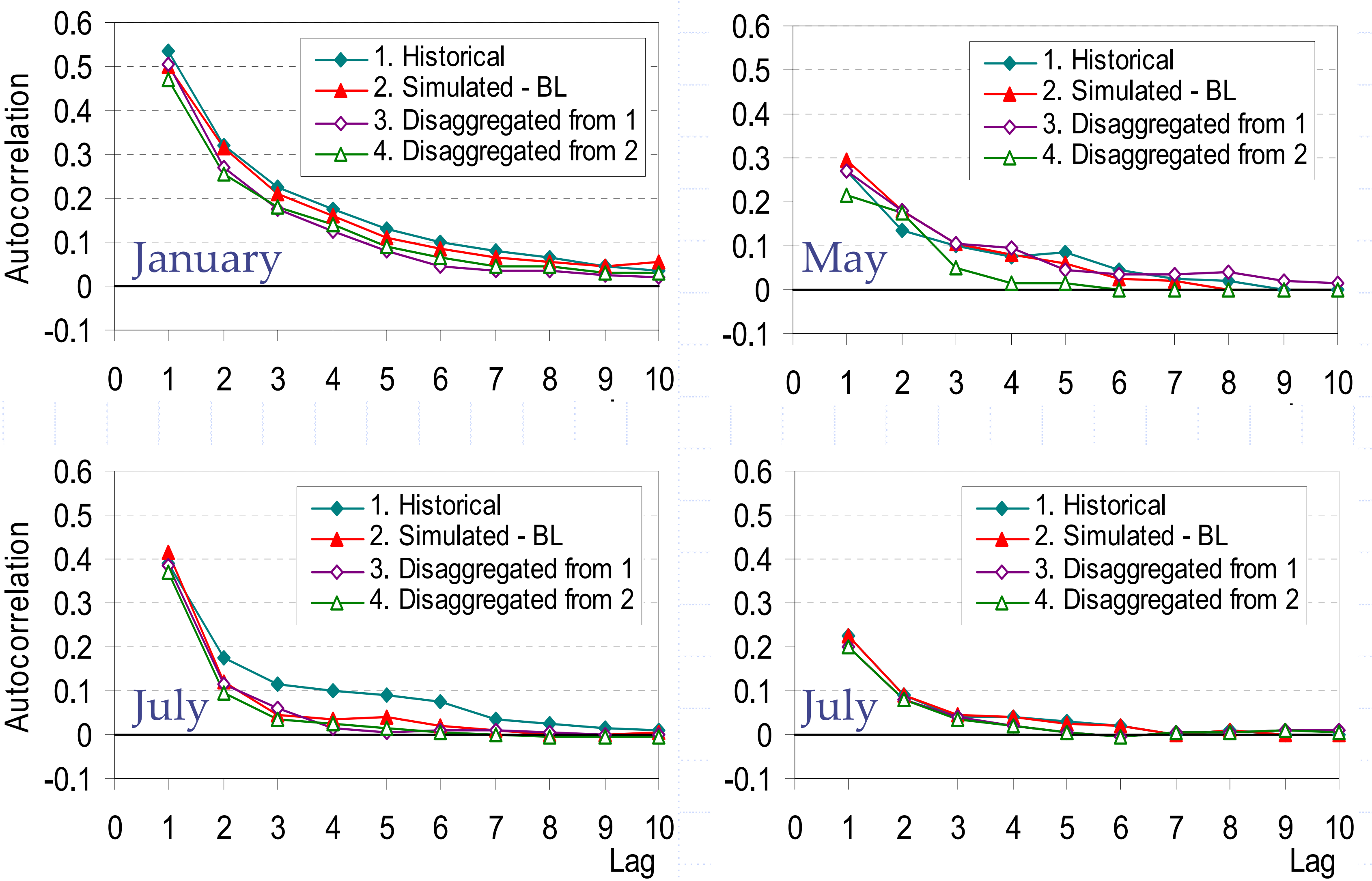
May



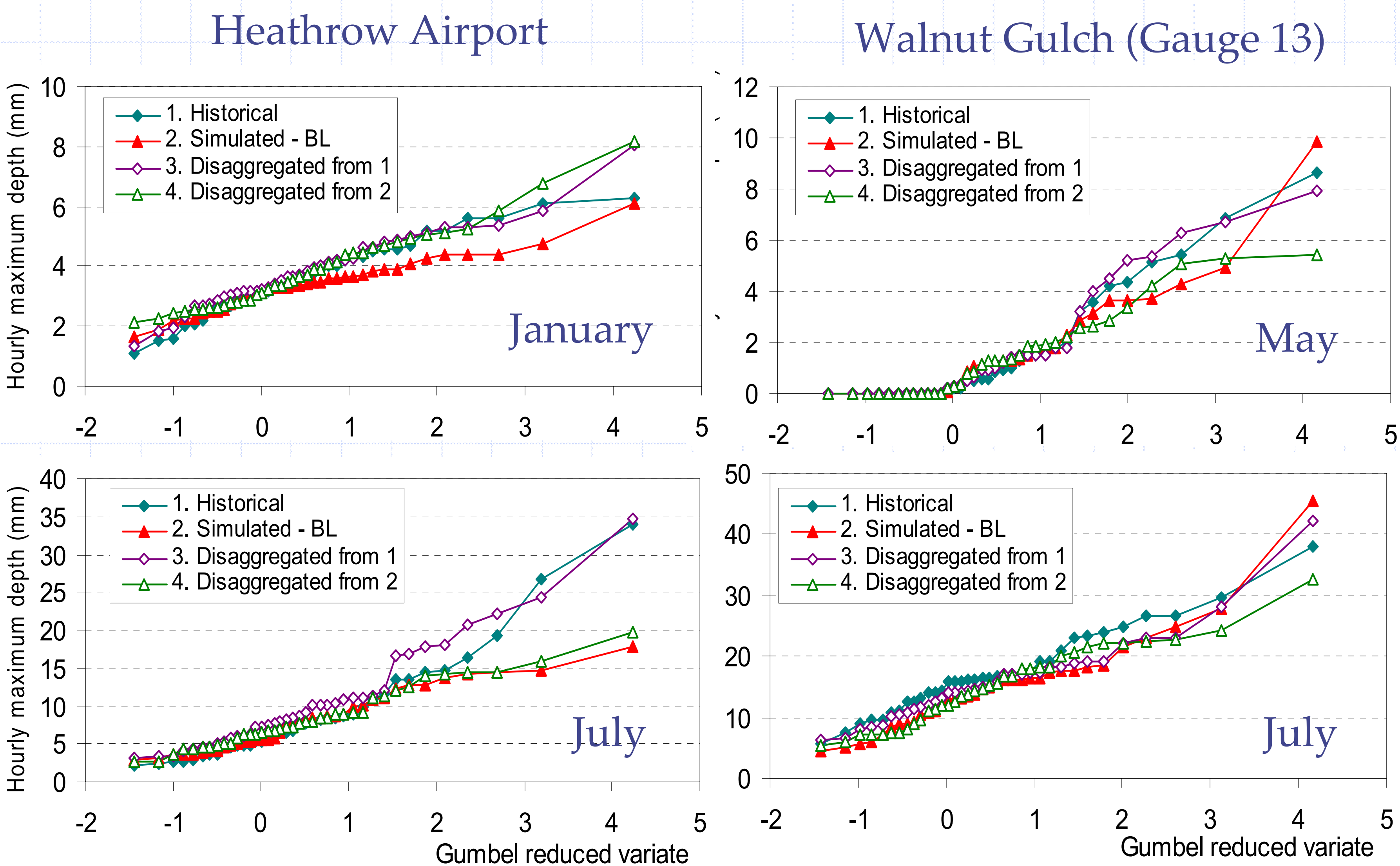
July

3. Preservation of autocorrelations

Heathrow Airport | Walnut Gulch (Gauge 13)

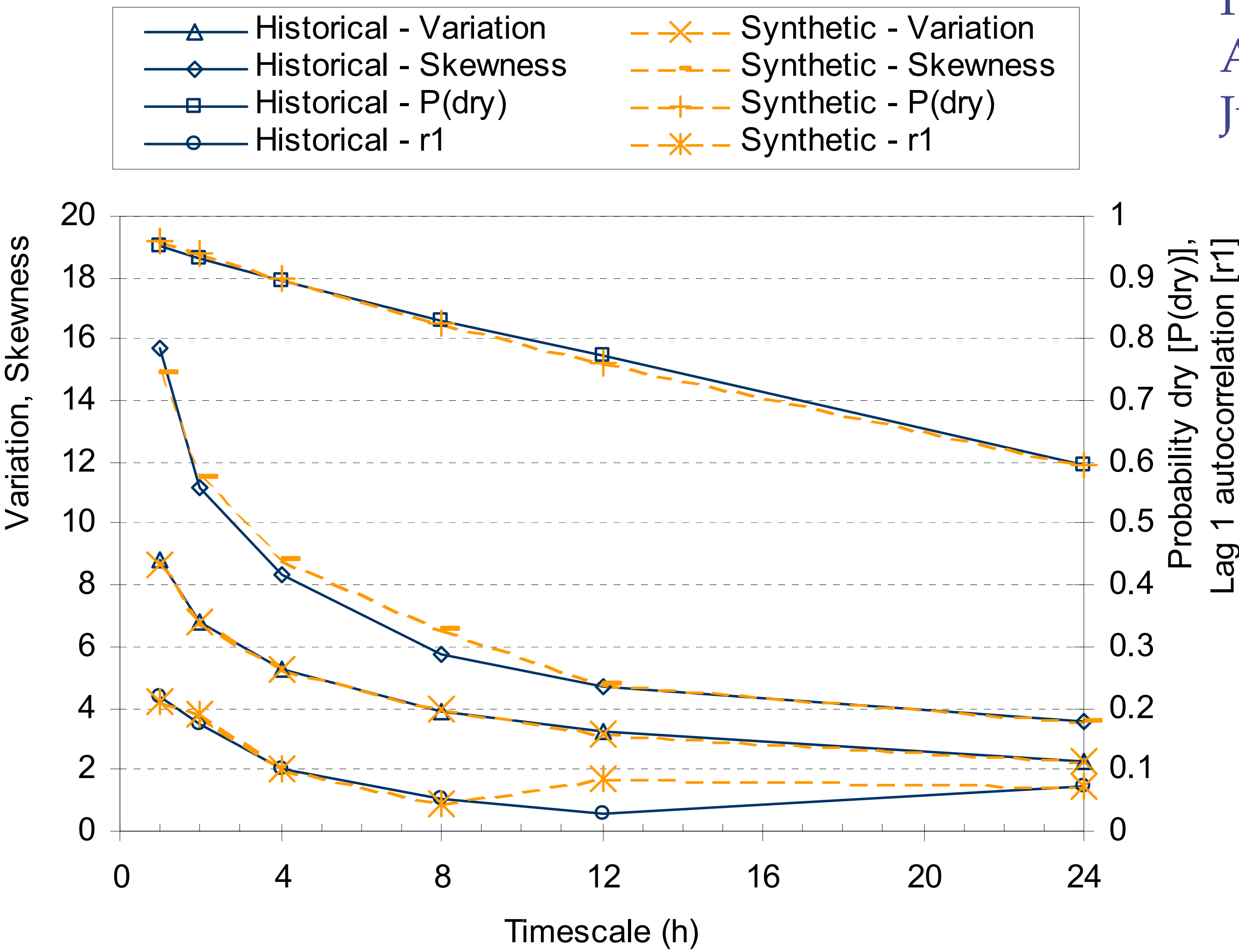


4. Distribution of hourly maximum depths



5. Preservation of statistics at intermediate scales

Heathrow Airport, July



MuDRain: A model for multivariate disaggregation of rainfall at a fine time scale

Basic assumptions

- The disaggregation is performed at *n* sites simultaneously
- At all *n* sites there are higher-level (daily) time series available, derived either
  - from measurement or
  - from a stochastic model (daily)
- At one or more of the *n* sites there are lower-level (hourly) series available, derived either
  - from measurement or
  - from a stochastic model (hourly, e.g. Hyetos)
- The lower-level rainfall process at the remaining sites can be generated by a simplified multivariate AR(1) model ($\mathbf{X}_s = \mathbf{a}\ \mathbf{X}_{s-1} + \mathbf{b}\ \mathbf{V}_s$) utilising the cross-correlations among the different sites

MuDRain = multivariate AR(1) + repetition + coupling transformation

Can a simple AR(1) model describe the rainfall process adequately?

Exploration of the distribution function of hourly rain depths during wet days

Data: a 5-year time series of January (95 wet days, 2280 data values)

Location: Gauge 1, Brue catchment, SW England

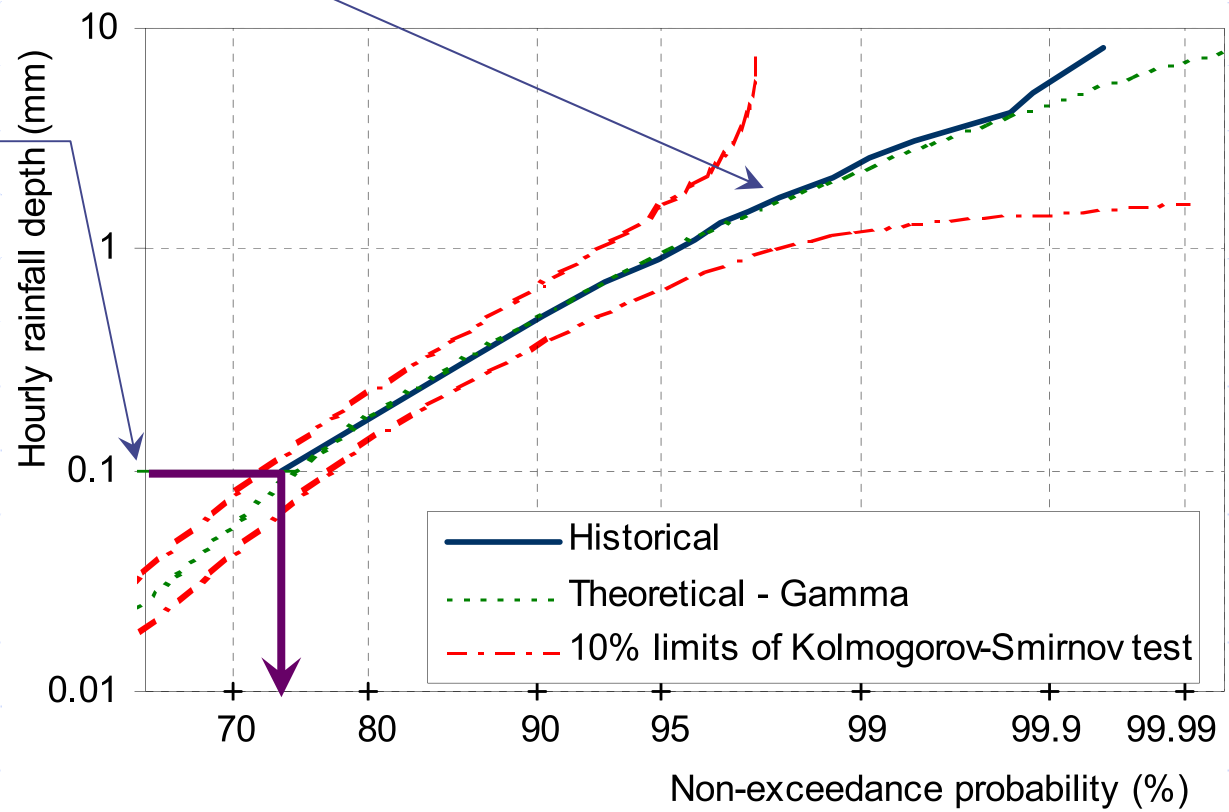
Highly skewed distribution

73.6% of values (1679 values) are zeros

The smallest measured values are 0.2 mm

Measured zeros can be equivalently regarded as < 0.1 mm

With this assumption, a gamma distribution can be fitted to the entire domain of the rainfall depth

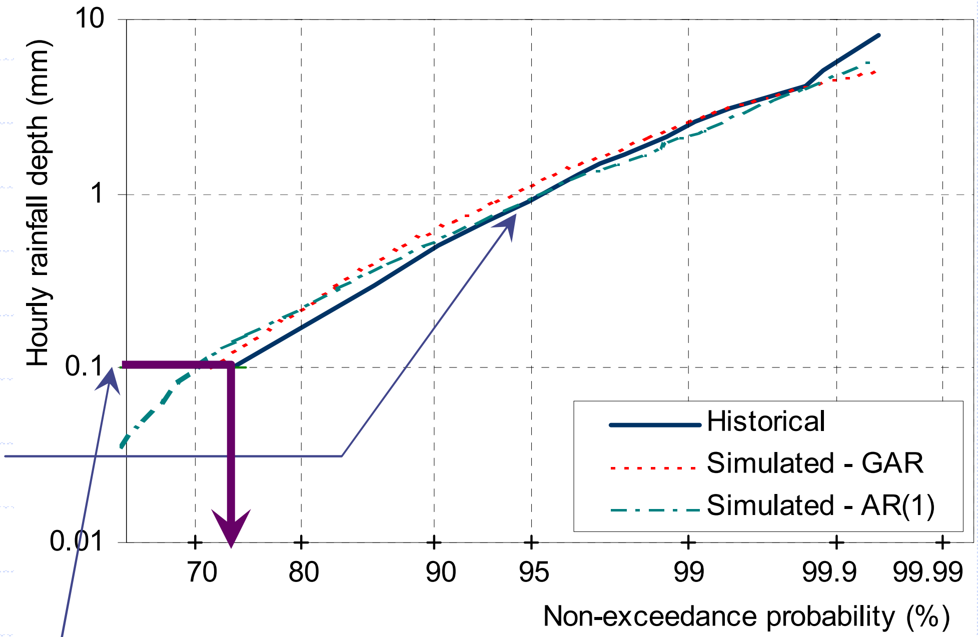


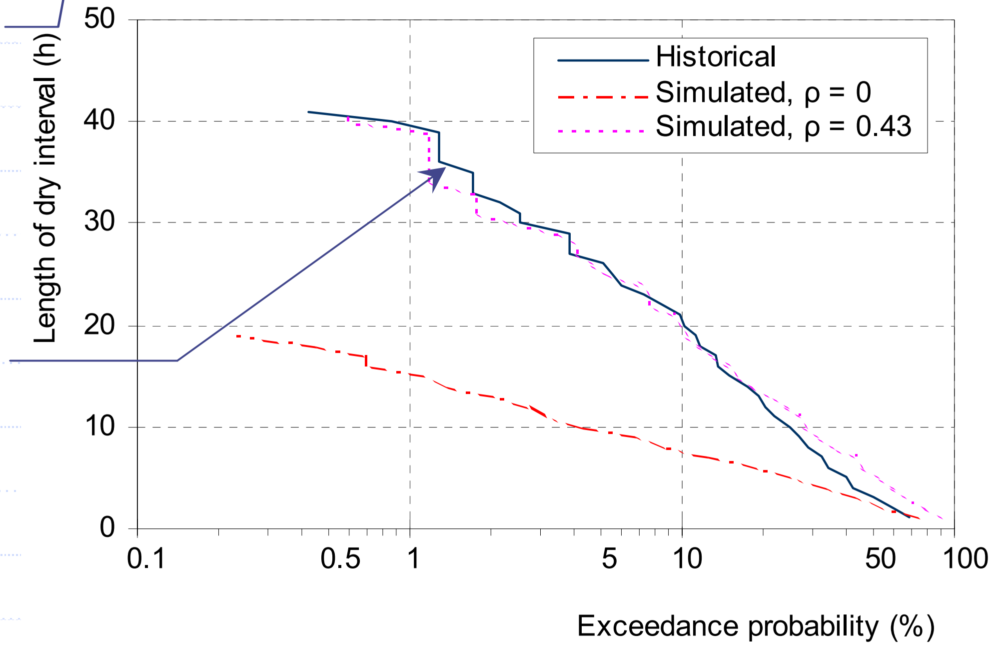
Simulation results using a GAR and an AR(1) model

The distribution of hourly rainfall depth is represented adequately both by the GAR and the AR(1) models

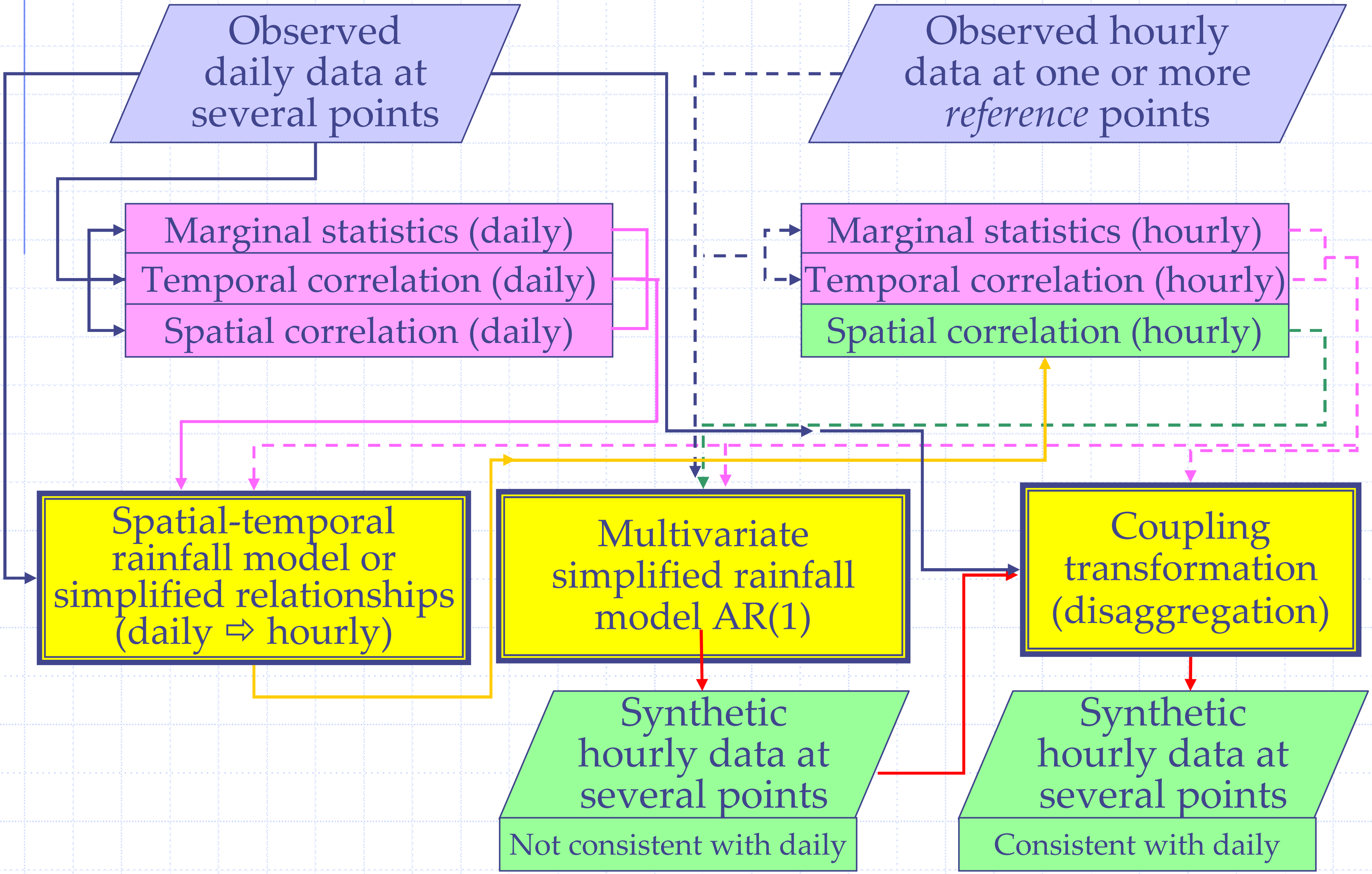
Intermittency is reproduced well if we truncate to zero all generated values that are smaller than 0.1 mm

The distribution of the length of dry intervals is represented adequately if the historical lag-1 autocorrelation is used in simulation

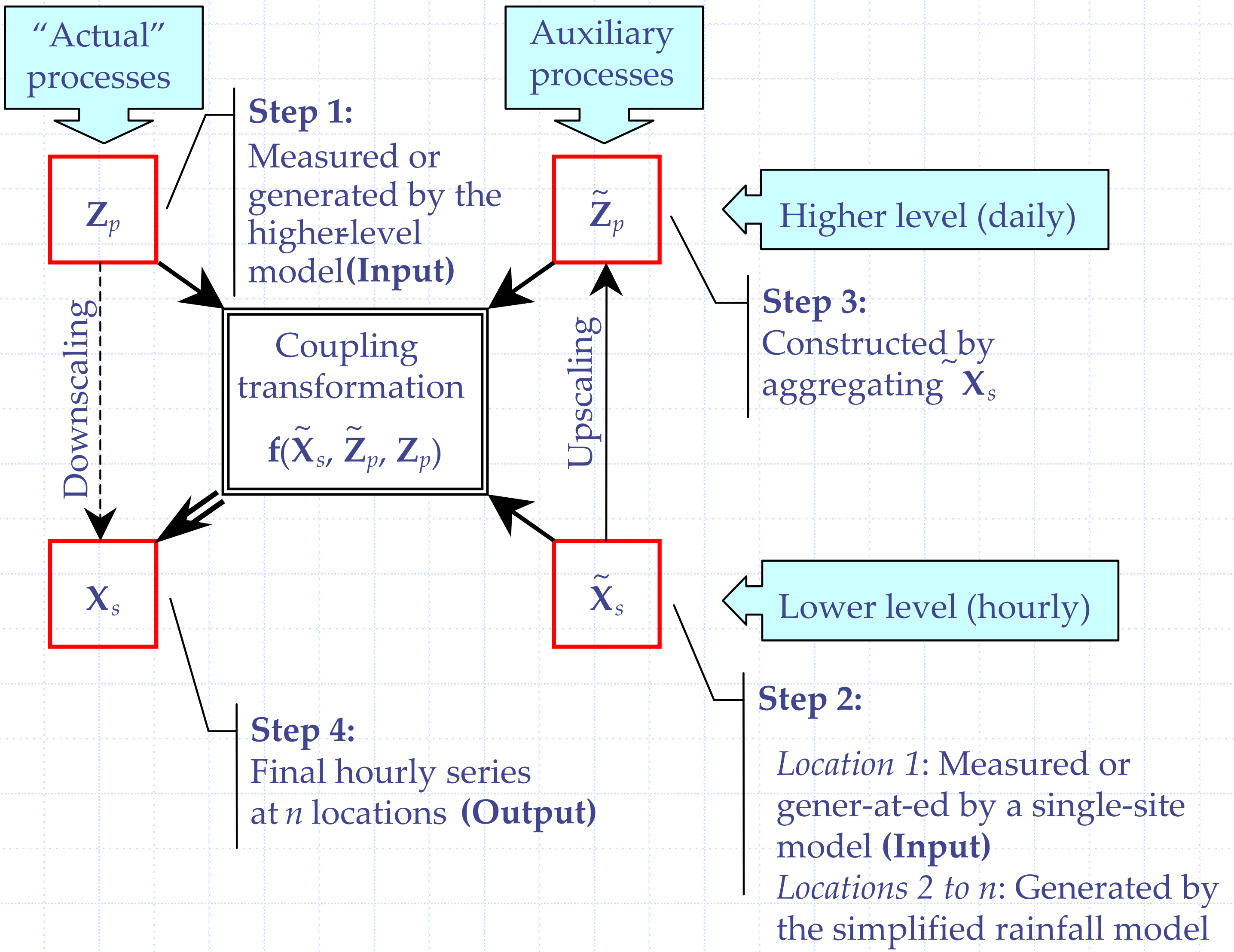




MuDRain: The modelling approach



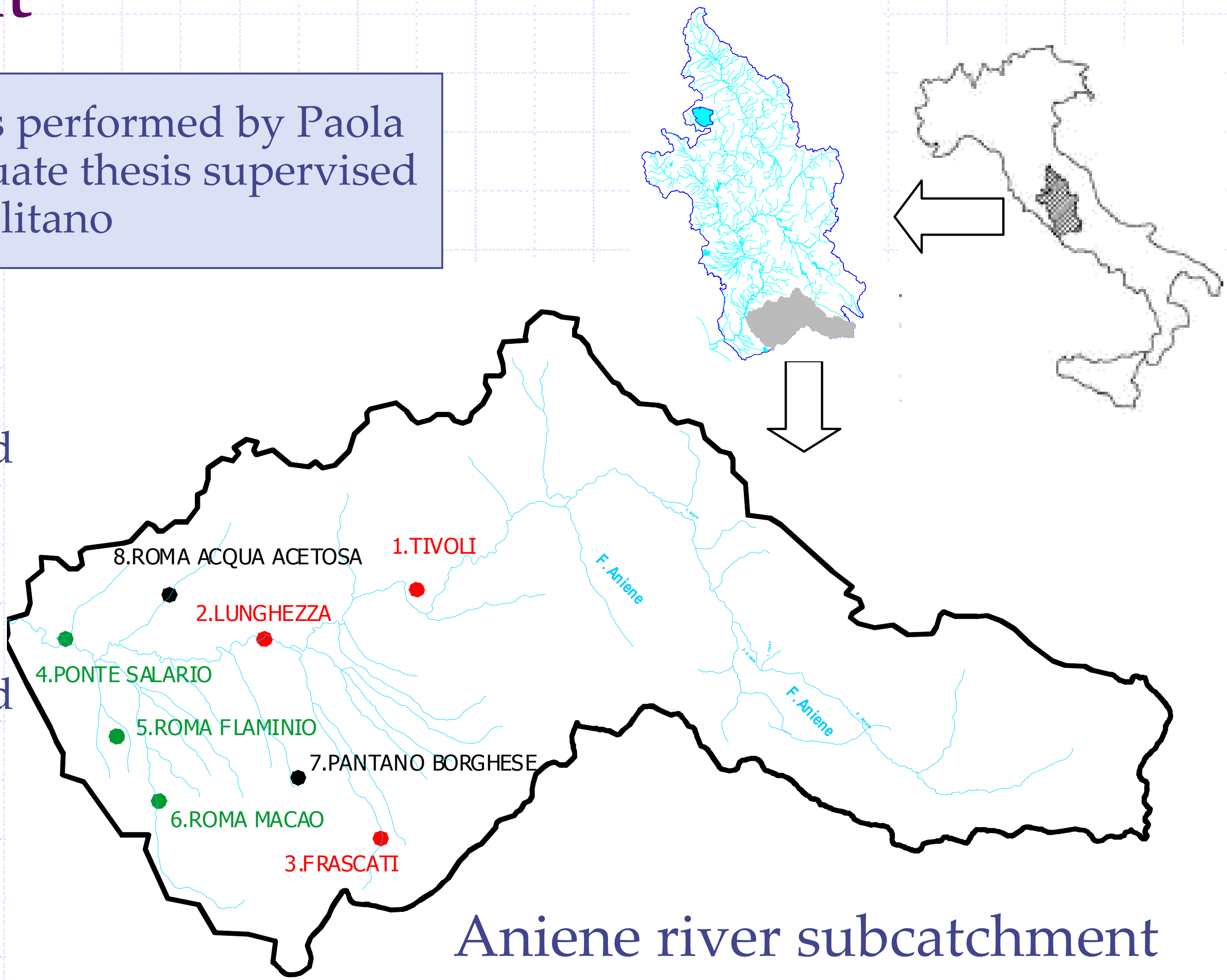
MuDRain: The simulation approach



MuDRain: Case study at the Tiber river catchment

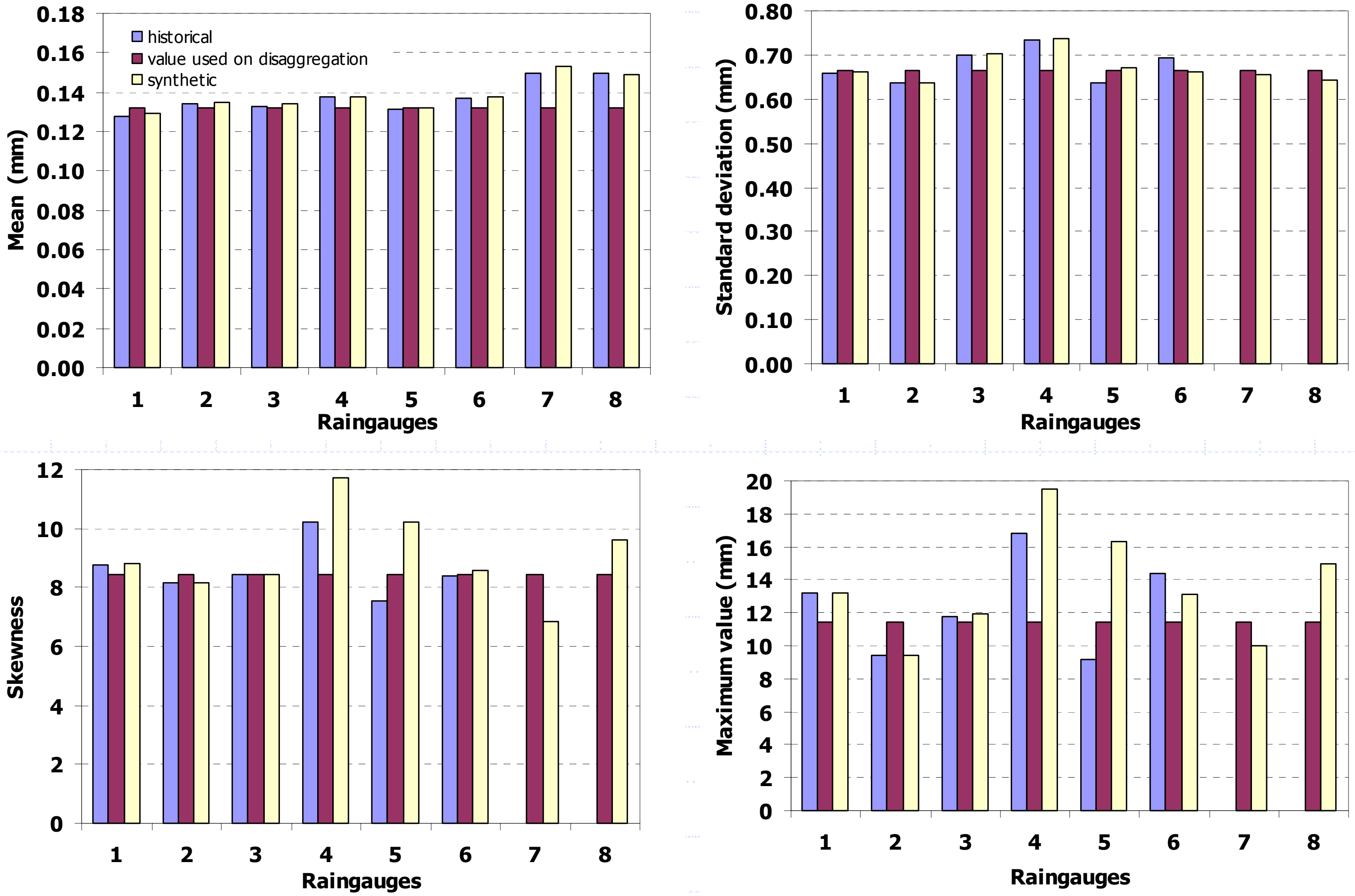
The case study was performed by Paola Fytilas in her graduate thesis supervised by Francesco Napolitano

- Reference stations
  Hourly data available and used in simulations
- Test stations
  Hourly data available and used in tests only
- Disaggregation stations
  Only daily data available

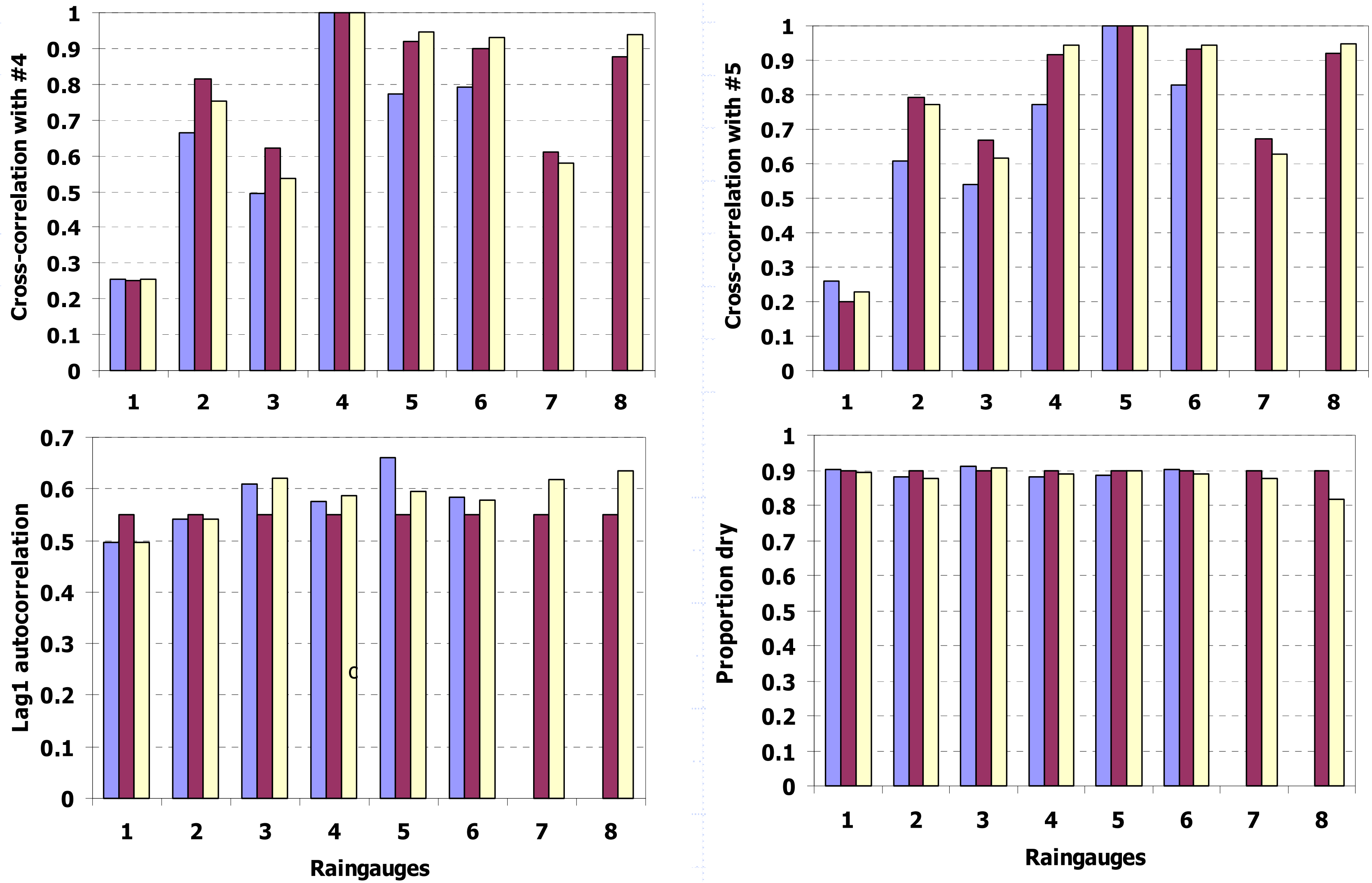


Aniene river subcatchment

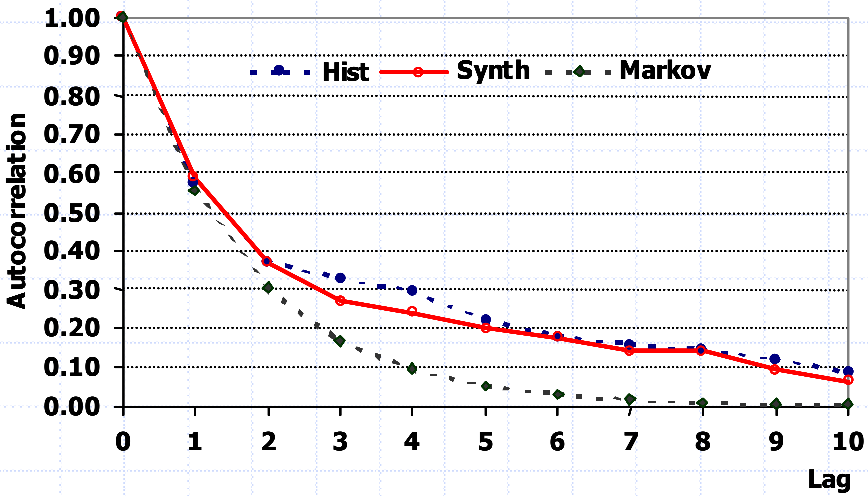
1. Preservation of marginal statistics of hourly rainfall

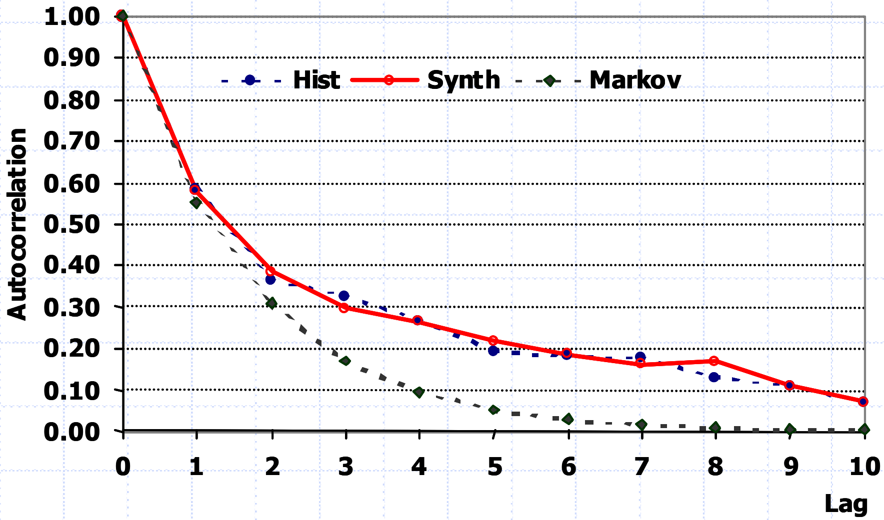


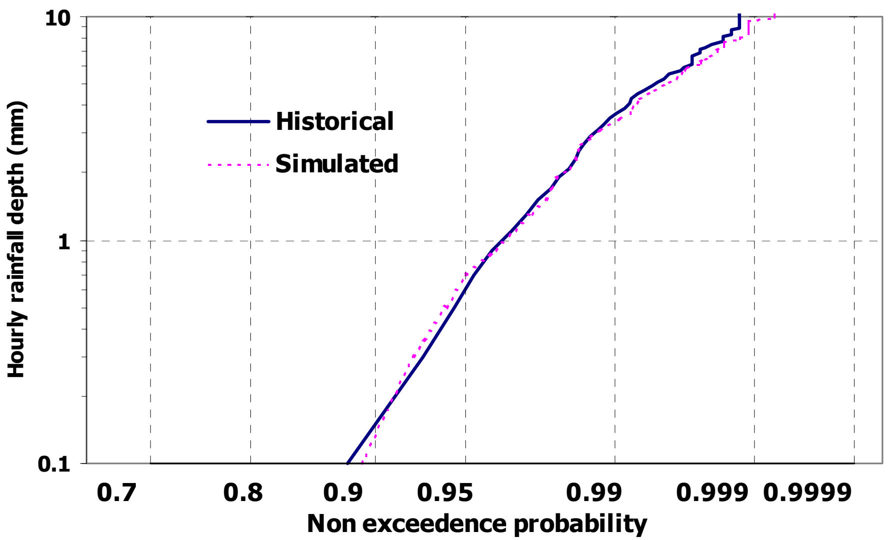
2. Preservation of cross-correlations, autocorrelations and probabilities dry

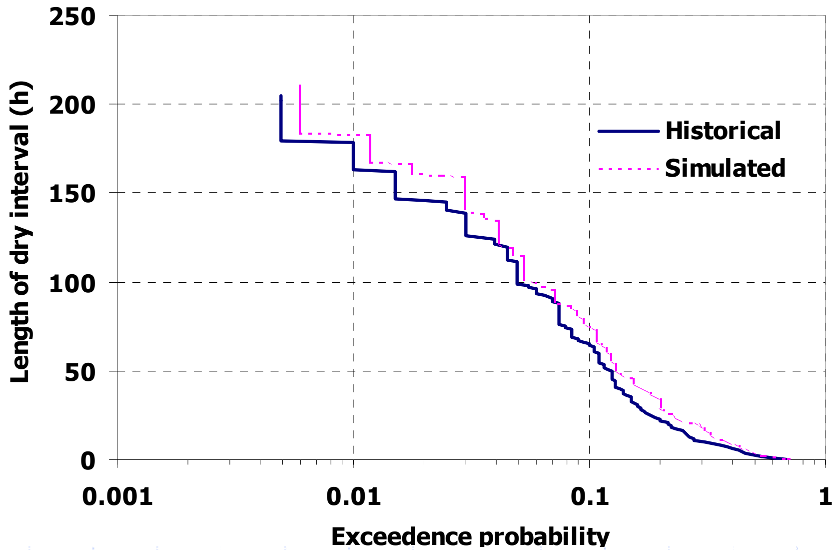


3. Preservation of autocorrelation functions and distribution functions

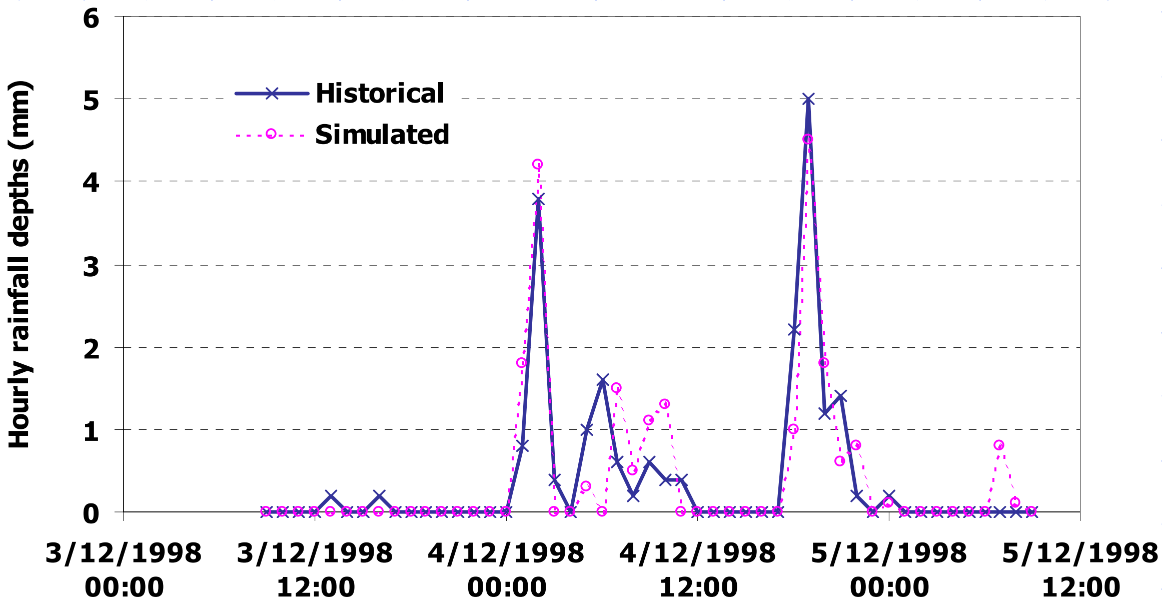


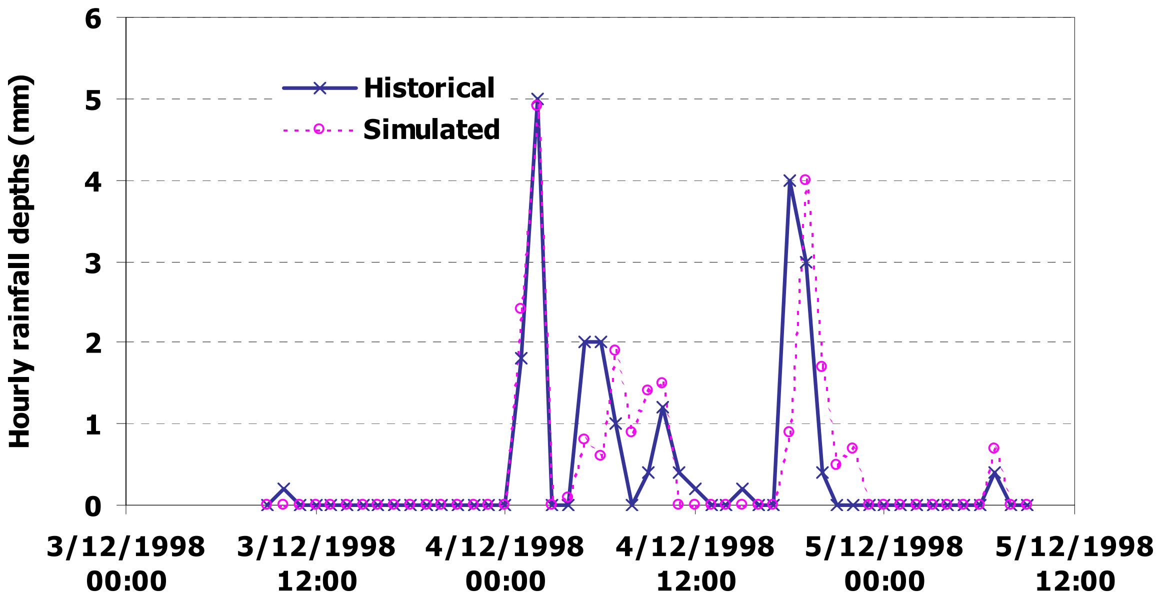




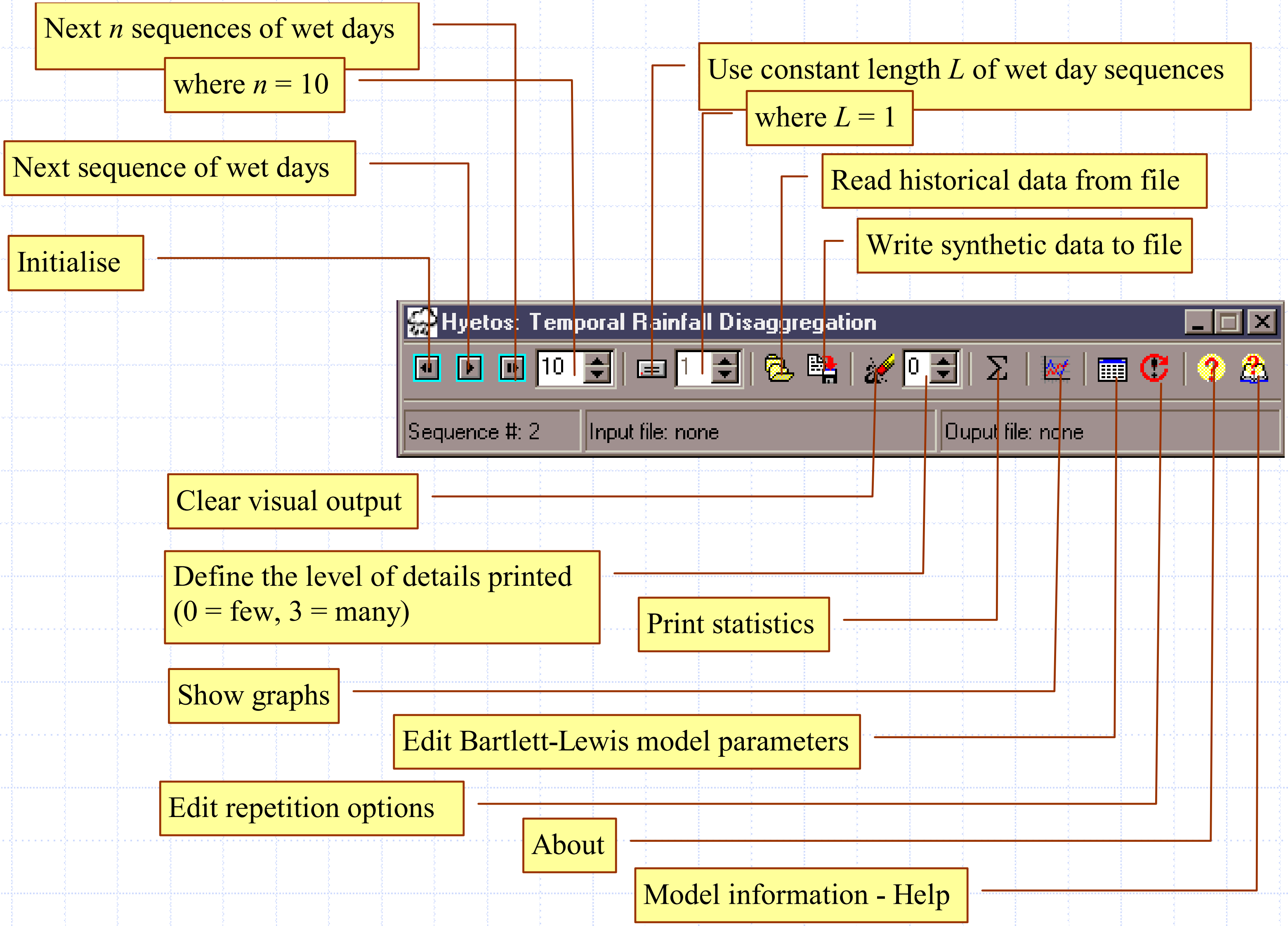


4. Preservation of actual hyetographs at test stations

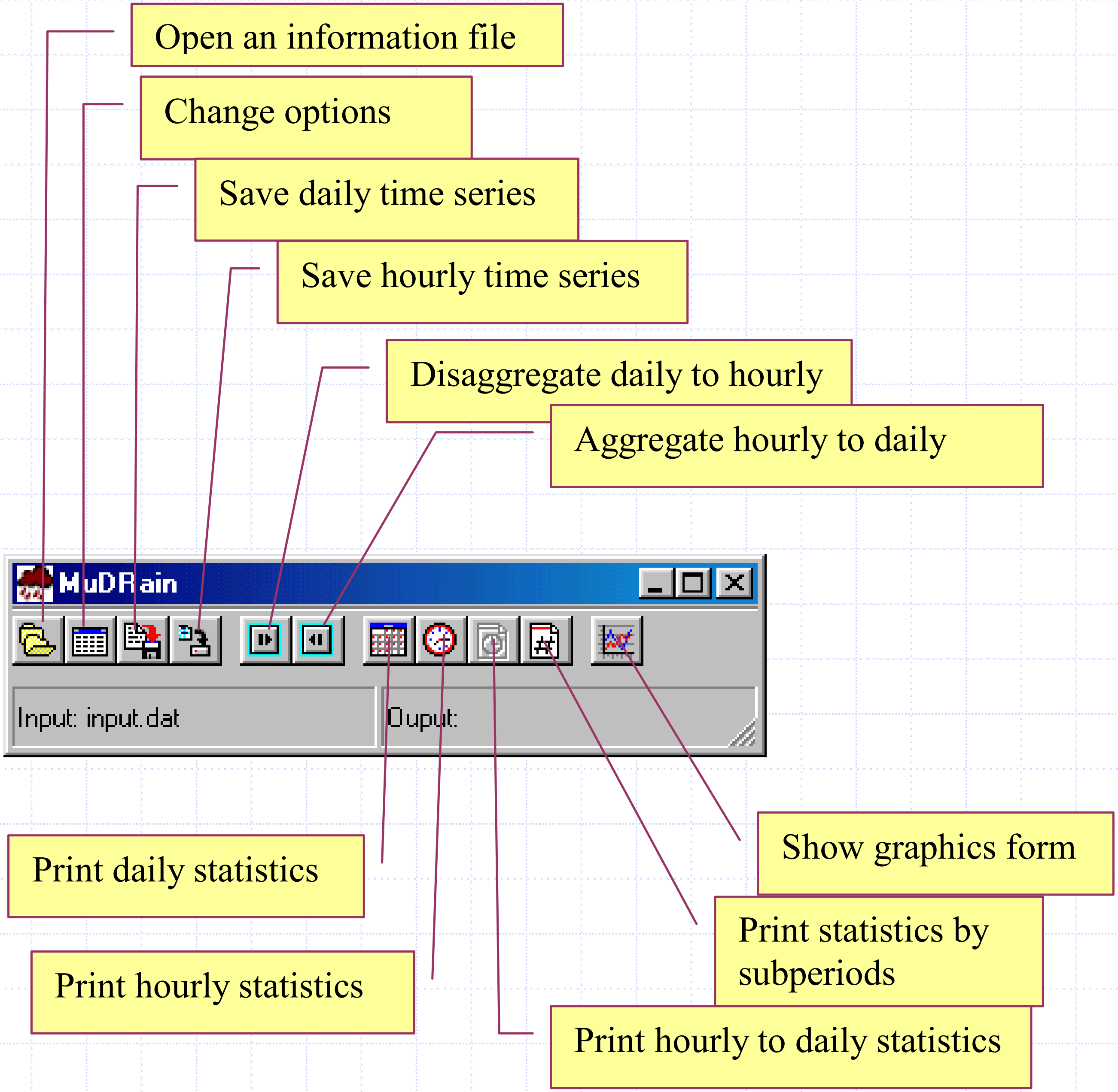




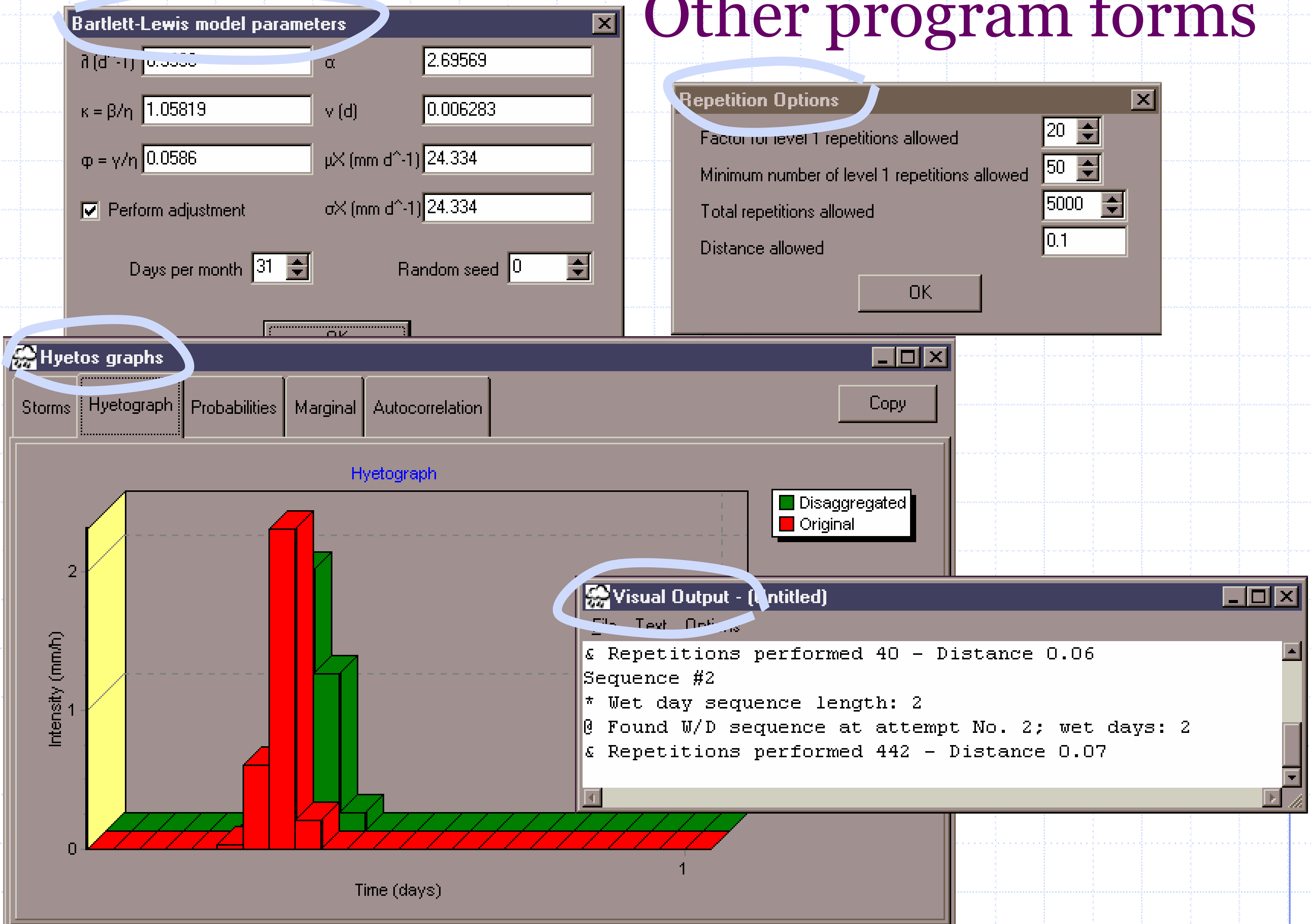
Hyetos main program form



MuDRain main program form



Other program forms



Conclusions (1) Historical evolution

- After more than thirty years of extensive research, a large variety of stochastic disaggregation models have been developed and applied in hydrological studies
- Until recently, there was a significant divergence between general-purpose disaggregation methods and rainfall disaggregation methods
- The general-purpose methods are generally multivariate while rainfall disaggregation models were only applicable in single variate setting
- Only recently there appeared developments in multivariate rainfall disaggregation along with implementation of general-purpose methodologies into rainfall disaggregation

Conclusions (2) Single-variate versus multivariate models

- Simulations
  - Current hydrological modelling requires spatially distributed information
  - Thus, multivariable rainfall disaggregation models have greater potential as they generate multivariate fields at fine temporal resolution
- Enhancement of data records
  - Provided that there exists at least one fine resolution raingauge at the area of interest, multivariate models have greater potential to disaggregate daily rainfall at finer time scales, because
    - they can derive spatially consistent rainfall series in number of raingauges simultaneously, in which only daily data are available
    - they can utilise the spatial correlation of the rainfall field to derive more realistic hyetographs

Conclusions (3) The Hyetos and MuDRain software applications

- Hyetos combines the strengths of a standard single-variate stochastic rainfall model, the Bartlett-Lewis model, and a general-purpose disaggregation technique
- MuDRain combines the simplicity of the multivariate AR(1) model, the faithfulness of a more detailed single-variate model (for one location) and the strength of a general-purpose multivariate disaggregation technique
- Both models are implemented in a user-friendly Windows environment, offering several means for user interaction and visualisation
- Both programs can work in several modes, appropriate for operational use and model testing

Conclusions (4) Results of case studies using Hyetos and MuDRain

- The case studies presented, regarding the disaggregation of daily historical data into hourly series, showed that both Hyetos and MuDRain result in good preservation of important properties of the rainfall process such as
  - marginal moments (including skewness)
  - temporal correlations
  - proportions and lengths of dry intervals
  - distribution functions (including distributions of maxima)
- In addition, the multivariate MuDRain provides a good reproduction of
  - spatial correlations
  - actual hyetographs

Future developments and applications

- Application of existing methodologies in different climates
- Refinement and remediation of weaknesses of existing methodologies
- Development of new methodologies

Future challenges

- Modern technologies in measurement of rainfall fields, including weather radars and satellite images, will
  - improve our knowledge of the rainfall process
  - provide more reliable and detailed information for modelling and parameter estimation of rainfall fields
- In future situations the need for enhancing data sets will be limited but simulation studies will ever require investigation of the process at many scales
- As a single model can hardly be appropriate for all scales simultaneously, it may be conjectured that there will be space for disaggregation methods even with the future enhanced data sets
- Future disaggregation methods should need to give emphasis to the spatial extent of rainfall fields in order to incorporate information from radar and satellite data

This presentation is available on line at

http://www.itia.ntua.gr/e/docinfo/570/

Programs Hyetos and MuDRain are free and available on the web at

http://www.itia.ntua.gr/e/softinfo/3/

and http://www.itia.ntua.gr/e/softinfo/1/

The references shown in the presentation can be found in the Workshop proceedings