

Η ΧΡΗΣΗ ΔΙΕΡΕΥΝΗΤΙΚΗΣ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΟΝ ΠΡΟΣΔΙΟΡΙΣΜΟ ΤΗΣ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ ΔΙΕΡΓΑΣΙΩΝ

X. ΜΗΤΣΑΣ, Ζ. ΜΕΤΑΞΙΩΤΟΥ

**ΕΛΛΗΝΙΚΟ ΙΝΣΤΙΤΟΥΤΟ ΜΕΤΡΟΛΟΓΙΑΣ (ΕΙΜ), Δ/ΝΣΗ ΜΗΧΑΝΙΚΩΝ ΜΕΓΕΘΩΝ,
ΘΕΣ/ΝΙΚΗ 57022**

Email: chris.mitsas@eim.org.gr

Περίληψη: Στην εργασία αυτή παρουσιάζεται η συνεχιζόμενη προσπάθεια του εργαστηρίου Μάζας του ΕΙΜ να διασφαλίσει την αξιοπιστία των αποτελεσμάτων που παράγονται σε αυτό. Για την επίτευξη του σκοπού αυτού εφαρμόζεται η μεθοδολογία της διερευνητικής ανάλυσης δεδομένων (Exploratory Data Analysis, EDA) σε σειρές μετρήσεων μάζας στον συγκριτή του εργαστηρίου των 10 kg με βασικό σκοπό την ταυτοποίηση πιθανής απόκλισης από τις προϋποθέσεις της κατάστασης στατιστικού ελέγχου. Οι τεχνικές της EDA στηρίζονται στην γραφική απεικόνιση των σχέσεων μεταξύ των αποτελεσμάτων και αντιμετωπίζουν τόσο τα «κλασσικά» ερωτήματα του προσδιορισμού των βέλτιστων εκτιμητών και της αβεβαιότητάς των και του ελέγχου στατιστικών υποθέσεων όσο και της επιλογής και επικύρωσης του καταλληλότερου μοντέλου, και ανίχνευσης συσχετίσεων και πιθανών “outliers”.

1. Εισαγωγή

Η θέση και η διατήρηση μιας διεργασίας σε κατάσταση προβλεψιμότητας αποτελεί έναν από τους σημαντικότερους στόχους στην επιστήμη και την τεχνολογία και αποτελεί ένα βασικότατο κριτήριο της αξιοπιστίας μιας μετρητικής διεργασίας. Η εξαγωγή τεκμηριωμένων συμπερασμάτων σχετικά με την προβλεψιμότητα μιας μετρητικής διεργασίας θα ήταν εξαιρετικά δύσκολη εάν δεν βασιζόνταν στο γεγονός ότι ανεξάρτητα από την πολυπλοκότητα της κάθε διεργασίας, αυτή πάντα διέπεται από τέσσερις βασικές παραδοχές σύμφωνα με τις οποίες τα μετρητικά δεδομένα έχουν τα εξής χαρακτηριστικά:

- Είναι τυχαία
- Προέρχονται από συγκεκριμένη στατιστική κατανομή
- Η κατανομή αυτή χαρακτηρίζεται από ένα σταθερό μέτρο θέσης
- Η κατανομή χαρακτηρίζεται από σταθερή διασπορά.

Εάν και οι τέσσερις παραπάνω παραδοχές ισχύουν τότε η διεργασία βρίσκεται σε κατάσταση στατιστικού ελέγχου γεγονός που επιτρέπει την περιγραφή της συμπεριφοράς της βάσει πιθανοτήτων όχι μόνο στο παρελθόν αλλά κυρίως στο μέλλον [1]. Σε αντίθετη περίπτωση η διεργασία είναι μεταβαλλόμενη και μη. Προκειμένου να εξαχθούν αξιόπιστα συμπεράσματα σχετικά με τον χαρακτήρα και την προβλεψιμότητα μιας μετρητικής διεργασίας εφαρμόζονται διαφορετικές τεχνικές ανάλυσης δεδομένων που έχουν τον ίδιο τελικό στόχο, αλλά διαφέρουν σημαντικά ως προς τη μέθοδο επεξεργασίας των δεδομένων προκειμένου να οδηγηθούν σε ασφαλή συμπεράσματα.

Στην κλασσική προσέγγιση [2] τα δεδομένα αναλύονται υιοθετώντας “a-priori” ένα συγκεκριμένο μοντέλο οι παράμετροι του οποίου εκτιμώνται και ελέγχονται κάτω από την παραδοχή της ισχύος του παραπάνω μοντέλου. Το μειονέκτημα αυτής της προσέγγισης έγκειται στο γεγονός ότι η εγκυρότητα των συμπερασμάτων εξαρτάται από την ισχύ της παραδοχής του αρχικού μοντέλου. Επιπλέον, η κλασσική ανάλυση επικεντρώνεται στον χαρακτηρισμό της διεργασίας και των δεδομένων μέσα από συγκεκριμένους στατιστικούς εκτιμητές οι οποίοι περιγράφουν μεν κάποια σημαντικά χαρακτηριστικά της κατανομής (θέση, διασπορά) αλλά ταυτόχρονα αδυνατούν να

προσεγγίσουν άλλα εξίσου σημαντικά χαρακτηριστικά (λοξότητα, αυτοσυσχέτιση κλπ). Στην ανάλυση κατά Bayes [3] τόσο τα δεδομένα όσο και οποιαδήποτε προηγούμενη γνώση σχετικά με τη διεργασία συνδιάζονται προκειμένου να εξαχθούν συμπεράσματα ή να γίνουν έλεγχοι υποθέσεων σχετικά με τις παραμέτρους του ισχύοντος μοντέλου. Σε αντίθεση με τις παραπάνω τεχνικές, η διερευνητική ανάλυση δεδομένων (Exploratory Data Analysis) [4,5] δεν υιοθετεί εκ των προτέρων κάποιο μοντέλο. Αντιθέτως, χρησιμοποιώντας τις εξαιρετικές δυνατότητες της γραφικής απεικόνισης διερευνά τα δεδομένα με στόχο της αποκάλυψη της δομής που τα διέπει αφήνοντας τα ίδια να υποδείξουν το καταλληλότερο μοντέλο για τη συγκεκριμένη διεργασία. Επιπλέον, η τεχνική EDA χρησιμοποιεί τα δεδομένα στο σύνολό τους με αποτέλεσμα να μην υπάρχει απώλεια πληροφορίας. Με τον τρόπο αυτό η διερευνητική ανάλυση δεδομένων παρέχει τα μέσα προκειμένου ο αναλυτής να οδηγηθεί σε ασφαλή συμπεράσματα σχετικά με το μοντέλο που διέπει τη μετρητική διεργασία κάνοντας τις ελάχιστες δυνατές παραδοχές. Η διερευνητική ανάλυση δεδομένων αποτελεί ένα υποσύνολο των σύγχρονων τεχνικών ανάλυσης μεγάλων βάσεων δεδομένων με στόχο την αποκάλυψη ιδιαίτερων και χρήσιμων, αλλά συχνά κρυφών, δομών που τα διέπουν και που αποκαλούνται τεχνικές Data Mining (DM) [6].

2. Μεθοδολογία Ανάλυσης

Ένα κρίσιμο βήμα στην διερευνητική ανάλυση δεδομένων είναι να τεθούν, σε πρώιμο στάδιο της ανάλυσης, τα βασικά ερωτήματα τα οποία χρήζουν απάντησης στη συγκεκριμένη διεργασία και μάλιστα αυτά να ταξινομηθούν κατά σειρά φθίνουσας σημασίας. Για την παρούσα εργασία τα ερωτήματα αυτά είναι κατά σειρά προτεραιότητας τα εξής:

1. Βρίσκεται η μετρητική διεργασία σε κατάσταση στατιστικού ελέγχου;
2. Προκύπτει κάποια μη προφανής δομή από την ανάλυση των δεδομένων με τις τεχνικές EDA;
3. Υπάρχουν δεδομένα τα οποία θα μπορούσαν να χαρακτηριστούν “outliers”;
4. Ποιό είναι το μοντέλο το οποίο προσεγγίζει καλύτερα τη συγκεκριμένη διεργασία;
5. Ποιοί είναι οι βέλτιστοι εκτιμητές με βάση το παραπάνω προτεινόμενο μοντέλο;
6. Ποια είναι η σχέση των αποτελεσμάτων της ανάλυσης EDA με τα αντίστοιχα αποτελέσματα που προκύπτουν από την εφαρμογή των τεχνικών SPC στα ίδια δεδομένα όσον αφορά τη θέση και τη μεταβλητότητα της διεργασίας;

Το βασικό ερώτημα από την απάντηση του οποίου εν μέρη απορρέουν και οι απαντήσεις στα υπόλοιπα ερωτήματα είναι το πρώτο και η απάντησή του στηρίζεται στον έλεγχο της ισχύος των τεσσάρων βασικών παραδοχών μέσα από απλά αλλά ισχυρά και αποτελεσματικά εργαλεία των τεχνικών EDA. Συγκεκριμένα, σε πρώτο στάδιο χρησιμοποιούνται τέσσερις διαφορετικές τεχνικές γραφικής απεικόνισης δεδομένων κάθε μία από τις οποίες ελέγχει μία από τις βασικές τέσσερις υποθέσεις [6]. Οι τεχνικές αυτές είναι:

1. **Σειριακό διάγραμμα δεδομένων** (Run sequence plot) – Ανιχνεύει μετατοπίσεις στη θέση και στη διασπορά μιας κατανομής δεδομένων ενώ ταυτόχρονα μπορεί και υποδεικνύει πιθανά “outliers”. Για το λόγο αυτό χρησιμοποιείται στον έλεγχο της υποθέσης του σταθερού μέτρου θέσης και της σταθερής διασποράς. Το διάγραμμα φέρει στο κατακόρυφο άξονα τις τιμές του μετρούμενου μεγέθους, Y_i , και στον οριζόντιο άξονα τον αύξοντα αριθμό λήψης των τιμών, i ($i=1,2,3,\dots$).
2. **Διάγραμμα χρονικής υστέρησης** (Lag plot) – Ελέγχει την υπόθεση της τυχαιότητας των δεδομένων. Το διάγραμμα φέρει στον κατακόρυφο άξονα τις

τιμές του μετρούμενου μεγέθους, Y_i , για όλες τις τιμές i , ενώ στον οριζόντιο άξονα τις χρονικά μετατοπισμένες κατά ένα βήμα τιμές Y_{i-1} , για όλες τις τιμές i ($i=1,2,3,\dots$).

3. **Ιστόγραμμα δεδομένων (Histogram)** – Ελέγχει την υπόθεση της κανονικής κατανομής των δεδομένων σε συνδιασμό με την τέταρτη τεχνική του διαγράμματος κανονικότητας. Το συνολικό εύρος των τιμών του μετρούμενου μεγέθους διαχωρίζεται σε ισομεγέθη διαστήματα τα οποία ονομάζονται τάξεις. Για κάθε τάξη γίνεται καταμέτρηση των τιμών εκείνων από το σύνολο των δεδομένων που εμπίπτουν στην τάξη αυτή. Το πλήθος των τιμών που εμπίπτουν σε μία τάξη ονομάζεται συχνότητα. Έτσι το ιστόγραμμα φέρει στον κατακόρυφο άξονά του τις τιμές των συχνοτήτων ενώ στον οριζόντιο άξονα τις τιμές του μετρούμενου μεγέθους κατανεμημένες σε τάξεις.
4. **Διάγραμμα κανονικότητας (Normal probability plot)** – Το διάγραμμα αυτό είναι μια γραφική τεχνική μέσω της οποίας διερευνάται εάν τα δεδομένα προέρχονται από μια κατανομή κατά προσέγγιση κανονική. Τα δεδομένα απεικονίζονται στο διάγραμμα σε σχέση με μια υποθετικά ισχύουσα κανονική κατανομή. Το διάγραμμα δημιουργείται τοποθετώντας στον κατακόρυφο άξονα τις ταξινομημένες κατά αύξουσα σειρά τιμές του μετρούμενου μεγέθους ενώ στον οριζόντιο άξονα απεικονίζονται οι λεγόμενες στατιστικές διάμεσοι κανονικής τάξης (normal order statistic medians), $N(i)$, οι οποίες ορίζονται ως $N(i)=G(U(i))$.

Η συνάρτηση $U(i)$ ορίζεται ως:

$$U(i) = 1 - U(n) \text{ for } i = 1$$

$$U(i) = (i - 0.3175)/(n + 0.365) \text{ for } i = 2, 3, \dots, n-1$$

$$U(i) = 0.5^{(1/n)} \text{ for } i = n,$$

ενώ η συνάρτηση G είναι η αντίστροφη συνάρτηση κατανομής (inverse distribution function).

Οι παραπάνω έλεγχοι των βασικών υποθέσεων γίνονται με βάση την εμφάνιση των αντίστοιχων γραφικών απεικονίσεων των δεδομένων ως εξής:

1. **Σειριακό διάγραμμα δεδομένων (Run sequence plot)**. Εάν στο διάγραμμα αυτό τα δεδομένα κατανέμονται οριζόντια χωρίς ενδείξεις αλλαγής κλίσης ή συγκεκριμένης τάσης τότε η υπόθεση του σταθερού μέτρου θέσης είναι αληθής. Επιπλέον εάν η κατακόρυφη διασπορά των δεδομένων κατά μήκος του διαγράμματος παραμένει κατά προσέγγιση σταθερή τότε ισχύει και η υπόθεση της σταθερής διασποράς.
2. **Διάγραμμα χρονικής υστέρησης (Lag plot)**. Εάν η γραφική αυτή απεικόνιση δεν χαρακτηρίζεται από κάποια συγκεκριμένη δομή τότε θεωρούμε ότι τα δεδομένα ικανοποιούν την υπόθεση της τυχαιότητας.
3. **Ιστόγραμμα (Histogram)**. Εάν το ιστόγραμμα παρουσιάζει κατά προσέγγιση την χαρακτηριστική μορφή «καμπάνας» τότε η ισχύουσα κατανομή είναι συμμετρική και ενδεχομένως κατά προσέγγιση κανονική.
4. **Διάγραμμα κανονικότητας (Normal probability plot)**. Εάν το διάγραμμα αυτό απεικονίζει μια κατά προσέγγιση γραμμική συνάρτηση τότε η ισχύουσα κατανομή είναι περίπου κανονική.

Η παραπάνω προσέγγιση στην ανάλυση των δεδομένων εκτός από την απάντηση στο πρώτο ερώτημα δίνει επιπλέον τη δυνατότητα μέσα από τον έλεγχο των τεσσάρων βασικών υποθέσεων με τις αντίστοιχες γραφικές τεχνικές να αποκαλυφθούν ενδεχομένως και άλλα χαρακτηριστικά της διεργασίας τα οποία σε άλλη περίπτωση θα παρέμεναν κεκαλυμμένα. Ενδείξεις για τέτοια χαρακτηριστικά μιας διεργασίας αντιμετωπίζονται με την επιλογή κατάλληλου εργαλείου από το μεγάλο φάσμα τεχνικών που διαθέτει η μεθοδολογία EDA όπως θα καταδειχθεί στη συνέχεια [7].

3. Πειραματικό Μέρος

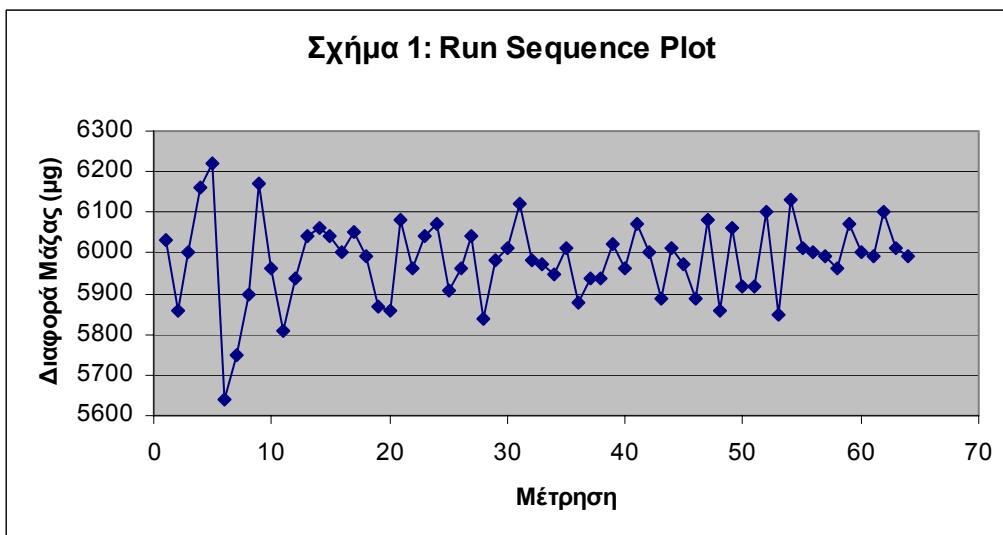
Τα πειραματικά δεδομένα που χρησιμοποιήθηκαν στην παρούσα ανάλυση προέκυψαν από τη συγκριτική ζύγιση ενός προτύπου ελέγχου του εργαστηρίου Μάζας του EIM, ονομαστικής τιμής 10 kg και τάξης ακρίβειας F1 κατά OIML, με πρότυπο αναφοράς ίδιας ονομαστικής τιμής και τάξης ακρίβειας E2. Για τις μετρήσεις χρησιμοποιήθηκε ο συγκριτής μάζας του εργαστηρίου Μάζας του EIM SARTORIUS CC 10000 U-L δυναμικότητας 100060 g. Η ζυγιστική κυψέλη του συγκριτή λειτουργεί με την αρχή της ηλεκτρομαγνητικής αντιστάθμισης δύναμης (electromagnetic force compensation) [8] με ικανότητα αντιστάθμισης βάρους που αντιστοιχεί σε μάζα 60 g. Τα χαρακτηριστικά του μετατροπέα A/D του συγκριτή επιτρέπουν τη μέτρηση μάζας με διακριτική ικανότητα 10 μg . Ο συγκριτής είναι, τέλος, εφοδιασμένος με ηλεκτρομηχανικό σύστημα αυτόματης εναλλαγής βαρών. Τα πρωτογενή δεδομένα αναφέρονται στις διαφορές ενδείξεων του συγκριτή οι οποίες μετά από κατάλληλες διορθώσεις [9] ανάγονται σε διαφορές μάζας Δm μεταξύ του προτύπου ελέγχου T, m_T , και του προτύπου αναφοράς R, m_R . Οι συγκρίσεις στο συγκεκριμένο συγκριτή πραγματοποιούνταν σύμφωνα με την ακολουθία ζυγίσεων RTTR. Η θερμοκρασία και η σχετική υγρασία του χώρου του εργαστηρίου καταγράφονται σε συνεχή βάση με ένα όργανο Ahlborn GmbH με αισθητήρια τύπου NTC και capacitive, αντίστοιχα, με συχνότητα δειγματοληψίας 2 λεπτά. Οι προδιαγραφές λειτουργίας του εργαστηρίου όσον αφορά τα μεγέθη της θερμοκρασίας, της σχετικής υγρασίας και της ταχύτητας του αέρα είναι $\pm 0,1\%$, $\pm 5\%$ και $< 0,05$ m/s, αντίστοιχα, και έχει βρεθεί μετά από σχετικούς ελέγχους από ανεξάρτητο φορέα πιστοποίησης κατά την παράδοση του εργαστηρίου ότι αυτές πληρούνται [10]. Τα πρωτογενή δεδομένα που συλλέχθηκαν σύμφωνα με τα παραπάνω σε χρονική διάρκεια περίπου 2 ημερών αποτελούν το υλικό εφαρμογής των τεχνικών EDA προκειμένου να απαντηθούν τα ερωτήματα που τέθηκαν σε σχέση με την παραπάνω μετρητική διεργασία.

4. Αποτελέσματα

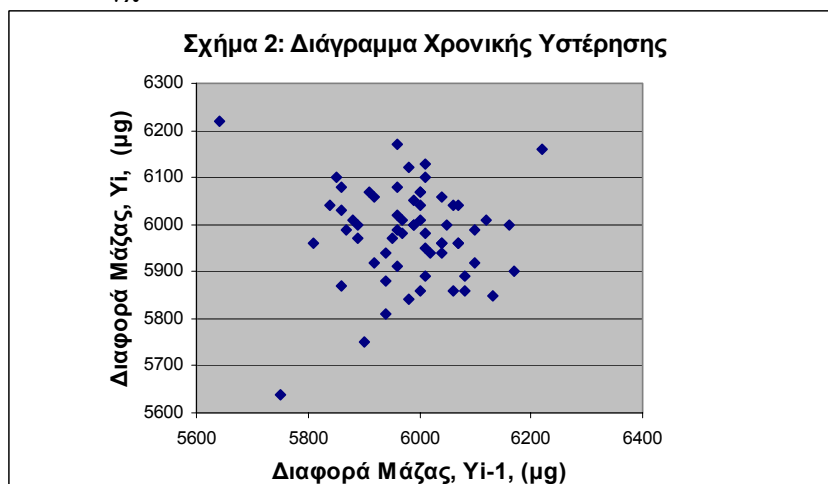
Στο σχήμα 1 παρουσιάζεται το διάγραμμα σειριακής απεικόνισης των πρωτογενών δεδομένων. Από το διάγραμμα αυτό δεν φαίνεται κάποια αξιοσημείωτη μεταβολή στη θέση τους καθ'όλη τη διάρκεια των μετρήσεων. Το ίδιο ισχύει και για τη διασπορά των τιμών η οποία μετά τις πρώτες 8 μετρήσεις παρουσιάζει αξιοσημείωτη σταθερότητα. Η εικόνα αυτή αποτελεί ισχυρή ένδειξη ότι τα δεδομένα προέρχονται από μια κατανομή με σταθερά μέτρα θέσης και διασποράς γεγονός που επιτρέπει να δεχθούμε την ισχύ της τρίτης και της τέταρτης παραδοχής σχετικά με την παρούσα μετρητική διεργασία του Εργαστηρίου Μάζας.

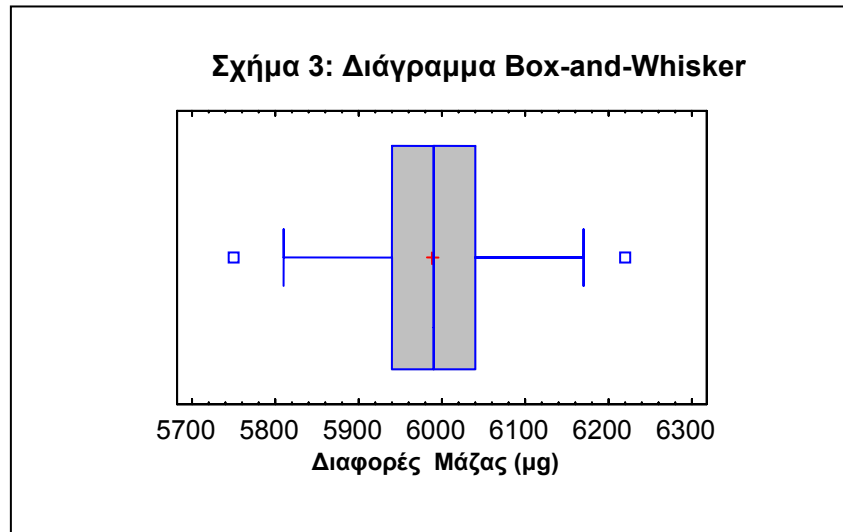
Στο σχήμα 2 παρουσιάζεται το διάγραμμα χρονικής υστέρησης (Lag Plot) κατασκευασμένο με διάστημα υστέρησης $i=1$. Το διάγραμμα αυτό, παρά το γεγονός ότι καταδεικνύει την ύπαρξη κάποιων μεμονωμένων τιμών που θα μπορούσαν να αποτελούν “outliers” της μετρητικής διεργασίας, στερείται κάποιας συγκεκριμένης δομής και παρουσιάζει μια εικόνα η οποία χαρακτηρίζει μια τυχαία κατανομή. Η ύπαρξη, ωστόσο, μεμονωμένων σημείων εκτός της κύριας μάζας των δεδομένων καθιστά απαραίτητο τον έλεγχο της κατανομής για την ύπαρξη “outliers”. Για την ανίχνευση τέτοιων σημείων εφαρμόζονται διάφορες τεχνικές όπως για παράδειγμα η δοκιμή Grubbs, η τιμή z (z-score) και η δοκιμή Rosner [11]. Για τα δεδομένα της παρούσας εργασίας επιλέχθηκε μια γραφική μέθοδος δοκιμής η οποία αποτελεί ένα ισχυρό εργαλείο της τεχνικής EDA για το σκοπό αυτό και είναι το διάγραμμα Box-and-Whisker [12]. Η κατασκευή του διαγράμματος Box-and-Whisker, που παρουσιάζεται στο σχήμα 3, σε αντίθεση με τις

άλλες δοκιμές, δεν προϋποθέτει την παραδοχή μιας συγκεκριμένης κατανομής και δίνει την δυνατότητα της ταυτόχρονης ταυτοποίησης όλων των πιθανών “outliers”. Από το διάγραμμα αυτό προκύπτει ότι τα σημεία με τιμές 6220 μg και 5750 μg , αντίστοιχα, βρίσκονται εκτός των οριακών τιμών και θα μπορούσαν να χαρακτηριστούν “outliers” σε επίπεδο εμπιστοσύνης 95%. Από τα σχήματα 2 και 3 προκύπτει συνεπώς ότι η παραδοχή για την τυχειότητα των δεδομένων φαίνεται να είναι αληθής ιδιαίτερα μετά τον αποκλεισμό δύο σημείων τα οποία χαρακτηρίζονται με βάση την παραπάνω ανάλυση ως “outliers”.

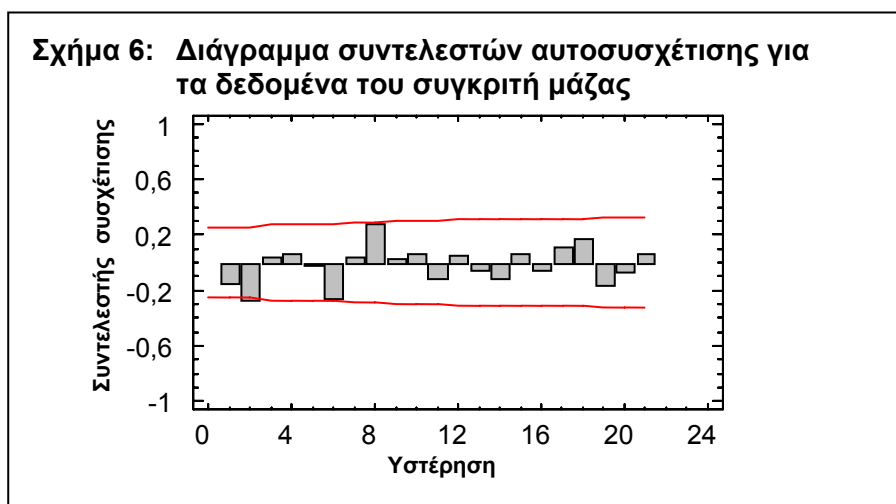
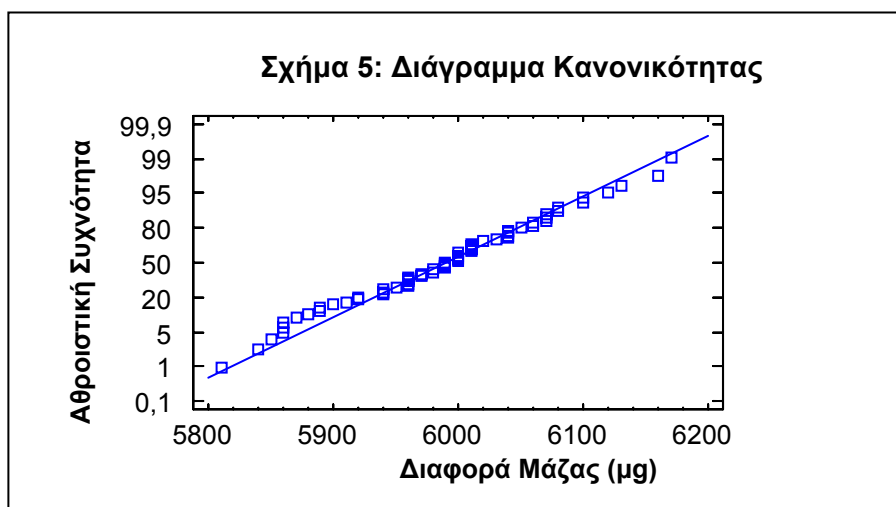
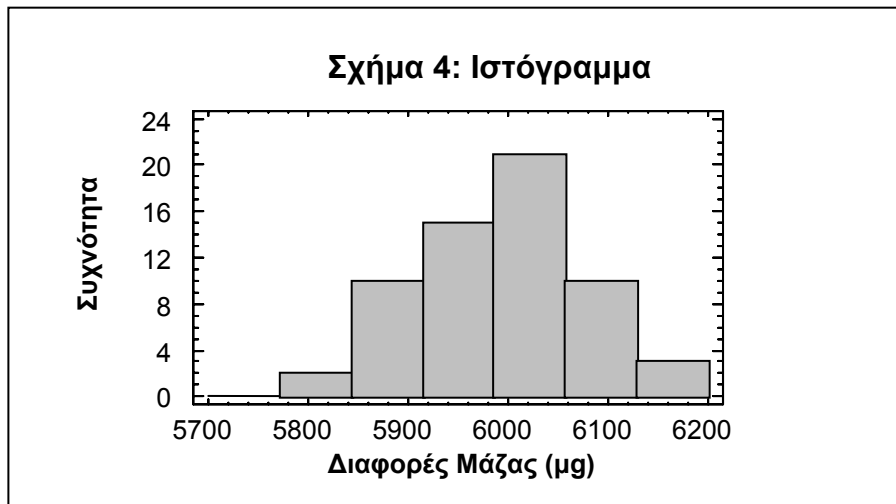


Στο σχήμα 4 παρουσιάζεται το ιστόγραμμα της κατανομής των δεδομένων το οποίο δεν περιλαμβάνει τα δύο ακραία σημεία (outliers). Από το ιστόγραμμα δεν προκύπτει κάποια σημαντική παρέκκλιση από μια κατά προσέγγιση συμμετρική και ενδεχομένως κανονική κατανομή, γεγονός που υποδηλώνει ότι η δεύτερη παραδοχή είναι πιθανόν να είναι αληθής. Ο παραπάνω ισχυρισμός ενισχύεται από το σχήμα 5 στο οποίο παρουσιάζεται το διάγραμμα κανονικότητας των δεδομένων. Στο διάγραμμα αυτό η ευθεία εκφράζει τη θέση πάνω στην οποία θα έπρεπε να βρίσκονται τα δεδομένα εάν περιγράφονταν από μια κανονική κατανομή. Όπως προκύπτει, τα δεδομένα ακολουθούν μια γραμμική σχέση η οποία δεν αποκλίνει σημαντικά από την θεωρητική, γεγονός που επιβεβαιώνει την υπόθεση ότι τα δεδομένα ακολουθούν μια περίπου συμμετρική και κατά προσέγγιση κανονική κατανομή. Από την παραπάνω ανάλυση των δεδομένων προκύπτει ότι η συγκεκριμένη μετρητική διεργασία στο εργαστήριο μάζας βρίσκεται σε κατάσταση στατιστικού ελέγχου.





Η γραφική απεικόνιση μιας διεργασίας σε κατάσταση στατιστικού ελέγχου γίνεται με τυπικές τεχνικές SPC (Statistical Process Control) οι οποίες βασίζονται στην υπόθεση της τυχειότητας των δεδομένων της διεργασίας. Σε περίπτωση, ωστόσο, που τα δεδομένα παρουσιάζουν κάποιο βαθμό αυτοσυσχέτισης απαιτείται η υιοθέτηση ενός παραμετρικού μοντέλου περιγραφής της διεργασίας το οποίο να είναι σε θέση να περιγράψει τον ενδεχομένως δυναμικό χαρακτήρα της [13]. Ένα ισχυρό εργαλείο της EDA προς αυτή την κατεύθυνση είναι το διάγραμμα αυτοσυσχέτισης το οποίο παρουσιάζεται στο σχήμα 6 για το σύνολο των δεδομένων για διάφορα βήματα χρονικής υστέρησης. Το διάγραμμα αυτοσυσχέτισης απεικονίζει την ύπαρξη και το μέγεθος της βραχυπρόθεσμης ή μακροπρόθεσμης «μνήμης» των δεδομένων σε μια διεργασία όπως αυτή εξελίσσεται στο χρόνο [14]. Από το σχήμα 6 προκύπτει ότι, με εξαίρεση την τιμή χρονικής υστέρησης $t=2$, κανένας συντελεστής αυτοσυσχέτισης δεν είναι στατιστικά σημαντικός σε επίπεδο εμπιστοσύνης 95%. Η εξαίρεση για χρόνο υστέρησης $t=2$ εξηγείται ως εξής. Ο συγκριτής μάζας των 10 kg που χρησιμοποιήθηκε για τη λήψη των δεδομένων ακολουθεί κύκλο ζύγισης R-T-T-R. Σε κάθε ακολουθία τεσσάρων ζυγίσεων ο συγκριτής μετρά κατά σειρά το πρότυπο αναφοράς, το πρότυπο ελέγχου, το πρότυπο ελέγχου ξανά και τέλος το πρότυπο αναφοράς. Από τα αποτελέσματα των ζυγίσεων αυτών προέκυψαν τα δεδομένα που παρουσιάζονται εδώ ως διαφορές μαζών $\Delta m_i = m_{Ti} - m_{Ri}$. Ο στατιστικά σημαντικός συντελεστής αυτοσυσχέτισης για $t=2$ καταδεικνύει την συσχέτιση κάθε διαφοράς μάζας με τη μεθεπόμενη της. Αυτό προκύπτει λογικά από την ακολουθία ζύγισης που εκτελεί ο συγκριτής μάζας δεδομένου ότι οι διαφορές που προκύπτουν από ζυγίσεις τύπου R-T σχετίζονται μεταξύ τους και το ίδιο ισχύει και για τις διαφορές που προκύπτουν από ζυγίσεις τύπου T-R. Το σημαντικότερο, ωστόσο, χαρακτηριστικό της συσχέτισης αυτής είναι ότι δεν μεταβάλεται με το χρόνο όπως προκύπτει από το σειριακό διάγραμμα δεδομένων (σχήμα 1). Συνεπώς, στο σύνολό της η παρούσα διεργασία περιγράφεται από μια κατανομή κατά προσέγγιση συμμετρική, με σταθερές παραμέτρους θέσης και διασποράς ενώ παράλληλα παρουσιάζει στοιχεία αυτοσυσχέτισης τα οποία ωστόσο δεν έχουν δυναμική συμπεριφορά. Για τους παραπάνω λόγους η διεργασία βρίσκεται σε κατάσταση στατιστικού ελέγχου και χαρακτηρίζεται από χρονική στασιμότητα (stationarity).



Η πλέον ενδεδειγμένη τεχνική SPC για την απεικόνιση της κατάστασης στατιστικού ελέγχου της παρούσας διεργασίας με τα προαναφερθέντα χαρακτηριστικά είναι η κατασκευή ενός διαγράμματος ελέγχου το οποίο βασίζεται σε ένα μοντέλο ARIMA (AutoRegressive, Integrated, Moving Average) το οποίο λαμβάνει υπόψη του και την αυτοσυσχέτιση των δεδομένων [13,15]. Η επιλογή των κατάλληλων παραμέτρων του μοντέλου ARIMA έγινε με τη βοήθεια του στατιστικού λογισμικού STATGRAPHICS Centurion XV, Version 15.0.04, StatPoint, Inc., [15]. Το διάγραμμα ελέγχου βάσει του

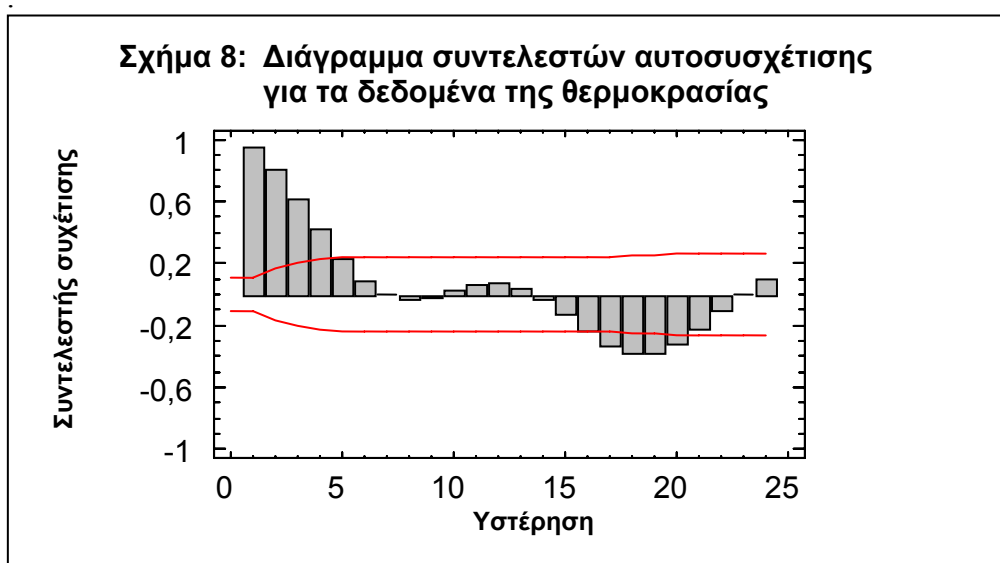
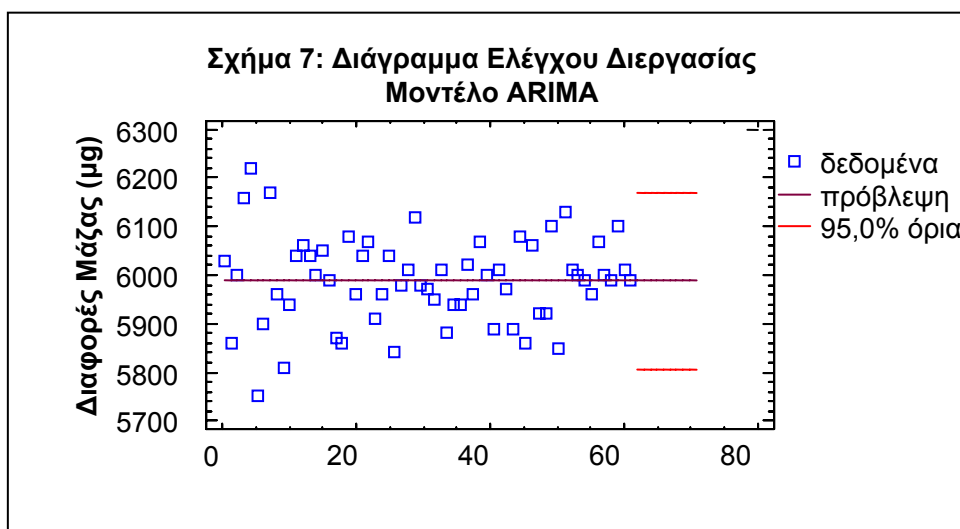
μοντέλου ARIMA απεικονίζεται στο σχήμα 7 στο οποίο φαίνεται και η πρόβλεψη του μοντέλου για 12 χρονικά διαστήματα. Οι παράμετροι της διεργασίας όπως προκύπτουν από την ανάλυση ARIMA παρουσιάζονται στον πίνακα 1. Στον ίδιο πίνακα παρουσιάζονται και οι αντίστοιχες παράμετροι όπως θα προέκυπταν από ένα τυπικό διάγραμμα ελέγχου τύπου EWMA (Exponentially Weighed Moving Average) [16]. Από την συγκριτική παράθεση των αποτελεσμάτων προκύπτει ότι η εκτιμώμενη τυπική απόκλιση της διεργασίας είναι μικρότερη όταν ληφθεί υπόψη η αυτοσυσχέτιση των δεδομένων ενώ η εκτίμηση της θέσης της διεργασίας είναι η ίδια και για τα δύο μοντέλα.

Πίνακας 1: Σύνοψη αποτελεσμάτων μοντελοποίησης της ζυγιστικής διεργασίας

Παράμετροι μοντέλου	Μοντέλο Διεργασίας	
	EWMA	ARIMA
Εκτιμώμενη μέση τιμή	5987,94	5987,94
Εκτιμώμενη τυπική απόκλιση	95,0666	90,3962

Ένα τελευταίο στοιχείο που εξετάζεται στα πλαίσια της παρούσας εργασίας είναι η συμπεριφορά της θερμοκρασίας του χώρου κατά τη διάρκεια των μετρήσεων και ο ρόλος της στη διατήρηση της μετρητικής διεργασίας σε κατάσταση στατιστικής προβλεψιμότητας. Είναι προφανές ότι μια ελεγχόμενη διεργασία όπως αυτή που παρουσιάζεται δεν μπορεί να εκτυλίσσεται σε ένα περιβάλλον μη σταθερών θερμοκρασιακών συνθηκών. Η θερμοκρασία του χώρου ρυθμίζεται μέσω του κλιματισμού ενώ υπάρχει συνεχής καταγραφή της ανά δύο λεπτά. Λόγω της ρύθμισης του κλιματισμού οι τιμές της θερμοκρασίας αναμένεται να ακολουθούν κάποιου είδους περιοδική συμπεριφορά γύρω από τη τιμή ρύθμισης (set point) και εάν αυτό ισχύει θα πρέπει να αποτυπώνεται στο αντίστοιχο διάγραμμα αυτοσυσχέτισης. Το διάγραμμα αυτό φαίνεται στο σχήμα 8 από το οποίο προκύπτουν τα εξής. Υπάρχει ισχυρή συσχέτιση μεταξύ των δεδομένων με χρόνους υστέρησης $t=1, 2, 3$ και 4 η οποία αμέσως μετά αποσβαίνει και εμφανίζεται στη συνέχεια στους χρόνους υστέρησης $t=16, 17, 18, 19$ και 20 .

Η παραπάνω εικόνα ταυτίζεται και επιβεβαιώνει τον τρόπο ρύθμισης της θερμοκρασίας του χώρου η οποία μεταβάλλεται γύρω από το set point ως μια περιοδική συνάρτηση με περίοδο περίπου 15 χρονικά διαστήματα τα οποία για τον ισχύοντα χρόνο δειγματοληψίας των 2 min αντιστοιχεί σε περίπου 30 λεπτά. Η συσχέτιση που εμφανίζεται σε μικρούς χρόνους υστέρησης εκφράζει τη συγγένεια των δεδομένων σε κάθε τέταρτο της περιόδου η οποία όμως λόγω της κυκλικής μεταβολής αποσβαίνει αμέσως μετά. Η επανεμφάνιση της συσχέτισης σε χρόνους υστέρησης ίσους προς 16 δηλώνει την επανέναρξη του κύκλου μεταβολής της θερμοκρασίας.



5. Συμπεράσματα

Τα συμπεράσματα της παρούσας αξιολόγησης της μετρητικής διαδικασίας του εργαστηρίου Μάζας του ΕΙΜ με εφαρμογή των εργαλείων της διερευνητικής ανάλυσης δεδομένων συνοψίζονται στα εξής:

1. Τα γραφικά εργαλεία της διερευνητικής ανάλυσης δεδομένων (EDA) αποτελούν μια εξαιρετικά εύχρηστη και αξιόπιστη μεθοδολογία αξιολόγησης δεδομένων διεργασιών η οποία, σε αντίθεση με τις κλασικές μεθόδους ανάλυσης, δεν υιοθετεί “a-priori” υποθέσεις σχετικά με τη δομή των δεδομένων αλλά αφήνει τα ίδια τα δεδομένα να κατευθύνουν την ανάλυση προς τις κατάλληλες υποθέσεις.
2. Με εφαρμογή των κατάλληλων εργαλείων της EDA προέκυψε ότι η μετρητική διεργασία του εργαστηρίου μάζας ικανοποιεί όλες τις παραδοχές οι οποίες χαρακτηρίζουν μια διεργασία σε κατάσταση στατιστικής προβλεψιμότητας.
3. Το ζυγιστικό σχήμα R-T-T-R που εφαρμόζει γενικά το εργαστήριο μάζας αποτυπώνεται στα δεδομένα έτσι ώστε να σχετίζονται σε βαθμό στατιστικά σημαντικό τα δεδομένα εκείνα που λαμβάνονται με όμοια ακολουθία ζυγίσεων R-T και T-R, αντίστοιχα. Η συσχέτιση αυτή δεν έχει δυναμικό χαρακτήρα και για το λόγο αυτό δεν αλλοιώνει την μορφή της κατανομής των δεδομένων η οποία είναι

κατά προσέγγιση κανονική και χαρακτηρίζεται από σταθερή θέση και σταθερή διασπορά.

4. Η διεργασία περιγράφεται με ένα παραμετρικό μοντέλο ARIMA το οποίο λαμβάνει υπόψη του και τη συσχέτιση των δεδομένων. Η εκτιμώμενη μεταβλητότητα της διεργασίας βάσει αυτού του μοντέλου είναι μικρότερη από την αντίστοιχη εκτίμηση βάσει ενός τυπικού SPC μοντέλου όπως για παράδειγμα ενός μοντέλου EWMA .
5. Τα θερμοκρασιακά δεδομένα αποκαλύπτουν την ύπαρξη στατιστικά σημαντικής αυτοσυσχέτισης σε χρόνους υστέρησης τέτοιους που χαρακτηρίζουν μια περιοδική μεταβολή γύρω από ένα προεπιλεγμένο σταθερό σημείο (set point). Παρόλα αυτά σε αυτό το στάδιο δεν φαίνεται να υπάρχει κάποια επίδραση αυτών των επιτρεπτών (εντός των προδιαγραφών λειτουργίας της κλιματιστικής μονάδας) διακυμάνσεων θερμοκρασίας με μετρητικά αποτελέσματα διαφορών μάζας.
6. Η χρήση και εξαγωγή συμπερασμάτων από την εφαρμογή των τεχνικών EDA είναι εύκολη και μπορεί να ενσωματωθεί εύκολα στις διαδικασίες λειτουργίας των εργαστηρίων διακριβώσεων με σκοπό τον έλεγχο της μεταβλητότητας σε συνδιασμό με άλλες τεχνικές SPC.

6. Βιβλιογραφία

1. J.K. Taylor and H.V. Oppermann, *Handbook for the Quality Assurance of Metrological Measurements*, NBS Handbook **145**, Gaithersburg, 1986.
2. J. Mandel, *The Statistical Analysis of Experimental Data*, Dover Publications, 1984.
3. J.W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, 1977.
4. D.C. Hoaglin, F. Mosteller and J.W. Tukey, *Understanding Robust and Exploratory Data Analysis*, Wiley-Interscience, 2000.
5. P.Lee, *Bayesian Statistics: An Introduction*, Oxford University Press, 1989.
6. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 1st edition, 2001.
7. P. Velleman and D.C. Hoaglin, *The ABC's of EDA: Applications, Basics and Computing of Exploratory Data Analysis*, 1981.
8. R. Schwartz, *Guide to Mass Determination with High Accuracy*, PTB-Bericht MA-40, 1995.
9. Χ. Μήτσας, *Προσδιορισμός Μάζας και Διακρίβωση Προτύπων Βαρών*, Τεχνική Οδηγία MM-MA-02a, Ελληνικό Ινστιτούτο Μετρολογίας, Θεσ/νίκη, 2003.
10. Εσωτερικό έγγραφο ΕΙΜ, *Πρωτόκολλο παράδοσης Εργαστηρίου Μάζας*, 2003.
11. F. Grubbs, *Procedures for Detecting Outlying Observations in Samples*, *Technometrics*, Vol **11**, No. 1, pp. 1-21, 1969.
12. J. Chambers, W. Cleveland, B. Kleiner and P. Tukey, *Graphical Methods for Data Analysis*, Duxbury Press, 1983.
13. C.W. Lu and M.R. Reynolds, *Control Charts for Monitoring the Mean and Variance of Autocorrelated Processes*, *Journal of Quality Technology*, **31**, pp. 259-274, 1999.
14. G.E.P. Box, G. Reinsel and G. Jenkins, *Time Series Analysis: Forecasting and Control*, Prentice Hall, 3rd edition, 1994.
15. N. W. Polhemus, *How to Construct a Control Chart for Autocorrelated Data Using STATGRAPHICS Centurion*, Statpoint, Inc., 2005.
16. Montgomery, Douglas C. *Introduction to Statistical Quality Control*. New York: John Wiley & Sons, 1991.