



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ  
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ  
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ  
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ  
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



## ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

### **Data Analytics in Batch Production**

**Σταύρος Α. Γιαγιάννος**

**Επιβλέπων :** Δημήτριος Ασκούνης  
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2019





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ  
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ  
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ  
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ  
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



## ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

### **Data Analytics in Batch Production**

**Σταύρος Α. Γιαγιάννος**

**Επιβλέπων :** Δημήτριος Ασκούνης  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 17<sup>η</sup> Ιουνίου 2019.

*(Υπογραφή)*

.....

Δημήτριος Ασκούνης  
Καθηγητής Ε.Μ.Π.

*(Υπογραφή)*

.....

Καθηγητής Ε.Μ.Π.

*(Υπογραφή)*

.....

Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2019

(Υπογραφή)

.....

**ΣΤΑΥΡΟΣ Α. ΓΙΑΓΙΑΝΝΟΣ**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Σταύρος Α. Γιαγιάννος

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

## ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της παρούσης εργασίας θα ήθελα να εκφράσω τις ευχαριστίες μου σε όσους βοήθησαν κατά την εκπόνησή της.

Ευχαριστώ θερμά τον επιβλέποντα της εργασίας, κ. Δημήτριο Ασκούνη, Καθηγητή ΕΜΠ, για την ευκαιρία που μου έδωσε να ασχοληθώ με το παρόν θέμα καθώς και για την υποστήριξη που μου παρείχε.

Θα ήθελα επίσης να ευχαριστήσω τον Υποψήφιο Διδάκτορα, κ. Μιχάλη Αυγουλή, για την πολύτιμη καθοδήγηση του σε όλα τα στάδια εκπόνησης της εργασίας, καθώς και για την πρόσβαση, την οποία μου παρείχε, στα δεδομένα που αξιοποιήθηκαν στην παρούσα εργασία. Τέλος, ευχαριστώ ιδιαιτέρως τους γονείς μου, Αθανασία και Αριστεΐδη και την αδερφή μου, Δάφνη, για την εμπιστοσύνη που μου δείχνουν καθημερινά, καθώς και τους φίλους μου για την αμέριστη στήριξή τους.



## Περίληψη

Σκοπός της παρούσας εργασίας είναι η ανάδειξη της προστιθέμενης αξίας που προσφέρει η εφαρμογή σύγχρονων μεθόδων ανάλυσης δεδομένων στη βιομηχανία και συγκεκριμένα στην Παραγωγή σε Παρτίδες. Σε αυτό το πλαίσιο, η εργασία επικεντρώνεται στην ανάδειξη των ιδιαιτεροτήτων του συγκεκριμένου κλάδου της βιομηχανίας, στην αναζήτηση και στην αξιολόγηση των καταλληλότερων μεθόδων και τεχνικών ανάλυσης δεδομένων σε μια διαδικασία παραγωγής σε παρτίδες καθώς και στην αξιοποίηση πραγματικών δεδομένων για τη δημιουργία, την εκπαίδευση, την εφαρμογή και την αξιολόγηση μοντέλων Ανάλυσης Δεδομένων και τελικά στην ερμηνεία και αξιοποίηση των αποτελεσμάτων με σκοπό τη βελτίωση μιας πραγματικής βιομηχανικής διαδικασίας παραγωγής. Αξίζει να σημειωθεί ότι η εργασία κινείται σε ακαδημαϊκό επίπεδο για να εξετάσει και να προσδιορίσει το πλαίσιο εφαρμογής της πολυμεταβλητής ανάλυσης σε μια παραγωγική διαδικασία, αξιοποιώντας πραγματικά δεδομένα για ρεαλισμό, χωρίς καμία ωστόσο δυνατότητα ή διάθεση εμπορικής εκμετάλλευσης των αποτελεσμάτων.

**Λέξεις Κλειδιά:** <<Ανάλυση Δεδομένων, Παραγωγή σε Παρτίδες, Ανάλυση Κυρίων Συνιστωσών, Παλινδρόμηση Μερικών Ελαχίστων Τετραγώνων, Διοίκηση Παραγωγής>>

## **Abstract**

The purpose of this thesis is to highlight the added value offered by the application of modern methods of data analysis to the industry and specifically to Batch Production. In this context, the work focuses on highlighting the peculiarities of this particular industry, while an extensive search for the most appropriate data analysis methods and techniques for a batch production process, is also being carried out. Furthermore, the thesis focuses on taking advantage of real data for the design, training, implementation and evaluation of Data Analysis models and ultimately pursues the interpretation and exploitation of the results in order to improve a real industrial production process. It is worth noting that this thesis is of strictly academic interest and it aims at examining the possibility of applying multivariable analysis to a manufacturing procedure. The use of real data is intended to provide a realistic dimension to the produced results, however, there is neither the possibility nor the intention of commercial use.

**Keywords:** <<Data Analysis, Batch Production, Batch Manufacturing, Principal Component Analysis, Partial Least Squares Regression, Production Management>>



## ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΚΕΦΑΛΑΙΟ 1: <b>Εισαγωγή</b> .....	1
1.1 Batch Production .....	1
1.1.1 Γενικά.....	1
1.1.2 Χαρακτηριστικά.....	1
1.1.3 Πλεονεκτήματα και μειονεκτήματα.....	3
1.2 Αντικείμενο της εργασίας.....	3
1.2.1 Προκλήσεις.....	3
1.2.2 Συνεισφορά.....	5
1.3 Δομή της εργασίας.....	5
ΚΕΦΑΛΑΙΟ 2: <b>Data Analytics</b> .....	7
2.1 Γενικά .....	7
2.2 Κατηγορίες Data Analytics.....	7
2.3 Data Analytics στη Βιομηχανία.....	10
2.4 Data Analytics στην Παραγωγή σε Παρτίδες.....	13
ΚΕΦΑΛΑΙΟ 3: <b>Μέθοδοι και Τεχνικές</b> .....	19
3.1 Γενικά .....	19
3.2 Πολλαπλή Γραμμική Παλινδρόμηση - MLR .....	22
3.2.1 Γενικά.....	22
3.2.2 Μαθηματικό Μοντέλο.....	23
3.2.3 Ερμηνεία.....	24
3.2.4 Εφαρμογές.....	25
3.2.5 Πλεονεκτήματα και Μειονεκτήματα.....	25
3.3 Ανάλυση Κυρίων Συνιστωσών - PCA.....	26
3.3.1 Γενικά.....	26
3.3.2 Μαθηματικό Μοντέλο.....	27
3.3.3 Ερμηνεία.....	29
3.3.4 Εφαρμογές.....	33
3.3.5 Πλεονεκτήματα και Μειονεκτήματα.....	34

3.4	Παλινδρόμηση Κυρίων Συνιστωσών - PCR.....	34
3.4.1	Γενικά.....	34
3.4.2	Μαθηματικό Μοντέλο.....	36
3.4.3	Ερμηνεία.....	37
3.4.4	Εφαρμογές.....	37
3.4.5	Πλεονεκτήματα και Μειονεκτήματα.....	38
3.5	Ανάλυση Μερικών Ελαχίστων Τετραγώνων– PLS/OPLS.....	39
3.5.1	Γενικά.....	39
3.5.2	Μαθηματικό Μοντέλο.....	40
3.5.3	Ερμηνεία.....	42
3.5.4	Εφαρμογές.....	42
3.5.5	Πλεονεκτήματα και Μειονεκτήματα.....	43
<b>ΚΕΦΑΛΑΙΟ 4: Το λογισμικό SIMCA .....</b>		<b>45</b>
4.1	Εισαγωγή.....	45
4.2	Γενικά.....	46
4.3	Εφαρμογή στην Παραγωγή σε Παρτίδες.....	51
4.3.1	Γενικά.....	51
4.3.2	<i>Batch Evolution Model</i> .....	52
4.3.3	<i>Batch Level Model</i> .....	54
<b>ΚΕΦΑΛΑΙΟ 5: Εφαρμογή .....</b>		<b>57</b>
5.1	Η Διαδικασία Παραγωγής.....	57
5.1.1	Γενικά.....	57
5.1.2	Περιγραφή.....	57
5.1.3	Παρακολούθηση της διεργασίας / <i>Monitoring</i> .....	62
5.1.4	<i>Dataset</i> .....	63
5.1.5	Διατύπωση του Προβλήματος.....	64
5.2	Μοντελοποίηση.....	64
5.2.1	Επεξεργασία του <i>Dataset</i> .....	64
5.2.2	Κατασκευή και ερμηνεία των Μοντέλων <i>BEM</i> και <i>BLM</i> .....	65

ΚΕΦΑΛΑΙΟ 6: Σύνοψη.....	81
6.1 Συμπεράσματα.....	81
6.2 Προτάσεις για μελέτη .....	83
ΚΕΦΑΛΑΙΟ 7: Βιβλιογραφία.....	85



# ΚΕΦΑΛΑΙΟ 1: *Εισαγωγή*

---

## 1.1 Batch Production

### 1.1.1 Γενικά

Με τον όρο Batch Production ή Παραγωγή σε παρτίδες, αναφερόμαστε σε εκείνον τον τρόπο οργάνωσης της παραγωγικής διαδικασίας μιας βιομηχανικής μονάδας παραγωγής, κατά τον οποίο μια περιορισμένη και αυστηρώς καθορισμένη ποσότητα, από κάθε είδος προϊόντος, τίθεται προς παραγωγή, ανά καθορισμένες χρονικές περιόδους. Το τελικό προϊόν προκύπτει από την εφαρμογή διακριτών βημάτων παραγωγής, κατά τη διάρκεια των οποίων οι πρώτες ύλες διέρχονται από μια σειρά σταθμών εργασίας, όπου υπόκεινται στην εκάστοτε απαραίτητη επεξεργασία. Η επανάληψη της διαδικασίας έχει σαν αποτέλεσμα την παραγωγή διακριτών παρτίδων του ίδιου τελικού προϊόντος, με αυστηρώς καθορισμένη αρχή και τέλος, τα ποιοτικά χαρακτηριστικά των οποίων, είναι ομοιόμορφα μεν σε κάθε παρτίδα, δύνανται όμως να παρουσιάζουν σημαντικές αποκλίσεις μεταξύ τους, στο πλαίσιο, πάντα, των τεχνικών προδιαγραφών που καθορίζονται από τις ανάγκες του καταναλωτή και των αντίστοιχων ανοχών απόκλισης από αυτές.

Μαζί με την παραγωγή εργασίας (one-off production) και τη μαζική παραγωγή (flow production or continuous production), η Παραγωγή σε Παρτίδες είναι μία από τις τρεις βασικές παραγωγικές μεθόδους. Η παραγωγή σε παρτίδες συναντάται συχνά σε αρτοποιεία, στην παραγωγή αθλητικών παπουτσιών, φαρμακευτικών υλικών, μελανιών, χρωμάτων και γενικώς φαρμακευτικών προϊόντων.

### 1.1.2 Χαρακτηριστικά

Ορισμένα από τα πιο χαρακτηριστικά γνωρίσματα της παραγωγής σε παρτίδες είναι τα ακόλουθα:

➤ Σύντομες περίοδοι παραγωγής:

Η παραγωγή σε παρτίδες χαρακτηρίζεται από σύντομες περιόδους παραγωγής καθώς και συχνές μεταβολές στη διάταξη του εξοπλισμού. Η αναδιάταξη του

εξοπλισμού δίνει μεγάλη ευελιξία στον όγκο παραγωγής και στην ποικιλία των προϊόντων που μπορούν να παραχθούν.

➤ Εξειδικευμένο εργατικό δυναμικό:

Από το εργατικό δυναμικό απαιτείται εξειδίκευση σε μια συγκεκριμένη διεργασία. Οι χειριστές των μηχανών μπορεί να υλοποιούν απλές διατάξεις, αλλά είναι οι εξειδικευμένοι εργάτες που υλοποιούν τις πιο σύνθετες.

➤ Εξειδικευμένοι επιβλέποντες των εργασιών:

Όπως και στην περίπτωση των εργατών, έτσι και οι επιβλέποντες των επιμέρους εργασιών οφείλουν να διαθέτουν βαθιά γνώση μιας συγκεκριμένης εργασίας

➤ Μηχανές γενικής χρήσης και ευέλικτη χωροταξική διάταξη:

Η προμήθεια και η διάταξη του εξοπλισμού γίνεται με κύριο γνώμονα την ευελιξία. Μηχανές και εξοπλισμός γενικής χρήσης που μπορούν, με ελάχιστες ρυθμίσεις, να χρησιμοποιηθούν για την υλοποίηση μιας ποικιλίας εργασιών, ομαδοποιούνται, οργανώνονται και διατάσσονται στο χώρο αναλόγως με τον τύπο της διεργασίας για τον οποίο προορίζονται.

➤ Χειρωνακτική διαχείριση υλικών:

Σε αντίθεση με άλλες μεθόδους παραγωγής, η διαχείριση και η μεταφορά των υλικών που απαιτείται στην παραγωγή σε παρτίδες, είναι μικρής κλίμακας και μπορεί να γίνεται με τη χρήση trolleys για παράδειγμα.

➤ Ουρές αναμονής:

Συχνά, παρτίδες εργασίας τείνουν να συνωστιζονται σε διάφορες μηχανές, σχηματίζοντας ουρές αναμονής, λόγω των διαφορετικών χρόνων επεξεργασίας, μεγεθών παρτίδας καθώς και λόγω της διαφορετικής αλληλουχίας εκτέλεσης κάποιων εργασιών κατά την παραγωγή διαφορετικών τελικών προϊόντων.

➤ Ευελιξία στον προγραμματισμό παραγωγής

Οι αναπάντεχες βλάβες και η απουσία προσωπικού δεν επηρεάζουν σε μεγάλο βαθμό την παραγωγή, καθώς η εργασία μπορεί να μεταφερθεί σε κάποια άλλη μηχανή και να ανατεθεί σε διαφορετικό χειριστή.

➤ Προγραμματισμός και έλεγχος παραγωγής:

Ο προγραμματισμός και ο έλεγχος της παραγωγής, στην περίπτωση της παραγωγής σε παρτίδες είναι πιο περίπλοκος απ' ό,τι σε άλλες μεθόδους παραγωγής. Συγκεκριμένα, ο έλεγχος των υλικών και των εργαλείων είναι

εξαιρετικά σημαντικός και θα πρέπει να εξασφαλίζεται η συνεχής και απρόσκοπτη επανατροφοδοσία., λεπτομερή λειτουργικά σχεδιαγράμματα πρέπει να ετοιμάζονται για κάθε επιμέρους τμήμα του προϊόντος και ο προγραμματισμός φόρτωσης/τροφοδοσίας και λειτουργίας της κάθε μηχανής είναι μοναδικός και θα πρέπει να υλοποιείται ξεχωριστά.

### **1.1.3 Πλεονεκτήματα και μειονεκτήματα**

Το μειωμένο αρχικό κεφάλαιο, η ευελιξία στην παραγωγή ποικιλίας προϊόντων και η μειωμένη ανάγκη για διατήρηση αποθέματος πρώτων υλών αποτελούν μερικά από τα κυριότερα πλεονεκτήματα της παραγωγής σε παρτίδες. Επιπλέον, το γεγονός ότι τα μηχανήματα αφενός δεν είναι διαρκώς σε λειτουργία και αφετέρου μπορούν να χρησιμοποιηθούν στην παραγωγή διαφορετικών προϊόντων, έχει σαν αποτέλεσμα τόσο τα μειωμένα κόστη κατανάλωσης ενέργειας όσο και το μειωμένο κόστος εξοπλισμού αντιστοίχως. Τέλος η παραγωγή σε παρτίδες διευκολύνει την παραγωγή δοκιμαστικών παρτίδων.

Σε αντιδιαστολή με τα προαναφερθέντα πλεονεκτήματα, η παραγωγή σε παρτίδες παρουσιάζει και μειονεκτήματα, όπως είναι ο «νεκρός χρόνος» που παρεμβάλλεται μεταξύ της παραγωγής δύο παρτίδων, κατά τη διάρκεια του οποίου η παραχθείσα παρτίδα θα πρέπει να ελεγχθεί ως προς τα ποιοτικά της χαρακτηριστικά και την ομοιομορφία της προτού παραχθεί η επόμενη και να πραγματοποιηθούν οι απαραίτητες ρυθμίσεις του εξοπλισμού. Επιπλέον στα μειονεκτήματα συμπεριλαμβάνονται οι αυξημένες απαιτήσεις σε χώρο επεξεργασίας των πρώτων υλών και αποθήκευσης μεγάλων παρτίδων τελικού προϊόντος καθώς και η έλλειψη κινήτρου και ενθουσιασμού από πλευράς των εργαζομένων, η οποία απορρέει από το μεγάλο βαθμό εξειδίκευσης τους και την ενασχόλησή τους με μια συγκεκριμένη εργασία.

## **1.2 Αντικείμενο της εργασίας**

### **1.2.1 Προκλήσεις**

Η λογική ελέγχου που χρησιμοποιείται στην επεξεργασία παρτίδων συχνά περιπλέκεται από το γεγονός ότι μπορεί να υπάρχουν πολλά προϊόντα που παράγονται με τον ίδιο εξοπλισμό. Το ευρύ φάσμα των συνθηκών λειτουργίας και οι συχνές αλλαγές στη διαδικασία, αποτελούν πρόκληση από άποψη μετρήσεων και ελέγχου.

Επιπλέον, συχνά υπάρχουν προκλήσεις κατά τη συγκέντρωση, τον «καθαρισμό», τον εξορθολογισμό και την ιεράρχηση των αναγκαίων και επαρκών δεδομένων για τη διεξαγωγή μιας λεπτομερούς ανάλυσης, η οποία θα αναδείξει τις πτυχές εκείνες της διαδικασίας, στις οποίες εντοπίζεται είτε περιθώριο βελτίωσης είτε κάποια πηγή συστηματικής παραγωγής σφαλμάτων. Ως συνέπεια των παραπάνω προκλήσεων, ιστορικά, η εφαρμογή των online analytics για ανίχνευση σφαλμάτων σε πραγματικό χρόνο καθώς και πρόβλεψη των τελικών τιμών των παραμέτρων ποιότητας δεν έχει αντιμετωπιστεί επαρκώς. Ωστόσο, η δυνατότητα αξιολόγησης της αποτελεσματικότητας της διαδικασίας παραγωγής και η δυνατότητα διαρκούς βελτίωσής της, παραμένουν προτεραιότητες για εταιρείες που χρησιμοποιούν την μέθοδο της παραγωγής παρτίδων.

Εξ ορισμού, κάθε είσοδος σε μια διαδικασία παραγωγής παρτίδας μπορεί δυνητικά να επηρεάσει μερικές ή όλες τις μετρήσεις που λαμβάνονται από το εκάστοτε τμήμα του εξοπλισμού. Αυτός ο διαδραστικός χαρακτήρας της διαδικασίας, σε συνδυασμό με τη βραδεία απόκριση στην ενσωμάτωση νέων εισόδων, που χαρακτηρίζει την παραγωγή σε παρτίδες, καθιστά δύσκολο, για τον χειριστή, τον εντοπισμό ανεπιθύμητων συνθηκών στην παρτίδα και την αξιολόγηση του αντίκτυπού τους στην ποιότητα του τελικού προϊόντος. Το γεγονός ότι οι ποιοτικές παράμετροι συχνά δεν είναι διαθέσιμες ως "on-line" μετρήσεις περιπλέκει τη δουλειά του χειριστή ακόμα περισσότερο. Επιπλέον, ιστορικά, η προσέγγιση την οποία έχουν ακολουθήσει οι μηχανικοί για την κατανόηση των διαδικασιών παραγωγής παρτίδων, συνοψίζεται στη χρήση απλών αναφορών, περιληπτικών στατιστικών στοιχείων που προκύπτουν μετά από την ολοκλήρωση της κάθε παρτίδας και απλή παρακολούθηση και αξιολόγηση των τάσεων που παρατηρούνται κατά τη διαδικασία. Παρότι χρήσιμες και ενημερωτικές, οι προσεγγίσεις αυτές αδυνατούν να ερμηνεύσουν και να αιτιολογήσουν τις διακυμάνσεις που μπορεί να παρουσιάζει η διαδικασία και να αποκωδικοποιήσουν την πολυπαραγοντική φύση που διέπει τις υπό εξέταση διεργασίες. Τα τελευταία χρόνια, έχει σημειωθεί σημαντική πρόοδος στην ανάπτυξη της τεχνολογίας και της κατανόησης που απαιτούνται για την αποτελεσματική εφαρμογή on-line στατιστικών αναλύσεων στην παραγωγή σε παρτίδες. Ενσωμάτωση τέτοιων εργαλείων στο σύστημα ελέγχου μιας διαδικασίας μπορεί να προσφέρει πολλά πλεονεκτήματα.



### 1.2.2 Συνεισφορά

Στην παρούσα διπλωματική εργασία επιχειρείται να δοθεί μια πρόταση αντιμετώπισης των παραπάνω προκλήσεων. Συγκεκριμένα η εν λόγω εργασία θα επικεντρωθεί στην αναζήτηση μεθόδων και τεχνικών για την εφαρμογή Data Analytics (D.A.) σε μια διαδικασία παραγωγής σε παρτίδες, στην αξιοποίηση πραγματικών δεδομένων για τη δημιουργία, την εκπαίδευση, την εφαρμογή και την αξιολόγηση μοντέλων D.A. και τέλος στην ερμηνεία και αξιοποίηση των αποτελεσμάτων με σκοπό τη βελτίωση της διαδικασίας.

Η συνεισφορά της διπλωματικής συνοψίζεται ως εξής:

1. Αναζήτηση και αξιολόγηση των επικρατέστερων αλγορίθμων Data Analytics για εφαρμογή σε διαδικασίες παραγωγής σε παρτίδες.
2. Περιγραφή υπολογιστικού λογισμικού για την εφαρμογή των προαναφερθεισών μεθόδων.
3. Συγκέντρωση πραγματικών δεδομένων βιομηχανικής διαδικασίας παραγωγής σε παρτίδες.
4. Υλοποίηση και εφαρμογή των μεθόδων Data Analytics και αξιολόγηση των αποτελεσμάτων τους.

### 1.3 Δομή της εργασίας

Όπως αναφέρθηκε προηγουμένως, στην παρούσα εργασία θα αναζητήσουμε εκείνες τις μεθόδους ανάλυσης δεδομένων, των οποίων η εφαρμογή μπορεί να προσφέρει κάποια προστιθέμενη αξία σε μια διαδικασία παραγωγής σε παρτίδες, θα τις εφαρμόσουμε σε πραγματικά δεδομένα τα οποία έχουν συλλεχθεί κατά τη διάρκεια μιας τέτοιας διαδικασίας, θα αξιολογήσουμε τα αποτελέσματά τους και θα προσπαθήσουμε να εξαγάγουμε κάποια συμπεράσματα για την εν λόγω διαδικασία παραγωγής, με απώτερο σκοπό την καλύτερη κατανόησή της και τελικά τη βελτίωσή της.

Σε αυτό το πλαίσιο, ο αναγνώστης αρχικά εισάγεται στην έννοια των Data Analytics και στην εφαρμογή τους στη βιομηχανία. Παρουσιάζονται τα οφέλη που μπορεί να αποφέρει η εφαρμογή μεθόδων ανάλυσης δεδομένων στη βιομηχανική παραγωγή και δίνεται ιδιαίτερη έμφαση στην εφαρμογή τους σε διαδικασίες παραγωγής σε παρτίδες.

Στη συνέχεια γίνεται μια βιβλιογραφική ανασκόπηση των πιο δημοφιλών μεθόδων ανάλυσης δεδομένων στην παραγωγή σε παρτίδες και παρουσιάζονται αναλυτικά, το μαθηματικό υπόβαθρο τους, η φυσική τους ερμηνεία, τα πλεονεκτήματα καθώς και οι αδυναμίες των μεθόδων αυτών.

Ακολούθως, παρουσιάζεται το εμπορικό λογισμικό με τη βοήθεια του οποίου θα υλοποιηθούν και θα εφαρμοσθούν ορισμένες από τις προαναφερθείσες μεθόδους ανάλυσης δεδομένων. Συγκεκριμένα παρουσιάζονται οι επιμέρους λειτουργίες του εν λόγω λογισμικού καθώς και τα διάφορα μοντέλα μέσω των οποίων υλοποιούνται οι προς εξέταση αλγόριθμοι.

Τέλος, γίνεται μια εκτενής παρουσίαση του πρακτικού τμήματος της παρούσας εργασίας, ήτοι της πρακτικής εφαρμογής των προαναφερθέντων μεθόδων και εργαλείων σε μια πραγματική βιομηχανική διαδικασία παραγωγής σε παρτίδες. Συγκεκριμένα, παρουσιάζονται η υπό εξέταση διαδικασία, αυτή καθαυτή, το σύνολο των προς ανάλυση δεδομένων που προκύπτουν κατά τη διάρκεια της διαδικασίας αυτής καθώς και οι πτυχές της εκείνης, την αποκωδικοποίηση και ερμηνεία των οποίων καλούμαστε να φέρουμε εις πέρας στην παρούσα εργασία. Επιπλέον, περιγράφεται λεπτομερώς η πειραματική διαδικασία η οποία εφαρμόζεται και αφορά στη συλλογή, επεξεργασία και ανάλυση των δεδομένων και την ερμηνεία των αντίστοιχων αποτελεσμάτων.

Τέλος, παρατίθενται τα συμπεράσματα που προέκυψαν και προτείνεται μια σειρά από θέματα που παρουσιάζουν ενδιαφέρον για περαιτέρω μελέτη.

## ΚΕΦΑΛΑΙΟ 2: *Data Analytics*

---

### 2.1 Γενικά

Με τον όρο Data Analytics, ελληνιστί Ανάλυση Δεδομένων, αναφερόμαστε στη διαδικασία εξέτασης συνόλων ακατέργαστων δεδομένων, προκειμένου να εξαχθούν χρήσιμα συμπεράσματα σχετικά με τις πληροφορίες που περιέχουν. Η διαδικασία αυτή συνίσταται στη συλλογή και κατηγοριοποίηση δεδομένων και στην εφαρμογή ποιοτικών και ποσοτικών τεχνικών και διαδικασιών, μέσω των οποίων μπορούν να ανιχνευθούν πρότυπα και τάσεις, να αποκαλυφθούν επαναλαμβανόμενα μοτίβα και να αποκωδικοποιηθούν συσχετίσεις μεταξύ των διαφορετικών παρατηρούμενων μεγεθών. Τα συμπεράσματα που προκύπτουν από μια τέτοια διαδικασία συμβάλλουν στη λήψη πιο εμπειριστατωμένων επιχειρηματικών αποφάσεων, κατά την οποία λαμβάνονται υπόψιν όλες οι πτυχές της ανάλυσης, ενώ παράλληλα μπορούν να αξιοποιηθούν από οργανισμούς και εταιρείες, προκειμένου να κατανοήσουν εις βάθος και να βελτιστοποιήσουν διαδικασίες, να εντοπίσουν αδυναμίες, πλεονεκτήματα και επιχειρηματικές ευκαιρίες, να επαληθεύσουν ή να θέσουν υπό αμφισβήτηση υπάρχουσες θεωρίες και μοντέλα, παρέχοντας τελικά, με αυτόν τον τρόπο μια σημαντική προστιθέμενη αξία στις λειτουργίες τους.

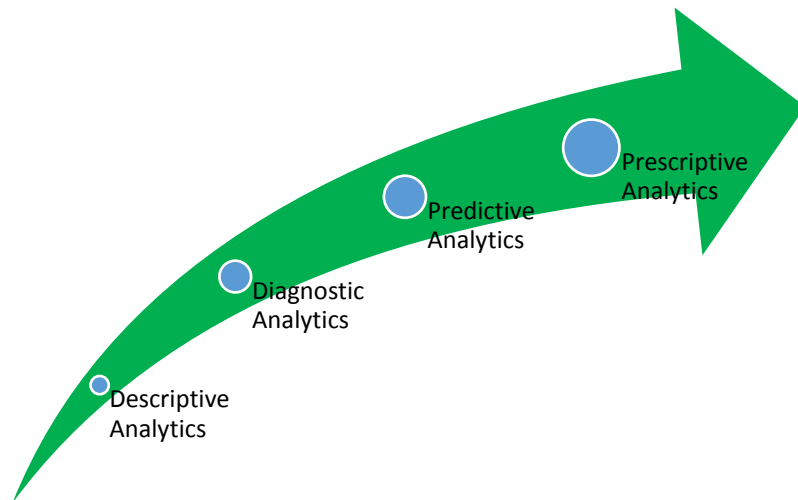
### 2.2 Κατηγορίες Data Analytics

Με βάση τη φάση της ροής εργασιών και το είδος της απαιτούμενης ανάλυσης, μπορούμε να διακρίνουμε τέσσερις βασικές κατηγορίες ανάλυσης δεδομένων που χρησιμοποιούνται σε όλες τις βιομηχανίες. Οι κατηγορίες αυτές παρότι είναι διακριτές, συνδέονται μεταξύ τους και χτίζονται η μια πάνω στην άλλη. Καθώς αρχίζουμε να μετακινούμαστε από τον απλούστερο τύπο ανάλυσης δεδομένων προς τους περισσότερο σύνθετους, ο βαθμός δυσκολίας εφαρμογής του καθώς και ο αριθμός των απαιτούμενων πόρων αυξάνεται. Ταυτόχρονα όμως, το επίπεδο της προστιθέμενης γνώσης και της αξίας αυξάνεται επίσης.

Οι τέσσερις βασικές κατηγορίες ανάλυσης δεδομένων είναι οι εξής:

1. Περιγραφική Ανάλυση Δεδομένων (Descriptive Analytics)
2. Διαγνωστική Ανάλυση Δεδομένων (Diagnostic Analytics)

3. Προγνωστική Ανάλυση Δεδομένων (Predictive Analytics)
4. Καθοδηγητική Ανάλυση Δεδομένων (Prescriptive Analytics)



Εικόνα 2-1 Σχέση πολυπλοκότητας – Προστιθέμενης αξίας των διαφόρων κατηγοριών Analytics

➤ Περιγραφική Ανάλυση Δεδομένων

Πρόκειται για το θεμέλιο λίθο οποιουδήποτε είδους ανάλυσης δεδομένων. Αποτελεί την απλούστερη και συνηθέστερη κατηγορία ανάλυσης δεδομένων σε μια επιχείρηση και έγκειται στην συλλογή, προ-επεξεργασία και σύνοψη ανεπεξέργαστων δεδομένων και την μετατροπή τους σε μια μορφή εύκολα κατανοητή από τους ανθρώπους. Η περιγραφική ανάλυση απαντά στο ερώτημα, «τι συνέβη;», με την περίληψη παλαιών δεδομένων χωρίς ωστόσο να παρέχει μια ερμηνεία στο ερώτημα «γιατί συνέβη;». Οι περιγραφικές αναλύσεις είναι χρήσιμες, επειδή διευκολύνουν την παρακολούθηση των κρίσιμων μετρικών απόδοσης και μας επιτρέπουν να μάθουμε από τις συμπεριφορές του παρελθόντος και να κατανοήσουμε πώς αυτές μπορούν να επηρεάσουν τις μελλοντικές εκβάσεις.

Τα ενδεικτικά βήματα της περιγραφικής ανάλυσης έχουν ως εξής:

1. Προσδιορισμός των επιχειρηματικών μετρικών έναντι σε ορισμένους στόχους και καθορισμός των KPI's για την παρακολούθηση της προόδου.
2. Συγκέντρωση και προετοιμασία των δεδομένων. Όταν ολοκληρωθεί η συλλογή των σωστών δεδομένων από τα συστήματα επιχειρήσεων, πρέπει να χρησιμοποιηθούν τεχνικές περιγραφικής ανάλυσης για την

προετοιμασία τους (για παράδειγμα, εξάλειψη διπλών εγγραφών και πιθανών εσφαλμένων μηδενικών μετρήσεων).

3. Ειδικοί εμπειρογνώμονες (αναλυτές δεδομένων, εμπειρογνώμονες Business Intelligence) πραγματοποιούν την ανάλυση των προετοιμασμένων δεδομένων.
4. Οπτικοποίηση / Παρουσίαση. Τα δεδομένα παρουσιάζονται στους εκάστοτε ενδιαφερόμενους στη μορφή γραφημάτων, πινάκων κλπ.

➤ Διαγνωστική Ανάλυση Δεδομένων

Η διαγνωστική ανάλυση παίρνει ως είσοδο τα αποτελέσματα που προκύπτουν από τις περιγραφικές αναλύσεις, με σκοπό να εντοπίσει την αιτία αυτού του αποτελέσματος. Σε αυτό το στάδιο λοιπόν, τα ιστορικά δεδομένα μπορούν να αντιπαραβληθούν σε σχέση με άλλα δεδομένα για να απαντήσουν στο ερώτημα γιατί συνέβη κάτι. Χάρη στις διαγνωστικές αναλύσεις, υπάρχει η δυνατότητα ανίχνευσης ανωμαλιών, εξεύρεσης εξαρτήσεων και ταυτοποίησης προτύπων. Οι εταιρείες προβαίνουν σε διαγνωστικές αναλύσεις, καθώς δίνουν σε βάθος γνώση για ένα συγκεκριμένο πρόβλημα, δημιουργώντας περισσότερες συσχετίσεις μεταξύ δεδομένων και εντοπίζοντας πρότυπα συμπεριφοράς. Η αναζήτηση της απάντησης συχνά περιλαμβάνει την συλλογή πληροφοριών από εξωτερικές πηγές. Όταν συλλεχθούν όλα τα απαραίτητα δεδομένα, οι αναλυτές καθορίζουν αιτιώδεις σχέσεις χρησιμοποιώντας διαφορετικές μεθόδους ανάλυσης δεδομένων (θεωρία πιθανότητας, ανάλυση παλινδρόμησης, φιλτράρισμα και άλλα) για να βρουν την απάντηση.

➤ Προγνωστική Ανάλυση Δεδομένων

Οι προγνωστικές αναλύσεις επιχειρούν να προβλέψουν τι είναι πιθανό να συμβεί. Οι αναλύσεις αυτές, χρησιμοποιούν τα ευρήματα των περιγραφικών και διαγνωστικών αναλύσεων για την ανίχνευση τάσεων, «συστάδων» γεγονότων με όμοια χαρακτηριστικά καθώς και αποκλίσεων από την αναμενόμενη έκβαση ενός γεγονότος, με στόχο την πρόβλεψη μελλοντικών τάσεων, γεγονός που τις καθιστά εξαιρετικά πολύτιμο εργαλείο για την επιχείρηση. Οι αναλύσεις αυτές βασίζονται κυρίως σε στατιστική μοντελοποίηση καθώς και σε πιο προηγμένες αριθμητικές μεθόδους, όπως είναι η Μηχανική Μάθηση και η Εξόρυξη Δεδομένων. Παρά τα πολυάριθμα πλεονεκτήματα που προσφέρει η προγνωστική ανάλυση, είναι σημαντικό να γίνει κατανοητό ότι η πρόβλεψη είναι απλώς μια εκτίμηση, η

ακρίβεια της οποίας εξαρτάται σε μεγάλο βαθμό από την ποιότητα των δεδομένων και τη σταθερότητα των συνθηκών εντός των οποίων παρατηρείται ένα γεγονός και συνεπώς απαιτείται προσεκτική επεξεργασία των δεδομένων και συνεχής ενασχόληση του εξειδικευμένου προσωπικού με τη βελτιστοποίηση των χρησιμοποιούμενων μοντέλων.

➤ Καθοδηγητική Ανάλυση Δεδομένων

Το σχετικά νέο πεδίο των καθοδηγητικών αναλύσεων επιτρέπει στους χρήστες να «συνταγογραφούν» έναν αριθμό διαφορετικών πιθανών δράσεων και να καθοδηγούνται προς μια λύση. Οι καθοδηγητικές αναλύσεις προσπαθούν να ποσοτικοποιήσουν την επίδραση των μελλοντικών αποφάσεων, προκειμένου να παρέχουν συμβουλές σχετικά με τα πιθανά αποτελέσματα πριν πραγματοποιηθούν οι αποφάσεις. Στην καλύτερη περίπτωση, οι καθοδηγητικές αναλύσεις προβλέπουν όχι μόνο τι θα συμβεί, αλλά και γιατί αυτό θα συμβεί παρέχοντας συστάσεις σχετικά με τις ενέργειες που πρέπει να γίνουν. Αυτές οι αναλύσεις υπερβαίνουν τις περιγραφικές και τις προγνωστικές αναλύσεις συνιστώντας μία ή περισσότερες πιθανές πορείες δράσης. Οι καθοδηγητικές αναλύσεις χρησιμοποιούν ένα συνδυασμό τεχνικών και εργαλείων, όπως αλγορίθμους μηχανικής μάθησης και αριθμητικές μεθόδους μοντελοποίησης. Αυτές οι τεχνικές εφαρμόζονται σε πολλά διαφορετικά σύνολα δεδομένων, συμπεριλαμβανομένων ιστορικών δεδομένων, δεδομένων σε πραγματικό χρόνο, και Big data.

## 2.3 Data Analytics στη Βιομηχανία

Ο βασικός παράγοντας ο οποίος οδηγεί στη βελτίωση της βιομηχανίας παραγωγής είναι η ανάγκη για ανταγωνιστικότητα και η ικανότητα αντιμετώπισης των απειλών και των ευκαιριών με ευέλικτο τρόπο. Οι βιομηχανικοί παραγωγοί βρίσκονται διαρκώς κάτω από συνεχή πίεση ώστε να χρησιμοποιούν πιο αποδοτικά τους πόρους τους. Σήμερα, όμως, η περαιτέρω βελτίωση στην παραγωγικότητά των βιομηχανιών φαντάζει δύσκολη, καθώς η εφαρμογή νέων καινοτόμων μεθόδων παραγωγής, η χρήση πληροφοριακών συστημάτων διαχείρισης των επιχειρησιακών λειτουργιών και η υιοθέτηση σύγχρονων τεχνικών διοίκησης της παραγωγής και διαχείρισης της ποιότητας, έχουν ήδη βελτιώσει κατά πολύ την απόδοση των βιομηχανικών παραγωγών.

Μια πιθανή απάντηση στην αναζήτηση πρακτικών με σκοπό την περαιτέρω βελτίωση των βιομηχανικών διαδικασιών, είναι η χρήση μεθόδων ανάλυσης δεδομένων. Οι οργανισμοί που δραστηριοποιούνται στον κλάδο της βιομηχανικής παραγωγής, παράγουν μέσω των διαδικασιών τους πολύ μεγάλο όγκο δεδομένων, ο οποίος όμως παραμένει στις περισσότερες περιπτώσεις ανεκμετάλλευτος. Ωστόσο, χάρη στο διαρκώς μειούμενο κόστος της υπολογιστικής ισχύος και στην ταχύτατη ανάπτυξη της επιστήμης στον κλάδο της ανάλυσης δεδομένων, οι βιομηχανικοί παραγωγοί μπορούν πλέον να αξιοποιήσουν αυτά τα δεδομένα, συλλέγοντας πληροφορίες από πολλαπλές πηγές δεδομένων και εκμεταλλευόμενοι τα μοντέλα μηχανικής μάθησης και τις πλατφόρμες απεικόνισης δεδομένων, να αποκαλύψουν νέους τρόπους βελτιστοποίησης των διαδικασιών τους, από την προμήθεια των πρώτων υλών έως την πώληση των τελικών προϊόντων τους.

Αν και κάθε περίπτωση είναι μοναδική και τα έχει τα δικά της ιδιαίτερα χαρακτηριστικά, είναι χρήσιμο να κατανοήσουμε το ευρύτερο πλαίσιο μέσω του οποίου μπορούν να χρησιμοποιηθούν οι μέθοδοι ανάλυσης δεδομένων στη βιομηχανία. Ορισμένοι τομείς οι οποίοι προσφέρονται για την εφαρμογή τέτοιων πρακτικών παρουσιάζονται στη συνέχεια.

➤ Προληπτική Συντήρηση:

Μία από τις ουσιαστικότερες συνεισφορές των μεθόδων ανάλυσης δεδομένων στη βιομηχανία, στην κατεύθυνση της βελτίωσης της παραγωγικότητας, είναι ο ρόλος που μπορούν να παίξουν στην προληπτική συντήρηση. Με δεδομένη την ευρύτατη χρήση των διαφόρων αισθητήρων και διασυνδεδεμένων συσκευών που περιλαμβάνονται στον σημερινό προηγμένο εξοπλισμό, οι βιομηχανικοί παραγωγοί μπορούν να χρησιμοποιήσουν προηγμένους αλγόριθμους ανάλυσης δεδομένων για να αποκαλύψουν τις πιθανές επιπλοκές στον εξοπλισμό, πριν αυτές εμφανιστούν και να διορθώσουν τα δευτερεύοντα ζητήματα προτού αυτά μετατραπούν σε δαπανηρά προβλήματα. Η προληπτική συντήρηση έχει τη δυνατότητα να εξοικονομήσει στους παραγωγούς εκατομμύρια δολάρια κατά τη διάρκεια ενός έτους, παρατείνοντας τη διάρκεια ζωής του εξοπλισμού και εξασφαλίζοντας την αποτελεσματική και απρόσκοπτη λειτουργία του. Χάρη στην ταχύτατη ανάπτυξη των διαφόρων εργαλείων διαχείρισης δεδομένων, καθίσταται ευκολότερη και πιο αποδοτική η συγκέντρωση και αξιοποίηση αυτών των στοιχείων.

➤ Ανάλυση Επιδόσεων:

Η λειτουργία του εξοπλισμού σε χαμηλότερη από την ονομαστική του δυναμικότητα καθιστά τη διαδικασία παραγωγής λιγότερο αποδοτική. Η εφαρμογή μεθόδων ανάλυσης δεδομένων συμβάλει στην ανάλυση των επιδόσεων μιας διαδικασίας και στην εφαρμογή αλλαγών και την υλοποίηση ρυθμίσεων με γνώμονα τον επιθυμητό όγκο παραγωγής. Προηγμένες αναλύσεις δεδομένων μπορούν να δώσουν στις εταιρείες μια συγκεκριμένη εικόνα για την επίδραση διαφορετικών μεταβλητών στη συνολική παραγωγικότητα

των παραγωγικών διεργασιών, να παρουσιάσουν πληροφορίες για τον πιο αποδοτικό τρόπο ελέγχου των παραγωγικών συστημάτων, δεδομένων των διαφόρων συστημικών περιορισμών(π.χ. Καλύτερη διαχείριση των bottlenecks στην παραγωγή) και να επισημάνουν τις μεγαλύτερες ευκαιρίες για βελτίωση της επίδοσης , λαμβάνοντας υπόψιν τις βασικές απώλειες του παραγωγικού συστήματος.

➤ Μείωση του χρόνου διακοπής της λειτουργίας (Downtime):

Λίγα πράγματα είναι πιο δαπανηρά για έναν κατασκευαστή από το χρόνο διακοπής της λειτουργίας. Σε ορισμένες βιομηχανίες, μπορεί να κοστίζει χιλιάδες δολάρια ανά λεπτό και εκατομμύρια δολάρια ετησίως. Με τη χρήση των κατάλληλων συστημάτων, τα οποία μπορούν να τροφοδοτούνται από δεδομένα και να τα αναλύουν σε πραγματικό χρόνο οι οργανώσεις αυτές μπορούν να μειώσουν σημαντικά το χρόνο διακοπής λειτουργίας και να εξασφαλίσουν τη μέγιστη παραγωγικότητα.

➤ Λήψη αποφάσεων:

Τελικά, η σημαντικότερη συνεισφορά της ανάλυσης δεδομένων εντοπίζεται στην βοήθεια που παρέχει τους παραγωγούς στη λήψη στρατηγικών αποφάσεων. Συγκεκριμένα, συμβάλει στη μείωση ή και στην εξάλειψη της ανάγκης για λήψη αποφάσεων βασιζόμενων σε αυθαίρετες παραδοχές και εικασίες και βοηθά στην στοιχειοθέτηση επιχειρημάτων για τη λήψη όσο το δυνατόν πιο εμπειρισταωμένων επιχειρηματικών αποφάσεων.



Η ιδέα ότι η ανάλυση δεδομένων αλλάζει δραστικά το τοπίο στην βιομηχανική παραγωγή είναι τόσο συναρπαστική όσο και γεμάτη προκλήσεις. Για εταιρείες που συλλέγουν δεδομένα για χρόνια, αλλά δεν τα έχουν ακόμα χρησιμοποιήσει, η ιδέα ότι θα χρειαστεί να αρχίσουν τελικά να δραστηριοποιούνται και στον εν λόγω τομέα είναι λίγο τρομακτική. Ωστόσο όταν υπερκεραστούν οι αρχικές προκλήσεις, διακρίνεται εύκολα η ύπαρξη μιας τεράστιας ευκαιρίας για ανάπτυξη και πρόοδο.

## 2.4 Data Analytics στην Παραγωγή σε Παρτίδες

Όπως είδαμε και προηγουμένως η εφαρμογή τεχνικών ανάλυσης δεδομένων στη βιομηχανία δύναται να προσφέρει σημαντικά οφέλη και υπολογίσιμη προστιθέμενη αξία στις λειτουργίες των οργανισμών που θα τις υιοθετήσουν. Συγκεκριμένα στην περίπτωση της βιομηχανικής παραγωγής σε παρτίδες, ωστόσο, συναντώνται σημαντικές προκλήσεις, οι οποίες καθιστούν την εφαρμογή τέτοιων τεχνικών ακόμα πιο επιτακτική.

Ιστορικά, οι διαδικασίες παραγωγής σε παρτίδες ήταν δύσκολο να ελεγχθούν και να αναλυθούν επειδή κάθε παρτίδα είναι μοναδική. Οι διάφορες παρτίδες δεν έχουν το ίδιο μέγεθος ή την ίδια διάρκεια, οι πρώτες ύλες και οι χρονικές υστερήσεις μεταξύ παρτίδων μπορεί να διαφέρουν ενώ υπάρχουν συχνά διαφορές στον εξοπλισμό, τις συνθήκες λειτουργίας και τις δραστηριότητες επεξεργασίας. Η κατανόηση του τρόπου με τον οποίο οι διαφορές αυτές επηρεάζουν την ποιότητα μιας παρτίδας, την ώρα που η παρτίδα αυτή βρίσκεται σε παραγωγή, μπορεί να προσφέρει τεράστια οφέλη.

Γίνεται λοιπόν φανερό ότι, κατά την εφαρμογή μεθόδων ανάλυσης δεδομένων σε διαδικασίες βιομηχανικής παραγωγής σε παρτίδες, θα πρέπει να ληφθούν σοβαρά υπόψιν οι ιδιαιτερότητες, οι οποίες ενυπάρχουν στις διαδικασίες αυτές και τις καθιστούν, μελέτη περίπτωσης με ξεχωριστά χαρακτηριστικά και αξία περαιτέρω ανάλυσης.

Ορισμένες από τις προαναφερθείσες ιδιαιτερότητες είναι οι εξής:

➤ Καθυστερήσεις διαδικασιών:

Διακοπές και επανεκκινήσεις της διαδικασίας που πραγματοποιούνται από τον χειριστή. Μερικές φορές, αυτές οι διακοπές και οι επανεκκινήσεις, μπορεί να είναι μέρος του σχεδιασμού της διαδικασίας παραγωγής κατά παρτίδες, όπως για παράδειγμα η προσθήκη κάποιου συστατικού. Άλλες φορές η εξέλιξη μιας παρτίδας μπορεί να καθυστερήσει λόγω των περιορισμών που επιβάλλονται από

την ανάγκη να περιμένουμε να γίνει ο κοινός εξοπλισμός διαθέσιμος. Ανεξαρτήτως από την πηγή των εν λόγω καθυστερήσεων, ο χρόνος ολοκλήρωσης μιας διαδικασίας παραγωγής παρτίδας μπορεί να διαφέρει μεταξύ παρτίδων, κάτι το οποίο επηρεάζει τον τρόπο με τον οποίο θα πρέπει να γίνεται η επεξεργασία των δεδομένων.

➤ Πρόσβαση σε εργαστηριακά δεδομένα:

Λόγω της φύσης του παραγόμενου προϊόντος με επεξεργασία σε παρτίδες, η online μέτρηση των παραμέτρων ποιότητας μπορεί να μην είναι τεχνικά εφικτή ή να μην δικαιολογείται από οικονομικής άποψης. Έτσι, είναι συνηθισμένη πρακτική η λήψη ενός δείγματος και η ανάλυση του στο εργαστήριο κατά τη διάρκεια διαφόρων σημείων στη διαδικασία παρτίδας. Σε πολλές εγκαταστάσεις σήμερα, τα δεδομένα του εργαστηρίου από τέτοια δείγματα παραμένουν μόνο εντός του εργαστηρίου. Τα εργαστηριακά αποτελέσματα κοινοποιούνται στον χειριστή της διαδικασίας μέσω τηλεφώνου ή μέσω κάποιου εργαστηριακού τερματικού που μπορεί να μην είναι συνδεδεμένο με το σύστημα ελέγχου. Προκειμένου να είναι δυνατή η εφαρμογή ανάλυσης δεδομένων σε πραγματικό χρόνο, τα εργαστηριακά δεδομένα θα πρέπει να είναι άμεσα διαθέσιμα στο αντίστοιχο εργαλείο.

➤ Αποκλίσεις στις πρώτες ύλες:

Η τροφοδότηση πρώτων υλών σε μια παρτίδα μπορεί να γίνει μέσω δεξαμενών αποθήκευσης των οποίων το περιεχόμενο αναπληρώνεται περιοδικά μέσω διαφόρων διαδικασιών μεταφοράς πρώτων υλών από εξωτερικούς προμηθευτές. Αλλαγές στις ιδιότητες των εισερχόμενων πρώτων υλών μπορούν να επηρεάσουν άμεσα τη διαδικασία παραγωγής σε παρτίδες και τα ποιοτικά χαρακτηριστικά του παραγόμενου προϊόντος. Παρόλο που ο προμηθευτής μπορεί να παρέχει τις ιδιότητες κάθε υλικού αποστολής, αυτά τα δεδομένα μπορεί να είναι διαθέσιμα μόνο στην αγορά ή στο εργαστήριο διασφάλισης ποιότητας. Εάν τα στοιχεία αυτά δεν είναι διαθέσιμα για χρήση στο εργαλείο ανάλυσης δεδομένων, τότε επηρεάζεται η ακρίβεια των αντίστοιχων προβλέψεων.

➤ Διαφορετικές συνθήκες λειτουργίας:

Η παραγωγή μιας παρτίδας μπορεί να διαιρεθεί σε πολλαπλές λειτουργίες. Οι συνθήκες επεξεργασίας μπορεί να διαφέρουν σημαντικά μεταξύ των διαφόρων επιμέρους λειτουργιών καθώς και μεταξύ των διαφορετικών τελικών προϊόντων. Έτσι, το μοντέλο ανάλυσης δεδομένων που εφαρμόζεται σε μια παρτίδα θα πρέπει να

λαμβάνει υπόψη το είδος του προϊόντος που παράγεται καθώς και τον τύπο της λειτουργίας που είναι ενεργή στη μονάδα επεξεργασίας.

➤ Σύγχρονες παρτίδες.

Τυπικά, οι περισσότερες εταιρείες θέτουν σε ταυτόχρονη παραγωγή πολλές παρτίδες στην ίδια επιμέρους λειτουργία. Μόλις ένα κομμάτι του εξοπλισμού καταστεί διαθέσιμο, μια νέα παρτίδα ξεκινάει. Οι εταιρείες το κάνουν αυτό προκειμένου να μεγιστοποιήσουν τη διακίνηση πρώτων υλών και προϊόντων. Έτσι, πολλαπλές παρτίδες του ίδιου υλικού μπορεί να βρεθούν εντός της μονάδας επεξεργασίας σε διαφορετικά στάδια ολοκλήρωσης μεταξύ τους. Η συλλογή δεδομένων και τα αντίστοιχα εργαλεία ανάλυσης τους, πρέπει να λειτουργούν αποτελεσματικά σε αυτό το περιβάλλον έτσι ώστε κάθε παρτίδα να αναλύεται και να μπορεί να επιθεωρηθεί από τον χειριστή.

Δεδομένων των ανωτέρω προκλήσεων, η πλειοψηφία των αναλύσεων περιορίζεται στην αξιολόγηση μιας διαδικασίας παραγωγής σε παρτίδες με βάση ένα πρότυπο που έχει οριστεί από την επονομαζόμενη ως "χρυσή" ή ιδανική παρτίδα. Τυπικά, μια χρυσή παρτίδα ορίζεται ως το χρονικό προφίλ των τιμών μέτρησης που καταγράφηκαν για μια συγκεκριμένη παρτίδα που πληρούσε όλους τους στόχους ποιότητας του προϊόντος. Όταν χρησιμοποιείται αυτό το πρότυπο, μια παρτίδα κρίνεται από το πόσο πιστά ακολουθείται το προφίλ της χρυσής παρτίδας μέσω της ρύθμισης των εισόδων της διεργασίας. Η εφαρμογή μιας τέτοιας συγκριτικής μεθόδου είναι πολύ εύκολη, μαστίζεται ωστόσο, από εγγενή προβλήματα. Η προσέγγιση αυτή έχει δύο σημαντικά μειονεκτήματα. Πρώτον, οι συνθήκες που υποδεικνύονται από κάθε μέτρηση ενδέχεται να επηρεάσουν την ποιότητα του προϊόντος με διαφορετικό τρόπο. Για παράδειγμα, μπορεί να είναι σημαντικός ο αυστηρός έλεγχος κάποιων παραμέτρων, ενώ κάποιες άλλες μετρήσεις μπορεί να διαφέρουν σημαντικά μεταξύ τους, χωρίς να επηρεάζεται η ποιότητα του προϊόντος. Δεύτερον, η μέθοδος της "χρυσής παρτίδας" είναι μια μονομεταβλητή προσέγγιση σε ένα πολυμεταβλητό πρόβλημα. Συνεπώς, δεν αποκτάται καμία απολύτως γνώση ως προς τις σχέσεις μεταξύ των διαφόρων παραλλαγών της διαδικασίας. Η μέθοδος αυτή συνίσταται ουσιαστικά στην στείρα μίμηση μιας παρτίδας χωρίς να υπάρχει κατανόηση του γιατί ή του πότε κάτι τέτοιο είναι αποτελεσματικό. Έτσι όταν δεν λαμβάνεται υπόψιν ο πολυμεταβλητός χαρακτήρας μιας διαδικασίας παραγωγής σε παρτίδες, είναι δυνατόν να οδηγηθούμε

σε λανθασμένες στρατηγικές ελέγχου και σε αναποτελεσματική επένδυση πόρων. Χρόνος και χρήματα μπορεί να δαπανηθούν κάπου όπου δεν είναι απαραίτητο και να απομακρυνθεί η προσοχή από τις πραγματικά κρίσιμες πτυχές της διαδικασίας. Αντιθέτως, μέσω της χρήσης πολυμεταβλητών στατιστικών τεχνικών, είναι δυνατόν να αποκωδικοποιούνται οι διάφορες μεταβολές τόσο εντός μιας παρτίδας, όσο και μεταξύ διαφορετικών παρτίδων, να διαπιστώνονται συσχετίσεις μεταξύ των μεταβλητών της διαδικασίας, να εντοπίζονται και να διαγιγνώσκονται σημαντικά γεγονότα κατά τη διάρκεια επεξεργασίας, να προβλέπονται και να προλαμβάνονται σφάλματα και τελικά να ελέγχονται τα ποιοτικά χαρακτηριστικά του τελικού προϊόντος.

Συνοψίζοντας, ο στόχος κάθε διαδικασίας παραγωγής σε παρτίδες είναι να παράγει με ασφάλεια και αξιοπιστία, την μέγιστη δυνατή ποσότητα τελικού προϊόντος που ικανοποιεί τις προδιαγραφές ποιότητας στο συντομότερο δυνατό χρονικό διάστημα. Η μη τήρηση οποιουδήποτε μέρους αυτού του στόχου μπορεί να έχει ως αποτέλεσμα την αύξηση του κόστους. Εφαρμόζοντας, ωστόσο, μια πολύ-παραμετρική ανάλυση της διαδικασίας, με τη χρήση εξειδικευμένων αλγορίθμων και μοντέλων ανάλυσης δεδομένων που παράγονται από μια συλλογή δεδομένων από ιστορικές παρτίδες, μπορούμε να εξασφαλίσουμε την επίτευξη του προαναφερθέντος στόχου. Συγκεκριμένα, ορισμένα από τα οφέλη εφαρμογής μιας τέτοιας ανάλυσης συνοψίζονται ως εξής:

➤ Πρόβλεψη ποιότητας παρτίδας:

Η εφαρμογή μιας πολυμεταβλητής ανάλυσης δεδομένων σε πραγματικό χρόνο δύναται να παρέχει ανά πάσα στιγμή μια πρόβλεψη των ποιοτικών χαρακτηριστικών της παραγόμενης παρτίδας, επιτρέποντας με αυτόν τον τρόπο στους χειριστές να πραγματοποιούν προσαρμογές, που κρίνονται απαραίτητες, προκειμένου να διατηρείται η ποιότητα της παρτίδας εντός των επιτρεπτών ορίων και ανοχών. Η προβλεψιμότητα μπορεί να εξοικονομήσει πολύτιμους πόρους και να συμβάλει καθοριστικά στην αύξηση της παραγωγικότητας.

➤ Ανίχνευση σφαλμάτων:

Μέσω μιας πολυμεταβλητής ανάλυσης της διαδικασίας σε πραγματικό χρόνο εντοπίζονται τυχόν παρεκκλίσεις από τη φυσιολογική εξέλιξη της διαδικασίας. Η ανίχνευση σφαλμάτων σε πραγματικό χρόνο επιτρέπει στους χειριστές να εντοπίζουν τους πιθανούς λόγους απόκλισης της διαδικασίας και να εστιάζουν την προσοχή τους στις πηγές παραγωγής συστηματικών σφαλμάτων. Μια προβολή στο

χρόνο της εξέλιξης της διαδικασίας, υποδεικνύει πότε μια παρτίδα μπορεί να εμφανίσει κάποιο πρόβλημα, ανακουφίζοντας τους χειριστές από την ανάγκη να παρακολουθούν όλα τα δεδομένα διαρκώς. Αυτή η έγκαιρη, στοχευμένη και άμεσα αξιοποιήσιμη πληροφορία μπορεί να σημαίνει τη διαφορά μεταξύ ενός προϊόντος που είναι εντός και ενός προϊόντος που είναι εκτός προδιαγραφών. Η αξιοποίηση των πληροφοριών αυτών θέτει τα θεμέλια για την μεγιστοποίηση της απόδοσης της διαδικασίας.

➤ Αυξημένη κατανόηση της διαδικασίας:

Η οπτικοποίηση των συσχετιζόμενων δεδομένων της διαδικασίας σε πραγματικό χρόνο βοηθά τους χειριστές να κατανοήσουν τις αλληλεπιδράσεις και τις συσχετίσεις μεταξύ των επιμέρους διαδικασιών και μεταβλητών, καθώς και την επίδρασή τους στην ποιότητα των τελικών προϊόντων. Καθίσταται έτσι δυνατή η λήψη καλύτερων αποφάσεων μέσω της κατανόησης της επίδρασης των μεταβλητών εισόδου στις μεταβλητές εξόδου.

➤ Ελαχιστοποίηση της μεταβλητότητας:

Η καλύτερη κατανόηση και η εξοικείωση με τις αλληλεπιδράσεις μεταξύ των παραμέτρων της διαδικασίας, που προκύπτει από την ανάλυση των δεδομένων, καθιστά δυνατές τις στοχευμένες προσαρμογές και το «fine tuning» της διαδικασίας ώστε να εξαλειφθούν οι διαφορές μεταξύ των παρτίδων.



## ΚΕΦΑΛΑΙΟ 3: *Μέθοδοι και Τεχνικές*

---

### 3.1 Γενικά

Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, υπάρχουν διάφορα είδη ανάλυσης, τα οποία μπορεί να εφαρμόσει μια επιχείρηση για να αξιοποιήσει πολύτιμα δεδομένα. Κάθε τύπος ανάλυσης δεδομένων θα έχει διαφορετικό αποτέλεσμα ή αντίκτυπο. Ο τύπος της ανάλυσης δεδομένων που θα πρέπει να χρησιμοποιηθεί στην εκάστοτε περίπτωση, εξαρτάται από το είδος του επιχειρηματικού προβλήματος που πρέπει να αντιμετωπιστεί. Οι διαφορετικές τεχνικές ανάλυσης δεδομένων θα μπορούσαν να οδηγήσουν σε διαφορετικά αποτελέσματα και συμπεράσματα, προσφέροντας έτσι διαφορετικές γνώσεις για την επιχείρηση. Ο βασικότερος στόχος οποιασδήποτε διεργασίας της ανάλυσης των δεδομένων είναι η αναζήτηση πληροφοριών, εντός συνόλων δεδομένων μεγάλης κλίμακας, οι οποίες μπορούν να οδηγήσουν σε εύληπτα συμπεράσματα για τα σύνολα αυτά.

Η βασικότερη διάκριση μεταξύ των διαφόρων μεθόδων και τεχνικών ανάλυσης δεδομένων είναι αυτή που τις κατηγοριοποιεί σε ποσοτικές και ποιοτικές μεθόδους. Όπως υποδηλώνει το όνομα, οι ποσοτικές μέθοδοι ανάλυσης δεδομένων αφορούν τις μετρήσιμες ποσότητες και τους «σκληρούς» αριθμούς, δηλαδή, ποσοτικά δεδομένα, τα οποία μπορούν να μετρηθούν αντικειμενικά. Έτσι, οι μέθοδοι ποσοτικής ανάλυσης βασίζονται στην ικανότητα ακριβούς καταμέτρησης και ερμηνείας δεδομένων βάσει γεγονότων. Σε αντιδιαστολή με τα ποσοτικά δεδομένα, τα ποιοτικά αφορούν τις πτυχές ενός προβλήματος που είναι πιο υποκειμενικές, όπως είναι για παράδειγμα πληροφορίες που προέρχονται από έρευνες πελατών και συνεντεύξεις με υπαλλήλους. Ως εκ τούτου, οι χρησιμοποιούμενες μέθοδοι ανάλυσης απομακρύνονται από τη λογική των αμιγώς στατιστικών και αριθμητικών μεθόδων και είναι λιγότερο δομημένες από τις ποσοτικές τεχνικές.

Στην παρούσα εργασία θα εστιάσουμε σε ποσοτικές μεθόδους ανάλυσης δεδομένων. Οι ποσοτικές μέθοδοι ανάλυσης δεδομένων βασίζονται κατά κύριο λόγο σε μαθηματικές και στατιστικές τεχνικές, σε τεχνικές που βασίζονται στην τεχνητή νοημοσύνη και την μηχανική μάθηση καθώς και σε τεχνικές οπτικοποίησης δεδομένων. Ορισμένες από τις σημαντικότερες κατηγορίες ποσοτικών μεθόδων ανάλυσης δεδομένων παρουσιάζονται ακολούθως.

➤ Περιγραφική Ανάλυση:

Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, η περιγραφική ανάλυση αποτελεί μια ματιά στο παρελθόν. Αυτή η στατιστική τεχνική κάνει ακριβώς αυτό που υποδηλώνει το όνομα της – «Περιγράφει». Εξετάζει τα δεδομένα και αναλύει παρελθόντα γεγονότα και καταστάσεις προκειμένου να προτείνει πιθανούς τρόπους με τους οποίους μπορούμε να προσεγγίσουμε το μέλλον. Οι περιγραφικές αναλύσεις εξετάζουν την παρελθούσα / ιστορική απόδοση για να κατανοήσουν τους λόγους πίσω από την αποτυχία ή την επιτυχία και μας επιτρέπουν να μαθαίνουμε από τις συμπεριφορές του παρελθόντος και πως αυτές θα μπορούσαν να επηρεάσουν τη μελλοντική απόδοση.

➤ Ανάλυση Παλινδρόμησης (Regression Analysis):

Η ανάλυση παλινδρόμησης επιτρέπει τη μοντελοποίηση της σχέσης μεταξύ μιας εξαρτημένης μεταβλητής και μιας ή περισσοτέρων ανεξάρτητων μεταβλητών. Οι διάφορες τεχνικές παλινδρόμησης χρησιμοποιούνται για την πρόβλεψη των τιμών, λαμβάνοντας υπόψη ένα συγκεκριμένο σύνολο δεδομένων και επισημαίνουν τις τάσεις και συσχετίσεις μεταξύ διαφόρων παραγόντων, συμβάλλοντας έτσι στην αποκάλυψη των πτυχών εκείνων των λειτουργιών μιας επιχείρησης, οι οποίες μπορούν να βελτιστοποιηθούν. Υπάρχει μια τεράστια ποικιλία διαφορετικών τύπων μοντέλων παλινδρόμησης, όπως μοντέλα γραμμικής παλινδρόμησης, πολλαπλή παλινδρόμηση, λογική παλινδρόμηση, μη γραμμική παλινδρόμηση κ.α..

➤ Ανάλυση Παραγόντων (Factor Analysis):

Η ανάλυση παραγόντων είναι μια τεχνική ανάλυσης δεδομένων βασισμένη σε παλινδρόμηση, η οποία χρησιμοποιείται για την εύρεση μιας βαθύτερης υποκείμενης δομής σε ένα σύνολο μεταβλητών. Έγκειται στην εύρεση νέων ανεξάρτητων παραγόντων (μεταβλητών) που περιγράφουν τα πρότυπα και τα μοντέλα σχέσεων μεταξύ αρχικών εξαρτημένων μεταβλητών, μειώνει τις διαστάσεις ενός πολυπαραγοντικού προβλήματος και αποτελεί ένα βασικό βήμα προς την αποτελεσματική ομαδοποίηση και ταξινόμηση.

➤ Ανάλυση Διασποράς (Dispersion Analysis):

Η ανάλυση διασποράς εξετάζει το εύρος στο οποίο εκτείνεται ένα σύνολο δεδομένων. Είναι συνεπώς μια τεχνική που περιγράφει πόσο εκτεταμένο είναι ένα σύνολο δεδομένων. Το μέτρο διασποράς βοηθά τους επιστήμονες δεδομένων να μελετήσουν τη μεταβλητότητα των διαφόρων μετρούμενων μεγεθών. Γενικά, η



διασπορά αντιπροσωπεύει τη μεταβολή των μεγεθών μεταξύ τους, καθώς και τη μεταβολή τους γύρω από τη μέση τιμή. Εάν η διαφορά μεταξύ της τιμής και του μέσου όρου είναι σημαντική, τότε η διασπορά είναι υψηλή. Διαφορετικά, είναι χαμηλή.

➤ Ανάλυση Ταξινόμησης-Ομαδοποίησης (Discriminant - Classification Analysis):

Η ανάλυση ταξινόμησης έγκειται στην ανίχνευση συνόλων δεδομένων με κοινά χαρακτηριστικά και χρησιμοποιεί μετρήσεις μεταβλητών σε διαφορετικές ομάδες αντικειμένων για να υπογραμμίσει σημεία που διακρίνουν και χαρακτηρίζουν τις ομάδες αυτές. Αυτές οι μετρήσεις χρησιμοποιούνται για την ταξινόμηση νέων στοιχείων στις αντίστοιχες ομάδες.

➤ Ανάλυση Χρονοσειρών (Time Series Analysis):

Η ανάλυση δεδομένων χρονοσειρών είναι η διαδικασία μοντελοποίησης και εξήγησης χρονικά εξαρτώμενων μεταβλητών. Ο στόχος είναι να αντληθούν όλες οι σημαντικές πληροφορίες, στατιστικά στοιχεία, κανόνες και πρότυπα από τη μορφή των δεδομένων. Οι πληροφορίες που εξάγονται χρησιμοποιούνται για την πρόβλεψη των μελλοντικών τιμών των εξεταζόμενων μεγεθών.

➤ Ανάλυση Υποθέσεων (Hypothesis Analysis):

Επίσης γνωστή ως "T Testing", αυτή η μέθοδος ανάλυσης επιτρέπει τη σύγκριση των διαθέσιμων δεδομένων έναντι υποθέσεων και παραδοχών που έχουν γίνει σχετικά με τις δραστηριότητες μιας επιχείρησης. Βοηθά επίσης στην πρόβλεψη του αντίκτυπου των αποφάσεων που μπορούν να ληφθούν. Η ανάλυση υποθέσεων επιτρέπει τη σύγκριση δύο μεταβλητών για την ανεύρεση συσχετίσεων και την στήριξη της λήψης αποφάσεων.

➤ Ανάλυση Monte Carlo (Monte Carlo Simulation):

Ως ένας από τους πιο δημοφιλείς τρόπους για τον υπολογισμό του αποτελέσματος των τυχαίων μεταβλητών σε ένα συγκεκριμένο παράγοντα, οι προσομοιώσεις Monte Carlo χρησιμοποιούν μοντέλα πιθανότητας για να προβλέψουν τον κίνδυνο και την αβεβαιότητα. Για τη δοκιμή μιας υπόθεσης ή ενός σεναρίου, μια προσομοίωση Monte Carlo θα χρησιμοποιήσει τυχαίους αριθμούς και δεδομένα για να οργανώσει μια ποικιλία πιθανών αποτελεσμάτων σε οποιαδήποτε κατάσταση με βάση οποιαδήποτε αποτελέσματα. Με τη δοκιμή μιας

ποικιλίας δυνατών εκβάσεων, γίνεται κατανοητή η επίδραση των τυχαίων μεταβλητών στο πρόβλημα.

Στην παρούσα εργασία, στόχος μας, όπως έχει ήδη αναφερθεί, είναι η ανάλυση των δεδομένων που προκύπτουν από την παρακολούθηση μιας διαδικασίας παραγωγής σε παρτίδες. Η ανάλυση που θα εφαρμοστεί θα πρέπει λάβει υπόψιν της, την πολυπαραγοντική φύση του προβλήματος της παραγωγής σε παρτίδες καθώς και τα διαθέσιμα ιστορικά στοιχεία, προκειμένου να είναι σε θέση να διακρίνει μεταξύ καλών και κακών παρτίδων, να φέρει στο φως τους λόγους, οι οποίοι μπορεί να οδηγήσουν στην αποτυχία μιας παρτίδας και να μπορεί να προβλέψει σε πραγματικό χρόνο την εξέλιξη των ποιοτικών χαρακτηριστικών μιας παρτίδας. Για το σκοπό αυτό επιλέγονται μέθοδοι, οι οποίες χρησιμοποιούνται κατά κόρον στη βιομηχανία παραγωγής σε παρτίδες και εμπίπτουν τόσο στην ανάλυση παλινδρόμησης όσο και στην ανάλυση παραγόντων και ταξινόμησης. Στις ακόλουθες ενότητες παρουσιάζεται το θεωρητικό υπόβαθρο που είναι απαραίτητο για την κατανόηση των μεθόδων αυτών.

## **3.2 Πολλαπλή Γραμμική Παλινδρόμηση - MLR**

### **3.2.1 Γενικά**

Η Παλινδρόμηση αναφέρεται στην ανάλυση ενός πλήθους μεταβλητών αναφορικά με τις ειδικές σχέσεις που αναπτύσσονται μεταξύ τους και αποτελεί μία από τις σημαντικότερες στατιστικές μεθόδους. Πολλά φαινόμενα των σύγχρονων επιστημών μοντελοποιούνται μέσω αυτής. Σε προβλήματα οικονομικής, βιολογικής φύσεως καθώς και σε άλλες επιστήμες, θεωρείται απαραίτητος ο καθορισμός της σχέσης μεταξύ μεταβλητών, δηλαδή ο σχεδιασμός ενός μοντέλου, μέσω του οποίου ερμηνεύεται ικανοποιητικά το εκάστοτε δοθέν πρόβλημα. Το μοντέλο αυτό, στην περίπτωση της γραμμικής παλινδρόμησης, προϋποθέτει τη θεώρηση μιας εξαρτημένης μεταβλητής  $Y$ , της οποίας η μεταβολή προκύπτει ως το γραμμικό αποτέλεσμα της μεταβολής ενός πλήθους ανεξάρτητων μεταβλητών  $X_i$ . Έτσι, δημιουργείται ένα μοντέλο που συσχετίζει τα διαθέσιμα δεδομένα (παρατηρήσεις) μεταξύ τους και μέσω αυτού καθίσταται δυνατή η πρόβλεψη μελλοντικών τιμών του  $Y$ , δεδομένων των μελλοντικών τιμών των ανεξάρτητων μεταβλητών  $X_i$ . Η συγκεκριμένη διαδικασία ονομάζεται Γραμμική Παλινδρόμηση και ο σκοπός της είναι ο καθορισμός του βέλτιστου για τα εκάστοτε δεδομένα, μοντέλου.

### 3.2.2 Μαθηματικό Μοντέλο

Το μαθηματικό μοντέλο της πολλαπλής γραμμικής παλινδρόμησης αποτελεί μια γενίκευση του απλού γραμμικού μοντέλου, όπου εξετάζεται η γραμμική σχέση μεταξύ μιας ανεξάρτητης ( $X$ ) και μιας εξαρτημένης μεταβλητής ( $Y$ ) και το οποίο εκφράζεται ως:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \varepsilon \quad ,$$

όπου σε ένα δείγμα  $n$  παρατηρήσεων οι τιμές των  $X$  και  $Y$  προκύπτουν ως:

$$Y_i = \beta_0 + \beta_1 \cdot X_{i1} + \varepsilon_i \quad , \quad i = 1, 2, \dots, n$$

Κατ' αντιστοιχία, το γενικευμένο μοντέλο για την περίπτωση της πολλαπλής γραμμικής παλινδρόμησης με  $k$  το πλήθος ανεξάρτητες μεταβλητές, προκύπτει ως:

$$Y_i = \beta_0 + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \dots + \beta_k \cdot X_{ik} + \varepsilon_i \quad , \quad i = 1, 2, \dots, n \quad ,$$

όπου:

- $X_{ij}$  είναι η τιμή της ανεξάρτητης μεταβλητής  $X_j$  για την παρατήρηση  $i$ , με  $j=1,2,\dots,k$
- $Y_i$  είναι η τιμή της εξαρτημένης μεταβλητής  $Y$  για την παρατήρηση  $i$
- $\beta_j$  είναι ο συντελεστής παλινδρόμησης που περιγράφει την επίδραση της μεταβλητής  $X_j$  στην μεταβλητή  $Y$ , με  $j=1,2,\dots,k$
- $\beta_0$  είναι ο σταθερός όρος
- $\varepsilon_i$  είναι το σφάλμα, δηλαδή η διαφορά της πραγματικής τιμής της μεταβλητής  $Y$ , από την τιμή που προβλέπεται από το μοντέλο για την παρατήρηση  $i$ .

Επομένως το παραπάνω μοντέλο μπορεί να γραφεί με τη μορφή πινάκων ως:

$$Y = X\beta + \varepsilon, \text{ όπου } Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & \dots & X_{1k} \\ 1 & X_{21} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{nk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \text{ και } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Το ζητούμενο είναι ο προσδιορισμός εκείνου του διανύσματος των συντελεστών παλινδρόμησης  $\beta$ , το οποίο ελαχιστοποιεί τα αντίστοιχα σφάλματα που συνθέτουν το διάνυσμα  $\varepsilon$ . Οι συντελεστές αυτοί υπολογίζονται με τη μέθοδο των ελαχίστων τετραγώνων, η οποία ελαχιστοποιεί το άθροισμα των τετραγώνων των σφαλμάτων ως:

$$\min \left( \sum_{i=1}^n \varepsilon_i^2 \right), \quad \text{όπου } \varepsilon_i = Y_i - \hat{Y}_i$$

θεωρώντας ότι  $Y_i$  είναι η πραγματική τιμή της εξαρτημένης μεταβλητής και  $\hat{Y}_i$  η τιμή που προβλέπεται από το μοντέλο της παλινδρόμησης.

Από την επίλυση του συστήματος εξισώσεων που προκύπτει, καταλήγουμε στην εκτιμήτρια του διανύσματος των συντελεστών παλινδρόμησης:

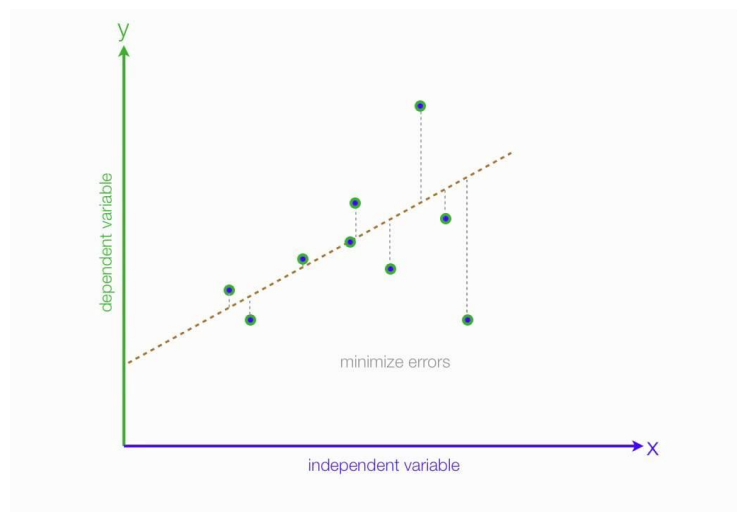
$$\beta = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

Στα προαναφερθέντα θα πρέπει να ληφθούν υπόψιν οι εξής παραδοχές:

1. Τα σφάλματα  $\varepsilon_i$  είναι ανεξάρτητες τυχαίες μεταβλητές και ακολουθούν την Κανονική Κατανομή
2. Οι αναμενόμενες τιμές των σφαλμάτων (μέσοι όροι) είναι μηδέν
3. Οι ανεξάρτητες μεταβλητές  $X$  είναι μεταξύ τους γραμμικώς ανεξάρτητες
4. Τα σφάλματα που αντιστοιχούν σε διαφορετικές επαναλήψεις του προβλήματος θεωρούνται μεταξύ τους ασυσχέτιστα.

### 3.2.3 Ερμηνεία

Η γεωμετρική ερμηνεία του μοντέλου της απλής γραμμικής παλινδρόμησης συνοψίζεται στην εύρεση μιας ευθείας, η οποία ελαχιστοποιεί το άθροισμα των τετραγώνων των αποστάσεων των παρατηρήσεων της εξαρτημένης μεταβλητής, από αυτή.



Εικόνα 3-1 Γεωμετρική ερμηνεία απλής παλινδρόμησης

Κατ' αυτή την έννοια, υπολογίζεται η ευθεία εκείνη, η οποία «ταιριάζει» στις παρατηρήσεις με τον καλύτερο δυνατό τρόπο και συνεπώς ερμηνεύει ικανοποιητικά, τη γραμμική σχέση που συνδέει την ανεξάρτητη με την εξαρτημένη μεταβλητή.

Στην γενική περίπτωση της πολλαπλής παλινδρόμησης, αναζητούνται οι επιμέρους γραμμικές σχέσεις μεταξύ της εξαρτημένης μεταβλητής και των διαφόρων ανεξάρτητων μεταβλητών. Συγκεκριμένα αναζητείται η εκάστοτε γραμμική σχέση που χαρακτηρίζει τη μεταβολή της εξαρτημένης μεταβλητής  $Y$ , όταν μεταβάλλεται η ανεξάρτητη μεταβλητή  $X_j$ , με δεδομένο ότι οι υπόλοιπες ανεξάρτητες μεταβλητές διατηρούνται σταθερές.

Έτσι, ο εκάστοτε συντελεστής παλινδρόμησης  $\beta_j$ , εκφράζει το πόσες μονάδες θα μεταβληθεί η εξαρτημένη μεταβλητή  $Y$ , δεδομένης, πρώτον, της μεταβολής της ανεξάρτητης μεταβλητής  $X_j$  κατά μια μονάδα και δεύτερον, δεδομένης της διατήρησης των υπολοίπων ανεξάρτητων μεταβλητών, σταθερών.

### 3.2.4 Εφαρμογές

Η μέθοδος της παλινδρόμησης βρίσκει εφαρμογή σε μεγάλο εύρος επιστημών, όπως η βιολογία, οι περιβαλλοντικές επιστήμες, η επιδημιολογία, η κοινωνιολογία, τα οικονομικά και η τεχνητή νοημοσύνη. Ενδεικτικά αναφέρονται, η πρόβλεψη δημοσιονομικών μεγεθών, η πρόβλεψη ζήτησης προϊόντων, η πρόβλεψη μακροοικονομικών μεγεθών, η μελέτη της επίδρασης των ανθρωπίνων δραστηριοτήτων στο περιβάλλον, η εξέλιξη της τιμής ενός αγαθού και γενικώς, η ανάλυση χρονοσειρών.

### 3.2.5 Πλεονεκτήματα και Μειονεκτήματα

Τα μεγαλύτερα πλεονεκτήματα στη χρήση γραμμικής παλινδρόμησης είναι η απλότητα της, η ευκολία εξήγησης του αποτελέσματος της εφαρμογής της, σε έναν άλλο άνθρωπο καθώς και η δυνατότητα καλύτερης κατανόησης του εξεταζόμενου συστήματος, την οποία μας προσφέρει, μέσω της ερμηνείας των συντελεστών παλινδρόμησης και της αξιολόγησης της βαρύτητάς τους. Αποτελεί επιπλέον, μια γρήγορη και υπολογιστικά μη απαιτητική μέθοδο, σε σύγκριση με πολλούς άλλους αλγόριθμους. Επιπλέον, επειδή είναι τόσο διαδεδομένη, ενσωματώνεται σχεδόν σε οποιοδήποτε υπολογιστικό λογισμικό ή γλώσσα προγραμματισμού.

Υπάρχουν επίσης πολλά μειονεκτήματα στην προσέγγιση αυτή. Το βασικότερο μειονέκτημα έγκειται στην παραδοχή ότι οι σχέσεις που διέπουν τις μεταβολές των εξεταζόμενων μεταβλητών είναι γραμμικές. Εάν υπάρχουν μη γραμμικές σχέσεις, οι προβλέψεις θα είναι αναξιόπιστες. Επιπλέον, οι αποκλίσεις τείνουν να έχουν μεγάλη επίδραση στο τελικό αποτέλεσμα. Τα περισσότερα συστήματα που καλούμαστε να εξετάσουμε, δεν πληρούν τις προϋποθέσεις που απαιτούνται προκειμένου η εφαρμογή της παλινδρόμησης να είναι αξιόπιστη. Ωστόσο, επειδή είναι τόσο απλό να εφαρμοστεί, χρησιμοποιείται συχνά ως σημείο αναφοράς για την αξιολόγηση της αποτελεσματικότητας και τη σύγκριση άλλων, πιο σύνθετων αλγορίθμων. Τέλος, αξίζει να σημειωθεί ότι, η εν λόγω μέθοδος τείνει να παράγει αξιόπιστα αποτελέσματα για μακροχρόνιο ορίζοντα πρόβλεψης, ενώ σε γενικές γραμμές δεν προτιμάται η χρήση της για την παραγωγή βραχυπρόθεσμων προβλέψεων.

### 3.3 Ανάλυση Κυρίων Συνιστωσών - PCA

#### 3.3.1 Γενικά

Στην πλειοψηφία των περιπτώσεων, το σημείο αφετηρίας μιας πολυμεταβλητής ανάλυσης δεδομένων είναι η προσπάθεια εξορθολογισμού ενός ευρύτατου συνόλου δεδομένων, το οποίο αποτελείται από ένα μεγάλο πλήθος παρατηρήσεων και μετρήσεων πολλών και διαφορετικών μεταβλητών, οι οποίες εκφράζουν τα βασικότερα χαρακτηριστικά του προς μελέτη φαινομένου. Το πλήθος των μεταβλητών αυτών καθορίζει τον αριθμό των διαστάσεων του προβλήματος που καλούμαστε να αντιμετωπίσουμε. Ένας από τους κυριότερους στόχους της διαδικασίας εξορθολογισμού του συνόλου των δεδομένων, είναι η ελαχιστοποίηση των διαστάσεων του προβλήματος, ώστε αυτό να καταστεί λιγότερο πολύπλοκο και απαιτητικό σε υπολογιστικούς πόρους. Η διαδικασία αυτή της ελαχιστοποίησης των διαστάσεων του προβλήματος έγκειται αφενός, στον εντοπισμό εκείνων των μεταβλητών, οι οποίες έχουν την ισχυρότερη επίδραση στην έκβαση του εκάστοτε πειράματος που πραγματοποιούμε και αφετέρου, στον εντοπισμό συσχετίσεων μεταξύ των μεταβλητών του προβλήματος και στην αντικατάσταση των συσχετιζόμενων αυτών μεταβλητών από νέες μεταβλητές, οι οποίες συνοψίζουν την επίδραση των επιμέρους συσχετιζόμενων μεταβλητών.

Πολύτιμο εργαλείο για την επίτευξη των προαναφερθέντων στόχων είναι η μέθοδος της Ανάλυσης Κυρίων Συνιστωσών. Πρόκειται για μία στατιστική διαδικασία

η οποία μετατρέπει μία ομάδα τιμών (παρατηρήσεων) δυνητικά συσχετιζόμενων μεταβλητών σε μία ομάδα νέων τιμών, μη γραμμικά συσχετιζόμενων μεταβλητών οι οποίες καλούνται κύριες συνιστώσες. Ο αριθμός των νέων μεταβλητών που προκύπτει είναι ίσος ή και συχνότερα πολύ μικρότερος από τον αριθμό των αρχικών μεταβλητών. Η μετάβαση αυτή, από το αρχικό στο νέο σύνολο μεταβλητών, πραγματοποιείται με τέτοιο τρόπο ώστε, η πρώτη συνιστώσα να ερμηνεύει τη μέγιστη δυνατή διακύμανση που αναπτύσσεται μεταξύ των αρχικών μεταβλητών, η δεύτερη, μη συσχετιζόμενη με την πρώτη, να εξηγεί ένα σημαντικό μέρος αυτής αλλά πάντα μικρότερο της πρώτης κοκ. Με άλλα λόγια, η Ανάλυση Κυρίων Συνιστωσών, επιδιώκει την ανεύρεση, από ένα πλήθος  $p$  μεταβλητών, ορισμένων νέων ολιγάριθμων μεταβλητών, οι οποίες έχουν την ιδιότητα να είναι γραμμικοί συνδυασμοί των αρχικών μεταβλητών και παράλληλα να μη συσχετίζονται μεταξύ τους. Κατ' αυτόν τον τρόπο, επιτυγχάνεται τόσο η μείωση των διαστάσεων του προβλήματος και συνεπώς η απλοποίησή του, όσο και ο προσδιορισμός των σημαντικότερων μεταβλητών του προβλήματος.

Οι στόχοι της Ανάλυσης Κυρίων Συνιστωσών συνοψίζονται ως εξής:

1. Η **εξαγωγή** των σημαντικότερων πληροφοριών από το αρχικό σύνολο δεδομένων.
2. Η “**συμπίεση**“ του συνόλου δεδομένων.
3. Η **απλοποίηση** της περιγραφής του συνόλου δεδομένων.
4. Η **ανάλυση** της δομής των παρατηρήσεων και των μεταβλητών.

### 3.3.2 Μαθηματικό Μοντέλο

Έστω ότι,  $x_{ij}$ , είναι η «επίδοση» του αντικειμένου  $i$  στη μεταβλητή  $j$ . Τότε, το σύνολο των υπό εξέταση δεδομένων αναπαρίσταται από έναν δισδιάστατο πίνακα  $\mathbf{X}$ , διαστάσεων  $n \times m$ , όπου  $m$  το πλήθος των παρατηρούμενων μεταβλητών για έναν αριθμό  $n$  αντικειμένων, τα οποία συνθέτουν το δείγμα κάποιου πληθυσμού.

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix}$$

Το πρώτο βήμα για την εφαρμογή της Ανάλυσης Κυρίων Συνιστωσών, αποτελεί η κανονικοποίηση των δεδομένων ως προς το μέσο όρο (mean centering) των παρατηρήσεων της κάθε μεταβλητής. Πιο συγκεκριμένα, υπολογίζεται ο μέσος όρος των παρατηρήσεων της κάθε μεταβλητής και εν συνεχεία αφαιρείται από την κάθε παρατήρηση. Κατ' αυτόν τον τρόπο η εκάστοτε μεταβλητή μετασχηματίζεται σε μια μεταβλητή με μηδενικό μέσο όρο και διασπορά ίση με την αρχική. Προκύπτει έτσι ο πίνακας  $\mathbf{X}^c$ , ως:

$$\mathbf{X}^c = \begin{bmatrix} x_{11}^c & \cdots & x_{1j}^c & \cdots & x_{1m}^c \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1}^c & \cdots & x_{ij}^c & \cdots & x_{im}^c \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1}^c & \cdots & x_{nj}^c & \cdots & x_{nm}^c \end{bmatrix}, \quad \text{όπου} \quad x_{ij}^c = x_{ij} - \frac{1}{n} \cdot \sum_{i=1}^n x_{ij}$$

Στη συνέχεια, υπολογίζεται ο τετραγωνικός πίνακας συνδιακύμανσης  $\mathbf{C}$ , διαστάσεων  $m \times m$  ως εξής:

$$\mathbf{c} = \begin{bmatrix} c_{11} & \cdots & c_{1j} & \cdots & c_{1m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{i1} & \cdots & c_{ij} & \cdots & c_{im} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{m1} & \cdots & c_{mj} & \cdots & c_{mm} \end{bmatrix}, \quad \text{όπου} \quad c_{ij} = \frac{1}{n-1} \cdot \sum_{k=1}^n x_{ki}^c \cdot x_{kj}^c$$

Το επόμενο βήμα της διαδικασίας επικεντρώνεται στον υπολογισμό των ιδιοτιμών  $\lambda_i$ , με  $i = 1 \dots m$  και των αντίστοιχων ιδιοδιανυσμάτων του πίνακα  $\mathbf{C}$ . Κάθε ένα από τα,  $m$  το πλήθος, ιδιοδιανύσματα αποτελεί μια συνιστώσα (component – feature) του νέου συστήματος συντεταγμένων στο οποίο μπορούμε να μετασχηματίσουμε τα αρχικά μας δεδομένα. Οι συνιστώσες αυτές είναι μεταξύ τους γραμμικώς ανεξάρτητες και προκύπτουν ως γραμμικοί συνδυασμοί των αρχικών μεταβλητών, στους οποίους η κάθε μια από τις αρχικές μεταβλητές συμμετέχει με διαφορετική βαρύτητα.

Στη συνέχεια τα ιδιοδιανύσματα ταξινομούνται με φθίνουσα σειρά των ιδιοτιμών στις οποίες αυτά αντιστοιχούν και επιλέγονται έστω τα  $k$  τον αριθμό πρώτα εξ αυτών. Σε αυτή την επιλογή, η οποία γίνεται με κριτήρια τα οποία θα εξηγηθούν στην επόμενη ενότητα, έγκειται και η μείωση των διαστάσεων του προβλήματος. Οι  $k$  τον αριθμό συνιστώσες που επιλέχθηκαν, καλούνται Κύριες Συνιστώσες (Principal Components) και αντικαθιστούν τις αρχικές μεταβλητές του προβλήματος. Από τις συνιστώσες αυτές



συντίθεται ο πίνακας φορτίων (loading matrix), διαστάσεων  $m \times k$ , με  $k \leq m$ , ο οποίος έχει για στήλες του τα επιλεχθέντα ιδιοδιανύσματα:

$$\mathbf{L} = \begin{bmatrix} l_{11} & \cdots & l_{1j} & \cdots & l_{1k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{i1} & \cdots & l_{ij} & \cdots & l_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{m1} & \cdots & l_{mj} & \cdots & l_{mk} \end{bmatrix}$$

Τέλος υπολογίζεται ο πίνακας των αποτελεσμάτων (scores matrix)  $\mathbf{S}$ , διαστάσεων  $n \times k$ , ο οποίος συντίθεται από τις προβολές των παρατηρήσεων των αρχικών μεταβλητών στο νέο σύστημα συντεταγμένων και αναπαριστά το νέο, μετασχηματισμένο και μειωμένων διαστάσεων, σύνολο δεδομένων:

$$\boxed{\mathbf{S} = \mathbf{X}^c \times \mathbf{L}}$$

$$\begin{bmatrix} s_{11} & \cdots & s_{1j} & \cdots & s_{1k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{i1} & \cdots & s_{ij} & \cdots & s_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{n1} & \cdots & s_{nj} & \cdots & s_{nk} \end{bmatrix} = \begin{bmatrix} x_{11}^c & \cdots & x_{1j}^c & \cdots & x_{1m}^c \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1}^c & \cdots & x_{ij}^c & \cdots & x_{im}^c \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1}^c & \cdots & x_{nj}^c & \cdots & x_{nm}^c \end{bmatrix} \times \begin{bmatrix} l_{11} & \cdots & l_{1j} & \cdots & l_{1k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{i1} & \cdots & l_{ij} & \cdots & l_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{m1} & \cdots & l_{mj} & \cdots & l_{mk} \end{bmatrix}$$

### 3.3.3 Ερμηνεία

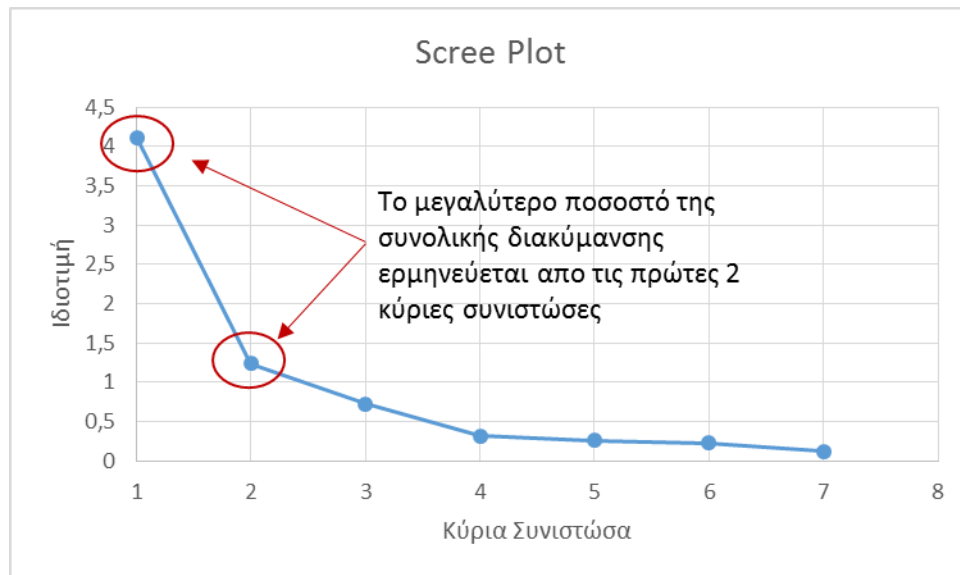
Όπως αναφέρθηκε στην προηγούμενη ενότητα, η γεωμετρική ερμηνεία της Ανάλυσης Κυρίων Συνιστωσών συνοψίζεται στο μετασχηματισμό ενός συνόλου δεδομένων σε ένα νέο σύστημα γραμμικώς ανεξάρτητων και ορθογωνίων αξόνων – συντεταγμένων. Πρακτικά, οι μεταβλητές του προβλήματος αντικαθίστανται από ένα νέο και ισομεγέθες με το αρχικό, σύνολο μεταβλητών, οι οποίες αποτελούν γραμμικό συνδυασμό των αρχικών μεταβλητών και είναι μεταξύ τους γραμμικώς ανεξάρτητες. Οι νέες αυτές μεταβλητές αποκαλούνται συνιστώσες και σε κάθε μια από αυτές αποδίδεται και ένα νέο λανθάνον χαρακτηριστικό, το οποίο αφενός μεν χαρακτηρίζει τα αντικείμενα του δείγματος που μελετάται, αφετέρου ωστόσο, μπορεί να στερείται προφανούς και εύληπτης φυσικής σημασίας. Κάθε μια από τις αρχικές μεταβλητές του

προβλήματος, συμμετέχει με διαφορετική βαρύτητα στη σύνθεση κάθε μιας εκ των νέων μεταβλητών – συνιστωσών. Οι συνεισφορές αυτές των αρχικών μεταβλητών στους γραμμικούς συνδυασμούς που συνθέτουν τις νέες συνιστώσες, καλούνται φορτία και συνθέτουν το πίνακα φορτίων, όπως αυτός παρουσιάστηκε στην προηγούμενη ενότητα. Τα φορτία αποκαλύπτουν τις σχέσεις που διέπουν τις μεταβλητές του προβλήματος και μας βοηθούν να κατανοήσουμε καλύτερα τους υποκείμενους μηχανισμούς που συγκροτούν το σύνολο των δεδομένων μας. Πράγματι, τα υψηλά φορτία των διαφορετικών μεταβλητών ως προς ένα συγκεκριμένο λανθάνον χαρακτηριστικό ή γνώρισμα, υποδεικνύουν ότι οι συγκεκριμένες μεταβλητές είναι συσχετισμένες μεταξύ τους. Αντιθέτως η συσχέτιση μεταξύ των μεταβλητών δε δύναται να αποδειχθεί εάν τα φορτία τους ως προς ένα χαρακτηριστικό είναι μικρά.

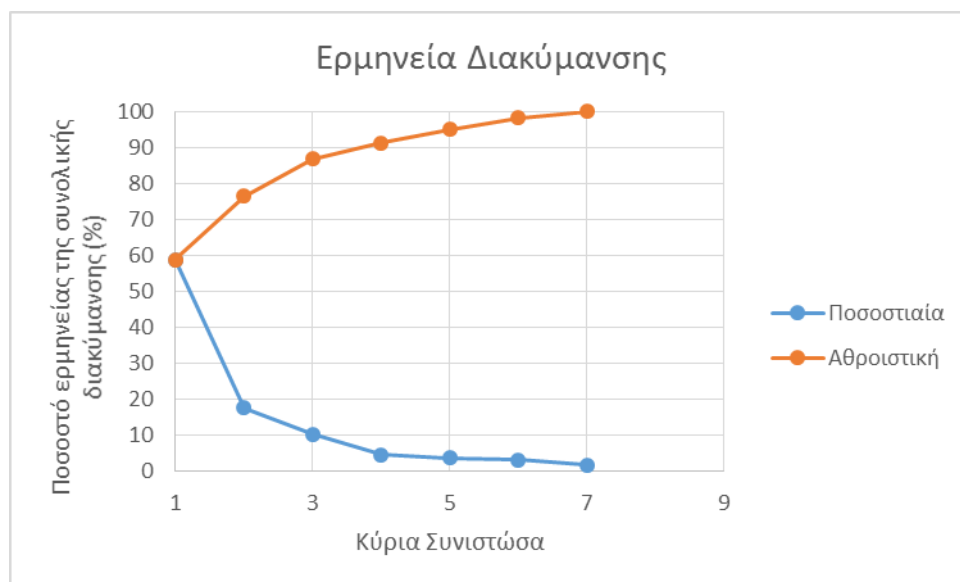
Όπως είδαμε, κάθε ιδιοδιάνυσμα του πίνακα συνδιακύμανσης και συνεπώς κάθε μια εκ των αντίστοιχων ιδιοτιμών, αντιστοιχεί σε ένα λανθάνον χαρακτηριστικό – συνιστώσα. Κάθε ιδιοτιμή χαρακτηρίζεται από το σχετικό μέτρο της ιδιοτιμής το οποίο για την  $i$ -οστή ιδιοτιμή ορίζεται ως:

$$\frac{\lambda_i}{\sum_1^m \lambda_j}$$

Το μέτρο αυτό εκφράζει το ποσοστό της συνολικής διακύμανσης, των αρχικών δεδομένων, που ερμηνεύει η  $i$ -οστή κύρια συνιστώσα. Έτσι, οι  $m$  το πλήθος κύριες συνιστώσες ερμηνεύουν το 100% της συνολικής διακύμανσης, όπως άλλωστε είναι αναμενόμενο, αφού αποτελούν έναν γεωμετρικό μετασχηματισμό των επίσης  $m$  το πλήθος αρχικών μεταβλητών. Η πρακτική αξία της μεθόδου της ανάλυσης των κυρίων συνιστωσών έγκειται στη δυνατότητα επιλογής εκείνων μόνο των κυρίων συνιστωσών, που ερμηνεύουν το μεγαλύτερο ποσοστό της συνολικής διακύμανσης των αρχικών δεδομένων. Με αυτό τον τρόπο επιτυγχάνεται μείωση των διαστάσεων του προβλήματος και συμπίεση των δεδομένων χωρίς σημαντικές απώλειες. Οι κύριες συνιστώσες του προβλήματος επιλέγονται συνήθως με τη βοήθεια ενός «scree plot», το οποίο απεικονίζει τις ιδιοτιμές των επιμέρους συνιστωσών, ταξινομημένων με φθίνουσα σειρά. Το ποσοστό της συνολικής διακύμανσης που ερμηνεύεται χαρακτηρίζεται ως ικανοποιητικά μεγάλο, σύμφωνα με την κρίση του αναλυτή και επιλέγεται αναλόγως ο απαιτούμενος αριθμός κυρίων συνιστωσών που θα χρησιμοποιηθούν. Συνήθως επιλέγονται οι συνιστώσες που βρίσκονται πριν από το σημείο απότομης αλλαγής της κλίσης της καμπύλης, όπως φαίνεται παρακάτω.



Εικόνα 3-2 Παράδειγμα Scree Plot

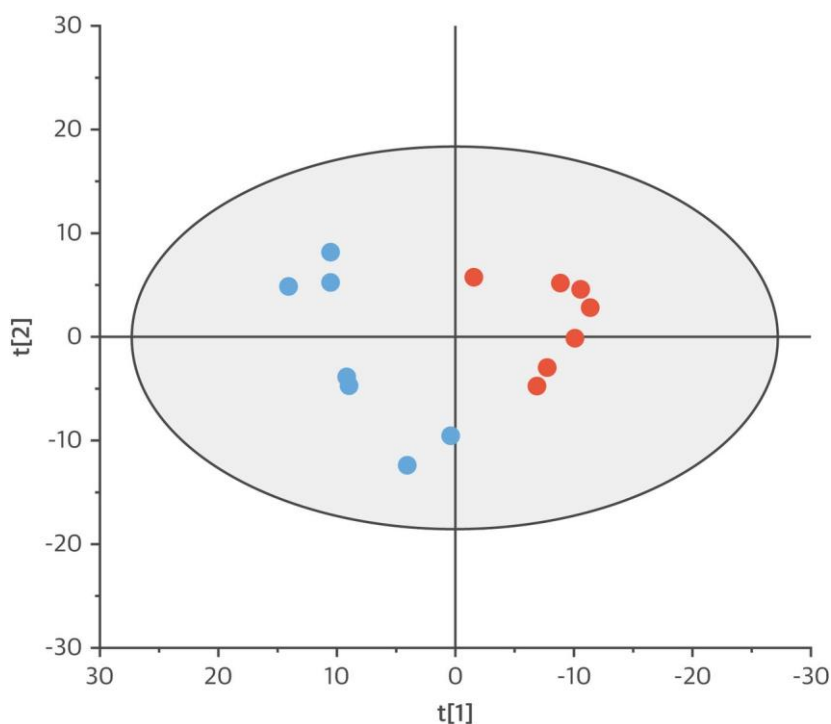


Εικόνα 3-3 Ερμηνεία συνολικής διακύμανσης

Κατ' αυτόν τον τρόπο, και συγκεκριμένα εξετάζοντας τη συνεισφορά τους στις κύριες συνιστώσες που ερμηνεύουν το μεγαλύτερο ποσοστό της διακύμανσης, μπορούμε να εξαγάγουμε συμπεράσματα για το πόσο σημαντικό ρόλο παίζει κάθε μια από τις αρχικές μεταβλητές του προβλήματος.

Τέλος, σημαντικές πληροφορίες μας παρέχουν και τα «scores», οι τιμές δηλαδή που λαμβάνουν οι νέες μεταβλητές του προβλήματος. Αν για ευκολία θεωρήσουμε μόνο τις δύο πρώτες κύριες συνιστώσες, τότε κάθε αντικείμενο του εξεταζόμενου δείγματος

και συνεπώς κάθε διακριτή επανάληψη του πειράματος, μπορεί να αναπαρασταθεί από ένα μοναδικό σημείο σε ένα δισδιάστατο διάγραμμα με οριζόντιο άξονα την κύρια συνιστώσα 1 και κάθετο άξονα την κύρια συνιστώσα 2. Ένα τέτοιο διάγραμμα καλείται «Score Plot».



**Εικόνα 3-4 Παράδειγμα Score Plot**

Τα διαγράμματα αυτού του τύπου αποτελούν ένα εύχρηστο και εύληπτο εργαλείο, μέσω του οποίου μπορούμε να διακρίνουμε συστάδες παρόμοιων επαναλήψεων του εξεταζόμενου πειράματος και συνεπώς να βγάλουμε συμπεράσματα για τους διαφόρους εξωγενείς ή ενδογενείς παράγοντες, η επίδραση των οποίων έχει σαν αποτέλεσμα κάποια συγκεκριμένη έκβαση του πειράματος.

Επιπλέον, μέσω των score plots καθίσταται δυνατός ο εντοπισμός αντικειμένων του δείγματος που παρουσιάζουν σημαντικές αποκλίσεις σε σχέση με τα υπόλοιπα, οι οποίες πιθανώς επηρεάζουν σημαντικά την αξιοπιστία των συμπερασμάτων στα οποία θα καταλήξουμε. Σε αυτήν τη περίπτωση τα αντικείμενα αυτά μπορούν να απομονωθούν από το αρχικό σύνολο δεδομένων και να επαναληφθεί εκ νέου η διαδικασία, με σκοπό την παραγωγή πιο αξιόπιστων αποτελεσμάτων.

### 3.3.4 Εφαρμογές

Η Ανάλυση Κυρίων Συνιστωσών αποτελεί μια από τις σημαντικότερες και δημοφιλέστερες στατιστικές μεθόδους πολυμεταβλητής ανάλυσης δεδομένων. Βρίσκει εφαρμογή σε ένα ευρύτατο πεδίο επιστημών, όπως είναι οι κοινωνικές επιστήμες, τα χρηματοοικονομικά, η επεξεργασία σημάτων και η τεχνητή όραση, η φαρμακολογία κ.α..

Όπως αναφέρθηκε στις προηγούμενες ενότητες, οι βασικότερες χρήσεις της μεθόδου είναι η οπτικοποίηση σύνθετων και πολύπλοκων συνόλων δεδομένων, η μείωση των διαστάσεων ενός πολυδιάστατου προβλήματος (dimensionality reduction), ο εντοπισμός συνόλων παρατηρήσεων που παρουσιάζουν μεγάλες αποκλίσεις από τις υπόλοιπες παρατηρήσεις που απαρτίζουν το εξεταζόμενο δείγμα (outlier – fault detection), η ομαδοποίηση και η ανίχνευση συστάδων παρατηρήσεων με παρόμοια χαρακτηριστικά (classification – clustering) και γενικώς ο εντοπισμός μοτίβων που χαρακτηρίζουν το εξεταζόμενο σύνολο δεδομένων (pattern recognition). Χάρη στην απλότητα της και το ευρύ φάσμα λειτουργιών τις οποίες μπορεί να επιτελέσει, η Ανάλυση Κυρίων Συνιστωσών μπορεί να χρησιμοποιηθεί είτε ως αυτοτελής μέθοδος ανάλυσης δεδομένων, είτε ως μέθοδος προ-επεξεργασίας και προετοιμασίας των δεδομένων για την περαιτέρω ανάλυσή τους από άλλες μεθόδους.

Συγκεκριμένα στην περίπτωση της βιομηχανικής παραγωγής σε παρτίδες, η οποία και αποτελεί το αντικείμενο της παρούσας εργασίας, η Ανάλυση Κυρίων Συνιστωσών επιτρέπει την αναγνώριση και την αξιολόγηση εκείνων των μεταβλητών που χαρακτηρίζουν τα παραγόμενα προϊόντα και τις αντίστοιχες διαδικασίες παραγωγής, οι οποίες μπορεί να είναι κρίσιμες για την ποιότητα του προϊόντος και την απόδοση της παραγωγικής διαδικασίας. Επιπλέον, αυτή η μέθοδος μπορεί να χρησιμοποιηθεί για να αναπτύξει μια κατανόηση της διαδραστικής σχέσης μεταξύ των εισροών και των μετρήσεων της διεργασίας και της ανάλυσης του τελικού προϊόντος. Όταν εφαρμόζεται σε πραγματικό χρόνο, η PCA μπορεί να χρησιμοποιηθεί για την αναγνώριση πιθανών καταστάσεων βλάβης και να ποσοτικοποιήσει τις επιπτώσεις τους στην ποιότητα των προϊόντων. Επιπλέον, με τη χρήση ιστορικών δεδομένων και τη βοήθεια της PCA, μπορούν να ομαδοποιηθούν οι διάφορες παρτίδες αναλόγως με τα χαρακτηριστικά τους και συνεπώς να ερμηνευθούν τυχόν αποκλίσεις κάποιων παρτίδων από αυτές που θεωρούνται ιδανικές και τελικά να εντοπιστούν εκείνοι οι

παράγοντες που μπορεί να οδηγήσουν την διαδικασία στην παραγωγή ελαττωματικών παρτίδων.

### 3.3.5 Πλεονεκτήματα και Μειονεκτήματα

Όπως είδαμε η μέθοδος της Ανάλυσης Κυρίων Συνιστωσών, είναι μια πολύ ισχυρή και εξαιρετικά χρήσιμη μέθοδος ανάλυσης δεδομένων, η οποία δύναται να προσδώσει σημαντική προστιθέμενη αξία στις αναλύσεις μας. Τα πλεονεκτήματα της εν λόγω μεθόδου συνοψίζονται ως εξής:

1. Μπορεί να εφαρμοστεί σε οποιοδήποτε σύνολο δεδομένων
2. Διαχειρίζεται αποτελεσματικά πολύ μεγάλα σύνολα δεδομένων
3. Αποτελεί μια ταχεία και αποτελεσματική μαθηματική τεχνική
4. Μειώνει τις διαστάσεις του εξεταζόμενου προβλήματος, ελαχιστοποιώντας παράλληλα την απώλεια πληροφορίας
5. Παρέχει πολύτιμες πληροφορίες για το σύνολο δεδομένων και τους υποκείμενους μηχανισμούς που το συγκροτούν.

Ωστόσο, όπως κάθε μέθοδος ανάλυσης δεδομένων, η PCA έχει και ορισμένα μειονεκτήματα, τα οποία θα πρέπει να λαμβάνονται σοβαρά υπόψιν κατά την εφαρμογή της:

1. Η μοντελοποίηση προβλημάτων με μη γραμμικά χαρακτηριστικά δεν είναι τόσο αποτελεσματική
2. Οι νέες συνιστώσες που προκύπτουν ως αποτέλεσμα της PCA στερούνται εύκολα ερμηνεύσιμης φυσικής σημασίας
3. Η εφαρμογή της συνεπάγεται απώλεια πληροφορίας, το ποσό της οποίας εξαρτάται από την κρίση του αναλυτή.

## 3.4 Παλινδρόμηση Κυρίων Συνιστωσών - PCR

### 3.4.1 Γενικά

Η μέθοδος της Παλινδρόμησης Κυρίων Συνιστωσών, όπως υποδηλώνει και η ονομασία της, αποτελεί ένα συνδυασμό των μεθόδων της Πολλαπλής Παλινδρόμησης και της Ανάλυσης Κυρίων συνιστωσών και εφαρμόζεται στις περιπτώσεις εκείνες στις οποίες η πολλαπλή παλινδρόμηση αποτυγχάνει να παραγάγει αξιόπιστα αποτελέσματα εξ' αιτίας της φύσης των ανεξάρτητων μεταβλητών του προβλήματος και συνεπώς του τρόπου με τον οποίο έχει δομηθεί το σύνολο των δεδομένων που καλούμαστε να εξετάσουμε.

Συγκεκριμένα, ερμηνεία ενός προβλήματος με τη χρησιμοποίηση της μεθόδου αναλύσεως της πολλαπλής παλινδρόμησης επιτυγχάνεται καλύτερα όταν οι ανεξάρτητες μεταβλητές που αποτελούν το μοντέλο είναι μεταξύ τους ασυσχέτιστες. Όταν παρατηρούνται έντονες συσχετίσεις μεταξύ των μεταβλητών είναι δύσκολο, αν όχι αδύνατο, να αξιολογηθεί η ουσιαστική συνεισφορά της μεταβολής μιας συγκεκριμένης ανεξάρτητης μεταβλητής, στην αντίστοιχη μεταβολή της εξαρτημένης μεταβλητής που οφείλεται αποκλειστικά στη συγκεκριμένη ανεξάρτητη μεταβλητή. Όταν οι ανεξάρτητες μεταβλητές δεν είναι γραμμικώς ανεξάρτητες (ορθογώνιες) μεταξύ τους, τότε οι εκτιμώμενοι συντελεστές παλινδρόμησης τείνουν να είναι εξαιρετικά ασταθείς και οι τιμές τους να υφίστανται δραματικές αλλαγές όταν κάποια νέα μεταβλητή προστίθεται ή απομακρύνεται ή όταν συμβαίνουν μικρές μεταβολές στα δεδομένα του προβλήματος. Το παραπάνω φαινόμενο, το οποίο έχει σαν αποτέλεσμα την παραγωγή αναξιόπιστων αποτελεσμάτων, μέσω του μοντέλου της πολλαπλής παλινδρόμησης καλείται πολυσυγγραμμικότητα (multicollinearity).

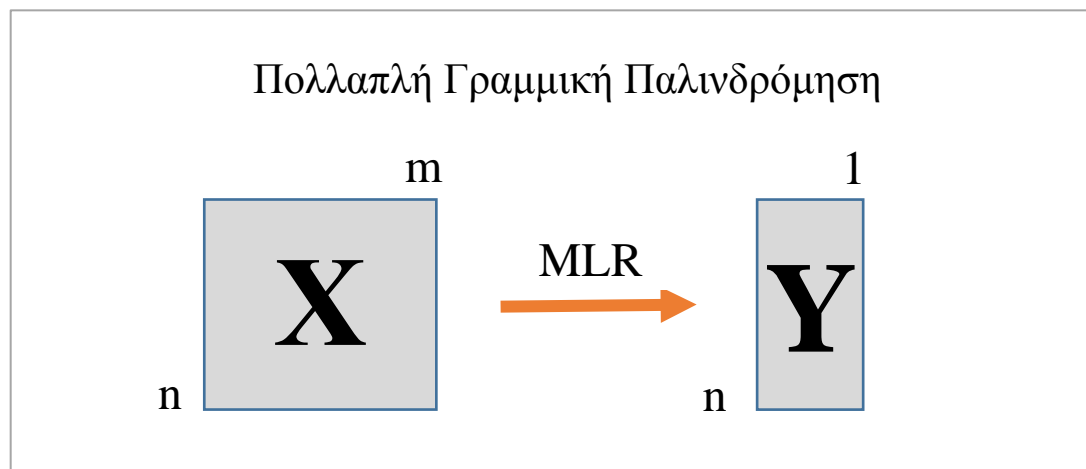
Η παρατήρηση του φαινομένου της πολυσυγγραμμικότητας μπορεί να οφείλεται τόσο σε εγγενή χαρακτηριστικά και περιορισμούς του προβλήματος, το οποίο καλούμαστε να εξετάσουμε και των μεταβλητών που το χαρακτηρίζουν, όσο και στον τρόπο με τον οποίο συλλέγονται τα δεδομένα. Αν για παράδειγμα το δείγμα που εξετάζεται δεν είναι απολύτως αντιπροσωπευτικό του πληθυσμού από τον οποίο προέρχεται, μπορεί να προκύψει το εσφαλμένο συμπέρασμα για τη συσχέτιση δύο μεταβλητών του προβλήματος. Επιπλέον υπάρχουν περιπτώσεις όπου η φυσική σημασία κάποιων μεταβλητών επιβάλλει την ύπαρξη έντονης γραμμικής συσχέτισης μεταξύ τους και η συμπερίληψή των μεταβλητών αυτών στο εξεταζόμενο σύνολο δεδομένων οδηγεί εξ' ορισμού στο φαινόμενο της πολυσυγγραμμικότητας.

Η αντιμετώπιση του φαινομένου της πολυσυγγραμμικότητας έγκειται στο μετασχηματισμό του αρχικού συνόλου δεδομένων, σε ένα νέο, το οποίο θα αποτελείται εξ' ολοκλήρου από μεταβλητές οι οποίες δε θα είναι γραμμικώς συσχετισμένες μεταξύ τους. Στο νέο, μετασχηματισμένο και απαλλαγμένο από το φαινόμενο της πολυσυγγραμμικότητας, σύνολο δεδομένων μπορεί εν συνεχεία να εφαρμοστεί η μέθοδος της Πολλαπλής Παλινδρόμησης.

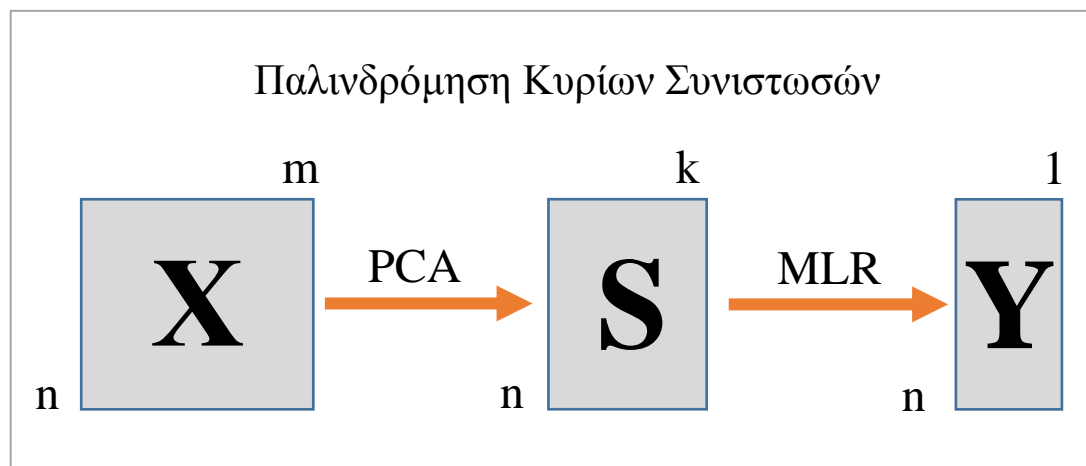
### 3.4.2 Μαθηματικό Μοντέλο

Όπως ήδη αναφέρθηκε, η μέθοδος της Παλινδρόμησης Κυρίων Συνιστωσών είναι ένας συνδυασμός της μεθόδου PCA και της πολλαπλής παλινδρόμησης. Έτσι η εφαρμογή της μεθόδου συνοψίζεται ως εξής:

1. Εντοπισμός φαινομένου πολυσυγγραμμικότητας
2. Εφαρμογή Ανάλυσης Κυρίων Συνιστωσών στο σύνολο των παρατηρήσεων των ανεξάρτητων μεταβλητών
3. Εφαρμογή του μοντέλου της Πολλαπλής Παλινδρόμησης με χρήση του νέου συμπιεσμένου συνόλου δεδομένων.



Εικόνα 3-5 Σχηματική απεικόνιση του μοντέλου MLR



Εικόνα 3-6 Σχηματική απεικόνιση του μοντέλου PCR

Κατά συνέπεια, αν το αρχικό γραμμικό μοντέλο, όπως αυτό παρουσιάστηκε στην ενότητα 3.2.2, είναι το:

$$Y = X\beta + \varepsilon$$



Τότε, κατόπιν εφαρμογής της PCA στο σύνολο των παρατηρήσεων των ανεξάρτητων μεταβλητών, που συνθέτει τον πίνακα  $\mathbf{X}$ , διαστάσεων  $n \times m$  και σύμφωνα με τα όσα παρουσιάστηκαν στην ενότητα 3.3.2, το νέο γραμμικό μοντέλο πολλαπλής παλινδρόμησης θα γράφεται ως:

$$\mathbf{Y} = \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \text{όπου } \boldsymbol{\gamma} = (\mathbf{S}^T \cdot \mathbf{S})^{-1} \cdot \mathbf{S}^T \cdot \mathbf{Y} \text{ και } \mathbf{S} = \mathbf{X}^c \times \mathbf{L}$$

Τέλος η εκτίμηση των συντελεστών παλινδρόμησης  $\beta_i$ , θα προκύπτει ως

$$\boxed{\boldsymbol{\beta} = \mathbf{L} \times \boldsymbol{\gamma}}$$

,όπου  $\mathbf{L}$  η μήτρα φορτίων η οποία συντίθεται από τα  $k$  επιλεγθέντα ιδιοδιανύσματα της μήτρας συνδιασποράς, όπως παρουσιάστηκε στην ενότητα 3.3.2.

### 3.4.3 Ερμηνεία

Όπως αναφέρθηκε σε προηγούμενη ενότητα, ένα από τα μειονεκτήματα της ανάλυσης κυρίων συνιστωσών, έγκειται στη δυσκολία της ερμηνείας της φυσικής σημασίας των νέων λανθανόντων χαρακτηριστικών, τα οποία αποδίδονται στο μελετώμενο πρόβλημα και αντικαθιστούν το αρχικό σύνολο των μεταβλητών. Η παλινδρόμηση κυρίων συνιστωσών αναζητά τις υποκείμενες γραμμικές σχέσεις μεταξύ του εκάστοτε λανθάνοντος χαρακτηριστικού, με την εξαρτημένη μεταβλητή, με δεδομένο ότι όλα τα υπόλοιπα λανθάνοντα χαρακτηριστικά διατηρούνται σταθερά. Με άλλα λόγια, όπως είδαμε, οι αρχικές μεταβλητές, κάποιες εκ των οποίων είναι γραμμικώς εξαρτημένες μεταξύ τους, αντικαθίστανται από τα γραμμικώς ανεξάρτητα λανθάνοντα χαρακτηριστικά και εφαρμόζεται το μοντέλο της πολλαπλής παλινδρόμησης, μεταξύ αυτών και της ανεξάρτητης μεταβλητής. Τέλος, προκειμένου να υπερκεραστεί η δυσκολία ερμηνείας των γραμμικών σχέσεων που προκύπτουν, οι συντελεστές παλινδρόμησης, μετασχηματίζονται εκ νέου στο αρχικό σύστημα συντεταγμένων.

### 3.4.4 Εφαρμογές

Η μέθοδος της Παλινδρόμησης Κυρίων Συνιστωσών, αποτελεί ένα συνδυασμό των μεθόδων της Πολλαπλής Παλινδρόμησης και της Ανάλυσης Κυρίων Συνιστωσών και ως εκ τούτου χρησιμοποιείται σε ένα ευρύ φάσμα εφαρμογών που μπορεί να μοντελοποιηθεί από γραμμικά μοντέλα και κυρίως σε περιπτώσεις στις οποίες κρίνεται απαραίτητη η αντιμετώπιση του φαινομένου της πολυσυγγραμμικότητας.

### 3.4.5 Πλεονεκτήματα και Μειονεκτήματα

Η Παλινδρόμηση Κυρίων Συνιστωσών είναι μια δημοφιλής και ευρέως χρησιμοποιούμενη μέθοδος. Πλεονεκτήματα της PCR περιλαμβάνουν τα ακόλουθα:

- ✓ Η PCR μπορεί να κάνει παλινδρόμηση όταν οι επεξηγηματικές μεταβλητές είναι εντόνως συσχετισμένες ή ακόμη και απολύτως συγγραμμικές.
- ✓ Η μόνη απόφαση που πρέπει να ληφθεί από τον αναλυτή, είναι πόσες βασικές συνιστώσες πρέπει να χρησιμοποιηθούν. Τα λανθάνοντα χαρακτηριστικά που αγνοούνται, δίνουν μια εικόνα για το ποιοι γραμμικοί συνδυασμοί μεταβλητών είναι υπεύθυνοι για την πολυσυγγραμμικότητα.
- ✓ Η PCR έχει μια μοναδική διακριτή παράμετρο, δηλαδή τον αριθμό των λανθανόντων χαρακτηριστικών που χρησιμοποιούνται.
- ✓ Η PCR μπορεί να εφαρμοστεί σε περιπτώσεις, όπου υπάρχουν περισσότερες μεταβλητές από τις παρατηρήσεις (ευρεία δεδομένα).

Τέλος, καθώς η PCR βασίζεται στην PCA, ένα μεγάλο πλεονέκτημα της παλινδρόμησης PCR έναντι της κλασικής παλινδρόμησης είναι τα διαθέσιμα γραφήματα που περιγράφουν τη δομή των δεδομένων. Χάρη στις γραφικές παραστάσεις συσχέτισης και φορτίων, είναι εύκολο να μελετήσουμε τη σχέση μεταξύ των μεταβλητών. Μπορεί να είναι σχέσεις μεταξύ των επεξηγηματικών μεταβλητών, καθώς και μεταξύ επεξηγηματικών και εξαρτημένων μεταβλητών. Το γράφημα των αποτελεσμάτων (score plot), παρέχει πληροφορίες σχετικά με την εγγύτητα των παρατηρήσεων του δείγματος και τη δομή του συνόλου δεδομένων.

Όσον αφορά στα μειονεκτήματα της μεθόδου, αξίζει να αναφέρουμε ότι, ο αλγόριθμος που είναι σήμερα γνωστός ως PCR είναι στην πραγματικότητα μια παρερμηνεία των αρχικών ιδεών πίσω από την PCR. Όταν οι Kendall και Hotelling πρότειναν για πρώτη φορά την PCR στη δεκαετία του 1950, πρότειναν «πλήρη» PCR, που σημαίνει την αντικατάσταση των αρχικών μεταβλητών από όλες τις κύριες συνιστώσες και όχι την επιλογή μόνο εκείνων οι οποίες ερμηνεύουν το μεγαλύτερο ποσοστό της διακύμανσης. Σήμερα, και ήδη από τη δεκαετία του 1890, ο όρος PCR έχει ταυτιστεί με την «ατελή» PCR, όπου επιλέγονται μόνο κάποιες από τις κύριες συνιστώσες. Το βασικότερο επιχείρημα κατά της χρήσης της ατελούς PCR συνοψίζεται στο γεγονός ότι κατά την εφαρμογή της και συγκεκριμένα στην απόφαση για το ποιες κύριες συνιστώσες θα χρησιμοποιηθούν, δεν λαμβάνεται υπόψιν η εξαρτημένη μεταβλητή απόκρισης. Στην πραγματικότητα, δεν υπάρχει λόγος να πιστεύουμε ότι τα χαρακτηριστικά που ερμηνεύουν το μεγαλύτερο ποσοστό της διακύμανσης, θα είναι και αυτά που προβλέπουν την εξαρτημένη μεταβλητή με τη μεγαλύτερη αξιοπιστία.

## 3.5 Ανάλυση Μερικών Ελαχίστων Τετραγώνων– PLS/OPLS

### 3.5.1 Γενικά

Όπως είδαμε στην προηγούμενη ενότητα, οι προκλήσεις τις οποίες συναντούμε κατά την ανάλυση πραγματικών δεδομένων και οι οποίες αποτελούν εγγενή χαρακτηριστικά των φαινομένων τα οποία επιχειρούμε να ερμηνεύσουμε, μας οδηγούν στην αναζήτηση νέων μεθόδων ανάλυσης ή στην επέκταση ήδη υπαρχόντων. Μια τέτοια περίπτωση επέκτασης των ήδη υπαρχόντων μεθόδων ανάλυσης, με σκοπό την διεύρυνση του πεδίου εφαρμογής τους σε προβλήματα αυξημένης πολυπλοκότητας, αποτελεί η μέθοδος της Ανάλυσης Μερικών Ελαχίστων Τετραγώνων ή αλλιώς η Προβολή σε Λανθάνουσες Δομές, όπως είναι ευρέως γνωστή.

Οι παραδοχές οι οποίες απαιτούνται για την εφαρμογή ορισμένων μεθόδων ανάλυσης, όπως για παράδειγμα η Πολλαπλή Γραμμική Παλινδρόμηση, συχνά δεν αντικατοπτρίζονται στο διαθέσιμο προς ανάλυση σύνολο δεδομένων. Χαρακτηριστικά παραδείγματα τέτοιων παραδοχών είναι η μη ύπαρξη πολυσυγγραμμικότητας, όπως αυτή εξηγήθηκε σε προηγούμενη ενότητα, καθώς και η ύπαρξη μιας και μοναδικής εξαρτημένης μεταβλητής. Όπως είδαμε, το πρόβλημα της πολυσυγγραμμικότητας αντιμετωπίζεται ικανοποιητικά από την μέθοδο της Παλινδρόμησης Κυρίων Συνιστωσών, η οποία ωστόσο, αφενός λειτουργεί για μονοδιάστατο πίνακα εξαρτημένων μεταβλητών και αφετέρου δεν λαμβάνει υπόψη της την εξαρτημένη μεταβλητή κατά την επιλογή των κυρίων συνιστωσών που χαρακτηρίζουν το σύνολο δεδομένων των ανεξάρτητων μεταβλητών. Τις δύο αυτές αδυναμίες της μεθόδου PCR, καλείται να αντιμετωπίσει η Ανάλυση Μερικών Ελαχίστων Τετραγώνων.

Συγκεκριμένα, όπως θα εξηγήσουμε στη συνέχεια, η μέθοδος PLS αποτελεί ένα σημαντικό εργαλείο ανάλυσης δεδομένων, το οποίο μας δίνει τη δυνατότητα να αντιμετωπίσουμε προβλήματα με πολυδιάστατα σύνολα δεδομένων τόσο ανεξάρτητων όσο και εξαρτημένων μεταβλητών, εξασφαλίζοντας τη μείωση των διαστάσεων του προβλήματός, αφενός με τρόπο ώστε να εξαλείφεται το φαινόμενο της πολυσυγγραμμικότητας και αφετέρου με κριτήριο τόσο την ταυτόχρονη ερμηνεία του μέγιστου δυνατού ποσοστού διακύμανσης στο σύνολο των ανεξάρτητων αλλά και στο σύνολο των εξαρτημένων μεταβλητών, όσο και την αξιόπιστη πρόβλεψη των εξαρτημένων μεταβλητών.

### 3.5.2 Μαθηματικό Μοντέλο

Η μέθοδος των Μερικών Ελαχίστων Τετραγώνων, παρουσιάζει αρκετές ομοιότητες με την μέθοδο της Παλινδρόμησης Κυρίων Συνιστωσών. Η ουσιαστικότερη διαφορά μεταξύ των δύο μεθόδων εντοπίζεται στο γεγονός ότι η PLS χρησιμοποιεί τα δεδομένα, τόσο του συνόλου των ανεξάρτητων μεταβλητών  $X$ , όσο και αυτά του συνόλου των εξαρτημένων μεταβλητών  $Y$ , προκειμένου να προσδιορίσει τις κύριες συνιστώσες του προβλήματος.

Όπως είδαμε σε προηγούμενες ενότητες, η εφαρμογή των PCA και PCR αποσκοπεί στον προσδιορισμό εκείνων των λανθανόντων μεταβλητών, των λεγόμενων Κυρίων Συνιστωσών, οι οποίες ερμηνεύουν την διακύμανση στο σύνολο των παρατηρήσεων των ανεξάρτητων μεταβλητών  $X$ , με τον καλύτερο δυνατό τρόπο. Στην μέθοδο PLS, τα κριτήρια με βάση την ταυτόχρονη ικανοποίηση των οποίων, προσδιορίζονται οι κύριες συνιστώσες του προβλήματος είναι τα εξής:

1. Η καλύτερη δυνατή ερμηνεία του συνόλου  $X$
2. Η καλύτερη δυνατή ερμηνεία του συνόλου  $Y$
3. Η μέγιστη δυνατή συσχέτιση μεταξύ των συνόλων  $X$  και  $Y$ .

Κατ' αναλογία με την μέθοδο PCA όπου υπολογίσαμε τον πίνακα αποτελεσμάτων ή αλλιώς τα scores, τα οποία ουσιαστικά συνθέτουν το σύνολο των παρατηρήσεων των μετασχηματισμένων μεταβλητών  $X$ , στην περίπτωση της μεθόδου PLS υπολογίζουμε δύο σύνολα scores και συγκεκριμένα τα  $X$ -scores και τα  $Y$ -scores που αντιστοιχούν στο σύνολο των ανεξάρτητων μεταβλητών  $X$  και των εξαρτημένων μεταβλητών  $Y$  αντιστοίχως, ως:

$$T = X \times W \quad \text{και} \quad U = Y \times C$$

Κάθε στήλη  $t_a$  του  $T$  αποτελεί την  $a$ -οστή κύρια συνιστώσα του  $X$  και αντιστοίχως κάθε στήλη  $u_a$  του  $U$  αποτελεί την  $a$ -οστή κύρια συνιστώσα του  $Y$ . Ο σκοπός της μεθόδου PLS είναι ο προσδιορισμός των προαναφερθέντων scores με τέτοιο τρόπο ώστε να μεγιστοποιείται η συνδιακύμανσή τους. Δηλαδή για κάθε κύρια συνιστώσα των  $X$  και  $Y$ , θέλουμε:

$$COV(t_a, u_a) = E\{(t_a - \bar{t}_a) \cdot (u_a - \bar{u}_a)\} = \max$$

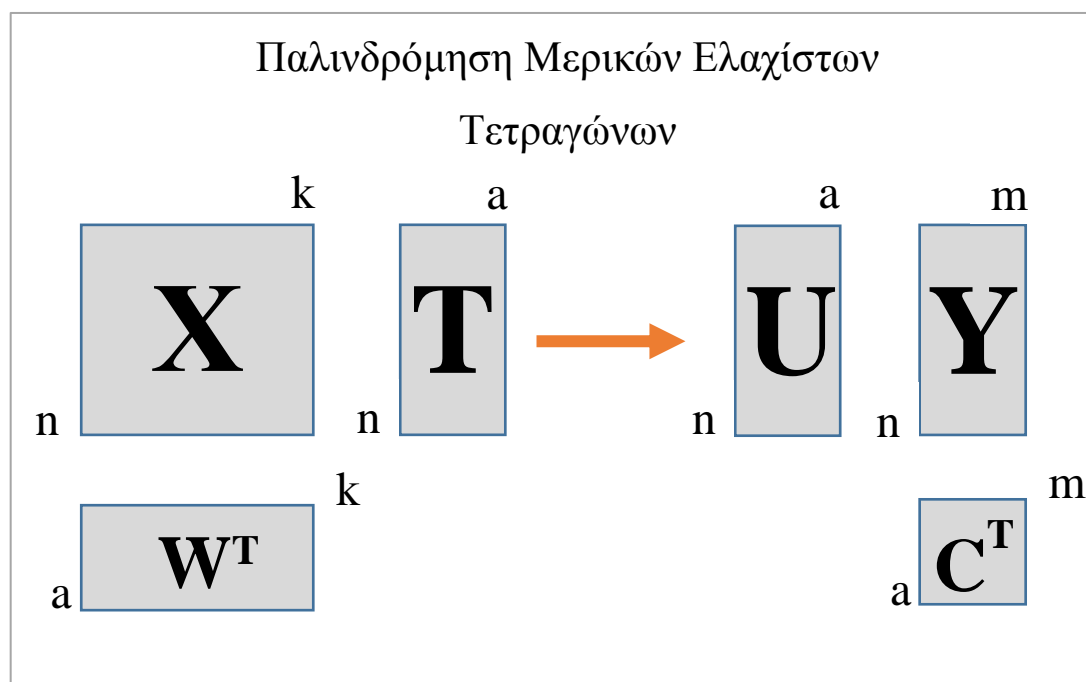
ή αλλιώς:

$$COV(t_a, u_a) = CORR(t_a, u_a) \times \sqrt{VAR(t_a)} \times \sqrt{VAR(u_a)} = \max$$

$$COV(t_a, u_a) = \sqrt{t_a^T t_a} \times \sqrt{u_a^T u_a} \times CORR(t_a, u_a) = max$$

Η μεγιστοποίηση του γινομένου αυτού έγκειται στη μεγιστοποίηση καθενός εκ των τριών όρων που το απαρτίζουν. Κάθε ένας από τους όρους αυτούς εκφράζει μαθηματικά και από ένα εκ των τριών κριτηρίων με βάση τα οποία προσδιορίζονται οι κύριες συνιστώσες, με τη σειρά που αυτά παρουσιάστηκαν στην αρχή της παρούσας υποενότητας. Τα  $t_a, u_a$  υπολογίζονται με τους περιορισμούς

$$w_a^T w_a = 1 \quad \text{και} \quad c_a^T c_a = 1$$



**Εικόνα 3-7 Σχηματική απεικόνιση της μεθόδου PLS**

Ο υπολογισμός των  $T, U, W$  και  $C$  πραγματοποιείται με τη βοήθεια του αλγορίθμου NIPALS, η εξήγηση του οποίου ξεφεύγει από το πλαίσιο της παρούσας εργασίας.

Τελικά προκύπτει ένα γραμμικό μοντέλο παλινδρόμησης της μορφής:

$$Y = T \times C^T + E = X \times W \times C^T + E,$$

όπου  $E$  οι αποκλίσεις (residuals) του μοντέλου.

### 3.5.3 Ερμηνεία

Όπως έχουμε ήδη περιγράψει, το Μοντέλο Παλινδρόμησης Μερικών Ελαχίστων Τετραγώνων σχηματίζει νέες  $X$  μεταβλητές, τις εκτιμήσεις  $t_a$ , οι οποίες είναι γραμμικοί συνδυασμοί των αρχικών μας  $X$  μεταβλητών και ακολούθως χρησιμοποιεί τις  $t_a$  ως προβλέπουσες του πίνακα  $Y$ . Κατ' αναλογία με την μέθοδο PCA, τα  $X$ -scores και  $Y$ -scores αποτελούν μια συλλογή από λανθάνοντα χαρακτηριστικά τα οποία ερμηνεύουν το μεγαλύτερο δυνατό ποσοστό της διακύμανσης εντός των συνόλων των ανεξάρτητων και εξαρτημένων μεταβλητών αντιστοίχως. Τα βάρη  $w_a$  και  $c_a$  παρέχουν την πληροφορία σχετικά με τον τρόπο που οι μεταβλητές συνδυάζονται προκειμένου να σχηματίσουν την ποσοτική σχέση ανάμεσα στους  $X$  και  $Y$  πίνακες. Έτσι κατορθώνουν να ερμηνεύσουν τα  $t_a$  και  $u_a$ . Επιπλέον ανάλογα με την τιμή τους, τα βάρη είναι απαραίτητα για την κατανόηση του ποιες  $X$  μεταβλητές είναι σημαντικές για την ερμηνεία του μοντέλου μας. Τέτοιες μεταβλητές έχουν αντίστοιχες υψηλές τιμές των  $w_a$ . Παρόμοιες τιμές των  $w_a$ , υποδεικνύουν ότι οι αντίστοιχες μεταβλητές τους παρέχουν παρόμοιες πληροφορίες. Όσον αφορά τώρα τα υπόλοιπα (residuals), πρόκειται για το μέρος των δεδομένων μας το οποίο αδυνατεί να εξηγήσει το μοντέλο μας. Παρ' όλα αυτά, τα υπόλοιπα παρέχουν διαγνωστικό ενδιαφέρον. Συγκεκριμένα, υψηλές τιμές των  $Y$  υπολοίπων δείχνουν ότι το μοντέλο είναι ασθενές, δεν προσαρμόζεται δηλαδή κατάλληλα στα δεδομένα.

### 3.5.4 Εφαρμογές

Η μέθοδος της Παλινδρόμησης Μερικών Ελαχίστων Τετραγώνων βρίσκει εφαρμογή, όπως εξηγήσαμε, σε εφαρμογές στις οποίες, αφενός πρέπει να αντιμετωπισθεί το φαινόμενο της πολυσυγγραμμικότητας και αφετέρου το σύνολο των εξαρτημένων μεταβλητών είναι πολυδιάστατο. Τέτοιου είδους εφαρμογές απαντώνται κατά κόρον στην χημειομετρία, την βιοπληροφορική και τις νευροεπιστήμες.

Συγκεκριμένα όσον αφορά τις διαδικασίες παραγωγής σε παρτίδες, η μέθοδος της Παλινδρόμησης Μερικών Ελαχίστων Τετραγώνων μπορεί να χρησιμοποιηθεί για την ανάλυση του αντίκτυπου που μπορεί να έχουν οι συνθήκες επεξεργασίας, των οποίων τα κύρια χαρακτηριστικά (πχ. θερμοκρασία, πίεση, υγρασία κλπ.) μετρούνται σε πραγματικό χρόνο κατά τη διάρκεια της παραγωγής, στις παραμέτρους ποιότητας του τελικού προϊόντος. Όταν η μέθοδος PLS εφαρμόζεται σε πραγματικό χρόνο, είναι δυνατόν να παρέχει στους χειριστές μια συνεχή πρόβλεψη των τελικών τιμών των

παραμέτρων ποιότητας του προϊόντος μετά από την ολοκλήρωση παραγωγής μιας παρτίδας, δίνοντας τους έτσι τη δυνατότητα να προχωρήσουν σε διορθωτικούς χειρισμούς, το αποτέλεσμα των οποίων μπορεί επίσης να προβλέπεται σε πραγματικό χρόνο. Κατ' αυτόν τον τρόπο μπορούν να αποφευχθούν σημαντικές αστοχίες οι οποίες θα μπορούσαν να θέσουν μια παρτίδα εκτός των προδιαγραφών ποιότητας.

### **3.5.5 Πλεονεκτήματα και Μειονεκτήματα**

Τα σημαντικότερα πλεονεκτήματα της μεθόδου της Παλινδρόμησης Μερικών Ελαχίστων Τετραγώνων έναντι των υπόλοιπων μεθόδων, στις οποίες έχουμε αναφερθεί, συνοψίζονται στη δυνατότητα της μεθόδου να παραγάγει ένα μοναδικό μοντέλο που να περιγράφει ένα πολυδιάστατο πρόβλημα, τόσο με πολλαπλές ανεξάρτητες όσο και πολλαπλές εξαρτημένες μεταβλητές, το οποίο αφενός μειώνει τις διαστάσεις του προβλήματος και αφετέρου μπορεί να προβλέπει ικανοποιητικά τις τιμές των εξαρτημένων μεταβλητών δεδομένων των παρατηρήσεων των ανεξάρτητων μεταβλητών.

Το βασικό μειονέκτημα της μεθόδου είναι ότι προϋποθέτει την ύπαρξη υποκείμενων γραμμικών σχέσεων μεταξύ των μεταβλητών εισόδου και των αποκρίσεων.





## ΚΕΦΑΛΑΙΟ 4: *Το λογισμικό SIMCA*

---

### 4.1 Εισαγωγή

Η εξέλιξη της τεχνολογίας καθώς και η αυξημένη δημοτικότητα και η ταχύτατη πρόοδος της Επιστήμης των Δεδομένων τα τελευταία χρόνια, είχε σαν αποτέλεσμα την ανάπτυξη μεγάλου αριθμού εργαλείων λογισμικού, τα οποία δίνουν έμφαση στην αυτοματοποίηση χρονοβόρων, περίπλοκων και υπολογιστικά απαιτητικών λειτουργιών διαχείρισης και επεξεργασίας μεγάλου όγκου δεδομένων. Έτσι, η υλοποίηση σύγχρονων, εξειδικευμένων και εξαιρετικά σύνθετων αλγορίθμων για την ανάλυση δεδομένων έχει καταστεί πιο εύκολη από ποτέ, επιτρέποντας στους αναλυτές να αφοσιώνονται στην όσο το δυνατόν καλύτερη κατανόηση του προβλήματος αυτού καθαυτού και δίνοντάς τους τη δυνατότητα να επιλέγουν τον εκάστοτε καταλληλότερο αλγόριθμο αφού πρώτα πειραματιστούν με πολλές διαφορετικές τεχνικές.

Στην παρούσα εργασία θα χρησιμοποιηθεί το λογισμικό Ανάλυσης Δεδομένων ονόματι SIMCA, το οποίο επικεντρώνεται στην πολυμεταβλητή ανάλυση διεργασιών και χρησιμοποιείται ευρέως στη βιομηχανία. Η επιλογή του συγκεκριμένου εργαλείου βασίστηκε κυρίως στις εξειδικευμένες λειτουργίες τις οποίες παρέχει για εφαρμογές διαδικασιών παραγωγής σε παρτίδες. Οι λειτουργίες αυτές συνοψίζονται στη δυνατότητα υλοποίησης δημοφιλών τεχνικών ανάλυσης, όπως αυτές που παρουσιάστηκαν στο προηγούμενο κεφάλαιο, με τις κυριότερες εξ αυτών να είναι οι μέθοδοι Ανάλυσης Κυρίων Συνιστωσών (PCA) και Παλινδρόμησης Μερικών Ελαχίστων Τετραγώνων (PLS) καθώς και στη δυνατότητα υλοποίησης μιας ευρύτατης ποικιλίας διαδραστικών και εξαιρετικά ενημερωτικών και χρήσιμων οπτικοποιήσεων.

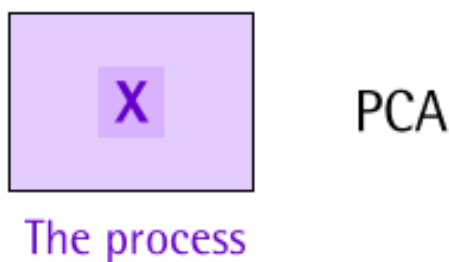
Στις ενότητες που ακολουθούν θα γίνει μια παρουσίαση της λογικής με βάση την οποία δομείται ένα πρόβλημα καθώς και της ροής των εργασιών που εκτελούνται κατά την υλοποίηση μιας πολυμεταβλητής ανάλυσης μιας διαδικασίας παραγωγής σε παρτίδες με τη βοήθεια του λογισμικού SIMCA. Αντίστοιχη λογική ακολουθείται στην πρακτική εφαρμογή την οποία πραγματεύεται η παρούσα εργασία και παρουσιάζεται σε επόμενο κεφάλαιο.

## 4.2 Γενικά

Είναι προφανές ότι το βασικότερο δομικό στοιχείο ενός προβλήματος ανάλυσης δεδομένων είναι το διαθέσιμο σύνολο δεδομένων αυτό καθαυτό. Η δομή του διαθέσιμου συνόλου δεδομένων καθώς και ο τρόπος με τον οποίο επιθυμούμε να το χειριστούμε και το είδος των συμπερασμάτων στα οποία θέλουμε να καταλήξουμε μας καθοδηγεί σε μεγάλο βαθμό στην επιλογή των τεχνικών που θα εφαρμόσουμε. Όπως ήδη αναφέρθηκε, το λογισμικό SIMCA εστιάζει κυρίως στις μεθόδους της Ανάλυσης Κυρίων Συνιστωσών και της Παλινδρόμησης Μερικών Ελαχίστων Τετραγώνων.

➤ Ανάλυση Κυρίων Συνιστωσών:

Στην περίπτωση που το σύνολο δεδομένων αποτελείται από ένα μπλοκ δεδομένων το οποίο απαρτίζεται από τις παρατηρήσεις των ανεξάρτητων μεταβλητών, ενδείκνυται η χρήση της μεθόδου PCA.

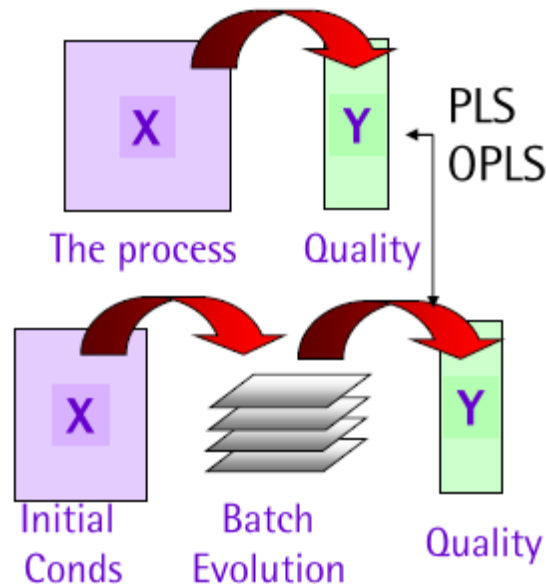


**Εικόνα 4-1** Σύνολο δεδομένων αποτελούμενο από ένα μπλοκ δεδομένων (Simca Documentation)

Στην περίπτωση αυτή, το ζητούμενο είναι η δυνατότητα αναγνώρισης και διάκρισης μεταξύ συνόλων δεδομένων, τα οποία αντλούνται από επιτυχημένες διαδικασίες παραγωγής και από διαδικασίες οι οποίες είχαν σαν αποτέλεσμα την αστοχία του τελικού προϊόντος. Θα πρέπει δηλαδή, μέσω της εκπαίδευσης ενός μοντέλου με τη χρήση δεδομένων από επιτυχημένες διαδικασίες παραγωγής, να είμαστε σε θέση να εντοπίσουμε μια διαδικασία η οποία έχει αποκλίνει από τις κανονικές και επιτρεπόμενες συνθήκες λειτουργίας. Στη συνέχεια, οι αποκλίσεις που εντοπίζονται θα πρέπει να ερμηνεύονται και να αποκαλύπτουν τα αίτια της αστοχίας, προκειμένου αυτά να αποφεύγονται στο μέλλον.

➤ Παλινδρόμηση Μερικών Ελαχίστων Τετραγώνων:

Στην περίπτωση που το σύνολο δεδομένων αποτελείται από δύο μπλόκ δεδομένων το πρώτο εκ των οποίων απαρτίζεται από τις παρατηρήσεις των ανεξάρτητων μεταβλητών και το δεύτερο απαρτίζεται από τις παρατηρήσεις των μεταβλητών απόκρισης, ενδείκνυται η χρήση της μεθόδου PLS.

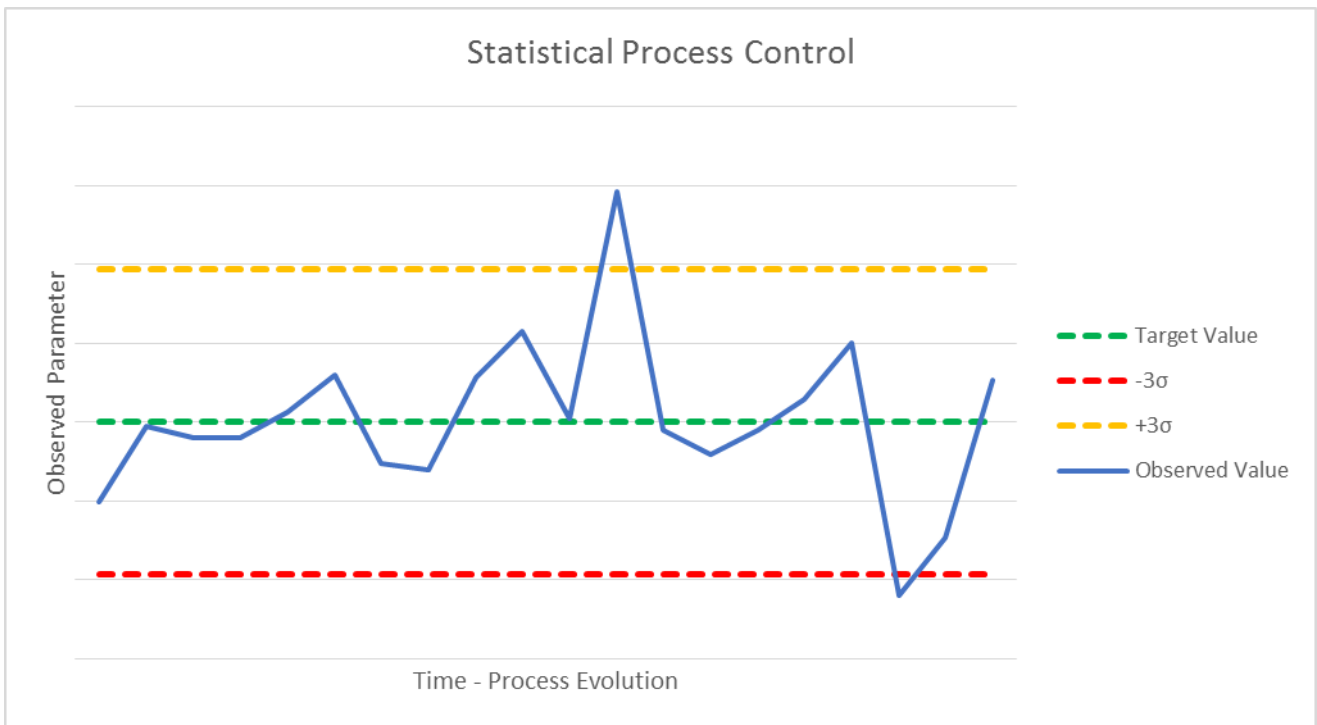


Εικόνα 4-2 Σύνολο δεδομένων αποτελούμενο από δυο μπλοκ δεδομένων (Simca Documentation).

Ειδικά δε στην περίπτωση της παραγωγής σε παρτίδες, το σύνολο δεδομένων έχει και μια τρίτη διάσταση, η οποία εκφράζει τον αριθμό των παρτίδων, ενώ συχνά λαμβάνονται υπόψη και οι αρχικές συνθήκες πριν την έναρξη της διαδικασίας. Σε αυτή την περίπτωση το ζητούμενο είναι η κατασκευή ενός μοντέλου το οποίο θα αξιοποιεί τις πληροφορίες που εμπεριέχονται στο μπλοκ δεδομένων των ανεξάρτητων μεταβλητών προκειμένου να παράγει προβλέψεις για τις μεταβλητές εξόδου της διαδικασίας.

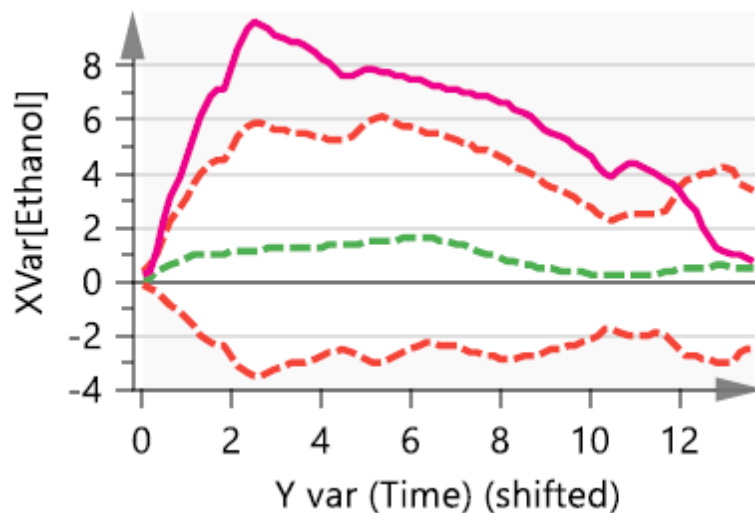
Έτσι, αναλόγως με τον τρόπο που δομείται το σύνολο δεδομένων μας, επιλέγουμε ποια από τις τεχνικές θα εφαρμόσουμε.

Το βασικό εργαλείο για τον έλεγχο της εξέλιξης παραγωγικών διαδικασιών είναι ο Στατιστικός έλεγχος ποιότητας και το διάγραμμα Shewhart, όπου ελέγχεται η εξέλιξη κάποιας παραμέτρου της διαδικασίας έναντι κάποιας τιμής στόχου και κάποιων άνω και κάτω ορίων εκτός των οποίων θεωρείται ότι η συγκεκριμένη παράμετρος έχει αστοχήσει. Η συγκεκριμένη μέθοδος είναι πολύ εύχρηστη σε προβλήματα μιας μεταβλητής και μας επιτρέπει το γρήγορο εντοπισμό και την αντιμετώπιση τυχόν αποκλίσεων.



Εικόνα 4-3 Στατιστικός Έλεγχος Ποιότητας – Διάγραμμα Shewhart

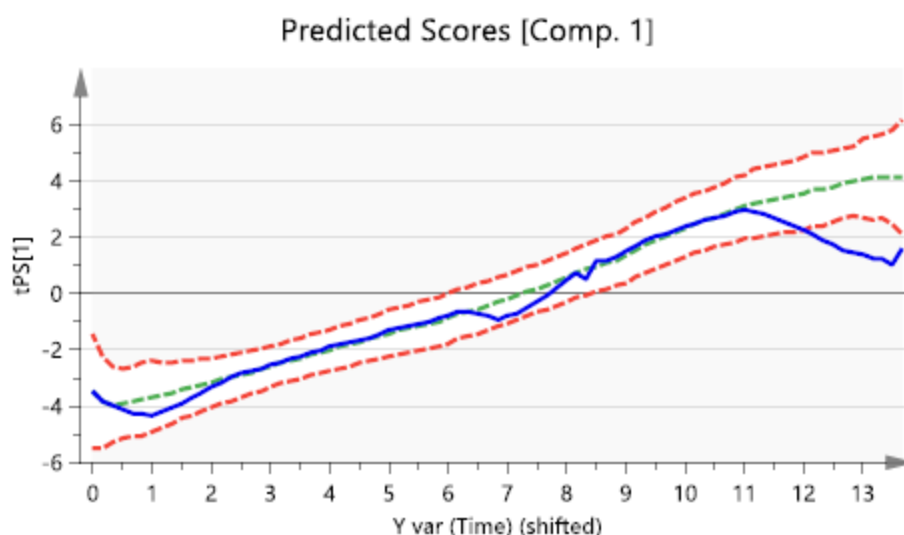
Αντίστοιχες οπτικοποιήσεις παρέχονται από το SIMCA για κάθε μια εκ των ανεξάρτητων μεταβλητών του προβλήματος. Τα εν λόγω γραφήματα αποτελούν μια κοντινή και λεπτομερή ματιά στην εξέλιξη της κάθε μεταβλητής.



Εικόνα 4-4 Παράδειγμα γραφικής απεικόνισης της εξέλιξης μιας μεταβλητής (SIMCA Documentations)

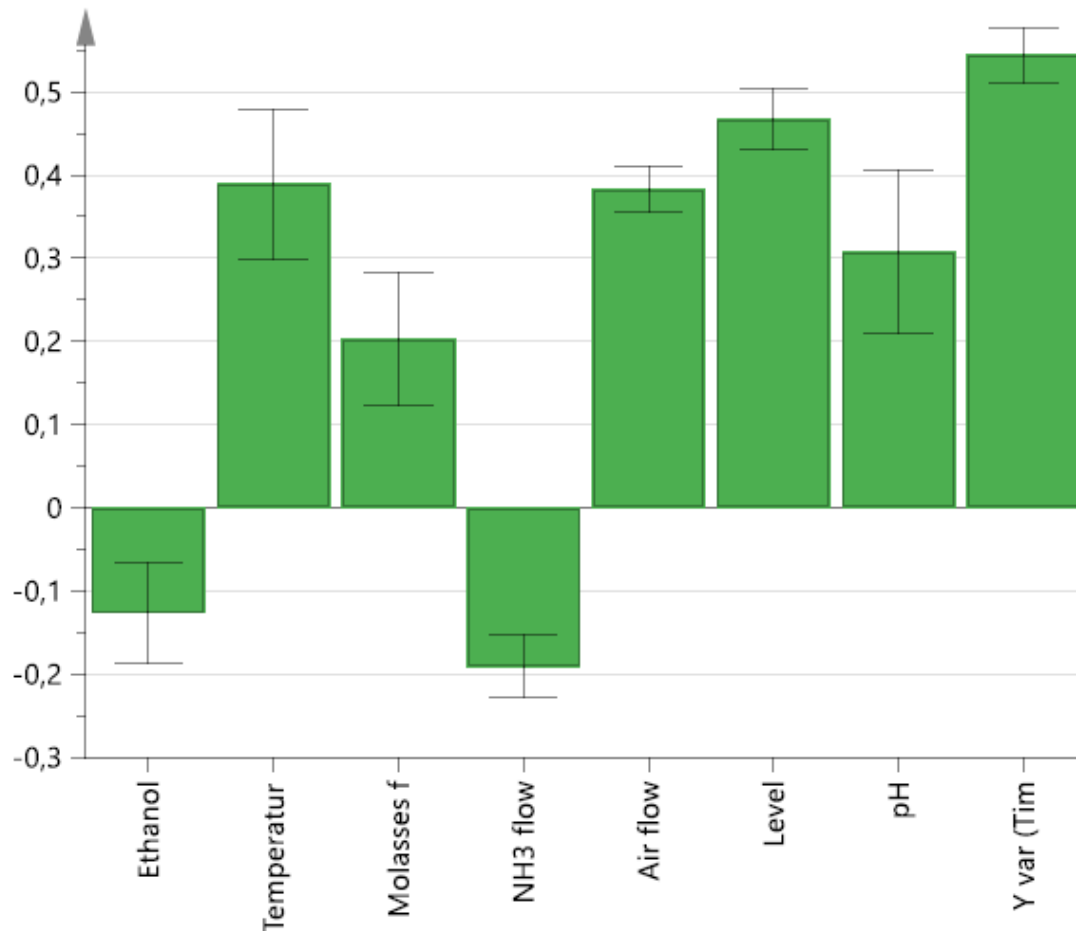
Ωστόσο, γίνεται εύκολα αντιληπτό ότι, όταν έχουμε να αντιμετωπίσουμε πολυμεταβλητά προβλήματα, αφενός η παρακολούθηση πολλαπλών τέτοιων διαγραμμάτων γίνεται δύσχρηστη και αφετέρου, όπως έχει ήδη εξηγηθεί σε προηγούμενο κεφάλαιο, θα πρέπει να είμαστε σε θέση να κρίνουμε ποιες από τις εντοπιζόμενες αποκλίσεις είναι πιθανό να επηρεάσουν τη ποιότητα του τελικού προϊόντος και σε τι βαθμό, ώστε να μπορούμε να εστιάσουμε καλύτερα τις προσπάθειες βελτίωσης της διαδικασίας παραγωγής στις πραγματικά κρίσιμες πτυχές του προβλήματος.

Μια χρήσιμη οπτικοποίηση την οποία παρέχει το Simca για να αντιμετωπίσει αυτό το πρόβλημα και η οποία αποτελεί μια περίληψη της εξέλιξης των παρακολουθούμενων παραμέτρων κατά τη διάρκεια της διαδικασίας παραγωγής, είναι η γραφική απεικόνιση της εξέλιξης των κυρίων συνιστωσών του προβλήματος.



**Εικόνα 4-5 Η προβλεπόμενη εξέλιξη της πρώτης κύριας συνιστώσας ενός πολυμεταβλητού προβλήματος (SIMCA Documentation)**

Επιπλέον, προκειμένου να είμαστε σε θέση να αξιολογήσουμε την επίδραση της κάθε μιας εκ των παρατηρούμενων παραμέτρων του προβλήματος, το SIMCA μας παρέχει τη δυνατότητα δημιουργίας γραφημάτων σημαντικότητας των μεταβλητών (VIP – Variable Importance Plot), τα οποία λαμβάνουν υπόψη τους τα βάρη με τα οποία συμμετέχει η κάθε μεταβλητή στη σύνθεση των κυρίων συνιστωσών του προβλήματος.



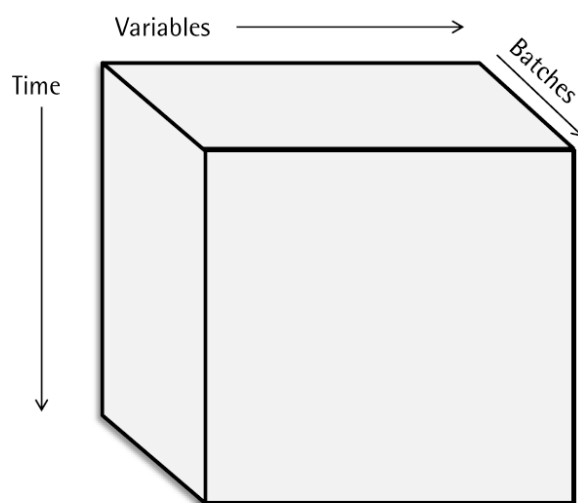
Εικόνα 4-6 Παράδειγμα γραφήματος σημαντικότητας μεταβλητών (SIMCA Documentation)

Στην περίπτωση που εξετάζουμε μια διαδικασία παραγωγής σε παρτίδες, η πολυπλοκότητα του προβλήματος αυξάνεται περαιτέρω λόγω της μιας επιπλέον διάστασης που προστίθεται σε αυτό. Στην ενότητα που ακολουθεί παρουσιάζεται η προσέγγιση του SIMCA στην πολυμεταβλητή ανάλυση διαδικασιών παραγωγής σε παρτίδες.

## 4.3 Εφαρμογή στην Παραγωγή σε Παρτίδες

### 4.3.1 Γενικά

Όπως έχει ήδη αναφερθεί, μια από τις ιδιαιτερότητες των προβλημάτων που έχουν να κάνουν με την ανάλυση διαδικασιών παραγωγής σε παρτίδες, είναι η δομή του συνόλου των δεδομένων αυτού καθαυτού. Το σύνολο αυτό αποτελείται από επιμέρους υποσύνολα δεδομένων, ένα για κάθε μια από τις εξεταζόμενες παρτίδες.

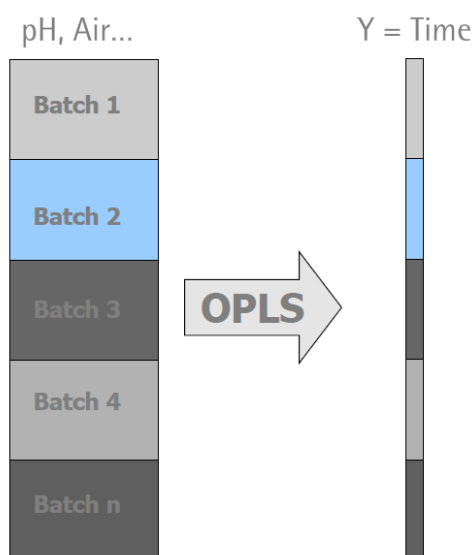


Εικόνα 4-7 Οι τρεις διαστάσεις του Συνόλου Δεδομένων (SIMCA Documentation)

Κάθε ένα από τα προαναφερθέντα επιμέρους υποσύνολα απαρτίζεται από στήλες οι οποίες αντιπροσωπεύουν τις παρατηρούμενες μεταβλητές του προβλήματος, ανεξάρτητες και εξαρτημένες, ενώ κάθε γραμμή αντιπροσωπεύει ένα σημείο στο χρόνο, ένα στιγμιότυπο δηλαδή της διαδικασίας παραγωγής, στο οποίο λαμβάνονται μετρήσεις για κάθε μια εκ των μεταβλητών. Θεωρούμε λοιπόν, ότι οι μετρήσεις όλων των παραμέτρων του προβλήματος λαμβάνονται με την ίδια συχνότητα και συνεπώς στο τέλος της παρτίδας θα έχουμε τον ίδιο αριθμό παρατηρήσεων για κάθε μια εκ των παραμέτρων. Καλούμαστε λοιπόν να συνδυάσουμε όλα αυτά τα διαθέσιμα δεδομένα με έναν αποτελεσματικό και εύληπτο τρόπο, ώστε να καταλήξουμε σε χρήσιμα συμπεράσματα. Το λογισμικό SIMCA προσεγγίζει την πρόκληση αυτή μέσω δύο διαφορετικών μοντέλων, ο συνδυασμός των οποίων επιτρέπει την όσο το δυνατόν αποτελεσματικότερη αξιοποίηση των δεδομένων μας. Τα μοντέλα αυτά παρουσιάζονται ακολούθως.

### 4.3.2 Batch Evolution Model

Η πρώτη προσέγγιση την οποία υιοθετεί το SIMCA για την ανάλυση συνόλων δεδομένων που προέρχονται από διαδικασίες παραγωγής σε παρτίδες, είναι αυτή του Μοντέλου Εξέλιξης της Παρτίδας (BEM). Σε αυτό το μοντέλο, το σύνολο δεδομένων διατάσσεται έτσι ώστε, κάθε γραμμή να αντιπροσωπεύει ένα σημείο στο χρόνο και κάθε στήλη αντιπροσωπεύει μια από τις παρατηρούμενες παραμέτρους. Ως μεταβλητή απόκρισης της διαδικασίας επιλέγεται μια μεταβλητή η οποία να περιγράφει την πρόοδο, την ωριμότητα και τελικά την εξέλιξη της παρτίδας. Η μεταβλητή αυτή συνήθως είναι ο χρόνος. Στη συνέχεια εφαρμόζεται η μέθοδος των Μερικών Ελάχιστων Τετραγώνων (PLS).

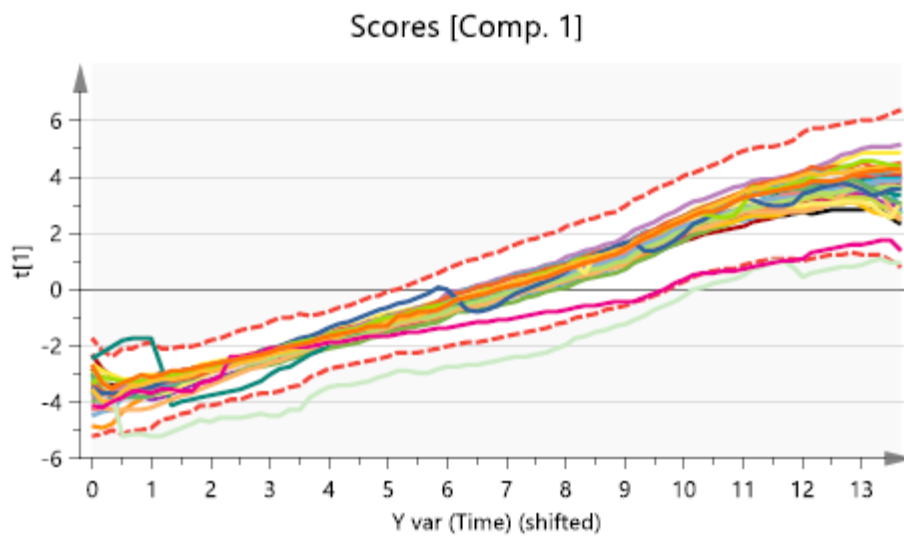


Εικόνα 4-8 Δομή του μοντέλου BEM (SIMCA Documentation)

Θεωρώντας ότι το σύνολο δεδομένων του προβλήματος απαρτίζεται τόσο από επιτυχείς όσο και από παρτίδες των οποίων τα ποιοτικά χαρακτηριστικά αποκλίνουν τα επιθυμητά, το επόμενο βήμα στην κατασκευή του μοντέλου είναι η απομόνωση των παρτίδων με κακές επιδόσεις. Αρχικά πραγματοποιείται μια πρώτη μοντελοποίηση η οποία λαμβάνει υπόψη της όλα τα δεδομένα. Στη συνέχεια με τη βοήθεια των διαγραμμάτων ελέγχου των παρτίδων που προκύπτουν, στα οποία παρατίθενται οι γραφικές απεικονίσεις της εξέλιξης στο χρόνο των κυρίων συνιστωσών όλων των παρτίδων (scores), είναι δυνατός ο εντοπισμός εκείνων των παρτίδων που παρουσιάζουν αποκλίνουσα συμπεριφορά σε σύγκριση με τη «μέση παρτίδα» και με τα αντίστοιχα άνω και κάτω όρια τα οποία επιλέγονται ως κάποιο πολλαπλάσιο της

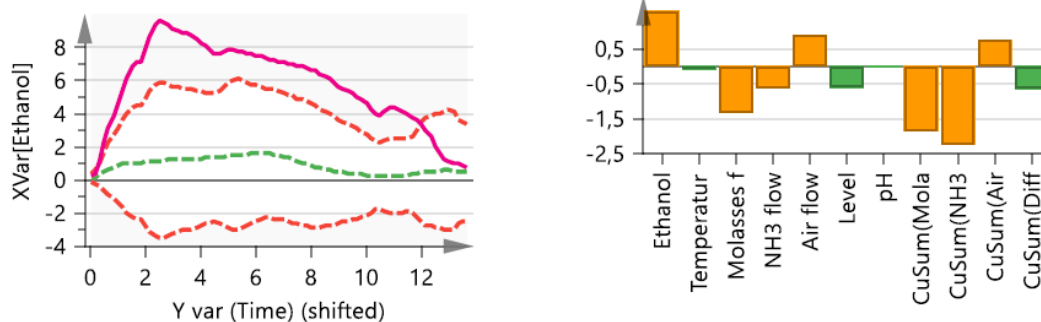


τυπικής απόκλισης (πχ.  $\pm 3\sigma$ ). Τα υποσύνολα δεδομένων που αφορούν τις εν λόγω παρτίδες μπορούν στη συνέχεια να απομακρυνθούν από το σύνολο δεδομένων ενώ επίσης δίνεται η δυνατότητα διερεύνησης των λόγων για τους οποίους οι παρτίδες αυτές αστόχησαν.



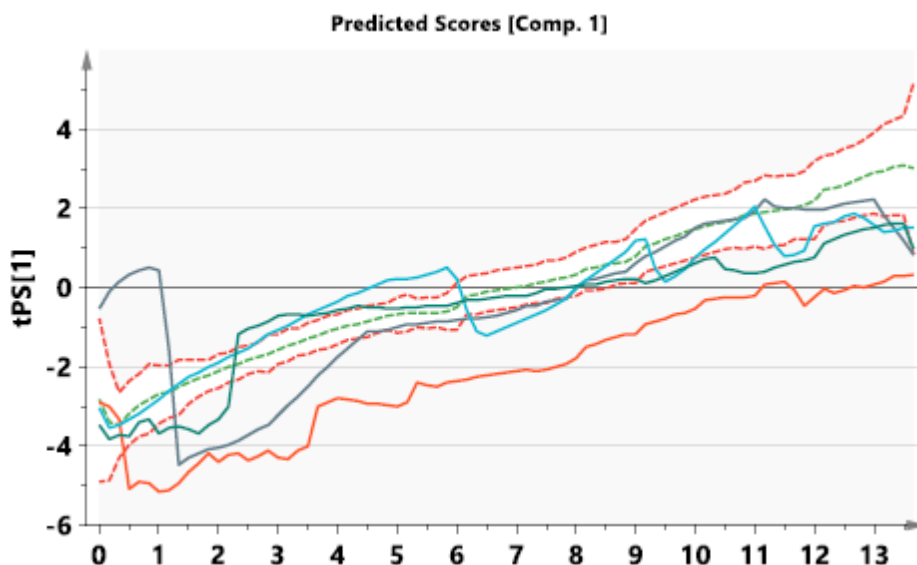
Εικόνα 4-9 Παράδειγμα Batch Control Chart (SIMCA Documentation)

Έτσι, με τη βοήθεια των διαγραμμάτων συμμετοχής (contribution plots) καθώς και μέσω της προσεκτικής και λεπτομερούς εξέτασης της εξέλιξης των επιμέρους παραμέτρων κατά τη διάρκεια της διαδικασίας, σε σύγκριση με την εξέλιξη των αντίστοιχων παραμέτρων, όπως αυτή παρατηρείται στις επιτυχείς παρτίδες, μπορούμε να διακρίνουμε ποιες από τις παραμέτρους, σε τι βαθμό και προς ποια κατεύθυνση αποκλίνουν από τις αναμενόμενες τιμές τους κάποια χρονική στιγμή, για κάποια από τις παρτίδες που αστόχησαν.



Εικόνα 4-10 Παράδειγμα Variable Control Chart (δεξιά) και Contribution Plot (αριστερά)

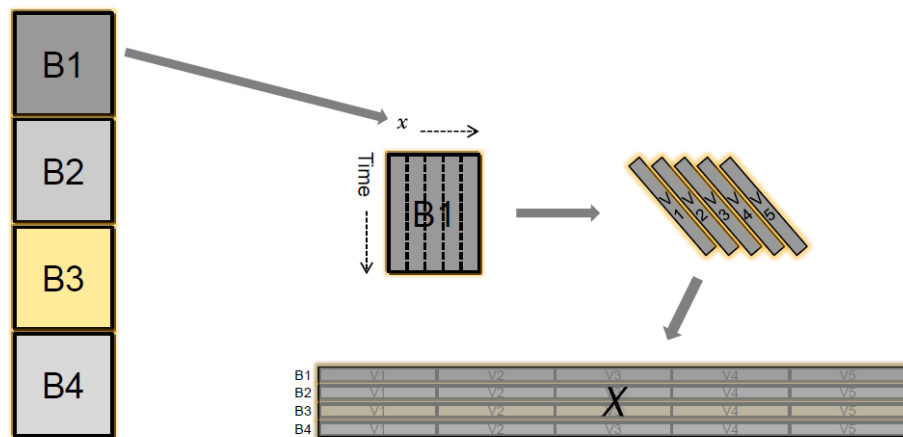
Στη συνέχεια, αφού απομακρυνθούν οι αποκλίνουσες παρτίδες, το μοντέλο μπορεί να κατασκευαστεί εκ νέου, χωρίς πλέον να επηρεάζεται από δεδομένα τα οποία μπορεί να προέρχονται από κακές μετρήσεις ή από παρτίδες με κακές επιδόσεις. Από τη διαδικασία αυτή προκύπτει το τελικό μοντέλο που εκφράζει την ομαλή εξέλιξη της εξεταζόμενης παραγωγικής διαδικασίας. Τα δεδομένα που απομακρύνθηκαν από το αρχικό σύνολο, μπορούν τώρα να χρησιμοποιηθούν για την επαλήθευση του μοντέλου.



Εικόνα 4-11 Επαλήθευση μοντέλου και εντοπισμός παρτίδων που προβλέπεται να αστοχήσουν (SIMCA Documentation)

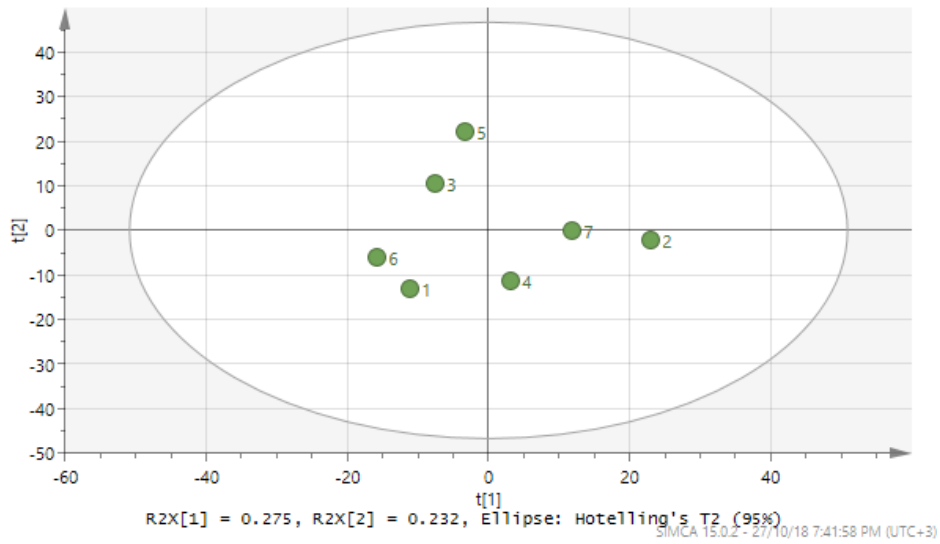
### 4.3.3 Batch Level Model

Ο δεύτερος τρόπος με τον οποίο προσεγγίζεται η ανάλυση μιας πολυμεταβλητής διαδικασίας παραγωγής σε παρτίδες, είναι αυτός του Μοντέλου Παρτίδας (BLM). Κατά την κατασκευή του εν λόγω μοντέλου, το σύνολο δεδομένων διατάσσεται με τρόπο ώστε κάθε γραμμή να αντιπροσωπεύει μια παρτίδα. Συγκεκριμένα, όλες οι μετρήσεις της κάθε μιας εκ των παρατηρούμενων παραμέτρων, που ελήφθησαν κατά τη διάρκεια της διαδικασίας παραγωγής μιας συγκεκριμένης παρτίδας, διατάσσονται με τρόπο ώστε να σχηματίζουν ένα διάνυσμα-γραμμή. Στη συνέχεια τα επιμέρους διανύσματα που προκύπτουν, διατάσσονται το ένα δίπλα στο άλλο. Η διαδικασία αυτή επαναλαμβάνεται για κάθε μια εκ των εξεταζόμενων παρτίδων. Κατ' αυτό τον τρόπο, κάθε γραμμή του συνόλου δεδομένων απαρτίζεται από όλες τις διαθέσιμες πληροφορίες για κάθε παρτίδα, διατεταγμένες χρονικά και ανά παρτίδα.



**Εικόνα 4-12 Διάταξη του συνόλου δεδομένων για την κατασκευή του μοντέλου BLM  
(SIMCA Documentation)**

Στη συνέχεια, το αναδιατεταγμένο σύνολο δεδομένων χρησιμοποιείται για την εφαρμογή της μεθόδου της Ανάλυσης Κυρίων Συνιστωσών (PCA). Ουσιαστικά, αν θεωρήσουμε ότι έχουμε  $n$  παρατηρήσεις για κάθε μια εκ των  $m$  μεταβλητών σε κάθε παρτίδα, τότε όλες οι  $k$ -οστές παρατηρήσεις της μεταβλητής  $V_i$ , όπου  $k=1\dots n$  και  $i=1\dots m$ , συνθέτουν μια νέα μεταβλητή με μήκος ίσο με τον αριθμό των παρτίδων, η οποία μετά από την εφαρμογή της PCA θα μετασχηματιστεί στο χώρο των Κυρίων Συνιστωσών σύμφωνα με τα όσα παρουσιάστηκαν στο Κεφάλαιο 3. Κατ' αυτόν τον τρόπο μετά από την ολοκλήρωση της διαδικασίας, κάθε παρτίδα θα μπορεί να αναπαρίσταται από ένα σημείο στο διάγραμμα αποτελεσμάτων (score plot), οι συντεταγμένες του οποίου θα είναι τα scores των δύο πρώτων και σημαντικότερων κύριων συνιστωσών.



**Εικόνα 4-13 Απεικόνιση 7 παρτίδων στο Score Plot**

Με τη βοήθεια αυτών των απεικονίσεων, μας δίνεται η δυνατότητα να διακρίνουμε συστάδες παρτίδων οι οποίες παρουσιάζουν ομοιότητες καθώς και να εντοπίσουμε παρτίδες οι οποίες αποκλίνουν από την αναμενόμενη συμπεριφορά.

## ΚΕΦΑΛΑΙΟ 5: *Εφαρμογή*

---

### 5.1 Η Διαδικασία Παραγωγής

#### 5.1.1 Γενικά

Όπως έχει ήδη αναφερθεί, αφετηρία για την εκπόνηση της παρούσας εργασίας υπήρξε η ανάγκη για τη μελέτη, την βαθιά κατανόηση και τελικά τη βελτίωση μιας πραγματικής βιομηχανικής διαδικασίας παραγωγής, μέσω της αξιοποίησης δεδομένων, τόσο ιστορικών όσο και «on line», που σχετίζονται με τη διαδικασία αυτή.

Συγκεκριμένα, στο πλαίσιο της εργασίας, το απαραίτητο πεδίο για την εφαρμογή των όσων παρουσιάστηκαν στα προηγούμενα κεφάλαια, καθώς και τα αντίστοιχα μέσα και τους ανάλογους πόρους, παρείχε γνωστή πολυεθνική εταιρεία καλλυντικών στη Νότια Ευρώπη.

Στο εργοστάσιο της εν λόγω εταιρείας παράγονται προϊόντα ομορφιάς και προσωπικής περιποίησης με τη μέθοδο της παραγωγής σε παρτίδες. Κάθε γραμμή παραγωγής αποτελείται από πολλά μηχανήματα, το σημαντικότερο από τα οποία είναι το δοχείο ανάμιξης, όπου δύναται να παραχθεί μια μεγάλη ποικιλία προϊόντων, αναλόγως με τη ζήτηση της αγοράς. Η ανάμιξη είναι ένα σημαντικό βήμα της παραγωγικής διαδικασίας, καθώς είναι το στάδιο στο οποίο δημιουργείται το τελικό προϊόν. Το εκάστοτε δοχείο παρακολουθείται και ελέγχεται από ένα PLC, το οποίο συλλέγει τα δεδομένα, τα οποία καλούμαστε να αναλύσουμε.

Συγκεκριμένα, στο πλαίσιο της παρούσας εργασίας, τα δεδομένα τα οποία θα αναλύσουμε προέρχονται από ένα συγκεκριμένο δοχείο ανάμιξης, το οποίο βρίσκεται στη γραμμή παραγωγής παρτίδων γαλακτωμάτων.

#### 5.1.2 Περιγραφή

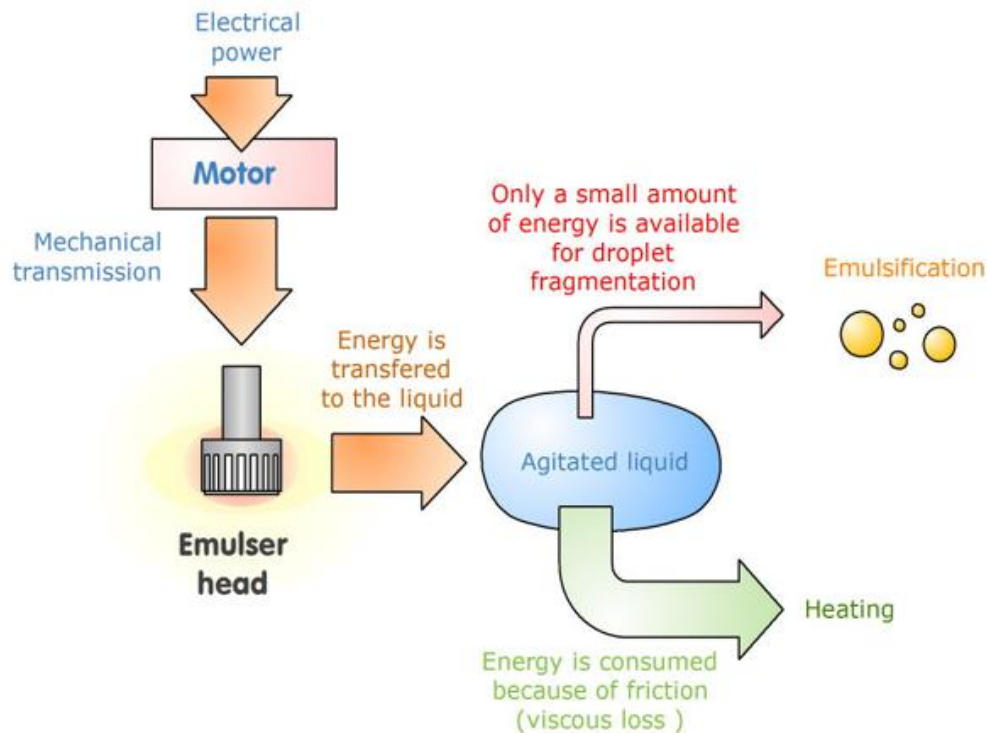
Η παραγωγική διαδικασία ουσιαστικά χωρίζεται σε δύο βασικά τμήματα, την παραγωγή αυτή καθ' αυτή των παρτίδων προϊόντος και τη διαδικασία του πλυσίματος/απολύμανσης των μηχανημάτων για εκ νέου παραγωγή προϊόντος άλλης σύστασης.

Η παραγωγή μιας παρτίδας αποτελείται από τα διακριτά βήματα της Ανάμιξης (Mixing), όπου αναμιγνύονται οι πρώτες ύλες και του Γεμίσματος (Filling), όπου το τελικό προϊόν, αφότου πρώτα ζυγιστεί, διοχετεύεται στα μπουκάλια.

Στην παρούσα εργασία εστιάζουμε κυρίως στη διαδικασία της Ανάμιξης, κατά τη διάρκεια της οποίας, μια σειρά από διεργασίες επιτελούνται μέσα στο δοχείο παραγωγής, με σκοπό να παραχθεί το τελικό προϊόν. Τέτοιες επί μέρους διεργασίες είναι η θέρμανση/ψύξη του μείγματος, η ανάδευση, η εισαγωγή νερού και πρώτων υλών κ.α. Επιπλέον επιτελούνται, δευτερευόντως, επί μέρους ενέργειες, όπως συντήρηση μηχανημάτων, δοκιμές νέων προϊόντων ή μηχανημάτων κ.α. Η σημαντικότερη επί μέρους διεργασία της διαδικασίας της Ανάμιξης είναι αυτή της γαλακτωματοποίησης, η δημιουργία και ανάμιξη δηλαδή, γαλακτωμάτων.

Η γαλακτωματοποίηση είναι η διαδικασία κατά τη διάρκεια της οποίας, δαπανάται ενέργεια με σκοπό τη μετατροπή των σταγόνων ελαίων σε μικρότερα σταγονίδια ώστε να καθίσταται δυνατή η ανάμιξή τους με νερό και άλλες πρώτες ύλες. Η δαπάνη ενέργειας και η άσκηση κάποιας μηχανικής δύναμης, όπως για παράδειγμα η ανάδευση, επιδιώκει τον διαχωρισμό των ελαίων σε μικρά σταγονίδια ούτως ώστε να υπερνικηθεί η επιφανειακή τάση των αναμιγνυόμενων υγρών, δίνοντας τους έτσι τη δυνατότητα να αυξήσουν την έκταση της επιφάνειας επαφής τους και τελικά να σχηματιστεί ένα όσο το δυνατόν σταθερότερο διάλυμα.

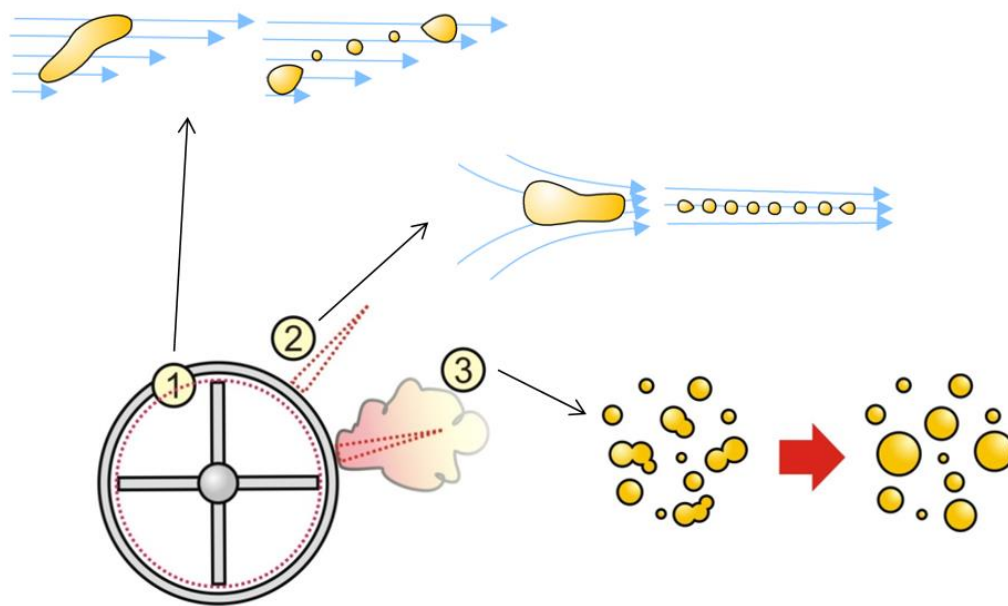
Υδροδυναμικές ροές οδηγούν στη διάσπαση των σταγονιδίων, η οποία υλοποιείται είτε μέσω ελαστικής διάτμησης είτε μέσω διάτμησης λόγω επιμήκυνσης. Η δαπανώμενη ενέργεια θα πρέπει να δημιουργήσει τόση υδροδυναμική ροή, όση είναι αναγκαία, ώστε να αποφεύγεται η επανένωση των σταγονιδίων. Η διαδικασία αυτή μπορεί να κρατήσει από μερικά λεπτά έως μερικές ώρες, αναλόγως με τις ποσότητες των πρώτων υλών που αναμιγνύονται.



Εικόνα 5-1: Διαδικασία γαλακτωματοποίησης

Στην εξεταζόμενη γραμμή παραγωγής, χρησιμοποιούνται μίκτες με ομογενοποιητές υψηλής πίεσης και καύσης, για την υλοποίηση συμπυκνωμένης γαλακτωματοποίησης. Τα δοχεία αποτελούνται από έναν κινητήρα που μεταφέρει ενέργεια μέσω μιας κεφαλής γαλακτωματοποιητή στο υγρό. Μόνο ένα μικρό μέρος της δαπανώμενης ενέργειας χρησιμοποιείται για τη διάσπαση των σταγονιδίων, ενώ το μεγαλύτερο μέρος της δαπανάται σε απώλειες λόγω τριβών.

Αυλακωτός μύλος προσαρμόζεται στο ρότορα του κινητήρα και δημιουργεί όλες τις απαραίτητες προϋποθέσεις για να ολοκληρωθεί η διάσπαση σταγονιδίων (ελαστική διάτμηση, διάτμηση μέσω επιμήκυνσης, υδροδυναμικό πεδίο). Το υγρό κατευθύνεται στο ρότορα και μέσω της κίνησής του οδηγείται, μέσα από τις αυλακώσεις, στο υδροδυναμικό πεδίο, που έχει ήδη δημιουργήσει μέσω της κίνησής του.



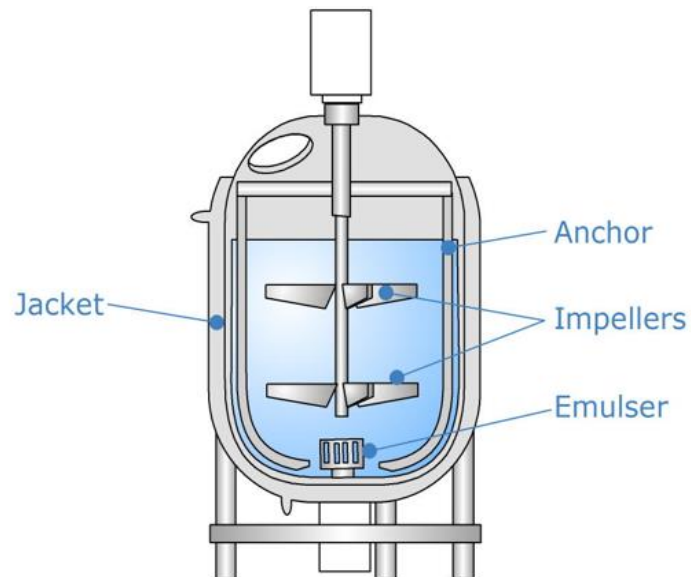
Εικόνα 5-2: Τρόποι διάσπασης ελαίων σε σταγονίδια

Η όλη διαδικασία συνίσταται στην εισαγωγή πρώτων υλών στο δοχείο, στην διάσπαση των ελαίων και τελικά στην ανάμιξη των πρώτων υλών και υλοποιείται με τον ακόλουθο εξοπλισμό:

- Τυπικός αντιδραστήρας παρτίδων (δοχείο ανάμιξης) με δυνατότητα γαλακτωματοποίησης (Batch Reactor).
- Αντλία εισαγωγής πρώτων υλών εντός του αντιδραστήρα (Product pump). Πρόκειται για μια λοβωτή αντλία η οποία βρίσκεται στο κύκλωμα σωληνώσεων που συνδέει τον πυθμένα με την έξοδο αλλά και με το άνω μέρος του δοχείου. Ο ρόλος της είναι τόσο η εισαγωγή πρώτων υλών όσο και η ανακυκλοφορία του μίγματος, από κάτω προς τα πάνω, για ταχύτερη ομογενοποίηση του υγρού.
- Άγκυρα (Anchor). Αναδευτήρας, ο ρόλος του οποίου είναι η ομοιόμορφη ανάδευση του μίγματος πλησίον των τοιχωμάτων του αντιδραστήρα και η απομάκρυνση υλών από αυτά.
- Εσωτερικός αναδευτήρας (Spiral) με πτερύγια (implellers), για την ανάδευση, και ανάμιξη του μίγματος.
- Ομογενοποιητής ή κεφαλή γαλακτωματοποίησης (emulsifier head) για τη διάσπαση των ελαίων σε σταγονίδια και τη διασπορά του υγρού.



- Σύστημα ψύξης/θέρμανσης για την ρύθμιση της θερμοκρασίας εντός του αντιδραστήρα. Η ρύθμιση της θερμοκρασίας επιτυγχάνεται με την παροχή ατμού ή νερού εντός διακένου (Jacket), περιμετρικά του δοχείου ανάμιξης.



**Εικόνα 5-3: Τυπικό δοχείο ανάμιξης – Batch Reactor**

Η διαδικασία παραγωγής της κάθε παρτίδας συνοψίζεται στην υλοποίηση εννέα τον αριθμό, διακριτών βημάτων, κάθε ένα από τα οποία με τη σειρά του, συνίσταται στην εκτέλεση επιμέρους χειρισμών. Τα εννέα διακριτά αυτά βήματα είναι τα εξής:

1. Εισαγωγή νερού
2. Εισαγωγή γλυκερίνης
3. Εισαγωγή λιπαρής φάσης
4. Γαλακτωματοποίηση
5. Έναρξη ψύξης
6. Δεύτερη Εισαγωγή νερού
7. Εισαγωγή αρώματος
8. Ρύθμιση φυσικοχημικών στοιχείων
9. Άδειασμα

### 5.1.3 Παρακολούθηση της διεργασίας / Monitoring

Ο έλεγχος και η παρακολούθηση της διαδικασίας παραγωγής πραγματοποιείται μέσω ενός ολοκληρωμένου και εξειδικευμένου συστήματος Βιομηχανικού Προγραμματιζόμενου Λογικού Ελεγκτή (PLC). Το εν λόγω σύστημα ελέγχου αποτελείται από έναν επεξεργαστή βιομηχανικού τύπου, ο οποίος αποτελεί την κεντρική μονάδα επεξεργασίας (CPU) και φέρει ενσωματωμένες κάρτες ψηφιακών εισόδων και εξόδων (DI/DO), από τη μονάδα τροφοδοσίας καθώς και από επιμέρους κάρτες ψηφιακών και αναλογικών εισόδων και εξόδων.

Οι διάφοροι χειρισμοί που απαιτούνται για την διεξαγωγή της διαδικασίας παραγωγής επιτελούνται μέσω εντολών του χειριστή προς το PLC. Παραδείγματα τέτοιων χειρισμών είναι τα ακόλουθα:

- Θέση της αντλίας πρώτων υλών σε λειτουργία
- Εισαγωγή πρώτων υλών
- Ρύθμιση των στροφών της αντλίας
- Έναρξη λειτουργίας ψύξης/θέρμανσης
- Ρύθμιση θερμοκρασίας
- Θέση της Άγκυρας σε λειτουργία
- Ρυθμιση στροφών Άγκυρας
- Θέση Spiral σε λειτουργία
- Ρύθμιση στροφών Spiral
- Θέση ομογενοποιητή σε λειτουργία
- Ρύθμιση στροφών ομογενοποιητή
- Ρύθμιση πίεσης δοχείου ανάμιξης
- Άνοιγμα – κλείσιμο βαννών

Επιπλέον, το σύστημα του PLC συλλέγει μετρήσεις προερχόμενες από μετρητικά όργανα καθώς και ενδείξεις που έχουν να κάνουν με την κατάσταση λειτουργίας των διαφορετικών τμημάτων του εξοπλισμού (πχ ενδείξεις ON/OFF χειροκίνητων βαννών).

Συγκεκριμένα οι μετρήσεις που λαμβάνονται και οι οποίες συνθέτουν το προς εξέταση σύνολο δεδομένων της διαδικασίας παραγωγής είναι οι εξής:

1. Θερμοκρασία εντός του δοχείου ανάμιξης (°C)
2. Πίεση εντός του δοχείου ανάμιξης (mbar)
3. Στροφές αναδευτήρα τύπου άγκυρας (RPM)
4. Στροφές αναδευτήρα τύπου spiral (RPM)
5. Στροφές ομογενοποιητή (RPM)
6. Στροφές αντλίας εισαγωγής πρώτων υλών (RPM)
7. Μάζα εισαγόμενης πρώτης ύλης (kg)
8. PH του μίγματος
9. Ιξώδες του μίγματος (cps)

### 5.1.4 Dataset

Το σύνολο δεδομένων που θα χρησιμοποιήσουμε στο πλαίσιο της παρούσας εργασίας συντίθεται από τη συλλογή μετρήσεων που ελήφθησαν και καταγράφηκαν από ένα συγκεκριμένο σύστημα PLC κατά τη διάρκεια της παραγωγής, οκτώ τον αριθμό διαφορετικών παρτίδων του ίδιου τελικού προϊόντος, σε ένα συγκεκριμένο δοχείο ανάμιξης. Οι επτά εκ των οκτώ παρτίδων, θεωρούνται επιτυχημένες, ενώ η μια δεν ανταποκρίνεται στα επιθυμητά χαρακτηριστικά ποιότητας. Συγκεκριμένα, το σύνολο δεδομένων μας αποτελείται από τις μετρήσεις εννέα τον αριθμό χαρακτηριστικών μεγεθών, όπως αυτά παρουσιάστηκαν στην αμέσως προηγούμενη παράγραφο, οι οποίες λαμβάνονται σε προκαθορισμένες χρονικές στιγμές κατά τη διάρκεια εξέλιξης της εκάστοτε παρτίδας. Για κάθε μια εκ των παρτίδων λαμβάνονται συνολικά 90 σετ μετρήσεων των 9 χαρακτηριστικών μεγεθών, τα οποία κατανέμονται στα επιμέρους στάδια της διαδικασίας παραγωγής, ως εξής:

A/A	Στάδιο Διαδικασίας Παραγωγής	Αριθμός σετ μετρήσεων	Σύνολικός Αριθμός Μετρήσεων
1	Εισαγωγή νερού	10	90
2	Εισαγωγή γλυκερίνης	2	18
3	Εισαγωγή λιπαρής φάσης	5	45
4	Γαλακτωματοποίηση	33	297
5	Έναρξη ψύξης	6	54
6	Δεύτερη Εισαγωγή νερού	6	54
7	Εισαγωγή αρώματος	4	36
8	Ρύθμιση φυσικοχημικών στοιχείων	4	36
9	Άδειασμα	20	180

Πίνακας 5-1 Κατανομή των λαμβανόμενων μετρήσεων κατά τη διαδικασία παραγωγής

Στη συνέχεια και σύμφωνα με τα παραπάνω, παρατίθεται ένα δείγμα του συνόλου δεδομένων, που αφορά το στάδιο της «Δεύτερης Εισαγωγής Νερού» της παρτίδας - 1:

A/A	Batch ID	Step	T (°C)	P (mbar)	Anchor RPM	Spiral RMP	Homo RPM	Product Pump RPM	Raw Material Mass (kg)	PH	Viscosity (cps)
57	1	6	55,72	-27,78	19,92	49,89	0,00	0,84	1,50	5,7	3100
58	1	6	52,29	-20,83	19,92	24,73	0,00	0,84	896,50	5,7	3100
59	1	6	53,30	-20,83	19,91	24,73	0,00	8,30	0,00	5,7	2823
60	1	6	46,91	-20,83	19,91	24,73	15,67	61,05	0,00	5,7	2823
61	1	6	45,05	-21,99	19,92	24,76	64,03	62,24	0,00	5,7	2823
62	1	6	45,83	-21,99	19,92	24,76	0,00	62,18	0,00	5,7	2823

Πίνακας 5-2 Δείγμα του Συνόλου Δεδομένων

### 5.1.5 Διατύπωση του Προβλήματος

Όπως έχει αναφερθεί και σε προηγούμενες ενότητες, ο σκοπός της ανάλυσης που θα εφαρμόσουμε είναι η αξιοποίηση ιστορικών δεδομένων, προερχόμενων από την παραγωγική διαδικασία, αφενός για την διερεύνηση της ύπαρξης, την αποκάλυψη και την αποκωδικοποίηση μοτίβων που διέπουν τη διαδικασία καθώς και συσχετίσεων μεταξύ των μεγεθών που χαρακτηρίζουν την εξέλιξή της και αφετέρου για την κατασκευή και εκπαίδευση ενός μοντέλου, το οποίο θα δύναται να προβλέπει τυχόν αποκλίσεις των ποιοτικών χαρακτηριστικών μιας παρτίδας από τις αποδεκτές τους τιμές, σε πραγματικό χρόνο, καθώς και να αναγνωρίζει και να υποδεικνύει εγκαίρως τις πιθανές αιτίες που θα μπορούσαν να οδηγήσουν στην αστοχία της παρτίδας. Η βαθύτερη κατανόηση αυτών των υποκείμενων διεργασιών που λαμβάνουν χώρα κατά τη διάρκεια της διαδικασίας παραγωγής, αποσκοπεί στην βελτιστοποίησή της κατά το βαθμό που αυτό είναι δυνατό και συνεπώς στη συνολική βελτίωση της παραγωγικότητας και τη μείωση του κόστους.

Στην προκειμένη περίπτωση, θα επιχειρήσουμε να αξιοποιήσουμε τα διαθέσιμα δεδομένα από οκτώ συνολικά παρτίδες, εκ των οποίων οι επτά θεωρούνται επιτυχημένες και θα χρησιμοποιηθούν για την κατασκευή και εκπαίδευση του μοντέλου μας. Η όγδοη παρτίδα, η οποία έχει αστοχήσει, θα χρησιμοποιηθεί, αφενός για τη δοκιμή του κατασκευασθέντος μοντέλου και αφετέρου για την άντληση πολύτιμων πληροφοριών σχετικά με τα αίτια που μπορούν να οδηγήσουν μια παρτίδα σε αστοχία.

## 5.2 Μοντελοποίηση

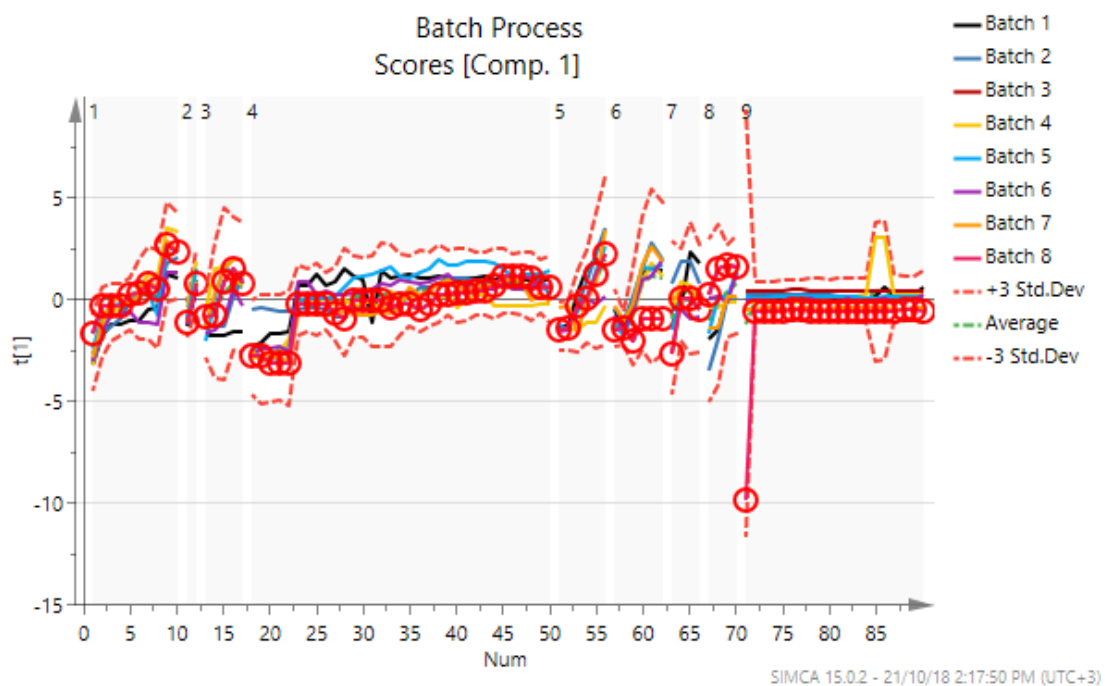
### 5.2.1 Επεξεργασία του Dataset

Προκειμένου να είμαστε σε θέση να χρησιμοποιήσουμε το σύνολο δεδομένων και να αντλήσουμε από αυτό αξιόπιστα συμπεράσματα, θα πρέπει πρώτα να το επεξεργαστούμε καταλλήλως. Το πρώτο στάδιο επεξεργασίας αφορά στον καθαρισμό του συνόλου δεδομένων από περιττές πληροφορίες. Το ανεπεξέργαστο σύνολο δεδομένων, όπως αυτό αντλείται από το PLC, περιλαμβάνει πληροφορίες που αφορούν στην επιλογή set points των διαφόρων ελεγχόμενων μεγεθών, από τους χειριστές, οι οποίες εμφανίζονται μεν σαν μετρήσεις, συνοδεύονται ωστόσο από μηνύματα τα οποία υποδεικνύουν την πραγματική τους φύση. Οι τιμές αυτές που αντιστοιχούν σε τιμές

στόχους, όπως αυτές καθορίζονται κατά τη διάρκεια της διαδικασίας, θα πρέπει να απομακρυνθούν από το σύνολο δεδομένων μας, καθώς δεν αποτελούν πραγματικές μετρήσεις. Το επόμενο στάδιο προεπεξεργασίας του συνόλου δεδομένων, αφορά στον εντοπισμό και τη συμπλήρωση τυχόν κενών τιμών ή «εσφαλμένων μηδενικών». Τέτοιες τιμές μπορεί, για παράδειγμα, να προκύψουν εξαιτίας κάποιας στιγμιαίας δυσλειτουργίας κάποιου μετρητικού οργάνου. Οι τιμές αυτές συνήθως επιλέγεται να συμπληρωθούν με τη μέθοδο της γραμμικής παρεμβολής. Τέλος, το λογισμικό SIMCA κανονικοποιεί με αυτοματοποιημένο τρόπο τα δεδομένα με τις τεχνικές του centering και scaling ώστε να είναι δυνατή η εφαρμογή των μεθόδων ανάλυσης.

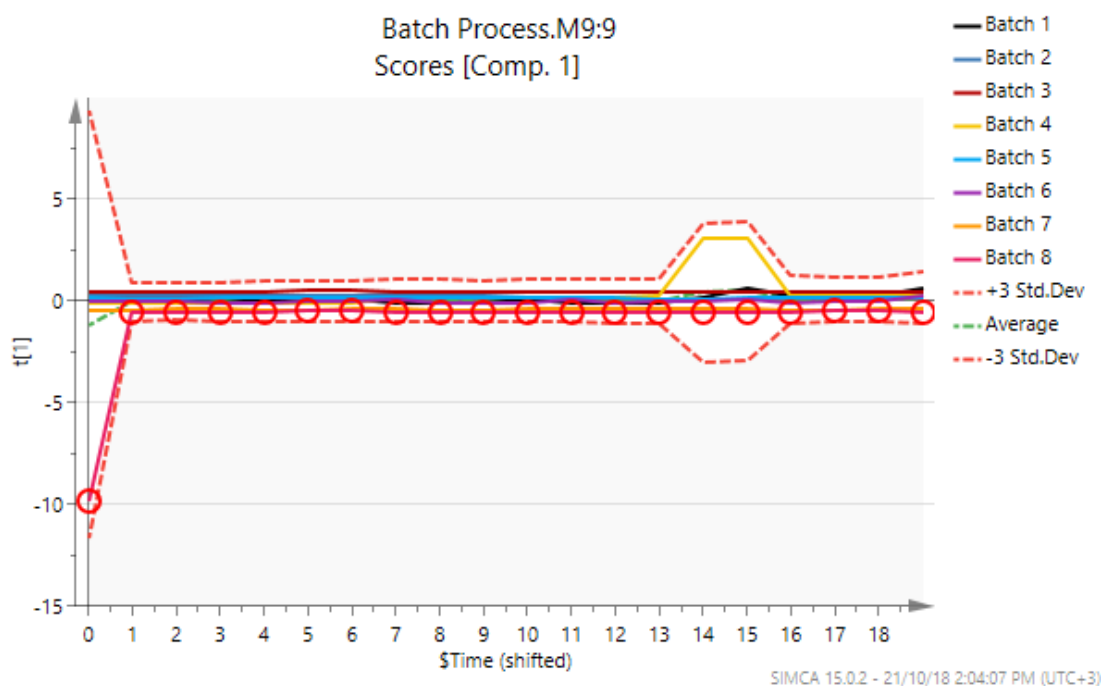
## 5.2.2 Κατασκευή και ερμηνεία των Μοντέλων BEM και BLM

Το πρώτο βήμα για την κατασκευή του μοντέλου έγκειται στην επιλογή των παρτίδων εκείνων οι οποίες θεωρούνται πετυχημένες. Για αυτό το σκοπό, όπως εξηγήθηκε και στο προηγούμενο κεφάλαιο, κατασκευάζεται σε πρώτη φάση ένα μοντέλο BEM, στο οποίο χρησιμοποιούνται τα δεδομένα όλων των παρτίδων. Το σύνολο  $X$  των ανεξάρτητων μεταβλητών συντίθεται από όλες τις μετρούμενες παραμέτρους, ενώ σαν μεταβλητή ενδεικτική της εξέλιξης της διαδικασίας παραγωγής λαμβάνεται ο χρόνος.



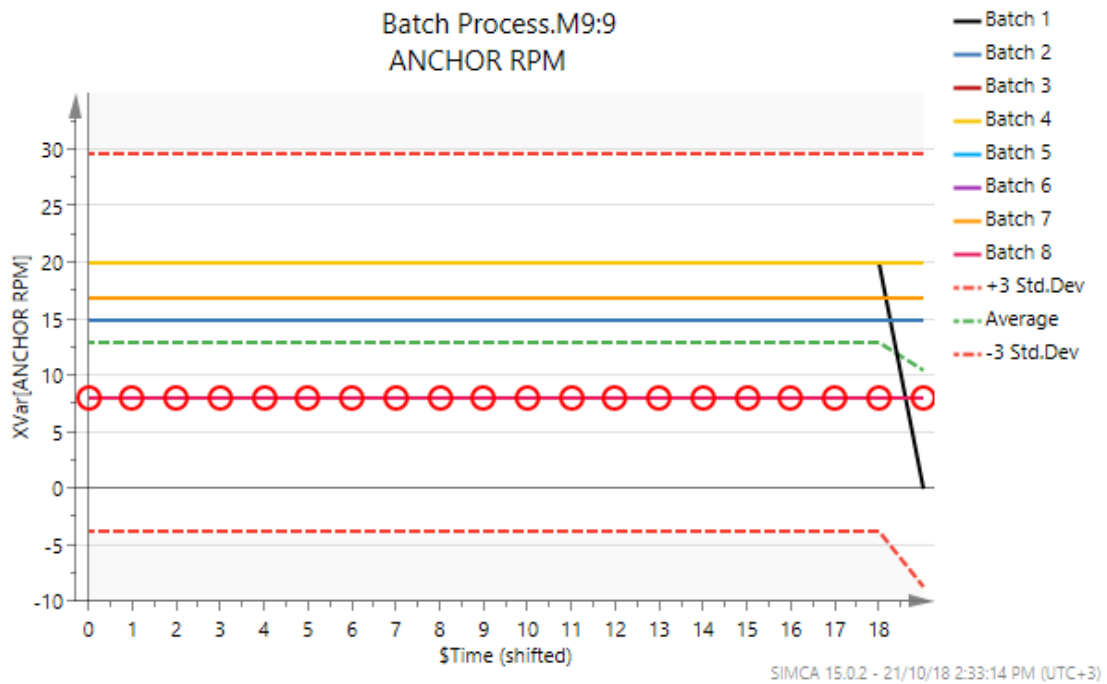
Εικόνα 5-4 Συνολικό Batch Control Plot για όλες τις εξεταζόμενες παρτίδες

Το παραπάνω διάγραμμα, στο οποίο απεικονίζεται η εξέλιξη στο χρόνο της πρώτης Κύριας Συνιστώσας για κάθε μια εκ των οκτώ εξεταζόμενων παρτίδων, μας δίνει μια συνολική εικόνα για τη συμπεριφορά της εκάστοτε παρτίδας κατά τη διάρκεια καθενός εκ των εννέα διακριτών σταδίων της διαδικασίας παραγωγής. Η εξέλιξη της κάθε παρτίδας αναπαρίσταται με μια γραφική παράσταση διαφορετικού χρώματος, ενώ η μέση αναμενόμενη διαδρομή μιας παρτίδας καθώς και τα άνω και κάτω αποδεκτά όρια αναπαρίστανται από μια πράσινη και δύο κόκκινες διακεκομμένες γραφικές παραστάσεις αντιστοίχως. Στο διάγραμμα αυτό παρατηρείται ότι, κατά κύριο λόγο, οι διαδρομές που ακολουθούν οι κύριες συνιστώσες των διαφόρων παρτίδων δεν παρουσιάζουν σημαντικές διαφορές μεταξύ τους. Στο 9<sup>ο</sup> και τελευταίο στάδιο της διαδικασίας ωστόσο, παρατηρείται μια σημαντική απόκλιση της 8<sup>ης</sup> παρτίδας από την αναμενόμενη πορεία εξέλιξης. Η απόκλιση αυτή γίνεται αμέσως αντιληπτή αν μεγεθύνουμε το επίμαχο σημείο του διαγράμματος, εστιάζοντας στο 9<sup>ο</sup> στάδιο της διαδικασίας.



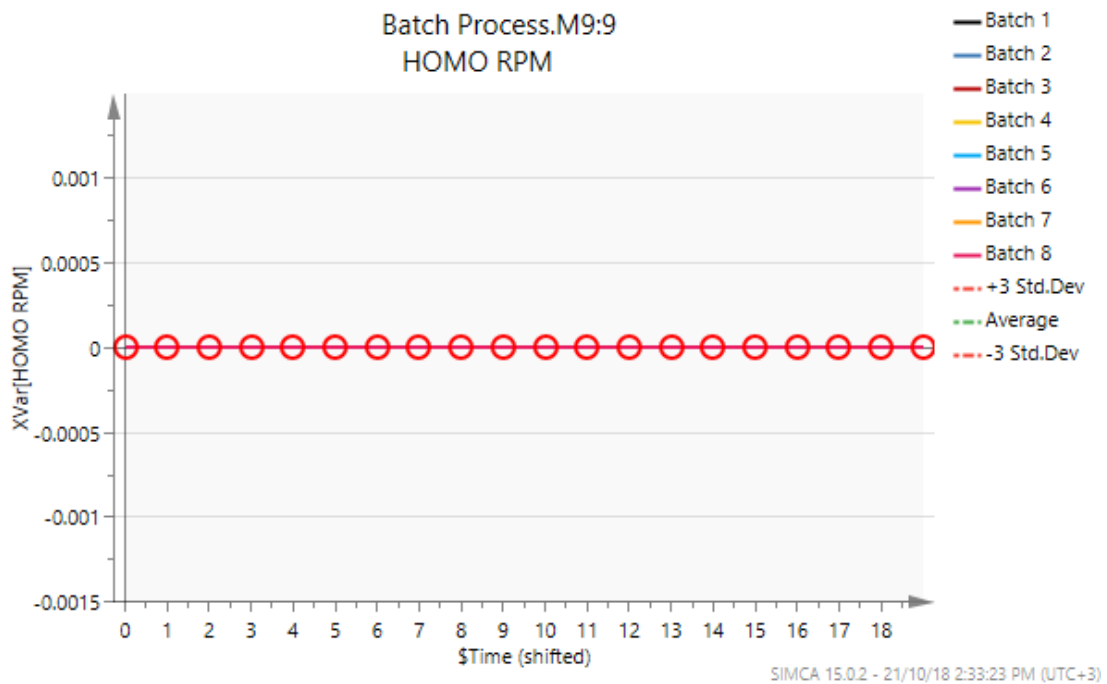
Εικόνα 5-5 Batch Control Plot για το 9<sup>ο</sup> στάδιο της διαδικασίας παραγωγής

Παρατηρούμε λοιπόν μια σημαντική απόκλιση της παρτίδας 8 στην αρχή του 9<sup>ου</sup> σταδίου της διαδικασίας παραγωγής. Μια περαιτέρω έρευνα για την εξέλιξη των διαφόρων παραμέτρων σε αυτό το στάδιο της διαδικασίας μπορεί να μας αποκαλύψει τους λόγους που οδήγησαν σε αυτή την απόκλιση.



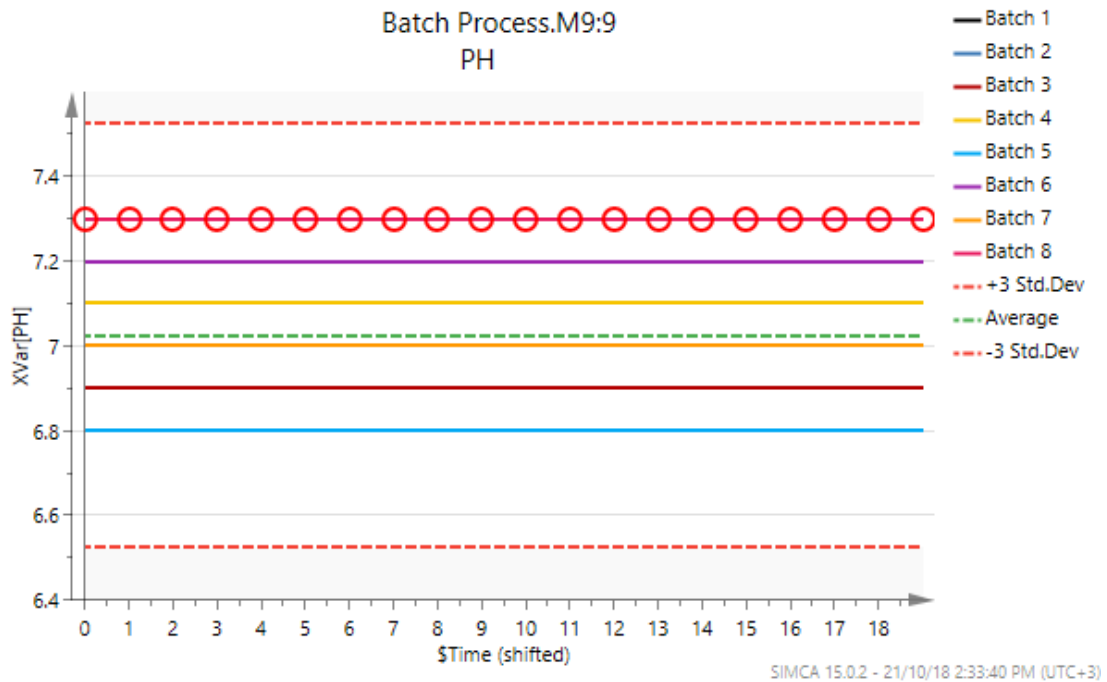
Εικόνα 5-6 Εξέλιξη της παραμέτρου «Στροφές αναδευτήρα τύπου Άγκυρας»

Στην περίπτωση των μετρούμενων στροφών του αναδευτήρα τύπου Άγκυρας της 8<sup>ης</sup> παρτίδας, παρατηρείται μη αξιολογήσιμη απόκλιση από την εξέλιξη της αντίστοιχης παραμέτρου στις υπόλοιπες παρτίδες.



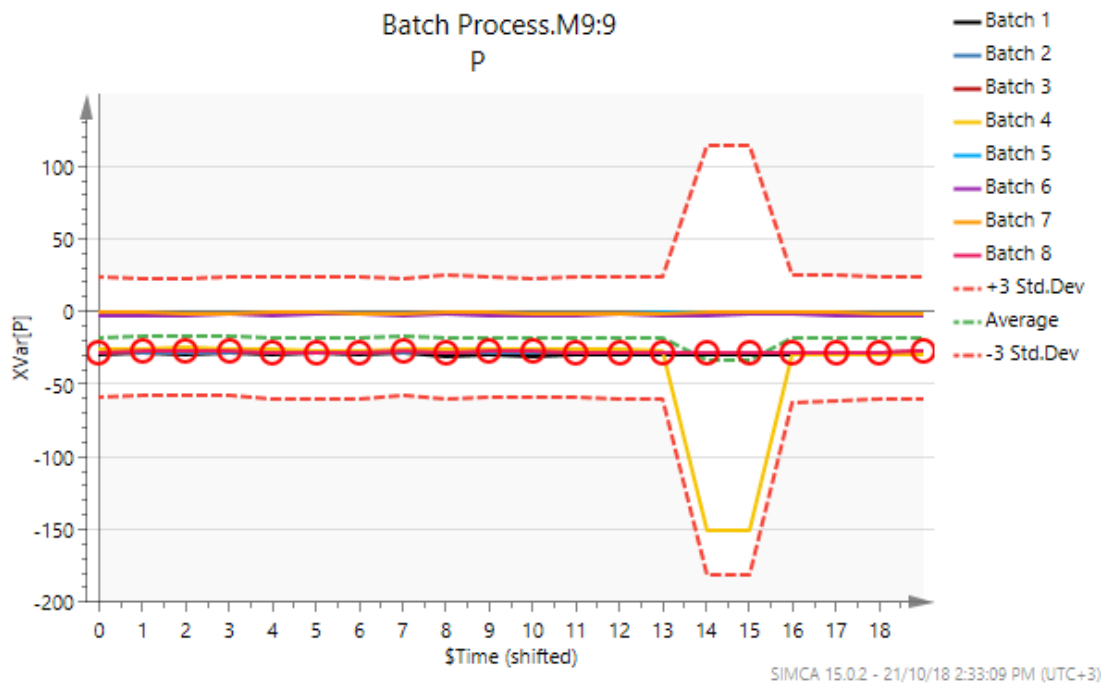
Εικόνα 5-7 Εξέλιξη της παραμέτρου «Στροφές Ομογενοποιητή»

Στην περίπτωση των στρωφών του ομογενοποιητή δεν παρατηρείται η παραμικρή αποκλιση



**Εικόνα 5-8 Εξέλιξη της παραμέτρου «PH»**

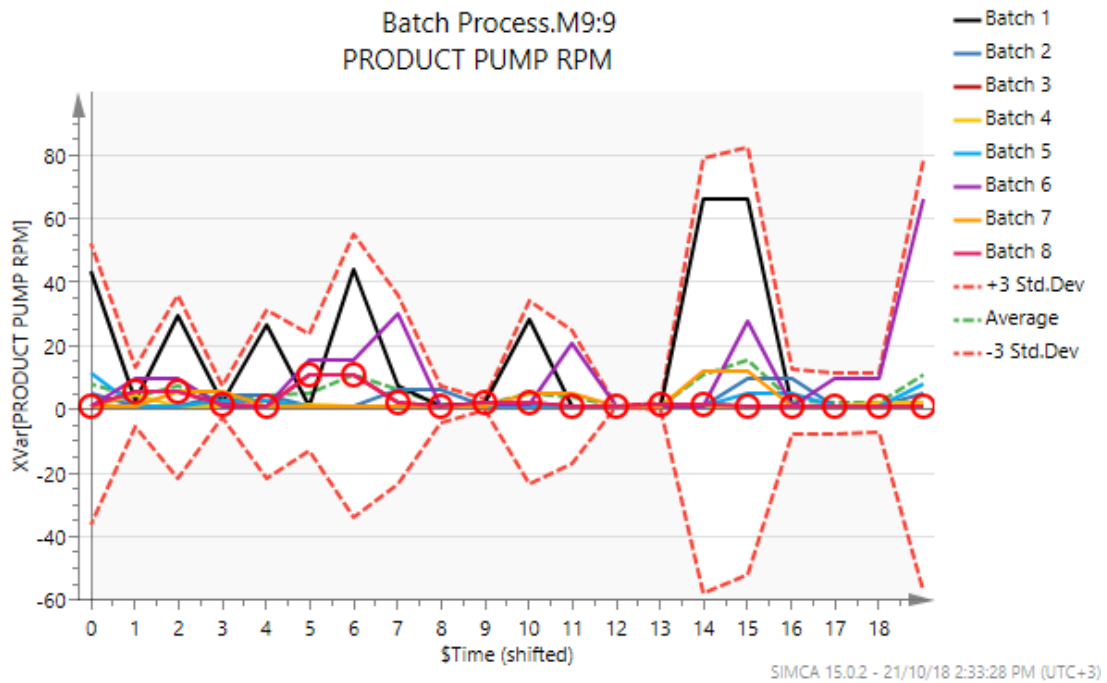
Η εξέλιξη του PH του μίγματος καθώς και της πίεσης εντός του δοχείου ανάμιξης, για την 8<sup>η</sup> παρτίδα, κρίνεται ως φυσιολογική και αναμενόμενη.



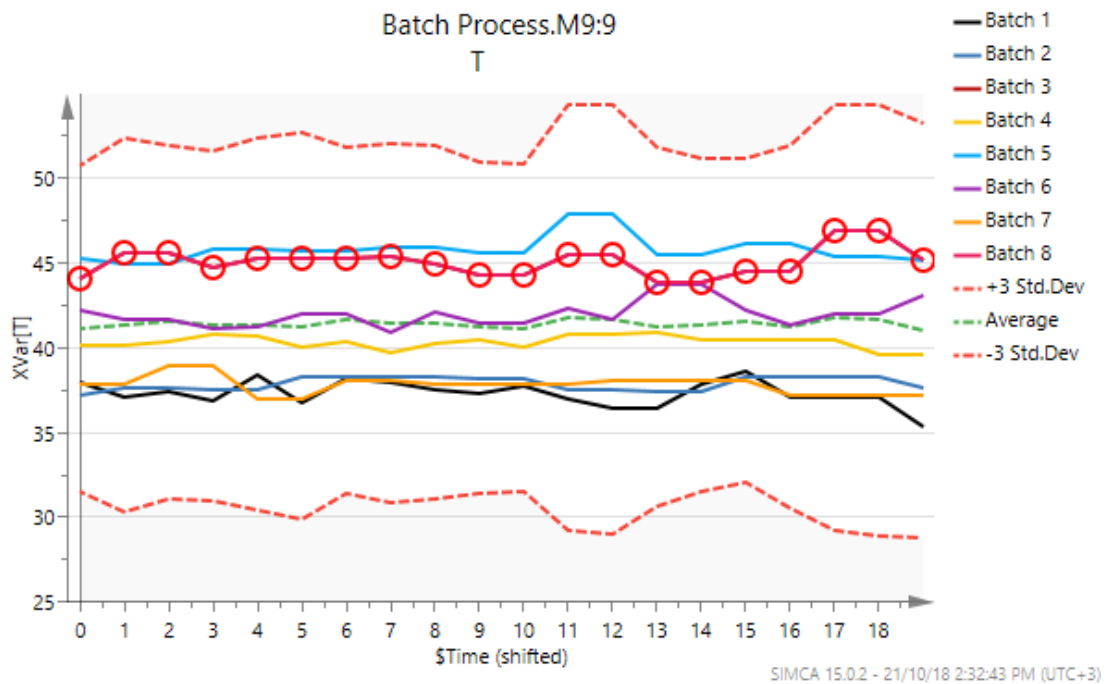
**Εικόνα 5-9 Εξέλιξη της παραμέτρου «Πίεση»**



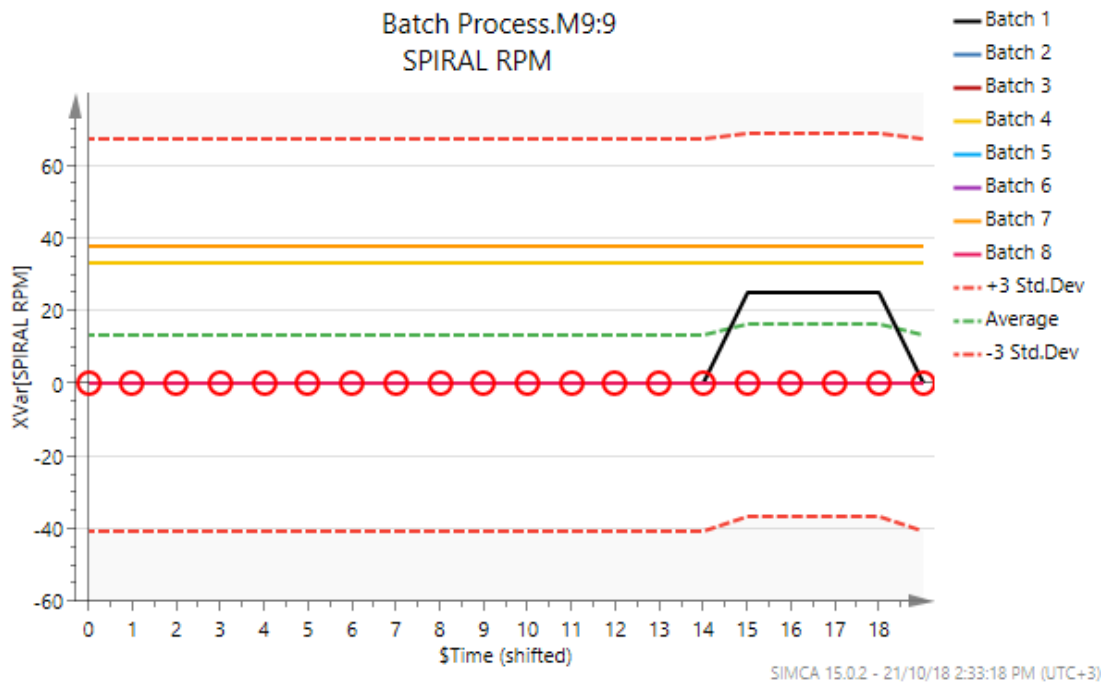
Εξίσου φυσιολογική και αναμενόμενη κρίνεται και η εξέλιξη της θερμοκρασίας εντός του δοχείου ανάμιξης καθώς και των στροφών της αντλίας εισαγωγής πρώτων υλών και του αναδευτήρα τύπου spiral.



Εικόνα 5-10 Εξέλιξη της παραμέτρου «Στροφές Αντλίας Εισαγωγής Πρώτων Υλών»

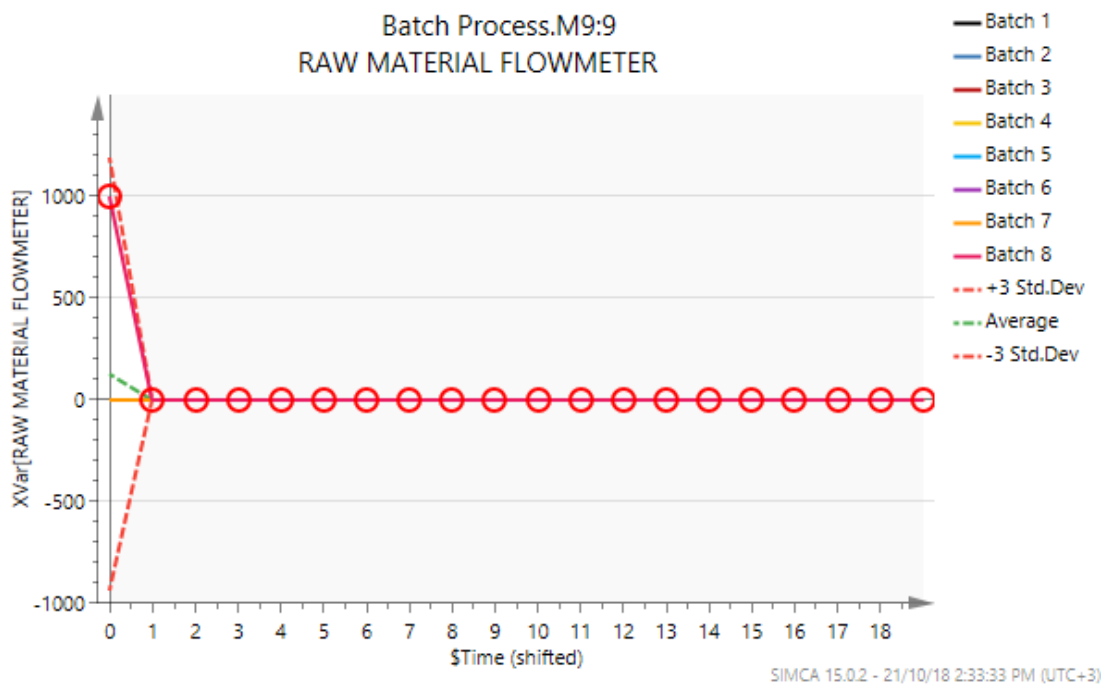


Εικόνα 5-11 Εξέλιξη της παραμέτρου «Θερμοκρασία»

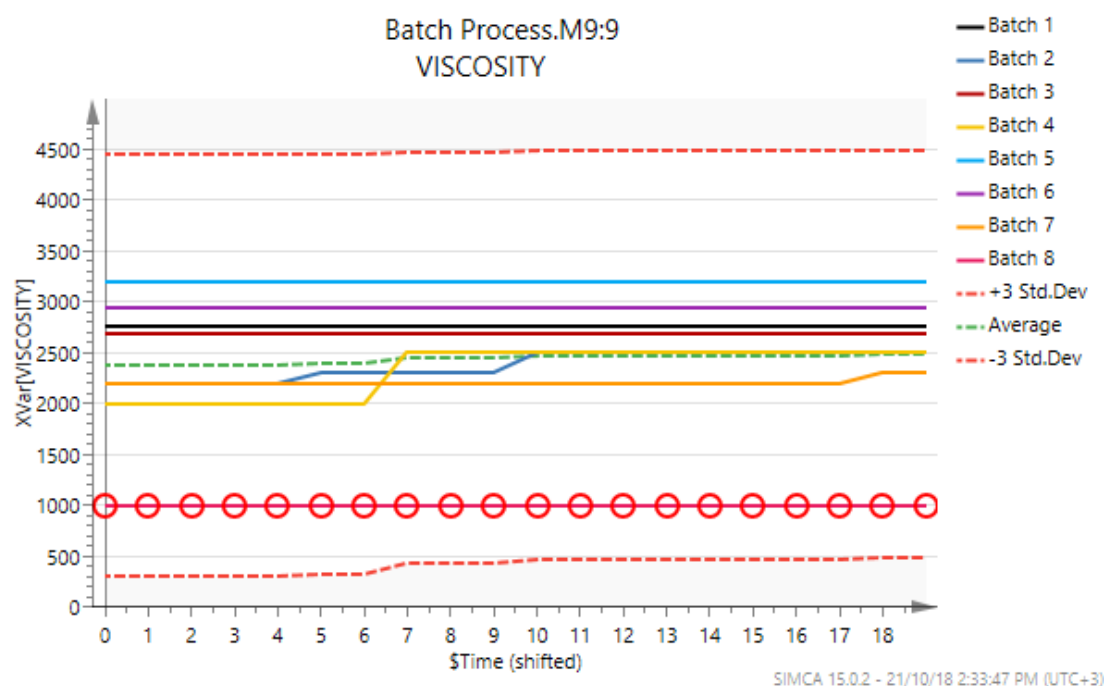


Εικόνα 5-12 Εξέλιξη της παραμέτρου «Στροφές Αναδευτήρα τύπου Spiral»

Τέλος, οι δύο παράμετροι που παρουσιάζουν τη μεγαλύτερη απόκλιση στην εξέλιξη τους από αυτή των υπολοίπων παρτίδων είναι η μάζα εισαγόμενης πρώτης ύλης και το ιξώδες.



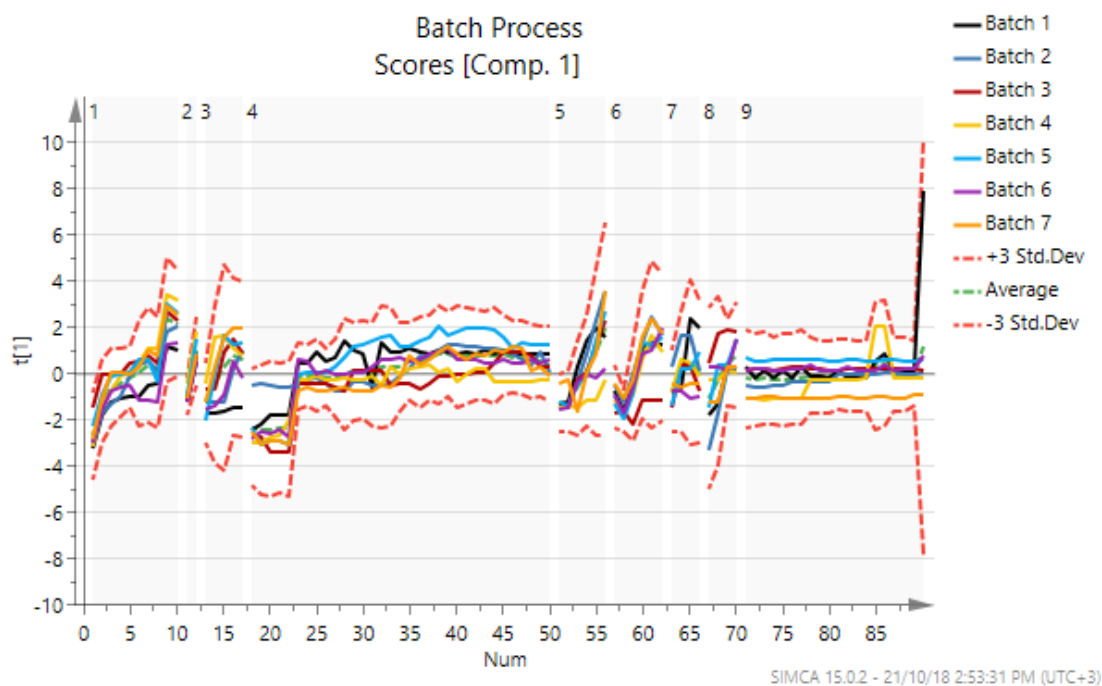
Εικόνα 5-13 Εξέλιξη της παραμέτρου «Μάζα εισαγόμενης Πρώτης Ύλης»



Εικόνα 5-14 Εξέλιξη της παραμέτρου «Ιξώδες»

Συγκεκριμένα παρατηρούμε ότι στην έναρξη του 9<sup>ου</sup> σταδίου της διαδικασίας παραγωγής, εμφανίζεται μια εξαιρετικά έντονη αύξηση στη μάζα της εισαγόμενης πρώτης ύλης, η οποία όπως είναι αναμενόμενο, προκαλεί με τη σειρά της μεγάλη διαφοροποίηση στο ιξώδες του μίγματος, το οποίο αποτελεί και μια από τις κρισιμότερες παραμέτρους για την αξιολόγηση της ποιότητας του τελικού προϊόντος. Μια τέτοια απόκλιση, η οποία κατά πάσα πιθανότητα οφείλεται σε λάθος χειρισμό, είναι ικανή να θέσει την παρτίδα εκτός προδιαγραφών. Συνεπώς, η παρτίδα 8 θα πρέπει να εξαιρεθεί από το σύνολο δεδομένων που θα χρησιμοποιηθεί για την κατασκευή του τελικού μοντέλου.

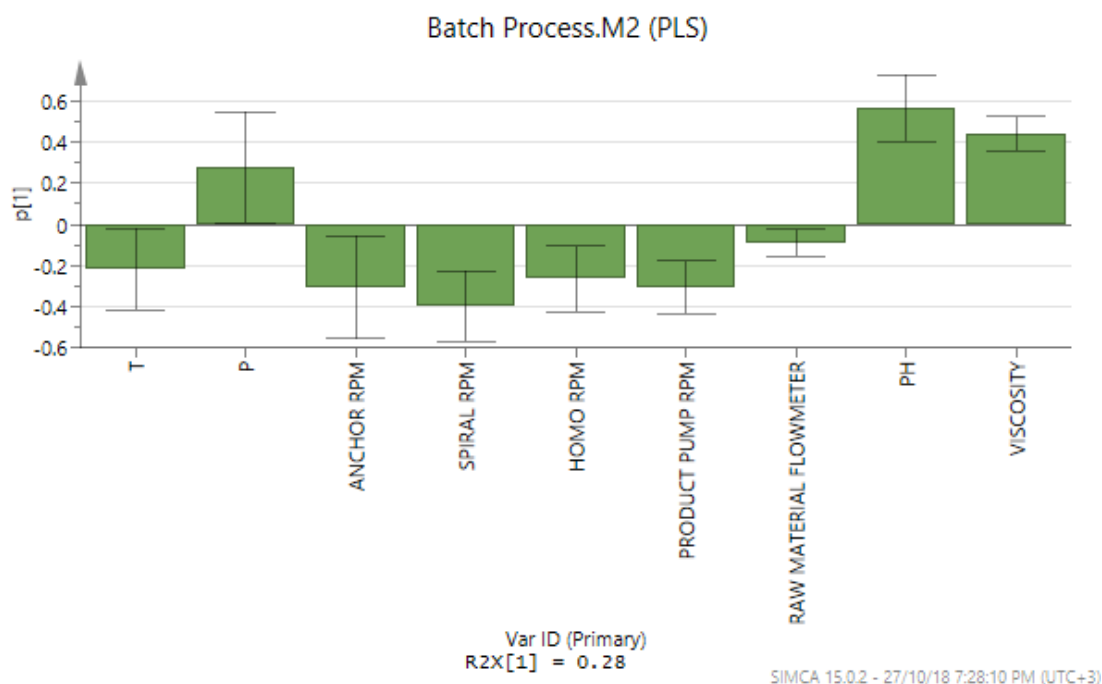
Έχοντας πλέον εντοπίσει και εξαιρέσει την προβληματική παρτίδα από το σύνολο δεδομένων, μπορούμε να προχωρήσουμε στην κατασκευή του τελικού μοντέλου BEM το οποίο θα αξιοποιεί τα δεδομένα που προκύπτουν από τις επτά πρώτες παρτίδες. Στη συνέχεια παρατίθεται το τελικό συνολικό διάγραμμα ελέγχου παρτίδων.



**Εικόνα 5-15** Τελικό συνολικό **Batch Control Plot** για όλες τις εξεταζόμενες παρτίδες

Στο παραπάνω διάγραμμα απεικονίζεται η εξέλιξη στο χρόνο της πρώτης Κύριας Συνιστώσας για κάθε μια εκ των πετυχημένων παρτίδων. Παρατηρούμε ότι κατά κύριο λόγο οι εξεταζόμενες παρτίδες ακολουθούν μια παρόμοια πορεία εξέλιξης στο χρόνο, η οποία μπορεί να συνοψιστεί από τη μέση τιμή που αναπαρίσταται από μια διακεκομμένη πράσινη γραμμή καθώς και από τις τυπικές αποκλίσεις  $\pm 3\sigma$ , οι οποίες αναπαρίστανται από κόκκινες διακεκομμένες γραμμές. Οι γραφικές αυτές παραστάσεις αντιπροσωπεύουν την επιθυμητή πορεία εξέλιξης μιας παρτίδας και τα άνω και κάτω αποδεκτά όρια αυτής αντιστοίχως.

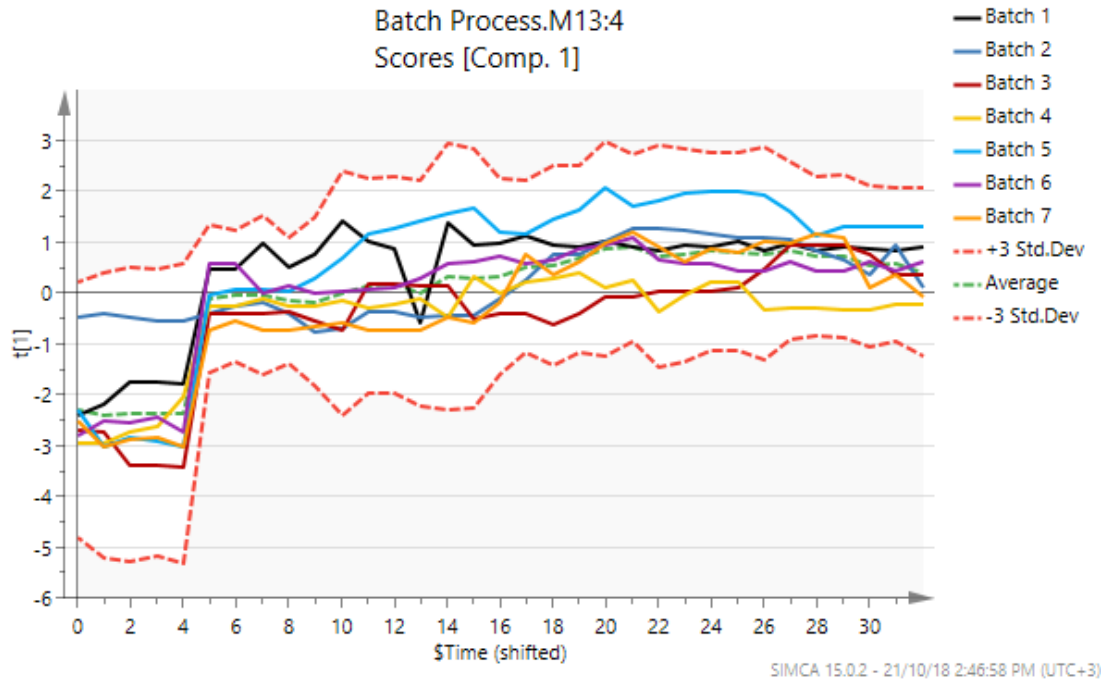
Επιπλέον, παρατίθεται το διάγραμμα «φορτίων» της πρώτης Κύριας Συνιστώσας, στο οποίο απεικονίζεται η βαρύτητα με την οποία συμμετέχει κάθε μια εκ των ανεξάρτητων μεταβλητών του προβλήματος στη σύνθεση της συνιστώσας αυτής. Κατ' αυτή την έννοια πρόκειται για ένα «διάγραμμα σημαντικότητας» (VIP) των μεταβλητών. Από το διάγραμμα αυτό συμπεραίνουμε ότι, στην περίπτωση που η παραγωγική διαδικασία αντιμετωπίζεται ως ενιαία και αδιαίρετη, οι δύο μεταβλητές, η εξέλιξη των οποίων επηρεάζει περισσότερο την εξέλιξη της πρώτης Κύριας Συνιστώσας, κατ' απόλυτη τιμή, είναι το PH και το ιξώδες.



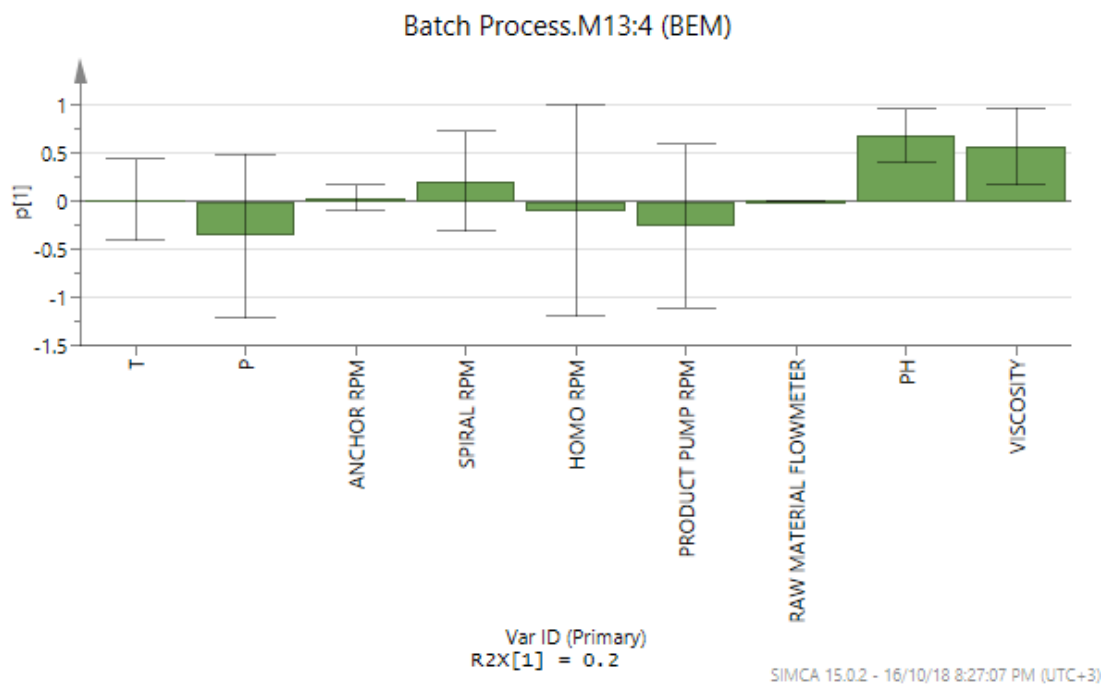
**Εικόνα 5-16 Διάγραμμα Σημαντικότητας των μεταβλητών (Variable Importance Plot)**

Το προαναφερθέν συμπέρασμα υποδεικνύει επί της ουσίας την κατεύθυνση προς την οποία θα πρέπει κινηθεί η επένδυση πόρων, προκειμένου να εξασφαλίζεται η αξιοπιστία της διαδικασίας. Έτσι, η επένδυση σε προηγμένα συστήματα παρακολούθησης και ελέγχου των παραμέτρων με τη μεγαλύτερη βαρύτητα μπορεί να προλαμβάνει μικρές αποκλίσεις αυτών οι οποίες όμως μπορεί να επιφέρουν σημαντικές επιπτώσεις στην τελική ποιότητα του τελικού παραγόμενου προϊόντος.

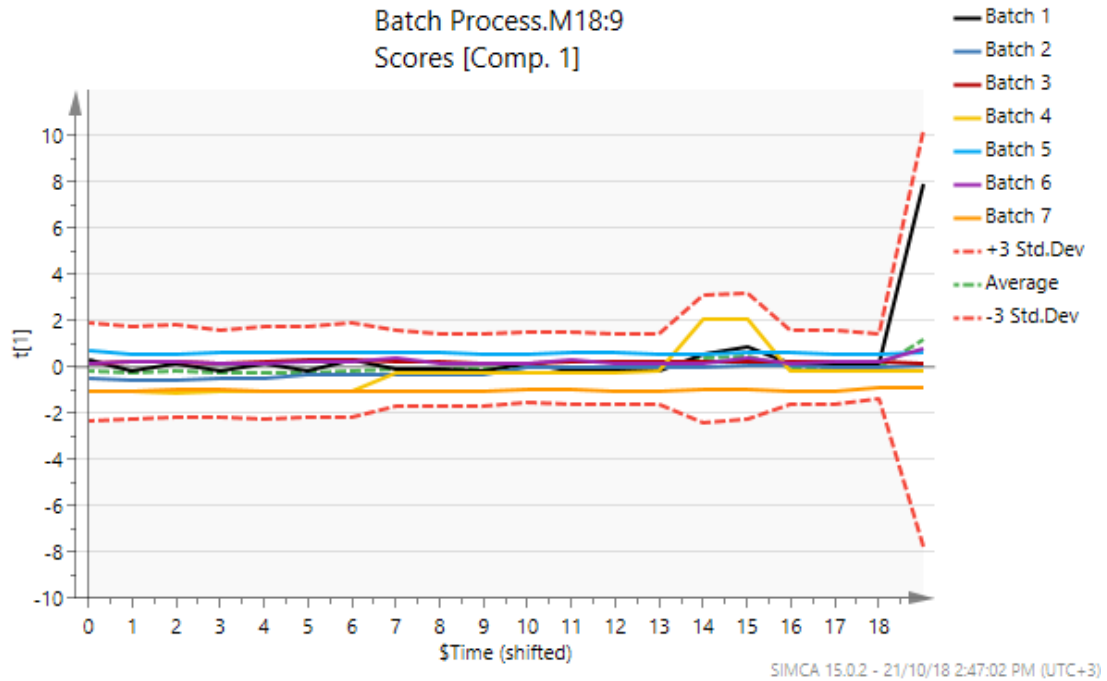
Όπως έχει ήδη αναφερθεί, στην περίπτωση που η διαδικασία παραγωγής αποτελείται από επιμέρους διακριτά βήματα, τότε το λογισμικό SIMCA μας δίνει τη δυνατότητα κατασκευής, τόσο ενός συνολικού μοντέλου για όλη τη διαδικασία, όσο και πολλών επιμέρους μοντέλων για κάθε ένα εκ των βημάτων αυτών. Η κατασκευή πολλαπλών επιμέρους μοντέλων μας δίνει τη δυνατότητα να μελετήσουμε διεξοδικά το κάθε στάδιο της διαδικασίας και να αξιολογήσουμε τη διαφορετική επίδραση που πιθανώς έχουν στο αποτέλεσμα της διαδικασίας οι διάφορες παρατηρούμενες παράμετροι καθώς και τις διαφορετικές σχέσεις που πιθανώς αναπτύσσονται μεταξύ τους στα διάφορα στάδια. Κατ' αυτόν τον τρόπο καθίσταται ευκολότερος και πιο στοχευμένος ο εντοπισμός πιθανών προβλημάτων. Στη συνέχεια παρατίθενται ενδεικτικά τα αντίστοιχα διαγράμματα για δύο επιμέρους στάδια της παραγωγικής διαδικασίας, συγκεκριμένα τα στάδια 4 και 9.



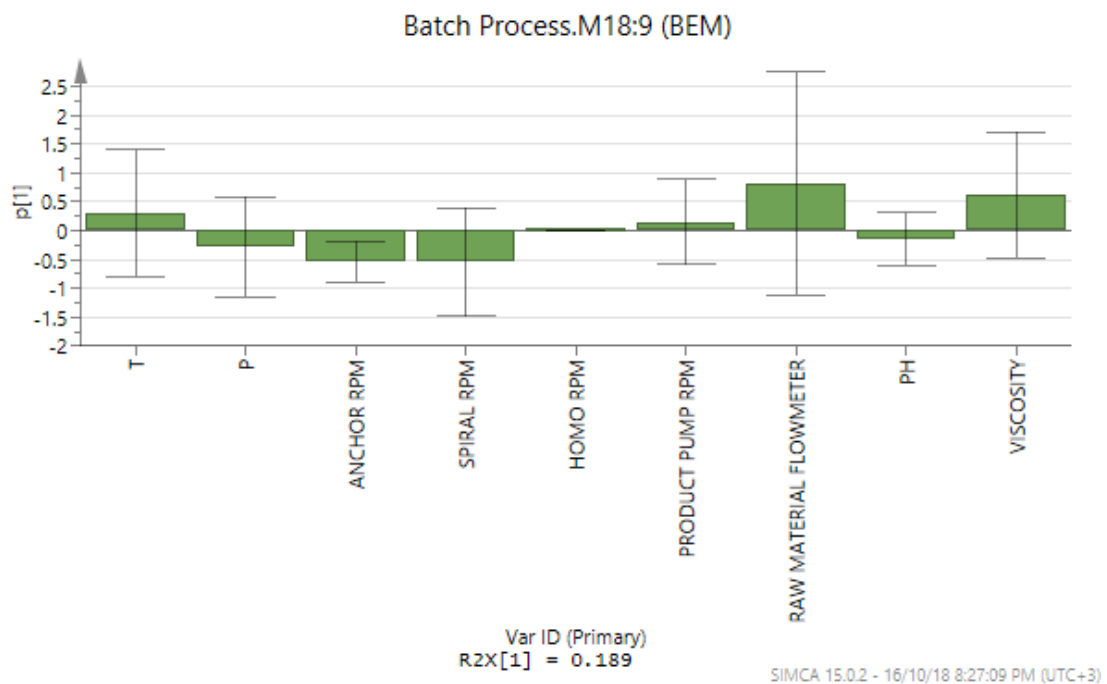
Εικόνα 5-17 Batch Control Plot για όλες τις εξεταζόμενες παρτίδες, για το στάδιο 4 της παραγωγικής διαδικασίας.



Εικόνα 5-18 Λιάγραμμα Σημαντικότητας των μεταβλητών (Variable Importance Plot) για το στάδιο 4 της παραγωγικής διαδικασίας.



Εικόνα 5-19 Batch Control Plot για όλες τις εξεταζόμενες παρτίδες, για το στάδιο 9 της παραγωγικής διαδικασίας.



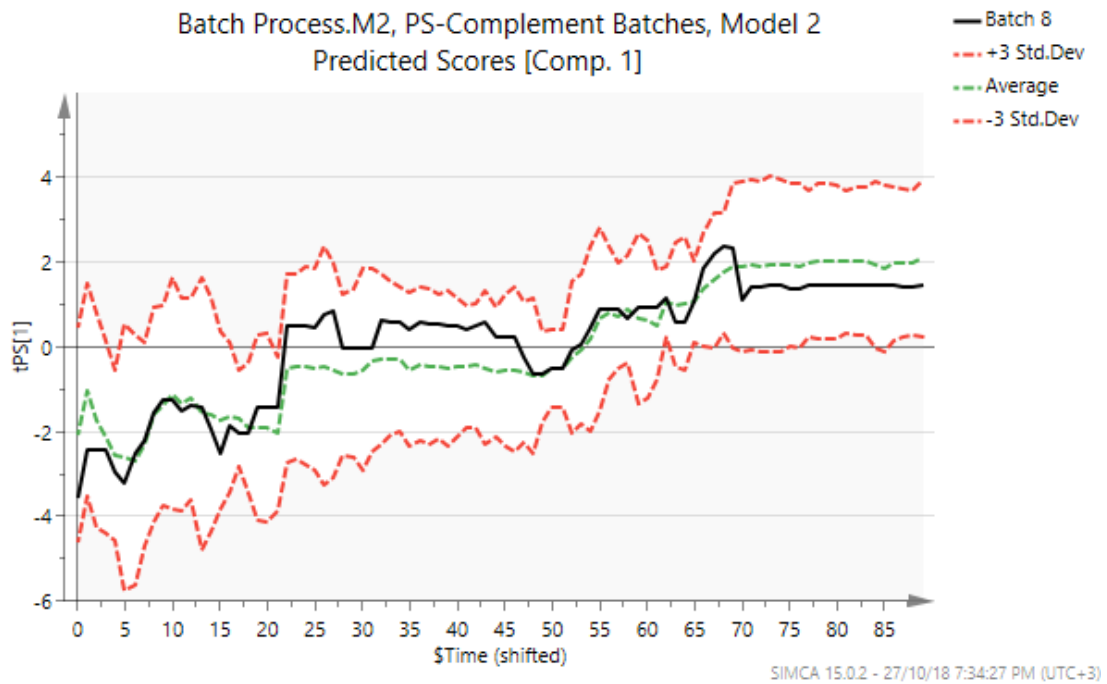
Εικόνα 5-20 Διάγραμμα Σημαντικότητας των μεταβλητών (Variable Importance Plot) για το στάδιο 9 της παραγωγικής διαδικασίας.

Από τα παραπάνω διαγράμματα γίνεται πλέον φανερή η διαφοροποίηση στην κατανομή της βαρύτητας μεταξύ των διαφόρων μεταβλητών του προβλήματος, όταν η παραγωγική διαδικασία διαιρείται σε επιμέρους στάδια. Βλέπουμε λοιπόν ότι σε κάθε στάδιο της διαδικασίας, η εξέλιξη της κάθε μεταβλητής επηρεάζει με διαφορετικό τρόπο την αντίστοιχη εξέλιξη της πρώτης Κύριας Συνιστώσας. Παρατηρούμε λοιπόν, για παράδειγμα, ότι στο 4<sup>ο</sup> στάδιο της παραγωγικής διαδικασίας οι παράμετροι με τη μεγαλύτερη επίδραση είναι το PH, το ιξώδες και η πίεση, ενώ στο 9<sup>ο</sup> στάδιο οι αντίστοιχες παράμετροι είναι η μάζα εισαγόμενης πρώτης ύλης, το ιξώδες και οι στροφές των αναδευτήρων τύπου άγκυρας και spiral. Το συμπέρασμα αυτό είναι συμβατό και με την εξέλιξη της παρτίδας 8 η οποία, όπως είδαμε, στο 9<sup>ο</sup> στάδιο της διαδικασίας παραγωγής, παρουσίαζε σημαντικές αποκλίσεις στην εξέλιξη της μάζας εισαγόμενης πρώτης ύλης και του ιξώδους και συνεπώς χαρακτηρίστηκε ως αποτυχημένη.

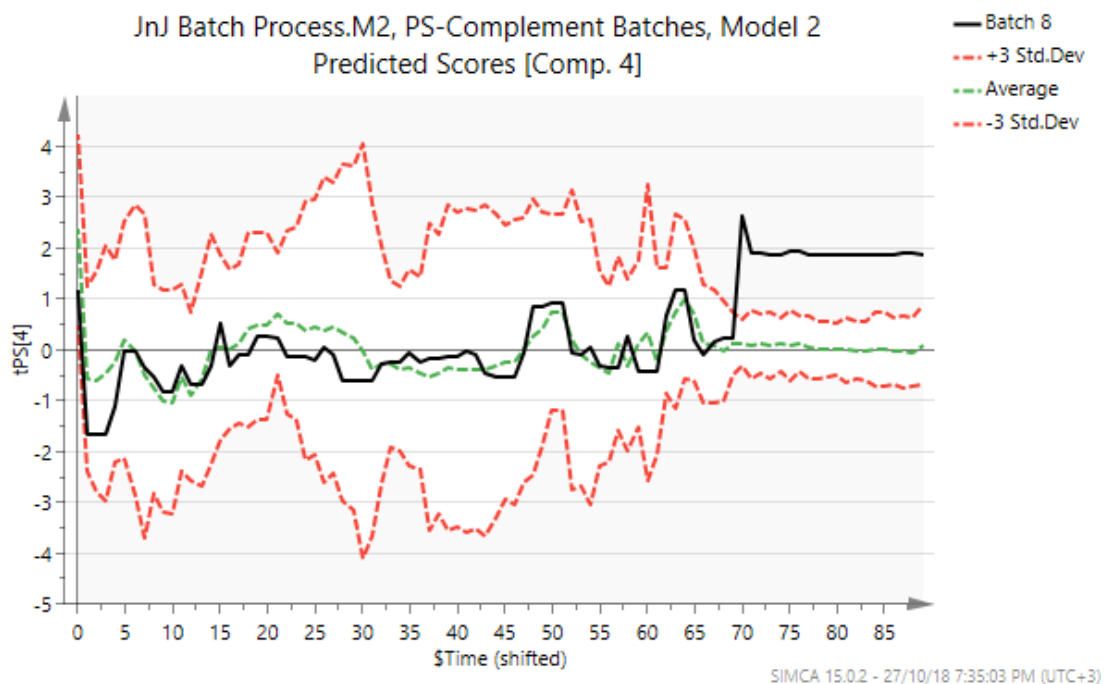
Προκειμένου τώρα, να αξιολογήσουμε τις επιδόσεις του μοντέλου, μπορούμε να χρησιμοποιήσουμε το σύνολο των παρτίδων που εξαιρέθηκαν από την κατασκευή του και να ελέγξουμε κατά πόσον προβλέπονται επιτυχώς οι αποκλίσεις οι οποίες τελικά οδήγησαν τις εν λόγω παρτίδες εκτός προδιαγραφών.

Για το σκοπό αυτό, αρχικά παρατίθεται το διάγραμμα των προβλεπόμενων scores της παρτίδας 8, σε αντιπαραβολή με την αναμενόμενη πορεία εξέλιξης της πρώτης Κύριας Συνιστώσας και τα αντίστοιχα άνω και κάτω όρια, στην περίπτωση που η παραγωγική διαδικασία αντιμετωπίζεται ως ενιαία. Σε αυτή την περίπτωση παρατηρούμε ότι το μοντέλο δεν εντοπίζει την απόκλιση της παρτίδας 8 με τρόπο ικανοποιητικό, καθώς η πορεία εξέλιξης της πρώτης Κύριας Συνιστώσας φαίνεται να βρίσκεται εντός των ορίων, όπως αυτά προέκυψαν από την εκπαίδευση του μοντέλου με βάση τις υπόλοιπες παρτίδες. Την ίδια ακριβώς συμπεριφορά μπορούμε να παρατηρήσουμε και για τις υπόλοιπες Κύριες Συνιστώσες, μέχρι και την 3<sup>η</sup>. Η απόκλιση της παρτίδας 8 αρχίζει να γίνεται αντιληπτή από την παρατήρηση της εξέλιξης της 4<sup>ης</sup> Κύριας Συνιστώσας του μοντέλου, όπου πράγματι, λίγο πριν την ολοκλήρωση της παρτίδας σημειώνεται σημαντική απόκλιση από την αναμενόμενη εξέλιξη. Ο λόγος που συμβαίνει αυτό είναι ότι, στην περίπτωση που η διαδικασία παραγωγής αντιμετωπίζεται ως ενιαία χωρίς να διαιρείται σε επιμέρους στάδια, η επίδραση της μάζας εισαγόμενης πρώτης ύλης, στην πορεία εξέλιξης της πρώτης και σημαντικότερης Κύριας Συνιστώσας είναι σχεδόν αμελητέα, όπως είδαμε στο στην Εικόνα 5-16.





Εικόνα 5-21 Batch Control Plot για την εξέλιξη της 1<sup>ης</sup> Κύριας Συνιστώσας, για την 8<sup>η</sup> παρτίδα.

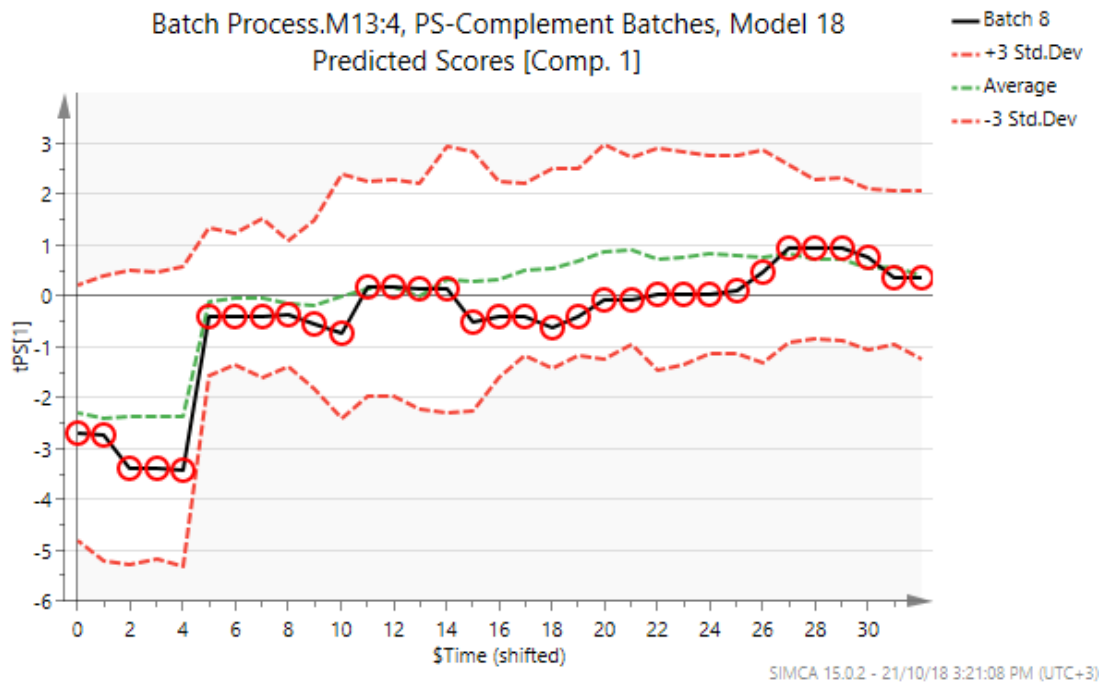


Εικόνα 5-22 Batch Control Plot για την εξέλιξη της 4<sup>ης</sup> Κύριας Συνιστώσας, για την 8<sup>η</sup> παρτίδα.

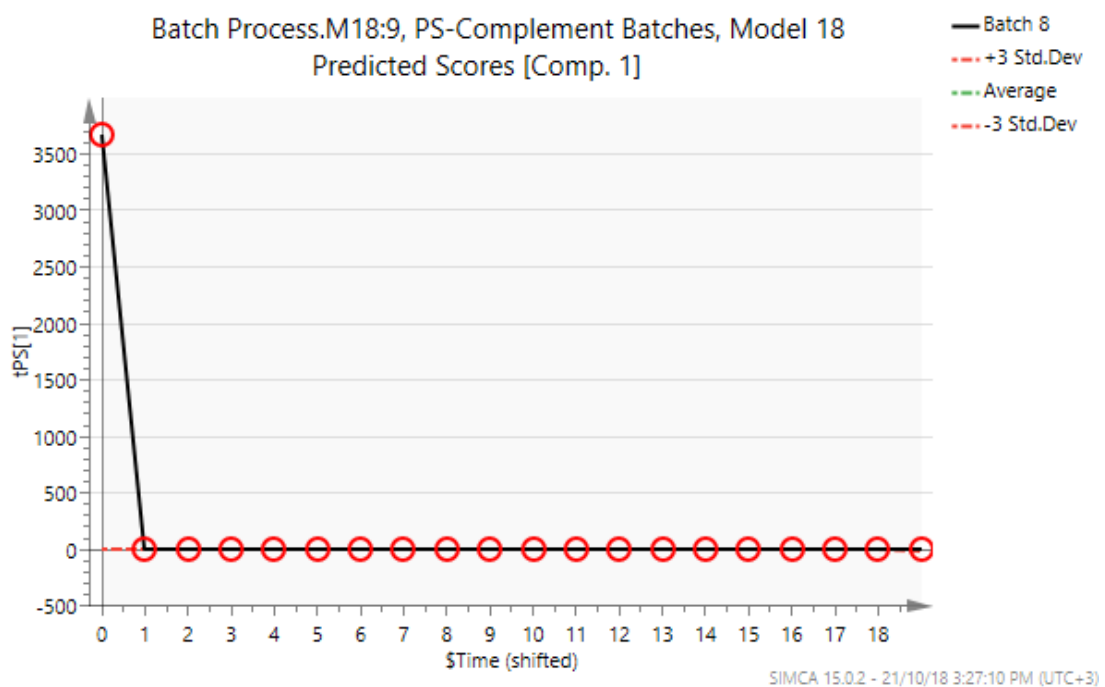
Προκύπτει λοιπόν το συμπέρασμα ότι στην εξεταζόμενη διαδικασία, η ύπαρξη έντονων διαφορών στις σχέσεις που εμφανίζονται μεταξύ των διαφόρων παραμέτρων κατά τη διάρκεια διαφορετικών φάσεων της διαδικασίας καθιστά μη αποδοτική την

αντιμετώπισή της ως ενιαία και υπαγορεύει την κατασκευή επιμέρους μοντέλων για κάθε ένα από τα διακριτά στάδια.

Για του λόγου το αληθές, στη συνέχεια παρατίθενται τα αντίστοιχα διαγράμματα, όπως αυτά προκύπτουν για τα επιμέρους στάδια 4 και 9 της διαδικασίας.



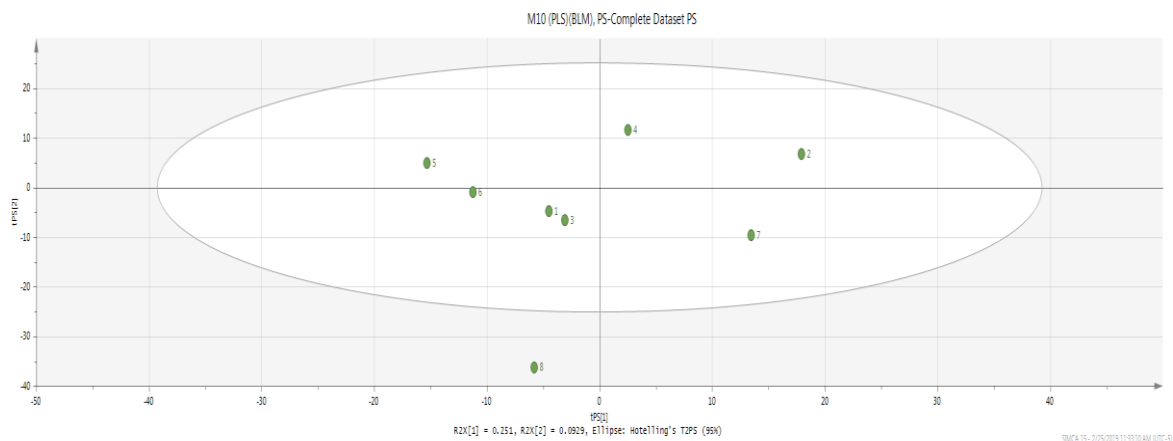
Εικόνα 5-23 Batch Control Plot για την εξέλιξη της 1ης Κύριας Συνιστώσας, για το 4<sup>ο</sup> στάδιο.



Εικόνα 5-24 Batch Control Plot για την εξέλιξη της 1ης Κύριας Συνιστώσας, για το 9<sup>ο</sup> στάδιο.

Στα παραπάνω διαγράμματα φαίνεται τόσο η φυσιολογική και αναμενόμενη εξέλιξη της παρτίδας κατά τη διάρκεια του 4<sup>ου</sup> σταδίου της παραγωγικής διαδικασίας, όσο και η εξίσου αναμενόμενη και έντονη απόκλιση της παρτίδας κατά τη διάρκεια του 9<sup>ου</sup> και τελευταίου σταδίου της διαδικασίας.

Τέλος, κατασκευάζεται το μοντέλο BLM, σύμφωνα με τα όσα παρουσιάστηκαν στην ενότητα 4.3.3 Batch Level Model. Στη συνέχεια παρατίθεται το αντίστοιχο score plot για τις δύο πρώτες Κύριες Συνιστώσες, όπου η κάθε παρτίδα αναπαρίσταται από ένα μοναδικό σημείο. Με τη βοήθεια αυτού του διαγράμματος, όπως έχει εξηγηθεί και σε προηγούμενο κεφάλαιο, μας δίνεται η δυνατότητα να διακρίνουμε τόσο τις παρτίδες οι οποίες παρουσιάζουν σημαντικές αποκλίσεις από όλες τις υπόλοιπες, όσο και συστάδες παρτίδων οι οποίες παρουσιάζουν ομοιότητες μεταξύ τους.



**Εικόνα 5-25 Score Plot μοντέλου BLM**

Έτσι για παράδειγμα, στη συγκεκριμένη περίπτωση μπορούμε να παρατηρήσουμε αφενός την αναμενόμενη απόκλιση της παρτίδας 8 και αφετέρου σημαντικές ομοιότητες μεταξύ των παρτίδων 1 και 3. Οι πληροφορίες αυτές είναι δυνατόν να αξιολογηθούν και να αποκαλύψουν άγνωστες πτυχές της διαδικασίας παραγωγής οι οποίες ξεφεύγουν από το πλαίσιο της παρούσης, μπορεί ωστόσο να αποδειχθούν εξαιρετικά χρήσιμες.



## ΚΕΦΑΛΑΙΟ 6: Σύνοψη

---

### 6.1 Συμπεράσματα

Στην παρούσα ενότητα επιχειρείται ένας συνοπτικός απολογισμός των συμπερασμάτων που προκύπτουν τόσο από την μελέτη της βιβλιογραφίας σχετικά με την εφαρμογή μεθόδων Ανάλυσης Δεδομένων στη Βιομηχανία και συγκεκριμένα σε συστήματα παραγωγής σε παρτίδες, όσο και από την υλοποίησή τους στην πράξη.

Όπως κατέστη φανερό, η Ανάλυση Δεδομένων μπορεί να αποδειχθεί ως ένα εξαιρετικά χρήσιμο εργαλείο για την κατανόηση, τον έλεγχο και τελικά τη βελτιστοποίηση μιας παραγωγικής διαδικασίας. Συγκεκριμένα, είδαμε πως με τη βοήθεια σύγχρονων και εξειδικευμένων μεθόδων ανάλυσης δεδομένων, δίνεται η δυνατότητα σε μια βιομηχανική επιχείρηση, να αξιοποιήσει με τον καλύτερο δυνατό τρόπο το, κατά κύριο λόγο, αδρανές κεφάλαιο των εκτενέστατων βάσεων ιστορικών δεδομένων που διαθέτει, καθώς και του μεγάλου όγκου δεδομένων που παράγει σε καθημερινή βάση και σε πραγματικό χρόνο, κατά τη διάρκεια της παραγωγής. Τα δεδομένα αυτά μετουσιώνονται σε ένα εξαιρετικά πολύτιμο αγαθό, καθώς, μέσω των αναλύσεων στις οποίες υπόκεινται, αποκαλύπτουν εκείνες τις πτυχές της παραγωγικής διαδικασίας, η βελτιστοποίηση των οποίων, θα αποφέρει σημαντική προστιθέμενη αξία στη λειτουργία της επιχείρησης. Στην περίπτωση των διαδικασιών παραγωγής σε παρτίδες είδαμε ότι, λόγω της φύσης τους, αφενός, η ανάγκη για αξιοποίηση των δεδομένων που παράγονται, καθίσταται επιτακτική και αφετέρου ότι η αξιοποίησή τους παρουσιάζει σημαντικές προκλήσεις. Διατρέχοντας τη βιβλιογραφία, μελετήσαμε διάφορες μεθόδους ανάλυσης δεδομένων και καταλήξαμε σε εκείνες, οι οποίες μπορούν να αντιμετωπίσουν τις προκλήσεις αυτές ικανοποιητικά. Έτσι, η Ανάλυση Κυρίων Συνιστωσών και η Παλινδρόμηση Μερικών ελαχίστων Τετραγώνων επιλέχθηκαν ως οι δύο επικρατέστερες μέθοδοι και εφαρμόστηκαν στην ανάλυση πραγματικών δεδομένων με τη βοήθεια του εμπορικού πακέτου λογισμικού SIMCA.

Η παραγωγική διαδικασία που αναλύθηκε στο πλαίσιο της παρούσας εργασίας, αποτελεί μια φυσικοχημική διεργασία, η οποία έχει ως αντικείμενο την παραγωγή καλλυντικών προϊόντων. Η παραγωγή πραγματοποιείται σε παρτίδες και κατά τη διάρκεια της συλλέγονται δεδομένα που αντικατοπτρίζουν την εξέλιξη κάποιων παραμέτρων, όπως είναι το ιξώδες και το PH του μίγματος, η πίεση και η θερμοκρασία

του δοχείου ανάμειξης, η μάζα εισαγόμενης πρώτης ύλης και οι στροφές αναδευτήρων διαφορετικών τύπων. Εξετάστηκαν 7 τον αριθμό επιτυχείς παρτίδες και μια παρτίδα η οποία παρουσιάζει σημαντική απόκλιση από τις προδιαγραφές ποιότητας του τελικού προϊόντος. Η εφαρμογή των μεθόδων PCA και PLS και η κατασκευή του μοντέλου παρτίδας (BLM) και του μοντέλου εξέλιξης παρτίδας (BEM) αποκάλυψαν τις υποκείμενες σχέσεις που αναπτύσσονται μεταξύ των παρατηρούμενων μεταβλητών κατά τη διάρκεια της παραγωγικής διαδικασίας, καθώς και την επίδραση που έχει η εξέλιξη της κάθε μιας από αυτές τις μεταβλητές στην συνολική εξέλιξη της παρτίδας. Επιπλέον, αποκαλύφθηκαν ομοιότητες και αποκλίσεις μεταξύ των εξεταζόμενων παρτίδων. Οι πληροφορίες αυτές είναι άμεσα αξιοποιήσιμες καθώς υποδεικνύουν εκείνες τις πτυχές της διαδικασίας, ο αποτελεσματικός έλεγχος των οποίων, θα πρέπει να θεωρείται ως κρίσιμος, προκειμένου να αποφεύγονται αστοχίες, να επιτυγχάνονται τα επιθυμητά χαρακτηριστικά του τελικού προϊόντος, να εξοικονομούνται πόροι και τελικά να μεγιστοποιείται η παραγωγικότητα και η αποδοτικότητα. Υποδεικνύεται συνεπώς, ο τρόπος με τον οποίο θα πρέπει να κατανεμηθούν οι τυχόν διαθέσιμοι επενδυτικοί πόροι ώστε να μεγιστοποιηθεί η απόδοση μιας επένδυσης που στοχεύει στην βελτίωση μιας υφιστάμενης παραγωγικής διαδικασίας. Σε αυτό το πλαίσιο, κρίνεται ως εξαιρετικά σημαντικό το συμπέρασμα που προκύπτει σχετικά με την αντιμετώπιση της διαδικασίας είτε ως ενιαία, είτε ως αποτελούμενη από επιμέρους στάδια. Συγκεκριμένα, διαπιστώθηκε ότι οι σχέσεις μεταξύ των διαφόρων μεταβλητών και η επίδραση τους στο τελικό αποτέλεσμα διαφοροποιείται σημαντικά μεταξύ των διαφόρων σταδίων της διαδικασίας. Μάλιστα, διαπιστώθηκε ότι τα μοντέλα που αντιμετωπίζουν τη διαδικασία ως ενιαία, αδυνατούν να εντοπίσουν σημαντικές αποκλίσεις της εξεταζόμενης παρτίδας από μια τυπική παρτίδα. Αντιθέτως, κατασκευάζοντας επιμέρους μοντέλα για κάθε διακριτό στάδιο της διαδικασίας, είμαστε σε θέση να αυξήσουμε την ευαισθησία τους και να εντοπίσουμε αποκλίσεις οι οποίες πιθανώς θα οδηγήσουν σε αστοχία της παρτίδας.

Σε αυτό το σημείο, αξίζει να σημειωθεί ότι η εργασία κινείται σε ακαδημαϊκό επίπεδο για να εξετάσει και να προσδιορίσει το πλαίσιο εφαρμογής της πολυμεταβλητής ανάλυσης σε μια παραγωγική διαδικασία, αξιοποιώντας πραγματικά δεδομένα για ρεαλισμό, χωρίς καμία ωστόσο δυνατότητα ή διάθεση εμπορικής εκμετάλλευσης των αποτελεσμάτων.

## 6.2 Προτάσεις για μελέτη

Η παρούσα εργασία θέτει τα θεμέλια και τις κατευθυντήριες γραμμές προκειμένου να πραγματοποιηθεί μια λεπτομερής και σε μεγαλύτερο βάθος ανάλυση της παραγωγικής διαδικασίας που μελετήθηκε. Σε αυτή την ενότητα λοιπόν, υποδεικνύονται ορισμένες πτυχές του εξεταζόμενου αντικειμένου, οι οποίες αξίζει να ληφθούν υπόψη σε μια μελλοντική του προσέγγιση.

Δεδομένου ότι βρίσκεται σε εξέλιξη μια διαδικασία αναβάθμισης του μετρητικού εξοπλισμού που χρησιμοποιείται για την συλλογή των δεδομένων, προτείνεται η δημιουργία ενός νέου συνόλου δεδομένων, αφενός αποτελούμενου από μετρήσεις υψηλότερης ποιότητας και αφετέρου αποτελούμενου από περισσότερες παρτίδες τελικού προϊόντος, τόσο επιτυχημένες όσο και αποκλίνουσες από τα επιθυμητά χαρακτηριστικά. Σε αυτό το πλαίσιο, προτείνεται και ο αυστηρότερος καθορισμός των επιθυμητών προδιαγραφών ποιότητας και των αντίστοιχων ανοχών. Επιπλέον, κρίνεται σκόπιμη η εισαγωγή περισσότερων παραμέτρων στο πρόβλημα, οι οποίες πιθανώς θα αποκαλύψουν ενδιαφέρουσες σχέσεις και αλληλεπιδράσεις συμβάλλοντας στην καλύτερη κατανόηση της διαδικασίας. Τέλος, προτείνεται η εφαρμογή του μοντέλου σε πραγματικό χρόνο εκτέλεσης της παραγωγικής διαδικασίας, η αξιολόγηση της δυνατότητας του, να προβλέπει επικείμενες αποκλίσεις της εξέλιξης μιας παρτίδας από την επιθυμητή πορεία και η αξιολόγηση της αποτελεσματικότητας των διορθωτικών χειρισμών που θα πρέπει να γίνουν, όπως αυτοί υπαγορεύονται από τα αποτελέσματα του μοντέλου, για να αποφευχθούν οι προαναφερθείσες αποκλίσεις. Προτείνεται δε, η εφαρμογή κλειστών βρόχων ανάδρασης για τον αυτόματο έλεγχο και ρύθμιση των παραμέτρων ενώ επισημαίνεται και η ανάγκη για σαφή ορισμό κριτηρίων εξόδου ανά βήμα παραγωγής τα οποία θα ορίζονται ακόμα και από την αρχή σχεδιασμού της διεργασίας βάση εργαστηριακών μετρήσεων. Ο προσδιορισμός των εν λόγω κριτηρίων ο οποίος θα μπορούσε να αποτελέσει αντικείμενο άλλης εργασίας, θα εξασφαλίσει μέγιστη αξιοπιστία και ελάχιστη ανθρώπινη συνδρομή στην παραγωγική διαδικασία.





## ΚΕΦΑΛΑΙΟ 7: *Βιβλιογραφία*

---

- [1] Real-Time Analytics for Batch Processes, By Janice Abel - <https://www.automationworld.com/article/technologies/quality-software/real-time-analytics-batch-processes>
- [2] How big data can improve manufacturing, By Eric Auschitzky, Markus Hammer, and Agesan Rajagopaul - <https://www.mckinsey.com/business-functions/operations/our-insights/how-big-data-can-improve-manufacturing>
- [3] Data Analytics in Batch Operations, By Robert Wojewodka and Terry Blevins - <http://modelingandcontrol.com/repository/Control0508.pdf>
- [4] Improving Data Analytics in Batch Manufacturing, By Jon Peterson - <https://www.chemengonline.com/improving-data-analytics-batch-manufacturing>
- [5] Κουτσογιαννακόπουλος Ιωάννης και Στρατήγης Χαρίλαος, Business Intelligence Application in Batch Process Performance Management, 2018
- [6] Σταυρινίδης, Σταύρος-Κωνσταντίνος, Παλινδρόμηση μερικών ελάχιστων τετραγώνων, 2011
- [7] Batch Production and It's Key Characteristics - <https://www.mbaknol.com/operations-management/batch-production-and-its-key-characteristics/>
- [8] Continuous vs. Batch Production, By Mark Muldowney - [https://www.contractpharma.com/issues/2018-0401/view\\_features/continuous-vs-batch-production/53360](https://www.contractpharma.com/issues/2018-0401/view_features/continuous-vs-batch-production/53360)
- [9] Lindsay I. Smith, A tutorial on Principal Components Analysis, 2002
- [10] Analyzing batch process data, a step-by-step guide, Lennart Eriksson
- [11] <https://en.wikipedia.org>
- [12] <https://learnche.org/>
- [13] <https://www.sartorius.com/en/knowledge/resources>

