Geostatistical tools for analyzing spatial extreme rainfall patterns over Tunis City

Zoubeida Bargaoui
National School of Engineers of Tunis (ENIT)

Tunis El Manar University

4th International workshop on hydrological extremes
Med-FRIEND group
University of Calabria 15-17 September 2011

Zoubeida.bargaoui@laposte.net

Introduction

- The identification of the structures of heavy rainfall spatial variability helps:
  - deriving rainfall patterns
  - quantifying rainfall amounts in unobserved sites
  - Areal rainfall estimations are quite important for flood and erosion risk assessment as well for economic damage evaluation

- Geostatistical tools are worth noting to address these concerns through
  - Assessment of variogram functions to identify the variability structure of a given rainfall pattern (as a random field).
  - And to perform Kriging as optimal interpolation method

Variogram assessment

- Commonly, geostatistical approach relies on the identification of the variogram function $\gamma(h)$ which represents the mean quadratic deviation in the random field
- $\gamma(h) = \frac{1}{2} E[(Z(x)-Z(x+h))^2]$

- Sample variogram $\gamma exp(h)$:
- $\gamma exp(h) = \frac{1}{2} \sum[(Z(x)-Z(x+h))^2]/N(h)$
- N(h) number of pairs for a given class of distance with average value h.

- Robust estimator (Cressie, 1993) adopts the absolute deviation
- $\gamma rob(h) = \frac{1}{2} (1/(0.457+0.494/N(h))\sum[| (Z(x)-Z(x+h)) |]$

Drawbacks of variogram function

- However the assessment of rainfall spatial dependence attached to heavy rainfall events needs specific approaches because variogram analysis assumes that the marginal distribution of the underlying random field is Gaussian
- Rivoirard (1991) proposed disjunctive kriging which corresponds to a cokriging of indicator function as a method of non linear geostatistic.
- This method was addressed as alternative to kriging of logarithmic transformation of data because of problems rising from the back transformation to original data

- Rivoirard J. (1991) Introduction au krigeage disjonctif et à la géostatistique non linéaire. Cours Centre de géologie appliquée. École des Mines de Paris.

- The gaussian type of dependence underestimates the dependence of extremes

- Cooley (2005) suggested that «the madogram $\nu(h)$ conveniently links geostatistical ideas to measures of dependence for extremes »

□ $\nu(h) = ½ \sum |(Z(x)-Z(x+h))| / N(h)$

- and that «the madogram could be used to model the spatial dependence as a function of distance, much like the variogram»
- The madogram was proposed by Matheron (1987) in linkage with the variogram of the indicator variable

- KazienKa and Pilz (2010) mentioned well-known methods to deal with non- gaussianity (kriging of logarithmic transformation of data ; disjunctive kriging; rank-order transformation)
  - Kazienka H.., J. Pilz., (2010), Copula based geostatistical modeling of continuous and discrete data including covariates. Stoch Environ Risk Assess 24: 661-673.
- Smith (1990) proposed multivariable *t*-transformation as extreme value process

- Copula function has been proposed as alternative and Bardossy and Li (2008) proposed a copula function based on Khi-Deux transformation.
  - Bàrdossy A, Li J (2008) Geostatistical interpolation using copulas. Water Resour Res 44:W07412
- **They demonstrated that Indicator kriging ; disjunctive kriging; simple kriging of the rank-transformed data are Copula based spatial interpolation models**

joint distribution (bivariate) and copula

- Copula is defined as a distribution function with marginal distributions uniforms on (0,1).
- Sklar theorem (1959) links copulas and multivariate distributions F with margines F_1 and F_2.
- $C(u_1, u_2)=\text{prob}\ (F_1(X)\leq u_1, F_2(Y) \leq u_2)$
- If the distribution is continue then the copula is unique

Aim of this work

Achieve a quantitative analysis of tail distribution of heavy rainfall eventsusing geostatistical tools

– and Generalized extreme value (GEV) bivariate distribution

- **Case study**: Analysis of Tunis daily precipitations patterns. September 17th 2003

Data

- maximum annual daily precipitation recorded at 12 stations sample sizes (16 to 116 years).
- *precipitation* records on 17/9/2003 at 75 stations max = 187 mm/day

FIG 1

Fig. 1 Réseau et totaux du 17-9-2003 (mm)

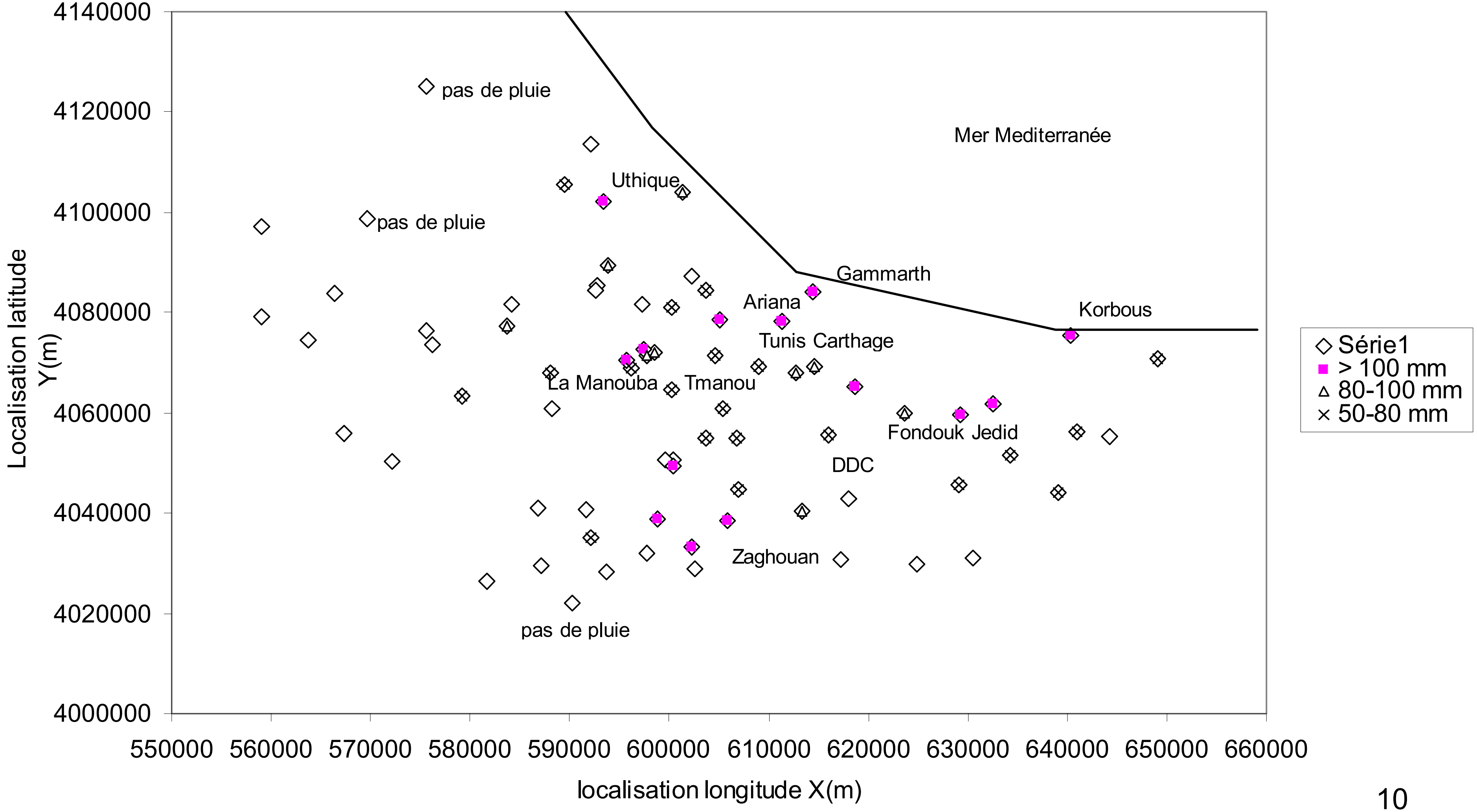
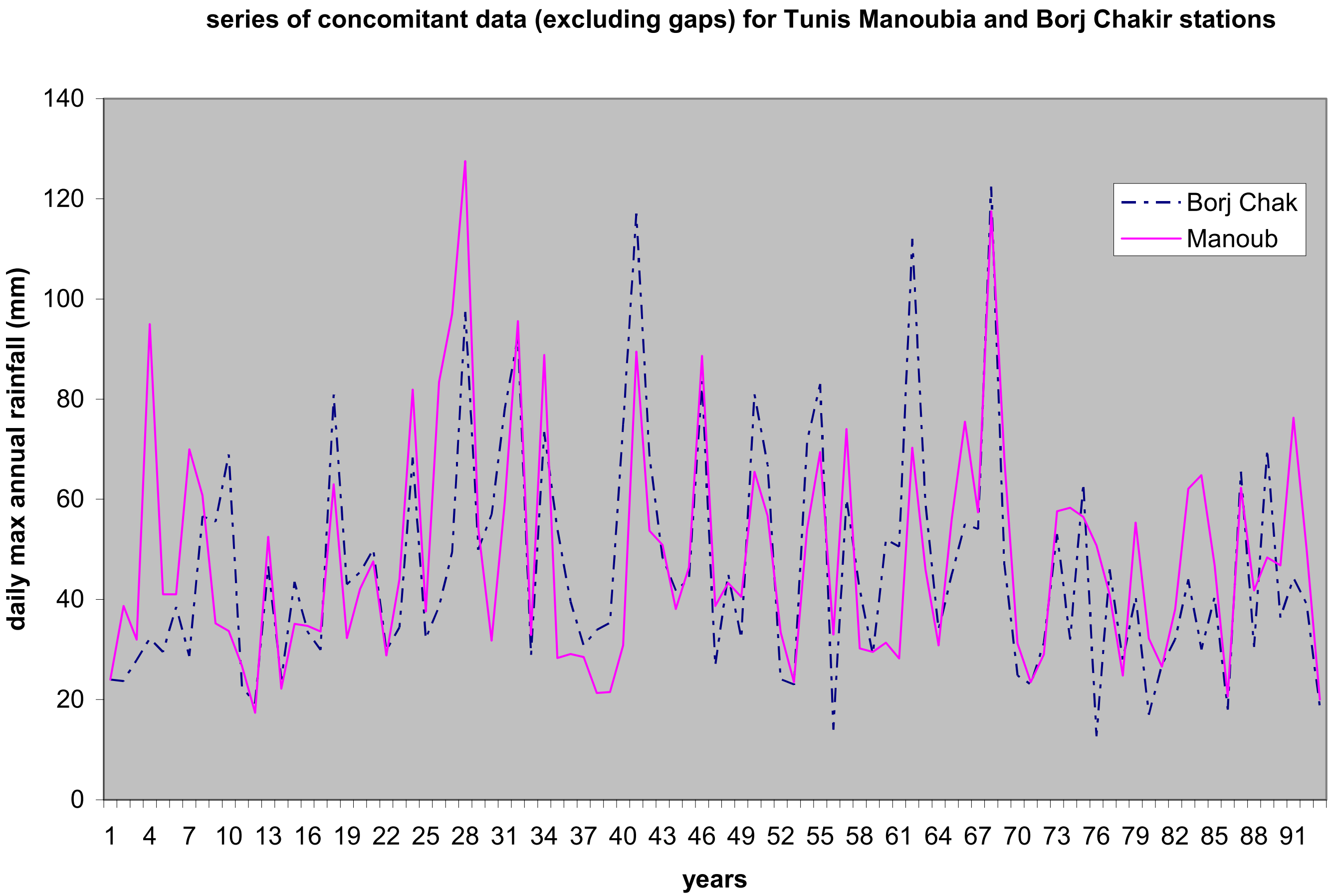


FIG 2


series of concomitant data (excluding gaps) for Tunis Manoubia and Borj Chakir stations
140
120
100
80
60
40
20
0
daily max annual rainfall (mm)
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91
years
Borj Chak
Manoub

A) Application of Extreme value theory

- $C(u,v) = \text{prob}\,(F(X) \le u,\ G(Y) \le v) = \exp\,[\log(uv)\ A\,(\log(u)/\log(uv))]$
- A is the Pickands dependence function
- (Gudendorf and Segers, 2009)

- This function has been extended in the bivariate case (Cooley;2005)
- Z(x) et Z(x+h)
- $C(u,v,h) = \text{prob}\,(F(Z(x)) \le u,\ F(Z(x+h)) \le v) = \exp\,[\log(uv)\ A\,(\log(u)/\log(uv),\ h)]$

Definition of extremal function θ

- $\theta(h)$ is defined as the tail of the joint bivariate distribution :
- **$Prob(F(Z(h)) \leq u, F(Z(x+h)) \leq u) = u^{\theta(h)}$**
- $\theta(h)$ gives a partial idea on the tail dependence structure (conversely to A);
- Extreme dependence for $\theta(h) = 1$;
- Extreme independence for $\theta(h) = 2$

Example of θ estimation

- Smith (1990) assumed local GEV model (μ,σ, ξ) for the variable X,
- he adopted a standardisation of data through the Frechet transformation

  $y = (1 - \xi(x - \mu)/\sigma))^{-1/\xi}$
- With $\Pr\{Yt \leq y\} = \exp(-1/y)$
- In this case, 1/Y has unit exponential distributions and
- 1/ $\max(y_1,y_2)$ has an exponential distribution

with mean 1/θ

- He plotted the extremal function as function of h.

Extremal dependence function and Pickands function

- In terms of the dependence function A, the extremal dependence function is (De Haan,1985)

□ **$\theta(h) = 2\,A(1/2,h)$ $(1 \leq \theta(h) \leq 2)$**

- **Etimation of A(t,h) and A(1/2,h) to derive** $\theta(h)$

Non parametric estimation of Hall and Tajvidi (2000) for A

- recommanded by Gudendorf and Segers (2009)

- $\varsigma_{HTi}(t) = \min [-\log(u)/LUM/ (1-t), -\log(v)/ LVM/ t]$; $0 \leq t \leq 1$
  - LUM average value of – log(u)
  - LVM average value of – log(v)
- Where u and v are Frechet transforms

$$1/A_{HTest}(t) = 1/k \sum_{i=1}^{k} \varsigma_{HTi}(t)$$

B) Analysis of scales of variability according to the results of the calibration of variograms and madograms

- Use of Simulated annealing (Monte Carlo simulation optimization method) with RMSE as calibration criteria.
- Changing the seed to initiate SA (replications) and multiplying the SA calibrations, it is possible to investigate the uncertainty about variogram (madogram) model.
- SA: annealing temperature decrease is according to a geometric distribution $T_n = T_0 \, (R_T)^n$
- Stopping criteria $T_n < T_{ratio}$ with $T_0 = 10^{-6}$
- R_T is computed according to the maximum number of iterations IT_{max} with no change in temperature
- $T_0 = 0.2$; $IT_{max} = 70$ so that $R_T = 0.7$.

C) Derivation of a regional GEV distribution assuming relationship between $\theta(h)$ and $\nu(h)$

- For the GEV distribution, $\theta(h)$ is related to the madogram (Cooley, 2005)
- So, $\theta(h)$ is modeled with a geostatistical spherical model.

- GEV $\xi \neq 0$:

□ $\theta(h)=[1+\xi\,\nu(h)/(\sigma\,\Gamma(1-\xi))]^{1/\xi}$

□ Γ Digamma function.

- Gumbel

□ $\theta(h) = \exp(\nu(h)/\sigma)$

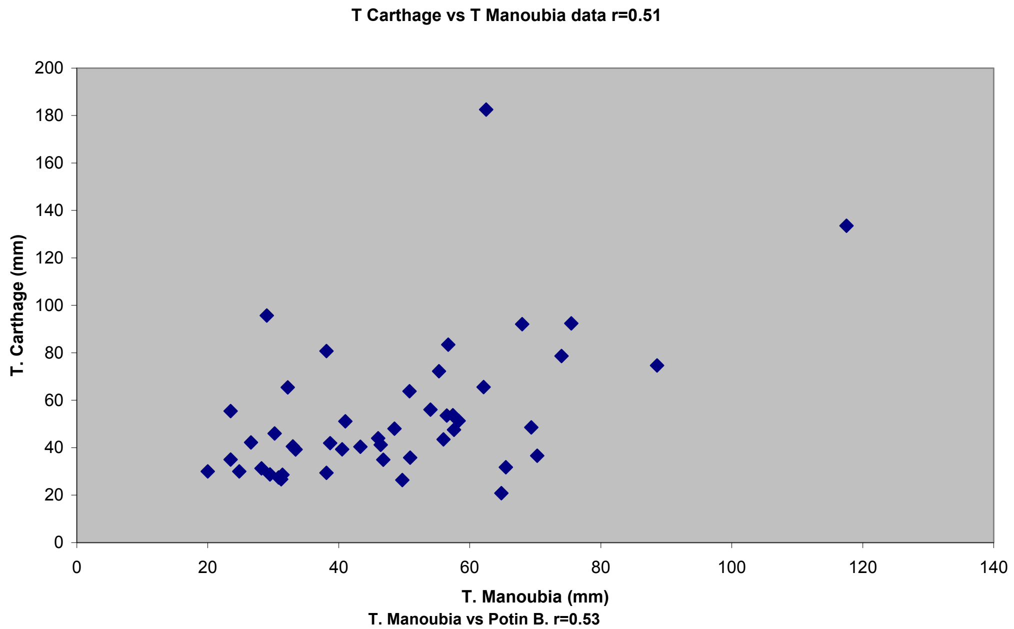
Results

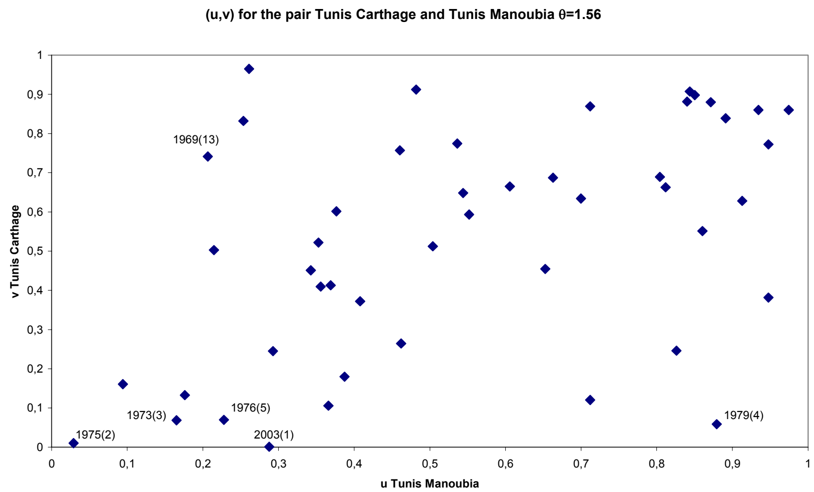
- Scatterplots (Fig. 2)
- Estimation of the dependence function (Fig. 3)
- Estimation and modeling of the extremale function (Fig. 4)
- Identification of the variogram and madogram of the rainfall field (Fig. 5a)
- Estimation of marginal GEV
- Identification of the variogram of GEV parameters separately (Fig. 5b,c)
- Comparison of scales of variability
- Mapping the conditional joint distribution for a given risk value (u=0.98) for a given location (Fig. 6)

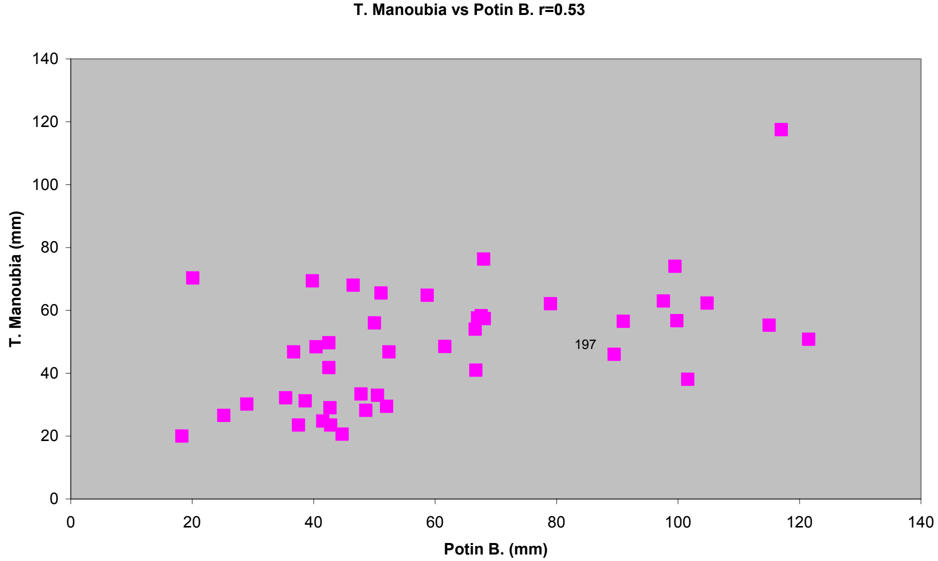
Practical aspects

- 1 km grid size (Number of nodes = 7793).
- inter distance for sampling variogram and madogram: 3.5 km.
- 18 classes of distances.

Scatterplots of data and Transformed data (FIG 2)







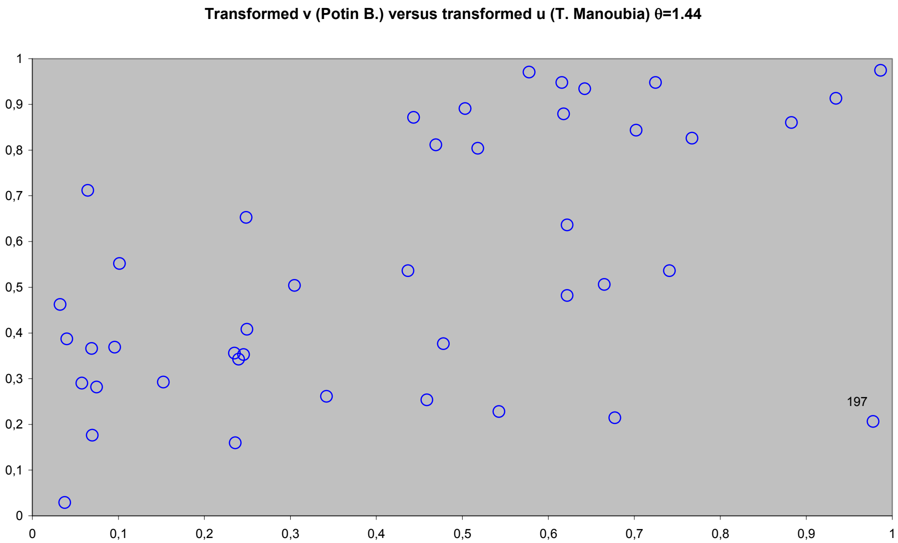
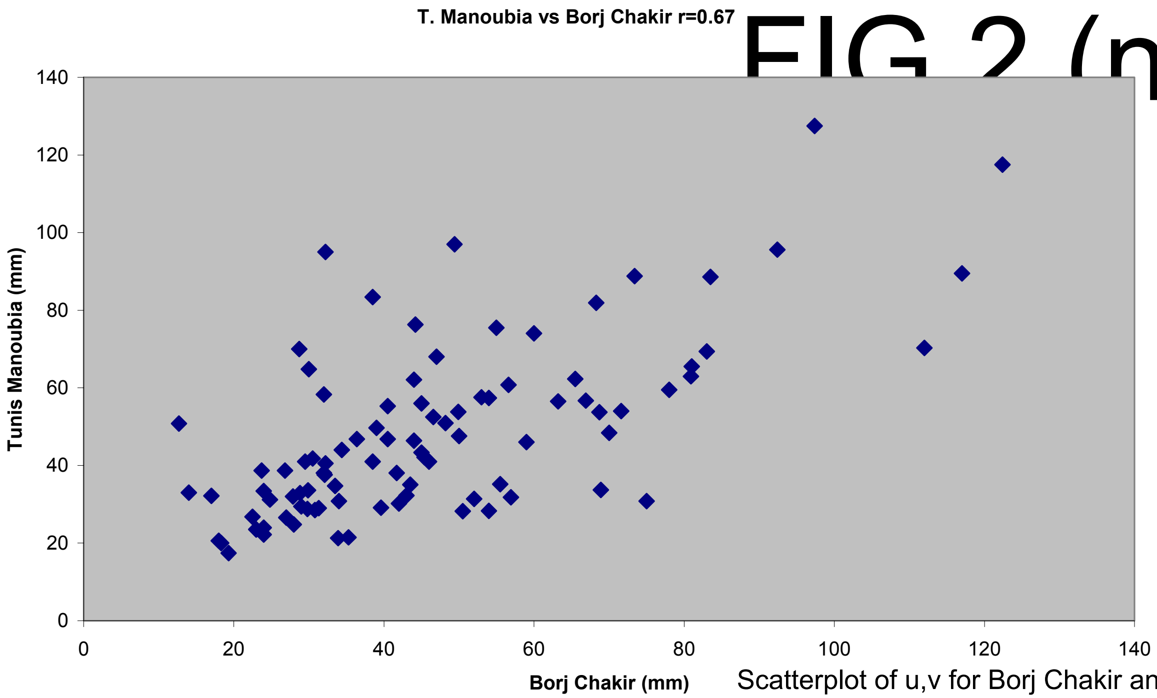


FIG 2 (next)



Scatterplot of u,v for Borj Chakir and Tunis Manoubia stations

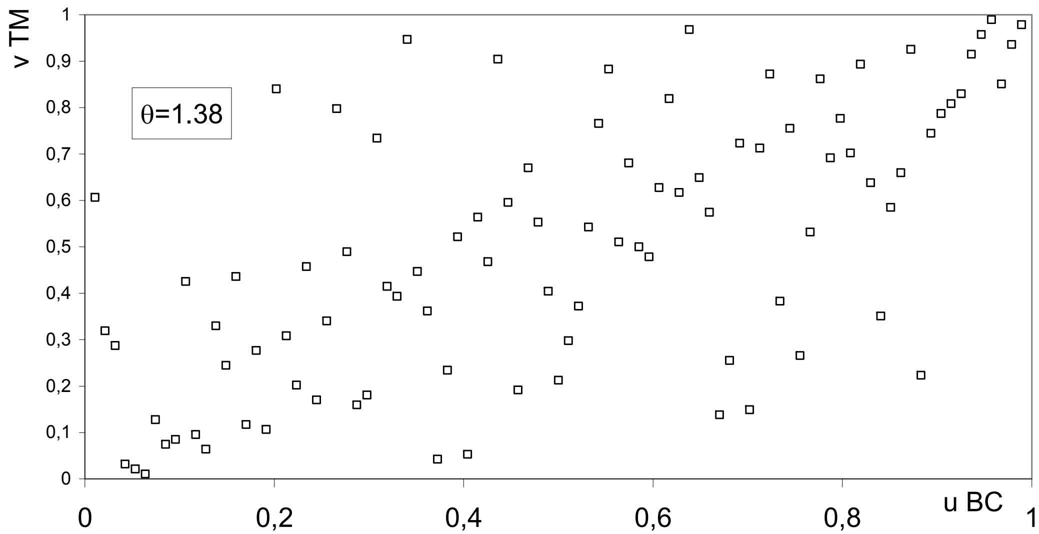


Fig. 3

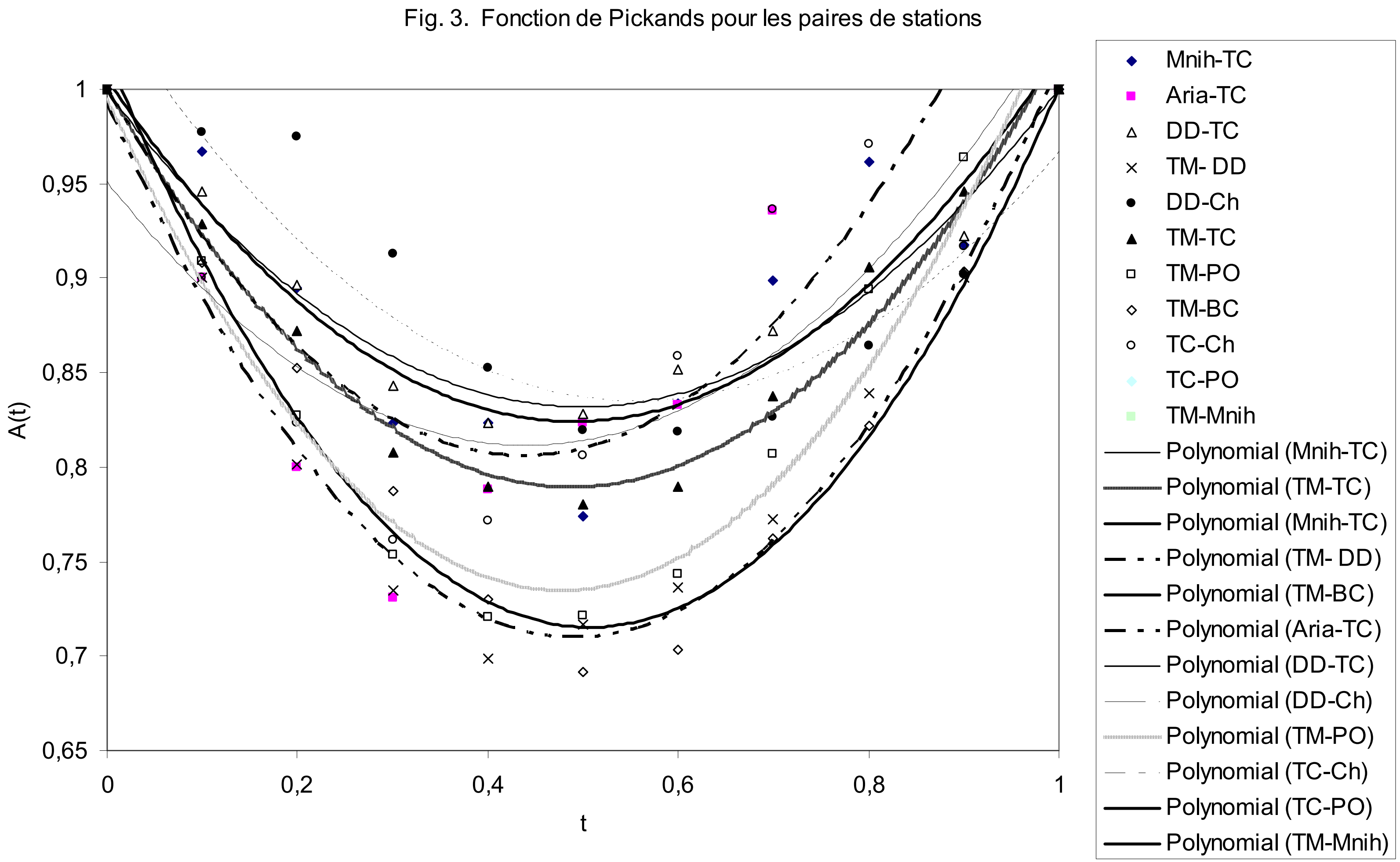
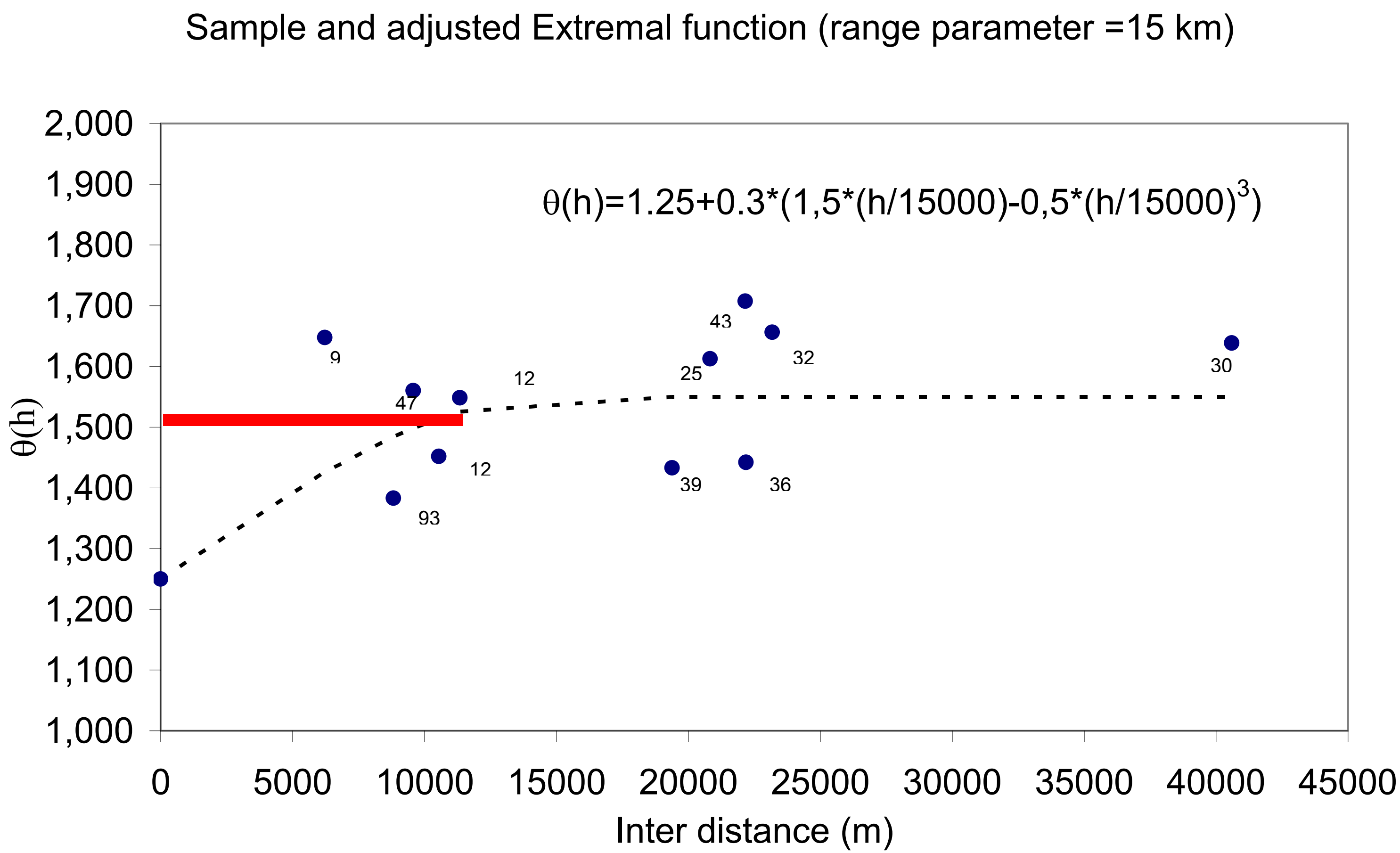


Fig. 3. Fonction de Pickands pour les paires de stations
A(t)
1
0,95
0,9
0,85
0,8
0,75
0,7
0,65
0
0,2
0,4
0,6
0,8
1
t
Mnih-TC
Aria-TC
DD-TC
TM- DD
DD-Ch
TM-TC
TM-PO
TM-BC
TC-Ch
TC-PO
TM-Mnih
Polynomial (Mnih-TC)
Polynomial (TM-TC)
Polynomial (Mnih-TC)
Polynomial (TM- DD)
Polynomial (TM-BC)
Polynomial (Aria-TC)
Polynomial (DD-TC)
Polynomial (DD-Ch)
Polynomial (TM-PO)
Polynomial (TC-Ch)
Polynomial (TC-PO)
Polynomial (TM-Mnih)

Fig. 4 (fitting of a spherical model; 10 km for extreme independency)



Partial conclusions on scales of variability according to 10 replicates of SA

- Classic and robust variograms lead to different results for the rainfall fied and the scale parameter σ field but there is insensitivity for ξ.
- Scales of variability (according to the decorrelation distance)
- Range parameter for ξ and rainfall field are of the same order (optimal values: 17 to 20 km; fluctuation of accepted solutions between 12 and 24 km).
- Conversely, for σ optimal range is 40 km and fluctuation is larger : 30 to 99 km).
- *No study of μ yet*

FIG 5

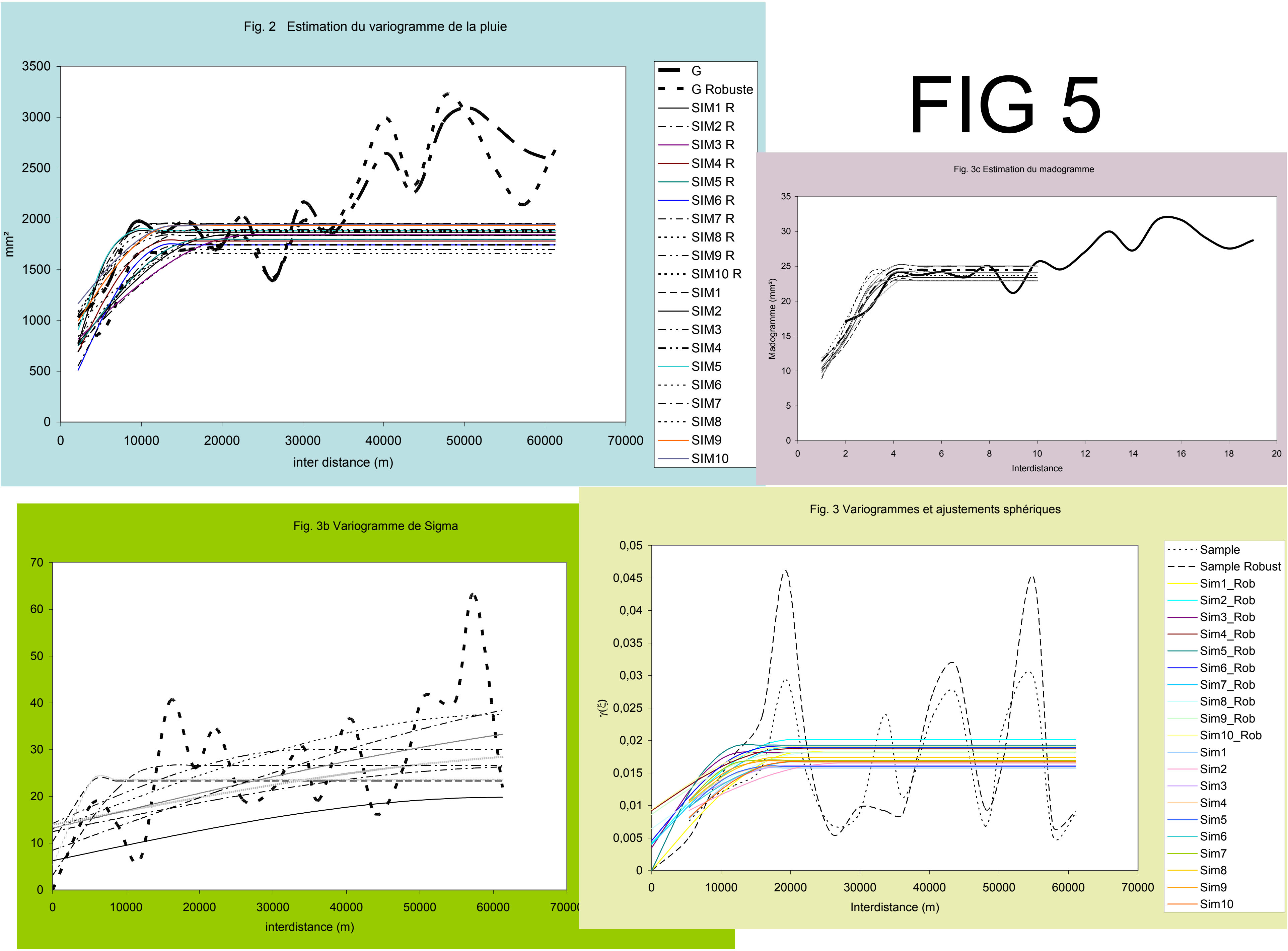


Fig. 2 Estimation du variogramme de la pluie
mm²
inter distance (m)
0
500
1000
1500
2000
2500
3000
3500
0
10000
20000
30000
40000
50000
60000
70000
G
G Robuste
SIM1 R
SIM2 R
SIM3 R
SIM4 R
SIM5 R
SIM6 R
SIM7 R
SIM8 R
SIM9 R
SIM10 R
SIM1
SIM2
SIM3
SIM4
SIM5
SIM6
SIM7
SIM8
SIM9
SIM10
Fig. 3c Estimation du madogramme
Madogramme (mm²)
Interdistance
0
5
10
15
20
25
30
35
0
2
4
6
8
10
12
14
16
18
20
Fig. 3b Variogramme de Sigma
interdistance (m)
0
10
20
30
40
50
60
70
0
10000
20000
30000
40000
50000
60000
70000
Fig. 3 Variogrammes et ajustements sphériques
γ(ξ)
Interdistance (m)
0
0,005
0,01
0,015
0,02
0,025
0,03
0,035
0,04
0,045
0,05
0
10000
20000
30000
40000
50000
60000
70000
Sample
Sample Robust
Sim1_Rob
Sim2_Rob
Sim3_Rob
Sim4_Rob
Sim5_Rob
Sim6_Rob
Sim7_Rob
Sim8_Rob
Sim9_Rob
Sim10_Rob
Sim1
Sim2
Sim3
Sim4
Sim5
Sim6
Sim7
Sim8
Sim9
Sim10

Scales of variability according to atmosphere dynamics

L_H	T	Stull	Pielke	Orlanski	Phénomènes atmosphériques
		macro	synoptique	macro-α	Circulation générale
2000km	1 mois				
		macro	synoptique	macro-β	Cyclones
200km	1 semaine				
		meso	meso	meso-α	Fronts
20km	1 jour				
		meso	micro	meso-β	Vents de montagne brises
2km	1 heure				
		meso/micro	micro	meso-γ	Circulations urbaines
200m	30 minutes				
		micro	micro	micro-α	cumulus
20m	1 minute				
		micro	micro	micro-β	panaches (cheminées), micro-turbulence

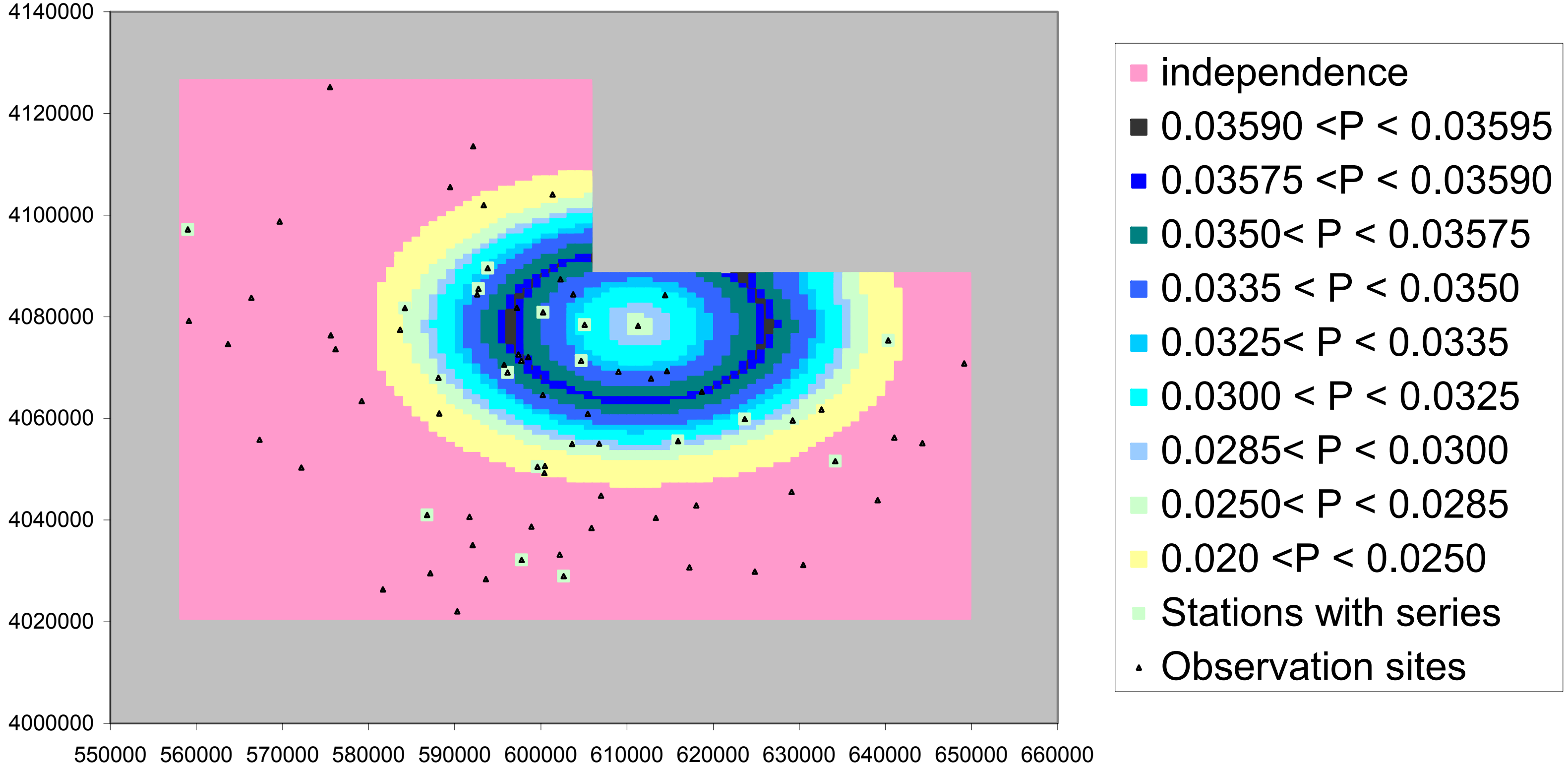
Variability of ξ and extreme 2003 rainfall field: micro scale; local effects

Variability of σ: mesoscale; dictated by General atmospheric circulation

Variability of θ (10 km) microscale effect

FIG. 6 (using the θ spherical model)

The tail of the joint bivariate GEV distribution Prob(F(Z(h))≤u, G(Z(x+h)) ≤u) = $u^{q(h)}$; Z(h):Tunis Carthage (Airport) station; u=0.98



Conclusions

- raw estimates of the extremal coefficients (not based on any model) were derived
- Estimation of standard error of θ may be achieved through Jackknife procedure (by omitting observations of a year, year by year)
- Attention is restricted to extremes based on annual maxima but the development of threshold methods should be investigated