

Interreg



ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ

Ελλάδα-Κύπρος

Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης



SmartWater2020



ΔΕΣΜΟΙ
ΑΝΑΠΤΥΞΗΣ

ΕΚΠΑΙΔΕΥΤΙΚΟ ΣΕΜΙΝΑΡΙΟ

**Τεχνολογίες Επικοινωνιών και Ανάλυσης Δεδομένων για
Ευφυή Δίκτυα Ύδρευσης**



Ηράκλειο, 30/11/2020



Θέματα Υψηλού Επιπέδου Ανάλυσης Δεδομένων

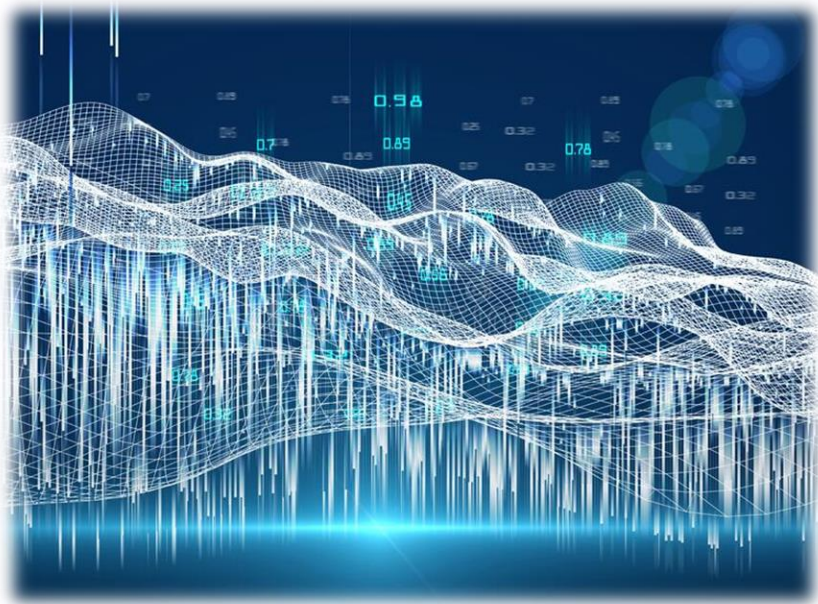


Γιώργος Τζαγκαράκης, Ερευνητής Β', ΙΤΕ-ΙΠ

Υψηλού Επιπέδου Ανάλυση Δεδομένων (High-Level Data Analysis)

ΥΕΑΔ κ' Λήψη Αποφάσεων

Ροές Δεδομένων
(Data Streams)



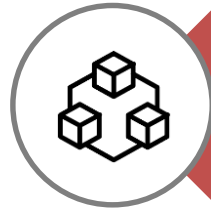
Υψηλού
Επιπέδου
Ανάλυση
Δεδομένων



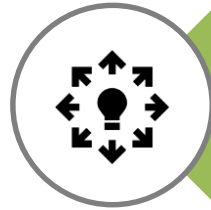
Λήψη Αποφάσεων



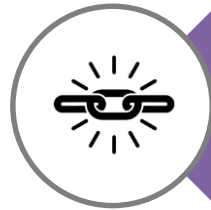
Χαρακτηριστικά Συστήματος



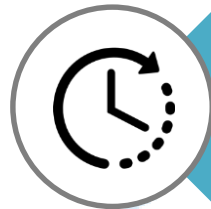
Αρθρωτό



Επεκτάσιμο



Ανθεκτικό



Λειτουργία σε Πραγματικό
Χρόνο

Μονάδα Ανάλυσης Δεδομένων

Κέντρο ελέγχου



Μονάδα Ανάλυσης Δεδομένων



Παρακολούθηση Συσχετίσεων



Ανάκτηση Ελλιπών Δεδομένων



Αύξηση Χρονικής Ανάλυσης



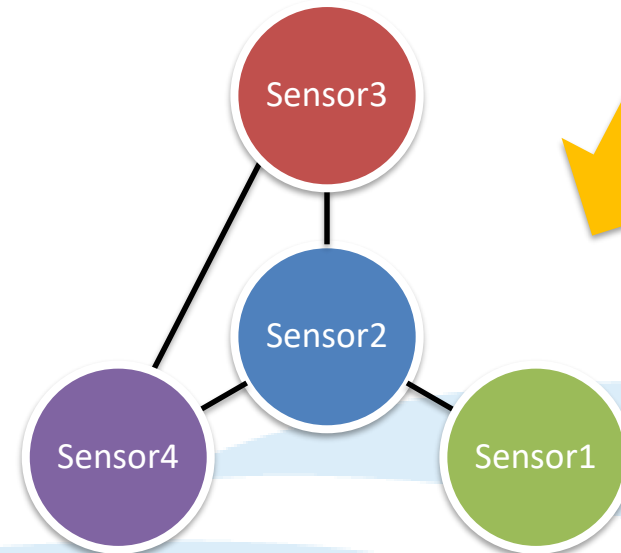
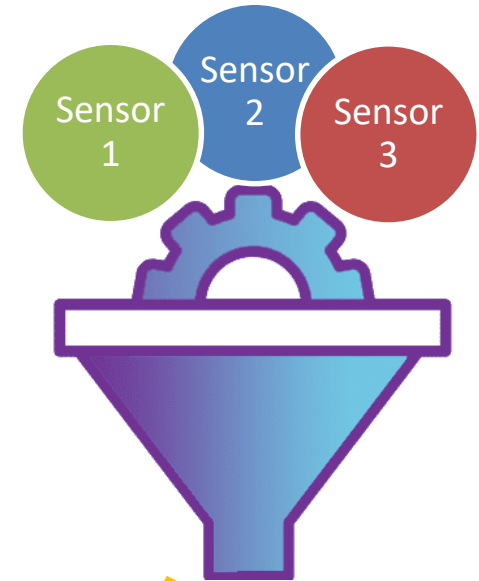
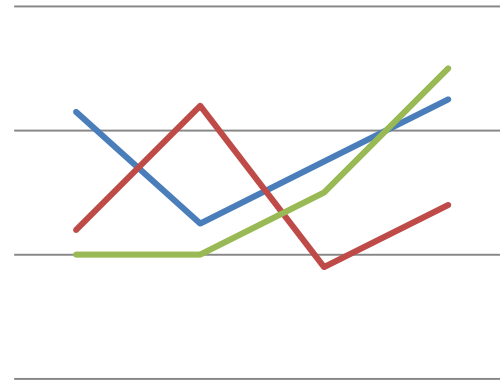
Ανίχνευση Ακραίων Συμβάντων



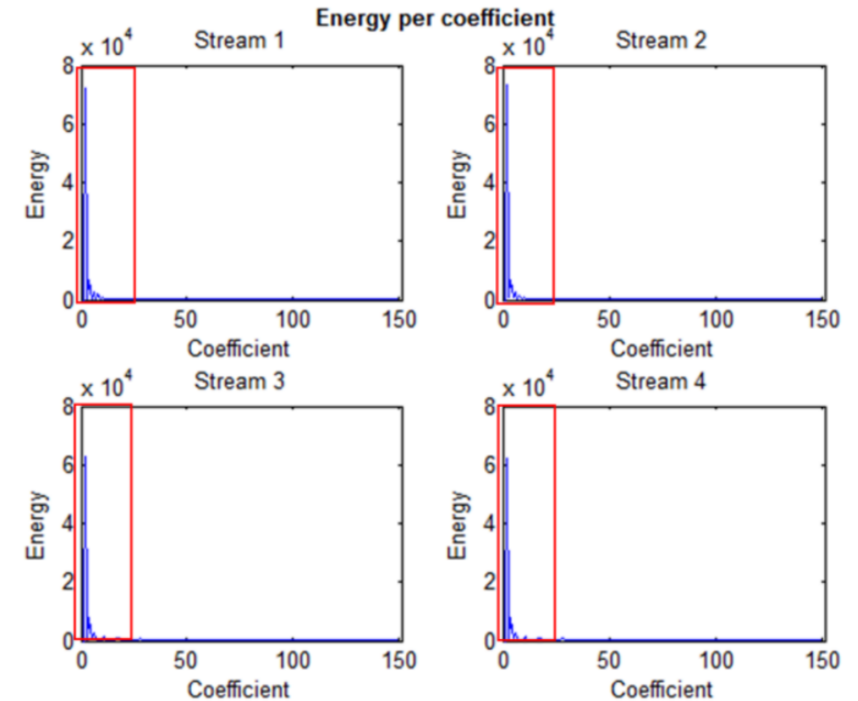
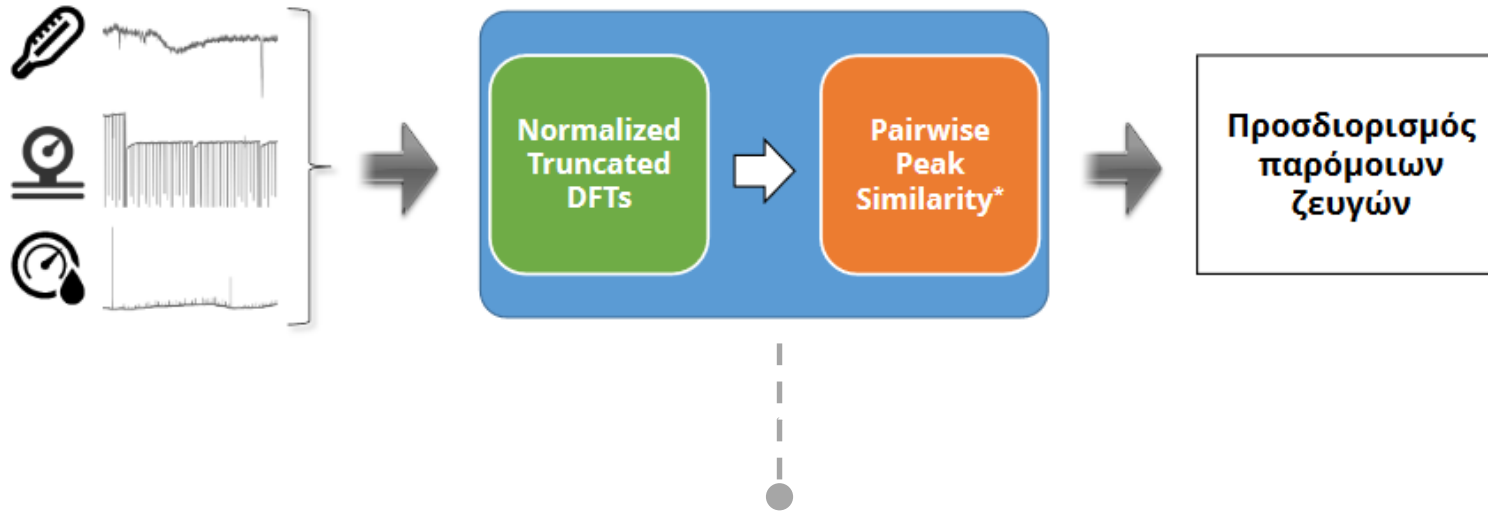
Οπτικοποίηση Δεδομένων

Παρακολούθηση Συσχετίσεων

- Ομαδοποίηση αισθητήρων
- Χρήση συσχετίσεων στην ανακατασκευή δεδομένων
- Αποκάλυψη συσχετίσεων μεταξύ φαινομενικά ανεξάρτητων δεδομένων



Παρακολούθηση Συσχετίσεων



- Προσφέρει καλύτερη χρονική απόδοση εκτέλεσης από άλλες μεθόδους
- Ο χρήστης μπορεί να καθορίσει το όριο «υψηλής συσχέτισης»

$$p_{sim}(\mathbf{s}, \mathbf{y}) = \frac{1}{M} \sum_{i=1}^M \left[1 - \frac{|S_i - Y_i|}{2 \cdot \max(|S_i|, |Y_i|)} \right]$$

- Υπολογισμός συσχετίσεων σε πραγματικό χρόνο

Function: `CorrelationMonitoring(X, threshold, thresholdCorrection, fourierCutoff, energyCutoff)`

Inputs:

@X: πίνακας με στήλες τα διανύσματα με τις μετρήσεις των αισθητήρων

@threshold: όριο πάνω από το οποίο δύο ροές δεδομένων χαρακτηρίζονται ως υψηλά συσχετισμένες

@thresholdCorrection: τιμή που προστίθεται στην υπολογισμένη συσχέτιση (στο πεδίο της συχνότητας) για τη διόρθωση της συσχέτισης στο πεδίο του χρόνου

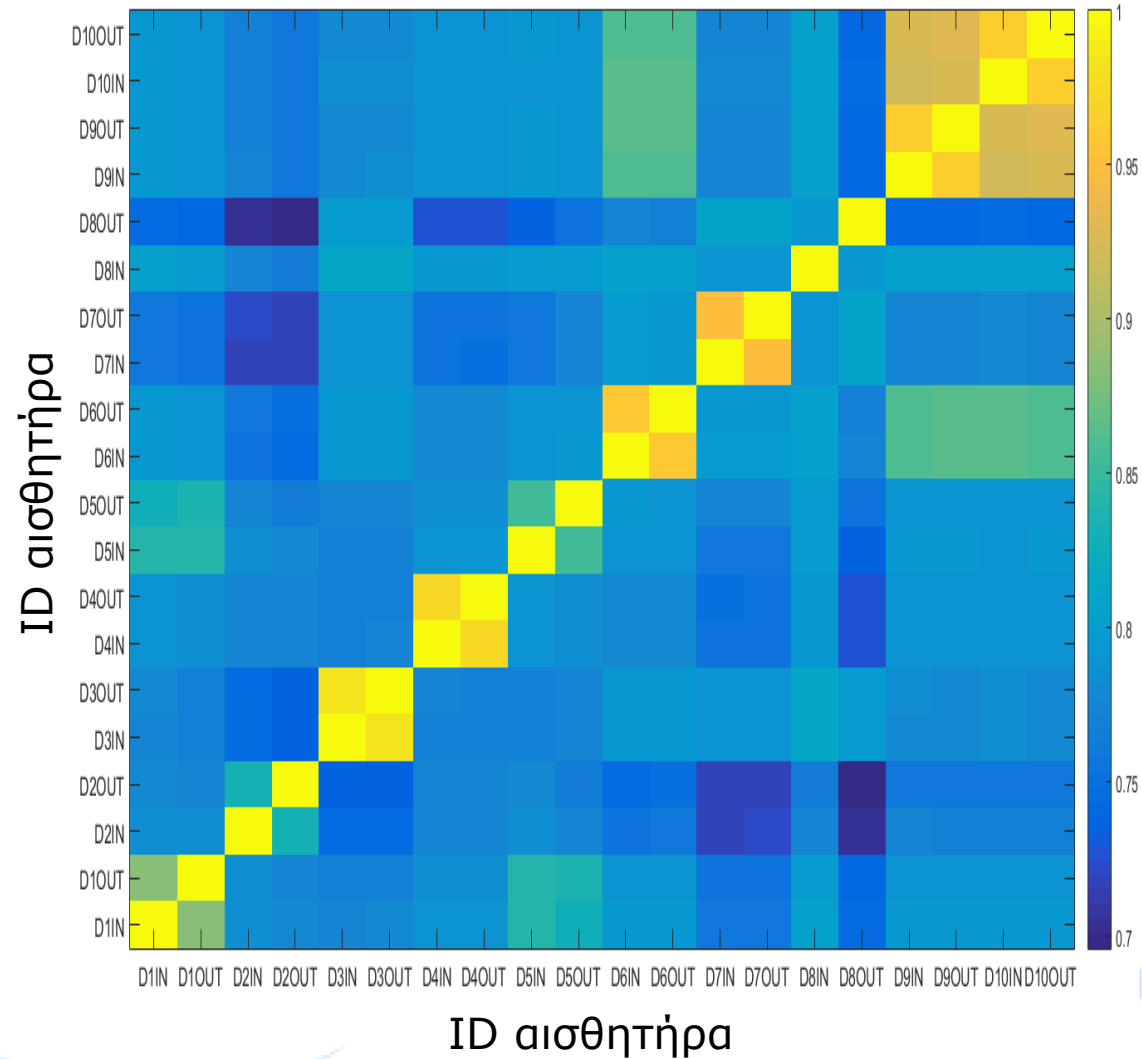
@fourierCutoff: τιμή που χρησιμοποιείται για τον προσδιορισμό του μήκους των διανυσμάτων, στο πεδίο της συχνότητας, που θα χρησιμοποιηθούν για τον υπολογισμό των συσχετίσεων (π.χ. 2 για περικοπή στο N/2 των συχνοτήτων, 4 για N/4, κλπ.)

@energyCutoff: ποσοστό της ενέργειας στο πεδίο των συχνοτήτων που καθορίζει το πλήθος των δεδομένων που θα χρησιμοποιηθούν για τον υπολογισμό της συσχέτισης

Output:

@Y: πίνακας με τις υπολογισμένες συσχετίσεις μεταξύ όλων των ζευγών ροών δεδομένων

Συσχετίσεις μεταξύ ζευγών ροών



Έλεγχος Χαμένων Μετρήσεων

- Έλεγχος ύπαρξης χαμένων μετρήσεων με σκοπό την ανακατασκευή τους

Function: `checkMissingData(x, missingID)`

Inputs:

@x: το αρχικό διάνυσμα μετρήσεων

@missingID: συμβολοσειρά που καθορίζει τον τρόπο αναπαράστασης των χαμένων μετρήσεων (π.χ. NaN, NA)

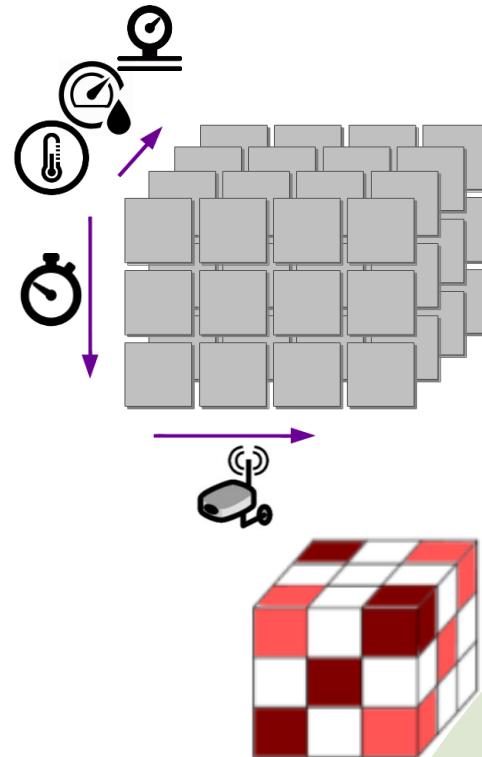
Output:

@y: διάνυσμα με τις θέσεις των χαμένων μετρήσεων

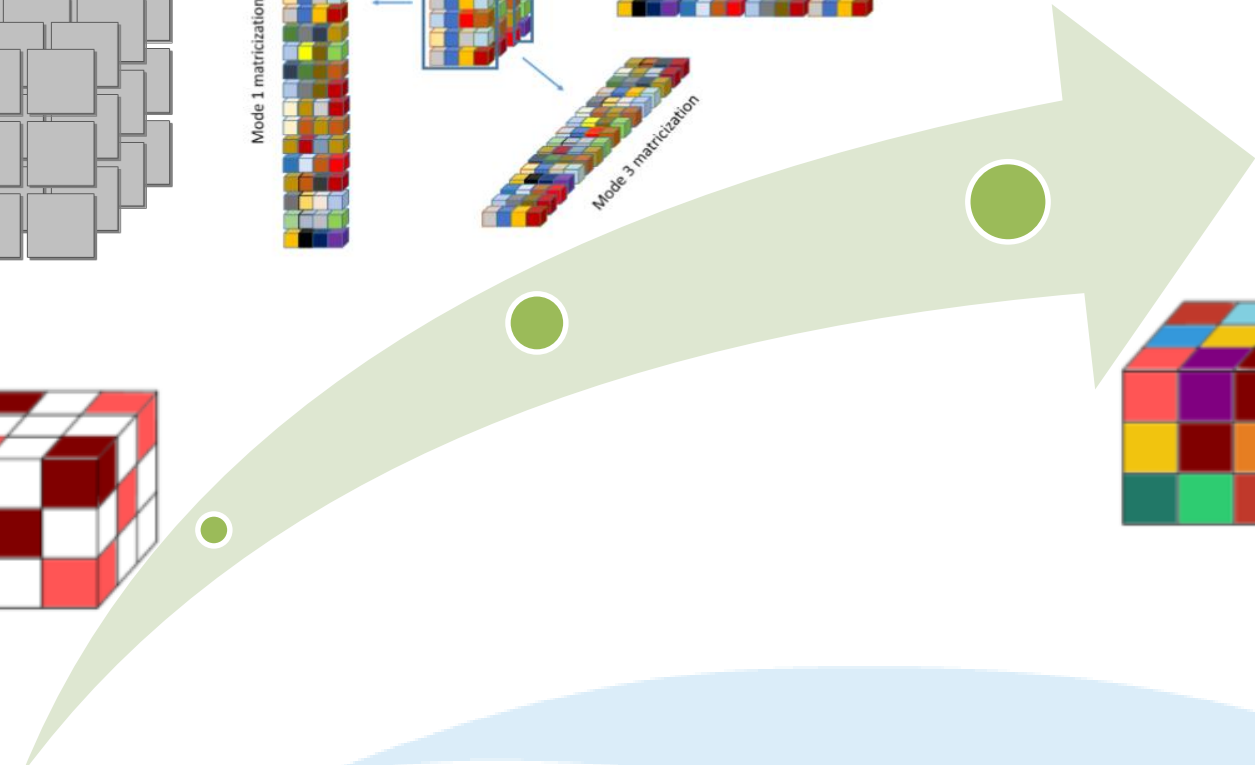
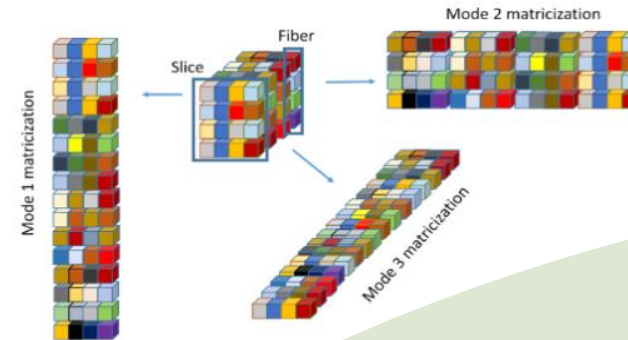
Ανάκτηση Ελλιπών Δεδομένων



Συμπλήρωση χαμένων
 μετρήσεων μεταξύ
 υψηλά συσχετισμένων
 ροών δεδομένων



Μέθοδοι συμπλήρωσης τανυστή



Λόγος Συμπλήρωσης
 (Filling Ratio (FR))

$\# \text{ Διαθέσιμων} / \# \text{ Συνόλου δεδομένων}$

Ανάκτηση Ελλιπών Δεδομένων

- Ο αλγόριθμος που υλοποιήσαμε για τη συμπλήρωση τανυστών βασίζεται στον Tmac (<https://xu-yangyang.github.io/TMac>)

Function: `TCRecovery(X, windowSize, stepSize)`

Inputs:

@X: πίνακας με στήλες τις ροές δεδομένων από υψηλά συσχετισμένους αισθητήρες

@windowSize: μέγεθος παραθύρου που χρησιμοποιείται για το Hankelization των δεδομένων

@stepSize: βήμα για τη μετάβαση στο επόμενο παράθυρο των Hankelized δεδομένων

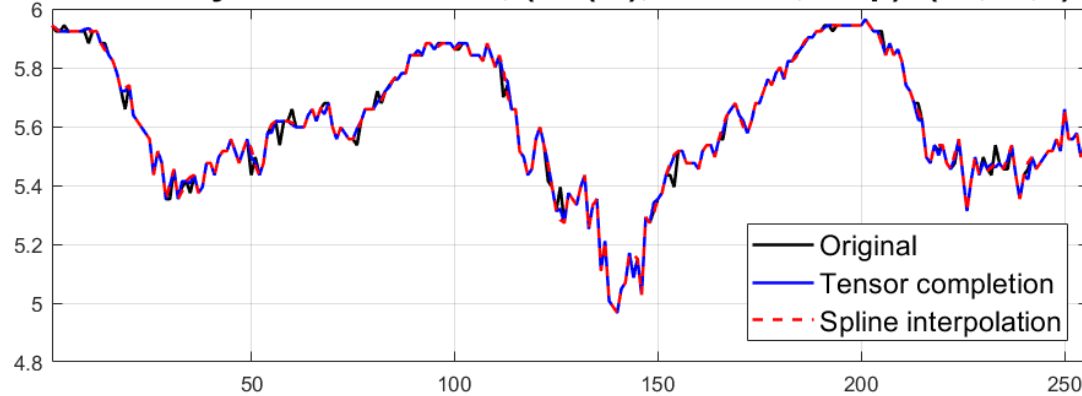
Output:

@Y: πίνακας με στήλες τις ροές με συμπληρωμένες τις χαμένες μετρήσεις

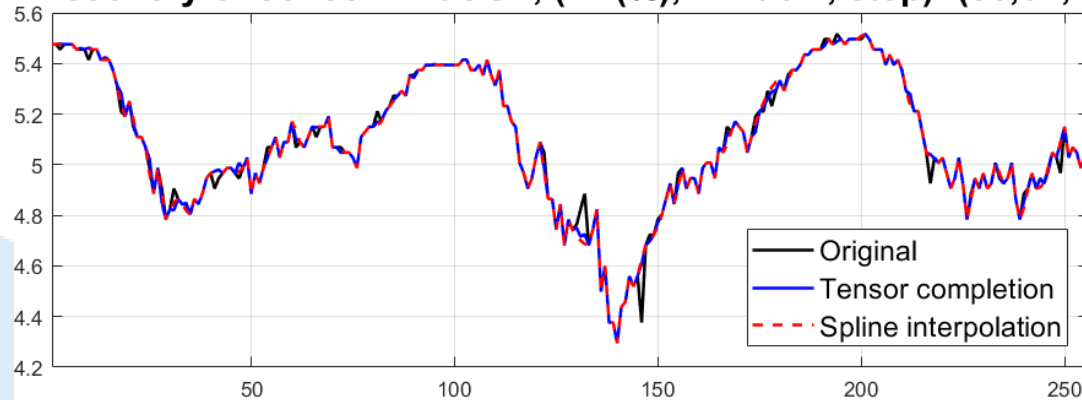


Δεδομένα ΔΕΥΑΜ

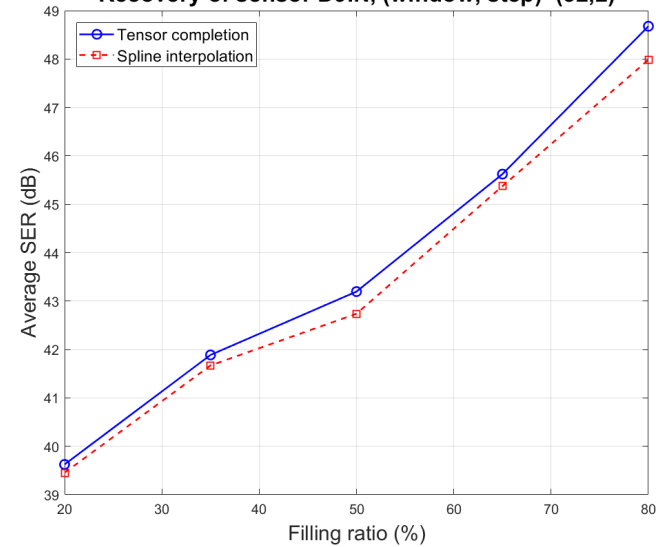
Recovery of sensor D9IN, (FR(%), window, step)=(80,32,2)



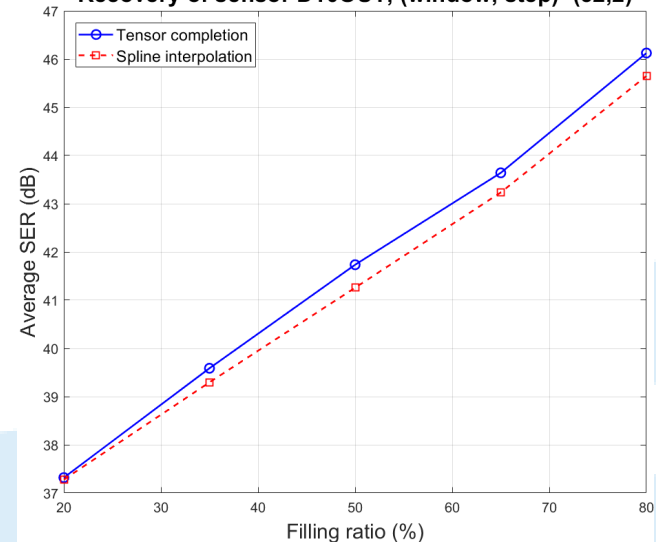
Recovery of sensor D10OUT, (FR(%), window, step)=(80,32,2)



Recovery of sensor D9IN, (window, step)=(32,2)



Recovery of sensor D10OUT, (window, step)=(32,2)



$$SER(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}$$

Αύξηση Χρονικής Ανάλυσης

π.χ. 1 μέτρηση / 15'

Τεχνητή Εισαγωγή
Ενδιάμεσων Μετρήσεων



Αύξηση αυτονομίας
αισθητήρων



Τελικό Διάνυσμα
Δεδομένων

π.χ. 1 μέτρηση / 30'



Αρχικό Διάνυσμα
Δεδομένων

Αύξηση Χρονικής Ανάλυσης

- Η επεξεργασία των δεδομένων σε μεγαλύτερες χρονικές αναλύσεις μπορεί να οδηγήσει στην εξαγωγή ακριβέστερων συμπερασμάτων για την κατάσταση του συστήματος

Function: *SuperResolution*(x, delta)

Inputs:

@**x**: αρχικό διάνυσμα μετρήσεων μήκους N

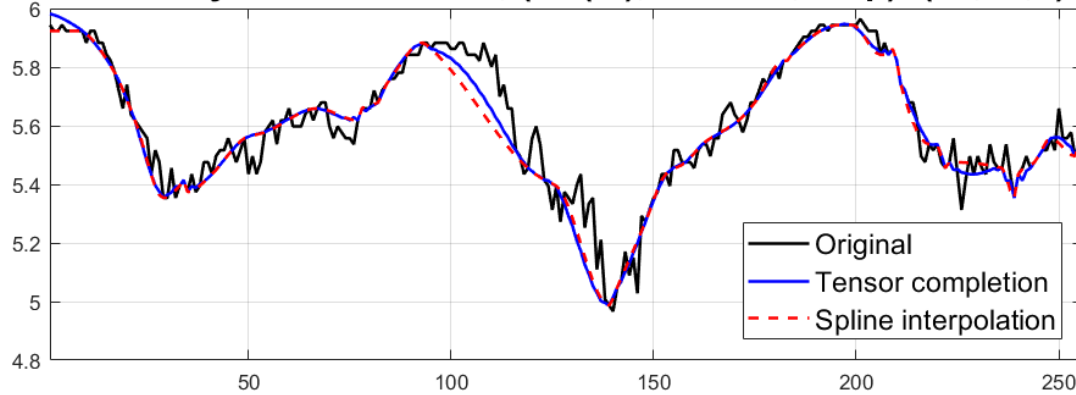
@**delta**: ακέραιο πολλαπλάσιο της αρχικής συχνότητας δειγματοληψίας των μετρήσεων

Output:

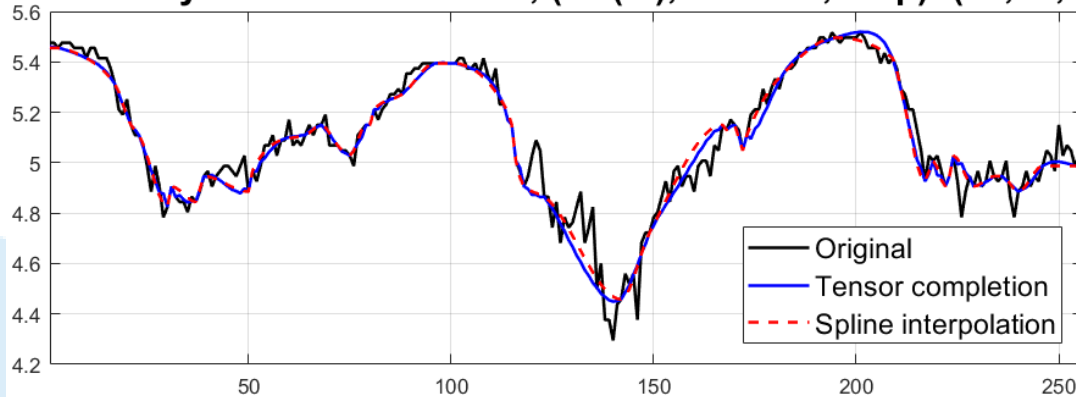
@**y**: ανακατασκευασμένο διάνυσμα με $\text{delta} \cdot N$ μετρήσεις

Δεδομένα ΔΕΥΑΜ

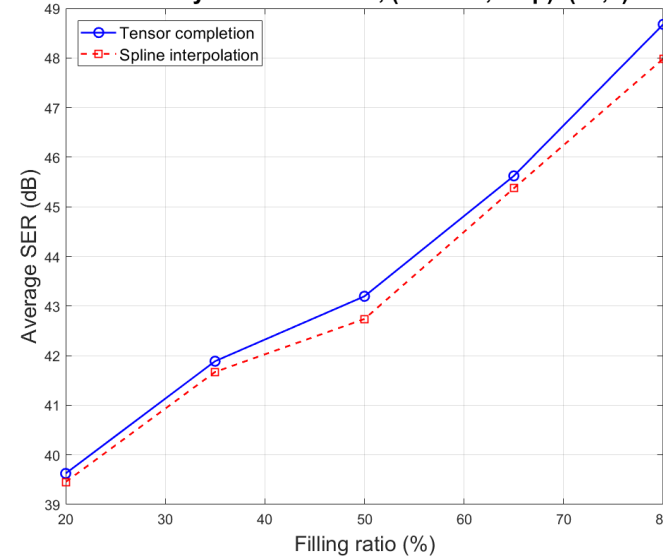
Recovery of sensor D9IN, (FR(%), window, step)=(20,32,2)



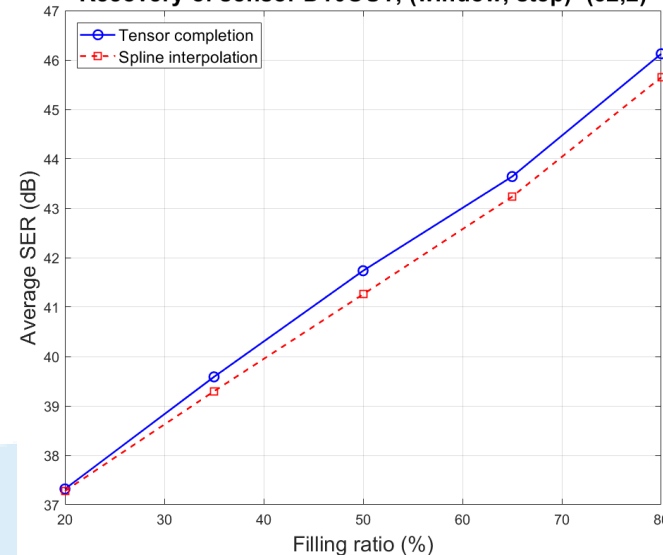
Recovery of sensor D10OUT, (FR(%), window, step)=(20,32,2)



Recovery of sensor D9IN, (window, step)=(32,2)



Recovery of sensor D10OUT, (window, step)=(32,2)

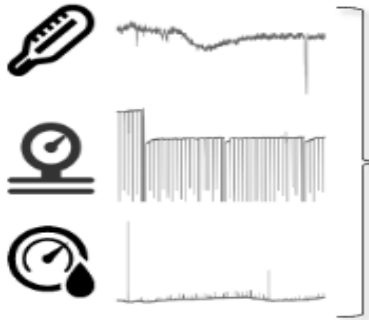


$$SER(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}$$

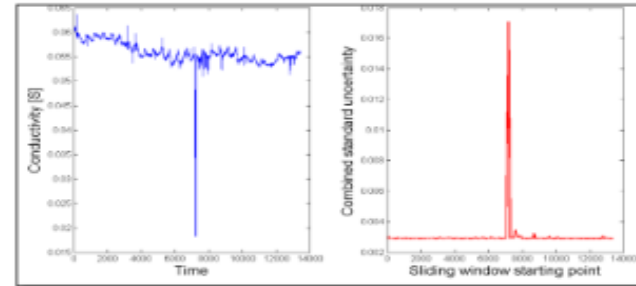
Ανίχνευση Ακραίων Συμβάντων με Αξιοποίηση Αβεβαιότητας



Έγκαιρη προειδοποίηση για σωστή λήψη αποφάσεων



Μέθοδος Spreadsheet
(εφαρμόζεται στα πρωτογενή δεδομένα, στιγμιαία αβεβαιότητα)



		Source of uncertainty	Value (\pm)	Probability distribution	Divisor	Standard uncertainty $u(x)$
Type B	Sensor	Calibration	C_1	Normal	2	$C_1 / 2$
		Precision (Resolution)	C_2	Rectangular	$\sqrt{3}$	$C_2 / \sqrt{3}$
		Sensitivity	C_3	Rectangular	$\sqrt{3}$	$C_3 / \sqrt{3}$
	ECA chip	Calibration	C_4	Normal	2	$C_4 / 2$
		Precision (Resolution)	C_5	Rectangular	$\sqrt{3}$	$C_5 / \sqrt{3}$
		Sensitivity	C_6	Rectangular	$\sqrt{3}$	$C_6 / \sqrt{3}$
Type A	Sensor density	C_7	Rectangular	$\sqrt{3}$	$C_7 / \sqrt{3}$	
	Sampling	C_8	Rectangular	$\sqrt{3}$	$C_8 / \sqrt{3}$	
		Combined standard uncertainty $u_{c,1}(y)$				
		Coverage factor k_2				
		Expanded uncertainty U_2				

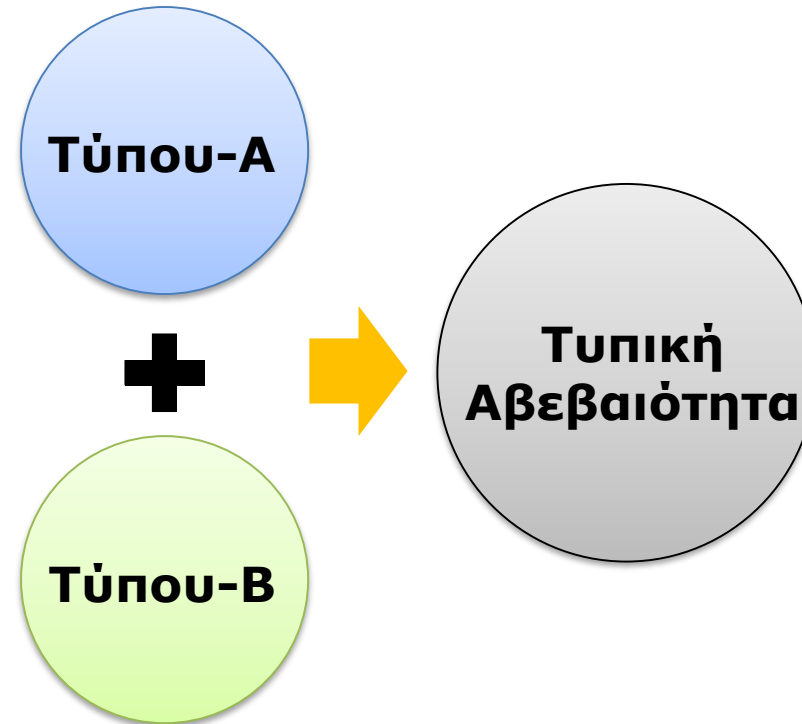
← Λίστα πιθανών πηγών αβεβαιότητας και αντίστοιχες κατανομές πιθανότητας

Αβεβαιότητα Τύπου-A

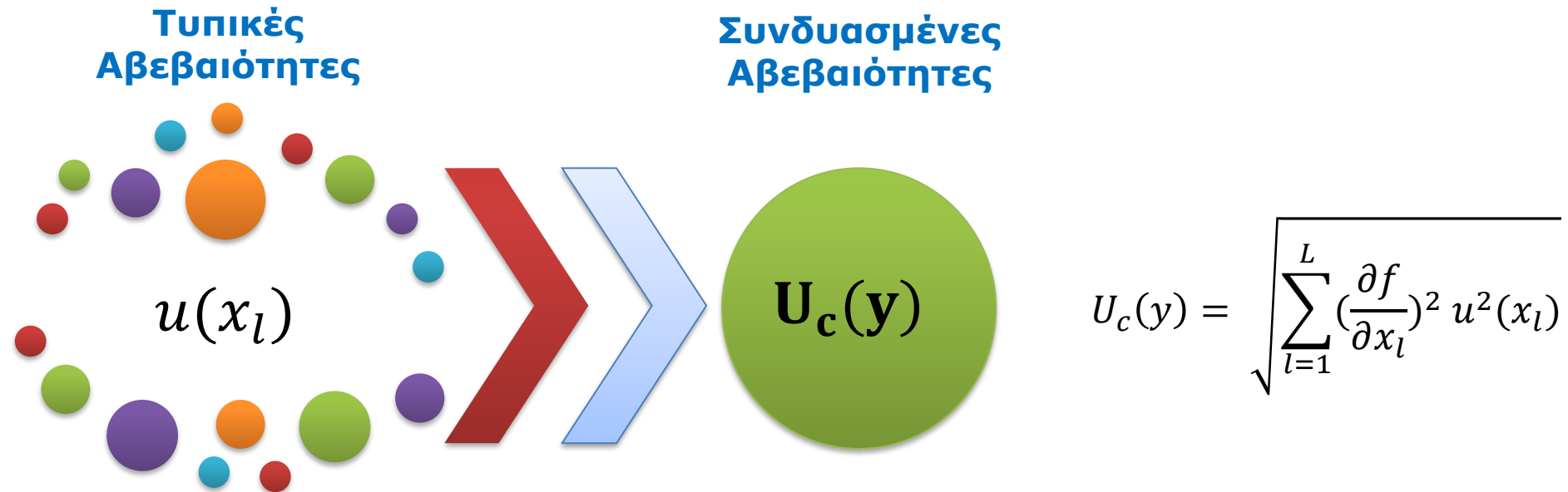
- Στατιστική αβεβαιότητα
- Συνθήκες λειτουργίας εξοπλισμού
- Κατασκευαστική ποιότητα
- ΔΕ μπορεί να ελαττωθεί

Αβεβαιότητα Τύπου-B

- Συστηματική αβεβαιότητα
- Μαθηματικά μοντέλα, μέθοδοι βαθμονόμησης
- ΜΠΟΡΕΙ να μειωθεί βελτιώνοντας τα αντίστοιχα μοντέλα



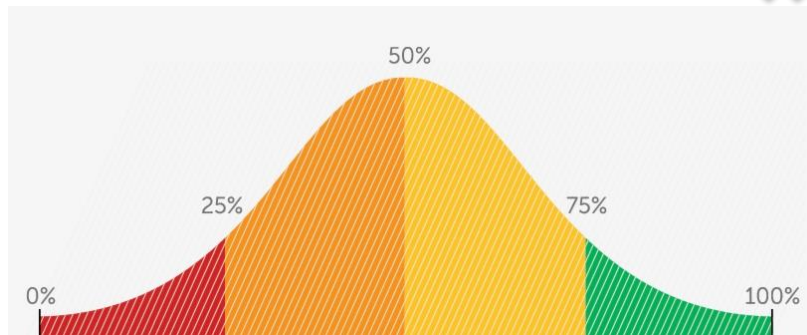
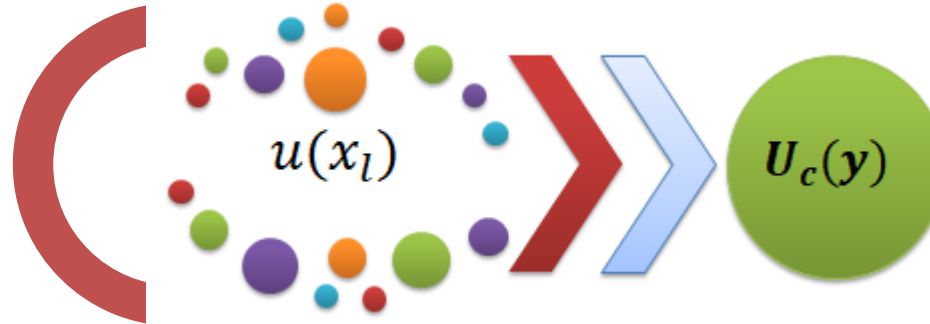
Συνδυασμένη Τυπική Αβεβαιότητα



- Υποθέτουμε ότι οι πηγές αβεβαιότητας είναι ανεξάρτητες
- L ανεξάρτητες τυπικές αβεβαιότητες $u(x_l)$
- Οι επιμέρους αβεβαιότητες πρέπει να εκφράζονται στις ίδιες μονάδες

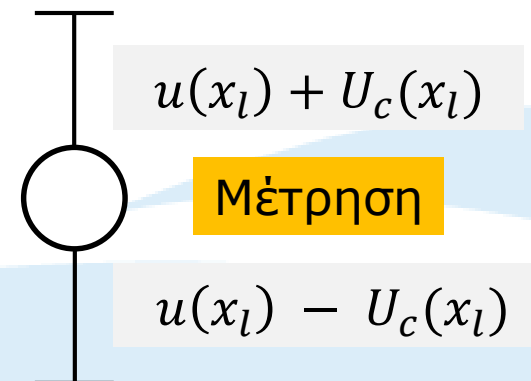
Εκτεταμένη Αβεβαιότητα

□ Συνδυασμένη Τυπική Αβεβαιότητα











□ **K**: καθορίζεται από το επιθυμητό επίπεδο εμπιστοσύνης της Gaussian κατανομής

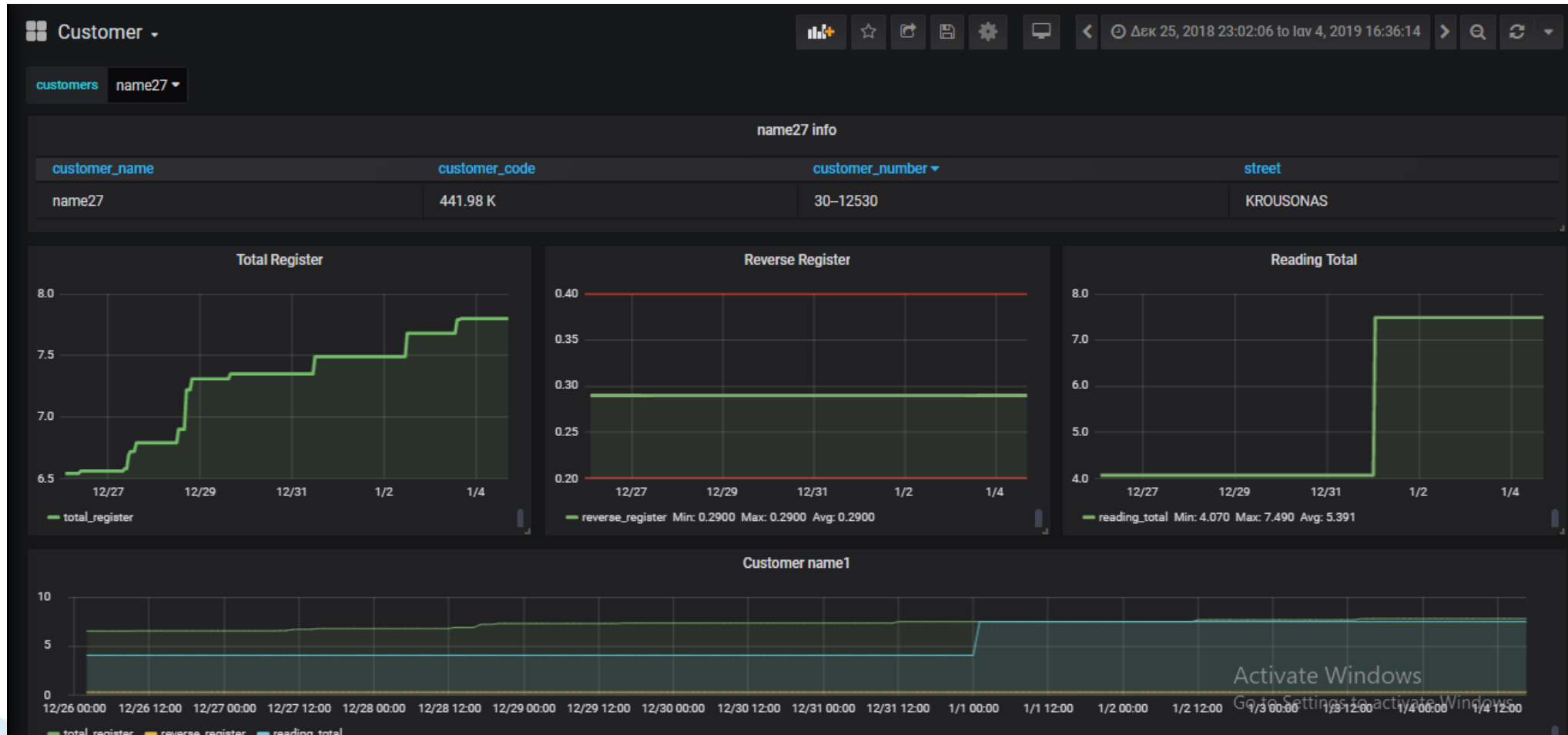
□ Η εκτεταμένη αβεβαιότητα **U** είναι το τελικό αποτέλεσμα της διαδικασίας εκτίμησης αβεβαιότητας



Ειδοποιήσεις Ακραίων Συμβάντων

				
Case				
Description	both the measurement and the expanded uncertainty interval are above the upper limit	the measurement is larger than limit and the expanded uncertainty interval contain	the measurement is lower than Limit and the expanded uncertainty interval contains it	both the measurement and the expanded uncertainty interval are below
Action	triggers clearly an alerting notification for the occurrence of an extreme event	possible divergences from normal operation	possible divergences from normal operation	is the only one that is in compliance with the specifications

Alerts	
	OK
	Warning
	Error



Δεδομένα ΔΕΥΑΜ



Σύνοψη

- Η μονάδα υψηλού επιπέδου ανάλυσης δεδομένων είναι αναπόσπαστο κομμάτι στις σύγχρονες πλατφόρμες ευφυών δικτύων ύδρευσης
- Προσαρμόσιμη στις απαιτήσεις και τις ανάγκες του εκάστοτε οργανισμού
- Ο σωστός σχεδιασμός εγγυάται εύκολη επέκταση με νέες λειτουργίες και αναβάθμιση των υπαρχουσών
- Εγγυάται την ομαλή λειτουργία των ευφυών δικτύων ύδρευσης, λαμβάνοντας υπόψη και την εγγενή αβεβαιότητα των δεδομένων

