



ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ

Δημοκρίτειο Πανεπιστήμιο Θράκης ΠΜΣ Διαχείριση Υδρομετεωρολογικών Καταστροφών

Ζητήματα αβεβαιότητας, και διαχείριση της, στην υδρολογία και τους υδατικούς πόρους

Μοντελοποίηση, Πρόβλεψη και Διαχείριση Κινδύνων Πλημμυρών

Τσουκαλάς Γιάννης - Δρ. Πολιτικός Μηχανικός ΕΜΠ

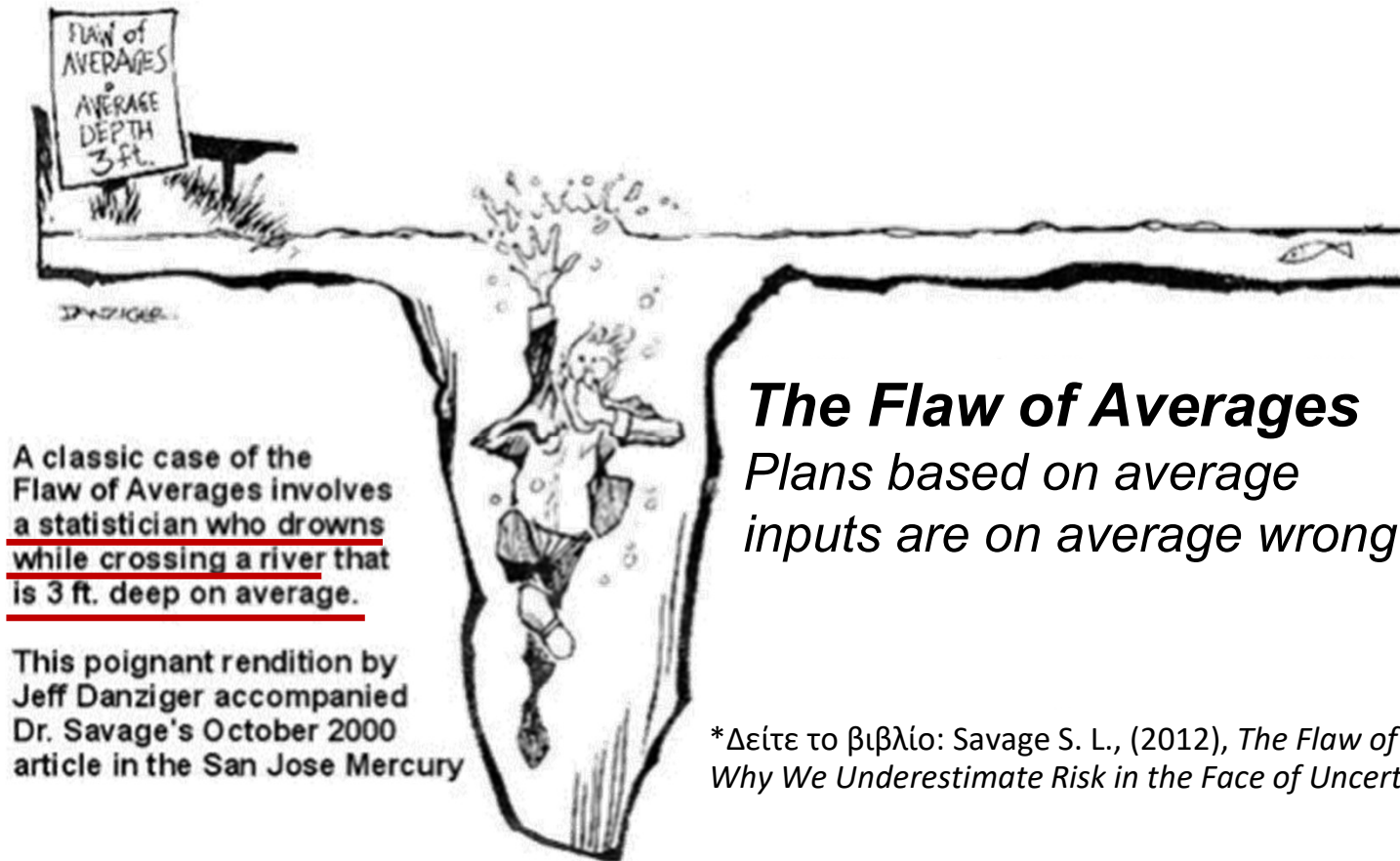
Επίκουρος Καθηγητής

Τμήμα Πολιτικών Μηχανικών ΔΠΘ

e-mail: itsoukal@civil.duth.gr

Αβεβαιότητα...

... Η απάντηση στο γιατί τόσα σχέδια/πλάνα βγαίνουν τελικά εκτός προγράμματος, εκτός προϋπολογισμού, και εκτιμήσεων.



*Δείτε το βιβλίο: Savage S. L., (2012), *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*

Κίνητρο

Πως μπορεί να βελτιωθεί η λήψη αποφάσεων όσον αφορά τον σχεδιασμό, τη διαχείριση και τη λειτουργία σύνθετων συστημάτων υδατικών πόρων;

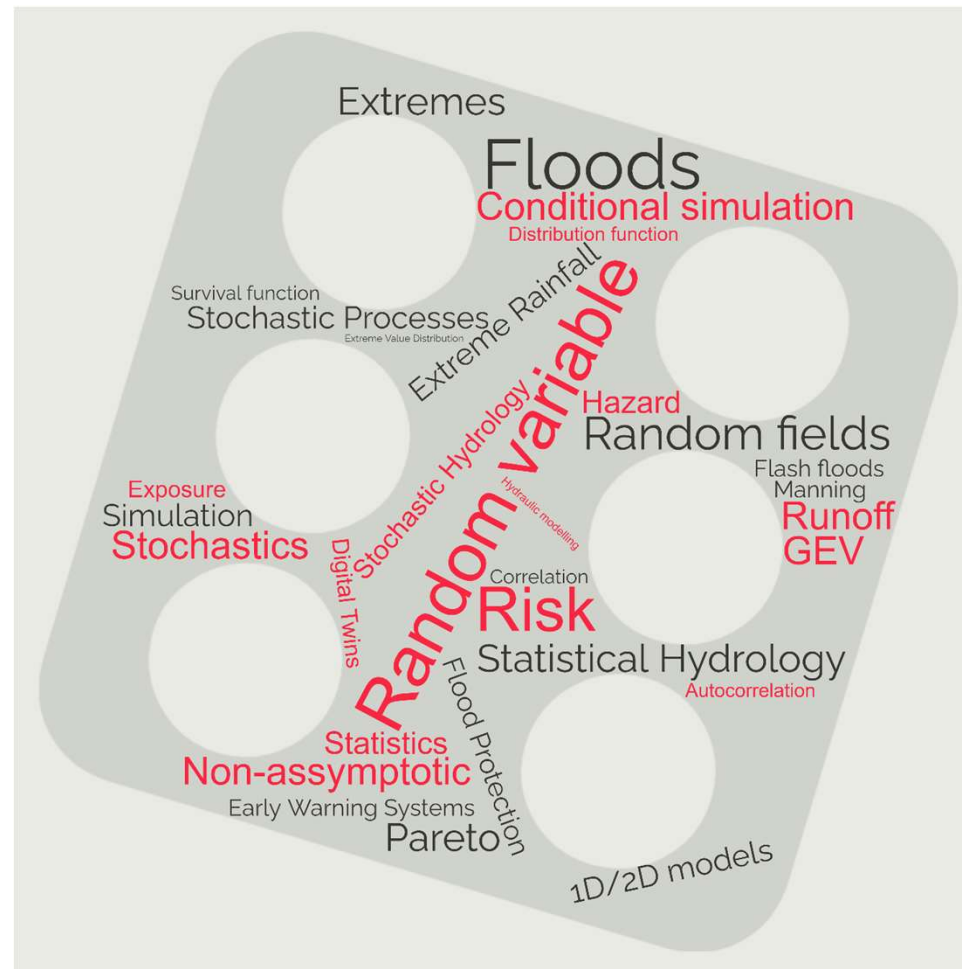
Με τέτοιο τρόπο όπου:

- Λαμβάνεται υπόψιν η **αβεβαιότητα**,
- Λαμβάνονται υπόψιν **εναλλακτικές μελλοντικές συνθήκες**,
- Λαμβάνεται υπόψιν η **βέλτιστη λύση**,
- Λαμβάνεται υπόψιν οι (τυπικές) **υπολογιστικές δυνατότητες**,
- Λαμβάνεται υπόψιν η **επεξηγησιμότητα** (explainability).

Βασικοί μεθοδολογικοί άξονες

- Ανάλυση εμπειρικών δεδομένων,
 - Θεωρητική τεκμηρίωση,
 - Χρήση μοντέλων και τεχνικών προσομοίωσης.
-

Ποια είναι τα εργαλεία για να αντιμετωπίσουμε την αβεβαιότητα;



Γιατί χρειαζόμαστε και τη στατιστική και τα «στοχαστικά»; Δεν αρκούν η τεχνητή νοημοσύνη (AI) και η μηχανική μάθηση (ML);

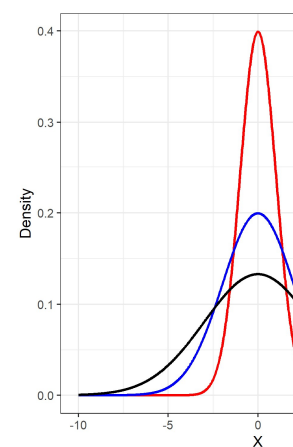
- Εξ ορισμού η «*γλώσσα της αβεβαιότητας*»: πιθανότητες, στατιστική και στοχαστικές μέθοδοι (στο χωροχρόνο).
- Επιτρέπουν τη δημιουργία ενός άπειρου αριθμού πιθανών πραγματοποιήσεων (και, ως εκ τούτου, εναλλακτικών μελλοντικών συνθηκών).
- Επιτρέπουν την ανάπτυξη μεθοδολογικών πλαισίων/προσεγγίσεων για τη λήψη βέλτιστων αποφάσεων σε συνθήκες αβεβαιότητας.
- Υπολογιστικά εφικτές μέθοδοι.
- Δεν απαιτούνται πρόσθετα εργαλεία ή μέθοδοι για την εξήγηση των προβλέψεων/προσομοιώσεων – σε αντίθεση με τις προσεγγίσεις black-box AI/ML.



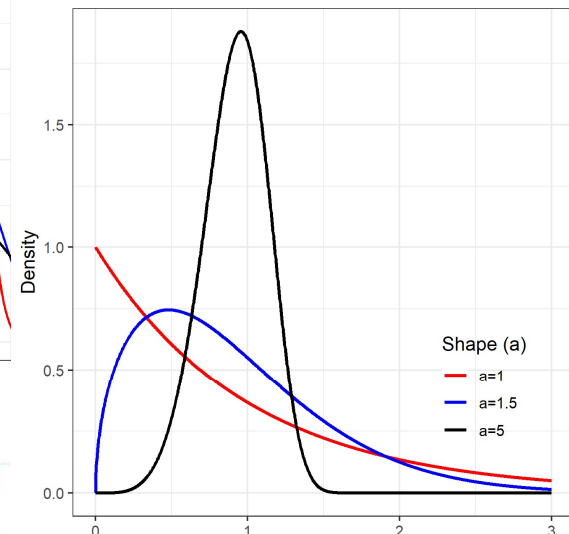
«Κλασσικά» μοντέλα κατανομών

- Ομοιόμορφη (Uniform)
- Κανονική/Γκαουσιανή (Normal/Gaussian)
- Exponential
- Gamma
- Weibull
- Generalized Extreme Value (extreme value)
- Gumbel (extreme value)
- Generalized Pareto (extreme value)
- Log-Normal
- Logistic
- Beta (φραγμένη στο $(0,1)$)
- Kumaraswamy (φραγμένη στο $(0,1)$)
- Bernoulli (διακριτή)
- Poisson (διακριτή)
- Binomial (διακριτή)

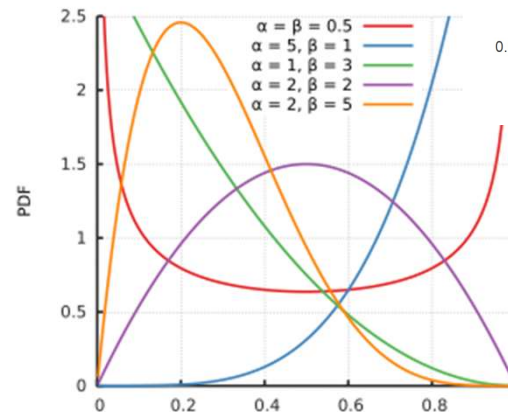
Gaussian



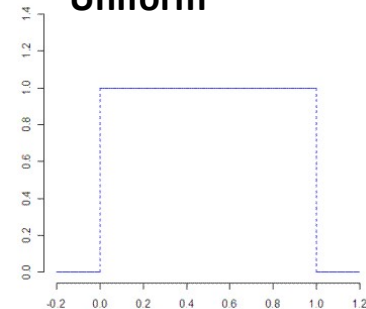
Weibull



Beta



Uniform



Περισσότερες (και πιο ευέλικτες) κατανομές

- Ικανές να περιγράψουν μεγάλο εύρος φαινομένων (φυσικών ή μη).

- Κατάλληλες για φυσικές διεργασίες αφού $X \in (0, \infty)$

Generalized Gamma

$$f_{GG}(x|b, a_1, a_2) = \frac{a_2}{b\Gamma(a_1/a_2)} \left(\frac{x}{b}\right)^{a_1-1} \exp\left(-\left(\frac{x}{b}\right)^{a_2}\right), \quad x > 0$$

Burr Type XII

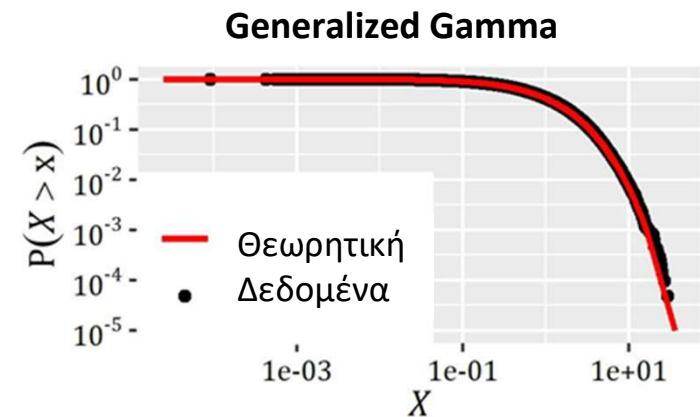
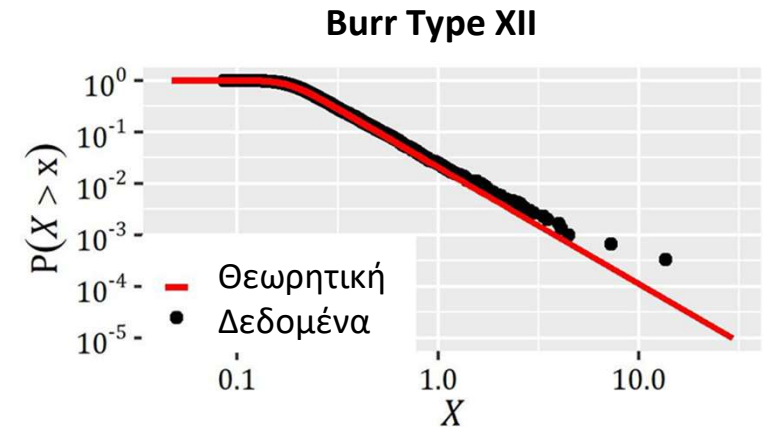
$$F_{Br_{XII}}(x|b, a_1, a_2) = 1 - \left(1 + \left(\frac{x}{b}\right)^{a_1}\right)^{-a_2}, \quad x > 0$$

Pareto II (γνωστή και ως Lomax) = $Br_{XII}(x|b, a_1 = 1, a_2)$

$$F_{PII}(x|b, a) = 1 - \left(1 + \frac{x}{b}\right)^{-a}, \quad x \geq 0$$

b : παράμετρος κλίμακας | a_i : παράμετρος σχήματος

Ενδεικτικά, εργασίες που χρησιμοποιούν τις παραπάνω κατανομές με αντικείμενο τη στατιστική/στοχαστική υδρολογία (π.χ., *Hao and Singh, 2008; Grimaldi et al., 2011; Tsoukalas et al., 2019; 2020*).



Διακριτή-συνεχής (μικτή) κατανομή με άτομο/ασυνέχεια στο μηδέν

Κατανομή με άτομο/ασυνέχεια στο μηδέν:

$$F_X(x) = \begin{cases} p_0, & x \leq 0 \\ p_0 + (1 - p_0)G_{\underline{X}}(x), & x > 0 \end{cases}$$

Ιδιαίτερα χρήσιμη για **διεργασίες** που χαρακτηρίζονται από **διαλείπουσα συμπεριφορά**.

(π.χ., Bell, 1987; Tsoukalas et al., 2019; 2022)

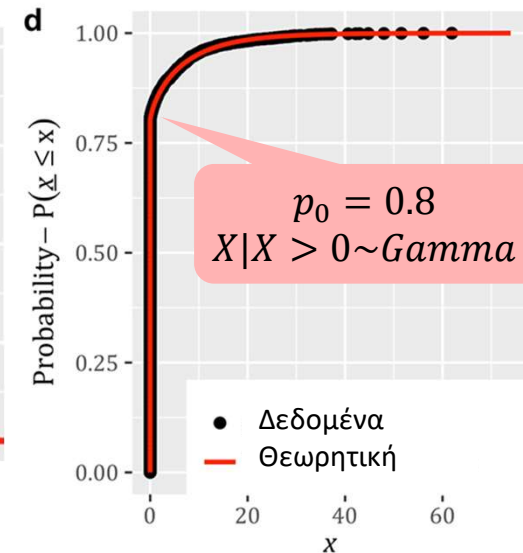
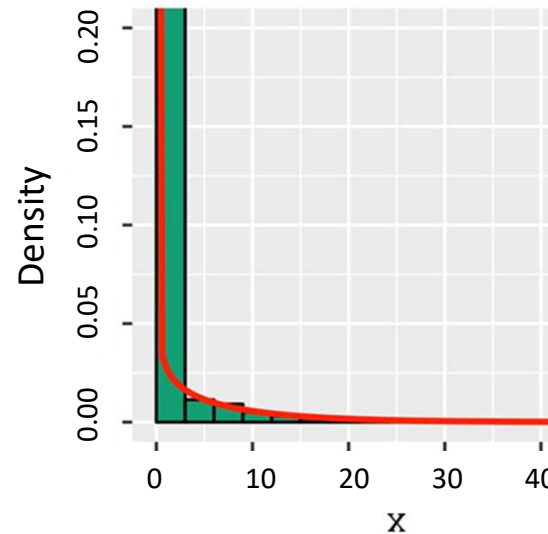
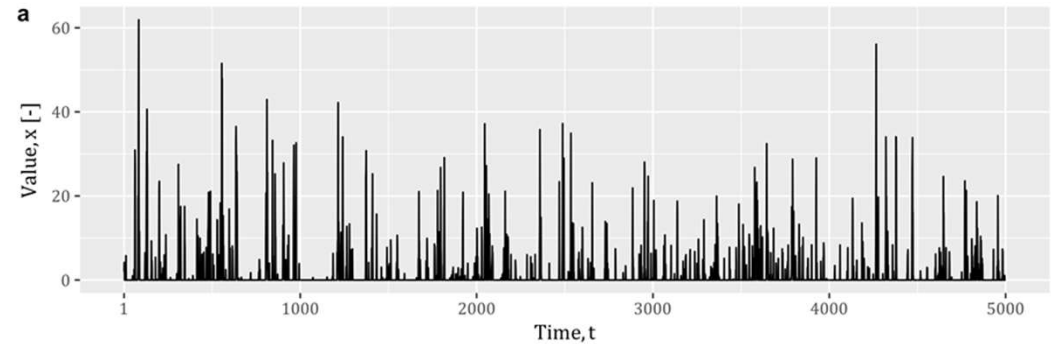
Πιθανότητα μηδενικής (και μη μηδενικής) τιμής

$$p_0 = P(X = 0) = 1 - p_1$$

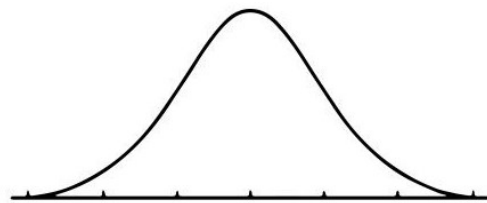
$$p_1 = P(X > 0) = 1 - p_0$$

Συνεχής κατανομή (θετικές τιμές)

$$G_X(x) := F_{X|X>0}(x) = P(X \leq x | X > 0)$$



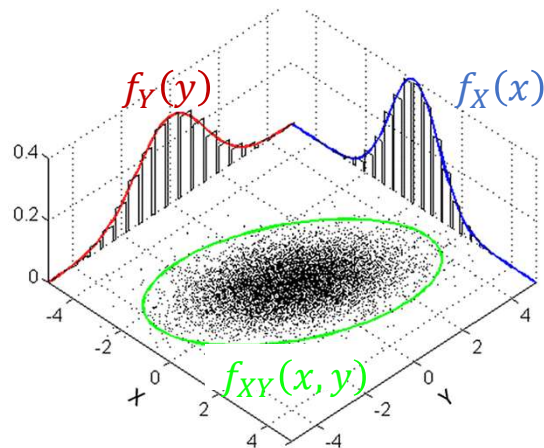
... και αν έχουμε δύο ή περισσότερες τυχαίες μεταβλητές;



NORMAL DISTRIBUTION



PARANORMAL DISTRIBUTION



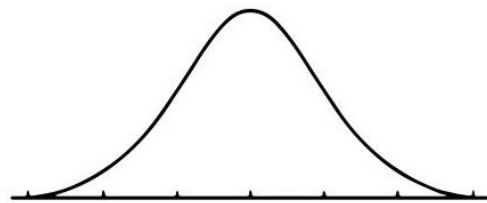
?

Συσχετισμένες μη-Γκαουσιανές τ.μ.

*Oh, Lord, please keep the world **linear** and **Gaussian**.*

~Chester Kisiel's [1967] pray to the theoretical hydrologist, Klemeš [1997 (p. 288)]

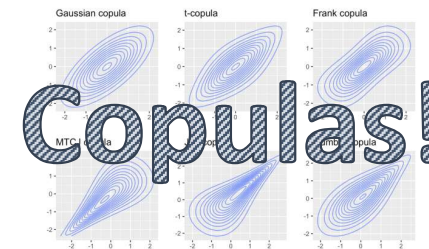
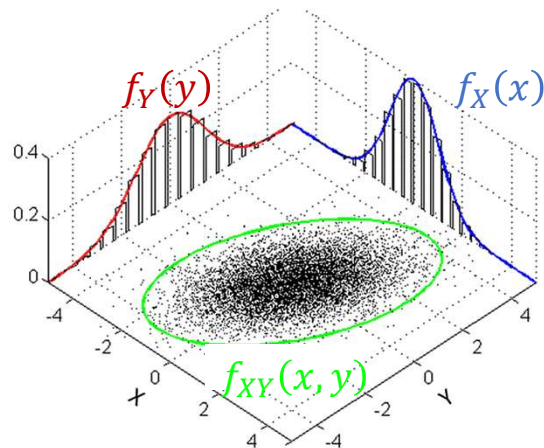
... και αν έχουμε δύο ή περισσότερες τυχαίες μεταβλητές;



NORMAL DISTRIBUTION



PARANORMAL DISTRIBUTION



Συσχετισμένες μη-Γκαουσιανές τ.μ.

*Oh, Lord, please keep the world **linear** and **Gaussian**.*

~Chester Kisiel's [1967] pray to the theoretical hydrologist, Klemeš [1997 (p. 288)]

Συναρτήσεις συζεύξεων (copula functions)

Τι είναι οι συναρτήσεις ζεύξης
(**copula** functions);

Με απλά λόγια, είναι μαθηματικά «εργαλεία» που επιτρέπουν την κατασκευή από-κοινού κατανομών (με 2 ή περισσότερες μεταβλητές) **μοντελοποιώντας ξεχωριστά τη δομή εξάρτησης** (πολυμεταβλητές **συναρτήσεις συζεύξεων** με ομοιόμορφη περιθώρια κατανομή) και τις περιθώριες κατανομές.

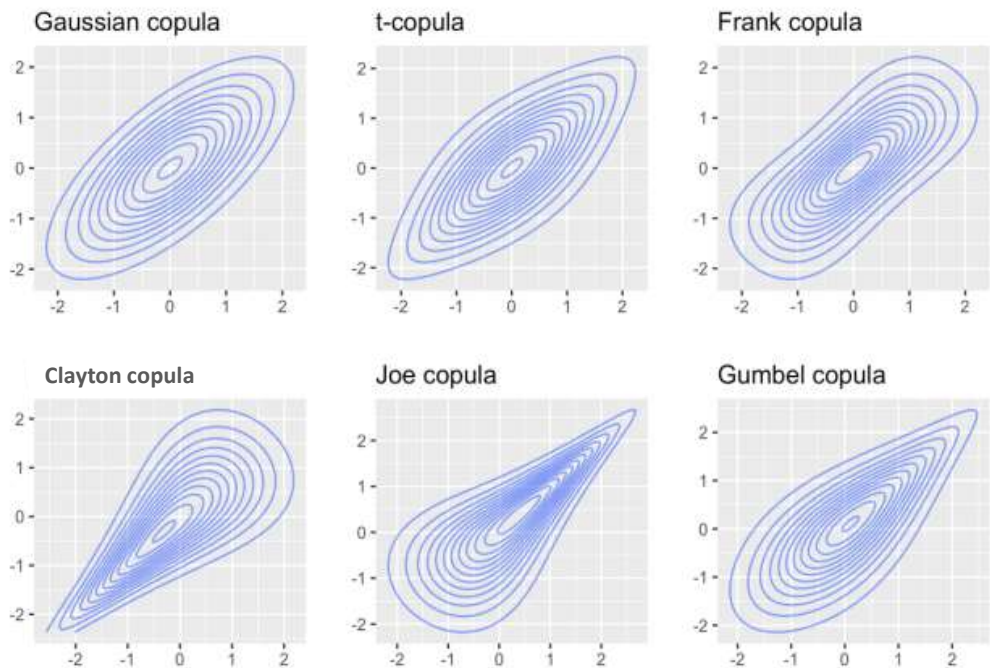
Με τη βοήθεια των **copula**, η συνάρτηση κατανομής $F_{X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$ δίνεται από τη σχέση:

$$F_{X_n}(x_1, \dots, x_n) = \mathcal{C}(u_1, \dots, u_n; \theta)$$

όπου $u_i = F_{X_i}(x_i)$, είναι ομοιόμορφα κατανομημένες τ.μ. στο διάστημα (0,1), ενώ το διάνυσμα θ εμπεριέχει τις παραμέτρους της συνάρτησης **copula** (συνήθως 1 ή 2) $\mathcal{C}(\cdot, \dots, \cdot)$.

Παράδειγμα

Έξι πολυμεταβλητές κατανομές, **όλες** με Γκαουσιανές (Gaussian/Normal) περιθώριες κατανομές, αλλά διαφορετική συνάρτηση σύζευξης (copula).



Προσομοίωση συσχετισμένων μη-Γκαουσιανών τυχαίων μεταβλητών

Παράδειγμα

Έστω ότι έχουμε 3 τ.μ. με κατανομές F_{X_i} :

$X_1 \sim \text{Gamma}$

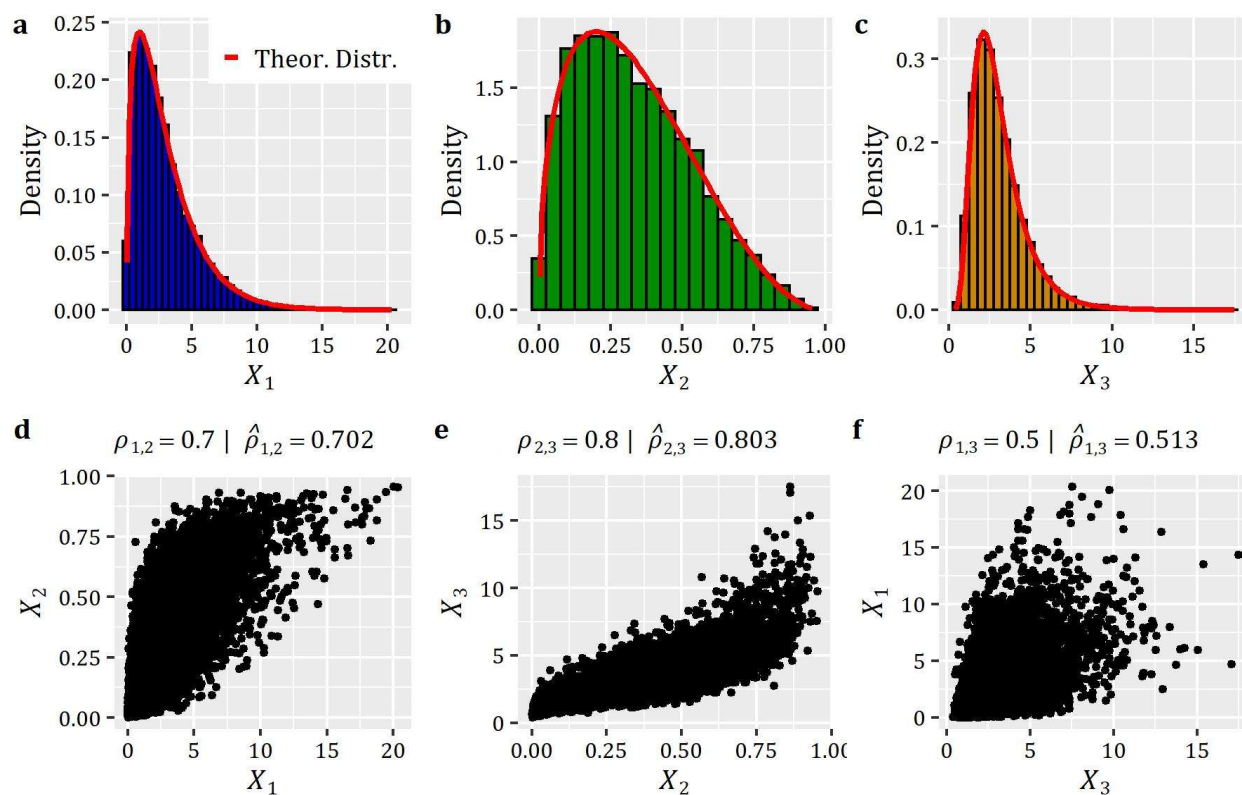
$X_2 \sim \text{Beta}$

$X_3 \sim \text{log-Normal}$

Ωστόσο, οι τ.μ. είναι
συσχετισμένες μεταξύ τους:

$$R = \begin{matrix} & \begin{matrix} X_1 & X_2 & X_3 \end{matrix} \\ \begin{matrix} X_1 \\ X_2 \\ X_3 \end{matrix} & \begin{bmatrix} 1 & 0.7 & 0.5 \\ 0.7 & 1 & 0.8 \\ 0.5 & 0.8 & 1 \end{bmatrix} \end{matrix}$$

Γένεση πραγματοποιήσεων συσχετισμένων μη-Γκαουσιανών τ.μ.



Δεσμευμένες μη-Γκαουσιανές κατανομές

Η χρήση των copula μπορεί να επεκταθεί και για την εξαγωγή δεσμευμένων (conditional) κατανομών (βλ. Tsoukalas [2018]).

Υπενθυμίζεται ότι:

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)) = C(u_X, u_Y)$$

όπου $F_X(x) = u_X$ και $F_Y(y) = u_Y$ οι περιθώριες CDF των τ.μ. X και Y .

Στην περίπτωση του Gaussian copula ισχύει ότι: $C(u_X, u_Y) = \Phi_2(u_X, u_Y; \theta)$, όπου Φ_2 η τυποποιημένη πολυμεταβλητή κανονική κατανομή 2 μεταβλητών.

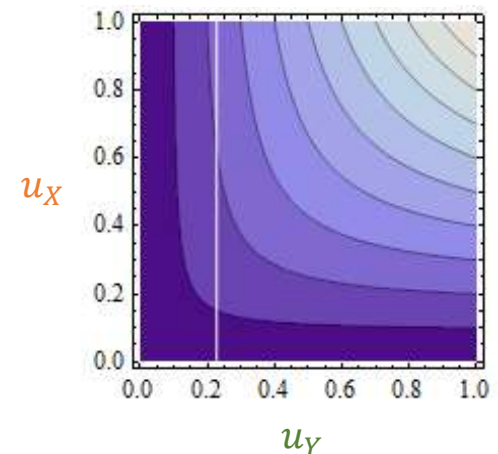
Conditional Gaussian Copula CDF

$$P(X \leq x | Y = y) = \Phi\left(\frac{\Phi^{-1}(u_X) - \theta \Phi^{-1}(u_Y)}{\sqrt{1 - \theta^2}}\right) := \alpha$$

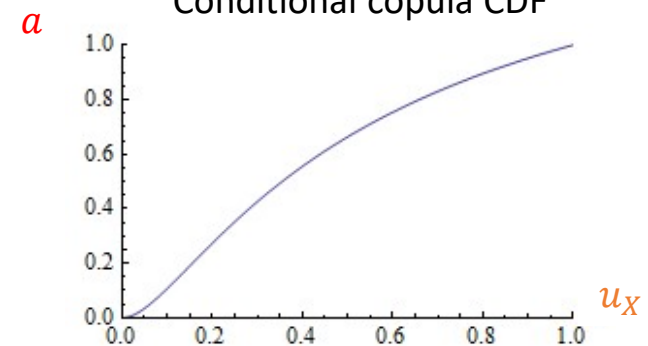
Conditional Gaussian Copula ICDF (conditional Quantile function)

$$x^{\alpha|Y=y} = F_{X|Y=y}^{-1}(\alpha) = F_X^{-1}\left(\Phi\left(\theta \Phi^{-1}(u_Y) + \sqrt{1 - \theta^2} \Phi^{-1}(\alpha)\right)\right)$$

Copula joint CDF



Conditional copula CDF

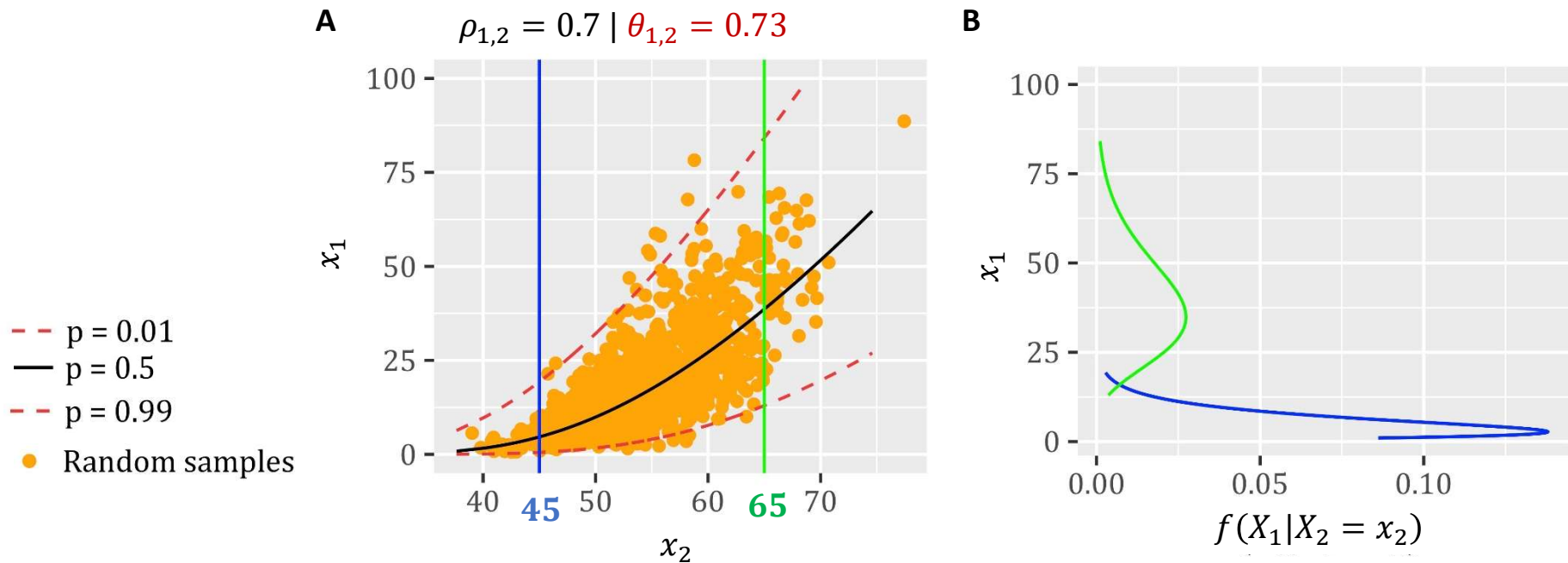


Προσομοίωση δεσμευμένων μη-Γκαουσιανών τυχαίων μεταβλητών

Παράδειγμα

Έστω $X_1 \sim \text{Gamma}(2, 10)$ και $X_2 \sim \text{logNormal}(0.10, 4)$ με $\rho_{1,2} = 0.7$, και Gaussian copula ($\theta_{1,2} = 0.73$).

(A) Ποσοστημόρια για $p = \{0.01, 0.5 \text{ και } 0.99\}$. (B) Δεσμευμένη PDF για $X_2 := x_2 = 45$ και $X_2 := x_2 = 65$.



Προβλήματα, προκλήσεις και ευκαιρίες!

Τα υδρομετεωρολογικά δεδομένα χαρακτηρίζονται από:

- **έλλειψη πληροφορίας** (π.χ. μετρήσεις) στην χρονική/χωρική **κλίμακα ενδιαφέροντος**,
- την ύπαρξη **ελλειπουσών τιμών**, και
- το **περιορισμένο μήκος των ιστορικών** δεδομένων.

Χαρακτηριστικά (και προκλήσεις!) υδρομετεωρολογικών διεργασιών

- Μη-Γκαουσιανή κατανομή
- Διαλείπουσα φύση (π.χ., βροχόπτωση)
- Χώρο-χρονική συσχέτιση
- Περιοδικότητα
- Διαφορετική πιθανοτική συμπεριφορά ανά χρονική κλίμακα συνάθροισης (δηλ. πρόβλημα συναθροισμένης κατανομής)

Προβλήματα και προκλήσεις που αποτελούν ακόμη σήμερα αντικείμενο έρευνας:

1. (Στοχαστική) προσομοίωση υδρομετεωρολογικών διεργασιών,
2. Συμπλήρωση κενών (συμπεριλαμβανομένης της χωρικής παρεμβολής),
3. Ποσοτικοποίηση της προγνωστικής αβεβαιότητας μοντέλων
4. Χωροχρονικός καταβιβασμός δεδομένων.

Τυπικές *εργασίες* σχεδόν σε όλα τις μελέτες και έργα μηχανικού!

Στοχαστική προσομοίωση υδρομετεωρολογικών διεργασιών

Χαρακτηριστικά φυσικών (και μη) διεργασιών

- Μη-Γκαουσιανή κατανομή
- Διαλείπουσα φύση (π.χ., βροχόπτωση)
- Χώρο-χρονική συσχέτιση
- Περιοδικότητα
- Διαφορετική πιθανοτική συμπεριφορά ανά χρονική κλίμακα συνάθροισης (δηλ. πρόβλημα συναθροισμένης κατανομής)

Πολλά μοντέλα!

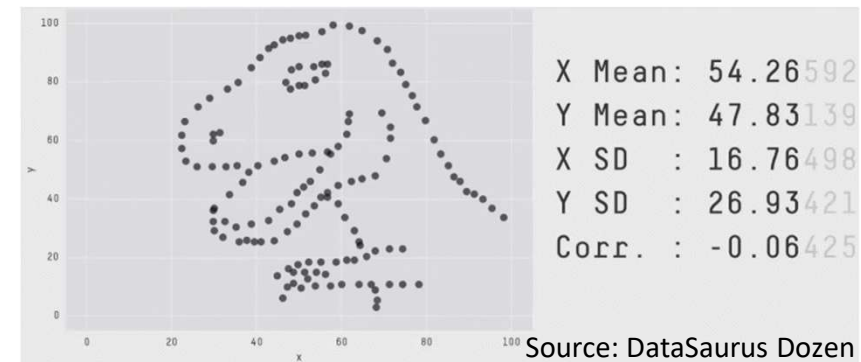
Μοντέλα στοχαστική προσομοίωσης

- Linear stochastic models
- Point process models
- Two-part models
- Resampling models
- **Copula-based models**

Συνήθης υπόθεση μοντέλων συνθετικών χρονοσειρών

Διατήρηση **βασικών στατιστικών μεγεθών** (π.χ., μέση τιμή, διασπορά, ασυμμετρία, και συσχετίσεις χαμηλής τάξης) των **ιστορικών δεδομένων** ([Matalas and Wallis, 1976](#)).

Ωστόσο κάτι τέτοιο **δεν** συνεπάγεται τη **διατήρηση της πιθανοτικής κατανομής τους**.



Κάτι που χαρακτηριστικά αναφέρεται από τους, [Klemeš and Borůvka \(1974\)](#):

“Simulation of a serially correlated series with a given marginal distribution is one of the important prerequisites of synthetic hydrology and of its applications to analysis of water resource systems”.

και Προσοχή...



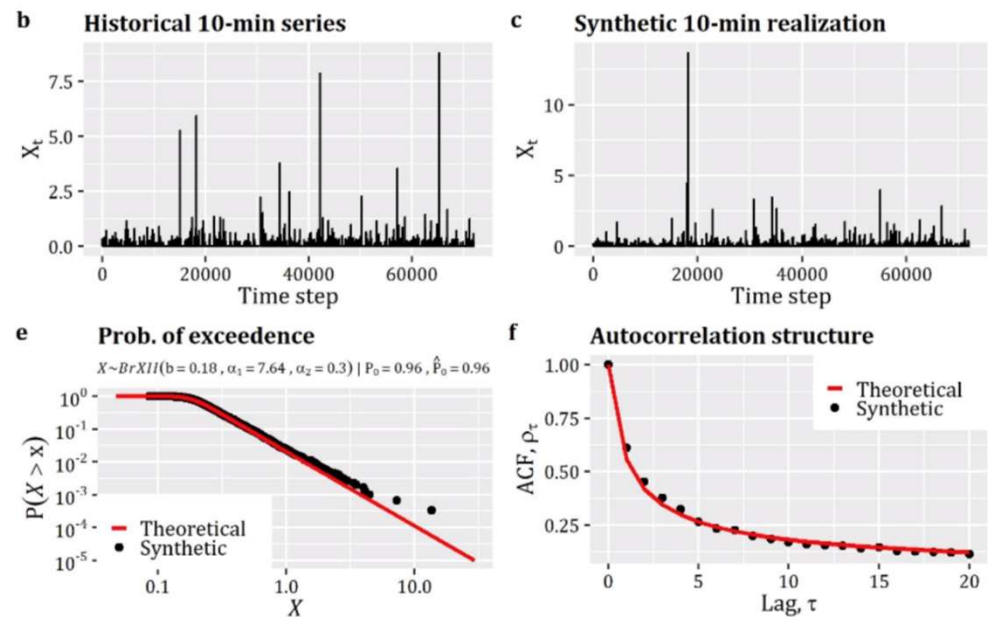
Ανάπτυξη μη-Γκαουσιανών στοχαστικών μοντέλων

Βασική ιδέα: Χρήση της έννοιας των **Copula** (Sklar, 1959; 1973) και συγκεκριμένα του **Gaussian copula**.

Γενικευμένη μεθοδολογία για τη στοχαστική προσομοίωση φυσικών διεργασιών με οποιαδήποτε κατανομή και δομή συσχέτισης.

...καθώς και για προσομοίωση σε **πολλαπλές κλίμακες** (μέσω επιμεριστικής διαδικασίας - **disaggregation**)!

- **Φειδωλή παραμετροποίηση** → Μέσω θεωρητικών μοντέλων (αυτό/έτερο-) συσχέτισης.
- **Περιοδικότητα** → Με τη χρήση κυκλο-στάσιμων μοντέλων (πχ., PAR).
- **Διαλείπουσα συμπεριφορά** → Με τη χρήση κατανομών με άτομο/ασυνέχεια στο μηδέν.



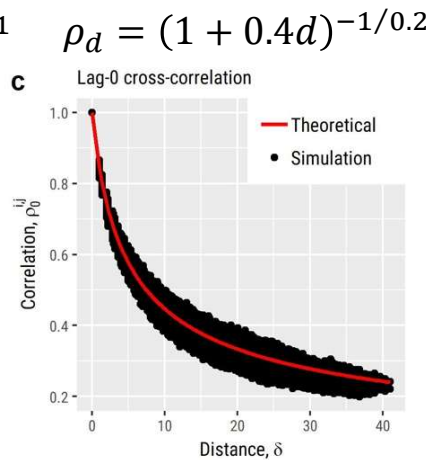
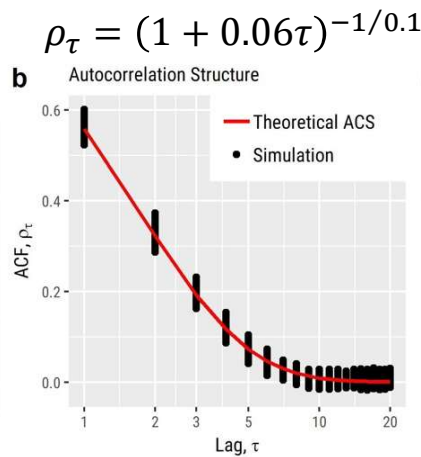
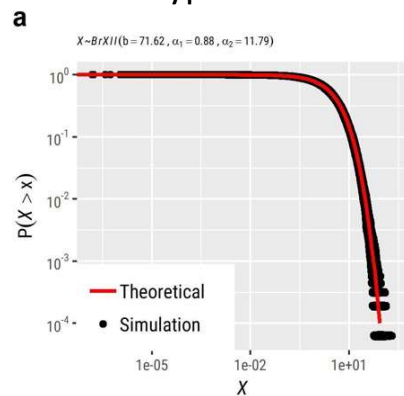
Βλέπε: (Tsoukalas et al., 2020; 2019; 2018a; 2018b; 2018c; 2017; Tsoukalas 2018; Kossieris et al. 2019)

Προσομοίωση μη-Γκαουσιανών τυχαίων πεδίων (non-Gaussian random fields)

Χώρο-χρονική προσομοίωση βροχόπτωσης

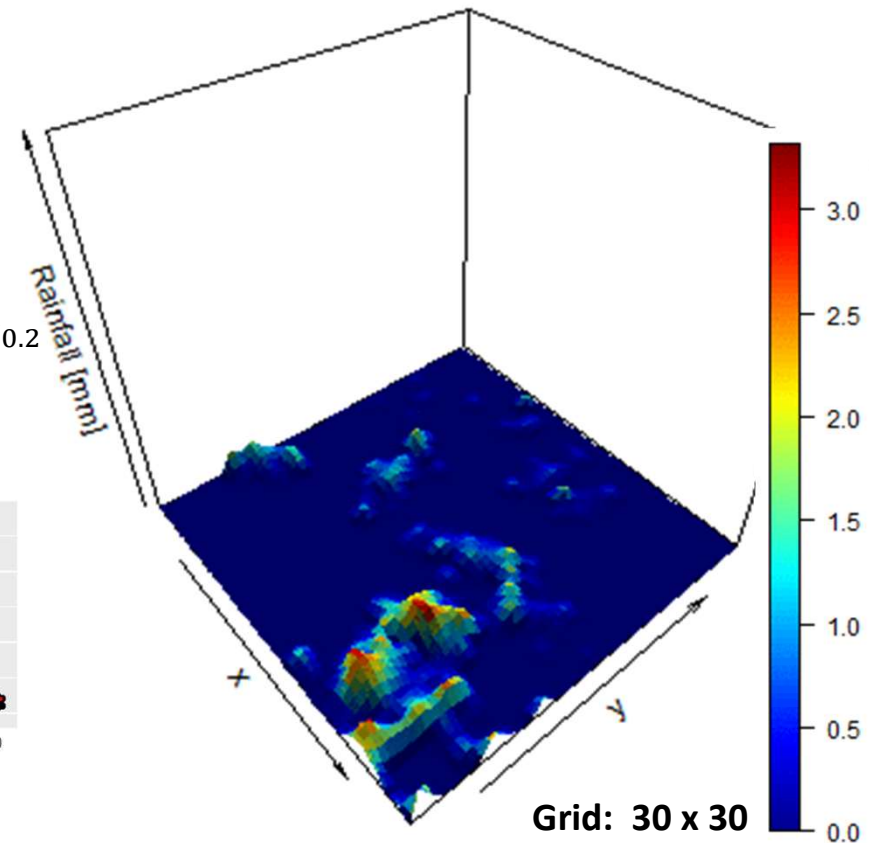
- **Δεδομένα:** Ημερήσια βροχόπτωσης
- **Κατανομή:** Burr type XII με άτομο στο μηδέν.
- **Δομή συσχέτισης στο χώρο (ρ_d) και χρόνο (ρ_τ):**

Burr type XII



Περισσότερες πληροφορίες: Tsoukalas et al. (2020).

Time-step: 1



Προσομοίωση δεσμευμένων μη-Γκαουσιανών τυχαίων πεδίων (conditional non-Gaussian random fields)

Γιατί; Οι περισσότερες υδρομετεωρολογικές μεταβλητές/διεργασίες χαρακτηρίζονται από μη-Γκαουσιανές κατανομές.

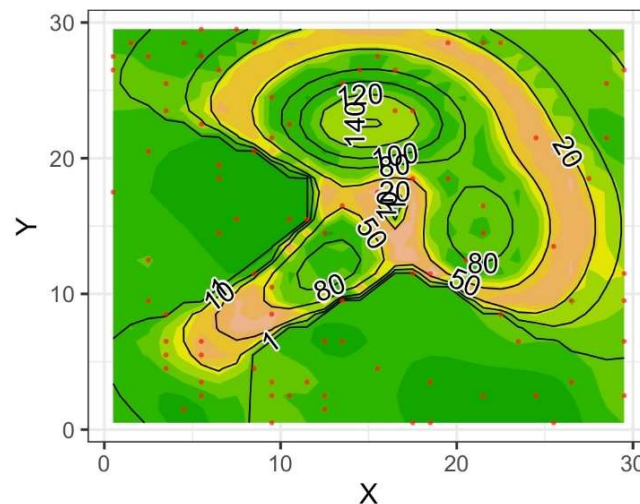
Παράδειγμα: Έστω μετρήσεις εδαφικής υγρασίας (0-100%) σε διάφορες θέσεις/σταθμούς

Πως? Geostatistics **+** Stochastics **+** Copulas

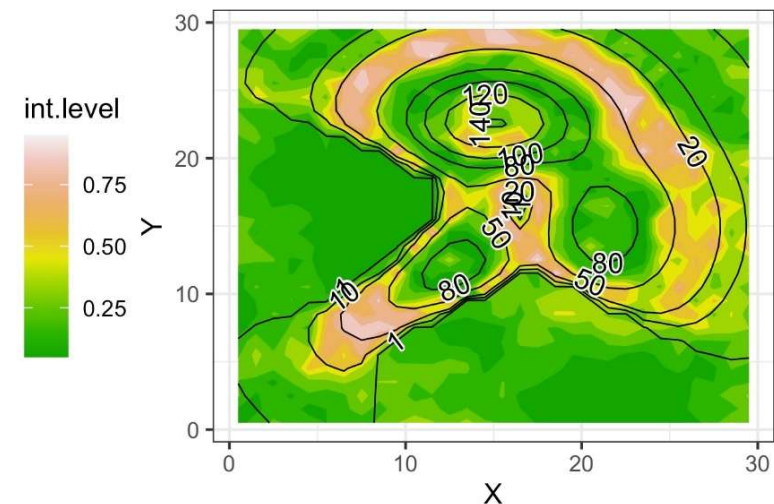
Παράδειγμα: Χωρική παρεμβολή ενός τυχαίου πεδίου με περιθώρια κατανομή Beta (δηλ. περιορισμένη στο (0, 1)).

Διαθέσιμα δεδομένα:
«**ΚΟΚΚΙΝΕΣ** ΚΟΥΚΚΙΔΕΣ».

Τελικό προϊόν: Μη-Γκαουσιανή χωρική παρεμβολή πλέγματος



Προσομοιωμένο πεδίο

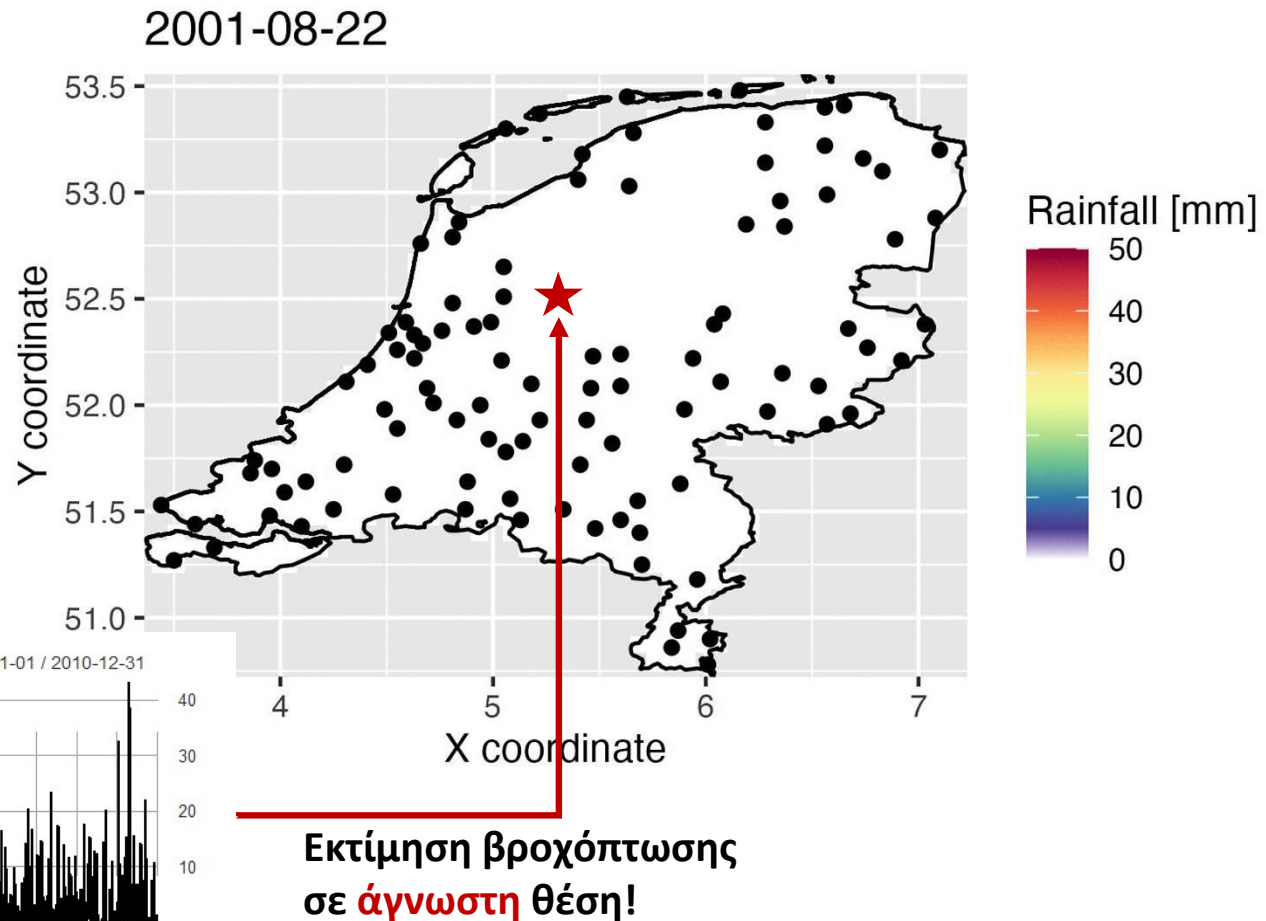
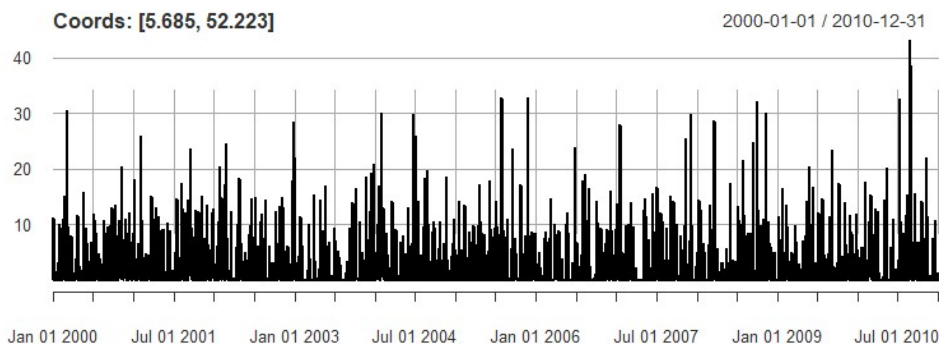


Πραγματικό πεδίο

Προσομοίωση δεσμευμένων μη-Γκαουσιανών τυχαίων πεδίων (conditional non-Gaussian random fields)

Τυπικές εφαρμογές

- Δημιουργία συνθετικών/στοχαστικών πεδίων με οποιαδήποτε κατανομή και δομή εξάρτησης.
- Συμπλήρωση και επέκταση δεδομένων.
- Χωρική παρεμβολή δεδομένων.

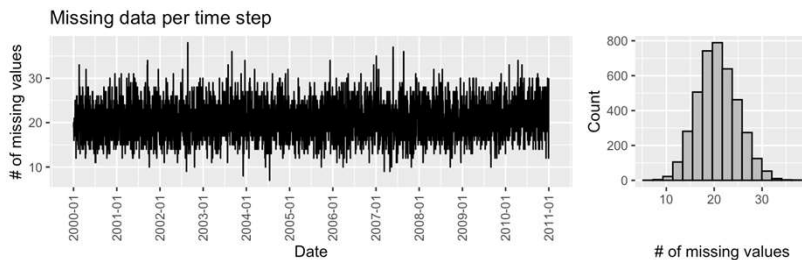
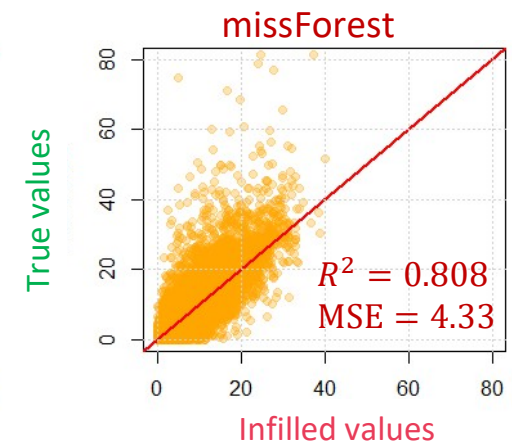
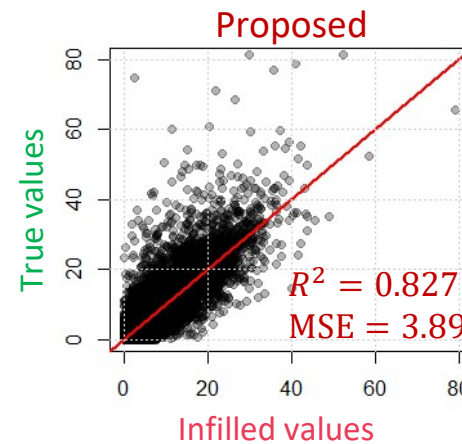


Προσομοίωση δεσμευμένων μη-Γκαουσιανών τυχαίων πεδίων (conditional non-Gaussian random fields)



Παράδειγμα

- Δεδομένα ημερήσιας βροχόπτωσης: 102 σταθμοί (NL)
- Τεχνητή εισαγωγή 20% κενών τιμών
- Εφαρμογή της μεθοδολογίας



Σύγκριση μεταξύ **πραγματικών** and **συμπληρωμένων** τιμών, με βάση (αριστερά) την προτεινόμενη μεθοδολογία και (δεξιά) το λογισμικό *missForest** (R package).

***Σημείωση:** Το λογισμικό *missForest* είναι μια σύγχρονη και ιδιαίτερα δημοφιλής (~6500 αναφορές) προσέγγιση (ML-based) για τη *συμπλήρωση* κενών.

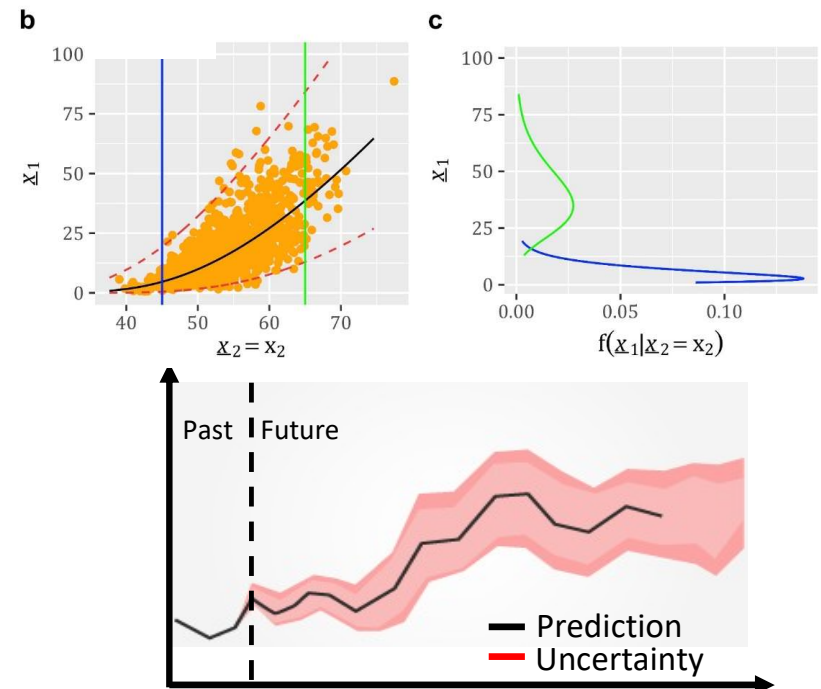
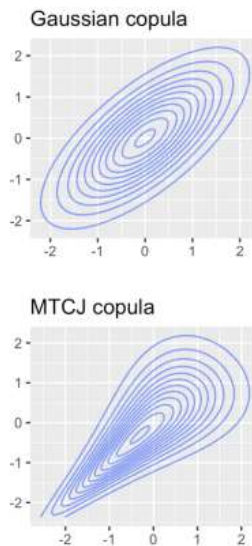
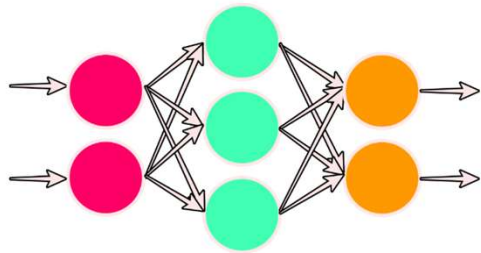
***Stekhoven D., Bühlmann P., (2012). *MissForest—non-parametric missing value imputation for mixed-type data*, Bioinformatics.**

Ποσοτικοποίηση της προγνωστικής αβεβαιότητας μοντέλων

Πως μπορούμε να μεταβούμε από προσδιοριστικές (ντετερμινιστικές) προβλέψεις/μοντέλα σε πιθανοτικά;

Deterministic models (inc. ML) $+$ Copulas $=$ Probabilistic models

- ARMA models
- Artificial Neural networks,
- Deep learning architectures,
- Support vector machines,
- Regression Trees
- Etc.

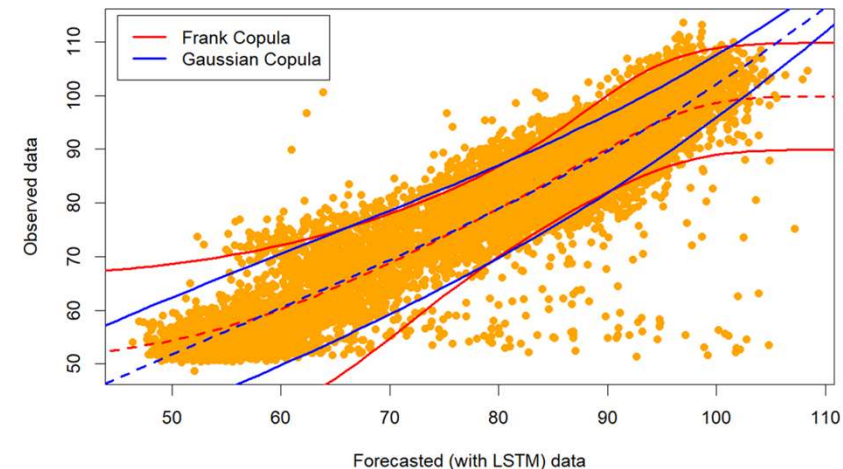
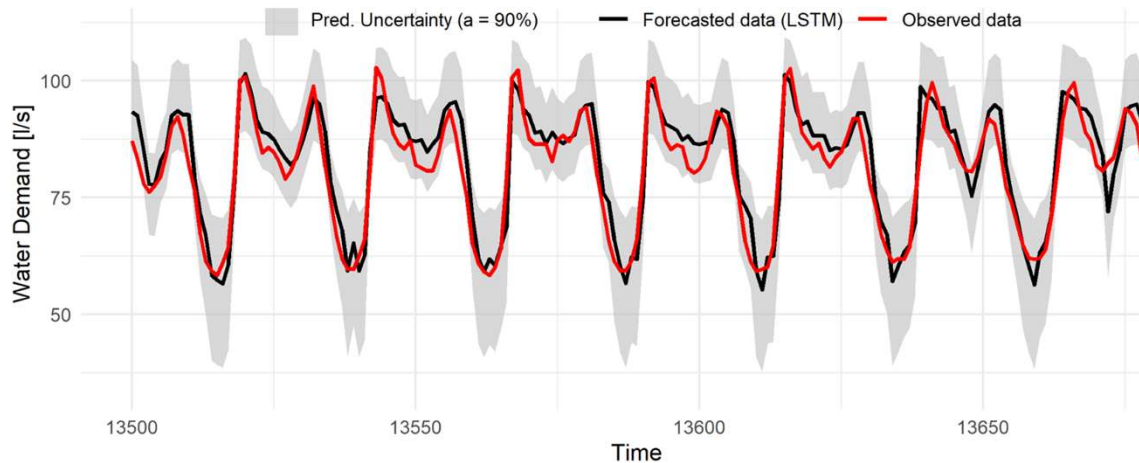


Ποσοτικοποίηση της προγνωστικής αβεβαιότητας μοντέλων

Εφαρμογή της προτεινόμενης μεθοδολογίας

- Ωριαία δεδομένα ζήτησης νερού (Battle of Water Demand Forecasting)
- Δεδομένα αστικής/εμπορικής περιοχής κοντά στο κέντρο της πόλης (anonymized)
- Βαθμονόμηση/εκπαίδευση προγνωστικού μοντέλου μηχανικής μάθησης:
Long Short-Term Memory (LSTM) neural network (1-step ahead forecasting)
Predictors: Past water demand values, meteorological parameters, and temporal features (month, day, and hour).

Μέρος παρουσίασης στο
WDSA/CCWI conference 2024

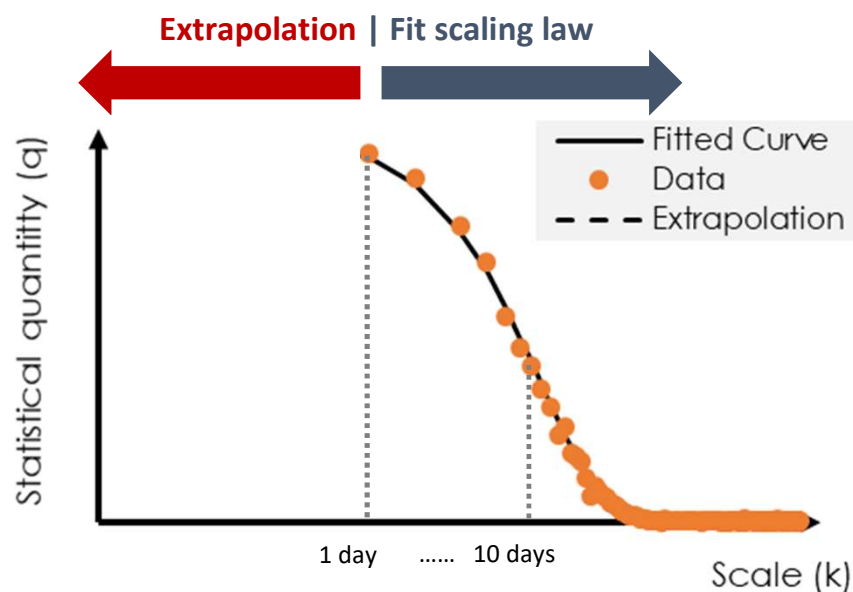


Όμως.. (όπως συνήθως στην Ελλάδα) δεν υπάρχουν δεδομένα...

ή **ενναλακτικά** τα διαθέσιμα δεδομένα είναι σε διαφορετική (χρονική ή χωρική) κλίμακα από την επιθυμητή (για τους σκοπούς της εκάστοτε μελέτης).

Γενικευμένη και φειδωλή προσέγγιση:

1. Στατιστική ανάλυση σε **πολλαπλές κλίμακες** (•) (χρονικές ή/και χωρικές)
2. Αναγνώριση και προσαρμογή **νόμων κλίμακας** (–) στο χρόνο ή/και χώρο.
3. Χρήση των παραπάνω νόμων κλίμακας για την **επέκταση** (– – –) της πληροφορίας σε **χαμηλότερες** κλίμακες.



Χρονικός καταβιβασμός
σε κλίμακα μικρότερη
από αυτή της μέτρησης.

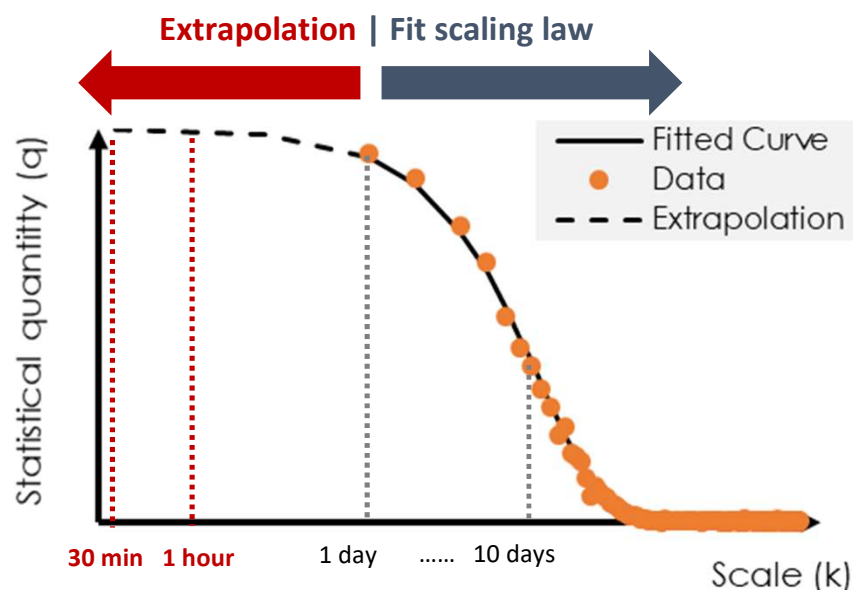
Kossieris P., Tsoukalas I., Efstratiadis A., Makropoulos C., *Downscaling of statistical quantities at fine time-scales, towards cost-effective enrichment of water demand records*. Water.

Όμως.. (όπως συνήθως στην Ελλάδα) δεν υπάρχουν δεδομένα...

ή **ενναλλακτικά** τα διαθέσιμα δεδομένα είναι σε διαφορετική (χρονική ή χωρική) κλίμακα από την επιθυμητή (για τους σκοπούς της εκάστοτε μελέτης).

Γενικευμένη και φειδωλή προσέγγιση:

1. Στατιστική ανάλυση σε **πολλαπλές κλίμακες** (•) (χρονικές ή/και χωρικές)
2. Αναγνώριση και προσαρμογή **νόμων κλίμακας** (–) στο χρόνο ή/και χώρο.
3. Χρήση των παραπάνω νόμων κλίμακας για την **επέκταση** (– – –) της πληροφορίας σε **χαμηλότερες κλίμακες**.



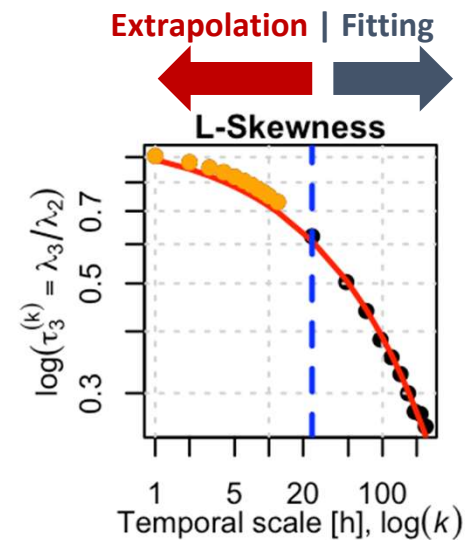
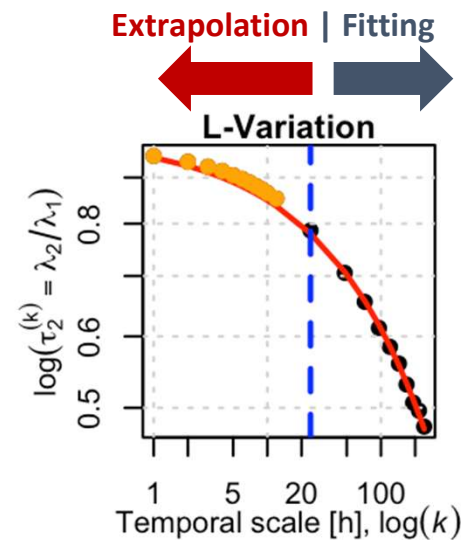
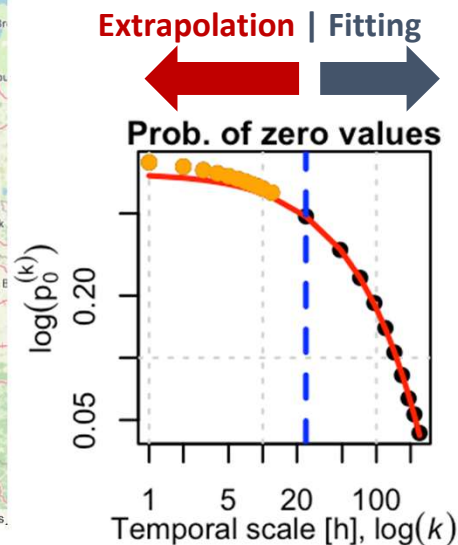
Χρονικός καταβιβασμός
σε κλίμακα μικρότερη
από αυτή της μέτρησης.

Kossieris P., Tsoukalas I., Efstratiadis A., Makropoulos C., *Downscaling of statistical quantities at fine time-scales, towards cost-effective enrichment of water demand records*. Water.

Χρονικός καταβιβασμός (downscaling) στατιστικών χαρακτηριστικών

Παράδειγμα

- Δεδομένα ωριαίας βροχόπτωσης από 33 μετεωρολογικούς σταθμούς στην **Ολλανδία**
- Χρήση δεδομένων **ΜΟΝΟ** ημερήσιας κλίμακας και, στη συνέχεια εφαρμογή της μεθόδου χρονικού καταβιβασμού (downscaling) στατιστικών χαρακτηριστικών
- Επικύρωση της μεθόδου μέσω των **δεδομένων ωριαίας κλίμακας**



Το λογισμικό anySim (R package - <https://github.com/itsoukal/anySim>)

Συνοπτική περιγραφή του θεωρητικού υπόβαθρου, και πολλά παραδείγματα εφαρμογής σε R.

Tsoukalas, I., Kossieris, P., Makropoulos, C. (2020). *Simulation of Non-Gaussian Correlated Random Variables, Stochastic Processes and Random Fields: Introducing the anySim R-Package for Environmental Applications and Beyond*. Water 12, 1645, doi: 10.3390/w12061645.

Water Resources Research

RESEARCH ARTICLE
10.1029/2017WR022462

Simulation of Stochastic Processes Exhibiting Any-Range Dependence and Arbitrary Marginal Distributions

Key Points:
• Simulation of processes with

Ioannis Tsoukalas¹, Christos Makropoulos¹, and Demetris Koutsoyiannis¹

Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University

Water Resources Research

RESEARCH ARTICLE 10.1002/2017WR021394 Stochastic Periodic Autoregressive to Anything (SPARTA): Modeling and Simulation of Cyclostationary Processes With Arbitrary Marginal Distributions

Key Points:
• Simulation of periodic processes with any marginal distribution
• Generation of synthetic time series in univariate or multivariate mode
• Accurate preservation of essential statistics and observed dependencies

Ioannis Tsoukalas¹, Andreas Efstratiadis¹, and Christos Makropoulos¹

¹Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Zografou, Greece



Simulating Marginal and Dependence Behaviour of Water Demand Processes at Any Fine Time Scale

Panagiotis Kossieris^{1,*}, Ioannis Tsoukalas¹, Christos Makropoulos^{1,2} and Dragan Savic^{2,3}

Article A Cautionary Note on the Reproduction of Dependencies through Linear Stochastic Models with Non-Gaussian White Noise

Ioannis Tsoukalas^{1,*}, Simon Michael Papalexiou², Andreas Efstratiadis¹ and Christos Makropoulos¹



Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

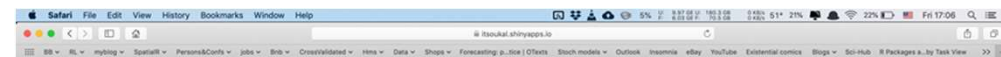
Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure

Ioannis Tsoukalas¹, Andreas Efstratiadis, Christos Makropoulos

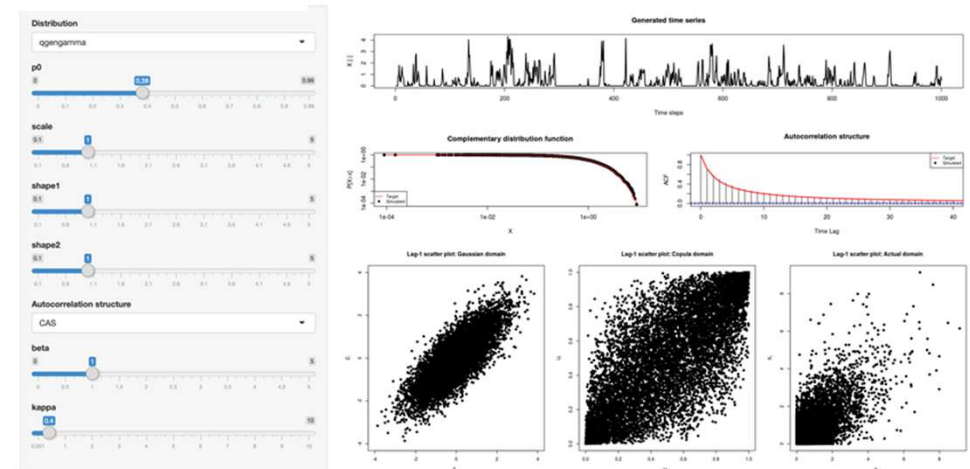
¹Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Heron Polytechnic 5, 15702 Zografou, Greece

Διαδικτυακή εφαρμογή R shiny

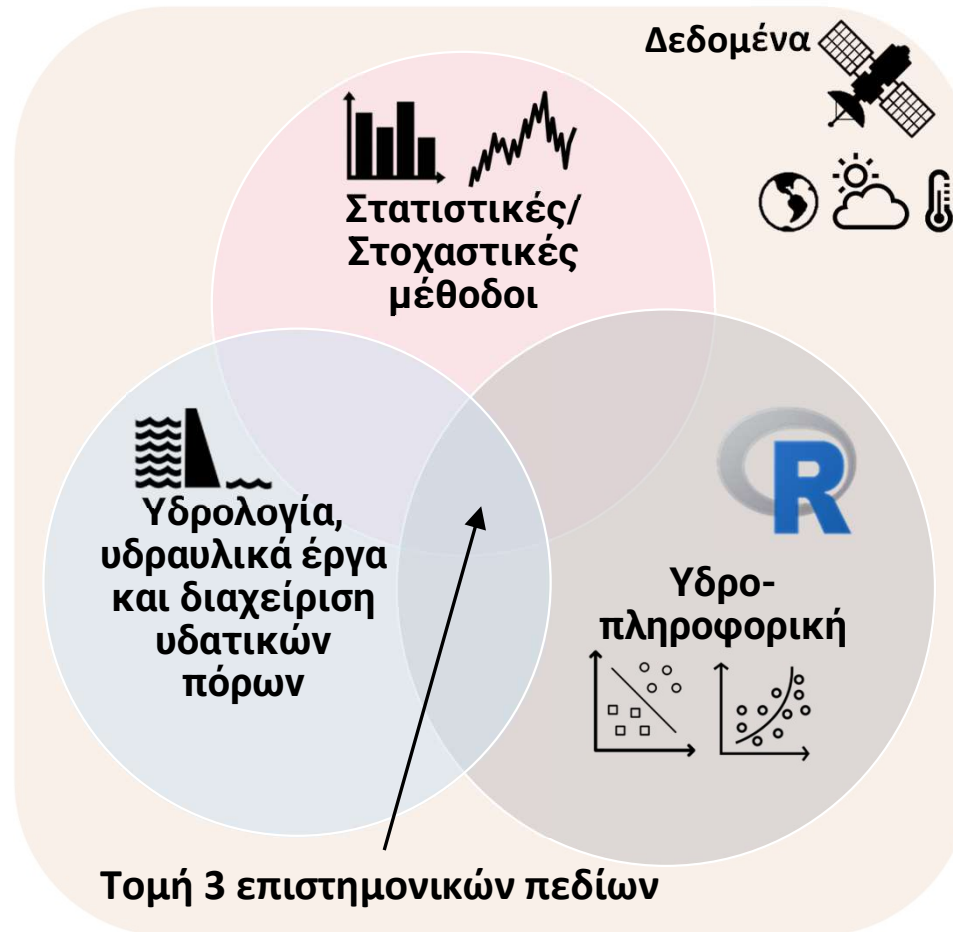
(https://itsoukal.shinyapps.io/anysim_app/)



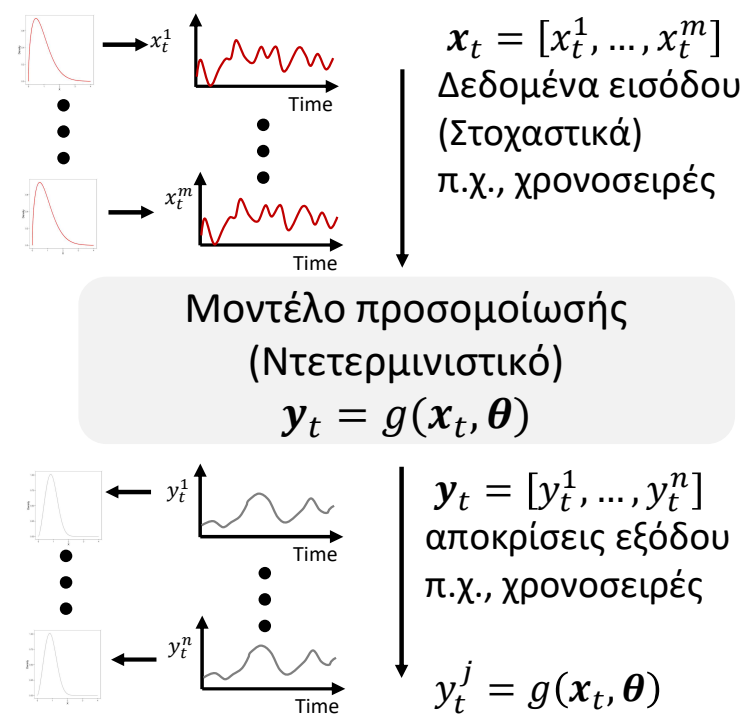
anySim Dashboard



Ανάλυση και διαχείριση έργων υδατικών πόρων υπό αβεβαιότητα...



Υπολογιστικά πειράματα τύπου Monte Carlo (MC)



$\theta = [\theta_1, \dots, \theta_d]$
Παράμετροι του μοντέλου

Βελτιστοποίηση

Εύρεση των βέλτιστων παραμέτρων θ που μεγιστοποιούν ή ελαχιστοποιούν κάποια κριτήρια (αντικειμενική συνάρτηση)

Εμπειρικά δεδομένα εισόδου

- Περιορισμένου μήκους.
- Δεν επιτρέπουν την εξαγωγή ασφαλών πιθανοτικών συμπερασμάτων.

Στοχαστικά δεδομένα εισόδου

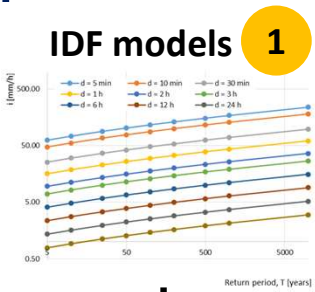
- Διάδοση της αβεβαιότητας στις χρονοσειρές εξόδου του μοντέλου (model outputs).
- Εξαγωγή ασφαλών πιθανοτικών συμπερασμάτων.
- Ανάπτυξη πλαισίων προσομοίωσης -βελτιστοποίησης (simulation-optimization) αλλά και λήψης αποφάσεων που ενσωματώνουν την αβεβαιότητα.

Σημείο κλειδί:

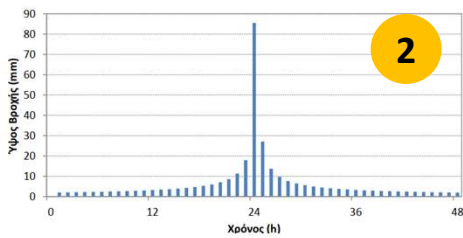
Ρεαλιστική (στοχαστική) μοντελοποίηση και προσομοίωση των διεργασιών εισόδου.

1 .Προσομοίωση πλημμυρικών ροών υπό αβεβαιότητα

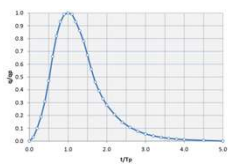
Κλασσική (ντετερμινιστική) προσέγγιση



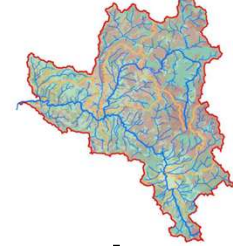
Alternate blocks



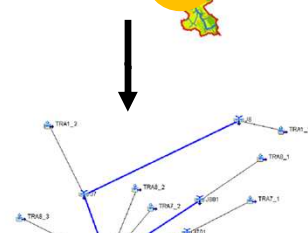
UH models



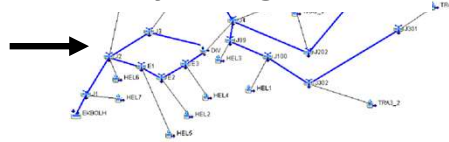
Corine



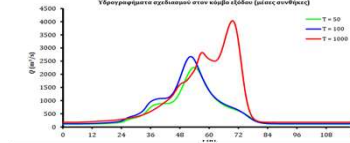
Soil type

SCS: CN 

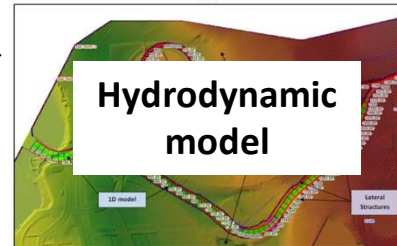
Hydrologic model



Hydrographs @ nodes



Hydrodynamic model



Depth maps



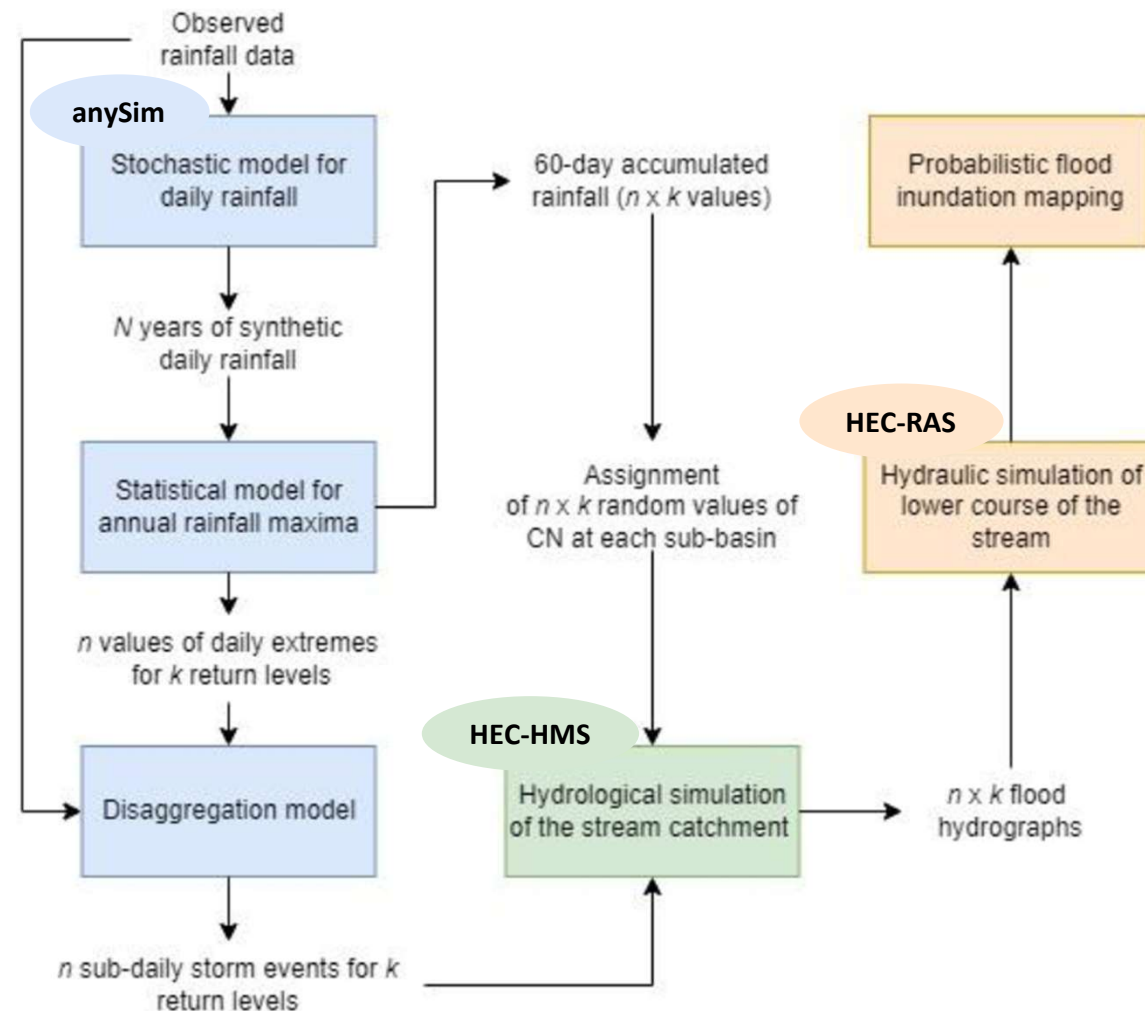
Ζητήματα & προβληματισμοί

- 1) **Περιορισμένη χρήση πιθανοτικών εννοιών**
(εκτός της έννοιας της περιόδου επαναφοράς, T)
- 2) **Μη ενσωμάτωση της στοχαστικότητας/ μεταβλητότητας που χαρακτηρίζει τα δεδομένα εισόδου** (πχ., βροχόπτωση προηγ. ημερών, και άρα CN), υετογράφημα εισόδου).
- 3) **Μη εκτίμηση και ποσοτικοποίηση της αβεβαιότητας των αποτελεσμάτων.**

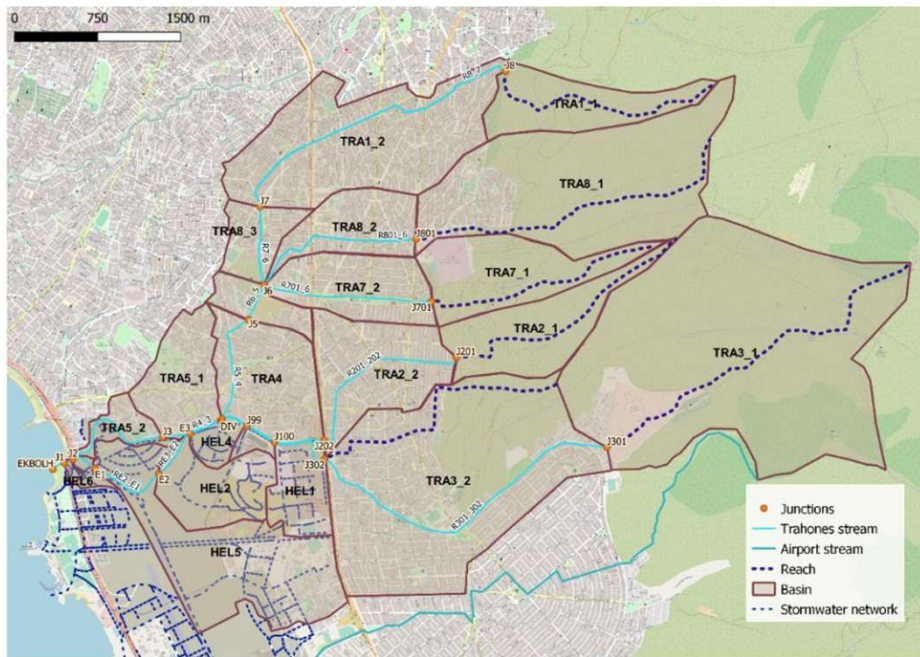
Προς ένα στοχαστικό πλαίσιο εκτίμησης πλημμυρικού κινδύνου/διακινδύνευσης

- Ανάλυση σε κλίμακα επεισοδίου
- Συνδυάζει ντετερμινιστικά μοντέλα (υδρολογικά και υδροδυναμικά) με τεχνικές στοχαστικής προσομοίωσης
- Οδηγεί στην παραγωγή πιθανοτικών μεγεθών/αποτελεσμάτων (π.χ., χάρτες με βάθη πλημμύρας, ταχύτητες ροής)
- Παρέχει τη δυνατότητα ενσωμάτωσης (και ποσοτικοποίησης) της αβεβαιότητας (υδρολογικής και υδραυλικής)
- Υπολογιστικά εφικτή διαδικασία

Efstratiadis A., Dimas P., Pouliasis G., Tsoukalas I., Kossieris P., Bellos V., Sakki G., Makropoulos C., Michas S., (2022). *Revisiting flood engineering practices and the concept of risk under a hybrid stochastic simulation framework*. Water.



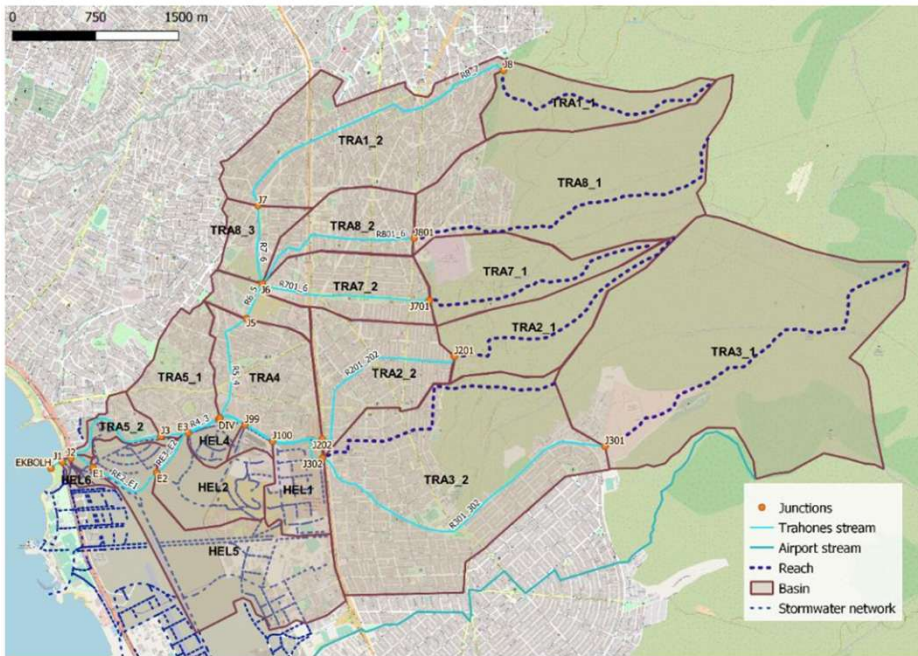
Αξιολόγηση ρίσκου και επιπτώσεων πλημμυρικών γεγονότων στην περιοχή του Ελληνικού



Πηγές αβεβαιότητας που λήφθηκαν υπόψιν

- I. Στατιστική αβεβαιότητα μέγιστης ετήσιας βροχόπτωσης
- II. Χρονική κατανομή των επεισοδίων βροχής
- III. Συνθήκες υγρασίας (βλ. Curve Number – CN) → εκτίμηση μέσω συνεχούς στοχαστικής προσομοίωσης της διεργασίας της ημερήσιας βροχόπτωσης.

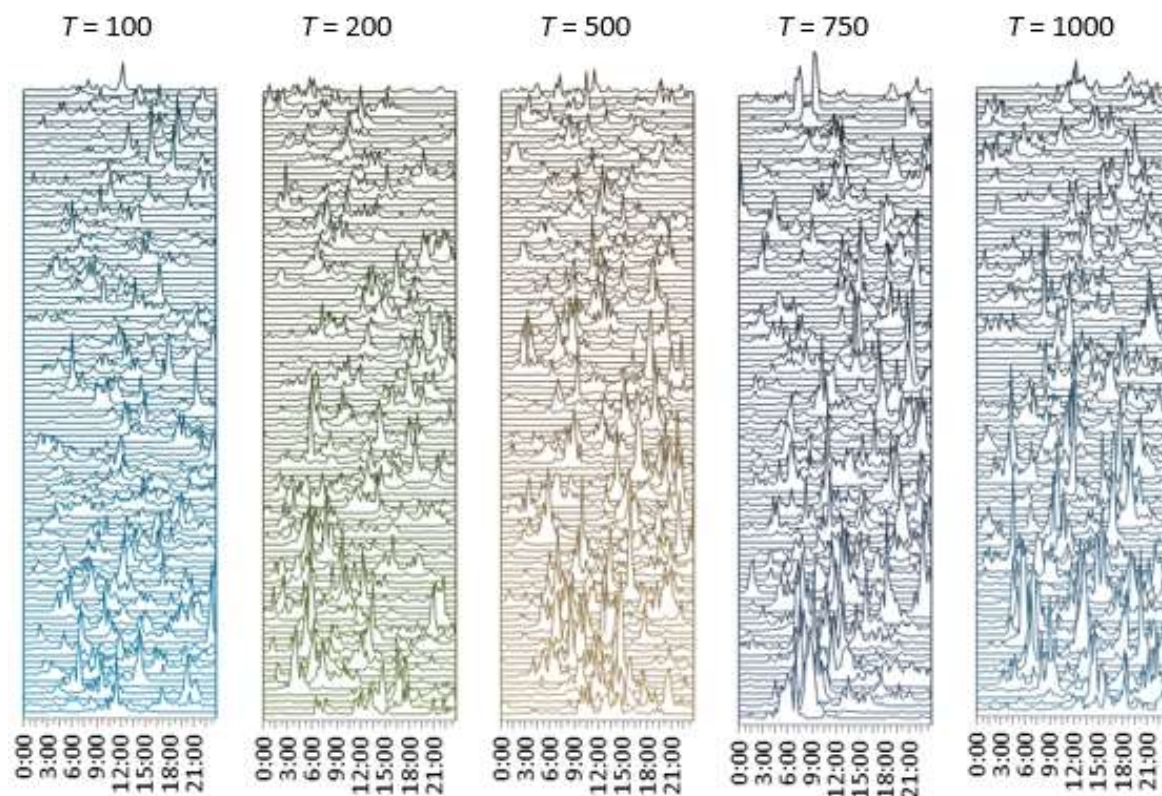
Αξιολόγηση ρίσκου και επιπτώσεων πλημμυρικών γεγονότων στην περιοχή του Ελληνικού



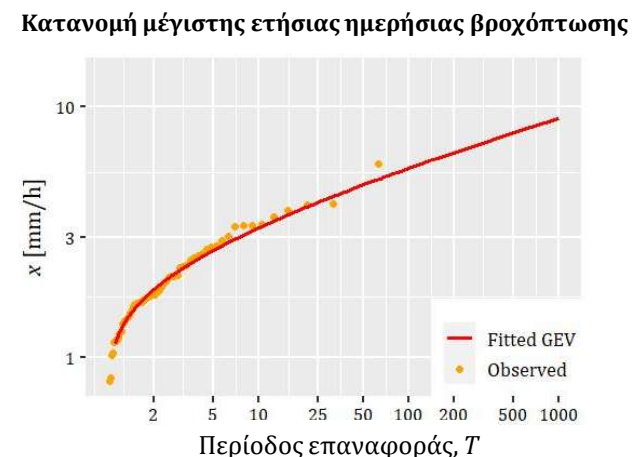
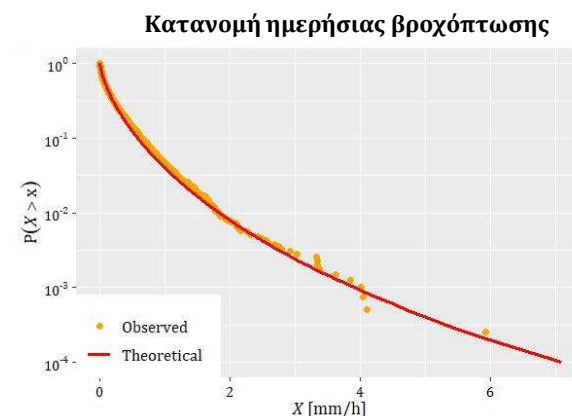
Πηγές αβεβαιότητας που λήφθηκαν υπόψιν

- I. Στατιστική αβεβαιότητα μέγιστης ετήσιας βροχόπτωσης
- II. Χρονική κατανομή των επεισοδίων βροχής**
- III. Συνθήκες υγρασίας (βλ. Curve Number – CN) → εκτίμηση μέσω συνεχούς στοχαστικής προσομοίωσης της διεργασίας της ημερήσιας βροχόπτωσης.

Σχεδιασμός έναντι πλημμυρικών φαινομένων υπο αβεβαιότητα



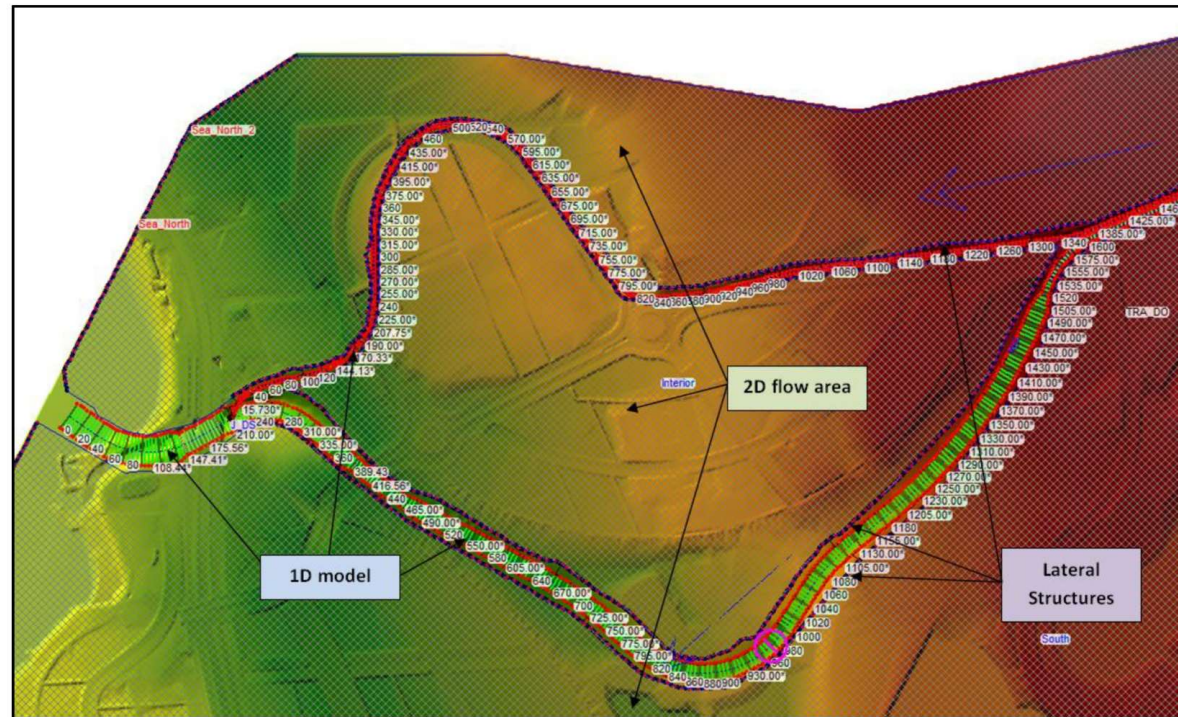
Παραγωγή, μέσω επιμερισμού, στοχαστικών υετογραφημάτων (με βήμα 15 min) για κάθε T , και συνεπών με τις πιθανοτικές κατανομές της 24h βροχόπτωσης.



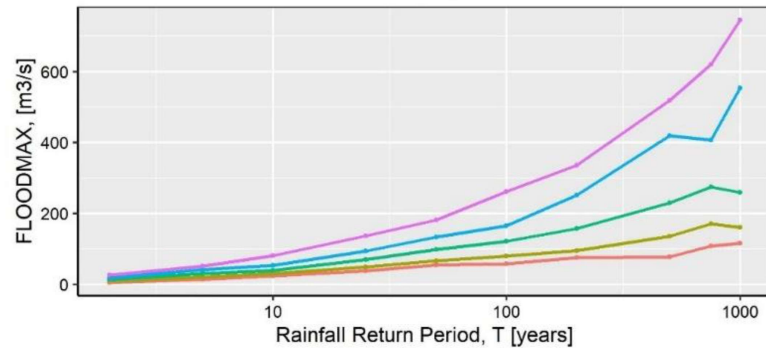
Συζευγμένη προσέγγιση 1D και 2D υδροδυναμικών μοντέλων

Συνδυασμός πλεονεκτημάτων των 1D και 2D μοντέλων

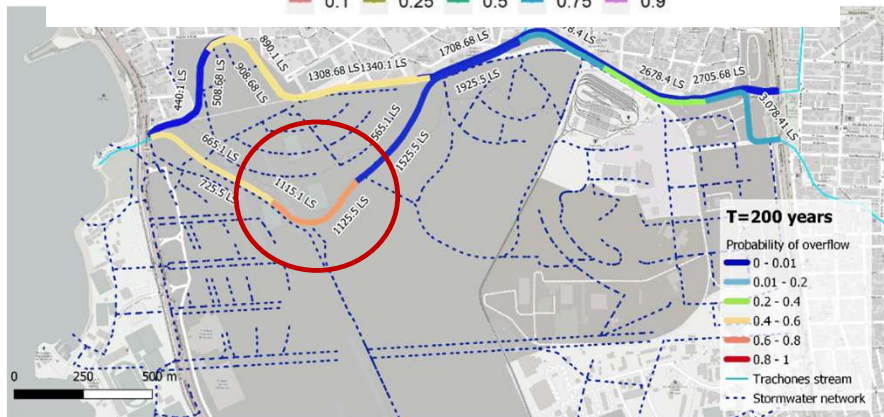
- Γρήγορη υδροδυναμική αναπαράσταση εκτεταμένων διασυνδεδεμένων ποτάμιων συστημάτων.
- Επίλυση μέσω ενός συζευγμένου αλγόριθμου 1D-2D: θεωρώντας ότι ένα ποτάμι/κανάλι έχει αναπαρασταθεί 1D με την περιοχή πίσω από ένα ανάχωμα να έχει αναπαρασταθεί σε 2D (συνδεδεμένα με ένα Lateral Structure).
- Υπερχειλίζει από το κυρίως ποτάμι προς τις εκτάσεις εκτεταμένης κατάκλυσης: Δυνατότητα στον μηχανικό να κάνει ένα ranking των σεναρίων και προκαταρκτικές αξιολογήσεις των επικίνδυνων περιοχών πριν καν προχωρήσει στην αναλυτική 2D προσομοίωση.
- **Bonus #1:** Μειώνουν δραματικά τους χρόνους εκτέλεσης των υπολογισμών σε σχέση με ένα αμιγώς 2D μοντέλο. Σημαντικό για πολλαπλά σενάρια!
- **Bonus #2:** Σταδιακή ανάπτυξη των επιμέρους μοντέλων.



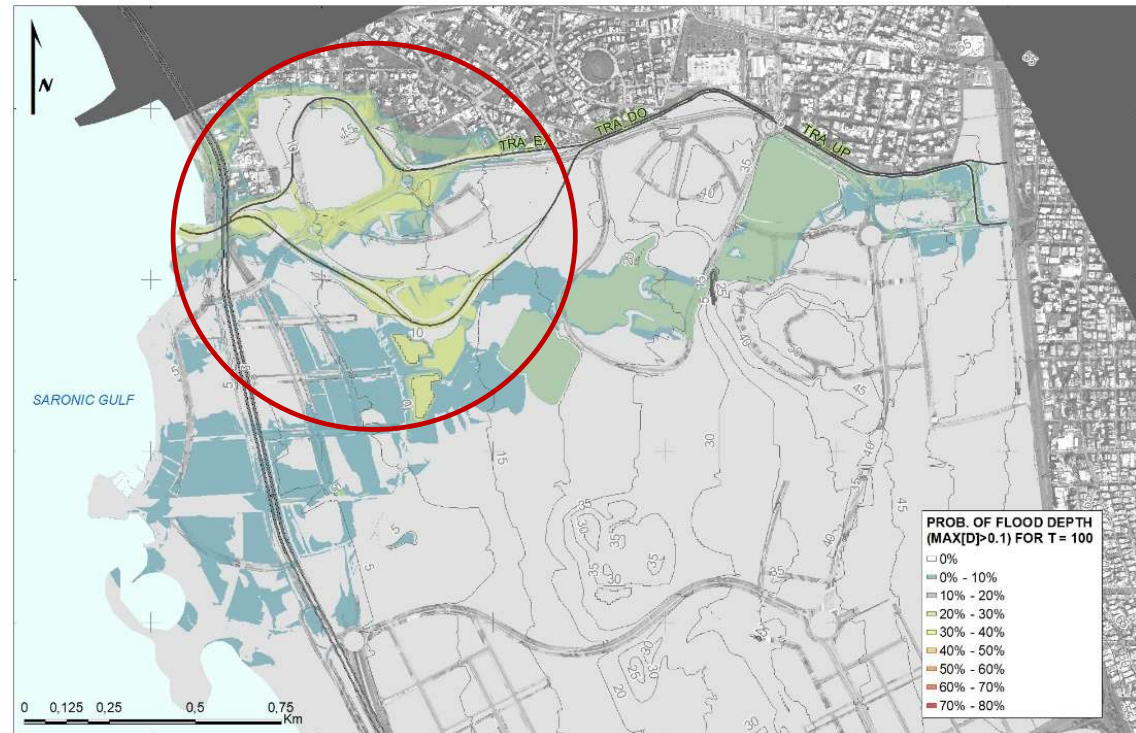
Στοχαστική προσομοίωση πλημμυρικών φαινομένων



— 0.1 — 0.25 — 0.5 — 0.75 — 0.9



Πιθανότητα υπερχείλισης των 24 πλευρικών κατασκευών (lateral structures), για επεισόδια βροχόπτωσης $T = 200$ έτη.



Χάρτης πιθανότητας υπέρβασης του ύψους του κρασπέδου (10 cm), για επεισόδια βροχόπτωσης με $T=100$ έτη.

Προσομοίωση πλημμυρικών ροών υπό αβεβαιότητα

Βασικά χαρακτηριστικά

- «Γεφυρώνει» ντετερμινιστικές και στοχαστικές προσεγγίσεις
 - Επεκτάσιμο σε διάφορες χωρικές κλίμακες
 - Επεκτάσιμο για την αντιμετώπιση διαφορετικών πηγών αβεβαιότητας
 - Εντοπισμός (πιθανοτικά) περιοχών υψηλού ρίσκου
 - Πιθανοτική εκτίμηση και απεικόνιση των πλημμυρικών μεγεθών (π.χ., βάθη ή όρια πλημμύρας για διάφορα T)
 - Υπολογιστικά εφικτή υπολογιστική διαδικασία
-

2. Βελτιστοποίηση συστημάτων υδατικών πόρων υπό αβεβαιότητα

Βελτιστοποίηση συστημάτων υδατικών πόρων υπό αβεβαιότητα

Βελτιστοποίηση! Υπολογιστικός φόρτος ☹

Συνήθη ζητήματα σε «πραγματικά» προβλήματα βελτιστοποίησης μηχανικού:

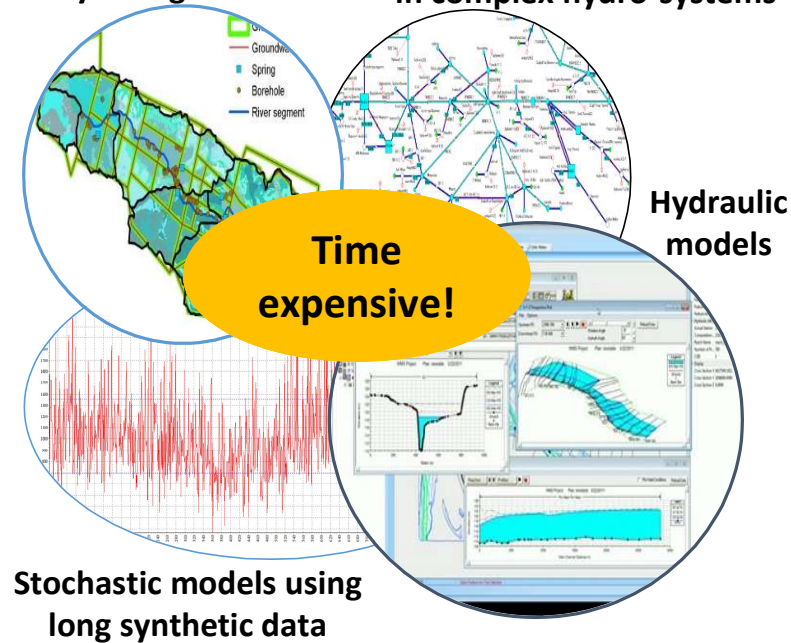
- ❑ **Ανάγκη για βελτιστοποίηση** ή ανάλυση ευαισθησίας του συστήματος
 - Χρονοβόρα (δαπανηρή) αντικειμενική συνάρτηση (π.χ., αποτίμιση μέσω μοντέλου προσομοίωσης)
 - Ασυνεχής και μη διαφορίσιμη αντικειμενική συνάρτηση
- ❑ Πολύπλοκα μοντέλα προσομοίωσης απαιτούν **μεγάλο υπολογιστικό φόρτο** (άρα και **χρόνο!**)

- ❑ Η **χρήση εξελικτικών αλγορίθμων** (evolutionary algorithms) βελτιστοποίησης καθίσταται **απαγορευτική** λόγω του **μεγάλου αριθμού** απαιτούμενων **επαναλήψεων**.

π.χ., **10 000 επαναλήψεις** ενός (γρήγορου!) **μοντέλου** που απαιτεί **5 λεπτά** για τον υπολογισμό της αντικειμενικής συνάρτησης → $10\,000 \times 300/60 \times 60 \times 24 = \mathbf{34.7 \text{ μέρες!!}}$

Calibration of physically-based hydrological models

Decision-making problems in complex hydro-systems



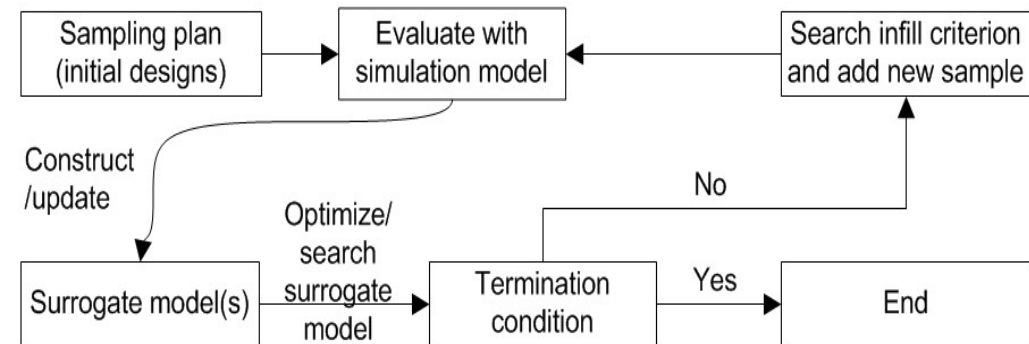
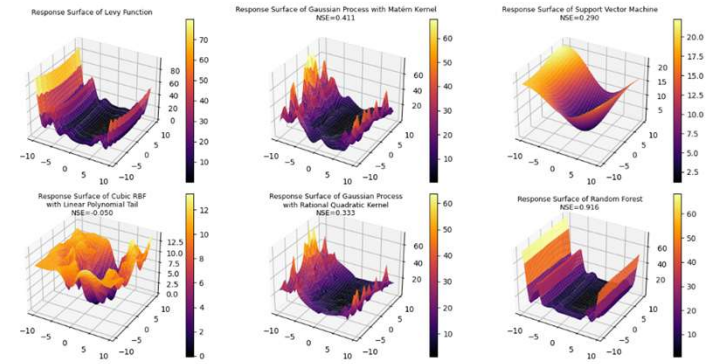
Βελτιστοποίηση συστημάτων υδατικών πόρων υπό αβεβαιότητα

Βελτιστοποίηση! Υπολογιστικός φόρτος ☹

Βασική ιδέα : Βελτιστοποίηση με την χρήση υποκατάστατων μοντέλων μηχανικής μάθησης (βλ. Surrogate-based Optimization)

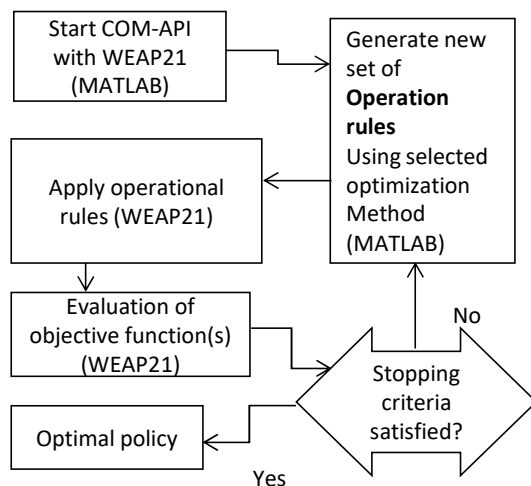
Χωρίς απαιτήσεις σε hardware! Μόνο έξυπνο software.

- **Κατασκευή** ενός μοντέλου μηχανικής μάθησης/ML (GPs, ANNs, SVMs) με το οποίο αντικαθιστάμε τις περισσότερες ακριβές προσομοιώσεις – βλ. υποκατάστατα μοντέλα.
- Οι περισσότερες **αξιολογήσεις** της **αντικειμενικής συνάρτησης** γίνονται με τη βοήθεια του υποκατάστατου μοντέλου (**ML**), ενώ το **ακριβό μοντέλο** χρησιμοποιείται σποραδικά για τη βελτίωση της ακρίβειας του υποκατάστατου μοντέλου (Razavi *et al.*, 2012).
- Τεχνική **κατάλληλη** και για **πολυκριτηριακά προβλήματα** (Tsoukalas and Makropoulos 2015a; 2015b)



Flowchart of a typical surrogate-based optimization procedure.

Παράδειγμα ΔΥΠ πολυκριτηριακής βελτιστοποίησης 4D: Η περίπτωση του Νέστου [1]



Διάγραμμα ροής που περιγράφει τη σύνδεση WEAP21-MATLAB

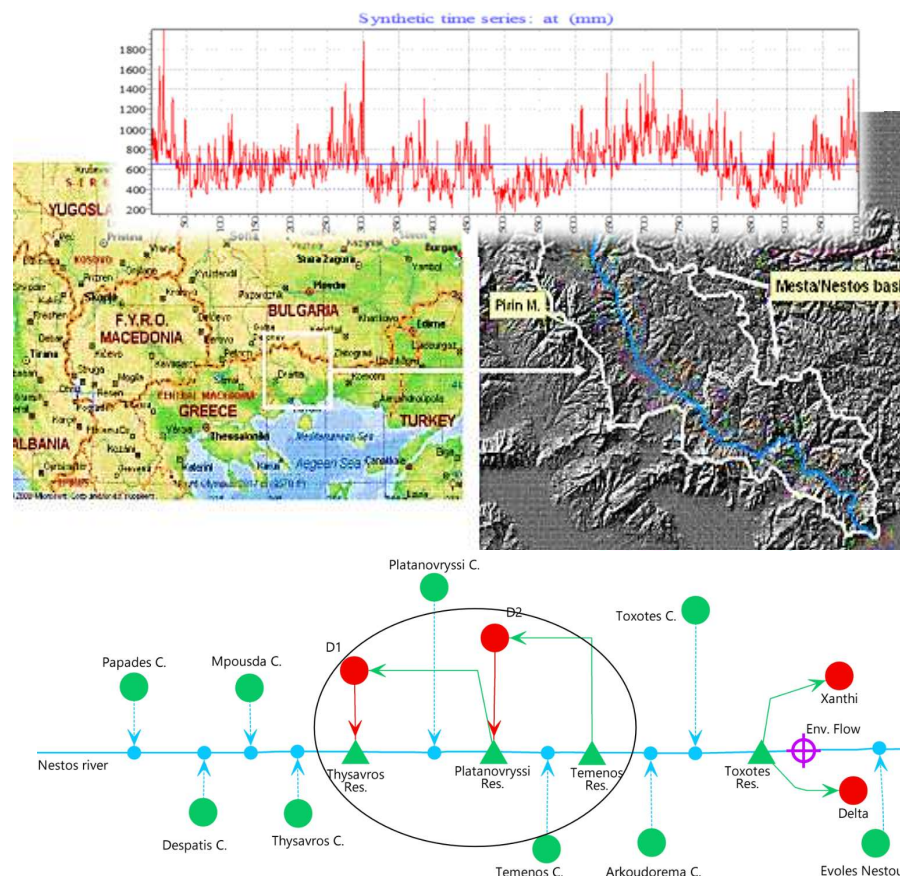
Στόχος:

- 1) Μεγιστοποίηση πρωτεύουσας ενέργειας συστήματος για αξιοπιστία 99%
- 2) Μεγιστοποίηση πρωτεύουσας ενέργειας συστήματος για αξιοπιστία 99% μόνο για μια συγκεκριμένη περίοδο (Μάιο-Σεπτέμβριο).

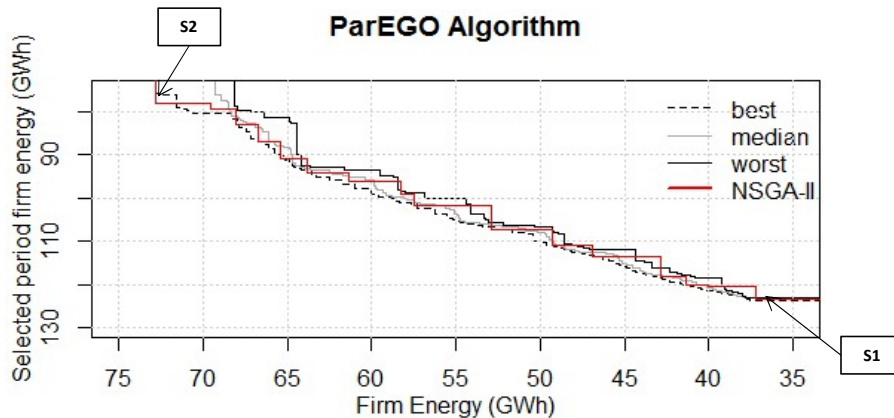
Μεταβλητές ελέγχου:

Ένας στόχος θ_i ανά ΥΗΕ, όπου αναφέρεται στη μηνιαία παραγωγή ενέργειας, επιπλέον στόχος για άντηληση.

Χρήση στοχαστικών χρονοσειρών 1000 ετών



Παράδειγμα ΔΥΠ πολυκριτηριακής βελτιστοποίησης 4D: Η περίπτωση του Νέστου [2]



Σύγκριση **ParEGO** best, median και worst EAF με **NSGA-II** για **400** και **10 000** υπολογισμούς του μοντέλου αντίστοιχα.

S1:

$$f_1 = 37 \text{ GWh}$$

$$f_2 = 123 \text{ GWh}$$

S2:

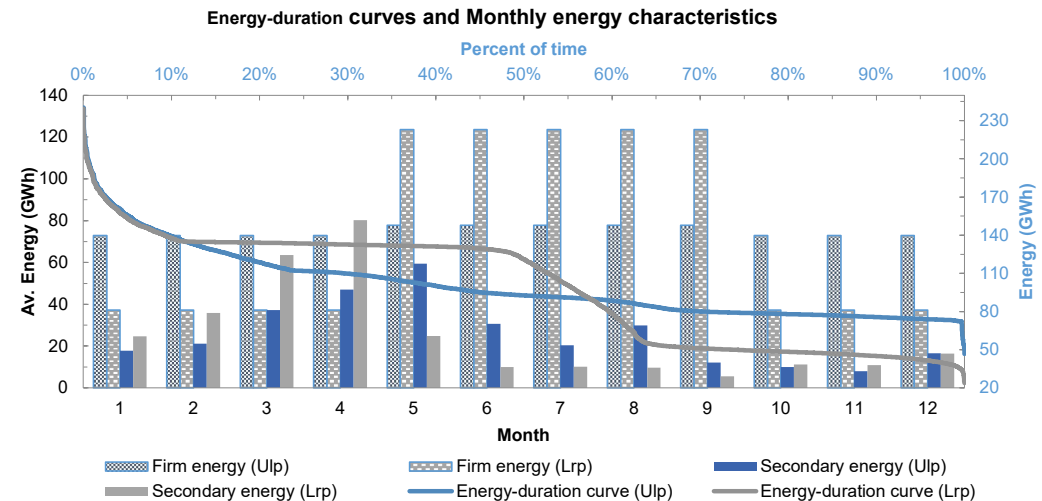
$$f_1 = 72 \text{ GWh}$$

$$f_2 = 77 \text{ GWh}$$

Single simulation run ~ 2 min →
10 000 × 2 min ~ 333 h ~ **14 days**

Setup:

- Ordinary Kriging model
- Gauss correlation function
- ParEGO algorithm (Knowles, 2005)



Αντί επιλόγου, μια λίστα ευχών

Λίστα ευχών

1. Αποδοχή πως όλα τα **μοντέλα είναι λάθος αλλά κάποια είναι χρήσιμα**. ~ George Box (1979 p. 2).
2. Αποδοχή πως η **αβεβαιότητα** είναι πανταχού παρούσα – ανεξάρτητα του προβλήματος, του μοντέλου, και του χρονικού ορίζοντα
3. Αποδοχή της ανάγκης για **συγχώνευση πληροφοριών από πολλαπλές πηγές** (αλλά και μοντέλα).
4. Αποδοχή ότι τα μοντέλα πρέπει να **επικαιροποιούνται συχνά**.
5. Αποδοχή πως η **αβεβαιότητα έχει δύο όψεις**.
 - **Την όμορφη** (μπορεί να μας προετοιμάσει καλύτερα για το απρόβλεπτο) και
 - **Την άσχημη** (δεν είναι απλά ένας αριθμός* – περιπλέκει αποφάσεις και πολιτική).

Δείτε επίσης, το βιβλίο: Leach P., (2006), *Why Can't You Just Give Me The Number? An Executive's Guide to Using Probabilistic Thinking to Manage Risk and to Make Better Decisions*



*Τέσσερις πραγματοποιήσεις 2Δ τυχαίου περίπατου (random walk)
(Ξεκινώντας πάντα από την κόκκινη τελεία.)*

Περισσότερες πληροφορίες – Ενδεικτική βιβλιογραφία

- Efstratiadis A., Dimas P., Pouliasis G., Tsoukalas I., Kossieris P., Bellos V., Sakki G., Makropoulos C., Michas S., (2022). Revisiting flood engineering practices and the concept of risk under a hybrid stochastic simulation framework. *Water*.
- Klemeš, V., and L. Borůvka (1974), Simulation of Gamma-Distributed First-Order Markov Chain, *Water Resour. Res.*, 10(1), 87–91, doi:10.1029/WR010i001p00087.
- Koutsoyiannis, D. (2000), A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resour. Res.*, 36(6), 1519–1533, doi:10.1029/2000WR900044.
- Kossieris P., Tsoukalas I., Efstratiadis A., Makropoulos C. (2019) Downscaling of statistical quantities at fine time-scales, towards cost-effective enrichment of water demand records. *Water*.
- Knowles, J., 2005. ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multi-objective optimization problems. *IEEE Trans. Evol. Comput.* 10 (1), 50e66.
- Koutsoyiannis, D. (2001), Coupling stochastic models of different timescales, *Water Resour. Res.*, 37(2), 379–391, doi:10.1029/2000WR900200.
- Matalas, N. C. (1967), Mathematical assessment of synthetic hydrology, *Water Resour. Res.*, 3(4), 937–945, doi:10.1029/WR003i004p00937.
- Nataf, A. (1962), Statistique mathématique-determination des distributions de probabilités dont les marges sont données, *C. R. Acad. Sci. Paris*, 255(1), 42–43.
- Tsoukalas, I., S. Papalexiou, A. Efstratiadis, and C. Makropoulos (2018a), A Cautionary Note on the Reproduction of Dependencies through Linear Stochastic Models with Non-Gaussian White Noise, *Water*, 10(6), 771, doi:10.3390/w10060771.
- Tsoukalas, I., A. Efstratiadis, and C. Makropoulos (2018b), Stochastic Periodic Autoregressive to Anything (SPARTA): Modeling and simulation of cyclostationary processes with arbitrary marginal distributions, *Water Resour. Res.*, 54(1), 161–185, doi:10.1002/2017WR021394.
- Tsoukalas, I., C. Makropoulos, and D. Koutsoyiannis (2018c), Simulation of stochastic processes exhibiting any-range dependence and arbitrary marginal distributions, *Water Resour. Res.*, doi:10.1029/2017WR022462.
- Tsoukalas, I., Efstratiadis, A., and Makropoulos, C. (2019). Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure, *Journal of Hydrology*, 575, 354–380, doi: 10.1016/j.jhydrol.2019.05.017.
- Tsoukalas, I., Kossieris, P., Makropoulos, C. (2020). Simulation of Non-Gaussian Correlated Random Variables, Stochastic Processes and Random Fields: Introducing the anySim R-Package for Environmental Applications and Beyond. *Water* 12, 1645, doi: 10.3390/w12061645.
- Tsoukalas, I., (2022). The tales that the distribution tails of non-Gaussian autocorrelated processes tell: Efficient methods for the estimation of the k-length block-maxima distribution. *Hydrological Sciences Journal*, doi:10.1080/02626667.2021.2014056.
-