



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ ΚΑΙ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

**Διερεύνηση του προβλήματος πρόγνωσης παραγωγής
ενέργειας σε διάφορες χρονικές κλίμακες από
Υδροηλεκτρικούς Ταμιευτήρες**

Βατσικουρίδης Ιωάννης

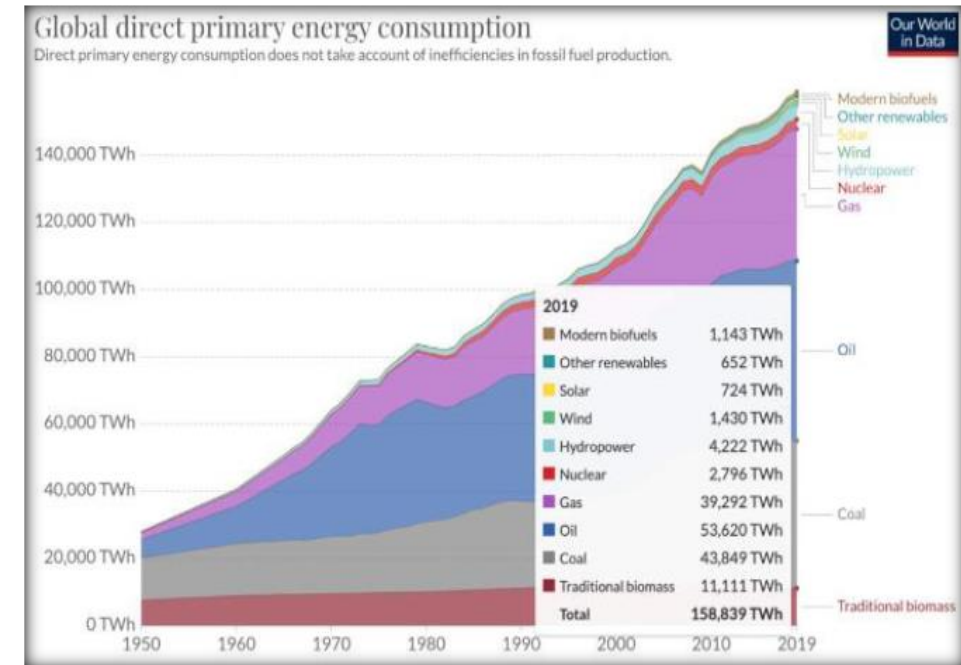
Επιβλέπων: Ανδρέας Ευστρατιάδης
Συνεπιβλέπων: Ιωάννης Τσουκαλάς

Αθήνα, Μάρτιος 2022

Υδροηλεκτρική Ενέργεια και Έργα

ΣΗΜΑΝΤΙΚΟΤΕΡΑ ΠΛΕΟΝΕΚΤΗΜΑΤΑ

- Η υδροηλεκτρική ενέργεια, από μεγάλης κλίμακας έργα (με δυνατότητα ταμίευσης), είναι χρονικά ρυθμιζόμενη, γεγονός που της προσδίδει σημαντική αξία.
- Η αναρρύθμιση των απορροών μέσω της αποθήκευσης, καθιστά τους ταμιευτήρες ικανούς να μπορούν να ανταποκριθούν άμεσα στις διακυμάνσεις της ζήτησης ηλεκτρικής ενέργειας. Συνεπώς, τα έργα μεγάλης ισχύος λειτουργούν για την κάλυψη των αιχμών του δικτύου κατά την διάρκεια των οποίων η αξία της KWh είναι πολλαπλάσια της αξίας της KWh βάσης.
- Οι βαθμοί απόδοσης των υδροστρόβιλων των υδροηλεκτρικών έργων μπορούν να υπερβούν το 90%, ποσοστό πολύ μεγάλο σε σχέση με άλλες πηγές ενέργειας. Για παράδειγμα, η μετατροπή της κινητικής ενέργειας του ανέμου σε ηλεκτρική έχει βαθμούς απόδοσης 30 – 40% και η μετατροπή της ηλιακής ακτινοβολίας σε ηλεκτρική ενέργεια 12 – 17%.



Αντικείμενο και Σκοπός μελέτης (1/2)

- Αντικείμενο της εργασίας αποτελεί η διενέργεια προγνώσεων απορροής σε μεγάλους υδροηλεκτρικούς ταμιευτήρες που αποσκοπούν στην εκτίμηση της παραγωγής ενέργειας σε μηνιαία κλίμακα. Το ενδιαφέρον προκύπτει ακριβώς από την αποθηκευτική ικανότητα των ταμιευτήρων που οδηγεί σε εξομάλυνση των σφαλμάτων της ενέργειας, όταν το ζητούμενο είναι μια σταθερή απόληψη.
- Για το σκοπό αυτό επιλέχθηκαν **3 ντετερμινιστικά μοντέλα πρόγνωσης** διαφορετικών οικογενειών και πολυπλοκότητας για 3 ταμιευτήρες διαφορετικών χαρακτηριστικών και αναρρυθμιστικής ικανότητας. Πρόκειται για τα φράγματα των Κρεμαστών, της Μεσοχώρας και του Ευήνου. Οι χρονικοί ορίζοντες πρόγνωσης είναι αυτοί των **4** και **12** μηνών αντίστοιχα.



AR(1)

Ένα φειδωλό, απλό αυτοπαλίνδρομο μοντέλο 1^{ης} τάξης, με μικρό υπολογιστικό φόρτο που είχε το ρόλο του baseline μοντέλου.



LSTM

Ένα σύνθετο μοντέλο επιβλεπόμενης μηχανικής μάθησης, τεχνητών νευρωνικών δικτύων ικανό για πρόγνωση σειριακών δεδομένων αλλά μεγάλου υπολογιστικού φόρτου .

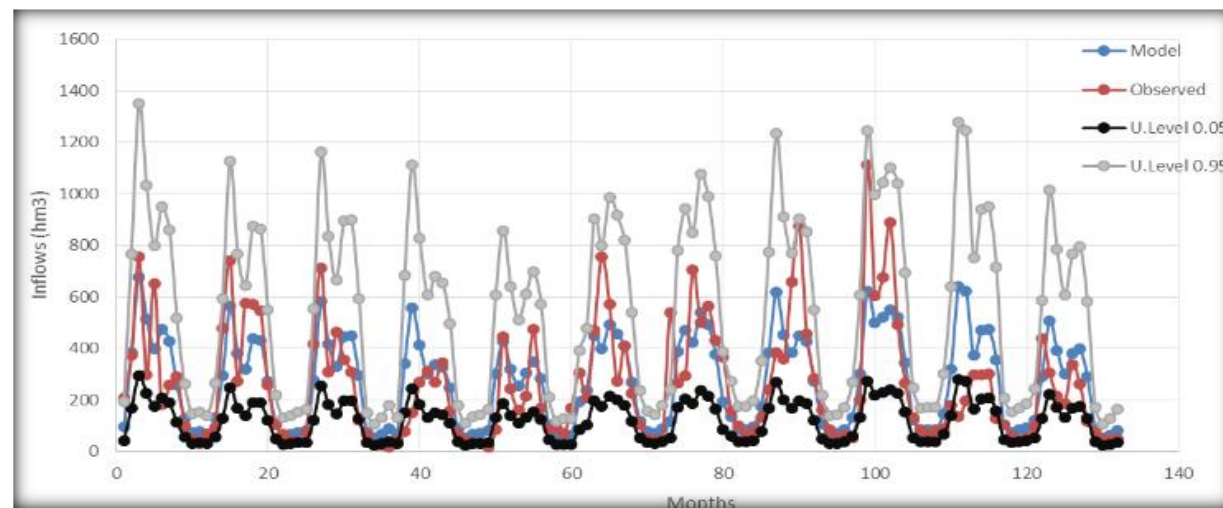
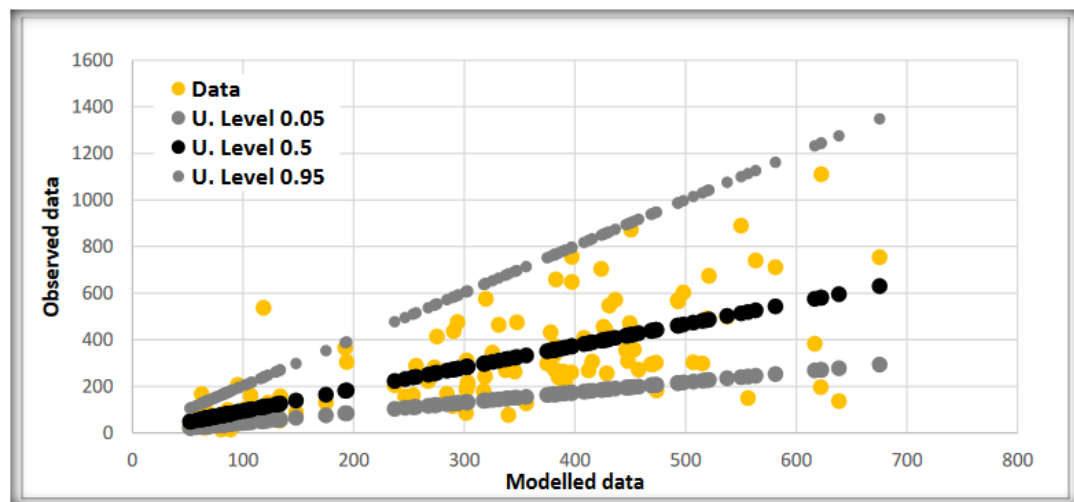


KNN

Ένα μοντέλο μη επιβλεπόμενης μηχανικής μάθησης που βασίζεται στην ομοιότητα των ιστορικών δεδομένων με τα προβλεπόμενα.

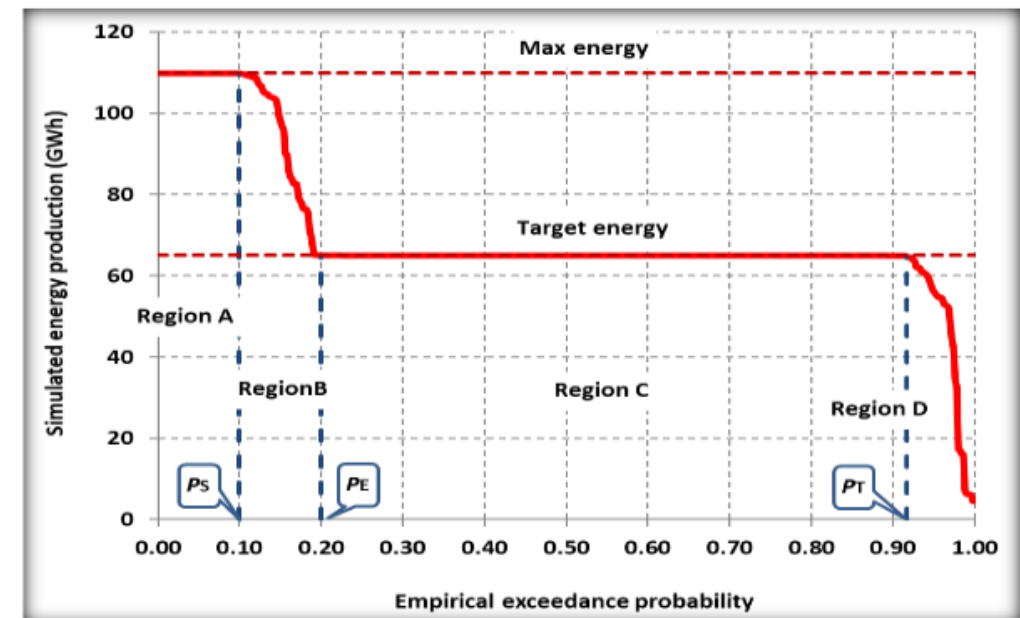
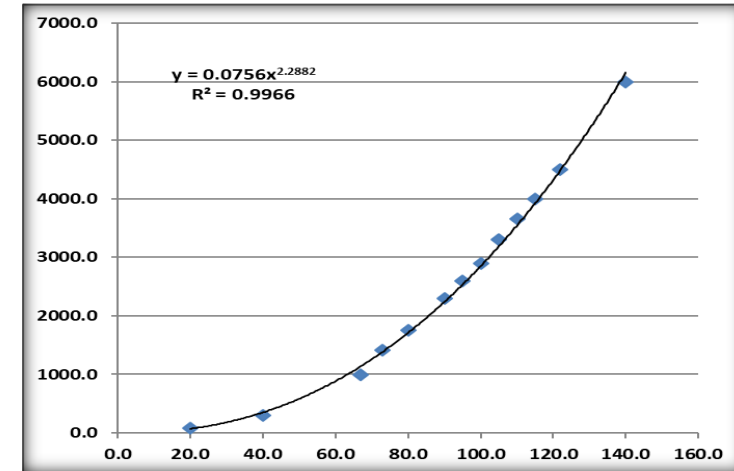
Αντικείμενο και Σκοπός μελέτης (2/2)

- Κατά τη διάρκεια εκπόνησης της εργασίας έγινε αντιληπτή η μεγάλη σημασία της ποσοτικοποίησης της αβεβαιότητας και η αναγκαιότητά της. Σε ένα αντικείμενο μελέτης που η κρίση και οι αποφάσεις του διαχειριστή του έργου είναι τόσο καθοριστικές δε νοείται να μη ληφθεί υπόψη η αβεβαιότητα.
- Έτσι λοιπόν στο 2^ο Μέρος της εργασίας γίνεται χρήση της μεθοδολογίας των συζεύξεων (copulas) για πρόγνωση απορροής. Οι συζεύξεις είναι συναρτήσεις που συνδέουν την πολυδιάστατη συνάρτηση κατανομής με τις μονοδιάστατες περιθώριες συναρτήσεις και επιτρέπουν την παραγωγή δεσμευμένων κατανομών.
- Τα αποτελέσματα προγνώσεων του μοντέλου-νικητή στις σημειακές προβλέψεις, χρησιμοποιούνται στην πιθανοτική μέθοδο πρόγνωσης.



Προσομοίωση Λειτουργίας Ταμιευτήρων

- Για τις προσομοιώσεις θα γίνει χρήση των εξισώσεων του υδατικού ισοζυγίου για κάθε ταμιευτήρα που περιγράφει τις σχέσεις μεταξύ προσφοράς και ζήτησης νερού για την κάλυψη αναγκών.
- Απαραίτητα δεδομένα εισόδου για την προσομοίωση είναι:
 1. Χρονοσειρά Εισροών
 2. ΑΣΛ-ΚΣΛ
 3. Στάθμη εξόδου φυγής
 4. Παροχτευτικότητα αγωγού προσαγωγής
 5. Σχέσεις μεταβολής της επιφάνειας και του αποθέματος
 6. Ζητήσεις Νερού (στόχος παραγωγής πρωτεύουσας ενέργειας-προκύπτει από βελτιστοποίηση)



Μοντέλο Αυτοπαλινδρόμησης 1^{ης} Τάξης AR(1) (1/5)

- Ανήκει στην οικογένεια των ARMA μοντέλων των οποίων η γενική μορφή είναι η εξής:

$$Y_t = \delta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

- Αν θεωρήσουμε μόνο μια υστέρηση της χρονοσειράς Y_t και μόνο το τρέχον σφάλμα ε_t , τότε λαμβάνουμε το αυτοπαλίνδρομο μοντέλο πρώτης τάξεως AR(1) στην εξής μορφή:

$$Y_t = \delta + \alpha_1 Y_{t-1} + \varepsilon_t$$

όπου $\delta = \mu(1 - \alpha)$

- Με τη μέθοδο των Ροπών και τη χρήση των εξισώσεων Yule-Walker εκτιμάμε την παράμετρο α , του συντελεστή του μοντέλου αυτοπαλινδρόμησης. Εν τέλει καταλήγουμε πως:

$$\alpha = \widehat{\rho}_1$$

όπου $\widehat{\rho}_1$ είναι ο δειγματικός συντελεστής αυτοσυσχέτισης 1^{ης} τάξης.

- Τελικά το μοντέλο θα πάρει τη μορφή:

$$Y_t = \mu(1 - \alpha_1) + \alpha_1 Y_{t-1}$$

αφού για τις ντετερμινιστικές προγνώσεις θεωρούμε την κεντρική τιμή της κατανομής των σφαλμάτων

Μοντέλο Αυτοπαλινδρόμησης 1^{ης} Τάξης AR(1) (2/5)

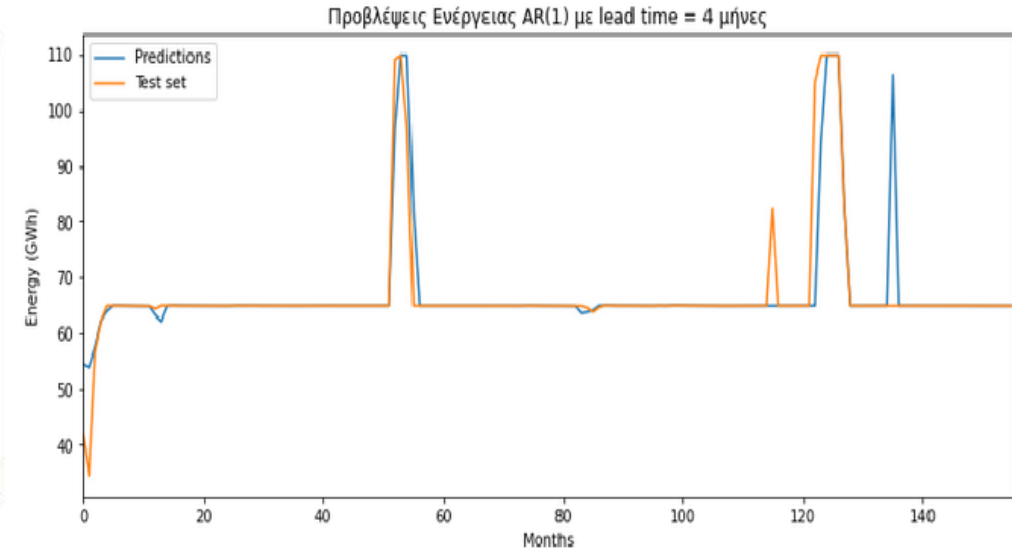
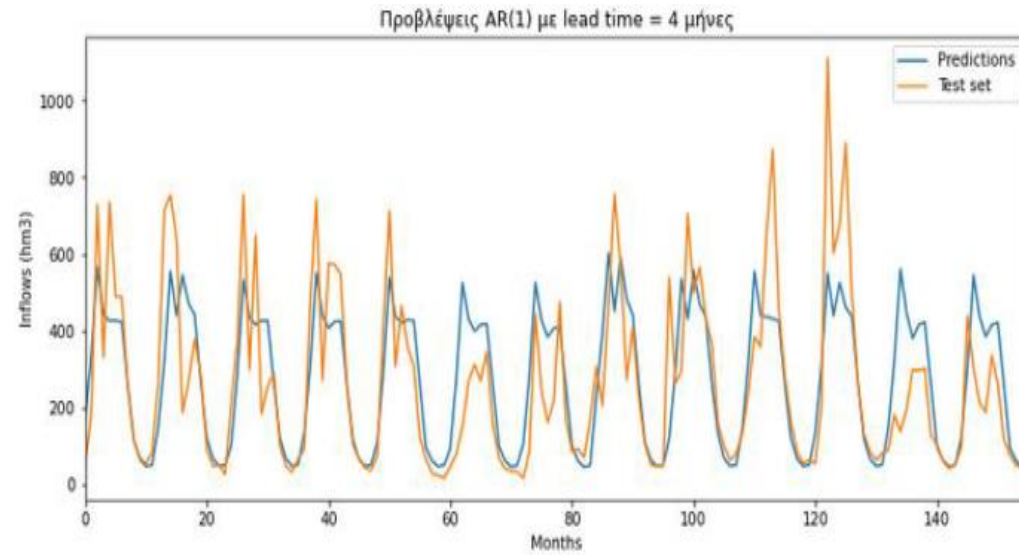
ΜΕΘΟΔΟΛΟΓΙΑ

1. Τυποποίηση χρονοσειράς απορροής \longrightarrow Διαίρεση με τη μέση τιμή κάθε μήνα
2. Χωρισμός του δείγματος σε training set και data set \longrightarrow αναλογία 3:1
3. Η εκτίμηση των παραμέτρων θα γίνει με τη χρήση του training set.
4. Ανά 4άδα ή 12άδα προβλέψεων γίνεται ενημέρωση του training set άρα και των τιμών των παραμέτρων του μοντέλου. Ενσωματώνεται η νέα στατιστική πληροφορία του δείγματος.
5. Αποτυποποίηση χρονοσειράς Εισροών \longrightarrow Πολλαπλασιασμός με τη μέση τιμή κάθε μήνα
6. Οι αριθμητικές τιμές των παραμέτρων αλλάζουν με το χρόνο. Παρόλο που η διαδικασία αυξάνει την ακρίβεια στις προβλέψεις αφού χρησιμοποιείται περισσότερη ιστορική πληροφορία, η επίδραση αναμένεται να είναι μακροπρόθεσμη.
7. Ενδεικτική μορφή του μοντέλου στην τελευταία 4άδα προβλέψεων:

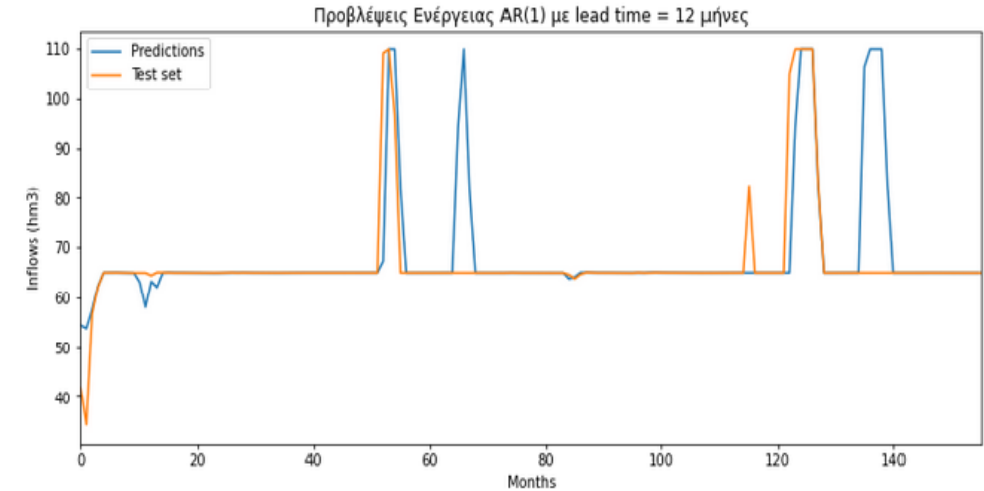
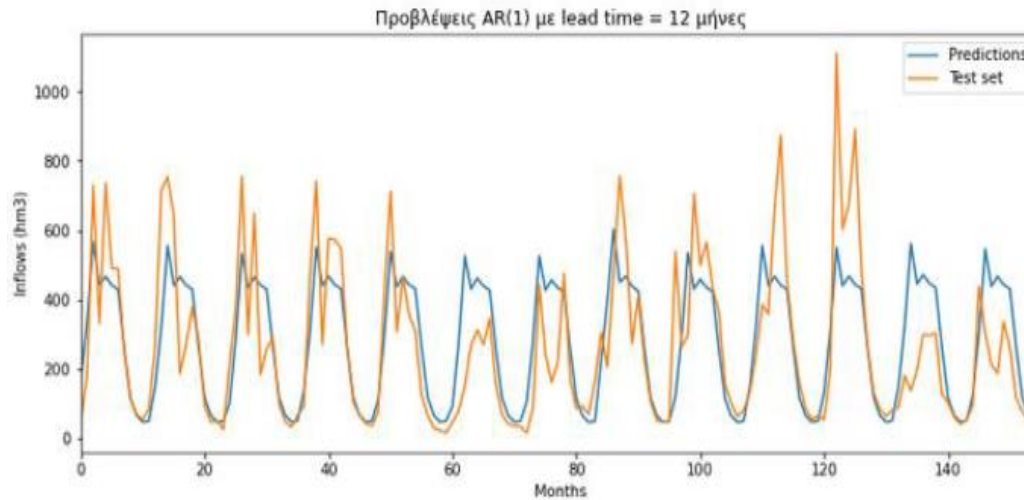
$$Y_t = 0.66 + 0.34 Y_{t-1}$$

Μοντέλο Αυτοπαλινδρόμησης 1ης Τάξης AR(1) (3/5)

ERROR METRICS	ΚΡΕΜΑΣΤΑ	
	AR(1)	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4	
MAPE (%)	51.69	1.86
MPE (%)	-30.34	-0.62
RMSE	149.69	5.69
R ²	0.57	0.7

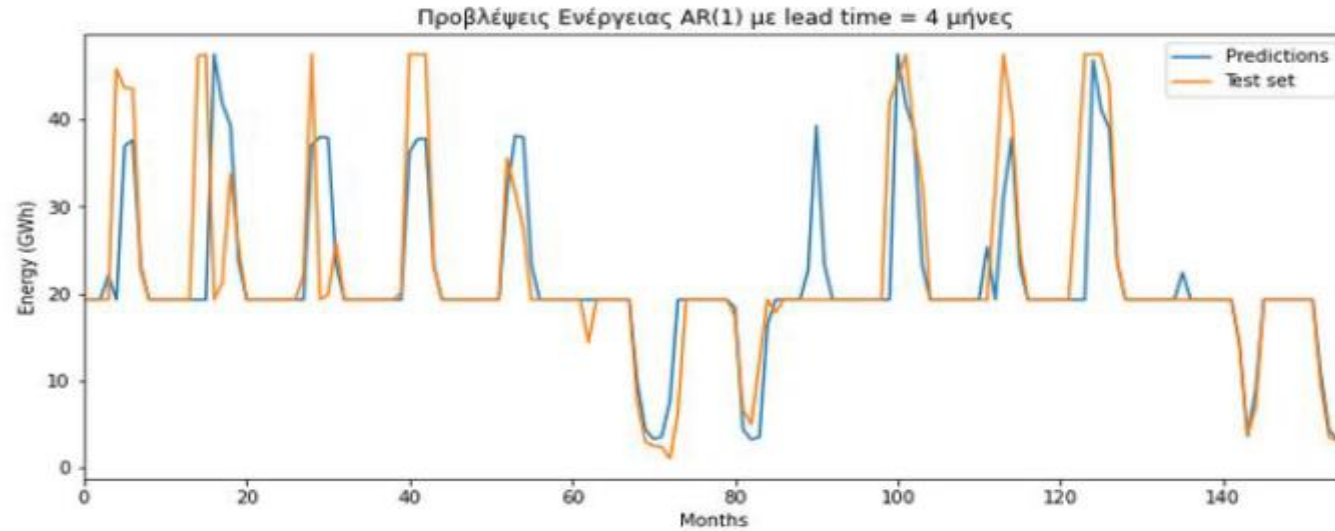


ERROR METRICS	ΚΡΕΜΑΣΤΑ	
	AR(1)	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12	
MAPE (%)	53.6	4.5
MPE (%)	-32.16	-2.78
RMSE	150.89	10.2
R ²	0.56	0.34

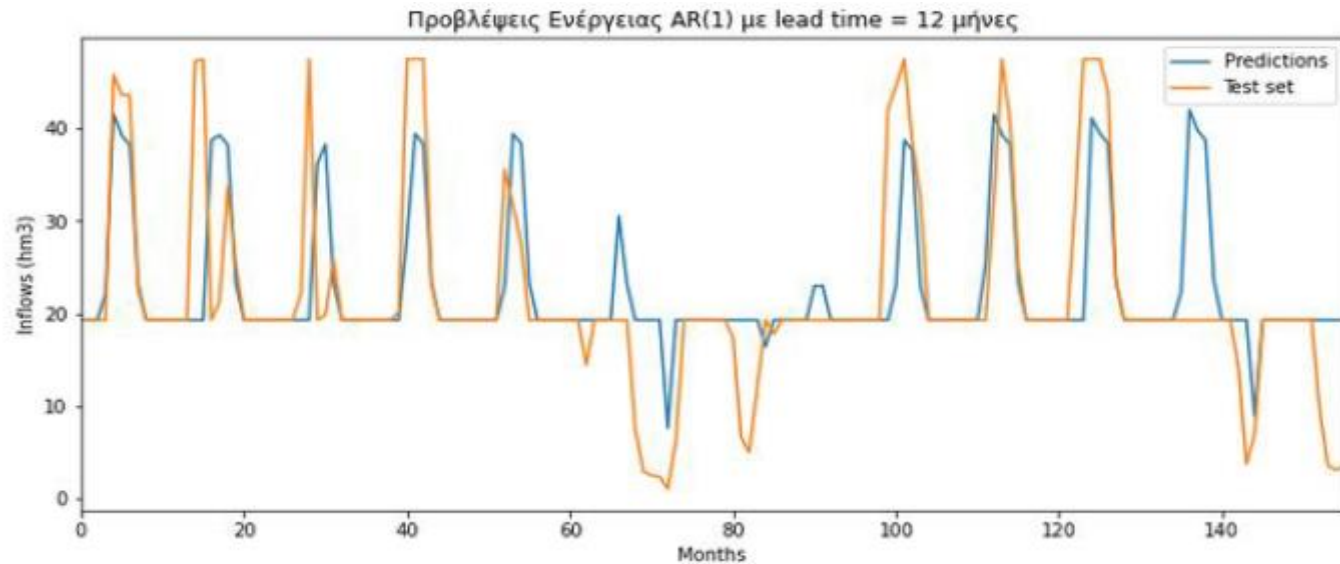


Μοντέλο Αυτοπαλινδρόμησης 1^{ης} Τάξης AR(1) (4/5)

ERROR METRICS	ΜΕΣΟΧΩΡΑ	
	AR(1)	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4	
MAPE (%)	51.69	16.87
MPE (%)	-30.34	-6.55
RMSE	26.49	7.08
R ²	0.57	0.55

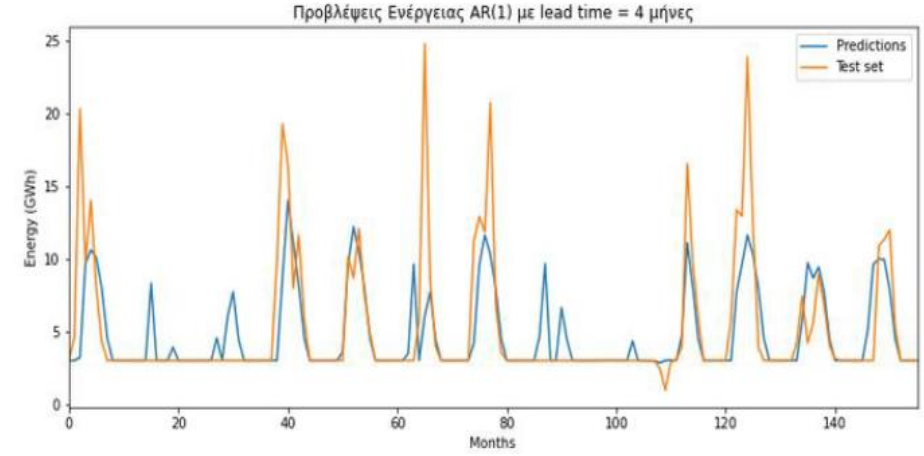
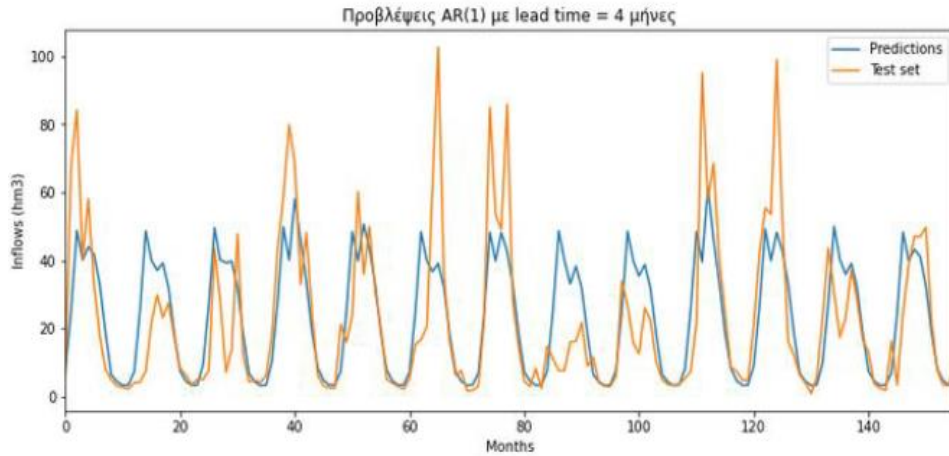


ERROR METRICS	ΜΕΣΟΧΩΡΑ	
	AR(1)	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12	
MAPE (%)	53.6	45.94
MPE (%)	-32.16	-37
RMSE	26.71	8.54
R ²	0.56	0.34

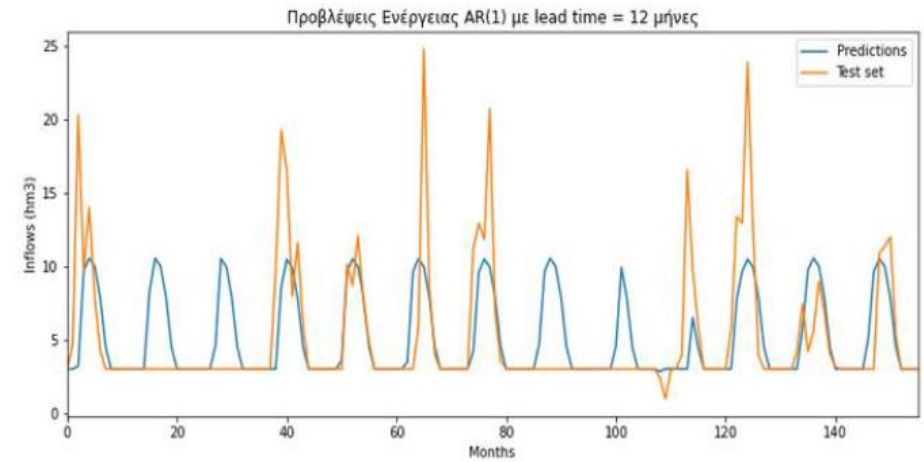
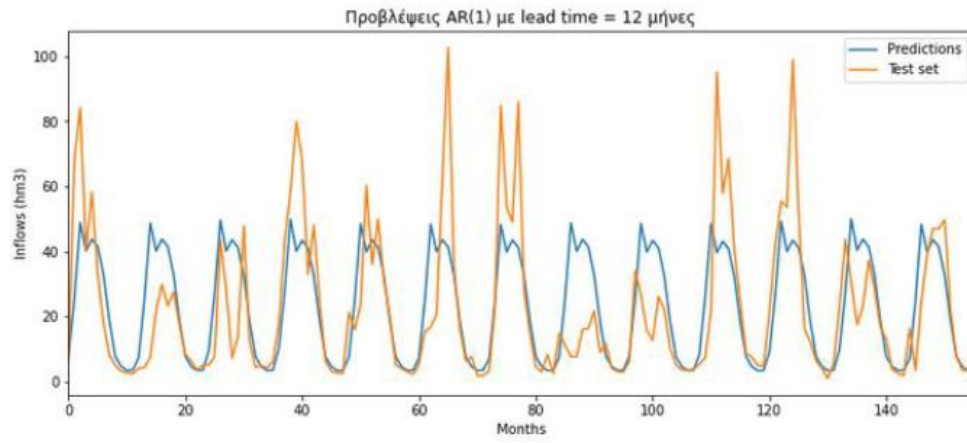


Μοντέλο Αυτοπαλινδρόμησης 1ης Τάξης AR(1) (5/5)

ERROR METRICS	ΕΥΗΝΟΣ	
	AR(1)	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4	
MAPE (%)	67.6	22.8
MPE (%)	-43.22	-9.7
RMSE	16.05	3.22
R ²	0.51	0.46



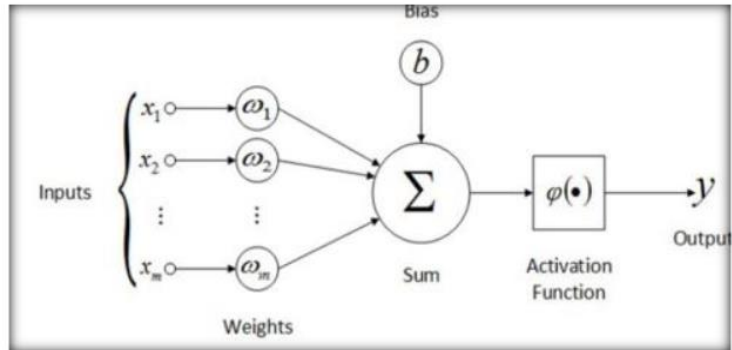
ERROR METRICS	ΕΥΗΝΟΣ	
	AR(1)	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12	
MAPE (%)	70.2	36.6
MPE (%)	-45.2	-22.6
RMSE	16.6	3.7
R ²	0.47	0.29



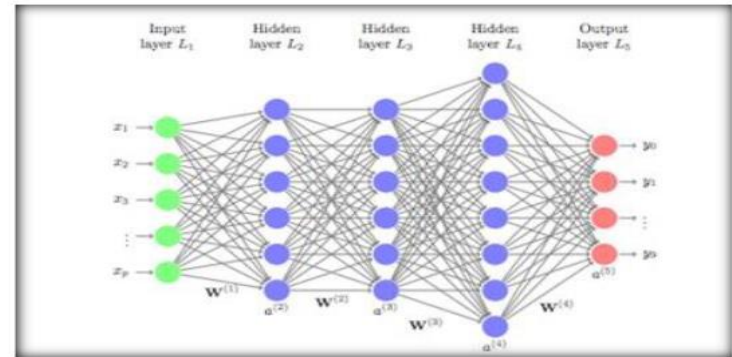
Μοντέλο LSTM (1/10)

ΘΕΩΡΗΤΙΚΑ ΣΤΟΙΧΕΙΑ

- Απλό Νευρωνικό Δίκτυο ενός επιπέδου



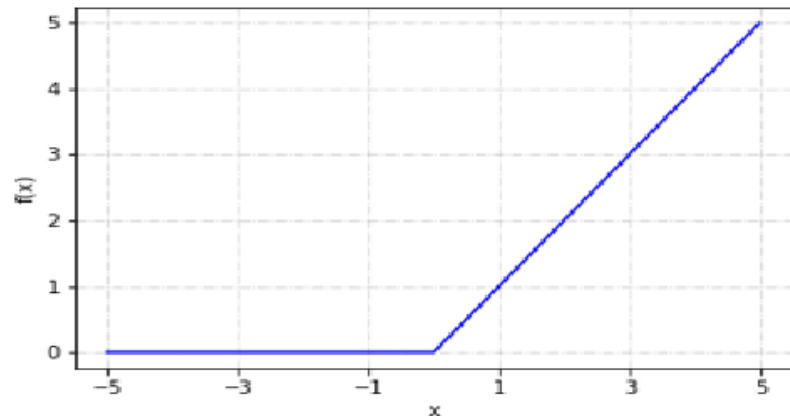
- Πολυεπίπεδο νευρωνικό δίκτυο (Multilayer Perceptron)



- Συνάρτηση Ενεργοποίησης
ReLU



$$f(x) = \max(x, 0)$$



Μοντέλο LSTM (2/10)

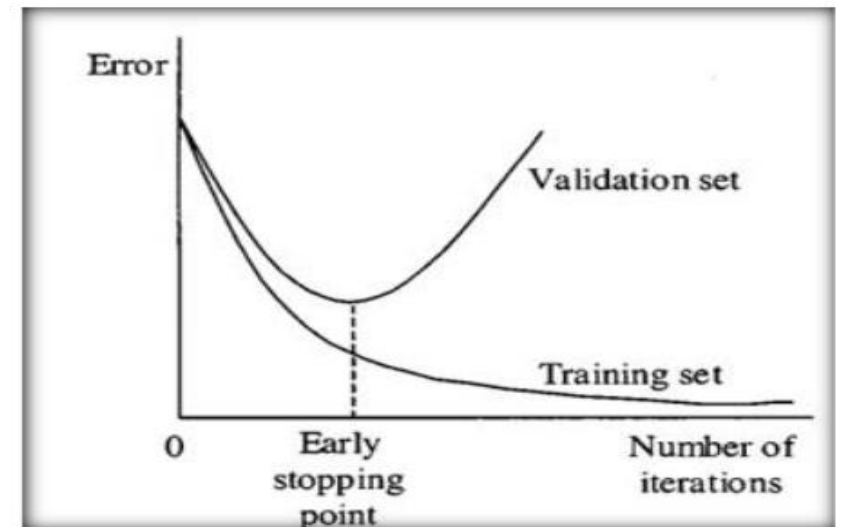
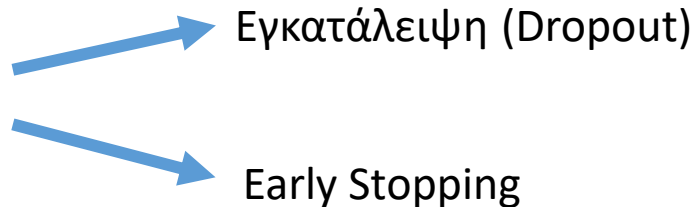
ΘΕΩΡΗΤΙΚΑ ΣΤΟΙΧΕΙΑ

- Για την μέτρηση της επίδοσής του αλγορίθμου μηχανικής μάθησης ,χρησιμοποιούμε συναρτήσεις που πρέπει να ελαχιστοποιήσουμε. Στην περίπτωση μας, για ένα πρόβλημα παλινδρόμησης πρέπει να ελαχιστοποιηθεί το μέσο τετραγωνικό σφάλμα (*loss function*).

$$MSE(\theta) = \frac{1}{N} \sum_{i=1}^N (f(x_i|\theta) - y_i)^2$$

- Χρησιμοποιείται ο αλγόριθμος βελτιστοποίησης *Adam* για την ελαχιστοποίηση της συνάρτησης κόστους.

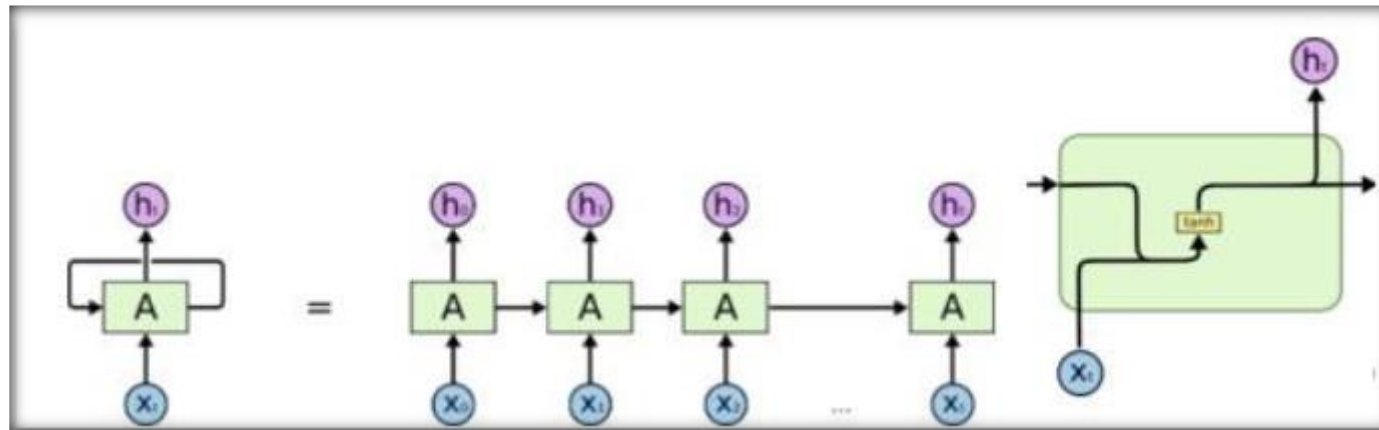
- Πρόληψη Υπερπροσαρμογής



Μοντέλο LSTM (3/10)

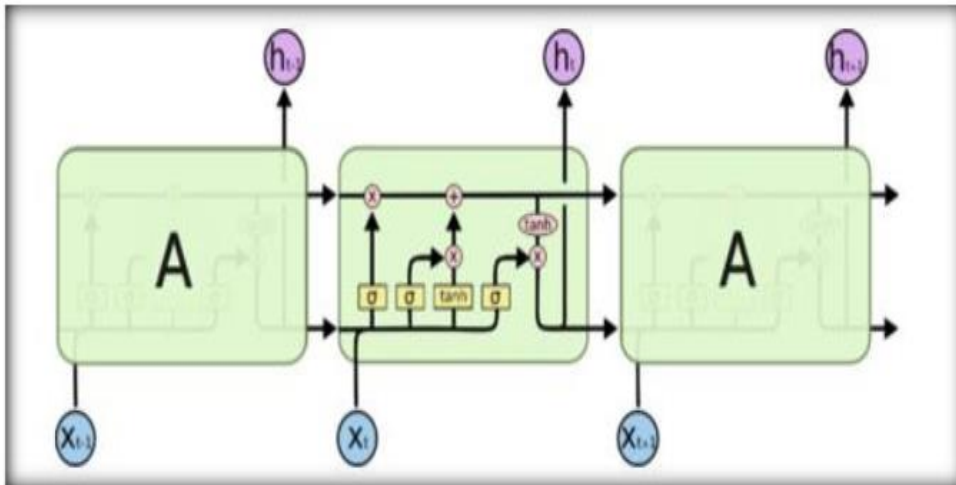
ΑΝΑΔΡΟΜΙΚΑ ΔΙΚΤΥΑ (RNN)

- Αποτελούνται από εσωτερικές συνδέσεις, οι οποίες επιτρέπουν την επεξεργασία χρονικών συσχετίσεων
- Η έξοδος των RNN σε κάθε χρονικό βήμα εξαρτάται από την προηγούμενη είσοδο και τους προηγούμενους υπολογισμούς του νευρωνικού δικτύου.
- Ενσωματώνουν παρελθοντικές πληροφορίες μαζί με πληροφορίες της κάθε εισόδου ώστε να προβλέψει μελλοντικές τιμές.
- Κατάλληλα για επεξεργασία σειριακών δεδομένων (χρονοσειρές).
- Ωστόσο, εμφανίζουν προβλήματα στην διαχείριση μεγάλου μήκους ακολουθιακών εισόδων

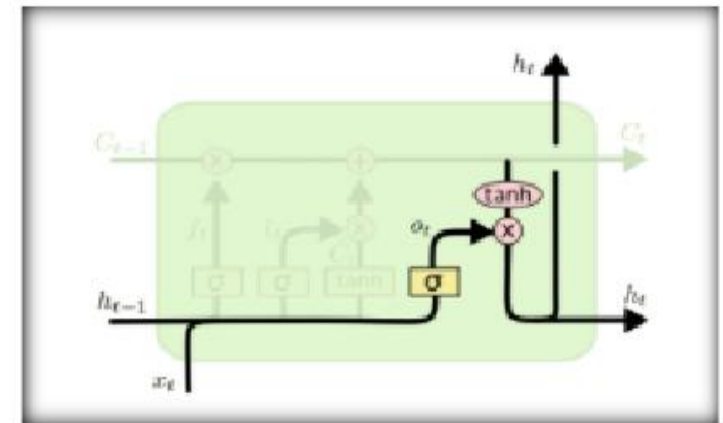
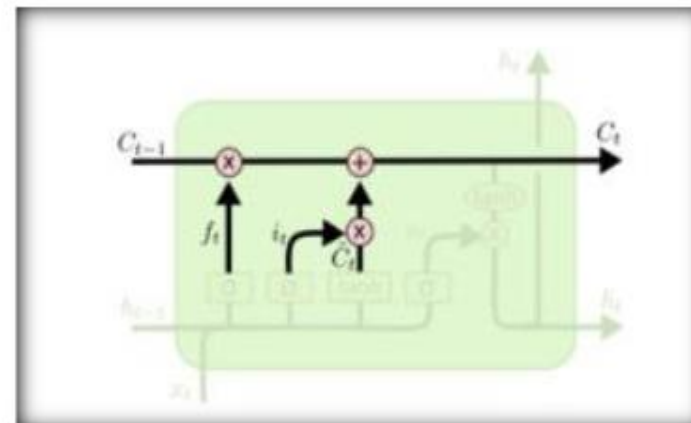
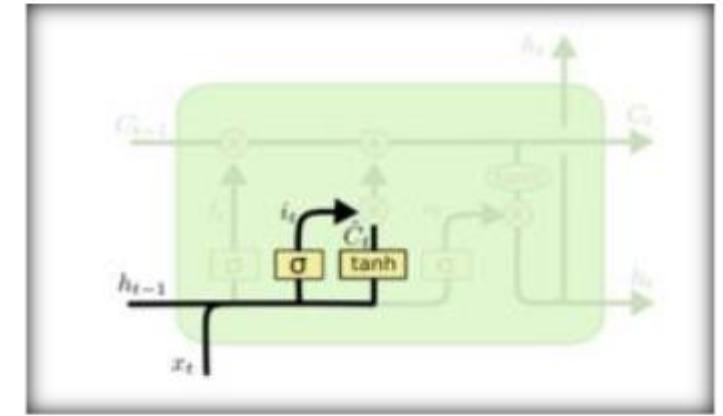
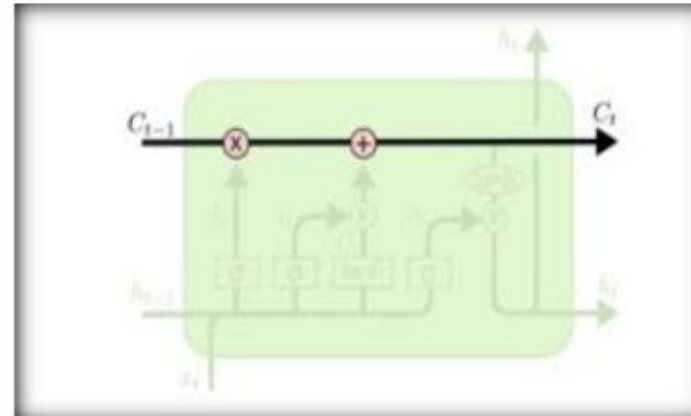


Μοντέλο LSTM (4/10)

- Για την επίλυση των παραπάνω προβλημάτων, αναπτύχθηκε το νευρωνικό δίκτυο μακράς-βραχείας μνήμης LSTM για την αποφυγή των προβλημάτων στην επεξεργασία μεγάλου μήκους εξαρτήσεων και την απόκτηση μνήμης για μεγάλες περιόδους.

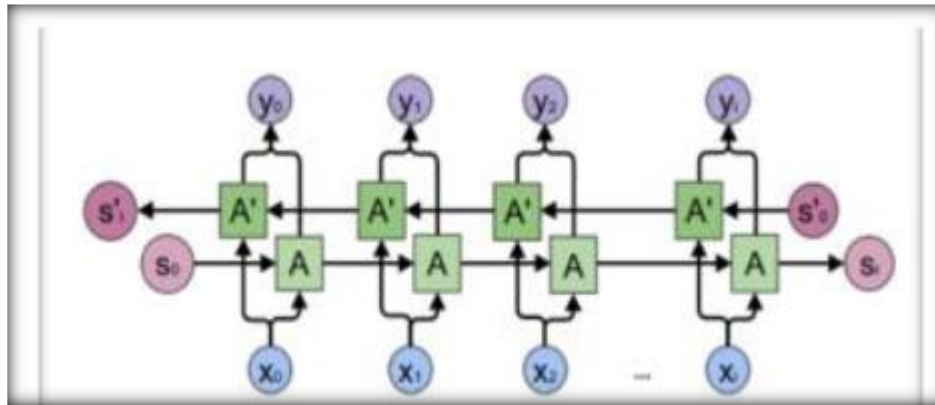


- Το κρυφό επίπεδο αποτελείται από το κύτταρο cell και τρεις πύλες (gates), οι οποίες ελέγχουν την ροή της πληροφορίας στην μονάδα του LSTM και ονομάζονται πύλη εισόδου (input gate), πύλη εξόδου (output gate), πύλη λυσμόνησης (forget gate)



Μοντέλο LSTM (5/10)

- Αντί του απλού LSTM χρησιμοποιήθηκε μια πιο εμπλουτισμένη εκδοχή του. Πρόκειται για το Νευρωνικό δίκτυο μακράς-βραχείας μνήμης διπλής κατεύθυνσης (**Bidirectional-LSTM**).
- Προκύπτει από τον διπλασιασμό του απλού LSTM. Το διάνυσμα εισόδου παρουσιάζεται μία φορά με την κανονική χρονική ακολουθία και μια με την ανάποδη. Μετά την επεξεργασία του κάθε επιπέδου, οι έξοδοι ενώνονται για κάθε χρονική στιγμή. Επιτρέπεται στο δίκτυο να επεξεργαστεί τα ακολουθιακά δεδομένα προς δυο κατευθύνσεις κάτι που βελτίωσε σημαντικά τα αποτελέσματα σε σχέση με το παραδοσιακό LSTM.

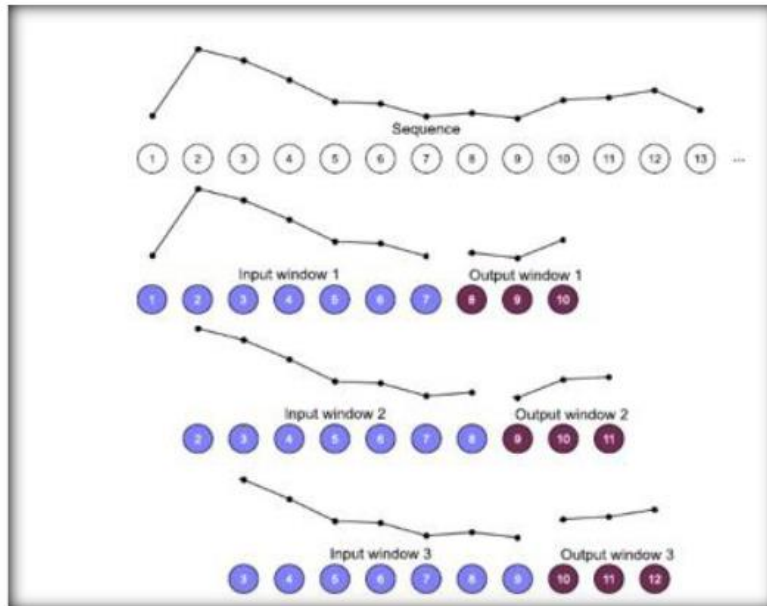


- Όλες οι αναλύσεις των σημειακών προβλέψεων για το μοντέλο LSTM πραγματοποιήθηκαν σε περιβάλλον *Python*.

Μοντέλο LSTM (6/10)

ΜΕΘΟΔΟΛΟΓΙΑ

- Η στρατηγική που χρησιμοποιήσαμε για τα μοντέλα μηχανικής μάθησης (LSTM, KNN) είναι αυτή των πολλαπλών εισόδων-πολλαπλών εξόδων (Multi- Input Multi-Output (MIMO)), όπου η έξοδος αποτελεί την συνολική πρόβλεψη όλου του ορίζοντα.
- Θα δημιουργηθεί ένα κινούμενο παράθυρο το οποίο θα ολισθαίνει στις παρατηρήσεις της χρονοσειράς, δημιουργώντας ζεύγη εισόδου-εξόδου που θα παρουσιαστούν στους αλγορίθμους μας.



```
X_train , y_train = [], []
```

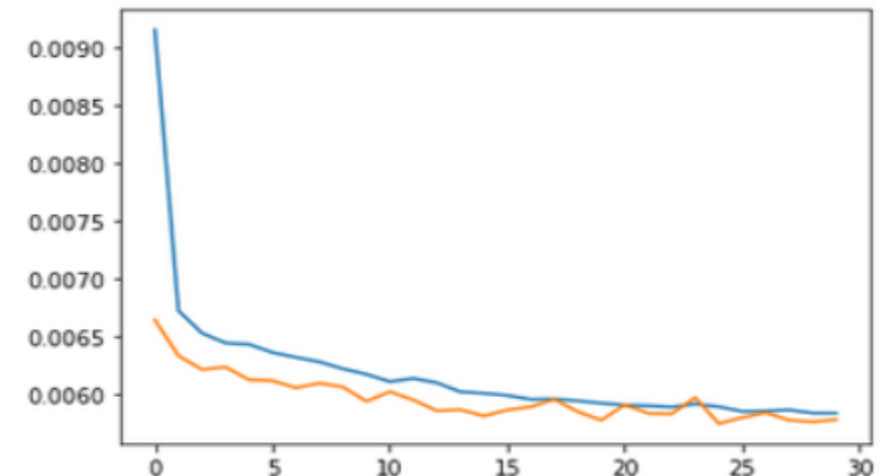
```
for i in range(24, len(data_train)-lead_time+1):  
    X_train.append(data_train[i-24:i])  
    y_train.append(data_train[i:i+lead_time])
```

Μοντέλο LSTM (7/10)

ΜΕΘΟΔΟΛΟΓΙΑ

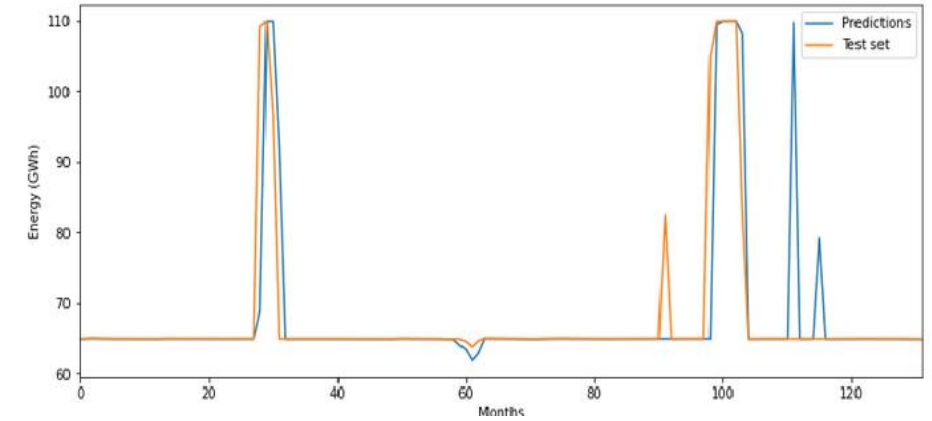
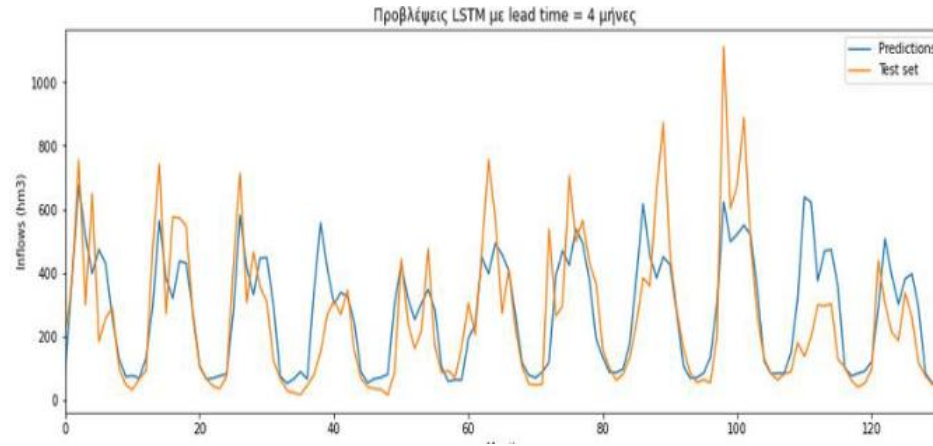
- Η εφαρμογή του LSTM γίνεται σε τέσσερις φάσεις. Στην πρώτη εκπαιδεύονται πάνω στην ίδια πληροφορία (συνθετικά δεδομένα) διάφορα μοντέλα, τα οποία βασίζονται μεν στον ίδιο αλγόριθμο, διαφέρουν δε ως προς τις υπερπαραμέτρους του συγκεκριμένου αλγορίθμου. Τα εκπαιδευμένα μοντέλα δοκιμάζονται στην συνέχεια στην πρόβλεψη ενός άλλου συνόλου δεδομένων, το οποίο καλείται σύνολο δεδομένων επικύρωσης (validation set). Από την σύγκριση αυτή προκύπτει ένα μοντέλο με το μικρότερο σφάλμα πρόβλεψης με βάση κάποιο μέτρο, συνήθως το RMSE.
- Στην τρίτη φάση ένα νέο μοντέλο με ίδιες υπερπαραμέτρους με αυτές του μοντέλου - νικητή της δεύτερης φάσης προσαρμόζει τις παραμέτρους του στο σύνολο της πληροφορίας που χρησιμοποιείται στις πρώτες δύο φάσεις, δηλαδή στο σύνολο των δεδομένων εκπαίδευσης και το σύνολο δεδομένων επικύρωσης μαζί. Το μοντέλο που προκύπτει δοκιμάζεται στην τελευταία φάση στην πληροφορία που προορίζεται για τον συγκεκριμένο σκοπό

```
BUILD LSTM MODEL  
  
In [22]: model = Sequential()  
         model.add(tf.keras.layers.Bidirectional(LSTM(units = 300,  
                                                    activation = 'relu', return_sequences=True, input_shape=(24,1))))  
         model.add(tf.keras.layers.Dropout(0.2))  
         model.add(tf.keras.layers.Bidirectional(LSTM(units = 300,  
                                                    activation = 'relu')))  
         model.add(tf.keras.layers.Dropout(0.2))  
         model.add(Dense(lead_time))  
  
In [23]: model.compile(loss='mse', optimizer='Adam')  
  
In [24]: callback = tf.keras.callbacks.EarlyStopping(monitor='val_loss', patience=5)  
         history = model.fit(X_train, y_train, epochs=50,  
                             batch_size=32, validation_split=0.2, callbacks=[callback])
```

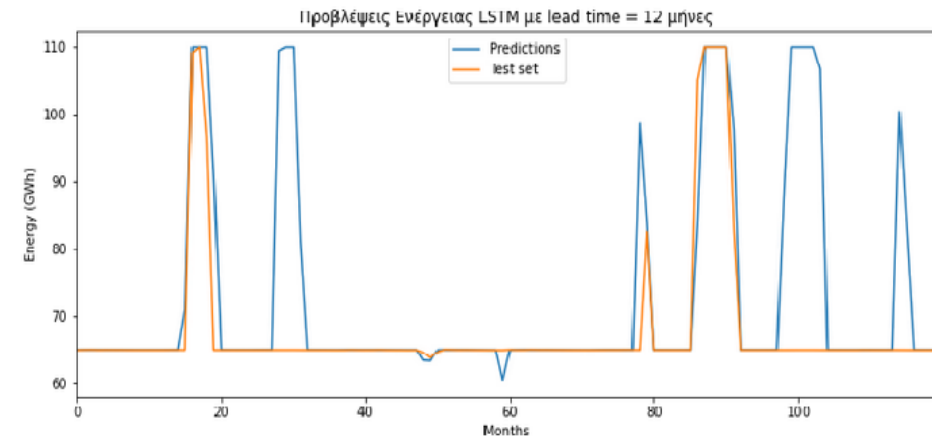
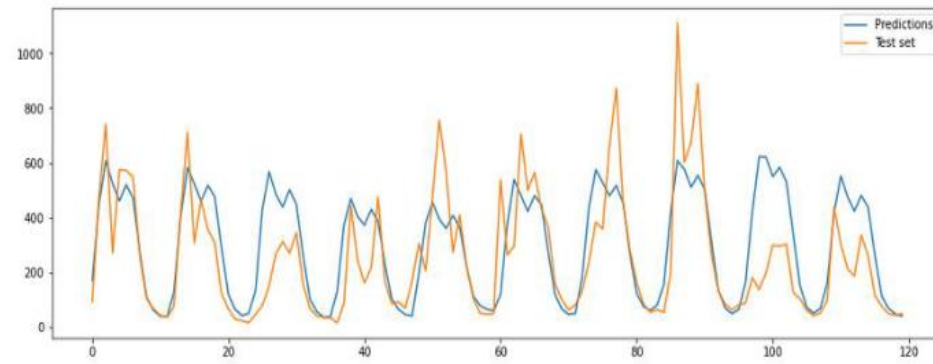


Μοντέλο LSTM (8/10)

ERROR METRICS	ΚΡΕΜΑΣΤΑ	
	LSTM	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4	
MAPE (%)	45.34	1.73
MPE (%)	-29.31	-0.6
RMSE	140.1	5.21
R ²	0.59	0.73

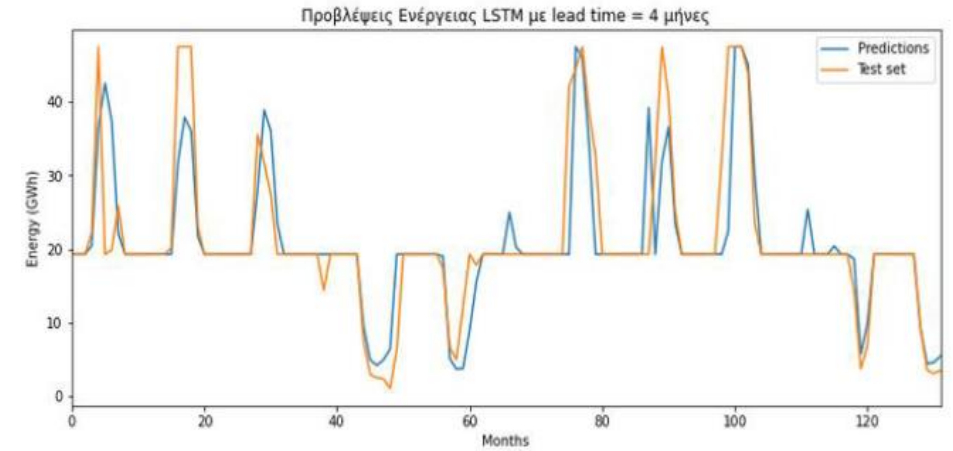
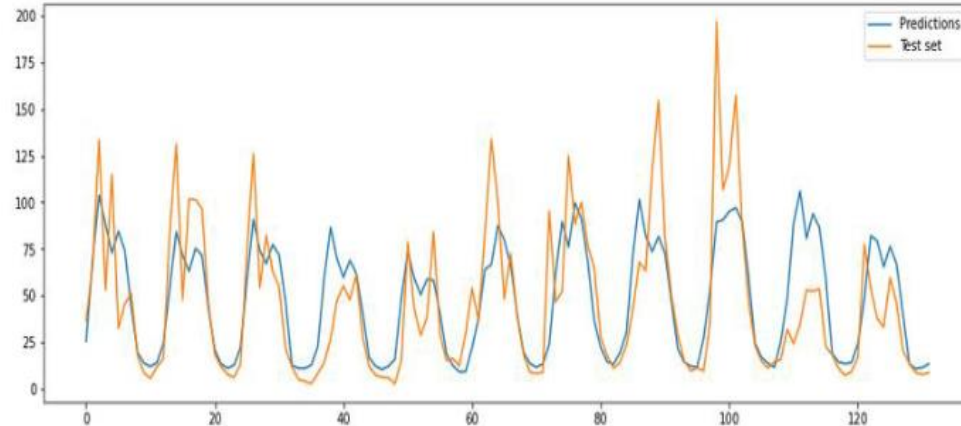


ERROR METRICS	ΚΡΕΜΑΣΤΑ	
	LSTM	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12	
MAPE (%)	65.6	7.14
MPE (%)	-49	-6.6
RMSE	164.1	13.2
R ²	0.52	0.37

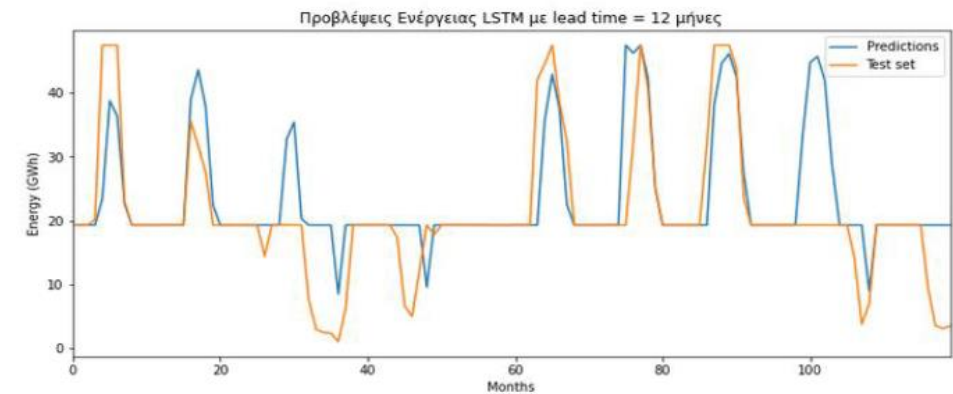
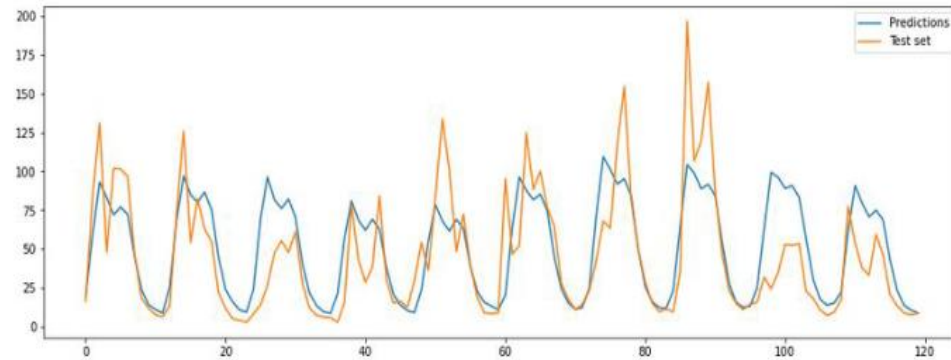


Μοντέλο LSTM (9/10)

ERROR METRICS	ΜΕΣΟΧΩΡΑ	
	LSTM	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4	
MAPE (%)	50.3	15.7
MPE (%)	-31.4	-6.8
RMSE	26.3	5.88
R ²	0.58	0.66

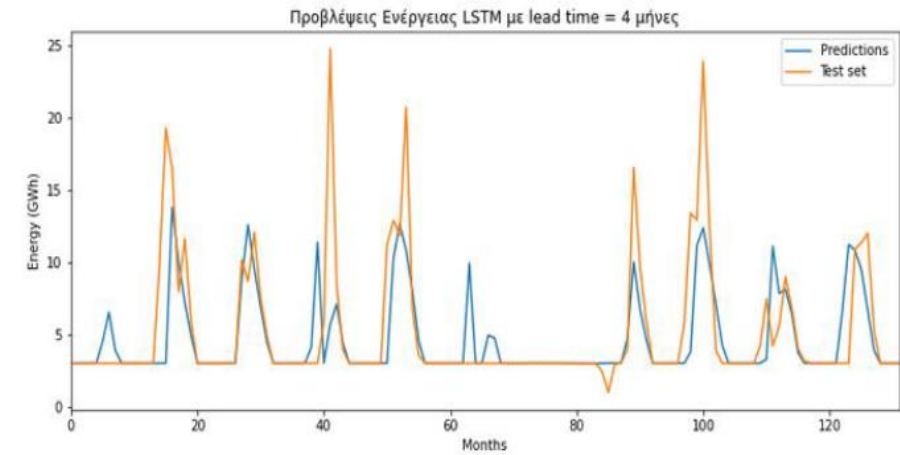
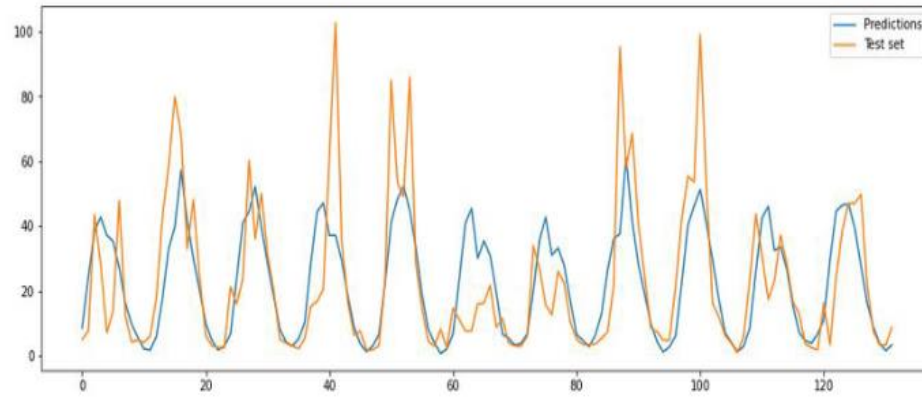


ERROR METRICS	ΜΕΣΟΧΩΡΑ	
	LSTM	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12	
MAPE (%)	64	57.1
MPE (%)	-49	-54.1
RMSE	27.24	8.4
R ²	0.54	0.43

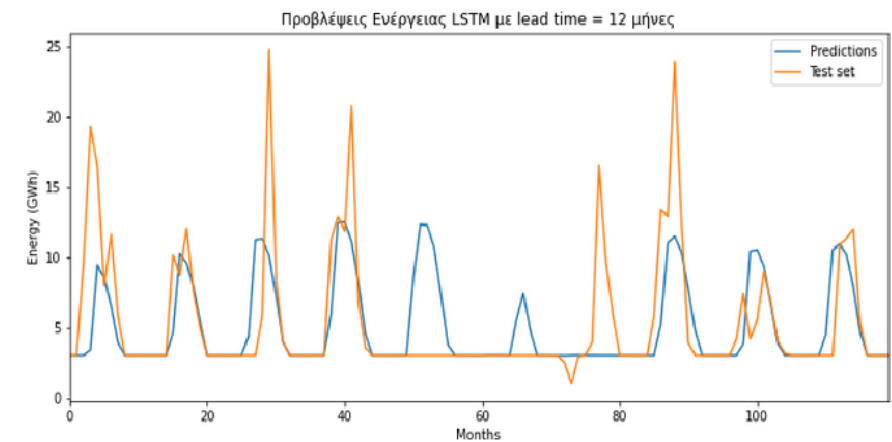
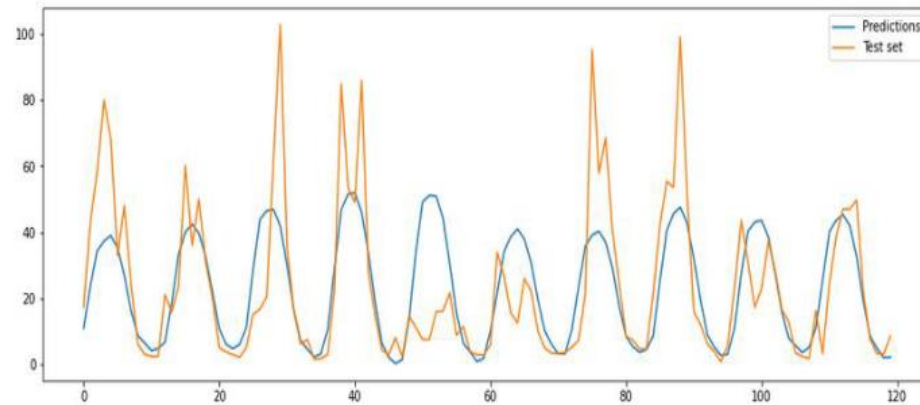


Μοντέλο LSTM (10/10)

ERROR METRICS	ΕΥΗΝΟΣ	
	LSTM	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4	
MAPE (%)	65.95	21.99
MPE (%)	-34.72	-7.02
RMSE	15.25	3.13
R ²	0.5	0.4



ERROR METRICS	ΕΥΗΝΟΣ	
	LSTM	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12	
MAPE (%)	69.08	33.7
MPE (%)	-40.4	-15.7
RMSE	16.2	3.5
R ²	0.49	0.3



Μοντέλο KNN (1/5)

ΘΕΩΡΗΤΙΚΑ ΣΤΟΙΧΕΙΑ


- Είναι ένας πολύ δημοφιλής αλγόριθμος που χρησιμοποιείται στην ταξινόμηση και στην παλινδρόμηση.
- Αποθηκεύει απλώς μια συλλογή παραδειγμάτων-δεδομένων.
- Κάθε παράδειγμα αποτελείται από ένα διάνυσμα αριθμητικών τιμών.
- Με ένα νέο παράδειγμα, το KNN βρίσκει τα K πιο όμοια παραδείγματά του (που ονομάζονται πλησιέστεροι γείτονες), σύμφωνα με μια μετρική απόστασης. Στην εργασία αυτή η μετρική είναι η Ευκλείδεια απόσταση.

$$D(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

- Η εκτίμηση λοιπόν της προβλεπόμενης τιμής δίνεται ως εξής:

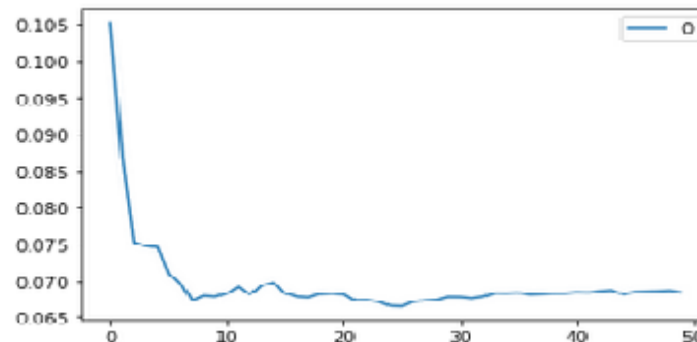
$$\hat{y} = \frac{1}{K} \sum_{xi \in N_k(x)} y_i$$

Μοντέλο KNN (2/5)

- Σημαντικό σημείο της μεθοδολογίας είναι η **χρήση συνθετικών χρονοσειρών** που διατηρούν τα στατιστικά χαρακτηριστικά των ιστορικών. Λόγω της φύσης του μοντέλου περισσότερη διαθέσιμη πληροφορία θα σημαίνει και μεγαλύτερη ακρίβεια.
- Παρόμοια προετοιμασία δεδομένων με αυτή του μοντέλου LSTM  Στρατηγική MIMO
- Σημαντική επιλογή για την αποτελεσματικότητα του μοντέλου είναι αυτή της παραμέτρου K

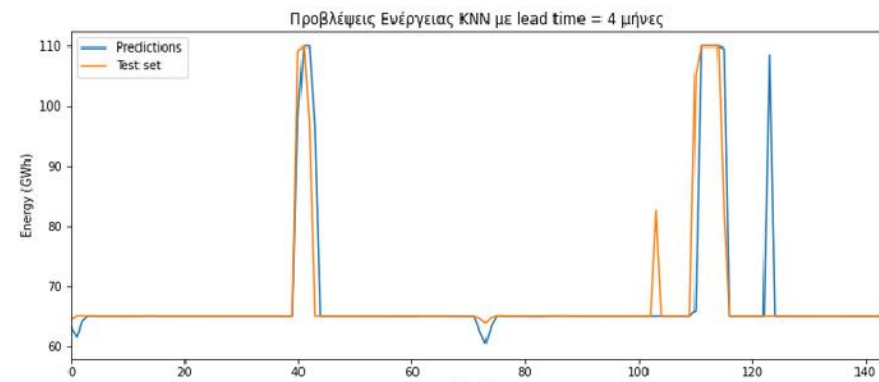
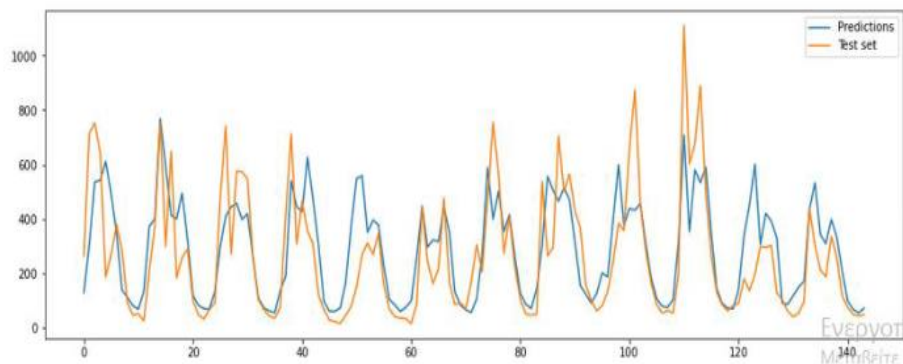

Χρήση εργαλείου GridSearchCV

- Το *GridSearchCV* προσαρμόζει επανειλημμένα τους αναδρομείς KNN σε ένα μέρος των δεδομένων και ελέγχει τις επιδόσεις στο υπόλοιπο μέρος των δεδομένων. Κάνοντας αυτό επανειλημμένα θα αποφέρει μια αξιόπιστη εκτίμηση της προγνωστικής απόδοσης καθεμιάς από τις τιμές για το K. Δοκιμάζεται στις τιμές από 1 έως 50. Στο τέλος, θα διατηρήσει την τιμή K με την καλύτερη προγνωστική απόδοση και την χαμηλότερη βαθμολογία σφάλματος.

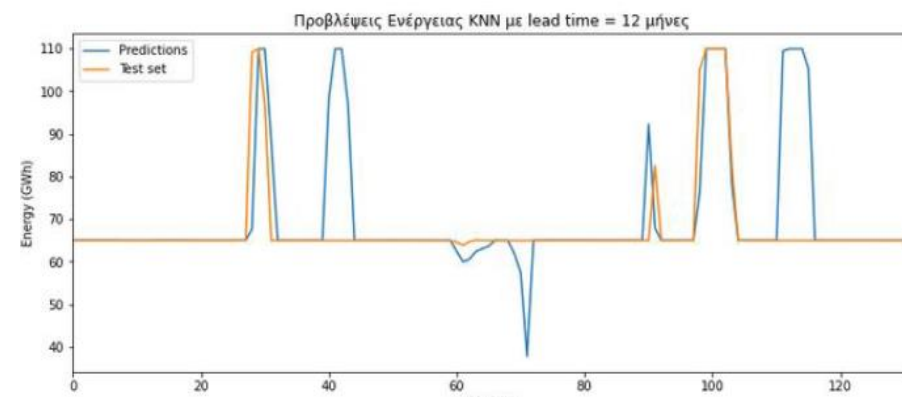
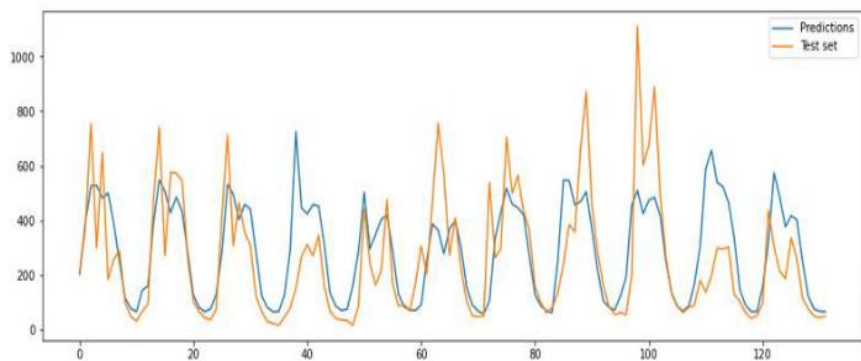


Μοντέλο KNN (3/5)

ERROR METRICS	ΚΡΕΜΑΣΤΑ	
	KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4	
MAPE (%)	67.4	1.74
MPE (%)	-51.3	-0.51
RMSE	155.9	6.32
R ²	0.54	0.67

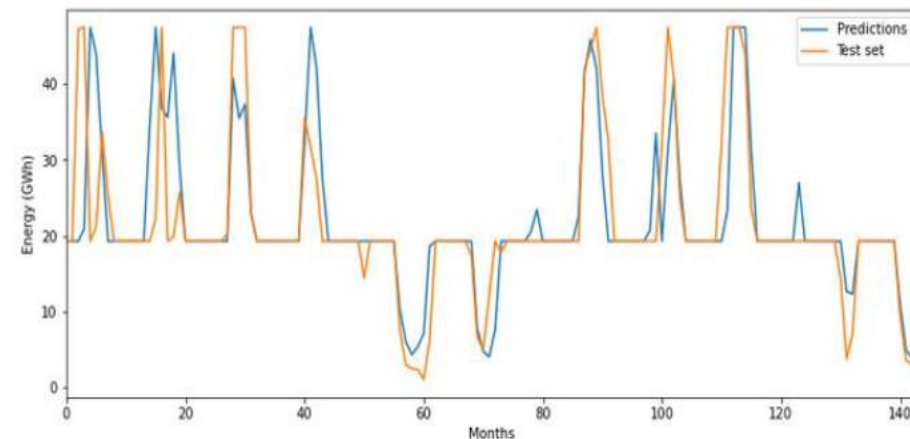
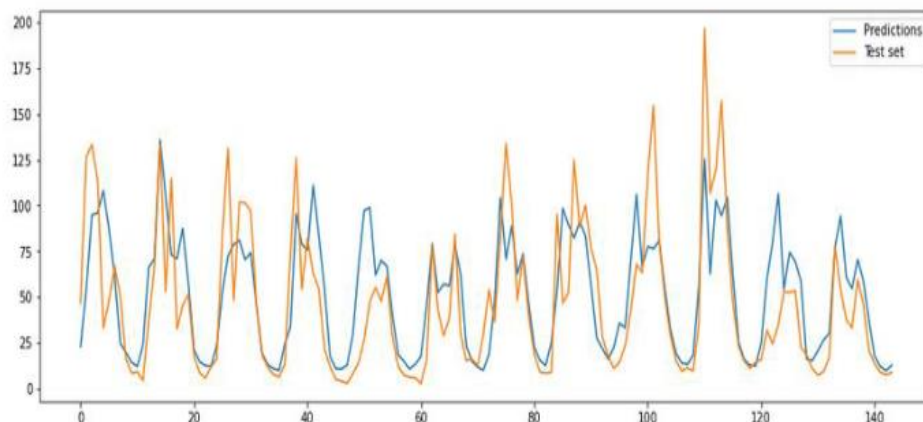


ERROR METRICS	ΚΡΕΜΑΣΤΑ	
	KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12	
MAPE (%)	72.3	6.36
MPE (%)	-55.2	-3.75
RMSE	167.8	12.57
R ²	0.46	0.29

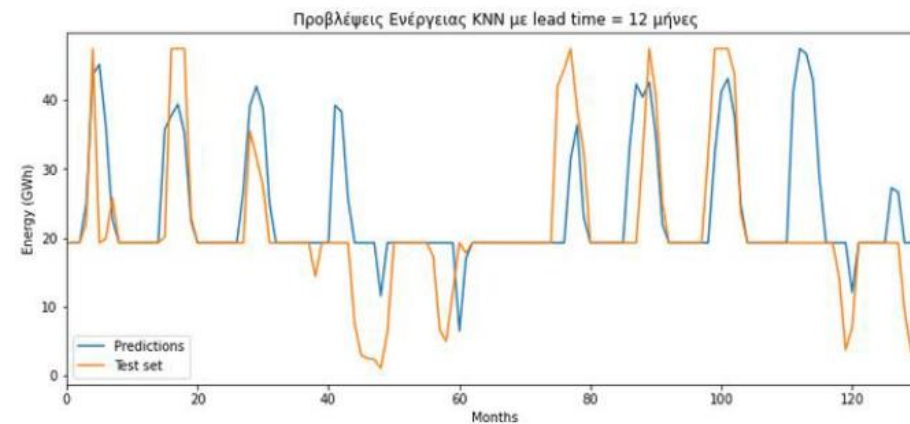
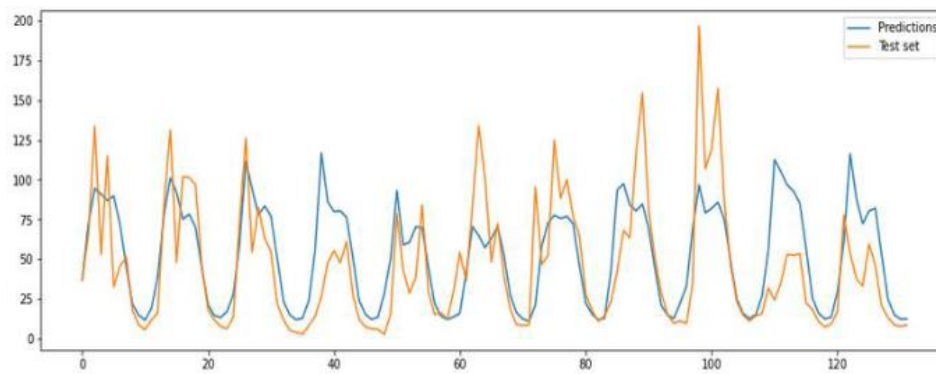


Μοντέλο KNN (4/5)

ERROR METRICS	ΜΕΣΟΧΩΡΑ	
	KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4	
MAPE (%)	67.4	22.8
MPE (%)	-51.3	-14
RMSE	27.6	7.4
R ²	0.54	0.52

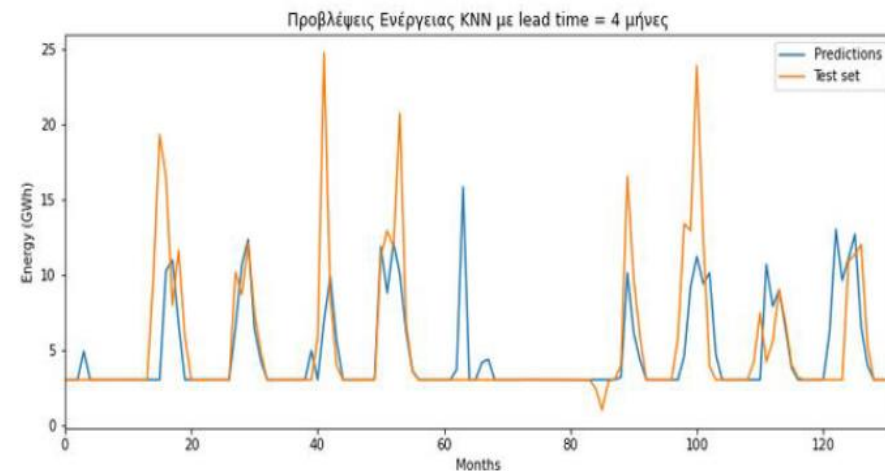
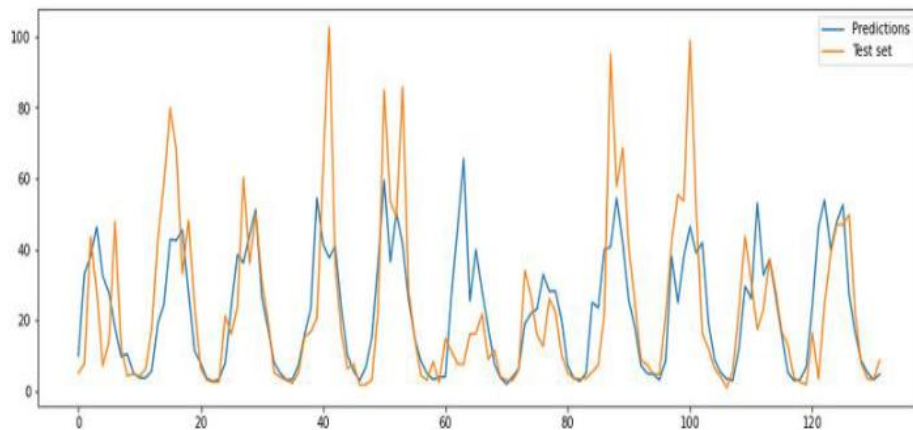


ERROR METRICS	ΜΕΣΟΧΩΡΑ	
	KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12	
MAPE (%)	74.8	59.8
MPE (%)	-59.2	-52.6
RMSE	29.7	9.3
R ²	0.47	0.32

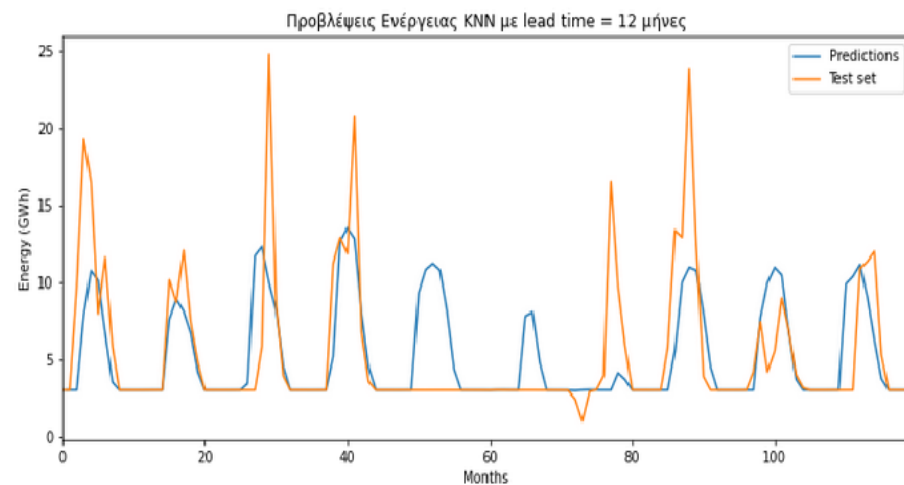
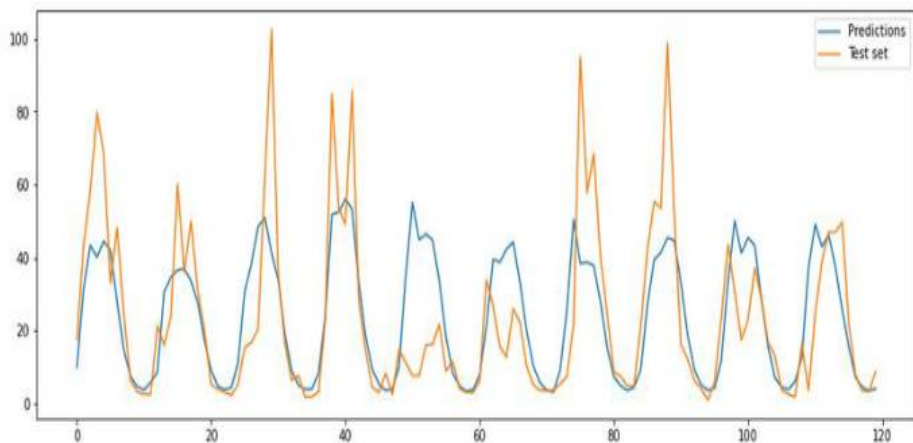


Μοντέλο KNN (5/5)

ERROR METRICS	ΕΥΗΝΟΣ	
	KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4	
MAPE (%)	78.7	24.1
MPE (%)	-48.3	-7.8
RMSE	17.4	3.6
R ²	0.43	0.36



ERROR METRICS	ΕΥΗΝΟΣ	
	KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12	
MAPE (%)	75	35.6
MPE (%)	-48	-18.2
RMSE	17.1	3.9
R ²	0.47	0.32



Συγκριτικά Αποτελέσματα και Οικονομική Αξιολόγηση (1/2)

ERROR METRICS	ΚΡΕΜΑΣΤΑ					
	AR(1)		LSTM		KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4		LEAD TIME = 4		LEAD TIME = 4	
MAPE (%)	51.69	1.86	45.34	1.73	67.4	1.74
MPE (%)	-30.34	-0.62	-29.31	-0.6	-51.3	-0.51
RMSE	149.69	5.69	140.1	5.21	155.9	6.32
R ²	0.57	0.7	0.59	0.73	0.54	0.67
ΚΕΡΔΟΣ(εκ. €/μήνα)	5.99		6.05		5.93	

ERROR METRICS	ΜΕΣΟΧΩΡΑ					
	AR(1)		LSTM		KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4		LEAD TIME = 4		LEAD TIME = 4	
MAPE (%)	51.69	16.87	50.3	15.7	67.4	22.8
MPE (%)	-30.34	-6.55	-31.4	-6.8	-51.3	-14
RMSE	26.49	7.08	26.3	5.88	27.6	7.4
R ²	0.57	0.55	0.58	0.66	0.54	0.52
ΚΕΡΔΟΣ(εκ. €/μήνα)	0.86		0.91		0.12	

ERROR METRICS	ΕΥΗΝΟΣ					
	AR(1)		LSTM		KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 4		LEAD TIME = 4		LEAD TIME = 4	
MAPE (%)	67.6	22.8	65.95	21.99	78.7	24.1
MPE (%)	-43.22	-9.7	-34.72	-7.02	-48.3	-7.8
RMSE	16.05	3.22	15.25	3.13	17.4	3.6
R ²	0.51	0.46	0.5	0.4	0.43	0.36
ΚΕΡΔΟΣ(εκ. €/μήνα)	-0.08		-0.02		-0.06	

Συγκριτικά Αποτελέσματα και Οικονομική Αξιολόγηση (2/2)

ERROR METRICS	ΚΡΕΜΑΣΤΑ					
	AR(1)		LSTM		KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12		LEAD TIME = 12		LEAD TIME = 12	
MAPE (%)	53.6	4.5	65.6	7.14	72.3	6.36
MPE (%)	-32.16	-2.78	-49	-6.6	-55.2	-3.75
RMSE	150.89	10.2	164.1	13.2	167.8	12.57
R ²	0.56	0.34	0.52	0.37	0.46	0.29
ΚΕΡΔΟΣ(εκ. €/μήνα)	4.42		2.25		3.4	

ERROR METRICS	ΜΕΣΟΧΩΡΑ					
	AR(1)		LSTM		KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12		LEAD TIME = 12		LEAD TIME = 12	
MAPE (%)	53.6	45.94	64	57.1	74.8	59.8
MPE (%)	-32.16	-37	-49	-54.1	-59.2	-52.6
RMSE	26.71	8.54	27.24	8.4	29.7	9.3
R ²	0.56	0.34	0.54	0.43	0.47	0.32
ΚΕΡΔΟΣ(εκ. €/μήνα)	-0.57		-1.36		-1.94	

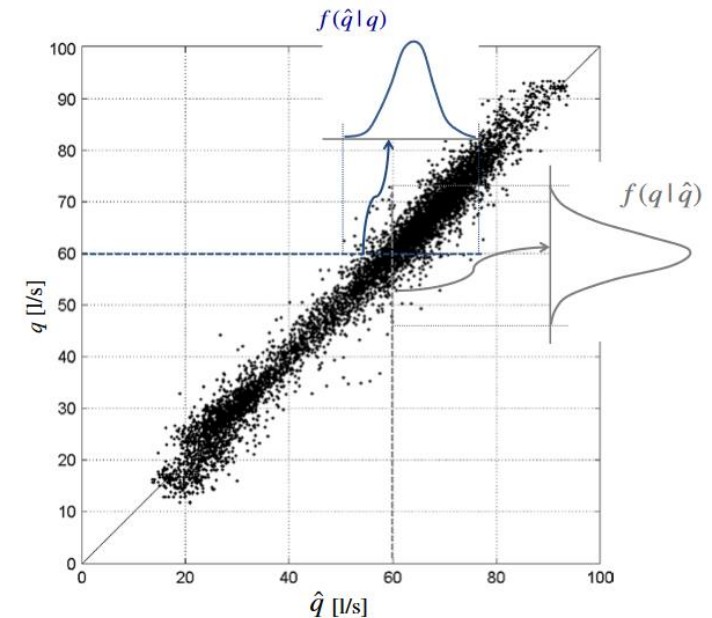
ERROR METRICS	ΕΥΗΝΟΣ					
	AR(1)		LSTM		KNN	
	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ	ΕΙΣΡΟΕΣ	ΕΝΕΡΓΕΙΑ
	LEAD TIME = 12		LEAD TIME = 12		LEAD TIME = 12	
MAPE (%)	70.2	36.6	69.08	33.7	75	35.6
MPE (%)	-45.2	-22.6	-40.4	-15.7	-48	-18.2
RMSE	16.6	3.7	16.2	3.5	17.1	3.9
R ²	0.47	0.29	0.49	0.3	0.47	0.32
ΚΕΡΔΟΣ(εκ. €/μήνα)	-0.49		-0.31		-0.4	

Ποσοτικοποίηση της αβεβαιότητας με τη χρήση συζεύξεων (copulas) (1/3)

- Οι συζεύξεις μπορούν να θεωρηθούν ως πολυδιάστατες συναρτήσεις κατανομών, των οποίων οι περιθώριες είναι ομοιόμορφες στο διάστημα $[0,1]$.
- Μπορούν να είναι χρήσιμα εργαλεία για την ποσοτικοποίηση της προγνωστικής αβεβαιότητας ενός μοντέλου, αφού επιτρέπουν την παραγωγή δεσμευμένων κατανομών.
- Η δεσμευμένη κατανομή $C_{O|M}$ μπορεί να αντιστραφεί και να λυθεί, για δεδομένη πιθανότητα (μη υπέρβασης) \underline{a} (π.χ., a μπορεί εκφράζει το επίπεδο αβεβαιότητας 5, 50 ή 95%).
- Γίνεται χρήση της Gaussian copula και η προς λύση εξίσωση τελικά λαμβάνει τη μορφή:

$$y_O^{a|F_{Y_M}(y_M)} = F_{Y_O}^{-1} \left(\Phi \left(\theta \Phi^{-1} \left(F_{Y_M}(u_M) \right) + \sqrt{(1 - \theta^2)} \Phi^{-1}(a) \right) \right)$$

όπου το Φ συμβολίζει την μονομεταβλητή Gaussian CDF και θ είναι η παράμετρος του Gaussian copula (και συνδέεται με των συντελεστή συσχέτισης των παρατηρημένων και μοντελοποιημένων δεδομένων).



Ποσοτικοποίηση της αβεβαιότητας με τη χρήση συζεύξεων (copulas) (2/3)

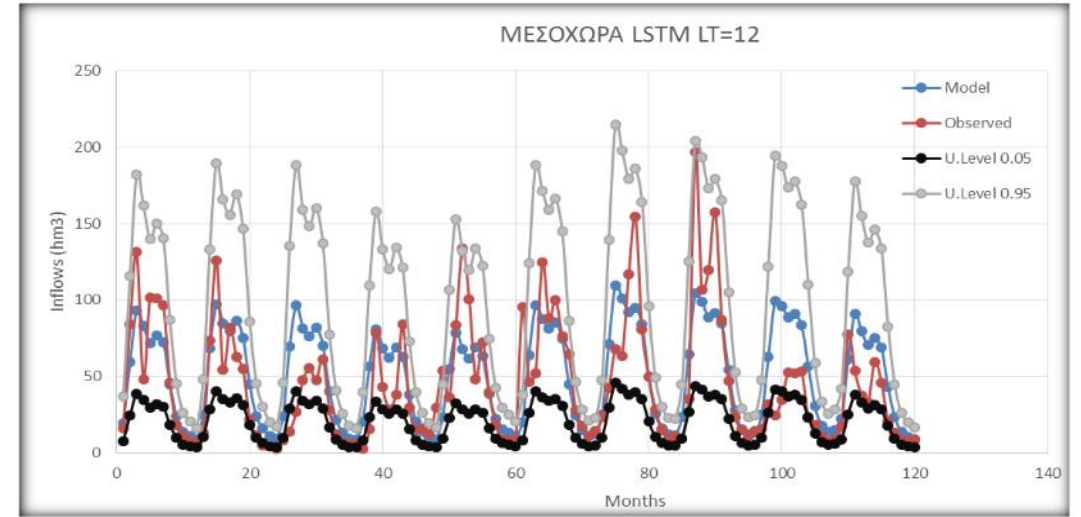
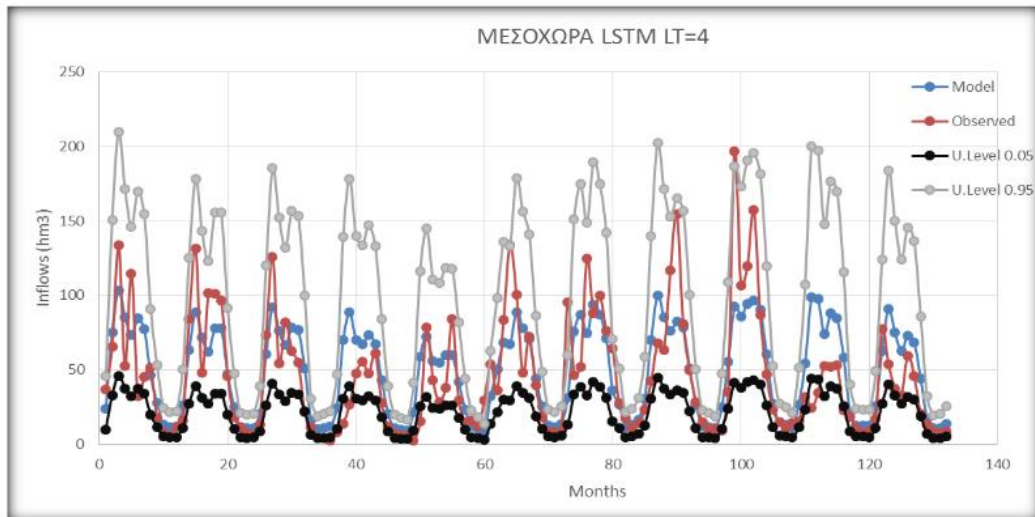
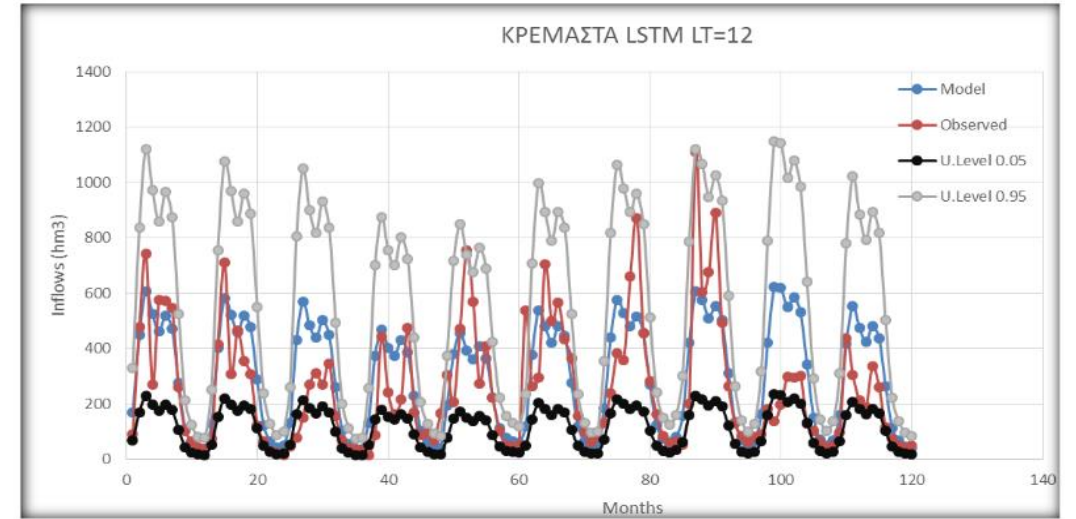
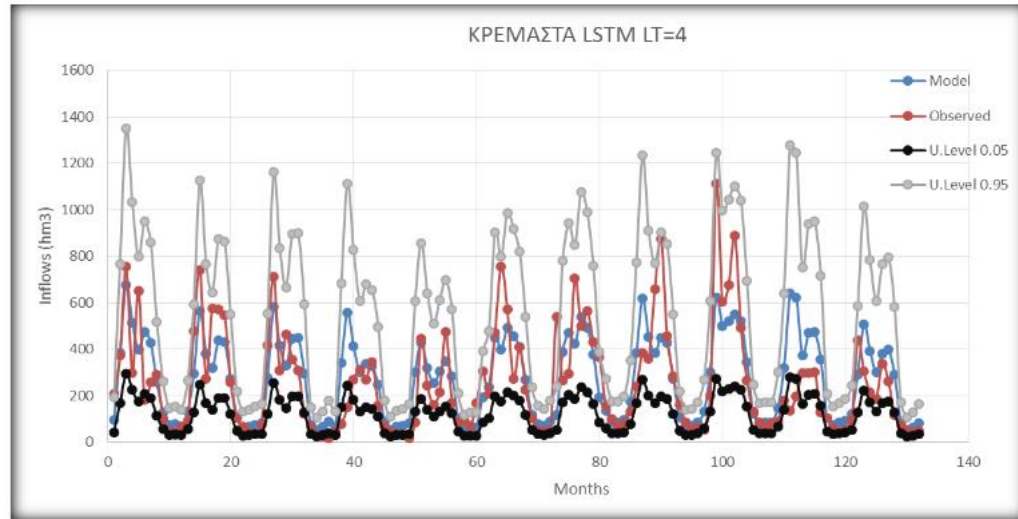
- Γίνεται η θεώρηση ότι τα παρατηρημένα και τα μοντελοποιημένα δεδομένα ακολουθούν λογαριθμοκανονική κατανομή. Στην περίπτωση αυτή, η παράμετρος θ εκτιμάται μέσω της χρήσης αναλυτικής σχέσης.

$$\theta = \frac{\ln(1+\rho\sqrt{(\exp(b_1^2)-1)(\exp(b_2^2)-1)})}{b_1 b_2}$$

- Σε αντίθετη περίπτωση θα χρειάζονταν αρκετά πιο σύνθετοι μαθηματικοί υπολογισμοί (μιας και στη γενική περίπτωση απαιτείται η επίλυση ενός διπλού ολοκληρώματος) που ξεφεύγουν από το σκοπό της παρουσίασης της συγκεκριμένης μεθοδολογίας.
- Χρησιμοποιήθηκαν τα αποτελέσματα του LSTM μοντέλου, όπως αυτά προκύψανε από τις σημειακές προγνώσεις.
- Για την εκτίμηση των παραμέτρων της λογαριθμοκανονικής κατανομής Χρησιμοποιήθηκαν οι εξής σχέσεις:

$$a := \mu_{\text{LogN}} = \ln\left(\frac{\mu}{\sqrt{1+\frac{\sigma}{\mu^2}}}\right) \qquad b^2 := \sigma_{\text{LogN}}^2 = \ln\left(1+\frac{\sigma^2}{\mu^2}\right)$$

Ποσοτικοποίηση της αβεβαιότητας με τη χρήση συζεύξεων (copulas) (3/3)



Συμπεράσματα και Προτάσεις (1/3)

- Το LSTM μοντέλο έχει την καλύτερη εικόνα σε όλα τα μέτρα επίδοσης στις αναλύσεις με χρονικό ορίζοντα πρόβλεψης τους 4 μήνες και για τους 3 ταμειευτήρες. Όσον αφορά στους 12 μήνες, εκεί το LSTM υπερέχει μόνο στην περίπτωση του ταμειευτήρα στον Εύηνο.
- Σημαντικό εξαγόμενο αποτελεί το γεγονός πως το συγκεκριμένο μοντέλο μπορεί να διακρίνει και να εντοπίσει την μεταβλητότητα του επιπέδου τιμών σε αρκετά πιο ικανοποιητικό βαθμό σε σχέση με τα άλλα 2 μοντέλα.
- Στις μέσες και χαμηλές ροές η επίδοση του είναι χειρότερη σε σχέση με αυτή του απλού AR(1).
- Γεννάται ιδέα δημιουργίας ενός υβριδικού μοντέλου που θα εκμεταλλεύεται τα δυνατά στοιχεία των 2 επιμέρους μοντέλων. Το μοντέλο αυτό αποτελεί πρόταση της εργασίας για περαιτέρω έρευνα.
- Η επίδοση του AR(1) ήταν μια ευχάριστη έκπληξη. Στην αρχή επιλέχτηκε με το ρόλο του baseline μοντέλου ως φειδωλό και απλό. Στην πορεία όμως διαδραμάτισε σημαντικό ρόλο στις αναλύσεις. Ειδικά για το φράγμα των Κρεμαστών, αν το ζητούμενο είναι αποκλειστικά η ενεργειακή παραγωγή, τα σφάλματα εξομαλύνονται σε τέτοιο βαθμό που δεν έχει νόημα η εφαρμογή ενός πιο σύνθετου μοντέλου όπως του LSTM, ειδικά αν αυτό έχει τόσες απαιτήσεις σε υπολογιστικό φόρτο.

Συμπεράσματα και Προτάσεις (2/3)

- Ο αλγόριθμος KNN είχε χειρότερη επίδοση από αυτή που αναμέναμε. Μπορεί τα συνθετικά δεδομένα που χρησιμοποιήθηκαν να βοήθησαν (λόγω της φύσης του μοντέλου) ωστόσο δεν στάθηκαν αρκετά σε σύγκριση πάντα με τα άλλα 2 μοντέλα.
- Στα θετικά του, η ικανότητα του να αντιλαμβάνεται σε ικανοποιητικό βαθμό την μεταβλητότητα των τιμών.
- Με τα εξαγόμενα αποτελέσματα της μελέτης μας, ο διαχειριστής του έργου θα έχει μία πληροφόρα-εκτίμηση για τους επόμενους 4 μήνες (ή παραπάνω). Θα είχε μεγάλο ενδιαφέρον το κατά πόσο και με ποιον τρόπο θα μπορούσε να διαχειριστεί τα τυχόντα ελλείμματα (ή ακόμα και πλεονάσματα) που θα προέκυπταν με βέλτιστο τρόπο. Θα μας ήταν ιδιαίτερα ευχάριστο να γίνει μία σχετική έρευνα στο μέλλον.
- Μόνο στους ταμιευτήρες Μεσοχώρας και Κρεμαστών για $LT=4$ και Κρεμαστών για $LT=12$, ο δείκτης οικονομικής αξιολόγησης λαμβάνει θετικές τιμές. Ίσως αυτό μας οδηγεί στο συμπέρασμα πως για υδροηλεκτρικούς ταμιευτήρες χωρίς μεγάλη αποθηκευτική ικανότητα δεν έχει νόημα η μακροπρόθεσμη πρόγνωση παρά μόνο βραχυπρόθεσμη και μεσοπρόθεσμη.
- Από την οικονομική αξιολόγηση προκύπτει επίσης πως και πάλι το LSTM έχει τους υψηλότερους δείκτες κέρδους για όλες τις αναλύσεις με $LT=4$ και για αυτή του Ευήνου με $LT=12$. Παρατηρώντας τους ίδιους δείκτες είναι εμφανές πως λανθασμένη επιλογή μοντέλου πρόγνωσης μπορεί να στοιχίσει σε ετήσια κλίμακα εκατομμύρια ευρώ

Συμπεράσματα και Προτάσεις (3/3)

- Σε ένα αντικείμενο μελέτης που η κρίση και οι αποφάσεις του διαχειριστή του έργου είναι τόσο καθοριστικές δε νοείται να μη ληφθεί υπόψη η αβεβαιότητα. Οι συζεύξεις, ως πολυδιάστατες συναρτήσεις κατανομών, είναι πολύ χρήσιμα εργαλεία και η μεθοδολογία που παρουσιάστηκε θα μπορούσε να βοηθήσει σημαντικά στη λήψη αποφάσεων. Το επίπεδο αβεβαιότητας ορίζεται από τον διαχειριστή και συνεπώς μπορεί να έχει στα χέρια του όλα τα πιθανά σενάρια. Αντικείμενο μελλοντικής έρευνας θα μπορούσε να είναι ο συνδυασμός δύο ή και περισσότερων ντετερμινιστικών μοντέλων που μέσω των συζεύξεων θα δημιουργούσαν μία ακόμα πιο αξιόπιστη πιθανοτική μέθοδο πρόγνωσης.
- Είναι πλέον θέμα του διαχειριστή το τι επίπεδο αβεβαιότητας θα επιλέξει, το πόσο θέλει να ρισκάρει. Είναι προετοιμασμένος για όλα τα σενάρια, αισιόδοξα ή απαισιόδοξα. Πεδίο έρευνας θα μπορούσε να αποτελέσει η διερεύνηση του ακριβούς επιπέδου αβεβαιότητας που θα είναι πιο κερδοφόρο για κάθε έργο.

ΕΥΧΑΡΙΣΤΩ ΓΙΑ ΤΗΝ ΠΡΟΣΟΧΗ ΣΑΣ !