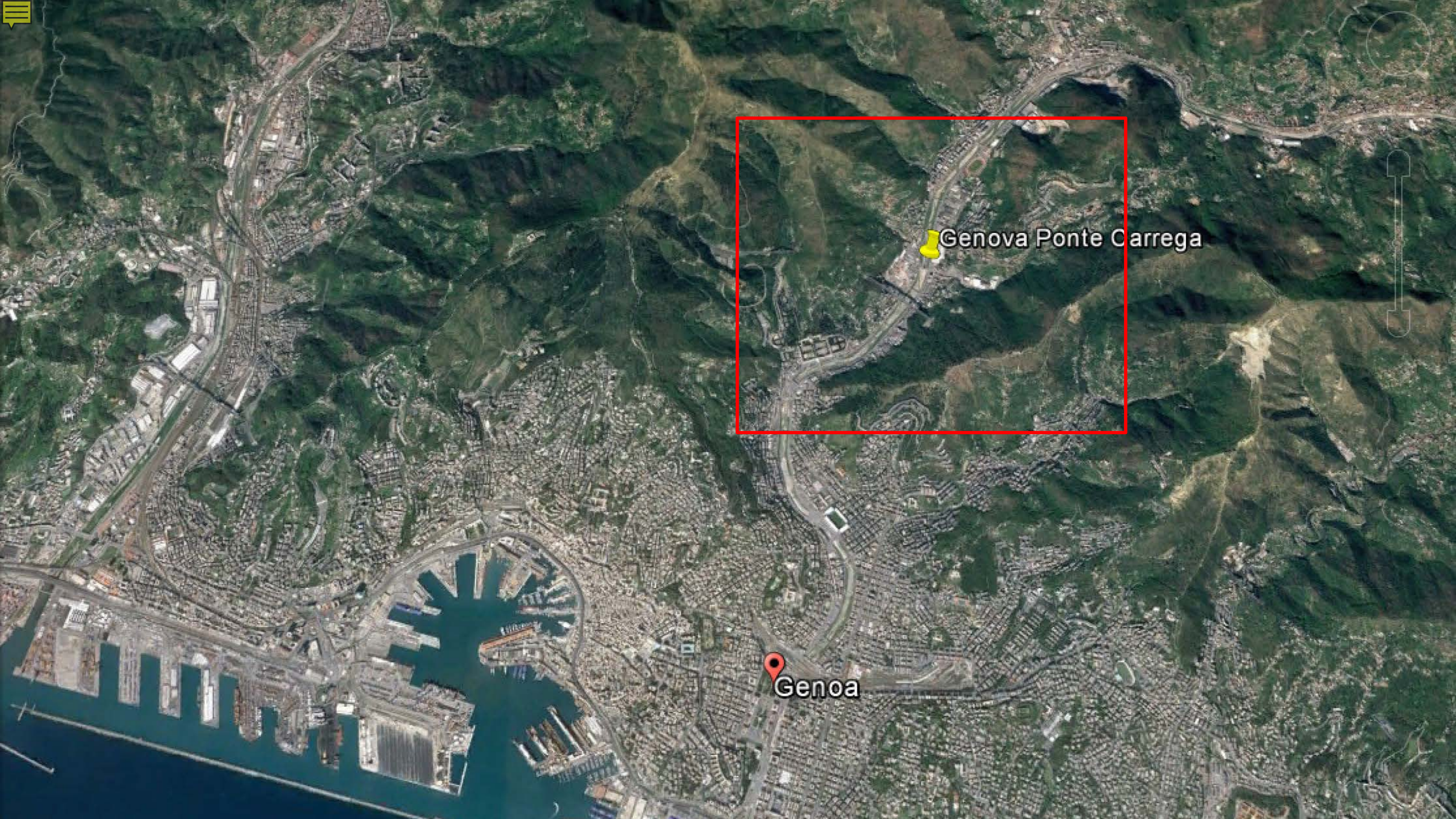
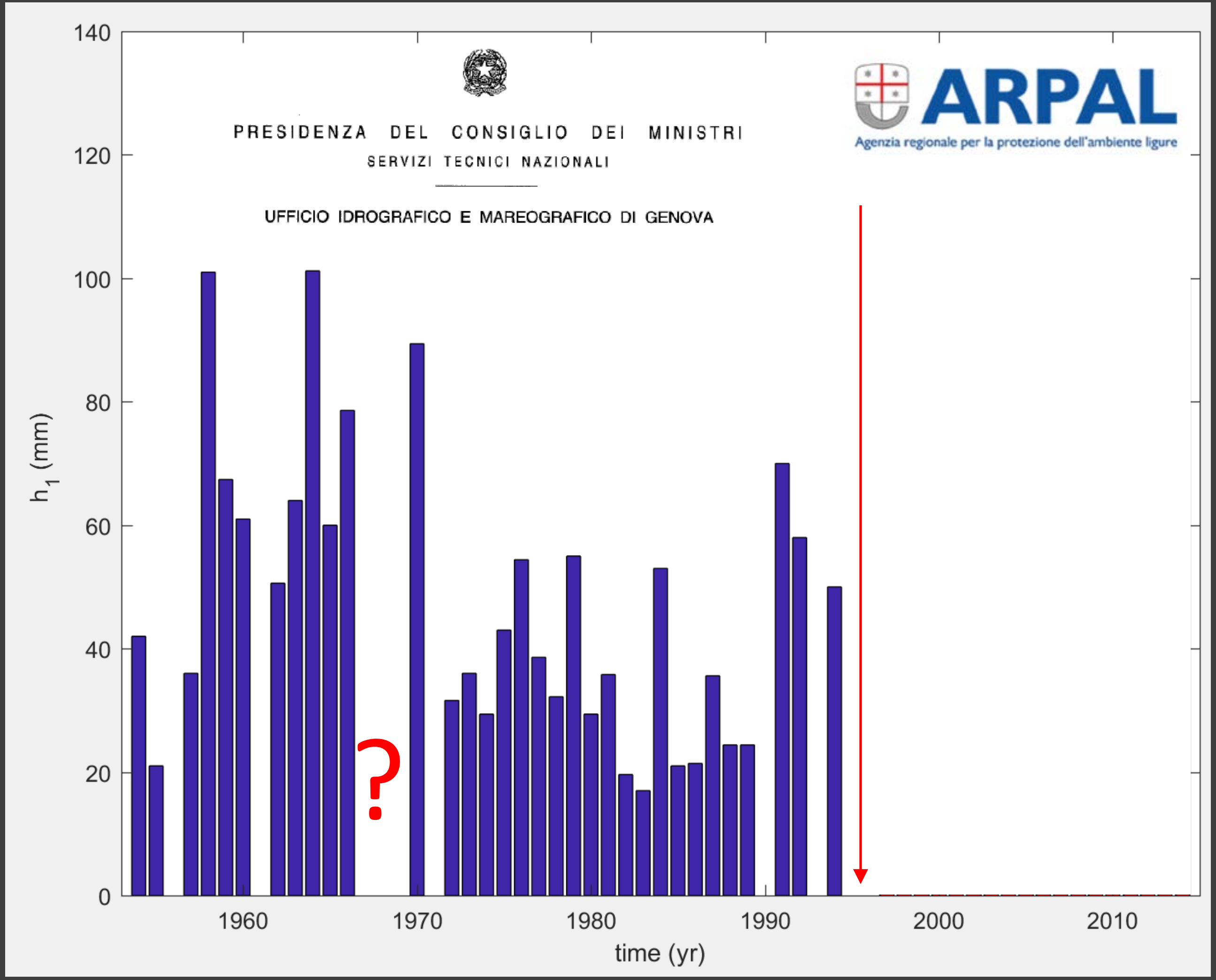
ADVANCES IN THE SPACE-TIME ANALYSIS OF RAINFALL EXTREMES

Spatial analysis of extreme rainfall in a data-rich fragmented framework

Andrea Libertino
andrea.libertino@polito.it
Politecnico di Torino


Genova Ponte Carrega
Genoa



Genova Ponte Carrega (GEPGA)

Genova Gavette (GEPGA)

TABLE OF CONTENTS

AIM:

Providing operational instruments for carrying out a robust regional rainfall frequency analysis in a data-rich fragmented framework.

OUTLINE:

INTRODUCTION: FREQUENCY ANALYSIS OF EXTREMES

h		T		P(h>h')
Rainfall depth (mm)	⟷	Return period (years)	⟷	Exceedance probability

The return period is the inverse of the probability of a rainfall depth to be exceeded in a year.
On average, a T-year rainfall depth is exceeded once in T years.

$$T = \frac{1}{P(h > h')} = \frac{1}{1 - F(h)}$$

Non-exceedance probability

→ THE INDEX METHOD[1]

h

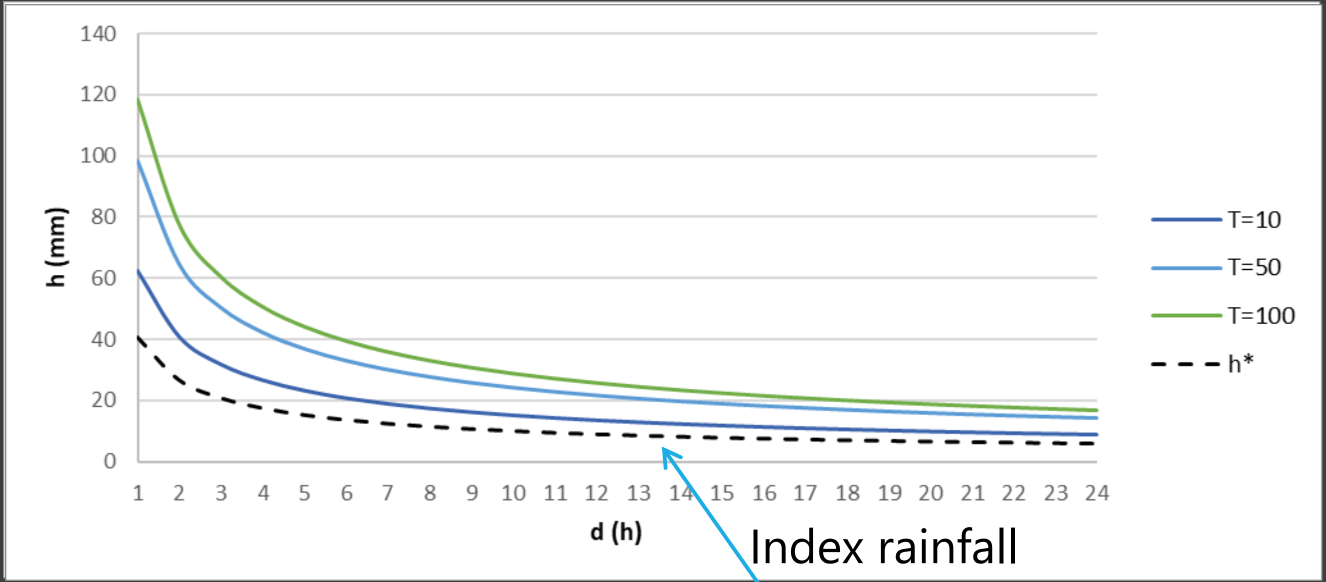
Rainfall depth (mm)

↔

T

Return period (years)

IDF curves



Index rainfall

$$h_{d,T} = \bar{h}_d \cdot K(T) = a \cdot d^n \cdot K(T)$$

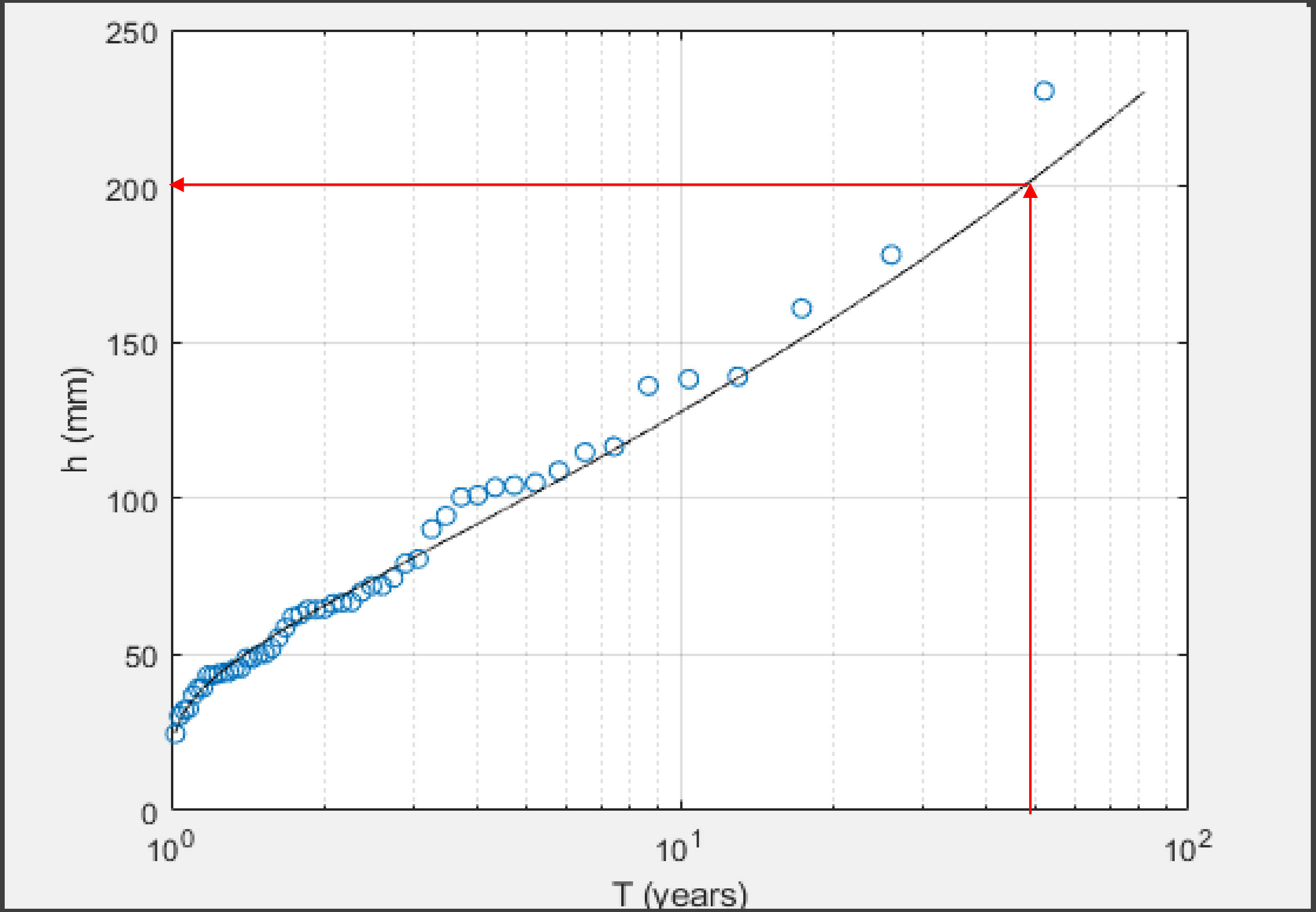
Growth curve

[1] *Dalrymple, 1960*

→ ESTIMATING THE GROWTH CURVE

STEPS:

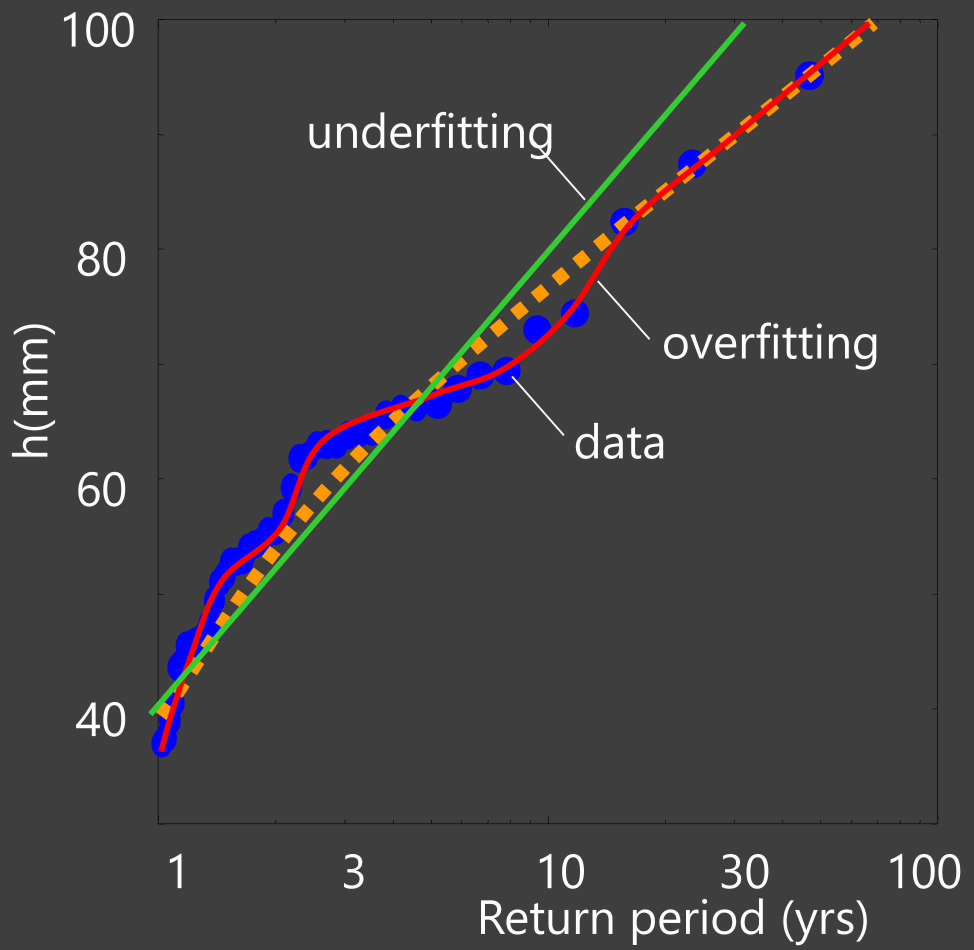
- Compile sample of rainfall depth
- Estimate empirical frequency
- Fitting a distribution function F(h)
- Reading off h_T



→ FITTING A DISTRIBUTION FUNCTION

1st STEP: Choice of the distribution function

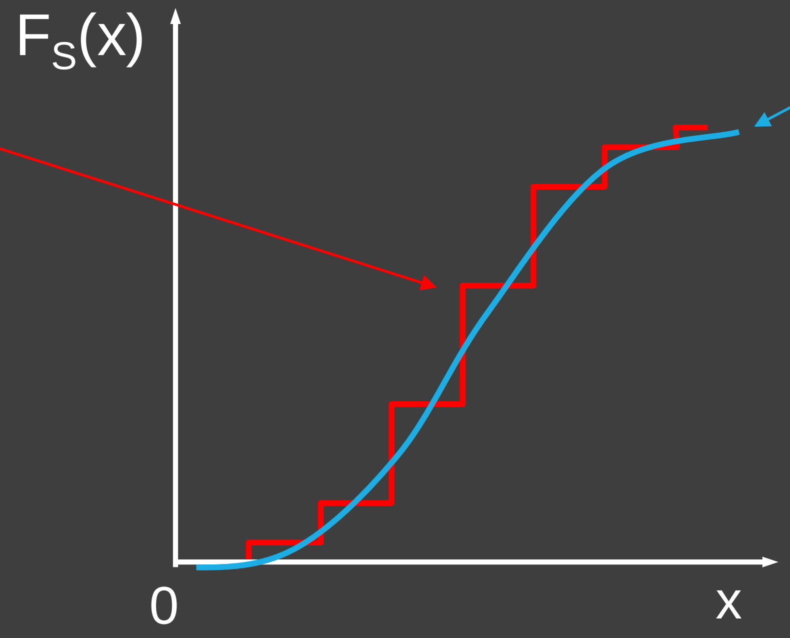
2 Parameter:	Gumbel Logistic Log-normal
3 Parameter:	General Extreme Value (GEV) Generalised Logistic Pearson Type III Log Pearson Type III
4 Parameter:	Kappa
5 Parameter:	Wakeby



→ FITTING A DISTRIBUTION FUNCTION
<u>2nd STEP: Parameters estimation</u>

Finding the characteristics of the population (all possible future rainfall depths) from the sample (I.e. the depths observed in the past)

Cumulative frequency distribution of the sample



Distribution function $F(x_i)$

E.g., Gumbel distribution

$$F(x) = \exp\left(-\exp\left(-\frac{x-c}{d}\right)\right)$$

Estimate the parameters of the formula so that the Cumulative Density Function of the distribution fits the empirical cumulative frequency of the data.

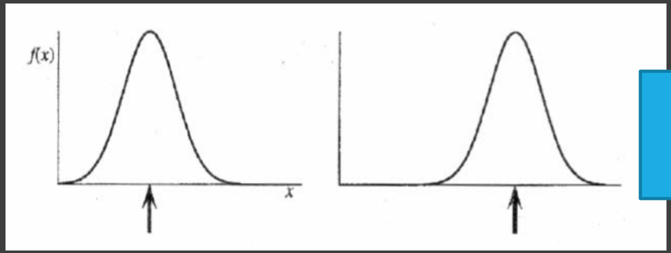
PARAMETERS ESTIMATION METHODS

Method of Moments

Moments of the population = moments of the sample

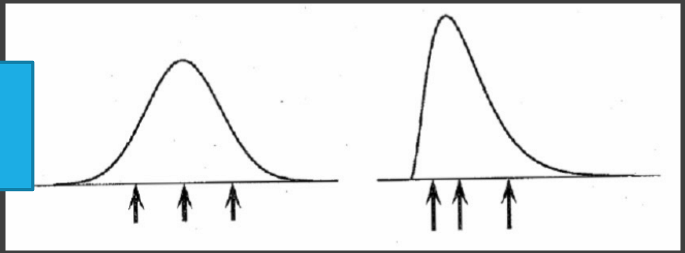
Method of L-Moments

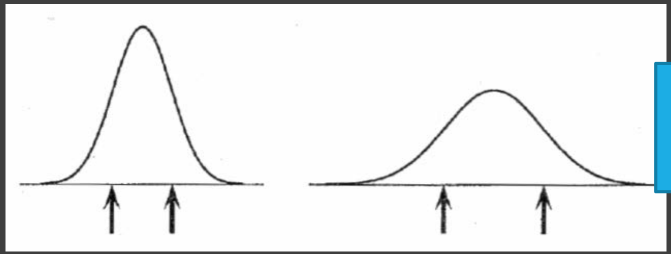
L-moments can summarize data as do conventional moments using linear combinations of the ordered observations.



λ_1

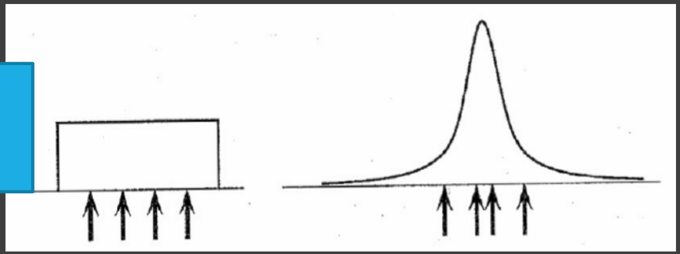
λ_3





λ_2

λ_4



→ THE L-MOMENTS METHOD

L-moments of
the distribution function

Sample L-moments
of the data series

$$l_{r+1} = \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} b_k$$

$$b_r = \frac{1}{n} \sum_{j=r+1}^{n} \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_j{:}n$$

Sample PWM

E.g., Gumbel distribution

$$F(x) = \exp\left(-\exp\left(-\frac{x-c}{d}\right)\right)$$

$$\lambda_1 = c + 0.5772 \cdot d$$
$$\lambda_2 = d \cdot \ln(2)$$

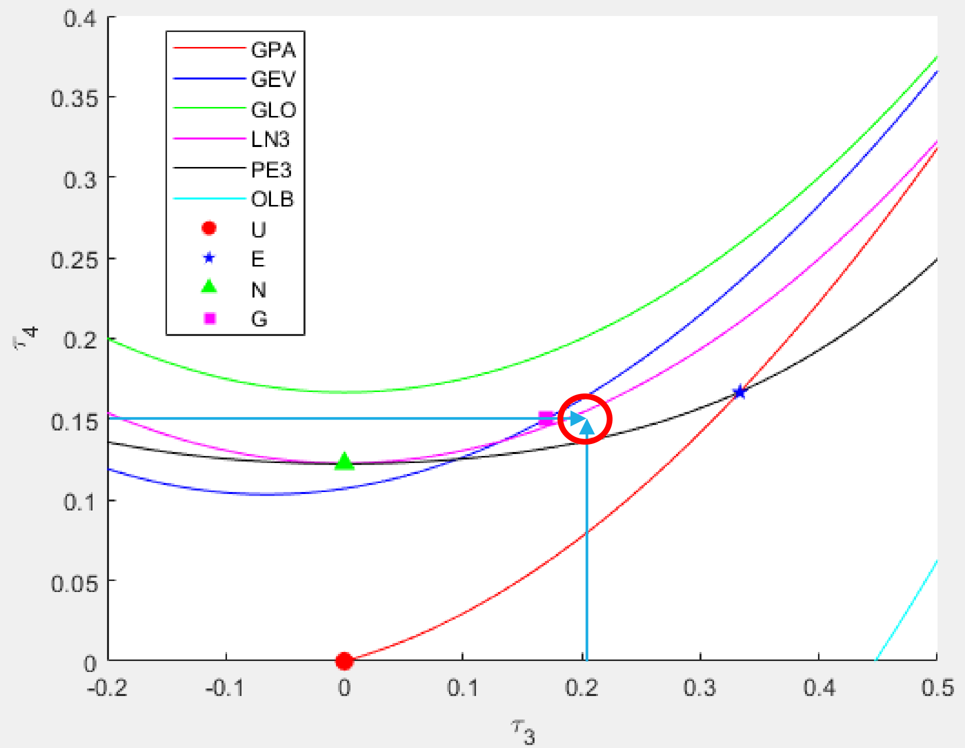
$$l_1 = b_0$$
$$l_2 = 2 \cdot b_1 - b_0$$

→ THE L-MOMENTS METHOD

Because L-moments avoid squaring and cubing the data, their ratios do not suffer from the severe bias problems encountered with product moments.

Dimensionless L-moments ratios give further information on the characteristic of the distribution:

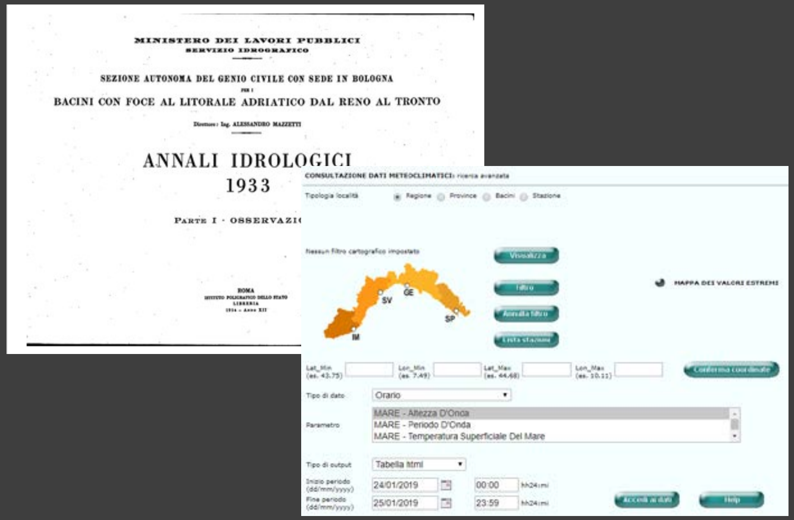
- L-coefficient of variation (L-CV):
  $\tau = \lambda_2/\lambda_1 = \lambda_2/\mu$
- L-coefficient of skewness (L-skewness or L-CA):
  $\tau_3 = \lambda_3/\lambda_2$
- L-coefficient of kurtosis (L-kurtosis or L-KUR):
  $\tau_4 = \lambda_4/\lambda_2$



L-Moments ratio diagram[2]

[2] *Hosking and Wallis, 1997*

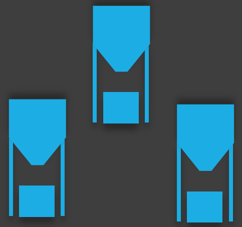
→ WHAT KIND OF DATA?



Annual maxima for 1-3-6-12 24 hours durations

High degree of accuracy at point location

Spatial variability of rainfall fields



Long temporal consistency

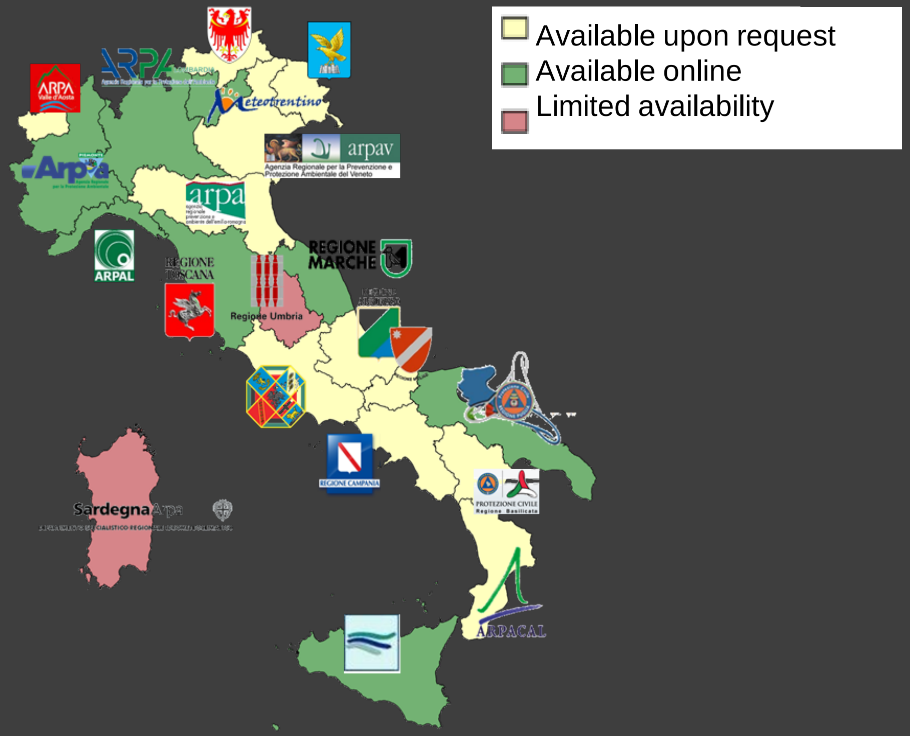
Data fragmentation

DEALING WITH SHORT AND FRAGMENTED RECORDS

Rainfall time series are often plagued with missing values creating sporadic and/or continuous gaps in their records. The fragmented behaviour traces back to the activation and dismissal of rain gauges, attributable to station relocation, service interruptions, replacement/renewal of the sensor, changes in the ownership of the station, etc.

The characteristics of the stations (location and elevation, type of sensor, etc.) may also change before and after the interruptions, with consequent problems in attributing the data to a unique homogeneous sample. Despite these problems are quite common, even in developed countries, many practical applications and statistical methodologies have little or no tolerance to missing values.

→ EXAMPLE: The Italian framework



National Hydrographic and Mareographic Service (SIMN) → ~1990 → 19 regional agencies
2 provincial agencies

→ EXAMPLE: The Italian framework

Annual maximum rainfall depths for 1-3-6-12-24 hours durations

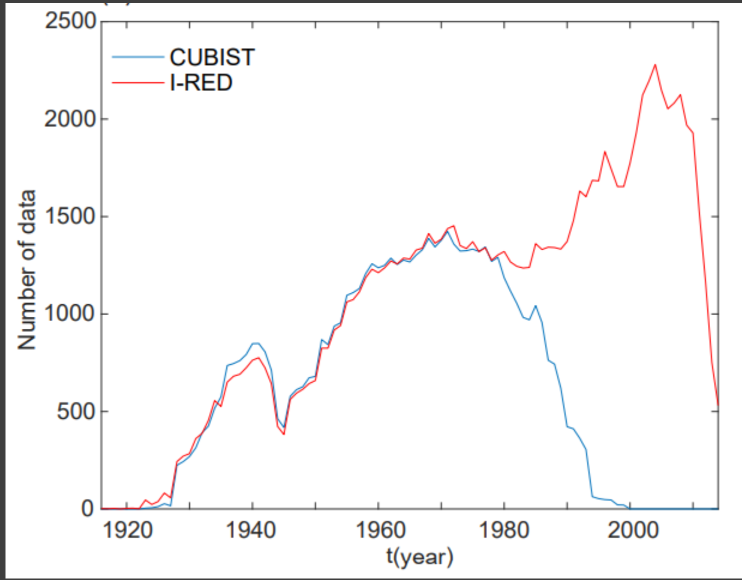
2748 stations

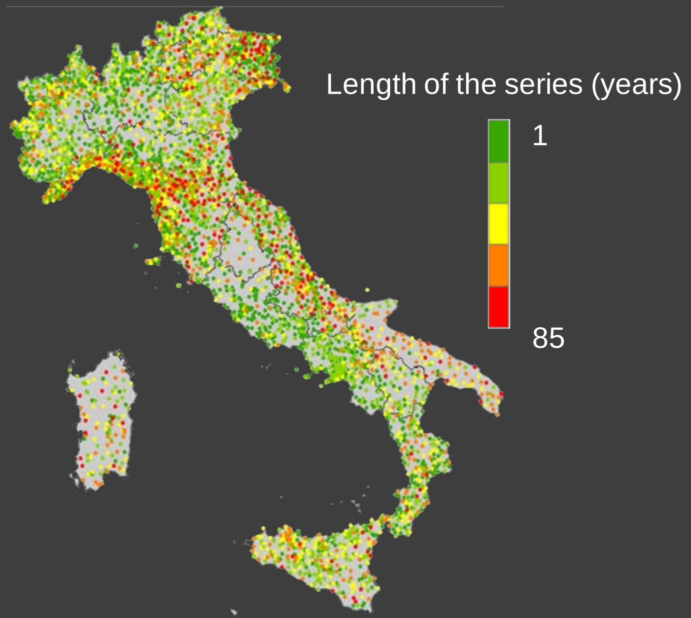
from 1916 to 2000

4688 stations

from 1916 to 2015

CUBIST [3]

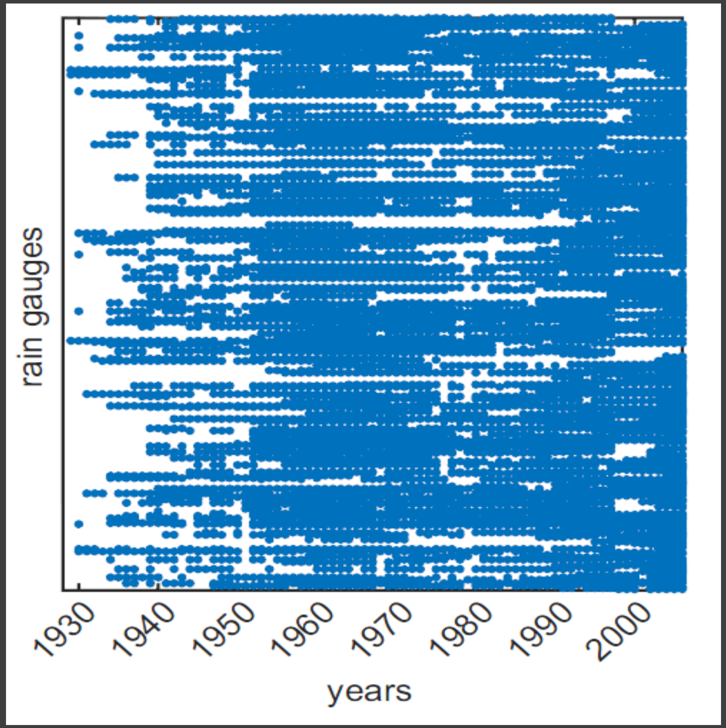




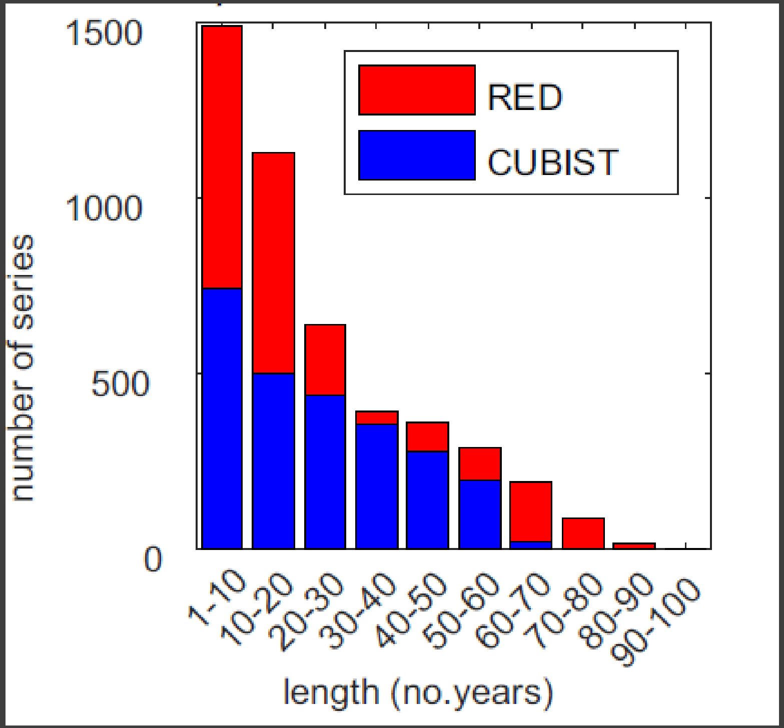
I-RED [4]

→ EXAMPLE: The Italian framework

Data availability in time in the Piemonte regional dataset



Number of series per lenght class in the I-RED and CUBIST databases



Two kind of problems:

- Statistical robustness of the estimations from short series
- Significance of the "lost information"

→ MISSING DATA MECHANISMS[5]

- *MAR (Missing At Random)*
  Data for a given variable (e.g., Y) are said to be MAR if the probability of missing data on Y is unrelated to the value of Y, after accounting for other variables (X).

- *MCAR (Missing Completely At Random)*
  Data on Y are said to be MCAR if the probability of missing data on Y is unrelated to the value of Y or any values of other variables (X) in a data set.

- *MNAR (Missing Not At Random)*
  Data on Y are said to MNAR if the probability of missing data on Y is related to value of Y or any values of other variables in a data set

Rainfall series missing data can be often attributed to the second classes (MCAR): the number and temporal occurrence of gaps (missing) in precipitation data a site (i.e., rain gauge) are not dependent on the data at the site or any other sites.

[5] *Little and Rubin, 2002*

→ HANDLING MISSING DATA[6]

- *Omitting missing records (OMR)*
  Procedures based on only complete records.

- *Infilling missing records (IMR)*
  Missing records are infilled and resultant complete data are analyzed by standard methods.

- *Accomodating missing records (AMR)*
  Procedures that use the series containing missing records (without infilling).

→ Often not suitable, for the significant loss of information

→ Often complex, computationally demanding, and can lead to errors when based on non-robust assumptions.

→ For the robustness of the estimates usually a minimum length threshold on the number of valid data has to be set.

[6] *Teegavarapu, 2012*

→ INFILLING MISSING RECORDS

STRONG TEMPORAL DEPENDENCE
or
NORMALLY DISTRIBUTED

e.g., daily streamflow, total annual rainfall

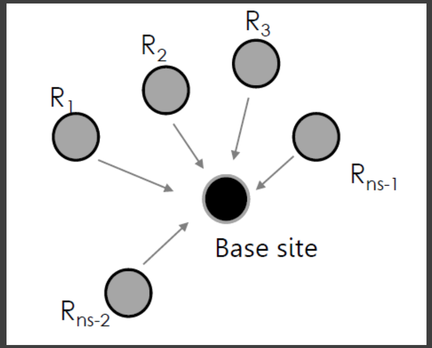
→ Temporal interpolation

When the gaps are frequent and systematic and when data show low auto-correlation in time, this approach is not effective.

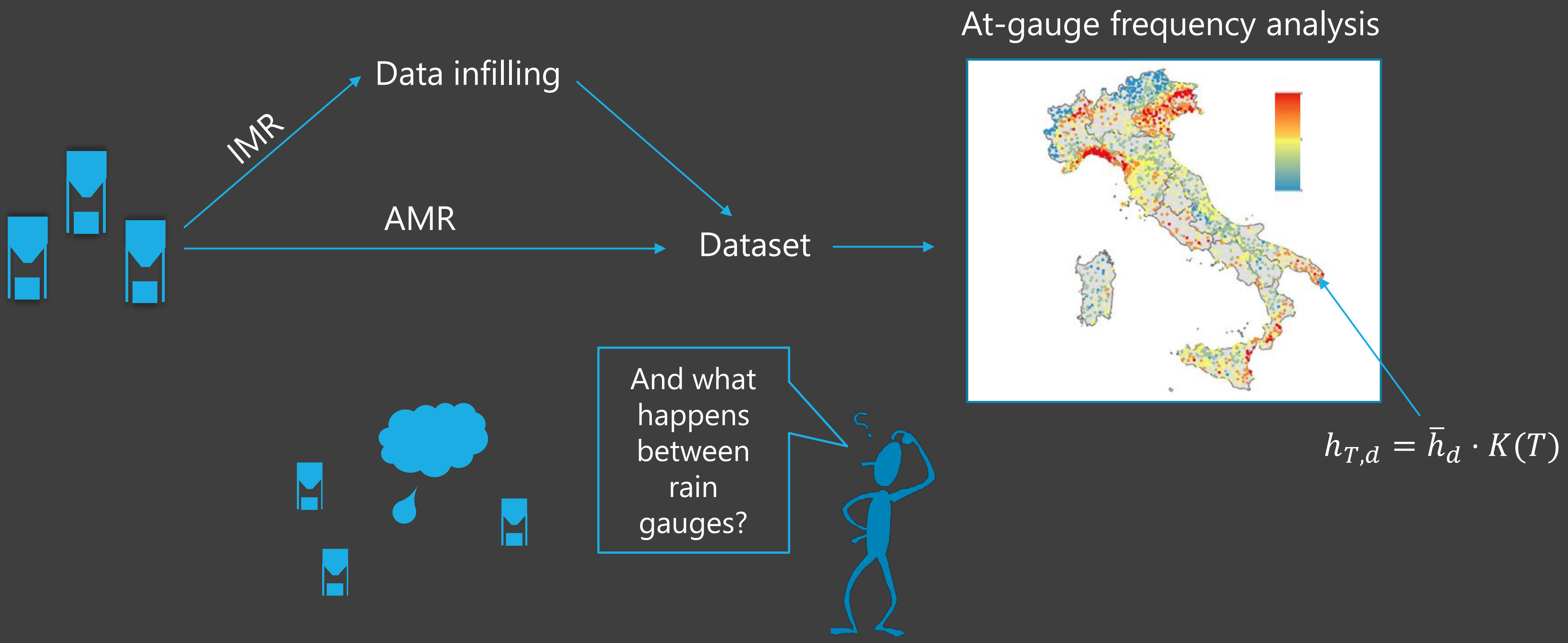
STRONG SPATIAL DEPENDENCE

e.g., precipitation data at multiple sites

→ Spatial interpolation



→ THE CHAIN OF RAINFALL FREQUENCY ANALYSIS WITH FRAGMENTED RECORDS



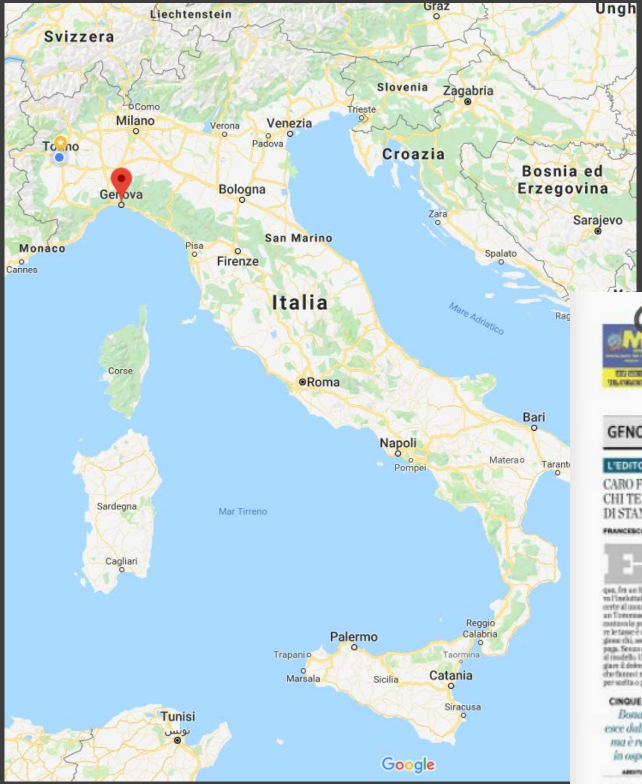
SPATIAL VARIABILITY OF RAINFALL FIELDS

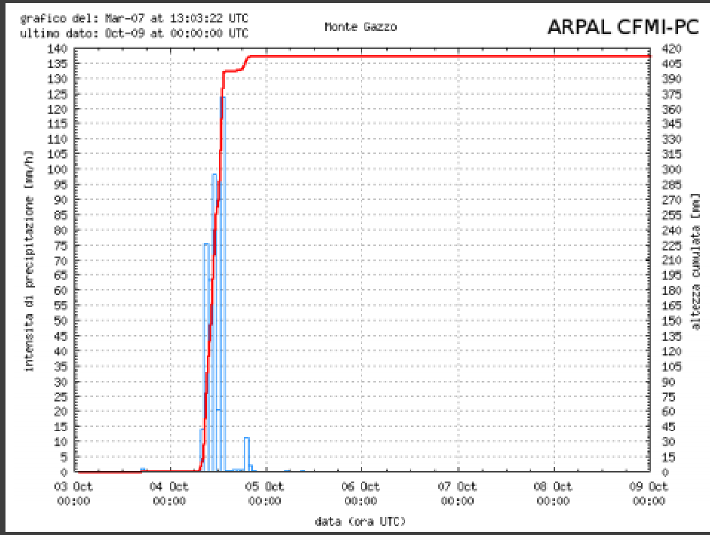
→ EXAMPLE: Self-regenerating Mesoscale Convective Systems (MCS)

GENOVA, 4 October 2010 [7]

Duration (h)	Rain Gauge	Rainfall (mm)	Return period (years)
1	Il Pero	140	>500
3	Monte Gazzo	243	>500
6	Monte Gazzo	396	>500
12	Monte Gazzo	411	>500
24	Monte Gazzo	411	200

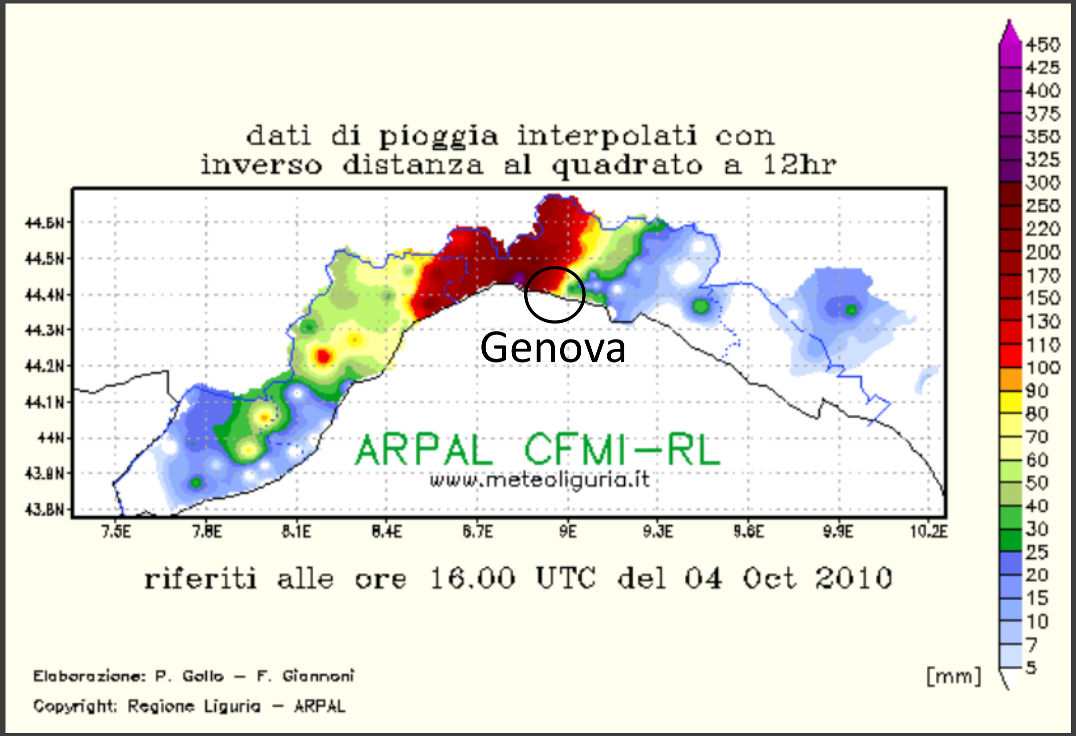
← *Maximum rainfall depths recorded during the event at the regional gauge network.*

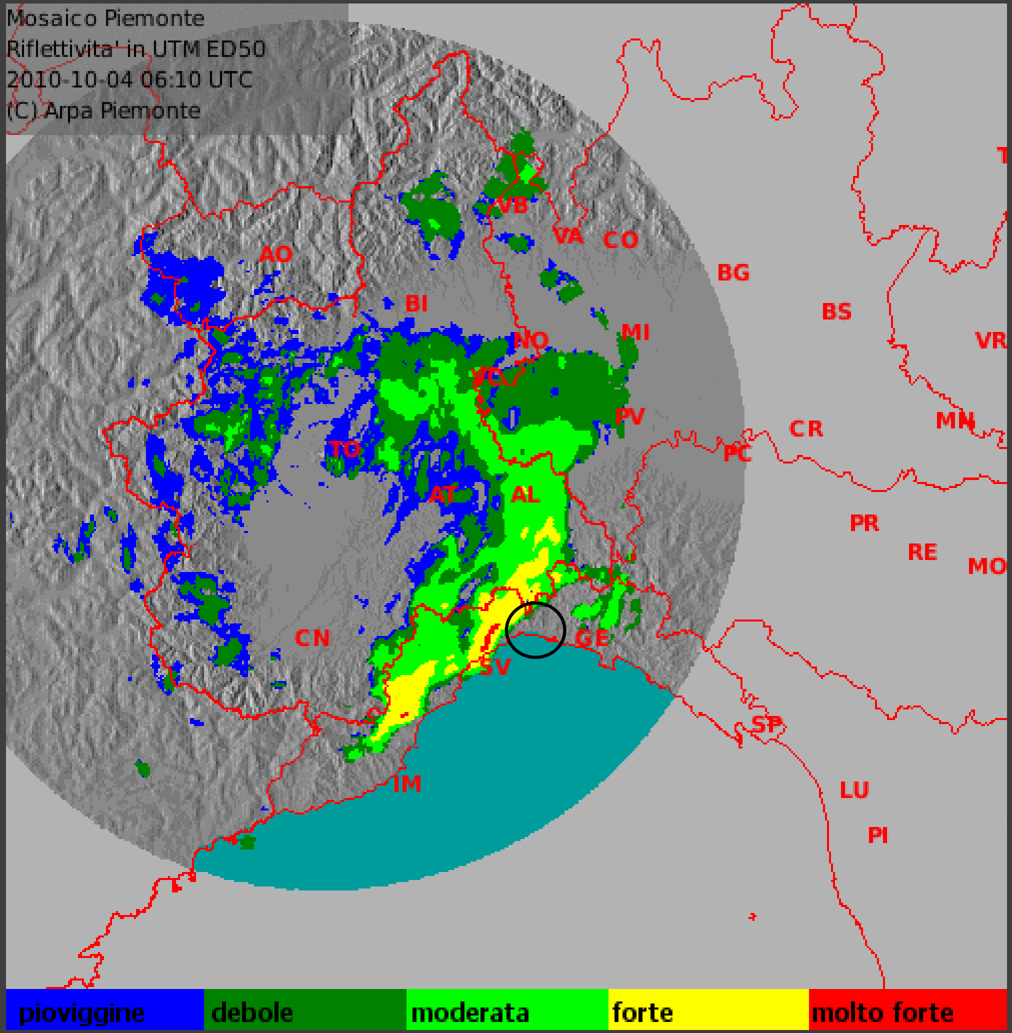




[7] Onorato et al., 2011

→ EXAMPLE: Self-regenerating Mesoscale Convective Systems (MCS)
GENOVA, 4 October 2010

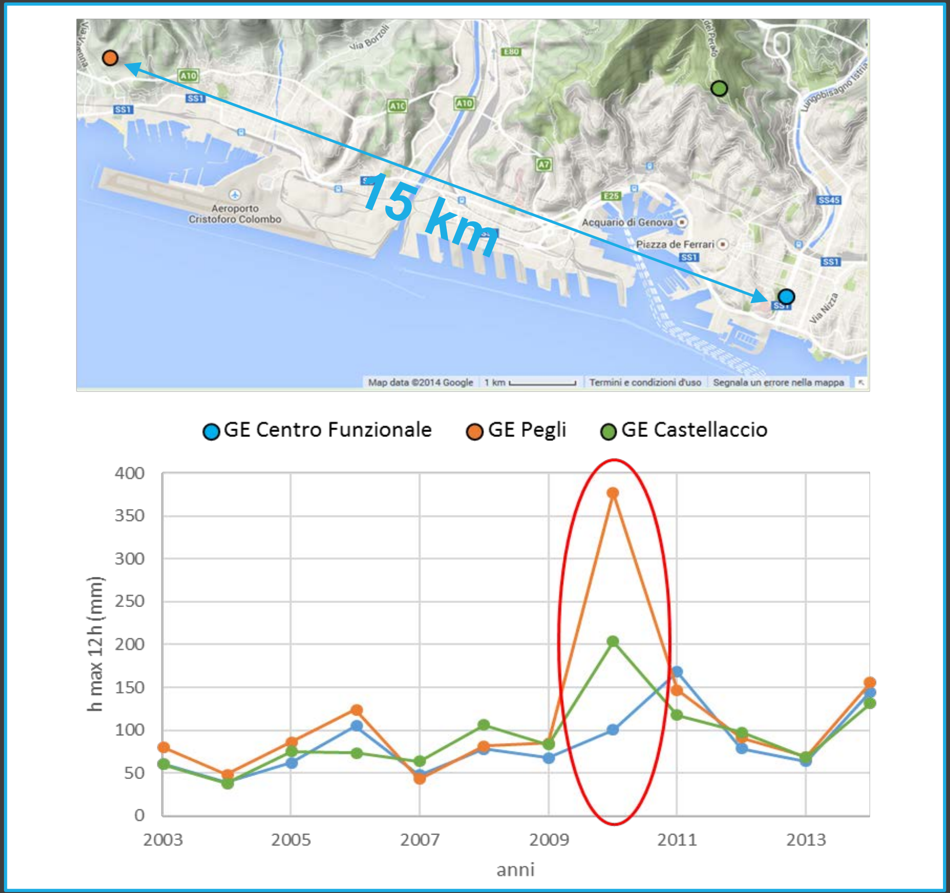




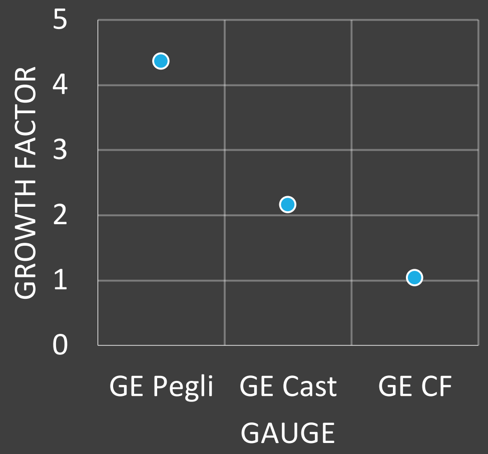
↑Cumulative rainfall depth from 4 to 16 UTC 04/10/2010 interpolated with IDW.
Rainfall estimated from the regional weather radars from 6 to 7 UTC 04/10/2010 →

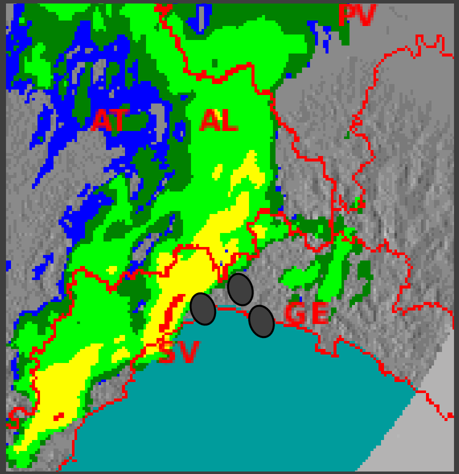
→ EXAMPLE: Self-regenerating Mesoscale Convective Systems (MCS)

<u>Assessing storm hazard at the urban scale</u>



Annual mixima for 12 hours duration.

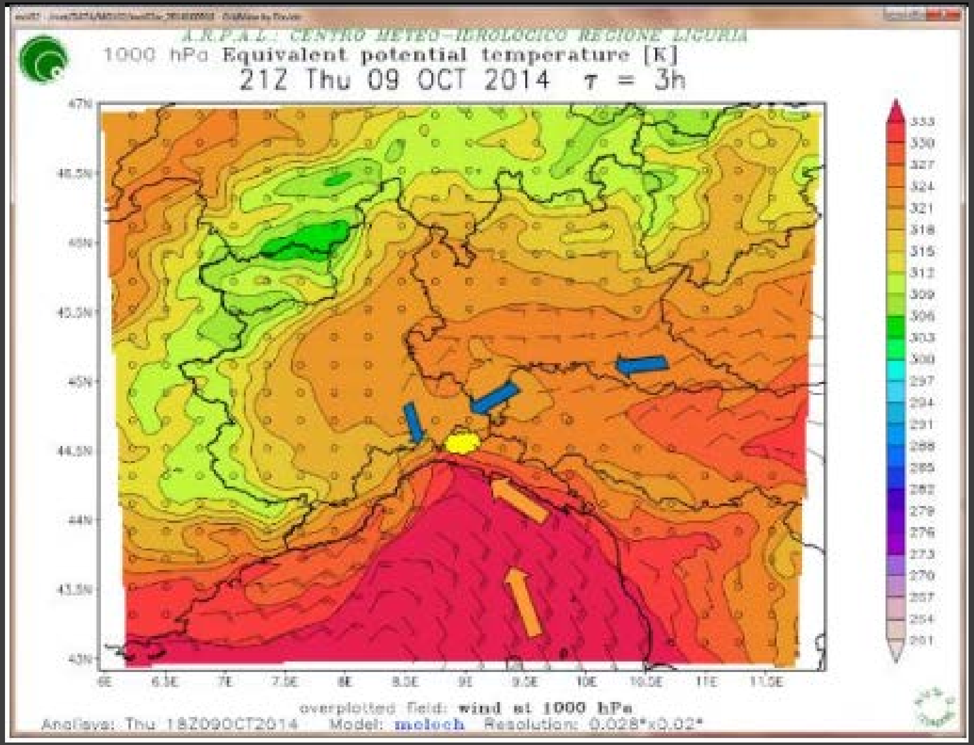




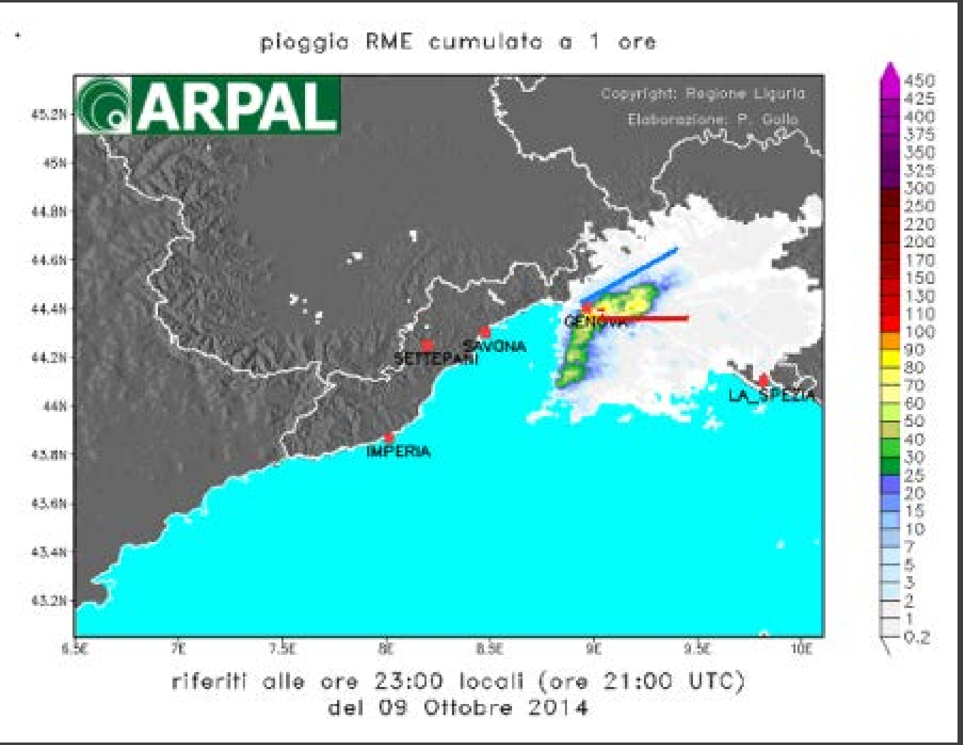
Is the western part of the city more prone to the development of extreme rainstorm than the eastern?

→ EXAMPLE: Self-regenerating Mesoscale Convective Systems (MCS)
The developement of MCSs in the Liguria sea

GENOVA, 9 October 2014 [8] - 21.00 UTC



Equivalent potential temperature at 1000 hPa

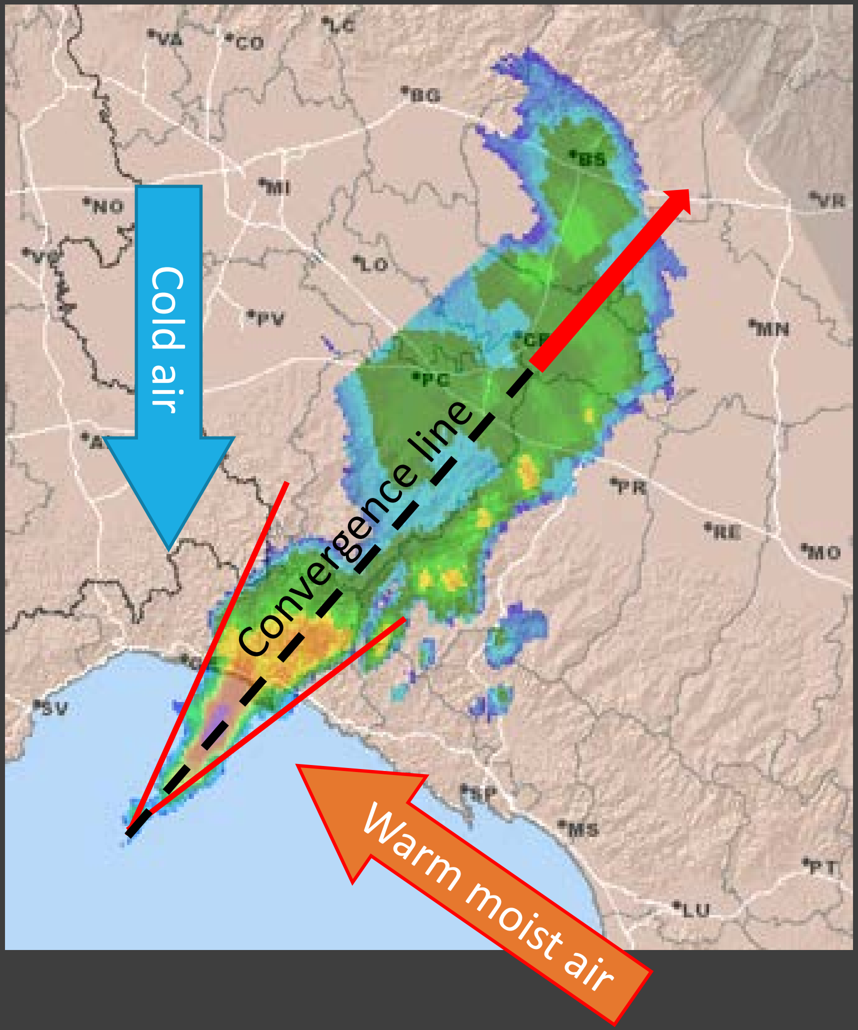


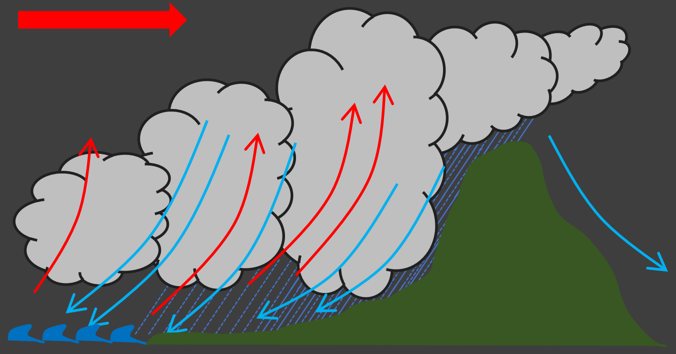
1-hour total radar estimated rainfall

[8] Onorato et al., 2014

→ EXAMPLE: Self-regenerating Mesoscale Convective Systems (MCS)

The developement of MCSs in the Liguria sea





MCS - Back-Building multicell storm

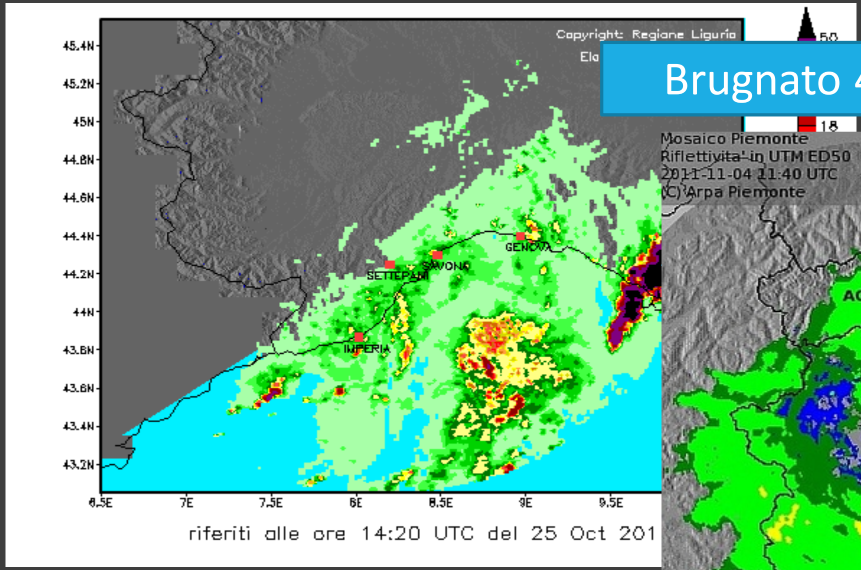
New convective cells continually regenerate at approximately the same rate at which the older ones are advected away. Regional radar and satellite imagery frequently reveal these stationary or backward regenerative systems that assume a characteristic V-shape [9]

[9] *Fiori et al., 2014*

→ EXAMPLE: Self-regenerating Mesoscale Convective Systems (MCS)

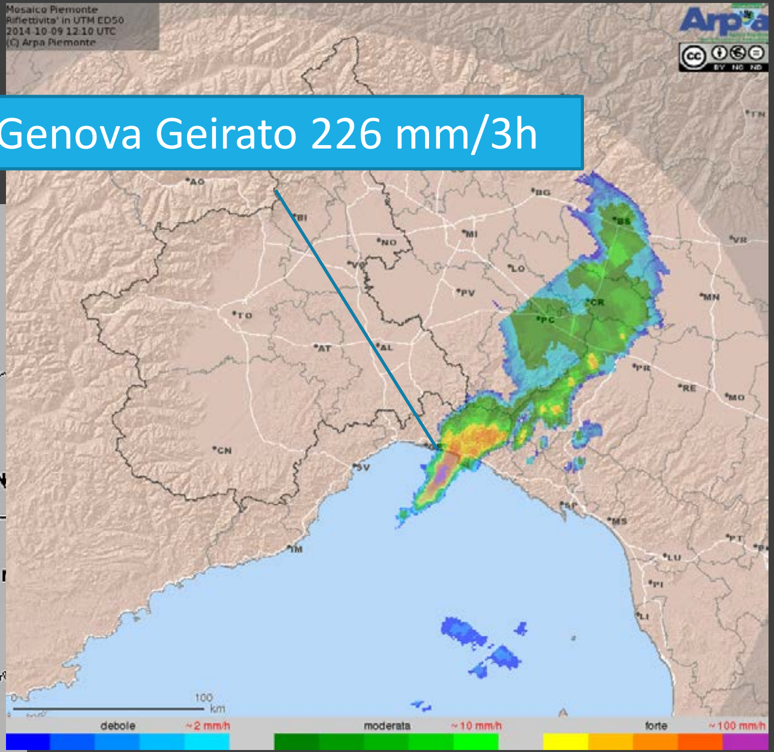
<u>From urban to regional hazard</u>

Cinque Terre (25/10/11)



Genova (4/11/2011)

Genova (9/10/2014)



HANDLING THE SPATIAL VARIABILITY

An *IDF* relation is basically valid only at the point where it is estimated. Rain gauges are generally not evenly distributed in space, and they allow only for a point estimation of the parameters of the rainfall distribution.

To extend estimates to ungauged locations, rainfall data are usually spatialized by:

- REGIONAL ESTIMATION, estimating the *IDF*s after pooling the available data within homogeneous areas defined by geographical boundaries, or centred around a location of interest [2].
- LOCAL ESTIMATION AND SPATIALIZATION, considering the distribution parameters estimated at the station locations and interpolating them in space with proper algorithms.

The choose of the best technique for spatial frequency analysis depends on different factors (spatial distribution of the network, length of the available series, aims of the analysis, etc.).

[2] *Hosking and Wallis, 1997*

→ CHOOSING THE BEST APPROACH

E.g., On the one hand the regional approach allows at increasing the available data by pooling up for the estimation of the growth curve, improving the robustness of the estimates for large return periods (the English manual FEH - Flood Estimation Handbook[10] suggests a station-year series with length $N>5T$)…

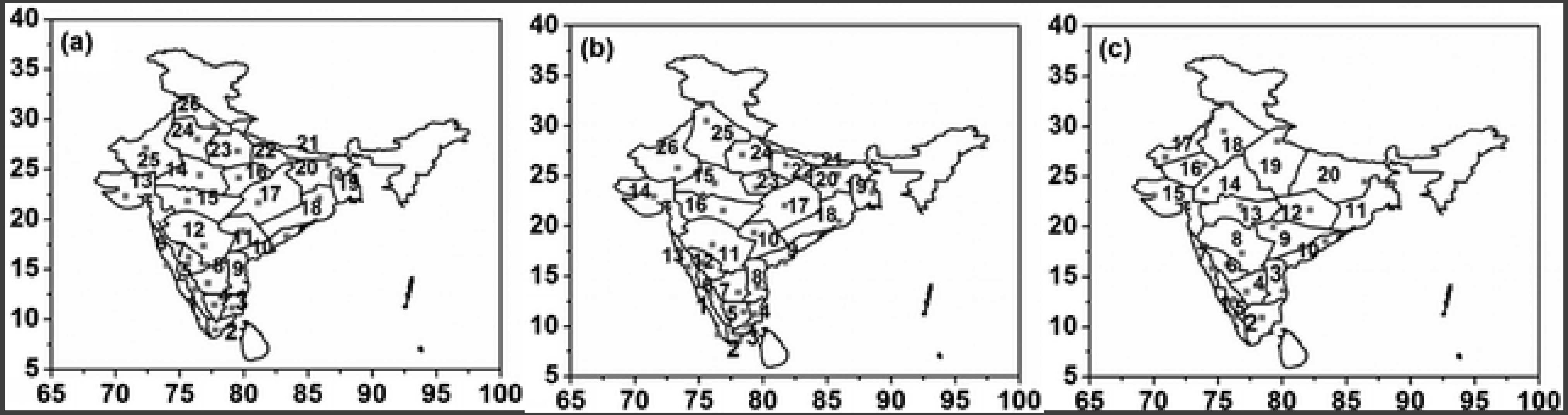
Length of record	Site analysis	Pooled analysis†	Shorthand description
< 14 years	No	Yes	Pooled analysis
14 to T years	For confirmation	Yes	Pooled analysis prevails
T to $2T$ years	Yes	Yes‡	Joint (site and pooled) analysis
> $2T$ years	Yes	For confirmation‡	Site analysis prevails

† Size of pooling-group chosen to provide $5T$ station-years of record
‡ Subject site excluded from pooled analysis

Reccomended methods for grouth curve estimation when T>27 years [10]

[10] *Reed et al., 1999*

…but on the other end the difficulties in identifying homogeneous regions and the arising of "border effects" due to the regional boundaries often leads to preferring a spatially-smooth approach.



Delineated homogeneous rainfall regions in India when (a) annual, (b) southwest, and (c) northeast monsoon rainfall are used for the correlation analysis.

[11] *Saikranthi et al., 1999*

→ REGIONAL FREQUENCY ANALYSIS

Under the hypotheses of:

- Hydrological homogeneity of the study region
- Independence of observed events**

Regional rainfall Frequency Analysis (*RFA*) enables one to **substitute space for time**.

- *RFA* improves the estimation accuracy for short samples

  If *M* annual sequences are available over a study area, for which the sample length is equal to *N1*, *N2*, … *NM* respectively then the size of the regional sample is equal to:

  $NReg. = N1 + N2 + \dots + NM$

- *RFA* enables one to predict *h(T)* in ungauged basins

** The 2nd hypothesis is often violated in practice, nevertheless the intersite correlation affects the variability of the regional estimator, but does not introduce bias[7]

[12] *Hosking and Wallis, 1988*

→ REGIONAL FREQUENCY ANALYSIS

Identification of homogeneous regions (pooling-group of sites) within which the flood frequency distribution is invariant except for a site-dependent scale factor termed index rainfall. Therefore:

$$h_{d,T} = \bar{h}_d \cdot K(T) = a \cdot d^n \cdot K(T)$$

$h_{d,T}$ T-year rainfall for duration d

$\bar{h}_d$ index-rainfall *scale factor (LOCAL)*

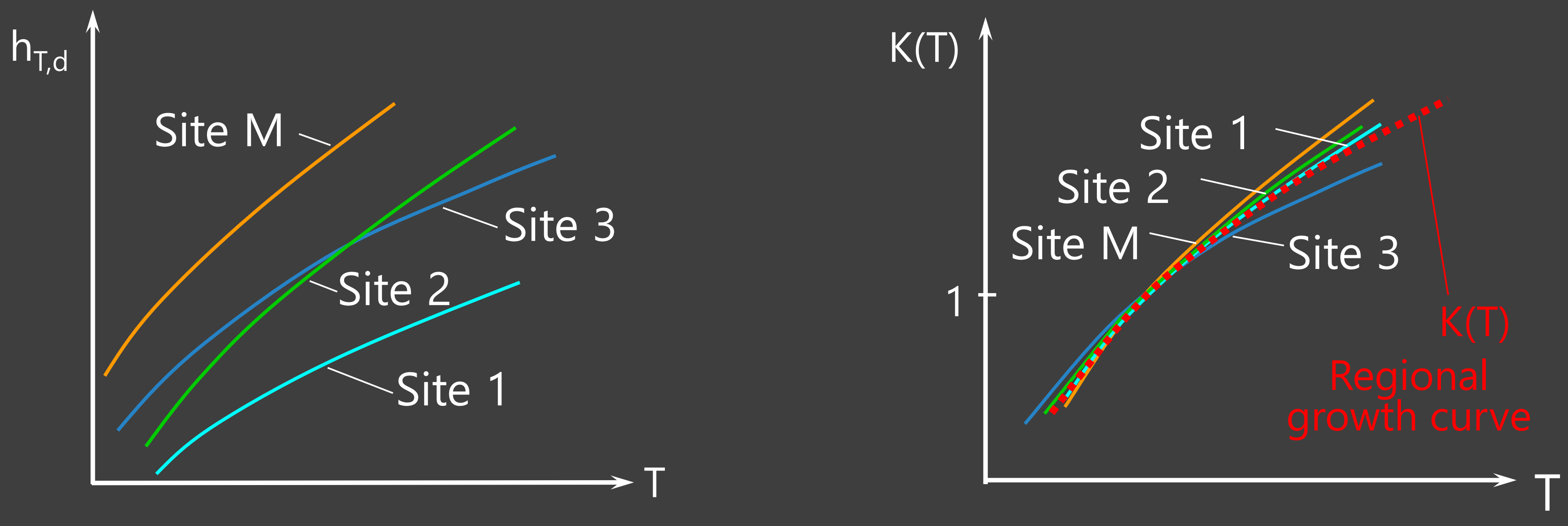
$K(T)$ dimensionless regional quantile *growth factor (REGIONAL)*

- 1st STAGE: Estimation of the regional growth curve K(T)
- 2nd STAGE: Estimation of the index-rainfall (gauged/ungauged sites)

→ REGIONAL FREQUENCY ANALYSIS

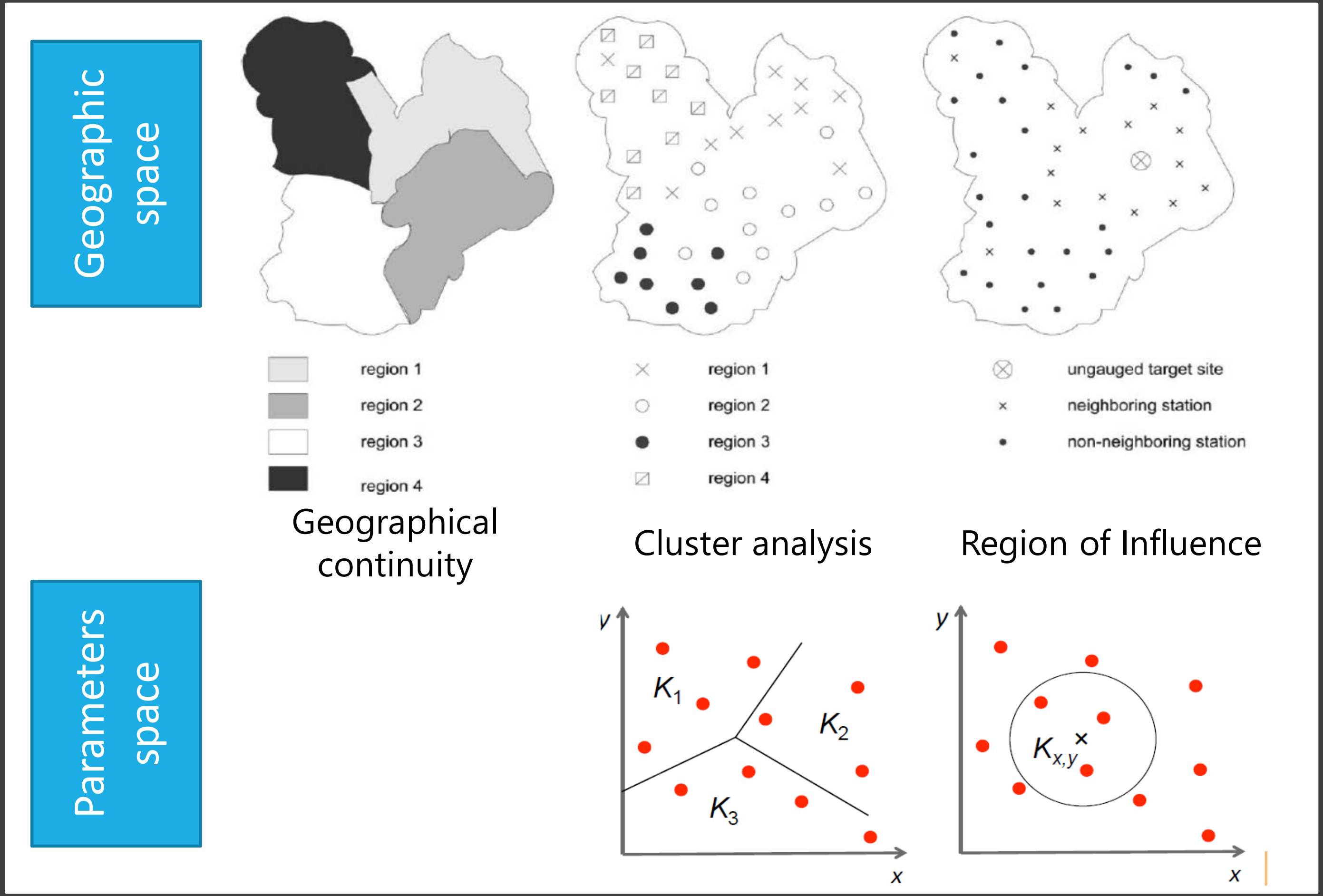
1st STAGE : Estimation of the regional growth curve

The traditional station-year method for regional frequency analysis involves pooling all the data into one long data series, and proceeding under the assumption that the observations at the stations are independent of each other. An extreme value distribution ("Regional growth curve") can then be fitted to this one long series.



→ REGIONAL FREQUENCY ANALYSIS

1st STAGE : Estimation of the regional growth factor



Approaches for the delineation of the homogeneous regions:

- Classical approach

  Fixed and geographically contiguous regions

- Focused pooling (Region of Influence - ROI)

  Pooling groups of sites identified on the basis of hydrological similarity with the target site[8]

[1] *Dalrymple, 1960*
[13] *Burn, 1990*

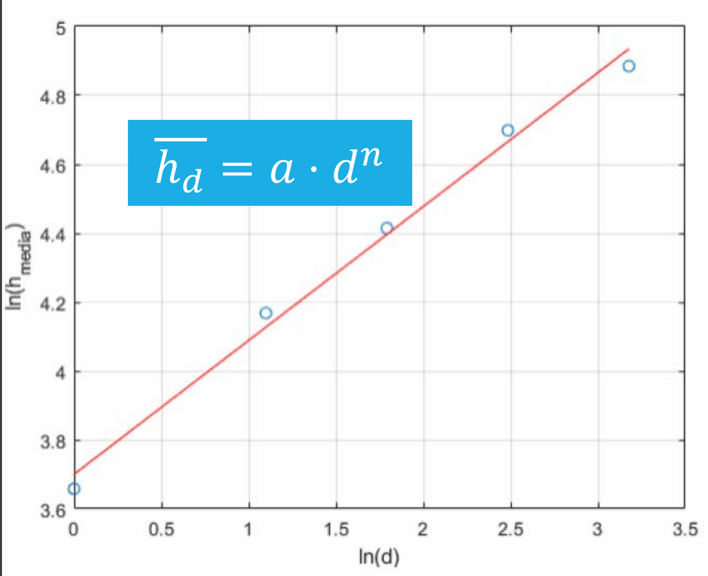
→ REGIONAL FREQUENCY ANALYSIS

2nd STAGE: Estimation of the site-dependent scale factor

Gauged target-site

Direct estimation + Log-interpolation

$$\overline{h_d} = \frac{1}{n}\sum_{i=1}^{n} h_{i,d}$$



Ungauged target-site

Indirect estimation

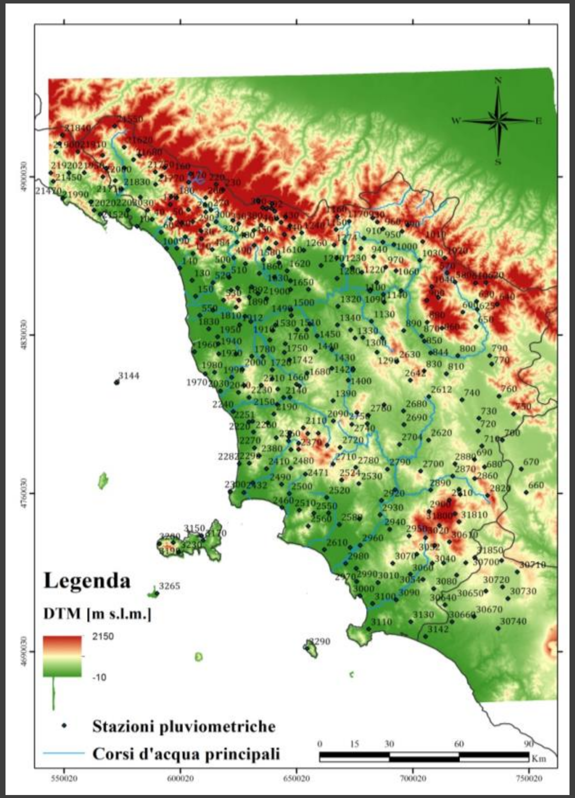
E.g., multi-regression model

$$\overline{h_d} = A_0 + A_1 \cdot \omega_1 + \cdots + A_n \cdot \omega_n + \epsilon$$

with

A_i parameters of the model
ω_i explanatory variables
ϵ model residuals

→ EXAMPLE: The Toscana region Regional Frequency Analysis[14]



Homogeneous regions identification → Preliminary hypotesis for the identification of the homogeneous regions based on the sample L-moments (L-skewnwss and L-CV)

Growth factor → TCEV Distribution
Gerarchical approach for parameters estimation based on 3 subsequential levels

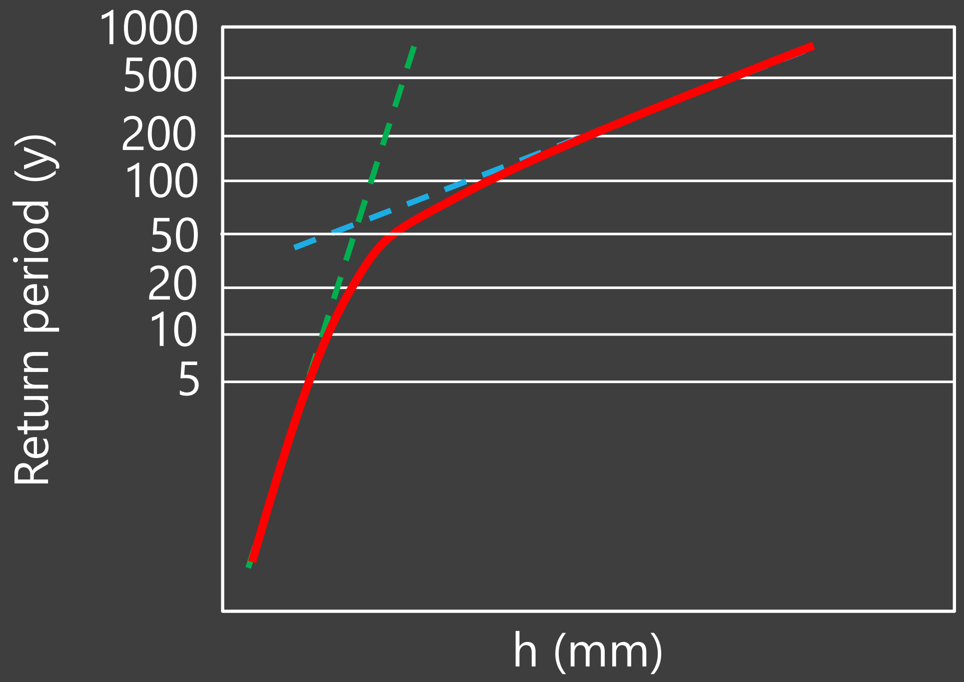
Index rainfall → For each region and each duration multi-parametric model considering morpho-climatic descriptors.

[14] *Caporali et al., 2008*

→ EXAMPLE: The Toscana region Regional Frequency Analysis
The TCEV Two Component Extreme Value Model[15]

The maximum rainfall / flows are generated by two different types (mechanisms) of events (Ordinary and Extraordinary), which generate annual maxima according to the Gumbel distribution.

$$F_x(x) = \exp[-\Lambda_1 e^{-\alpha_1 x}] \cdot \exp[-\Lambda_2 e^{-\alpha_2 x}]$$



[15] *Rossi et al., 1984*

→ EXAMPLE: The Toscana region Regional Frequency Analysis
The TCEV Two Component Extreme Value Model

Estimation of the growth curve

1st level: Estimation of the α_2e $\Lambda_2$parameters of the extraordinary component.

Can not be estimated from a single series, or even a few series of data. It is necessary to consider a very large area (region), Example: Italy, excluding the Po and the Alpine basins.

Indicative width 10^4 km^2

2nd level: Estimation of the Λ_1 parameter of the ordinary component.

A less extensive area (sub-region) homogeneous with respect to Λ_1 is sufficient for the estimation. Good estimate of Λ_1 can be also obtained from a sufficiently long (local) data set

Indicative width 10^3 km^2

3rd level: Estimation of the index value $\bar{h}_d$.

The annual average varies a lot for each location depending on the climatic and physiographic parameters characteristic of the location (also short series can provide a good estimate).

→ EXAMPLE: The Toscana region Regional Frequency Analysis

1st STAGE : Identification of the regions and sub-regions

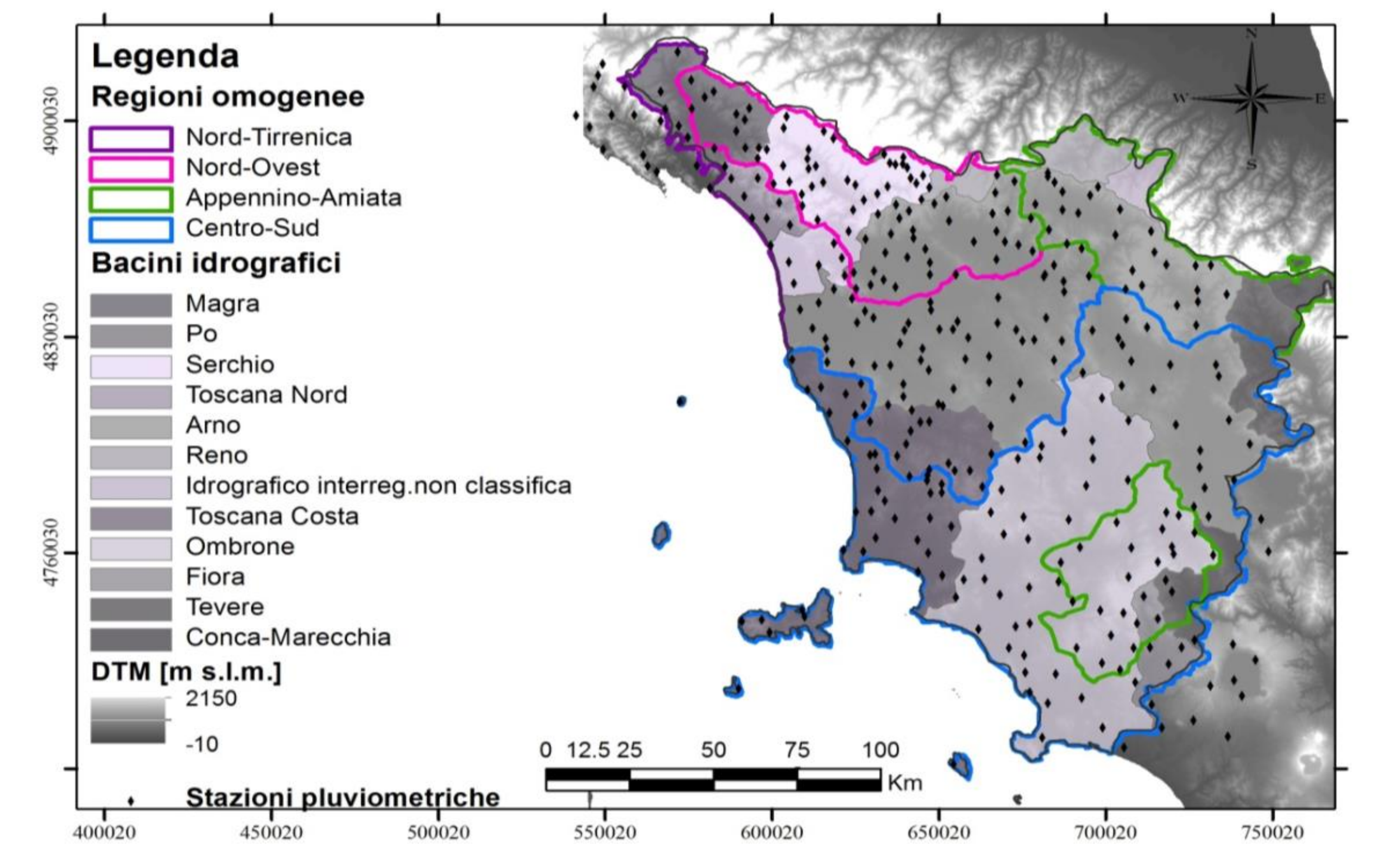
HYP 1	1 homogeneous region and 1 homogeneous subregion.
HYP 2	1 homogeneous region and 3 subregions
HYP 3	3 homogeneous region coincindent with the 3 subregions
HYP 4	4 homogeneous region coincindent with the 4 subregions

For all the hypothesis the TCEV parameters have been estimated and tested with :

- Differences between the mean and standard deviation calculated on the observed and theoretical series (Monte Carlo)
- Application of the Student's t and of the Wilcoxon's tests for the mean, $\chi 2$ test.
- Application of D-discordance and H-homogeneity tests
- Graphical comparison of the empirical growth curve of the observed with the theoretical one of the TCEV model on the Gumbel probability paper.

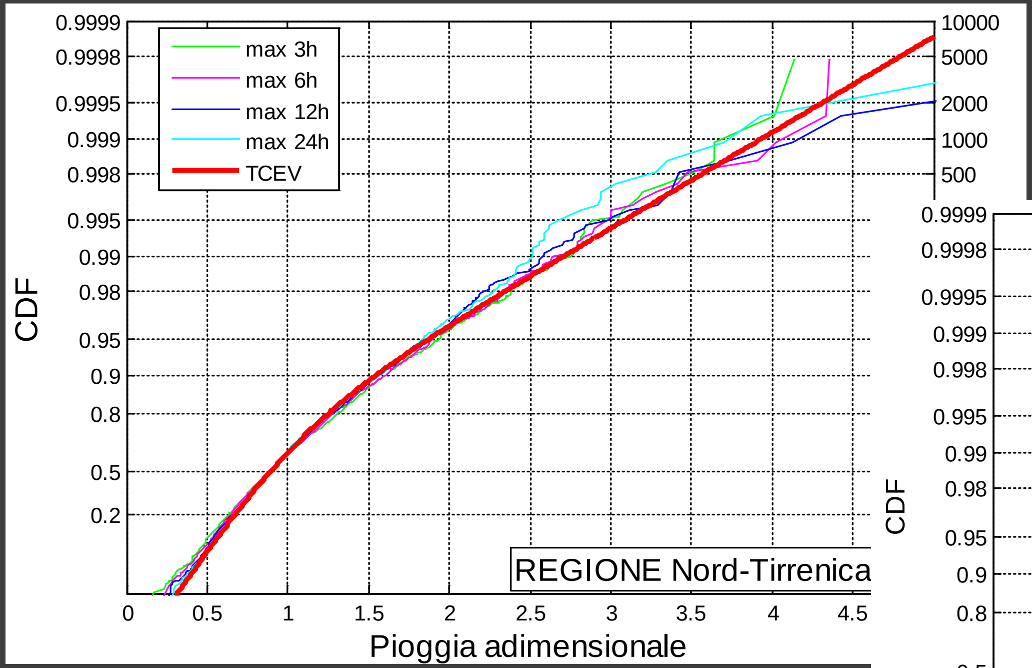
→ EXAMPLE: The Toscana region Regional Frequency Analysis

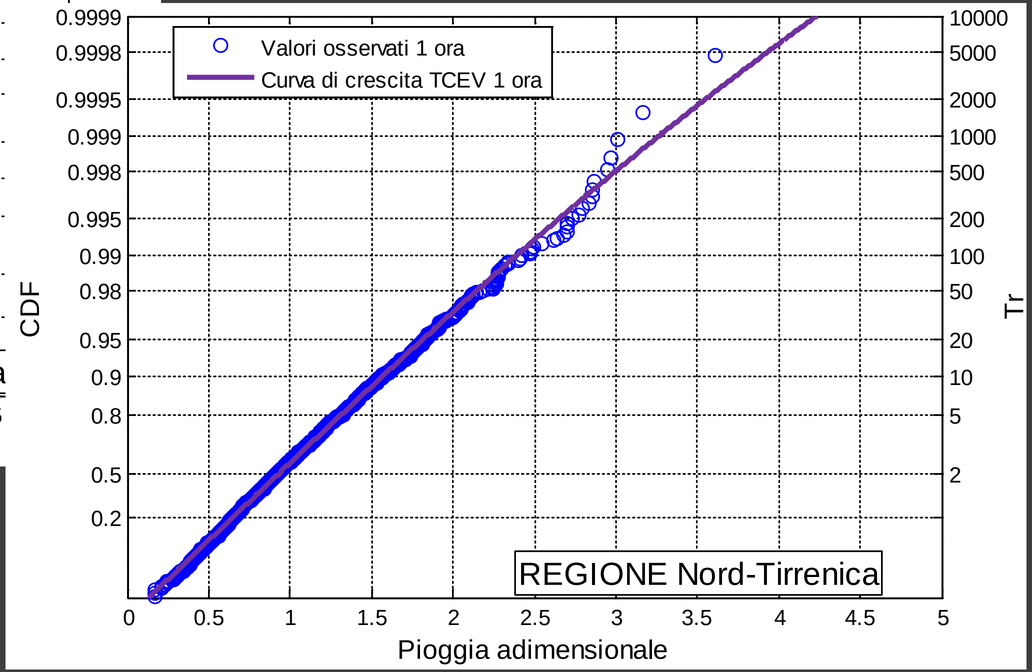
1st STAGE : Identification of the regions and sub-regions



Identified Homogeneous Regions

→ EXAMPLE: The Toscana region Regional Frequency Analysis
2nd STAGE : Growth curve estimation





Growth Curves for the Nord Tirrenica region. Growth curve for 1h duration is evaluated separately.

→ EXAMPLE: The Toscana region Regional Frequency Analysis
2nd STAGE : Growth curve estimation

TCEV parameters for the homogeneous regions.

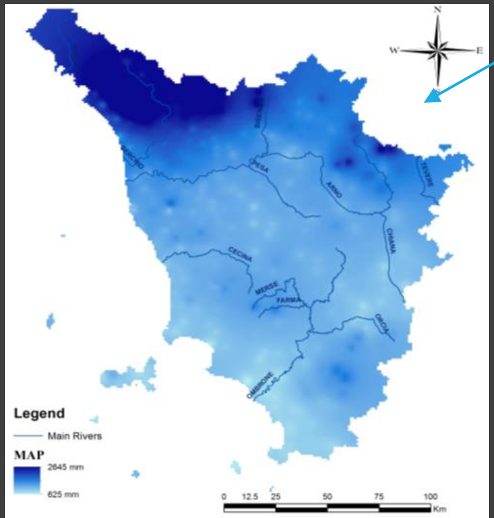
Regioni	θ^*	Λ^*	Λ_1	η	K_T	Note
Nord-Tirrenica	1.533	0.075	10.840	3.061	-0.5217+0.501·Ln T	Valida per d=1 ora
	2.634	0.438	31.195	4.937	0.2558+0.533·Ln T	Valida per d≥ 3 ore ed 1 g
Nord-Ovest	2.347	0.077	15.956	3.503	-0.9315+0.670·Ln T	Valida per d=1 ora
	2.600	0.176	22.755	4.091	-0.3397+0.636·Ln T	Valida per 3 ore≤d≤24 ore
	2.129	0.129	19.232	3.769	-0.3705+0.565·Ln T	Valida per 1 giorno
Appennino-Amiata	1.010	0.027	22.078	3.698	-0.1529+0.273·Ln T	Valida per 1 ora≤d≤12 ore
	2.456	0.127	33.292	4.350	-0.3605+0.565·Ln T	Valida per d=24 ore ed 1 g
Centro-Sud	1.844	0.100	13.686	3.342	-0.4901+0.552·Ln T	Valida per d=1 ora
	2.481	0.718	24.020	5.086	0.4634+0.488·Ln T	Valida per d=3 ora
	3.381	0.206	28.325	4.516	-0.4421+0.749·Ln T	Valida per d≥ 6 ore ed 1 g

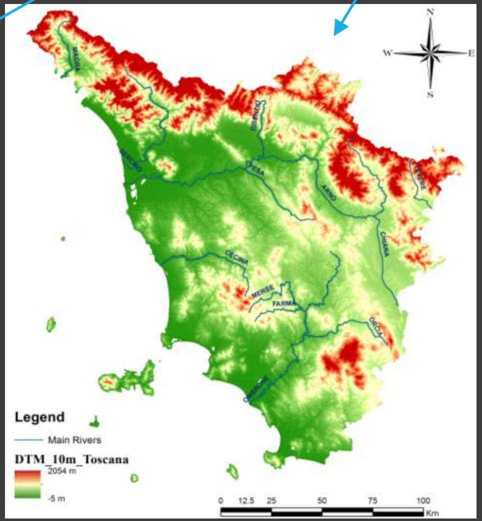
→ EXAMPLE: The Toscana region Regional Frequency Analysis

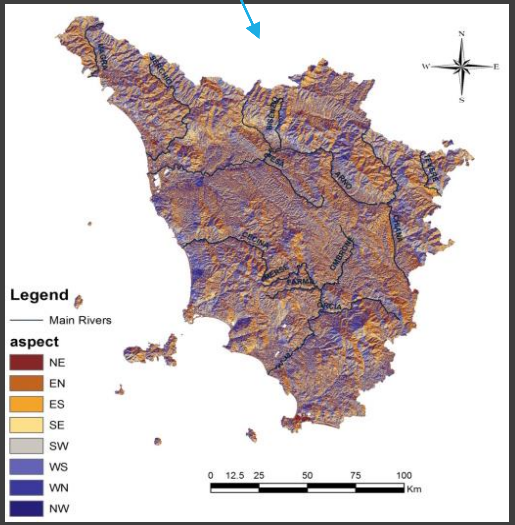
3rd STAGE : Index rainfall evaluation

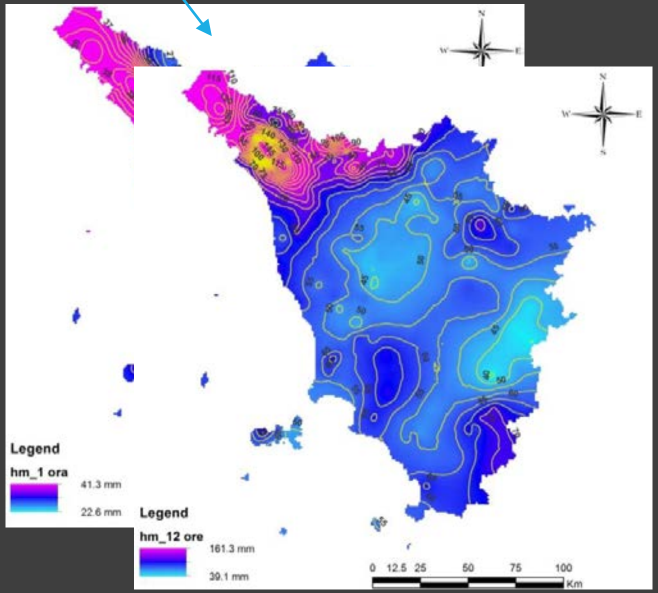
The index rainfall is estimated throughout the territory of the Toscana Region. For each homogeneous region and for each duration of rain a multivariate model is used according to the expression of *Caporali et al. (2008)*:

$$\mu = a_0 + a_1 \cdot \ln(MAP) + a_2 \cdot z + a_3 \cdot \left[\sin\left(\frac{Asp}{2} - \frac{\pi}{2}\right) + \pi\right] \cdot |Asp| + a_4 hm$$



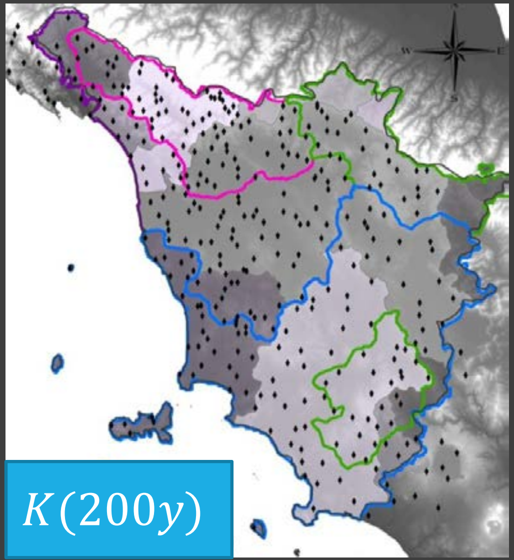




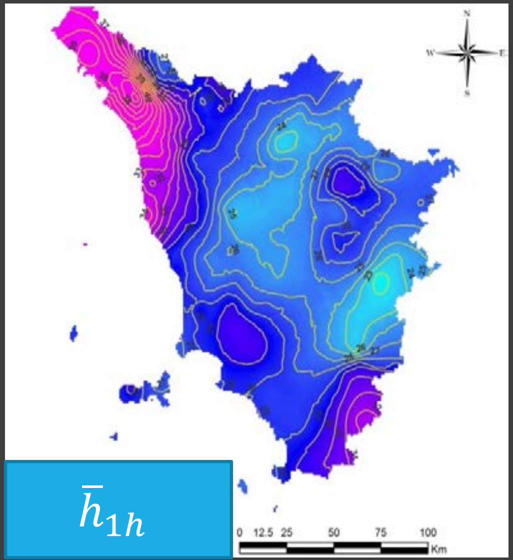


→ EXAMPLE: The Toscana region Regional Frequency Analysis

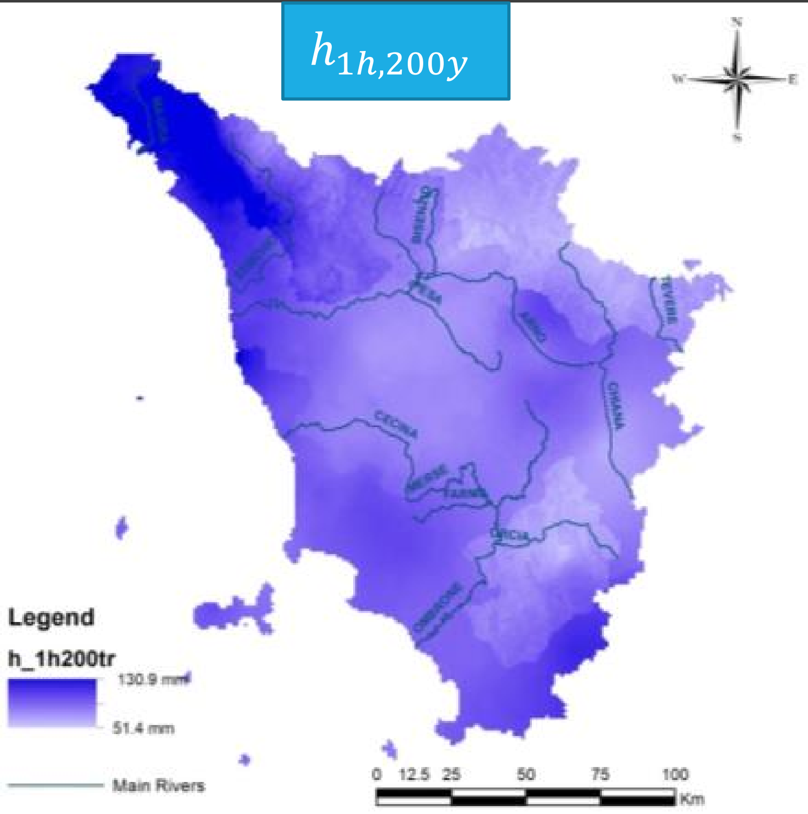
<u>Results</u>



X



=



Design rainfall with T=200 years for the durations1 hour.

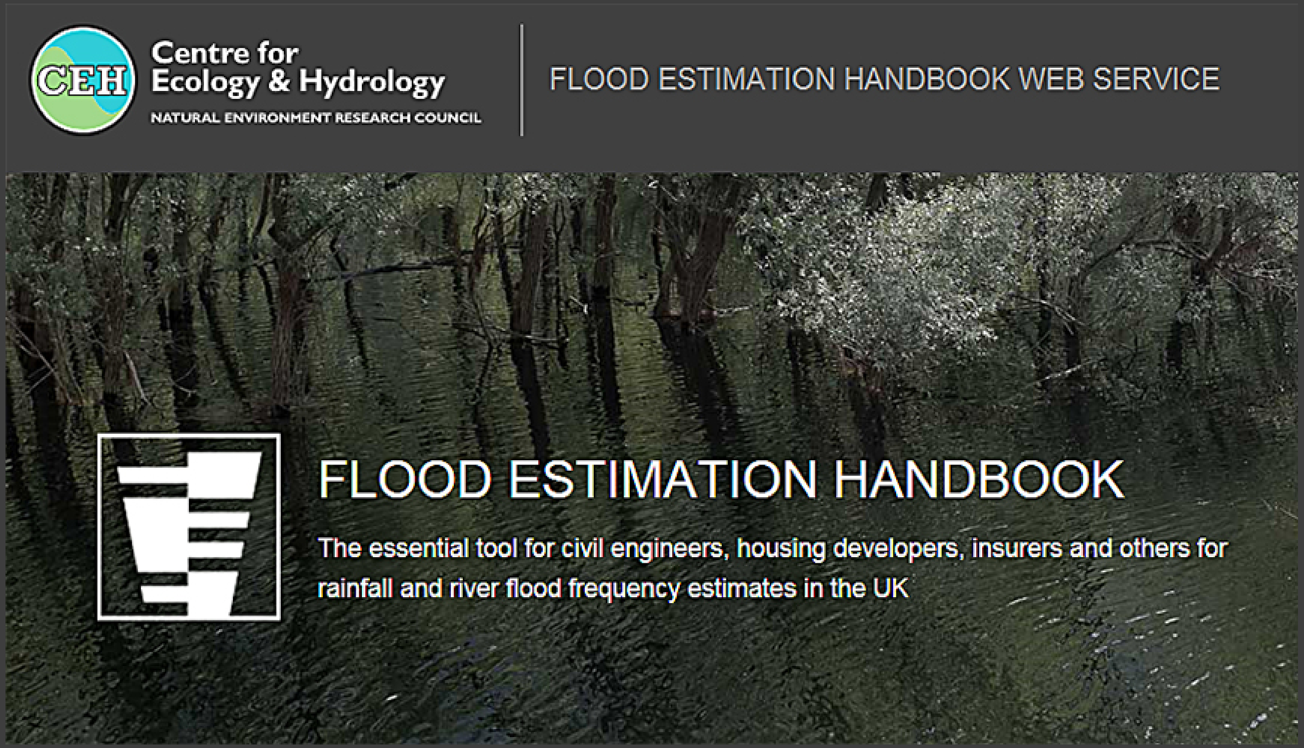
- Robust estimations

- Difficulties in the identification of the homogeneous regions
- Border effects

→ EXAMPLE: The FORGEX method[16]

Towards a «more spatially-smoothed approach» to regional analysis.

The rainfall frequency estimation method of the Flood Estimation Handbook (FEH) uses the FORGEX method for rainfall frequency estimation followed by the fitting of a depth-duration-frequency (DDF) model.



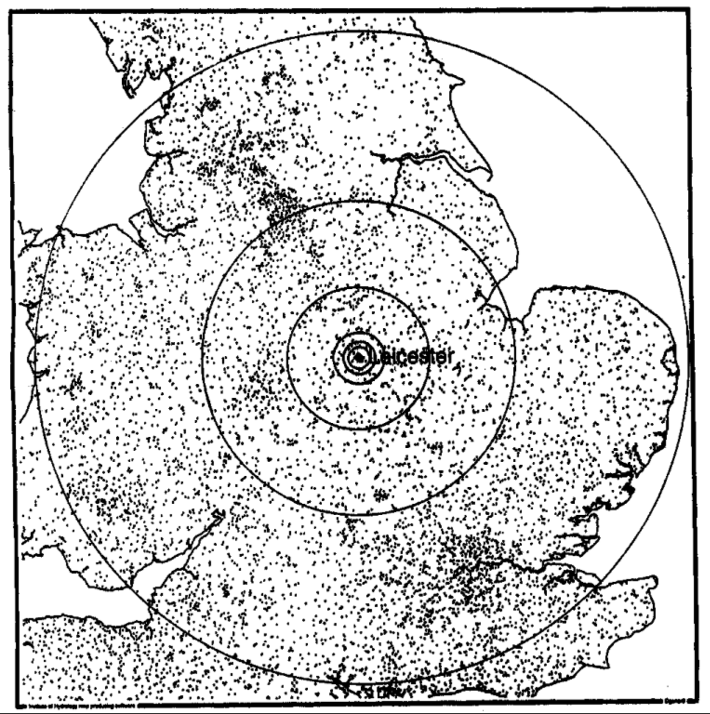
16 *Reed et al., 1999*

→ EXAMPLE: The FORGEX method

The methodology

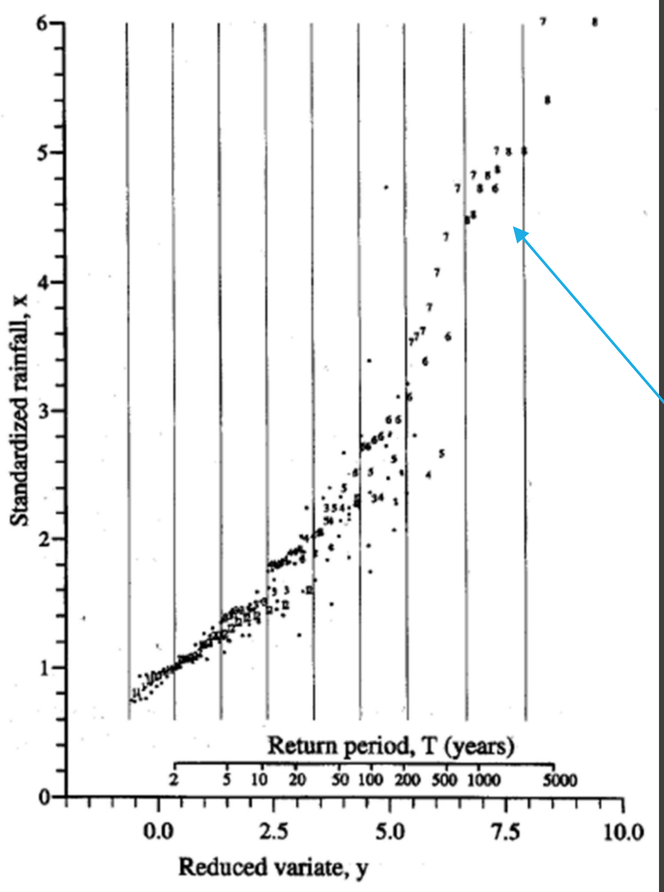
FORGEX

- INDEX RAINFALL: is the median AM rainfall at the site, spatialized with georegression on topographical and other variables.
- GROWTH CURVE: data are pooled from a hierarchy of expanding circular regions centred on the point of interest. Data from smaller networks are used to estimate the growth curve for short return periods and data from the larger networks are used for the longer return periods.



→ EXAMPLE: The FORGEX method

The methodology



The growth curve is plotted on a “sliced” y-x space. Each section, or y-slice, has width 1.0 on the Gumbel reduced variate scale. Data points from within the jth network are only plotted if their plotting position falls within the jth section of the growth curve.

Two kind of series are considered:

- Standardised values from individual stations.
- Network maximum (*netmax*) series, defined as the AM series of the largest standardized value recorded anywhere within the region.

Because of spatial dependence in the network of rain gauges, the plotting positions for the *netmax* points have been modified using a spatial dependence model.

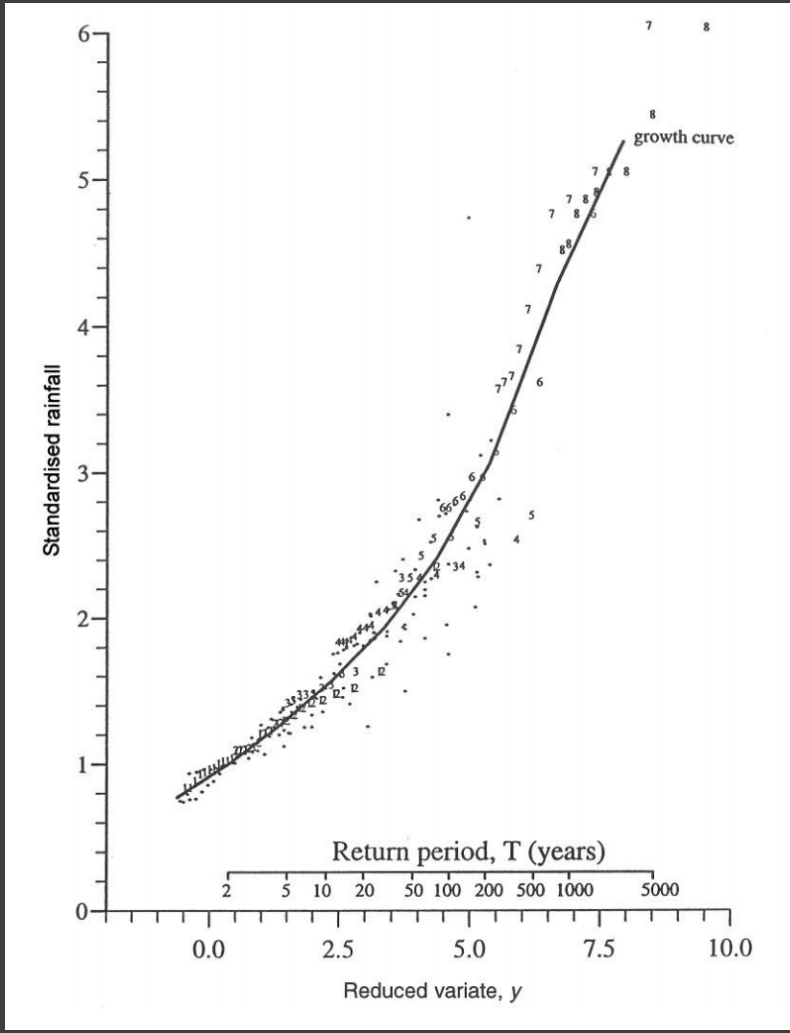
→ EXAMPLE: The FORGEX method

Results

For a given duration, an empirical growth curve consisting of concatenated linear segments is fitted to the plotted points (both individual and *netmax*) through a least squares routine.

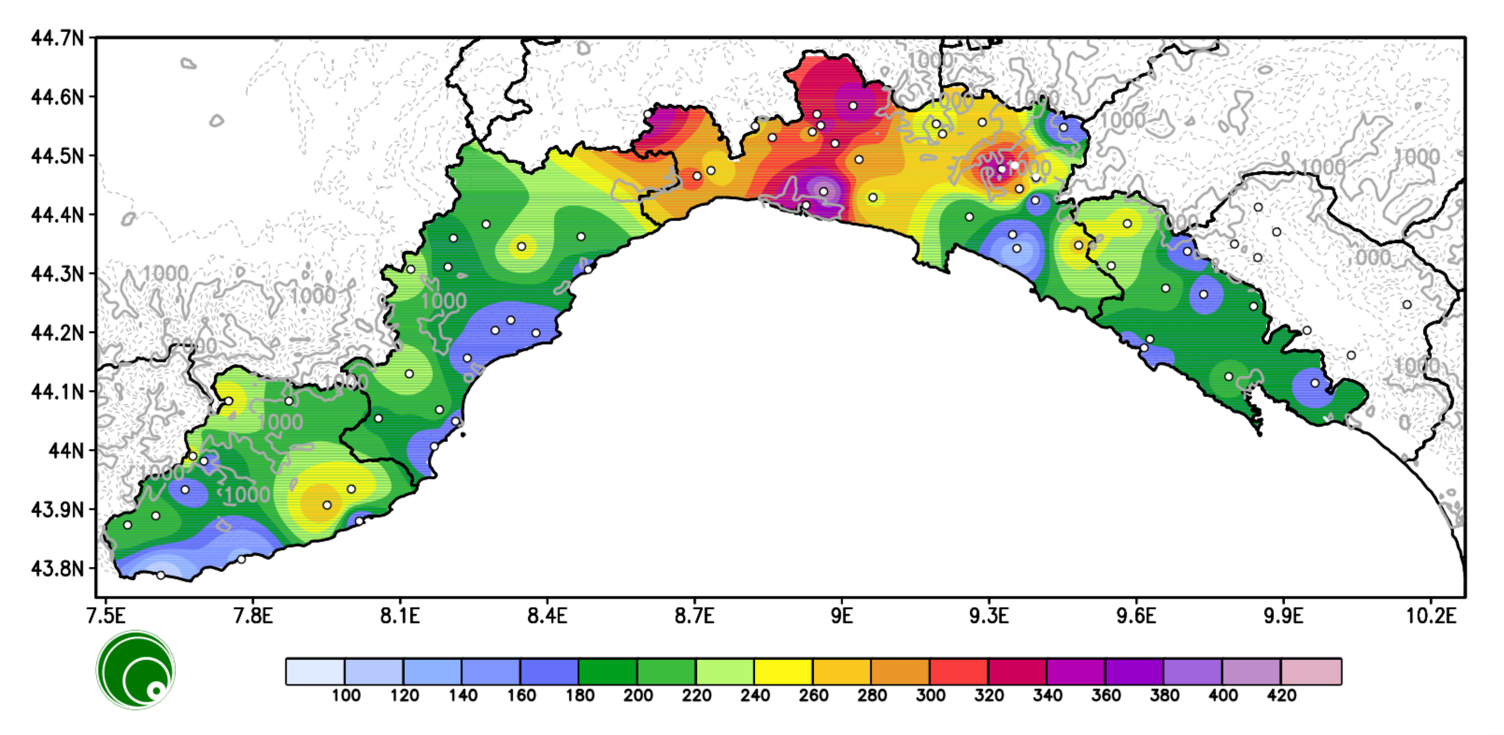
- Good use of local data and integration with regional information for large T
- No boundary problems

- The Gumbel distribution is not always the best choice
- The model does not allow spatial dependence to vary with return period



→ SPATIALIZATION OF LOCAL ESTIMATES

When local reliable estimations of the distribution parameters are available the spatialization to ungauged area can be carried out using spatial interpolation.



Design rainfall for T=50 years for the Liguria region. Quantile estimated at-site and interpolated with the Inverce Distance Weigth methodology[17]

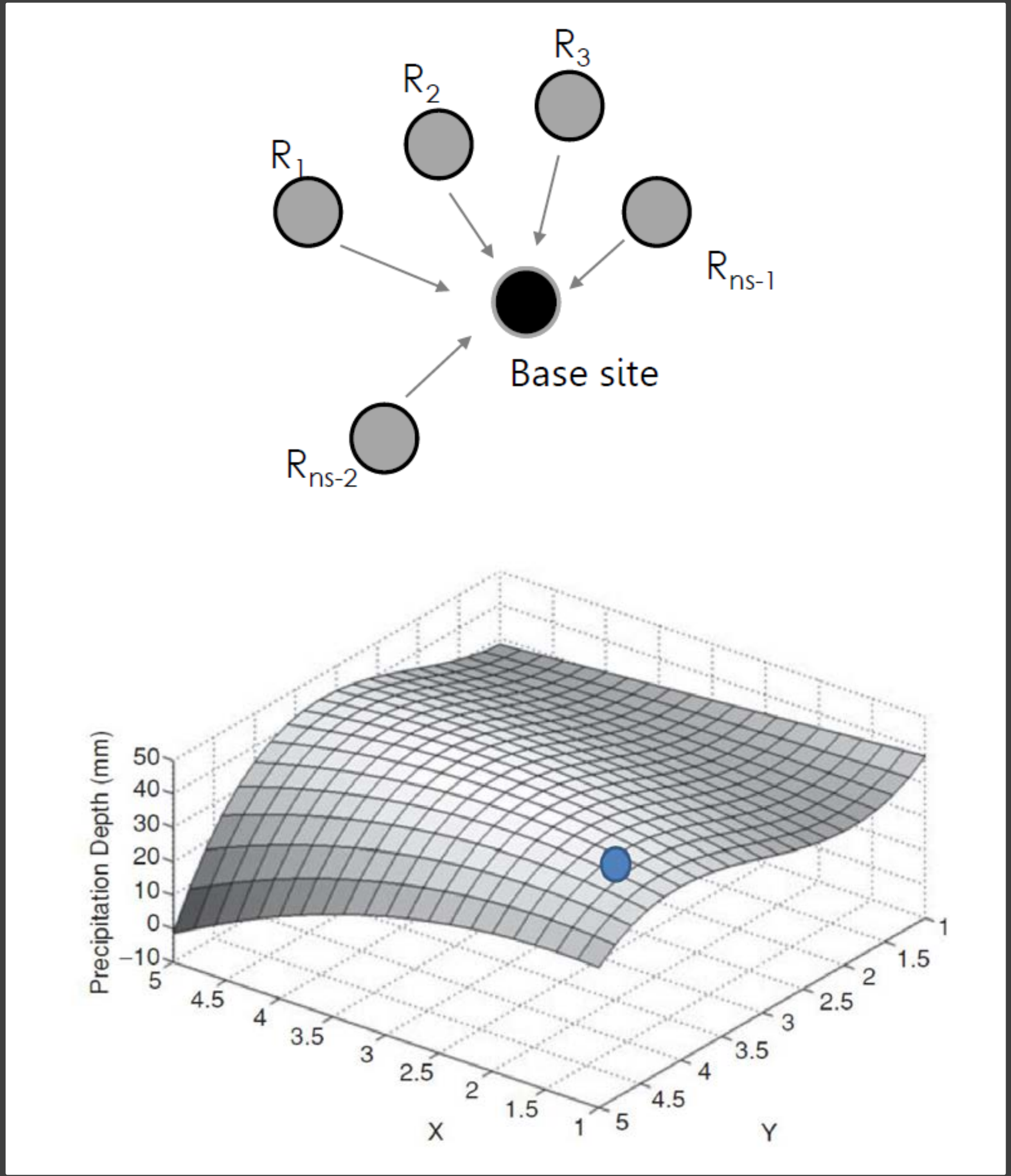
[17] *Agrillo and Bonati, 2013.*

→ SPATIAL INTERPOLATION

Estimation of missing data at a single site (base site) using available data at other observation sites (control points).

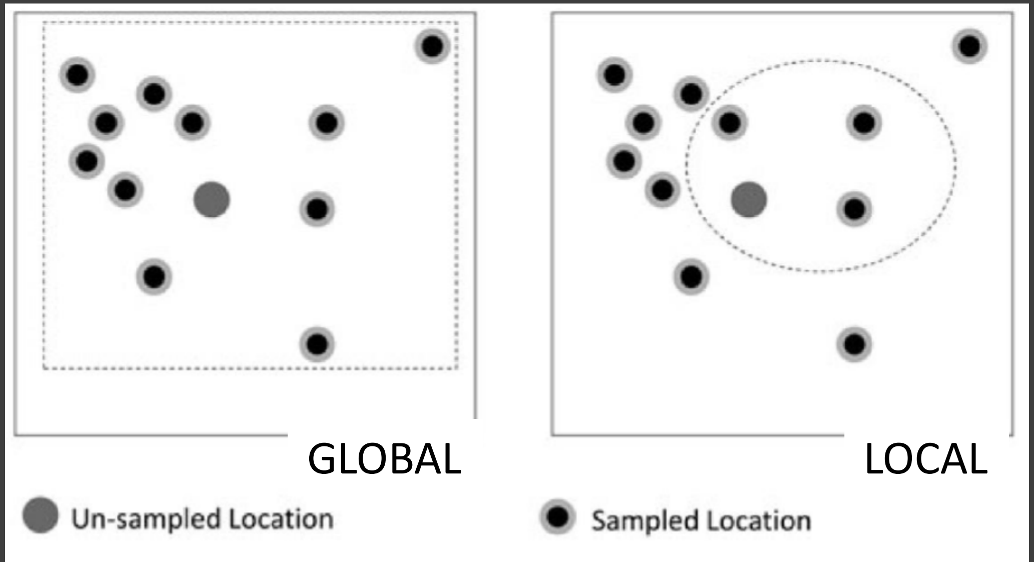
Different interpolation techniques are available. The best one depends on the characteristics of the network and of the records[5].

- Inverse Distance Weighting Method (also NWS method)
- Normal Ratio Method
- Quadrant Method
- Different forms of Kriging
- Trend surface Models
  - Local and Global
- Thin Plate Splines

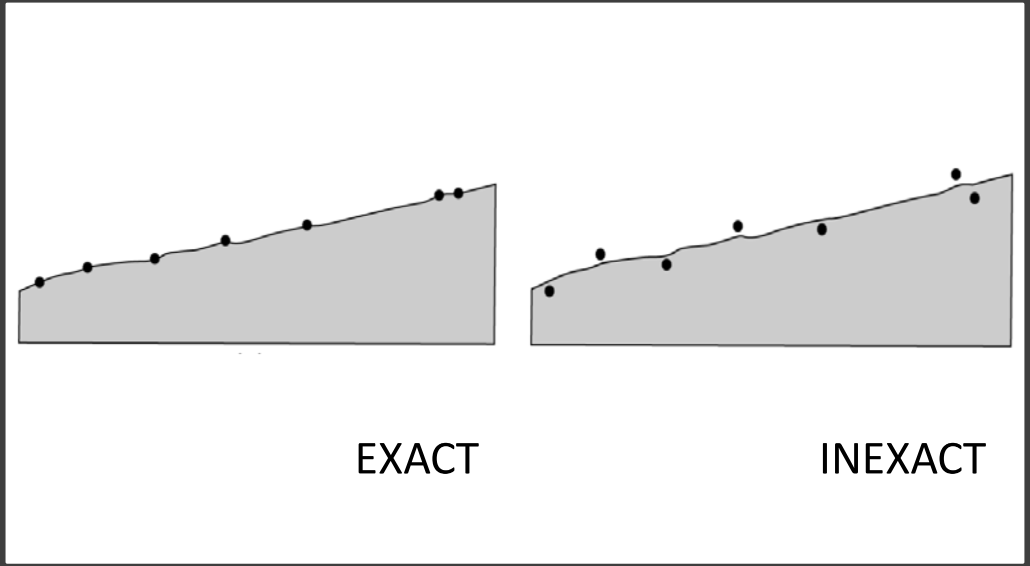


→ SPATIAL INTERPOLATION

Global VS Local Interpolation



Exact VS Inexact Interpolation



Deterministic VS Stochastic Interpolation[18]

Global		Local	
Deterministic	**Stochastic**	**Deterministic**	**Stochastic**
• Trend surface (*inexact*)	• Regression (*inexact*)	• Thiessen (*exact*) • Inverse Distance Weigthed (*exact*) • Splines (*exact*)	• Kriging (*exact*)

- Deterministic interpolation techniques create surfaces from measured points, based on either the extent of similarity or the degree of smoothing.

- Stochastic interpolation techniques utilize the statistical properties of the measured points, quantifying the spatial autocorrelation among measured points and accounting for the spatial configuration of the sample points around the prediction location.

[18] *Teegavarapu and Chandramouli, 2005.*

→ DETERMINISTIC INTERPOLATION: INVERSE DISTANCE WEIGHT

IDW is a deterministic (based on mathematical formulas) interpolation technique.
The assumption made for IDW is that the value of an attribute z at some unvisited point is a distance-weighted average of data points occurring within a neighborhood or window surrounding the unvisited point.

$$z(x) = \sum_{i=1}^{n} w_i z_i \qquad w_i = \frac{\frac{1}{d_{ik}}}{\sum_{i=1}^{n} \frac{1}{d_{ik}}}$$

The methods is simple and has a low computational cost. It assumes that nothing is known about the phenomenon being interpolated

STOCHASTIC INTERPOLATION: KRIGING[19]

Kriging is a stochastic interpolation method, based on the recognition that the spatial variation of any continuous attribute is often too irregular to be modelled by a simple mathematical function. The variation can be described better by a stochastic surface based on the relationships among the measured points.

Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. The method involves the fits to a mathematical function to a specified number of points, or all points within a specified radius, to determine the output value for each location.

Describe the spatial variation with variogram → Summarize the variation with a mathematical function → Use the function to determine interpolation weights

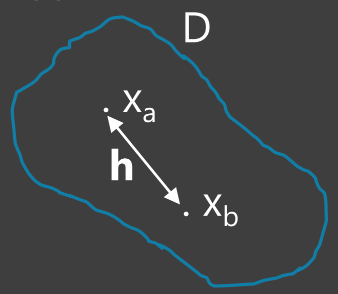
[19] *Isaaks and Srivastava, 1989.*

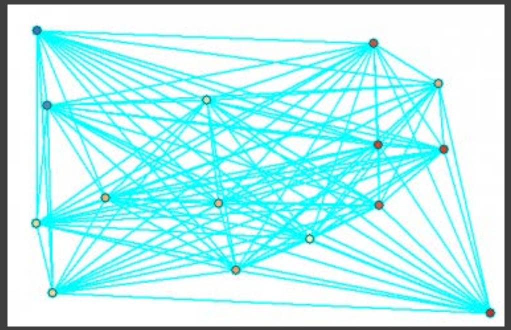
→ STOCHASTIC INTERPOLATION: KRIGING

1st STEP: Describe the spatial variation with variogram

The computation of a variogram involves plotting the relationship between the semivariance and the lag distance:

- Measure the strength of correlation as a function of distance
- Quantify the spatial autocorrelation





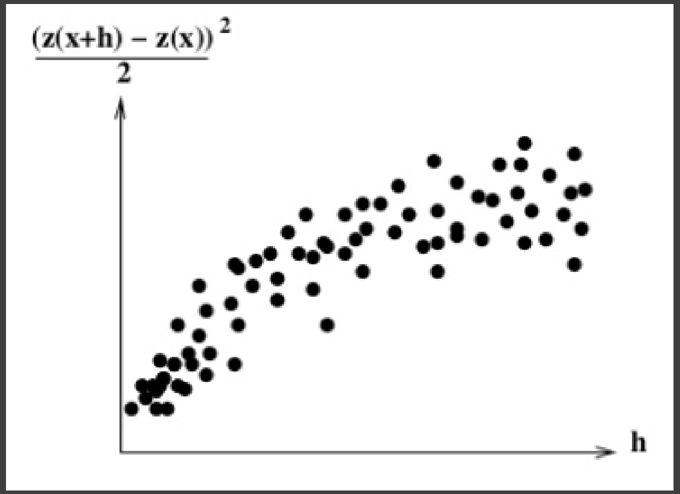
EXAMPLE:

- Consider the vector x=(x1 x2): coordinates of a point in 2D and $\mathbf{h}$ the vector separating 2 points

- Sample values z are compared according to the equation

$$\gamma(\mathbf{h}) = \frac{\left(\mathrm{z}(\mathbf{x}+\mathbf{h})-\mathrm{z}(\mathbf{x})\right)^2}{2}$$ for different lag distances $\mathbf{h}$

- The empirical variogram values $\gamma(\mathbf{h})$ are plotted

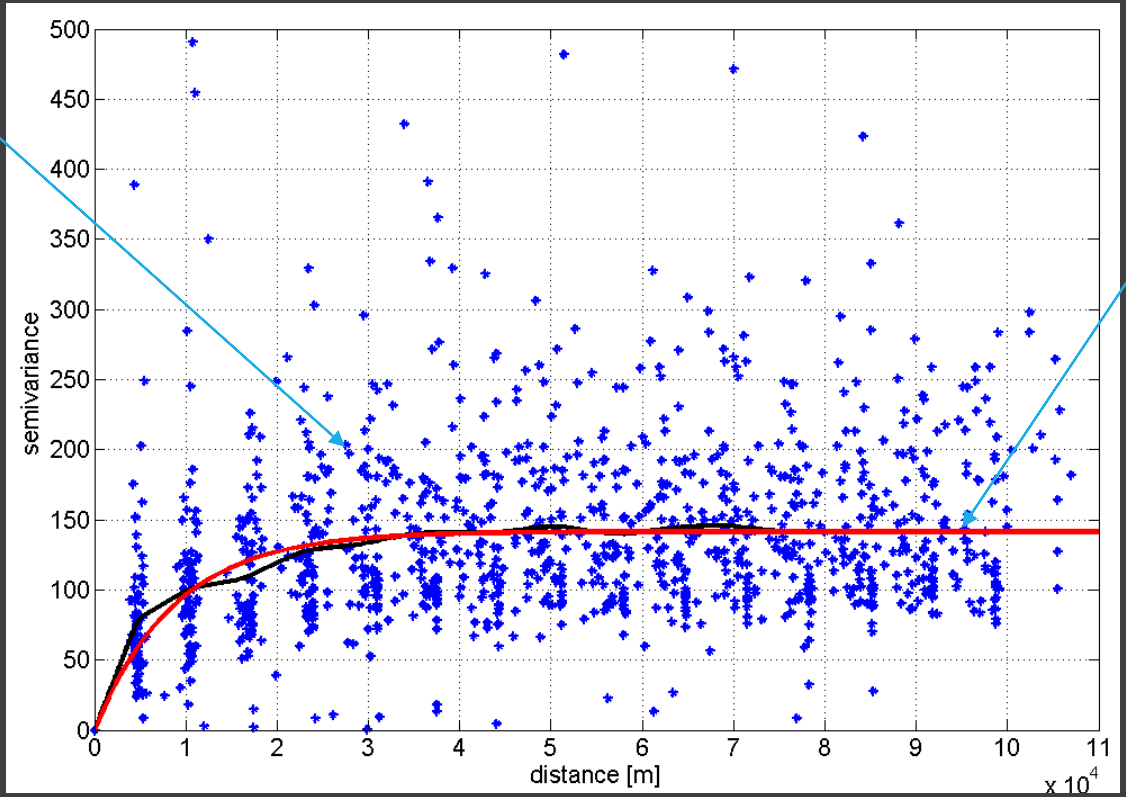


→ STOCHASTIC INTERPOLATION: KRIGING

2nd STEP: Summarize the variation with a mathematical function

Sample variogram cloud

Theoretical variogram



→ STOCHASTIC INTERPOLATION: KRIGING

2nd STEP: Summarize the variation with a mathematical function

SILL

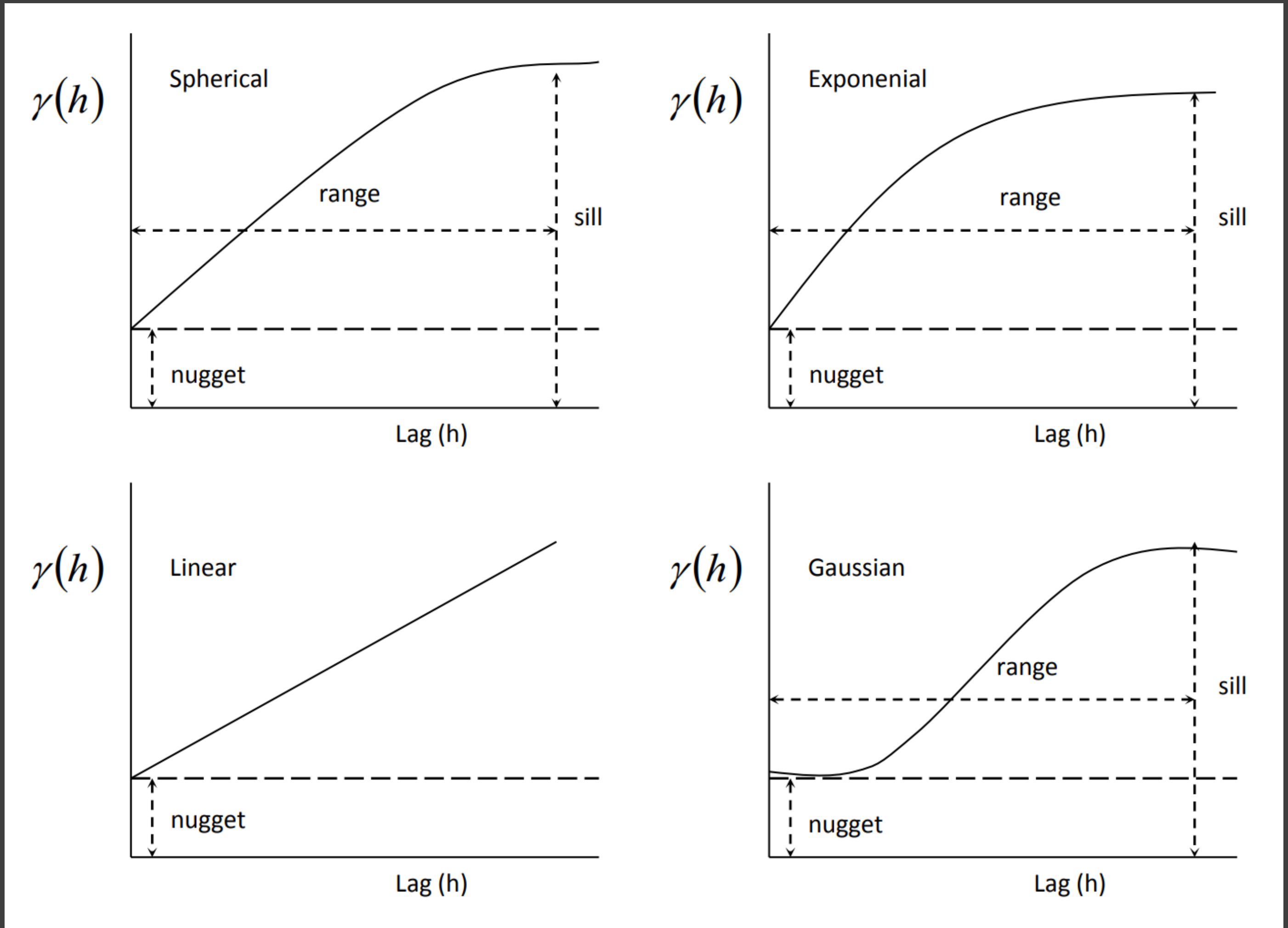
The value where the semivariogram first flattens off, the maximum level of semivariance.

RANGE

The point where the semivariogram reaches the sill on the lag-axis. Sample points that are farther apart than range are not spatially autocorrelated.

NUGGET

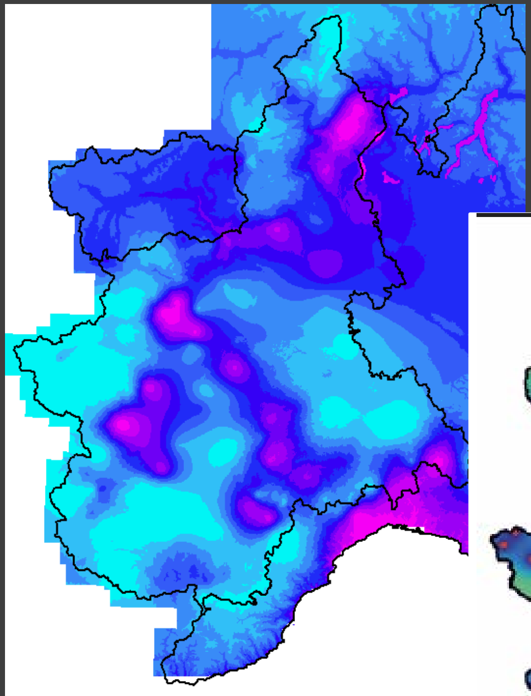
The value of the variogram with 0 lag; errors in measurements

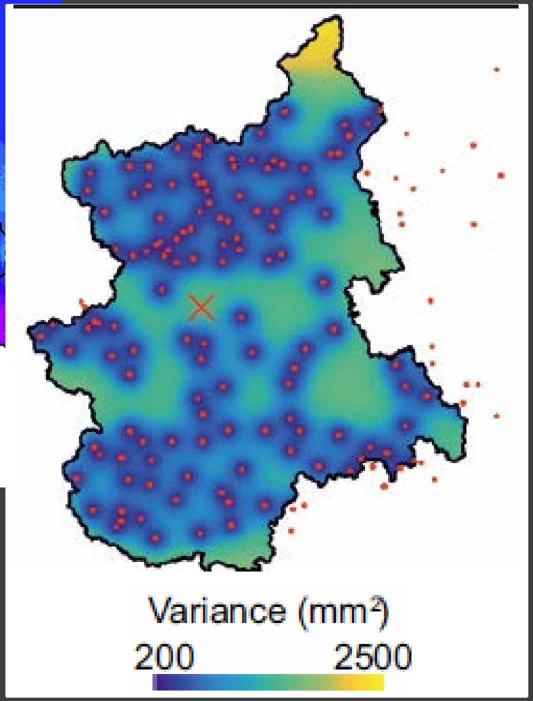


→ STOCHASTIC INTERPOLATION: KRIGING

3rd STEP: Use the function to determine interpolation weights

- The variogram model is used to determine the weights for unknown points.
- The calculation is rather complex, but once the weights are calculated, interpolation is the same as with IDW
- Kriging also produces kriging variance map which can be used for estimating the uncertainty of the interpolation





→ STOCHASTIC INTERPOLATION: KRIGING

Different types of kriging according to model structure…

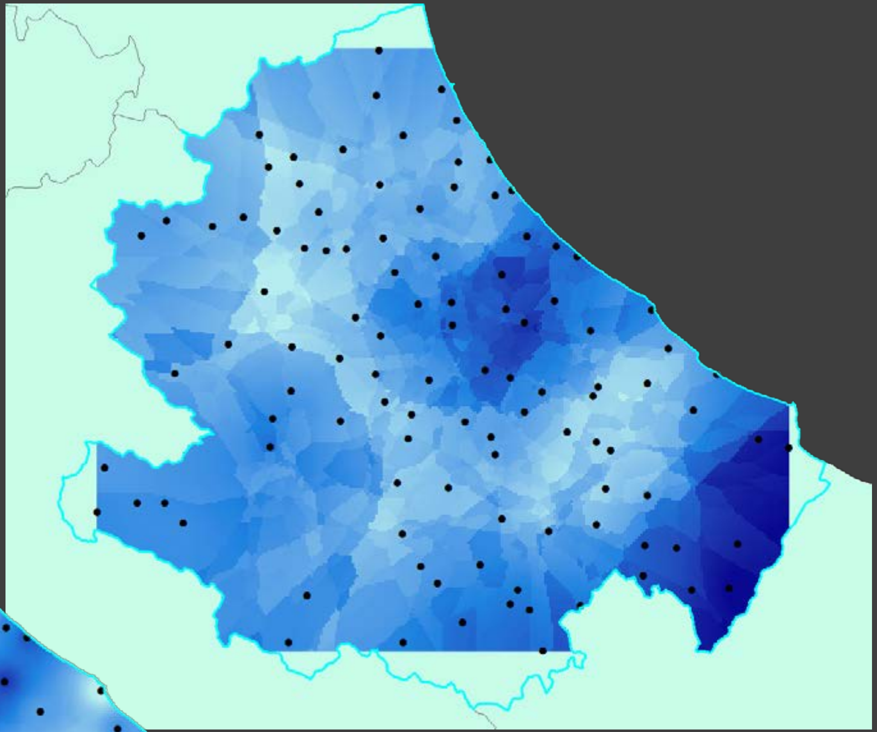
- Ordinary → mean is an unknown value estimated locally
- Simple → mean is a known constant, i.e. average of the entire data set
- Universal → drift in the data is modeled using trend surface analysis and the semivariogram is calculated using residual values from the surface

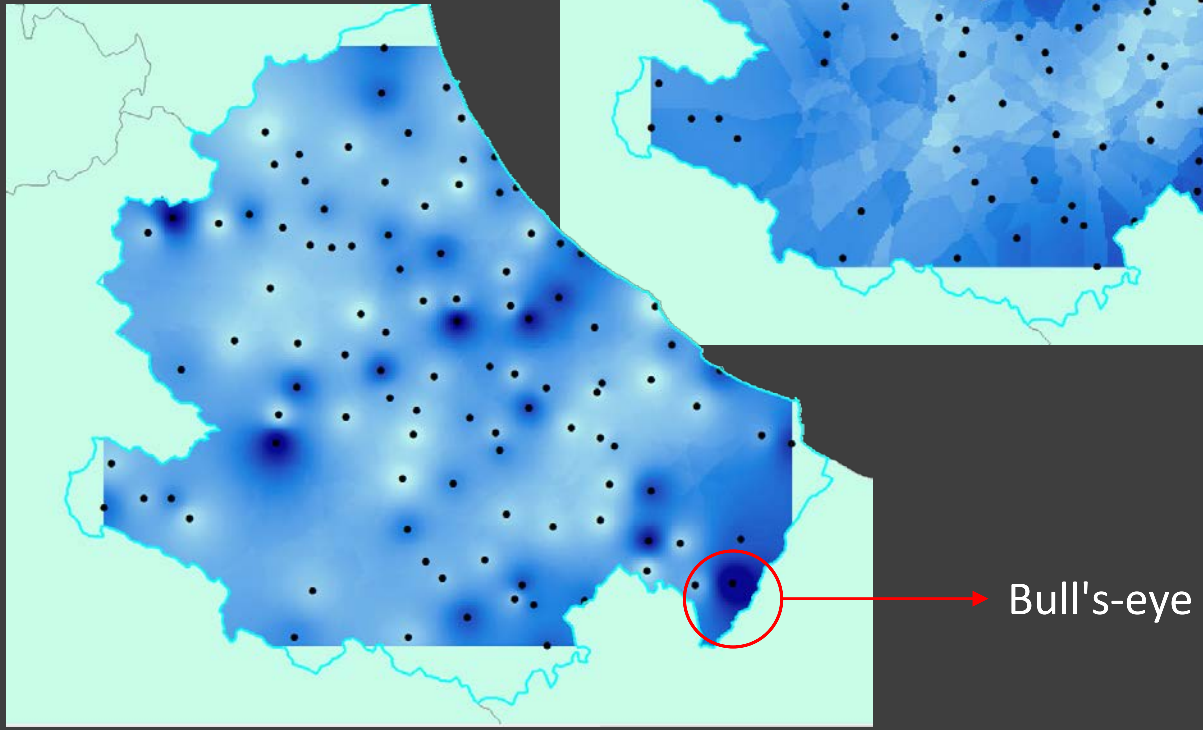
…or on the considered data

- Block → estimates an average value of a block
- Indicator → used when the interpolated value is binary
- Co-kriging → two or more interdependent variables are considered. The information contained in the associated variable is used to enable better estimations of the other variable.

→ SPATIALIZATION OF LOCAL ESTIMATES

Abruzzo 1-hour maxima for year 2010 interpolated with Ordinary Kriging →
and IDW ↓





- Provide estimations representative of the local variability

- Problem dealing with combined spatio-temporal fragmentations (interpolation totally relies on data)

A COMBINED SPACE-TIME APPROACH FOR *RFA*

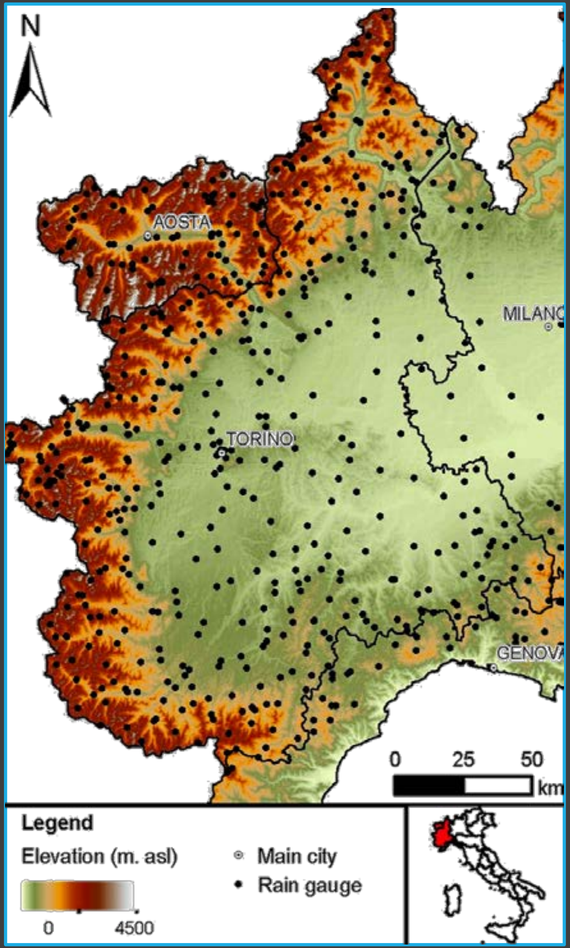
Dealing with extreme rainfall frequency analysis in data-rich environments is often necessary to tackle the space-time problems jointly, to preserve a robust statistical approach without discarding a significant amount of information which can be essential, especially when large return periods estimates are sought.

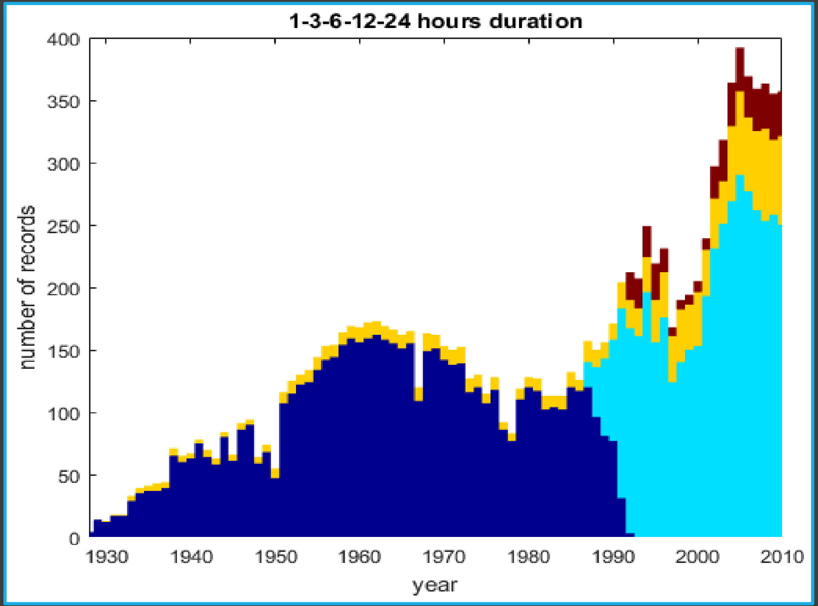
The "patched kriging"[20] techniques allows one to exploit all the information available from the recorded series, independently of their length, to provide extreme rainfall estimates in ungauged areas. The methodology has a low computational cost and does not require to work with stationary or significantly auto-correlated data, as it does not involve any interpolation along the time-axis. This feature proves to be particularly effective when dealing with frequent rain gauge relocations, allowing on the one hand to maximize the usable information at gauged sites, and on the other to extend the analysis to the ungauged ones.

[20] *Libertino et al., 2017*

→ EXAMPLE: The Piemonte region

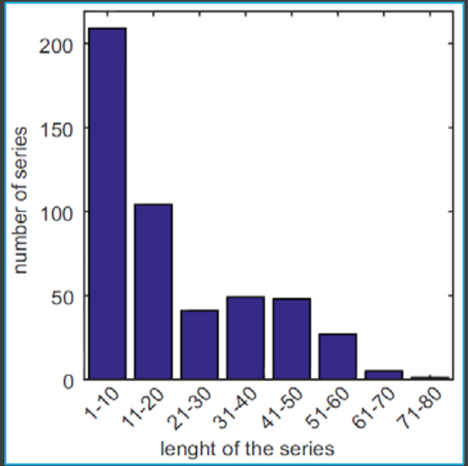
Dataset

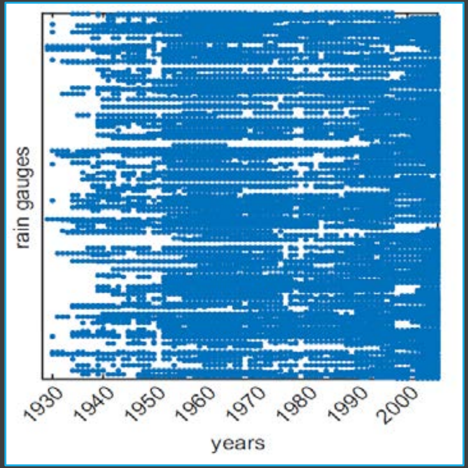




SIMN (Piemonte + VdA until 1992)
ARPA Piemonte (from 1987)
ARPA Lombardia (from 1930)
C.F. ARPA Vda (from 1992)

> 550 gauge locations in 70 years





→ EXAMPLE: The Piemonte region

<u>The hidden storm of Caselle</u>

CASELLE (TO), 13 September 2008[21]

Diluvio nel torinese e in Svizzera, gelo in Scandinavia

Mi piace | Condividi | Follow @meteogiornaleit | G+

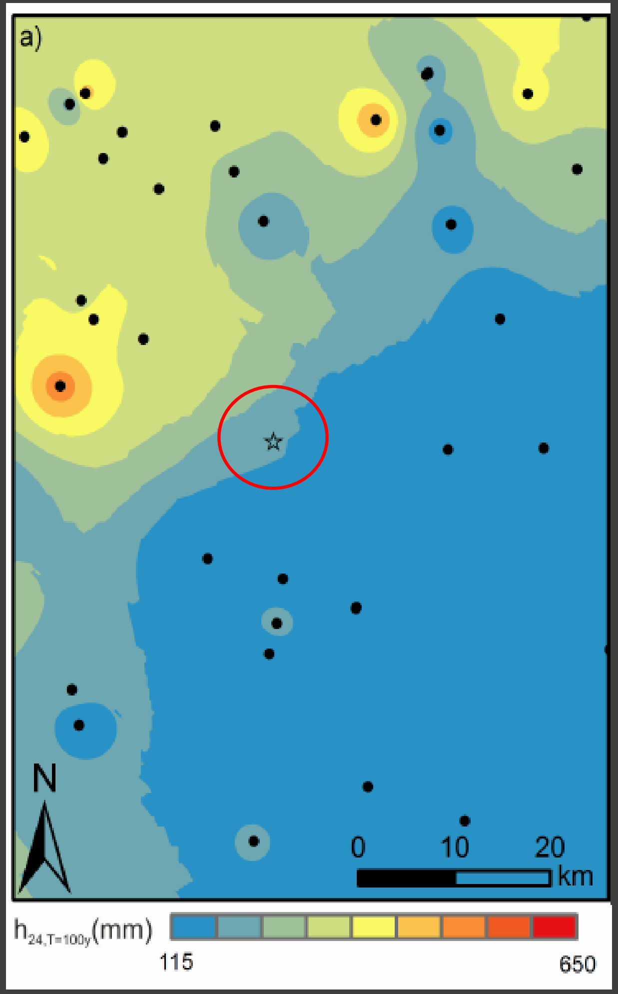
Pubblicato da:

Giovanni Staiano

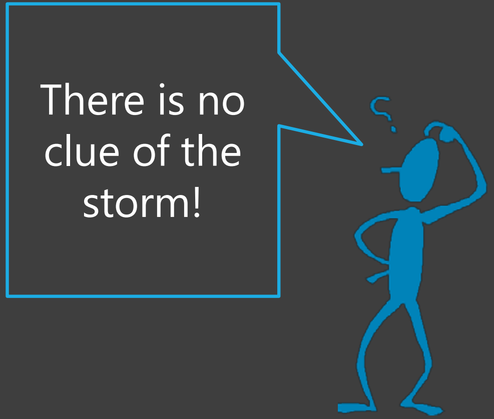
14-09-2008 ore 09:00

Nubifragi si sono abbattuti sabato nel torinese. A Torino Caselle sono caduti 228 mm di pioggia nell'arco di sole 12 ore! Sabato molto piovoso anche in Svizzera. Tra le 18 GMT di venerdì e la stessa ora di sabato, questi alcuni accumuli: Robiei 82 mm, Ulrichen 73, Locarno-Monti 69, Guetsch 67, Grimsel-Hospiz 66, San Bernardino 65, Hinterrhein 62, Lucerna 54. Oltre i 2000 metri le precipitazioni sono state in parte in forma nevosa.

Forti gelate sabato in Finlandia e nei settori settentrionali di Svezia e Norvegia. In Finlandia,



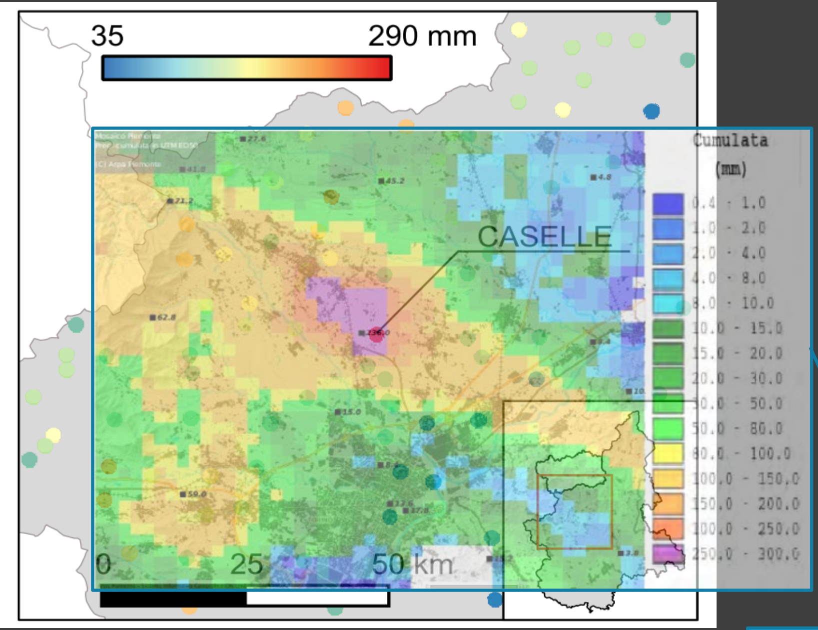
Design rainfall for T=100 years estimated at-gauge with the *I-RED* database and interpolated with Inverse Distance Technique. A threshold on the series length L=20 years has been set.



[21] *ARPA Piemonte, 2008*

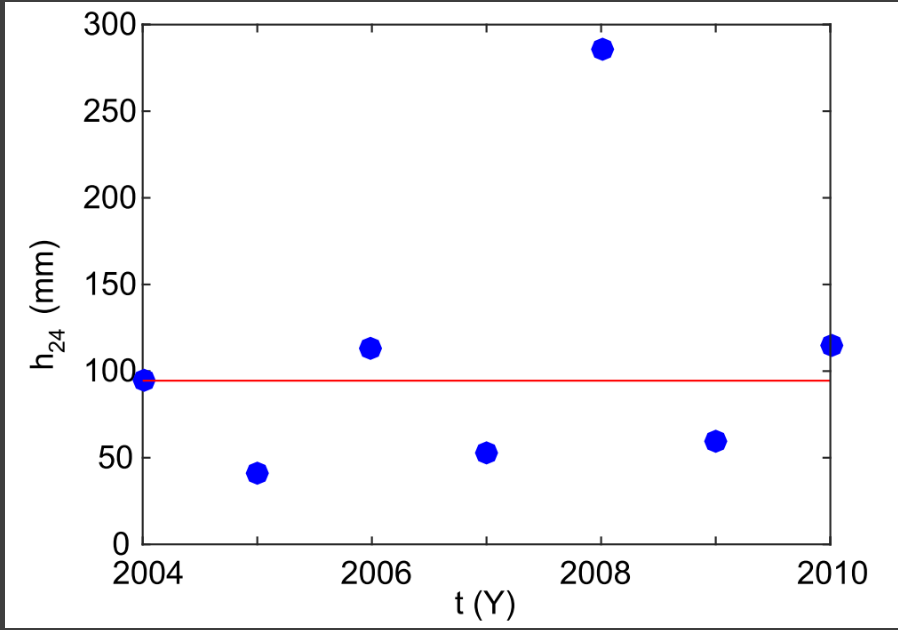
→ EXAMPLE: The Piemonte region

The hidden storm of Caselle



Annual maxima for 24 hours duration for the year 2008.

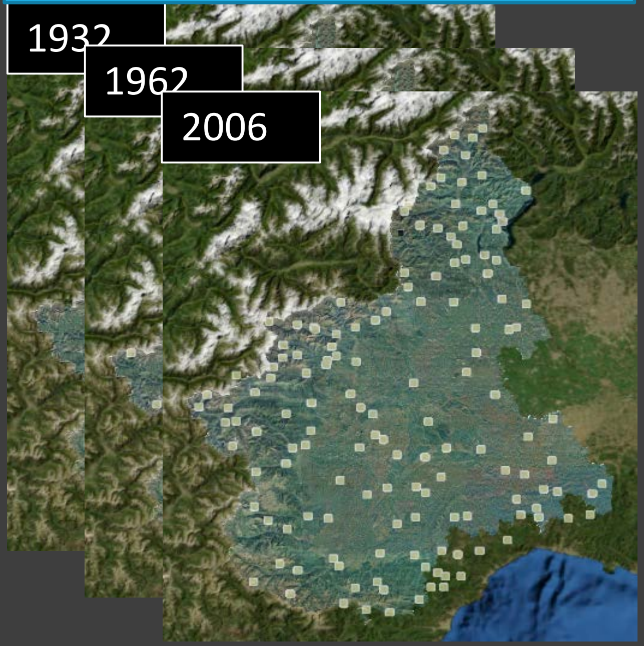
Radar cumulative rainfall map from 6 to 18 UTC 13/9/2008



Annual maxima for 24 hours duration for the CASELLE rain gauge in the I-RED database.

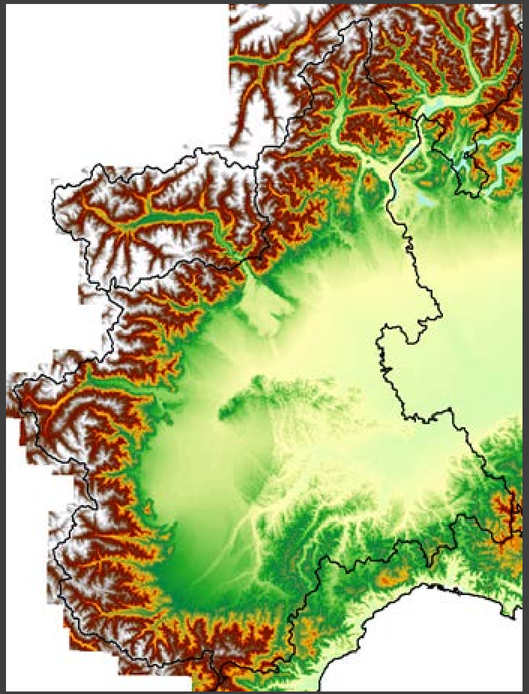
→ THE PATCHED KRIGING METHODOLOGY

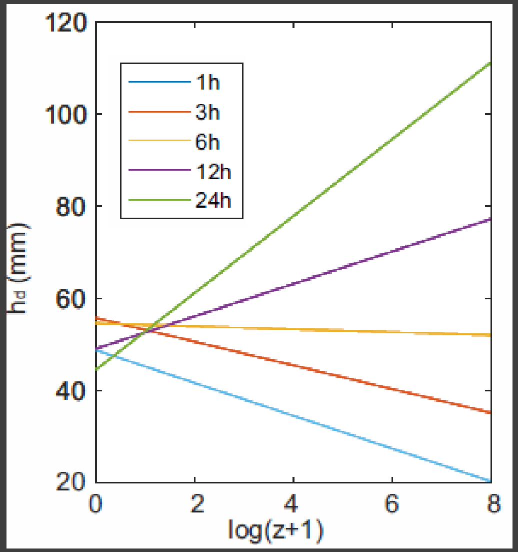
Every year…
Durations 1-24 h



Ordinary Kriging relies on the assumption that the covariance between any two random errors depends only on the distance → Need to remove trend with elevation.

$$h_d = m \cdot \ln(z+1) + m_0 + \epsilon_d$$

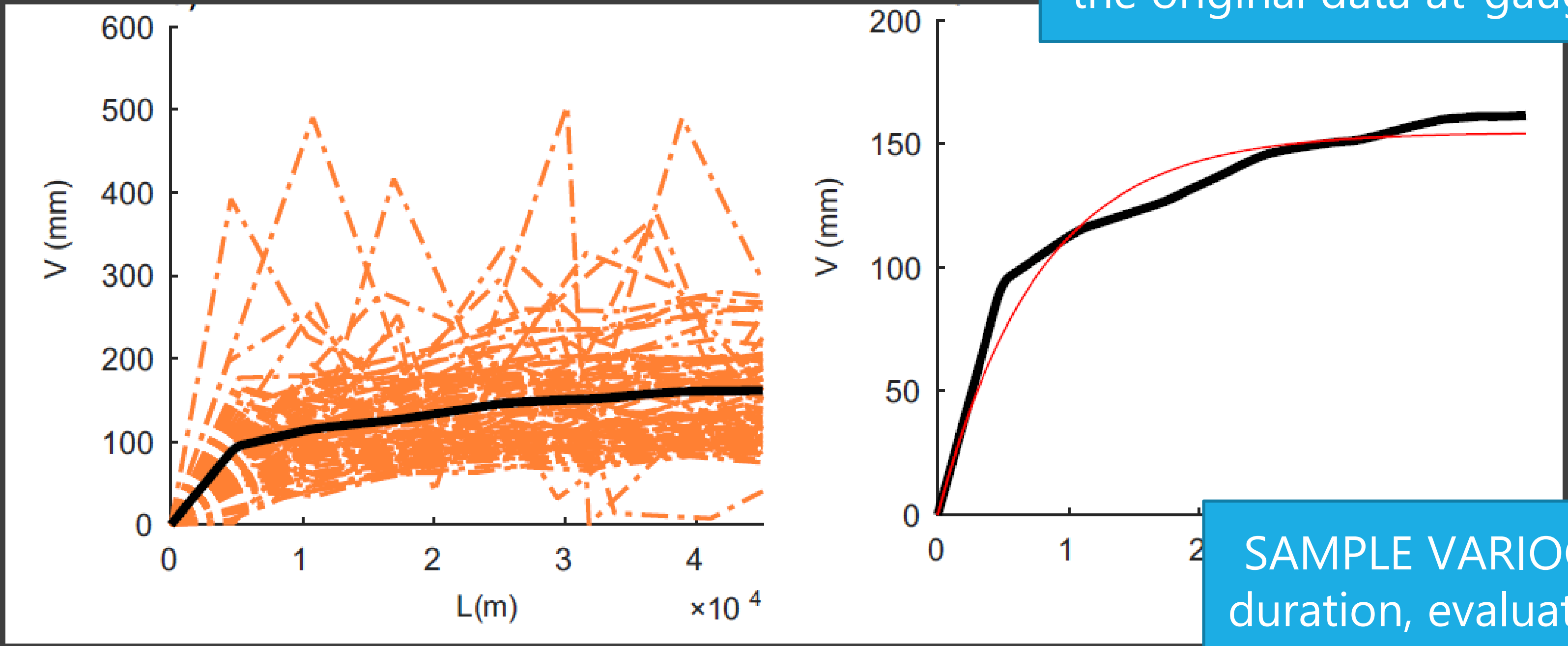




Detrending with elevation

→ THE PATCHED KRIGING METHODOLOGY

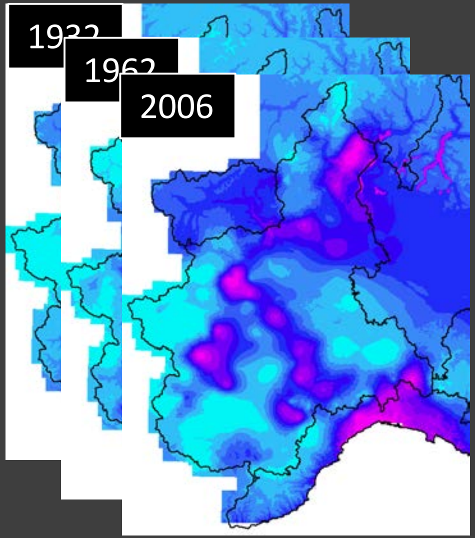
NUGGET is set to 0 to preserve the original data at-gauge points.



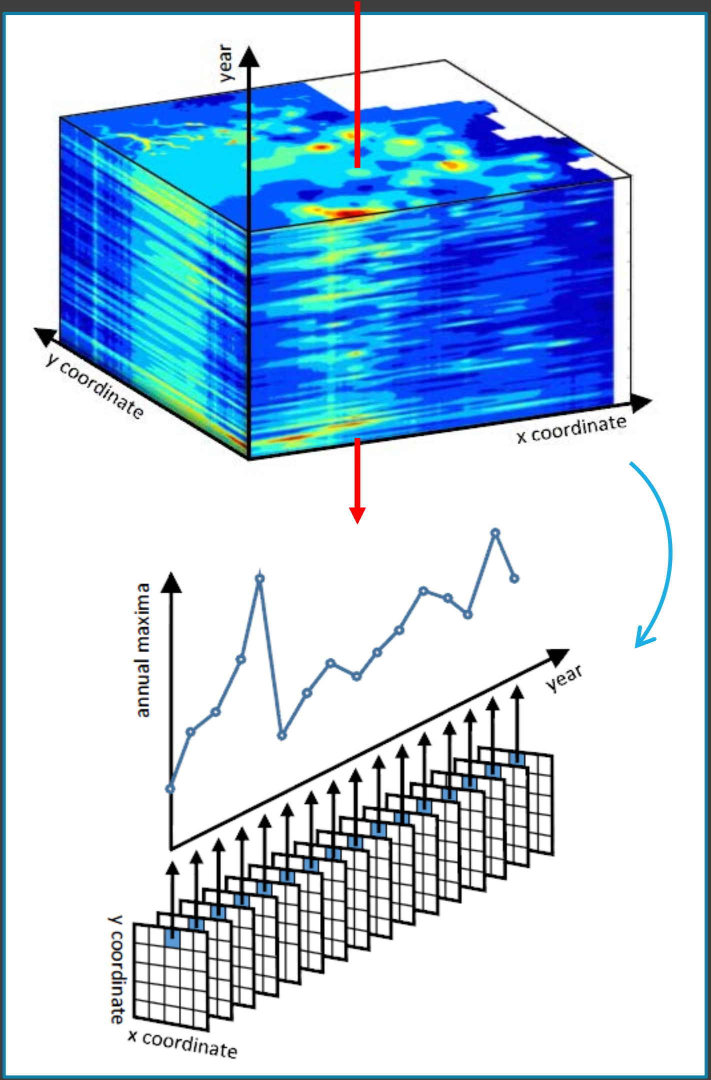
Theoretical and exponential variograms for the 1-hour duration. The orange dashed lines refer to the annual sample variograms, the black curve is the average sample variogram and the red one the theoretical fitted one.

SAMPLE VARIOGRAM for each duration, evaluated as the mean of the annual variograms weighted on the number of active rain gauges every year. Exponential THEORETICAL VARIOGRAM.

→ THE PATCHED KRIGING METHODOLOGY



Ordinary Kriging equations are applied considering th
10 nearest stations and results re-trended

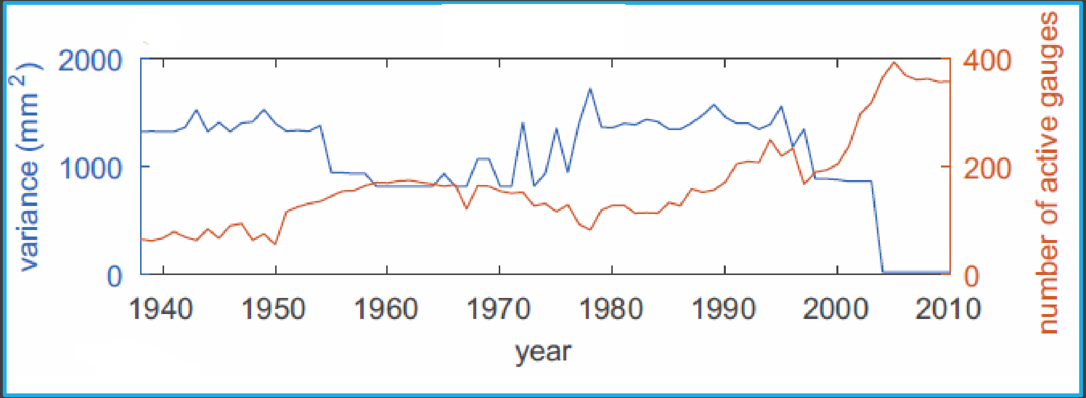


Rainfall cube
+
Kriging variance cube

«Coring» the cube along the time axis a set of complete «cored series» can be obtained

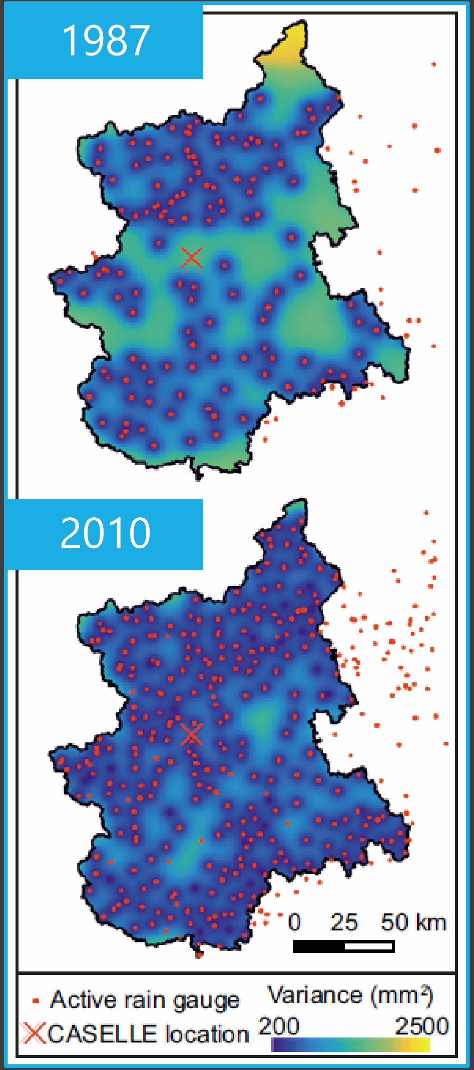
→ WEIGHTING THE L-MOMENTS

Kriging variance is larger in cells far from a gauged location and for a fixed cell increases/decreses when the number of station in the area decreases/increases

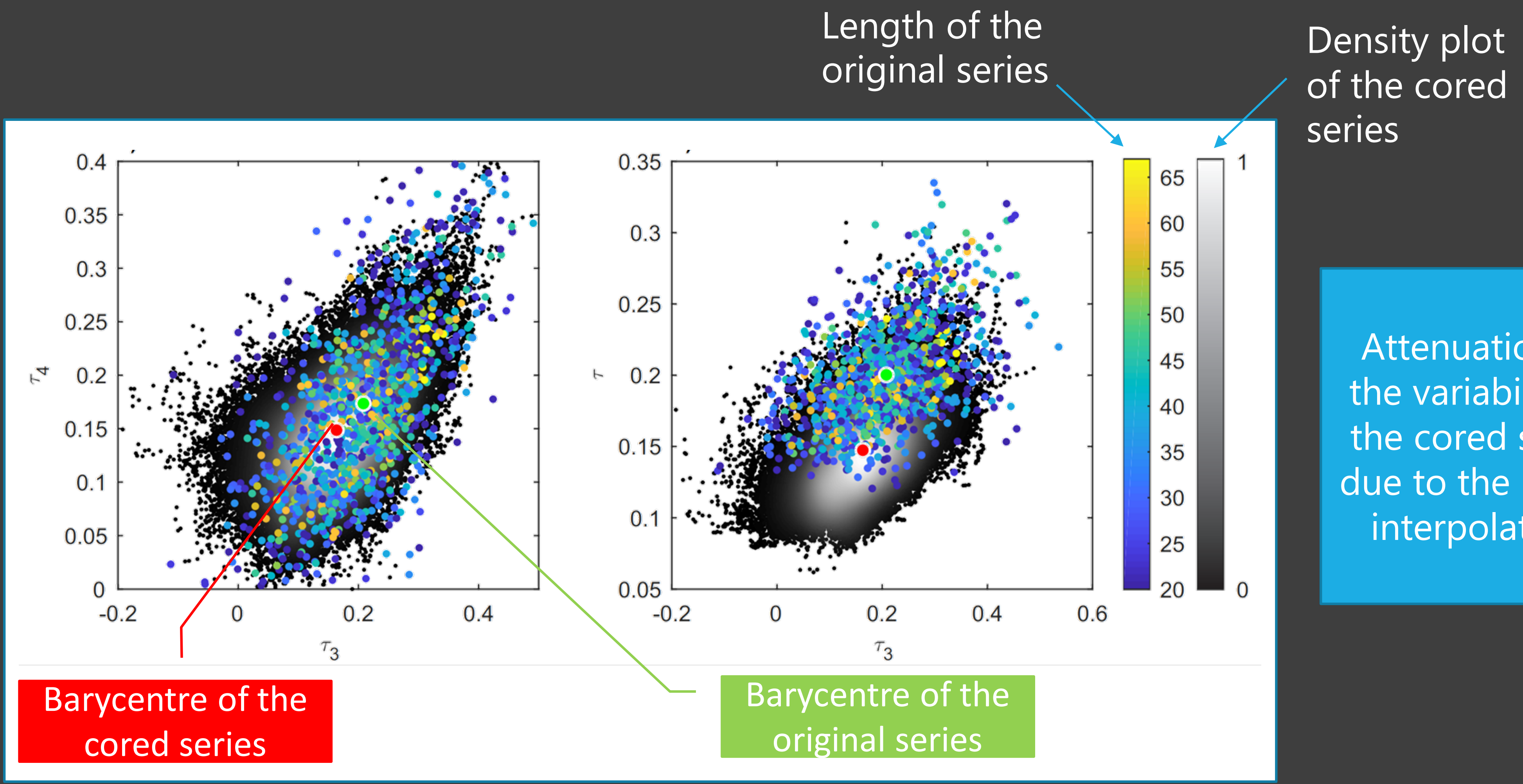


L-moments are weighted on the kriging variance, giving lower weight to the estimated values and to the years poor in data.

$$w_i = \frac{\sigma^2_{\max}}{\sigma^2_i} \quad w_{i,max} = 10$$



→ PRELIMINARY RESULTS



$\tau_3 - \tau_4$ and $\tau_2 - \tau$ plots. The grey dots represent the cored series and the coloured ones the original ones.

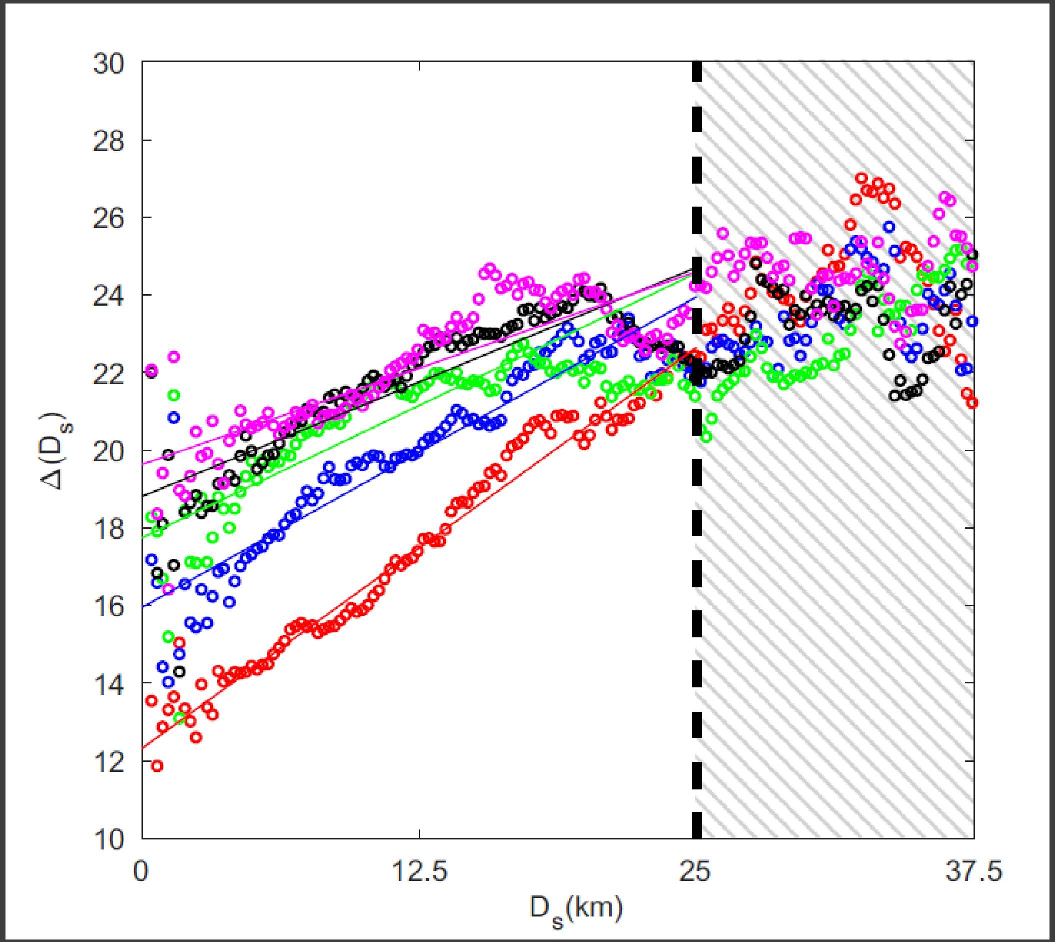
Attenuation of the variability in the cored series due to the use of interpolation.

→ BIAS CORRECTION

A bias correction factor factor *K* is introduced, as an increasing function of the distance from the nearer rain gauges, stemming from the assumptions that:

- If the target point is close to a gauging station, the distribution of the cored series will likely be very similar to the one of the original series, and then correction should be very limited.

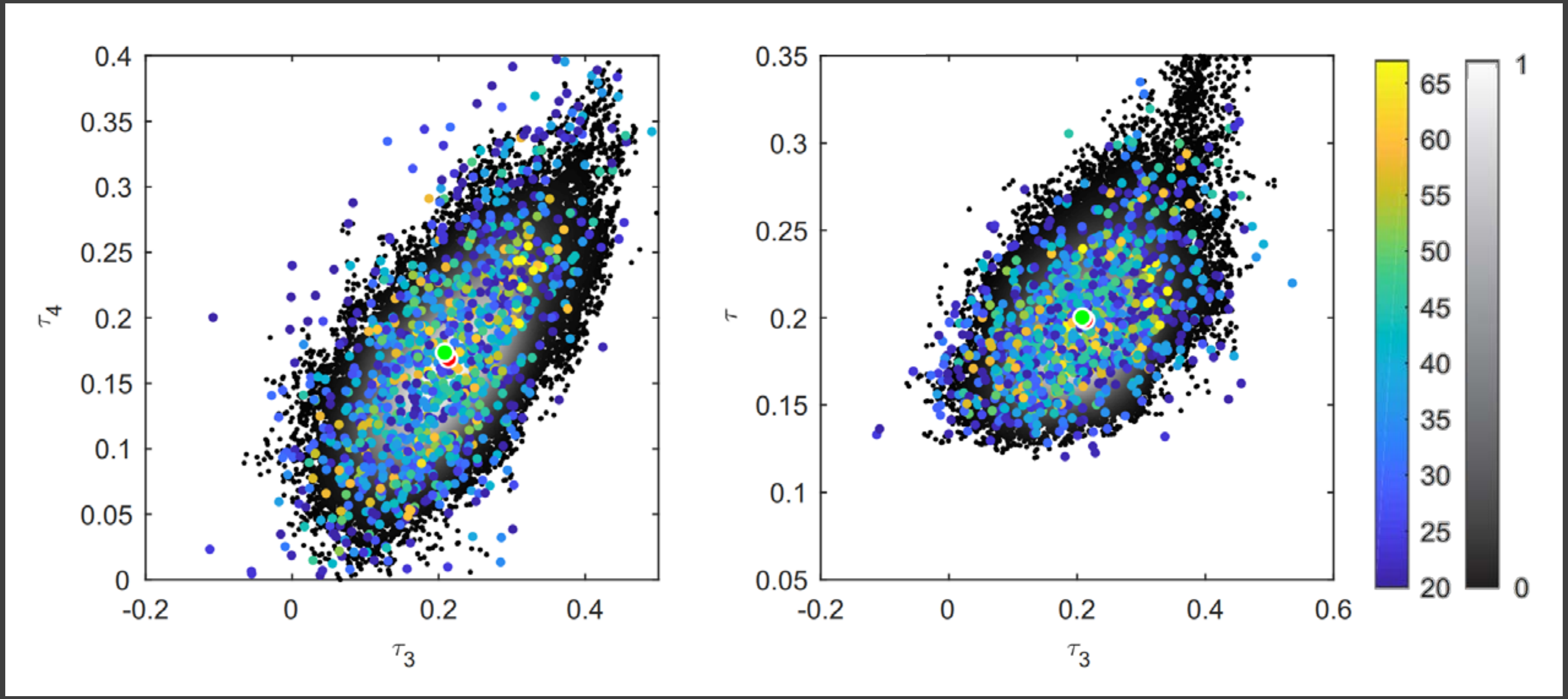
- When the target point moves further away from the gauging stations, the smoothing effect becomes very relevant and the correction becomes essential.

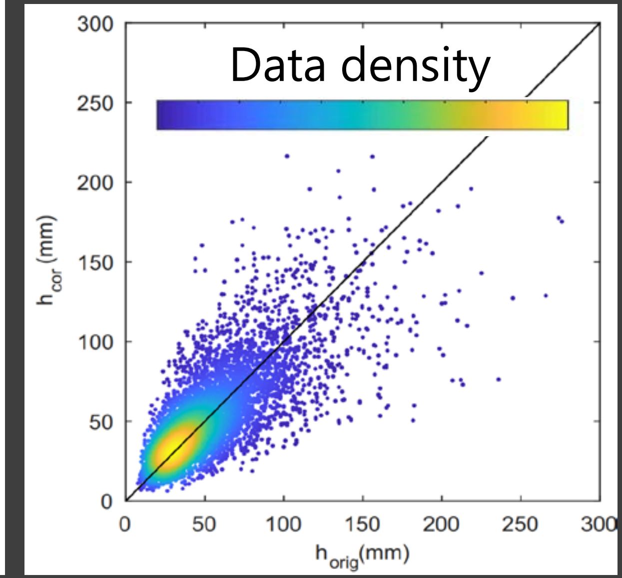


→ RESULTS

The patched kriging is able to provide not only series with L-moments consistent with those of the original ones, but also to reconstruct reliable annual maxima at ungauged areas preserving the information contained in the short series.

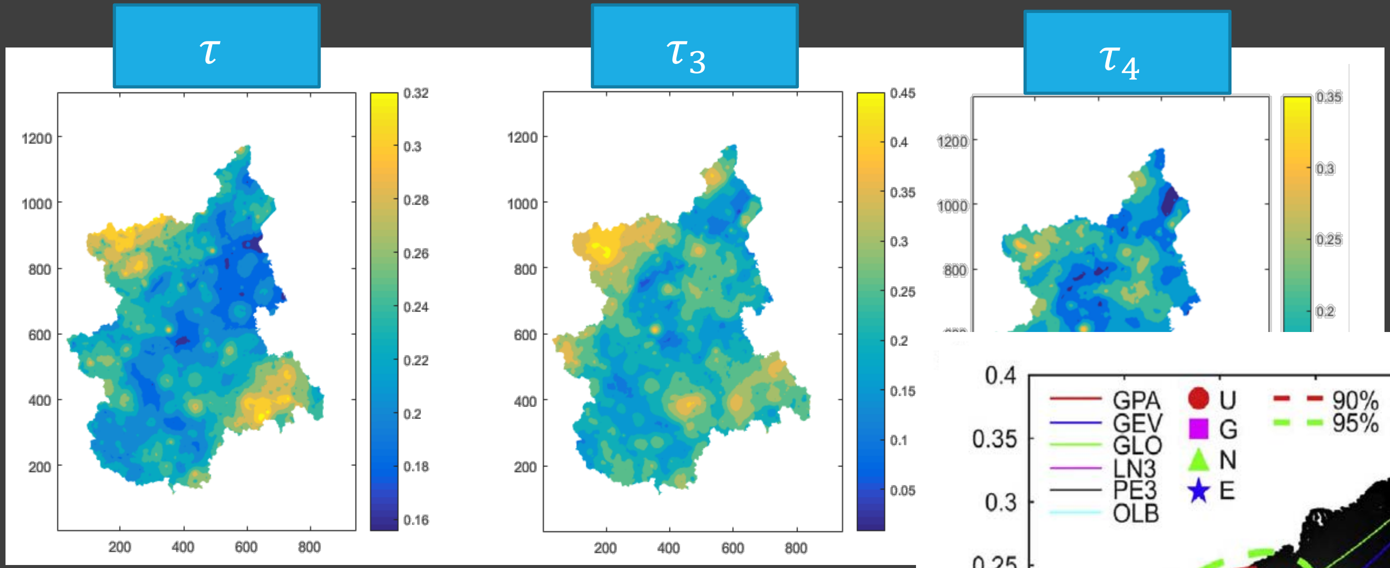
CROSSVALIDATION
original VS cored values



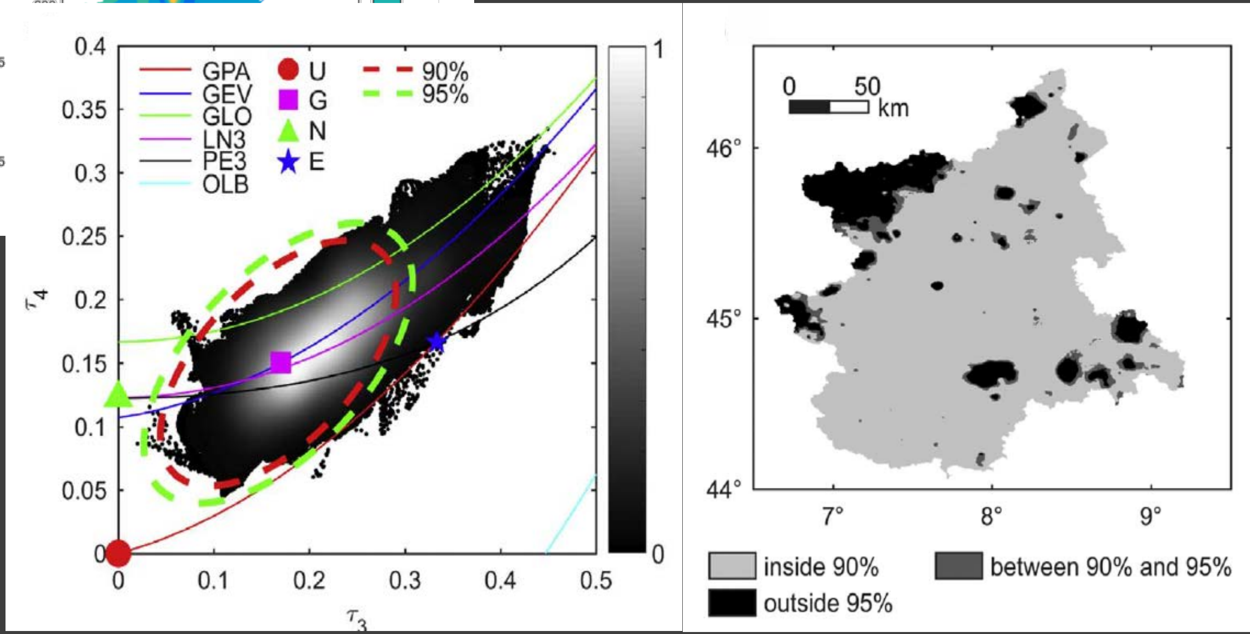


$\tau_3 - \tau_4$ and $\tau_2 - \tau$ plots. The grey dots represent the cored series and the coloured ones the original ones.

→ RESULTS



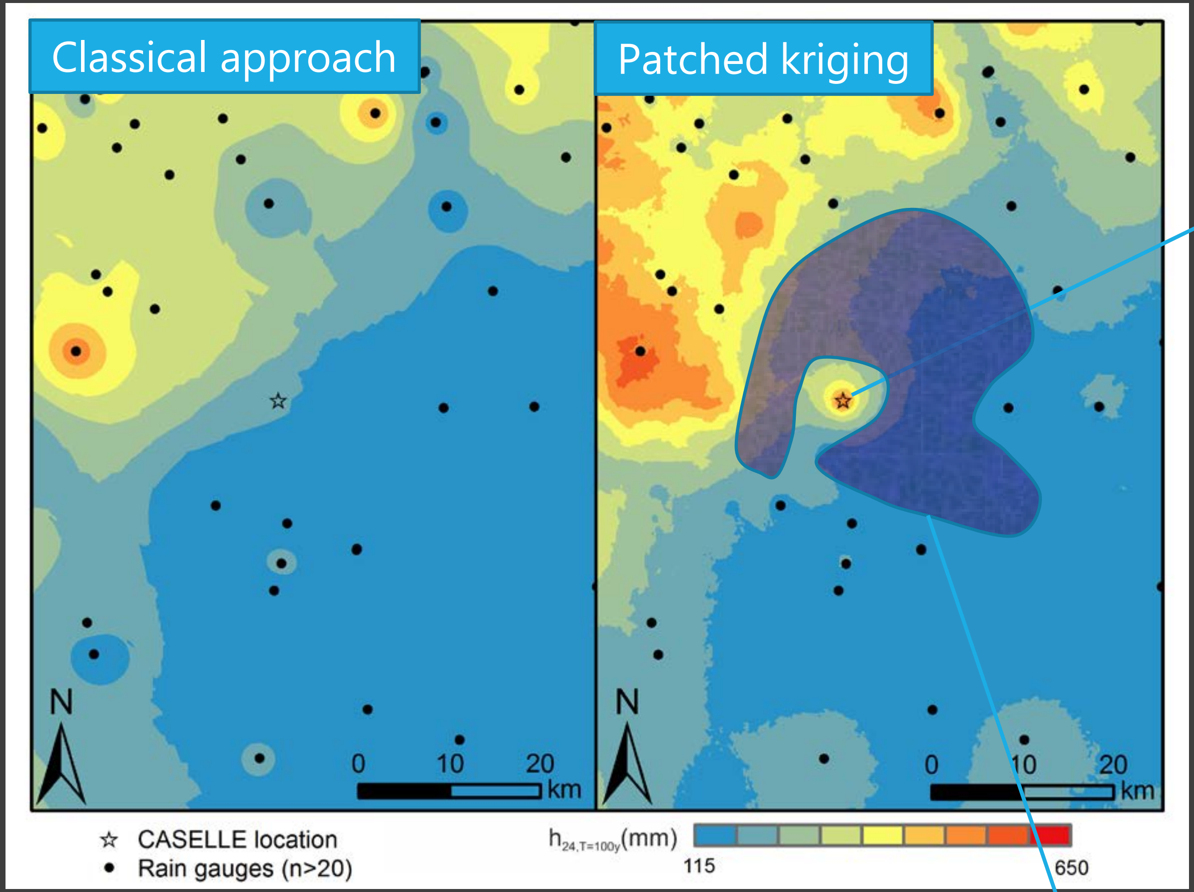
Distributed frequency analysis



(Left) L-Moments ratio diagram[2] related to the cored series. The ellipses represent the 95% and the 90% acceptance areas, defined by bootstrapping from a Gumbel distribution. The colour scale is proportional to the density of the points. (Right) Spatial distribution of the cells falling inside the 90% and 95% acceptance area of the Gumbel distribution.

[2] *Hosking and Wallis, 1997*

→ RESULTS

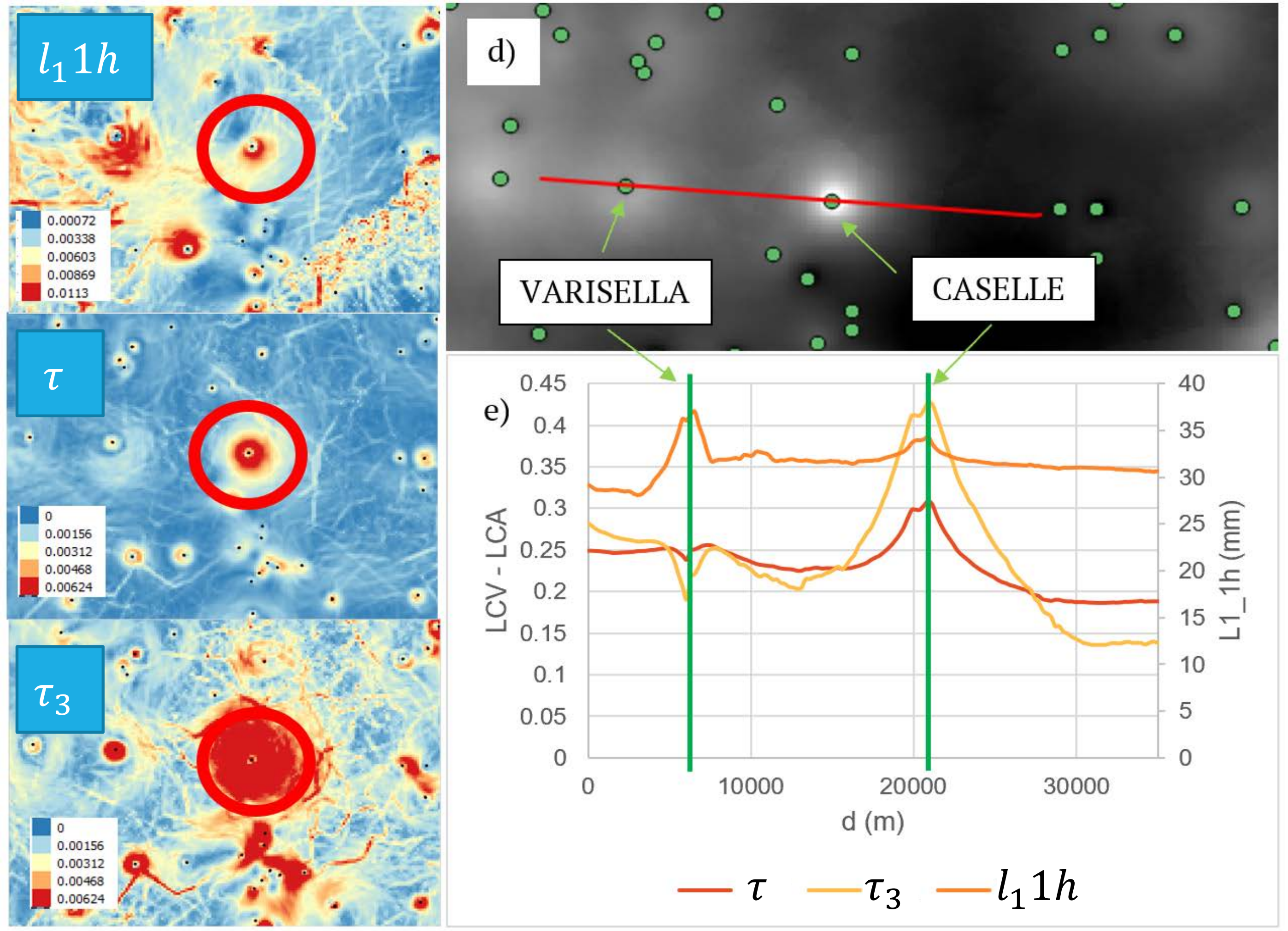


A clue that something happened here is now evident.

Interpolation techniques can only represent the estimation variance determined by the spatial and temporal resolution of the data, no clues of what happened here («Bull eyes effect»)

→ SPATIALIZATION OF THE LOCALIZED INFORMATION[22]

Analyzing the spatial distribution of the spatial derivative of the L-moments it can be seen that the spatial influence of the anomalies is more significant when the order of the L-moments increases. This is directly linked to reasons of sample variability of the L-moments which, although more robust than the classic moments, lose strength when the order increases.

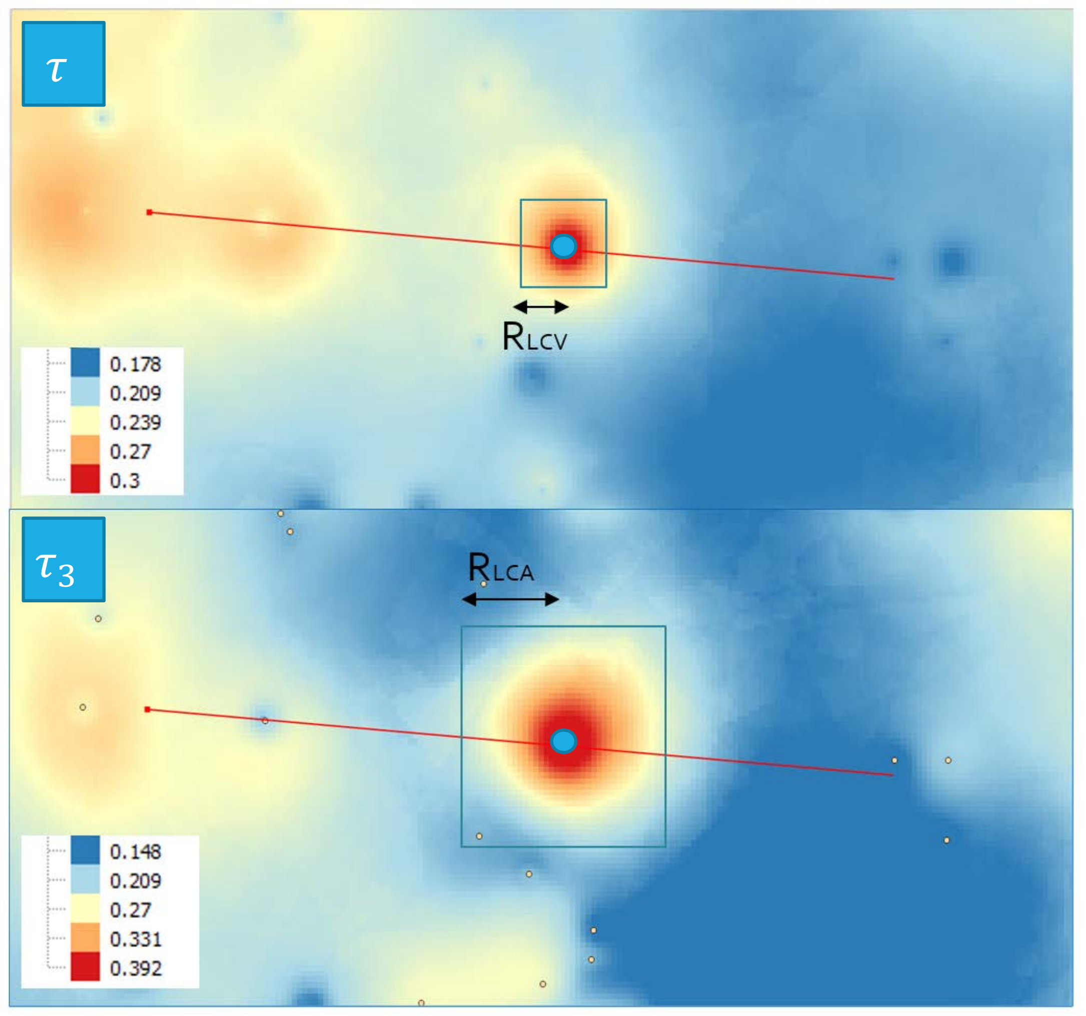


→ SPATIALIZATION OF THE LOCALIZED INFORMATION[22]

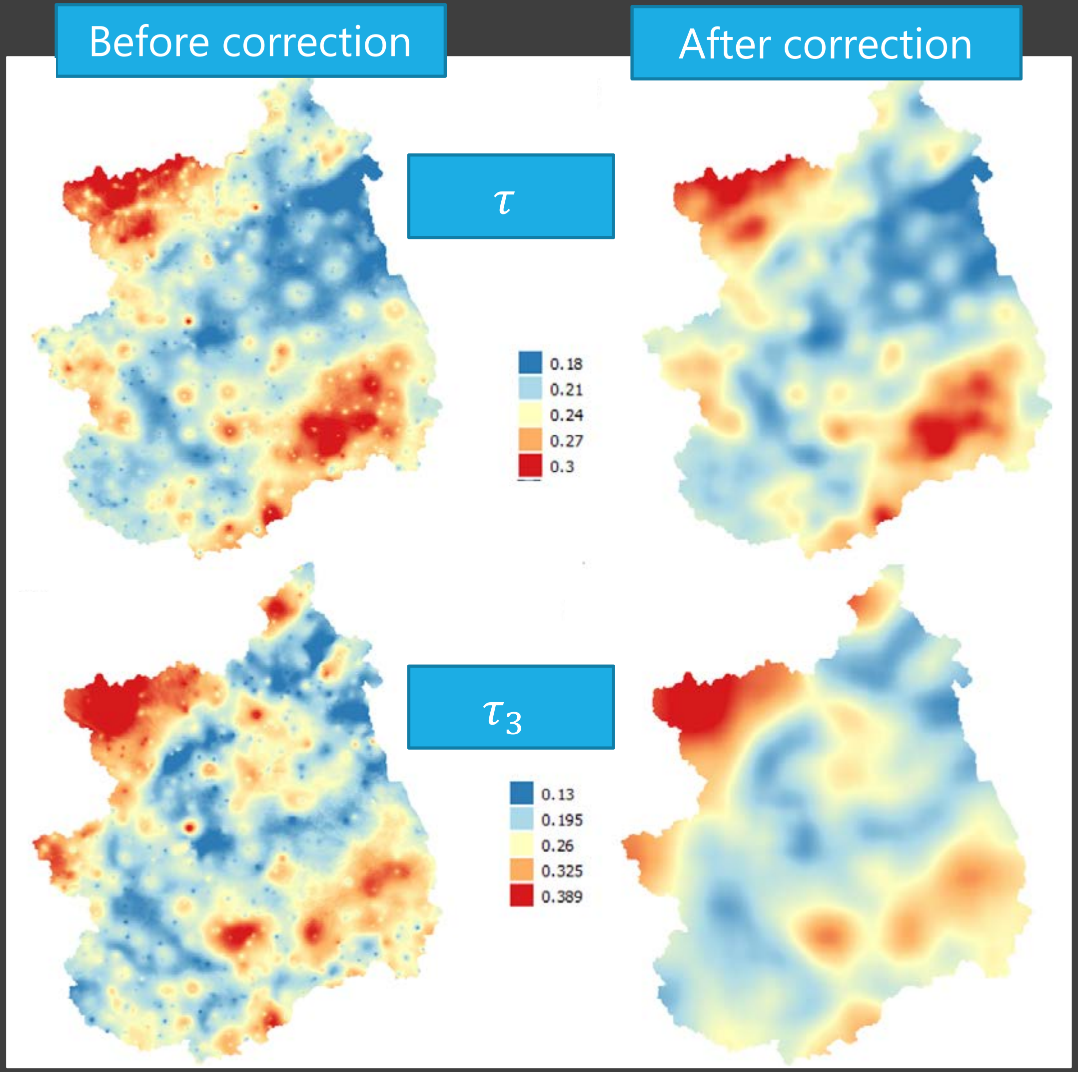
For the estimation of the parameters of the distributions, the L-moments are spatially filtered on areas of increasing radius as the order of moments increased. In detail, for each cell are considered:

- The mean relative to the cell itself
- τ spatially averaged on a $2R_{LCV}$ x$2R_{LCV}$ square
- τ_3 spatially averaged on a $2R_{LCA}$ x$2R_{LCA}$ square

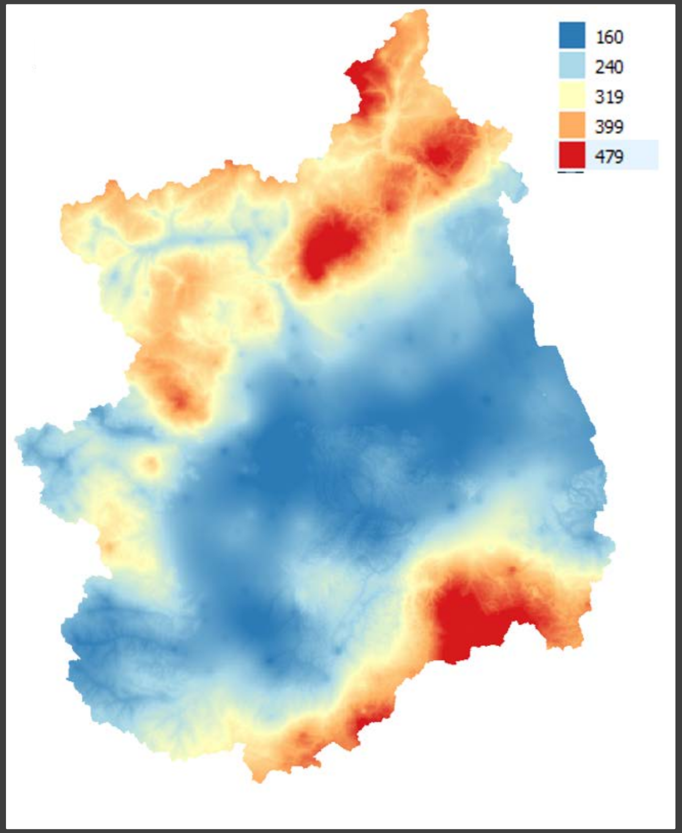
with $R_{LCA} > R_{LCV}$ estimated with reference to the spatial correlogram of the data.

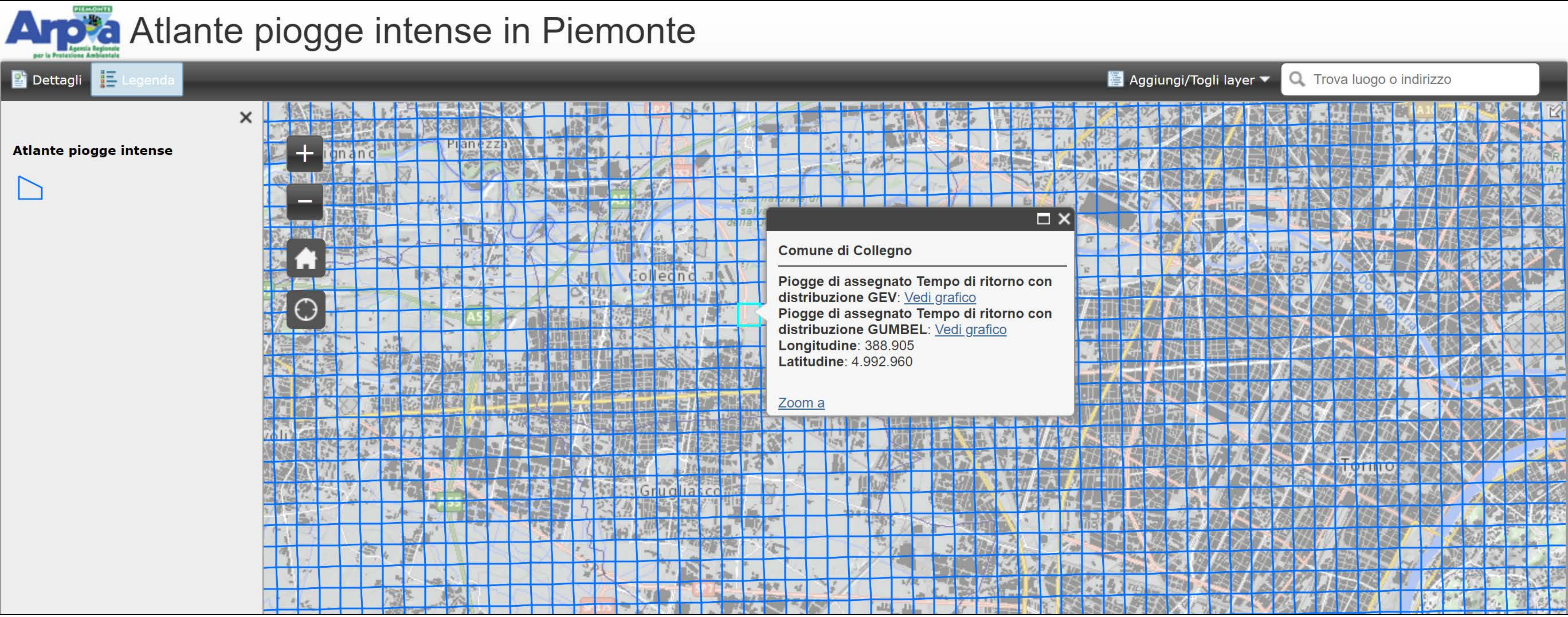


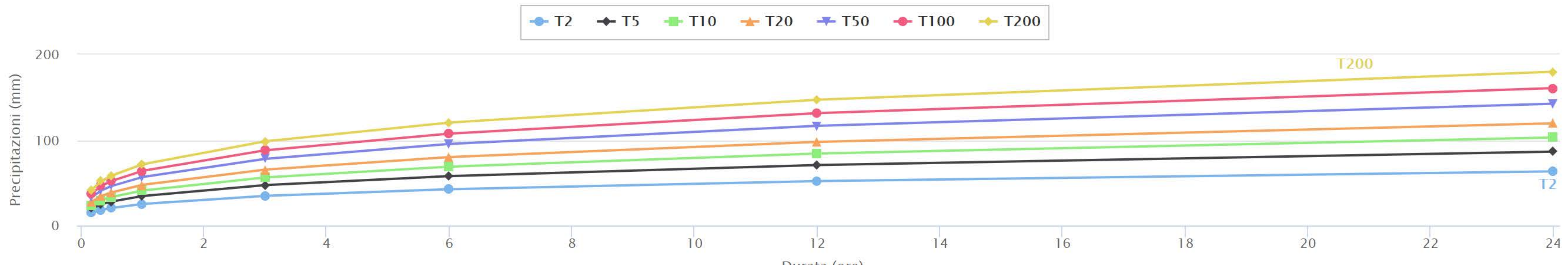
→ SPATIALIZATION OF THE LOCALIZED INFORMATION[22]



Design rainfall for T=200 years with a GEV distribution.




Arpa
Piemonte
Agenzia Regionale per la Protezione Ambientale
Atlante piogge intense in Piemonte
Dettagli
Legenda
Aggiungi/Togli layer
Trova luogo o indirizzo
Atlante piogge intense
Pianezza
Collegno
Grugliasco
Torino
Comune di Collegno
Piogge di assegnato Tempo di ritorno con distribuzione GEV: Vedi grafico
Piogge di assegnato Tempo di ritorno con distribuzione GUMBEL: Vedi grafico
Longitudine: 388.905
Latitudine: 4.992.960
Zoom a


T2
T5
T10
T20
T50
T100
T200
Precipitazioni (mm)
200
100
0
0
2
4
6
8
10
12
14
16
18
20
22
24
T200
T2

OPEN ISSUES

RAIN GAUGES
“Patched Kriging”
-> need integration for
better representing rainfall fields

EXTREME RAINFALL
ON A WIDE COMPLEX
DOMAIN

DATA ASSIMILATION at
DIFFERENT SPATIO-
TEMPORAL SCALES

REFERENCES

1) **Dalrymple T.** Flood-frequency analyses, manual of hydrology: Part 3. *Technical notes.* USGPO, 1960.
2) **Hosking J. R.M. and Wallis. J. R.** *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, 1997.
3) **Claps P., Barberis C., De Agostino M., Gallo E., Laguardia G., Laio F., Miotto F., Plebani F., Vezzù G., Viglione A., and Zanetta M.** Development of an information system of the italian basins for the CUBIST project. *EGU General Assembly,* 2008, Vienna
4) **Libertino A., Ganora D., and Claps P.** Space–time analysis of rainfall extremes in Italy: clues from a reconciled dataset. Hydrology and Earth System Sciences 22.5 (2018): 2705-2715.
5) **Little R. J.A., and Donald B. R.** Statistical analysis with missing data. Vol. 333. John Wiley & Sons, 2014.
6) **Teegavarapu R. S.V**. Floods in a changing climate: extreme precipitation. Cambridge University Press, 2012.
7) **Onorato L., Giannoni F., Gollo P., and Turato B.** Rapporto di evento meteoidrologico del 04/10/2010. ARPAL, Genova, 2011.
8) **Onorato L., Bonati V., Cavallo A., and Turato B.** Rapporto di evento meteoidrologico del 09/10/2014. ARPAL, Genova, 2014.
9) **Fiori E., Comellas A., Molini L., Rebora N., Siccardi F., Gochis D. J., Tanelli S., and Parodi A.** Analysis and hindcast simulations of an extreme rainfall event in the Mediterranean area: The Genoa 2011 case. Atmospheric Research 138 (2014): 13-29.
10) **Reed D. W., Faulkner D., Robson A., Houghton-Carr H., & Bayliss A.** *Flood Estimation Handbook*. *Volume I: Overview*. 1999.
11) **Saikranthi K., Rao T. N., Rajeevan M., and Rao S. V. B.** Identification and validation of homogeneous rainfall zones in India using correlation analysis. J. Hydrometeor. 14 (2013): 304–317.
12) **Hosking J. R.M. and Wallis. J. R.** The effect of intersite dependence on regional flood frequency analysis, Water Resour. Res., 24.4 (1988): 588– 600.
13) **Burn D. H**. Evaluation of regional flood frequency analysis with a region of influence approach, Wat. Resour. Res. 26 (1990): 2257-2265.
14) **Caporali E., Cavigli E., and Petrucci A.** The index rainfall in the regional frequency analysis of extreme events in Tuscany (Italy). Environmetrics 19.7 (2008): 714-724.
15) **Rossi F., Fiorentino M., and Versace P.** Two-component extreme value distribution for flood frequency analysis. *Wat. Res. Res.* 20.7 (1984): 847–856.
16) **Reed D. W., Faulkner D. S. , and Stewart E. J.** The FORGEX method of rainfall growth estimation II: Description. Hydrology and Earth System Sciences Discussions 3.2 (1999): 197-203.
17) **Agrillo G., and Bonati V.** Atlante climatico della Liguria. ARPAL, Genova, 2013.
18) **Teegavarapu R. S.V, and Chandramouli V.** Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. Journal of Hydrology 312.1-4 (2005): 191-206.
19) **Isaaks E. H., and Srivastava R. M.** An introduction to applied geostatistics. Oxford university press, 1989.
20) **Libertino A., Allamano P., Laio F., and Claps P.** Regional-scale analysis of extreme precipitation from short and fragmented records. Advances in Water Resources 112 (2018): 147-159.
21) **ARPA Piemonte.** Eventi di precipitazione intensa dell'estate 2008. ARPA Piemonte, Torino, 2008.
22) **Claps P., Laio F., Allamano P., Libertino A., and Iavarone M..** Attività di ricerca nell'ambito del progetto STRADA 2.0, Modulo CAPPIO (CAPitalizzazione azione di Caratterizzazione delle PIOgge estreme) – CUP E96G15000010007- CIG Z2513A8348. Final Report, ARPA Piemonte – CiNiD, 2015.