



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**

**ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ**

**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ**

**ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΟΙΚΟΝΟΜΙΚΗΣ ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ  
ΜΕ ΜΕΘΟΔΟΥΣ ΑΝΑΛΥΣΗΣ ΜΕΓΑΛΩΝ ΔΕΔΟΜΕΝΩΝ**

Διπλωματική Εργασία

**Σαμαράς Καμηλαράκης Ηλίας**

**Επιβλέποντες Καθηγητές:** Ελευθέριος Τσουκαλάς

*Καθηγητής Π.Θ.*

Γεώργιος Σταμπουλής

*Επίκουρος Καθηγητής Π.Θ.*

**Βόλος 2018**



**UNIVERSITY OF THESSALY**

**SCHOOL OF ENGINEERING**

**DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEER**

**MODELLING OF ECONOMIC ACTIVITY WITH THE USE  
OF BIG DATA ANALYTICS**

Diploma thesis

**Samaras Kamilarakis Ilias**

**Supervisors:**

Eleftherios Tsoukalas

*Professor UTH*

Georgios Stamboulis

*Assistant Professor UTH*

**Volos 2018**

Υποβλήθηκε για την εκπλήρωση μέρους των  
απαιτήσεων για την απόκτηση του  
Διπλώματος Ηλεκτρολόγου Μηχανικού και Μηχανικού Ηλεκτρονικών Υπολογιστών  
2018

## Δήλωση Συγγραφής

Εγώ, ο Ηλίας Σαμαράς Καμηλαράκης δηλώνω πως αυτή η διπλωματική εργασία με τίτλο «Οικονομική Μοντελοποίηση Με Χρήση Μεθόδων Ανάλυσης και Επεξεργασίας Μεγάλων Δεδομένων» και η έρευνα και ανάλυση που παρουσιάζονται σε αυτή είναι δική μου.

Επιβεβαιώνω πως, εκτός λάθους ή παράλειψης:

- Αυτή η εργασία έγινε εξ ολοκλήρου για την εκπλήρωση μέρους των απαιτήσεων για την απόκτηση του Διπλώματος Ηλεκτρολόγου Μηχανικού και Μηχανικού Ηλεκτρονικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας.
- Όπου έχει χρησιμοποιηθεί υλικό από άλλη διπλωματική εργασία αυτού ή άλλου πανεπιστημίου σε αυτή την εργασία αυτό έχει δηλωθεί στην βιβλιογραφία.
- Όπου έχω αναφερθεί στην εργασία άλλων η πηγή πάντα δίνεται και με εξαίρεση αυτών των αναφορών η εργασία αυτή είναι εξ'ολοκλήρου δική μου.
- Έχω αναγνωρίσει όλες τις κύριες πηγές βοήθειας.

Υπογραφή

---

Ημερομηνία

---

© 2018 Ηλίας Σαμαράς Καμηλαράκης

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών της Πολυτεχνικής Σχολής του Πανεπιστημίου Θεσσαλίας δεν υποδηλώνει αποδοχή των απόψεων του συγγραφέα (Ν. 5343/32 αρ. 202 παρ. 2).

## Ευχαριστίες

Με την περαίωση αυτής της Διπλωματικής εργασίας θέλω να ευχαριστήσω τον Πρόεδρο του τμήματος και Καθηγητή Ελευθέριο Τσουκαλά για τις χρήσιμες υποδείξεις του και την δυνατότητα που μου έδωσε να ασχοληθώ με το θέμα της παρούσας διπλωματικής εργασίας και τον επιβλέποντα της διπλωματικής εργασίας μου, Επίκουρο Καθηγητή Γεώργιο Σταμπουλή, για την πολύτιμη βοήθεια και καθοδήγησή του κατά τη διάρκεια της εκπόνησης της εργασίας μου. Επίσης, ευχαριστώ για την βοήθεια και την στήριξη τους φίλους και συμφοιτητές μου που μαζί όλα αυτά τα χρόνια αγωνιστήκαμε για την περαίωση των σπουδών μας. Τέλος, είμαι ευγνώμων στους γονείς μου, Δημήτριο και Μαρία και τον αδερφό μου Σταύρο για την ολόψυχη αγάπη και υποστήριξή τους όλα αυτά τα χρόνια.

Σαμαράς Καμηλαράκης Ηλίας

# ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΟΙΚΟΝΟΜΙΚΗΣ ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ ΜΕ ΜΕΘΟΔΟΥΣ ΑΝΑΛΥΣΗΣ ΜΕΓΑΛΩΝ ΔΕΔΟΜΕΝΩΝ

Σαμαράς Καμηλαράκης Ηλίας

Πανεπιστήμιο Θεσσαλίας, Πολυτεχνική Σχολή, Τμήμα Ηλεκτρολόγων Μηχανικών και  
Μηχανικών Ηλεκτρονικών Υπολογιστών, 2018

Επιβλέποντες Καθηγητές: Ελευθέριος Τσουκαλάς, *Καθηγητής Π.Θ.*

Γεώργιος Σταμπουλής, *Επίκουρος Καθηγητής Π.Θ.*

## Περίληψη

Ο παγκόσμιος πληθυσμός συνεχώς αυξάνεται, μέχρι πρότεινος υποανάπτυκτες χώρες αναπτύσσονται ραγδαία με την εκμετάλλευση νέων τεχνολογιών και ως εκ τούτου ραγδαία αυξάνονται και οι ενεργειακές ανάγκες του πλανήτη. Ταυτόχρονα ο τομέας της ενέργειας βρίσκεται σε μια διαδικασία κοινωνικό-τεχνικής μετάβασης με σκοπό την ανάπτυξη της παραγωγής ηλεκτρικής ενέργειας από ανανεώσιμες πηγές και την απεξάρτηση του από τα ορυκτά καύσιμα. Η εισαγωγή των ΑΠΕ αλλά και νέων καινοτόμων τεχνολογιών στην αγορά έχει αλλάξει τον τρόπο με τον οποίο λειτουργεί το σύστημα ηλεκτρικής ενέργειας καθώς και το ίδιο το σύστημα μετατρέποντάς το σε ευφυές δίκτυο ηλεκτρικής ενέργειας. Η πράσινη ενέργεια έχει διαφορετική τιμολόγηση από την συμβατική ενώ όσο αλλάζει το ενεργειακό μίγμα παραγωγικής ισχύος, τόσο θα μεταβάλλονται και οι τιμές της αγοράς ηλεκτρικής ενέργειας.

Παράλληλα, τα τελευταία χρόνια, έχει σημειωθεί μια έκρηξη από δεδομένα κάνοντας απαραίτητη την γνώση εξειδικευμένων τεχνικών εξόρυξης και ανάλυσης των μεγάλων δεδομένων για οποιαδήποτε αξιόπιστη εξαγωγή συμπερασμάτων από μελέτες που γίνονται σε κάθε επιστημονικό τομέα.

Στην παρούσα διπλωματική εργασία θα εξορυχθούν τα ανοικτά μεγάλα δεδομένα από τον ΛΑΓΗΕ και με βάση αυτά θα δημιουργηθούν μοντέλα μηχανικής μάθησης για την πρόβλεψη της οριακής τιμής συστήματος.

# MODELLING OF ECONOMIC ACTIVITY WITH THE USE OF BIG DATA ANALYTICS

Samaras KamilarakisIlias

University of Thessaly, School of Engineering, Department of Electrical and Computer  
Engineering, 2018

Thesis tutors: Eleftherios Tsoukalas, *ProfessorUTH*

Georgios Stamboulis, *Assistant Professor UTH*

## Abstract

The world's population is steadily rising, and so far, underdeveloped countries are rapidly developing with the exploitation of new technologies and hence the energy needs of the planet are growing rapidly. At the same time, the energy sector is in a process of socio-technical transition aimed at developing the production of electricity from renewable sources and its decomposition from fossil fuels. The introduction of RES and new innovative technologies into the market has changed the way the electricity system works and the system itself by transforming it into an intelligent electricity grid. Green energy has a different pricing than the conventional one, and by changing the energy mix of production power, both electricity market prices will change.

At the same time in recent years there has been an explosion of data making necessary knowledge of specialized mining and analysis techniques of large data for any reliable extraction of conclusions from studies made in any scientific field.

In this diploma thesis, open large data will be extracted from the greek system administrator-LAGIE and based on these, mechanical learning models will be created to address the cost limit value..

## Πίνακας περιεχομένων

Κεφάλαιο 1	ΕΙΣΑΓΩΓΗ .....	1
1.1	Κίνητρο – Στόχος.....	1
1.2	Οργάνωση Διπλωματικής Εργασίας.....	1
Κεφάλαιο 2	ΤΟ ΣΥΣΤΗΜΑ ΗΛΕΚΤΡΙΚΗΣ ΕΝΕΡΓΕΙΑΣ.....	3
2.1	Εισαγωγή.....	3
2.2	Το Ελληνικό Σ.Η.Ε. ....	4
2.3	Τα μειονεκτήματα του παραδοσιακού ηλεκτρικού δικτύου .....	9
Κεφάλαιο 3	ΕΥΦΥΗ ΔΙΚΤΥΑ ΗΛΕΚΤΡΙΚΗΣ ΕΝΕΡΓΕΙΑΣ .....	16
3.1	Ορισμός .....	16
3.2	Χαρακτηριστικά του ΕΗΔ.....	18
3.3	Οφέλη και πιθανοί κίνδυνοι των ευφύων δικτύων ηλεκτρικής ενέργειας .....	19
3.4	Η δομή των ευφύων δικτύων ηλεκτρικής ενέργειας .....	24
3.5	Μετάδοση δεδομένων στο ευφύες δίκτυο .....	26
3.6	Χαρακτηριστικά των ευφύων δικτύων ηλεκτρικής ενέργειας .....	27
3.7	Παράδειγμα δικτύων στο έξυπνο δίκτυο .....	32
3.7.1	Μικροδίκτυο .....	32
Κεφάλαιο 4	ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ (BIG DATA) - ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΟΥΣ.....	33
4.1	Γενικά.....	33
4.2	Εξέλιξη των μεγάλων δεδομένων.....	34
4.3	State of the art εργαλεία ανάλυσης μεγάλων δεδομένων .....	36
4.4	Batch based εργαλεία ανάλυσης .....	37
4.5	Stream based εργαλεία ανάλυσης .....	39
	Apache Hadoop.....	40
4.6	Τεχνικές ανάλυσης δεδομένων .....	43
4.7	Ανάλυση δεδομένων με την γλώσσα Python.....	44
4.7.1	Ιστορική αναδρομή .....	44
4.7.2	Python και ανάλυση δεδομένων .....	45
4.7.3	Βιβλιοθήκες της Python .....	46
4.7.4	Το μοντέλο StatsModel .....	48
Κεφάλαιο 5	ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ .....	49
5.1	Γενικά.....	49
5.2	Εισαγωγή του dataset στο Jupyter Notebook .....	52
5.3	Προεπεξεργασία του dataset .....	53
5.3.1	Συσχέτιση δεδομένων .....	56
5.4	Οπτικοποίηση δεδομένων.....	57
5.4.1	Smp timeseries.....	57



5.4.2	Histogramms .....	59
5.4.3	density plots .....	60
5.4.4	Box and wisker plots .....	61
5.4.5	Scatter plot and correlation matrix .....	62
5.5	Regression Models (Παλινδρόμηση) .....	65
5.5.1	Απλή Γραμμική Παλινδρόμηση .....	68
5.5.2	Πολλαπλή γραμμική παλινδρόμηση .....	70
5.5.3	Κριτήρια καταλληλότερου μοντέλου .....	71
5.6	Τα μοντέλα παλινδρόμησης .....	74
Κεφάλαιο 6	ΣΥΜΠΕΡΑΣΜΑΤΑ .....	81
	<b>Βιβλιογραφία</b> .....	82

## Κατάλογος Εικόνων

Εικόνα 1 - Δομή του Ευφυούς Ηλεκτρικού Δικτύου .....	17
Εικόνα 2 - Τα αλυσιδωτά οφέλη του Ευφυούς Ηλεκτρικού Δικτύου .....	21
Εικόνα 3 - Η αρχιτεκτονική ενός ευφυούς δικτύου ηλεκτρικής ενέργειας .....	25
Εικόνα 4 - Χαρακτηριστικά ενός ευφυούς δικτύου ηλεκτρικής ενέργειας .....	27
Εικόνα 5 - Στο κατώτερο επίπεδο φαίνεται ένα απλό μικροδίκτυο. Στο μεσαίο η ροή της ενέργειας σε αυτό. Στο πάνω επίπεδο η ροή της πληροφορίας. ....	32
Εικόνα 6 - Μετάβαση τεχνολογίας (Πηγή: <a href="http://www.ttivanguard.com/realtime/bigdata.pdf">http://www.ttivanguard.com/realtime/bigdata.pdf</a> ).....	34

## Κατάλογος Πινάκων

Πίνακας 1- Συνολικά μήκη γραμμών μεταφοράς του Ελληνικού ΣΗΕ [4] .....	6
Πίνακας 2 - Χαρακτηριστικά συμβατικού δικτύου Η.Ε. – ευφυούς δικτύου Η.Ε. ....	19
Πίνακας 3 - Δεδομένα που παράγονται από κοινωνικά δίκτυα. (Πηγή: <a href="https://www.omnicoreagency.com/">https://www.omnicoreagency.com/</a> ) ..	36
Πίνακας 4 - Σύγκριση των batch based εργαλείων επεξεργασίας (Πηγή: Yaqoob, et al., 2016). ....	38
Πίνακας 5 - Σύγκριση stream based εργαλείων ανάλυσης (Πηγή: Yaqoob, et al., 2016). ....	40
Πίνακας 6 - Σύγκριση διαφορετικών τεχνικών ανάλυσης δεδομένων. (Πηγή: Yaqoob, et al., 2016) .....	44
Πίνακας 7 - Ημερήσια μέση θερμοκρασία έτους 2007 (Ε.Μ.Υ., σταθμός Αχαρών).....	58
Πίνακας 8 – Πίνακας συσχέτισης μεταβλητών .....	64

## Κατάλογος Διαγραμμάτων

Διάγραμμα 1 - Growth of and digitization of global information-storage capacity .....	33
Διάγραμμα 2 - Αρχιτεκτονική HDFS. (Πηγή: <a href="http://en.wikipedia.org/wiki/Google_File_System">http://en.wikipedia.org/wiki/Google_File_System</a> ) .....	41
Διάγραμμα 3 - Αρχιτεκτονική MapReduce. (Πηγή: <a href="http://architects.dzone.com/article/1es/how-hadoop-Map-Reduce-works">http://architects.dzone.com/article/1es/how-hadoop-Map-Reduce-works</a> ) .....	42

## Κατάλογος στιγμιότυπων οθόνης

Στιγμιότυπο οθόνης 1 - Μορφή δεδομένων ΛΑΓΗΕ - 1 .....	49
Στιγμιότυπο οθόνης 2 - Μορφή δεδομένων ΛΑΓΗΕ - 2 .....	50
Στιγμιότυπο οθόνης 3 - Μορφή δεδομένων ΛΑΓΗΕ - 3 .....	50
Στιγμιότυπο οθόνης 4 - Μορφή δεδομένων ΛΑΓΗΕ - 4 .....	51
Στιγμιότυπο οθόνης 5 - Μορφή Dataset - 1 .....	51
Στιγμιότυπο οθόνης 6 - Μορφή Dataset - 2 .....	52
Στιγμιότυπο οθόνης 7 - Εισαγωγή του dataset στο Jupyter Notebook .....	52
Στιγμιότυπο οθόνης 8 - Σύντομη οπτικοποίηση των δεδομένων του datasheet .....	53
Στιγμιότυπο οθόνης 9 – Χρήση εντολής <code>data = data.dropna(how = 'any',axis=1)</code> .....	54
Στιγμιότυπο οθόνης 10 – Χρήση εντολής <code>data1=data.drop(data.columns[8:], axis=1)</code> .....	55
Στιγμιότυπο οθόνης 11 - Smp timeseries .....	57
Στιγμιότυπο οθόνης 12 – Χρήση βιβλιοθήκης seaborn για τη δημιουργία ιστογραμμάτων. ....	59
Στιγμιότυπο οθόνης 13 - Ιστογράμματα .....	60
Στιγμιότυπο οθόνης 14 - Χρήση βιβλιοθήκης seaborn για τη δημιουργία densityplots .....	61
Στιγμιότυπο οθόνης 16 – Χρήση βιβλιοθήκης matplotlib για τη δημιουργία.Boxandwiskerplots.....	62
Στιγμιότυπο οθόνης 17 – Scatterplot .....	63
Στιγμιότυπο οθόνης 18 – Χρήση των βιβλιοθηκών matplotlib και seaborn για τη δημιουργία του πίνακα συσχέτισης .....	64

# Κεφάλαιο 1 ΕΙΣΑΓΩΓΗ

## 1.1 Κίνητρο – Στόχος

Στην παρούσα διπλωματική εργασία παρουσιάζονται το παραδοσιακό σύστημα ηλεκτρικής ενέργειας και τα ευφυή δίκτυα ηλεκτρικής ενέργειας. Ακόμη αναφέρεται το τι είναι τα μεγάλα δεδομένα και αναλύεται η γλώσσα προγραμματισμού Python και οι βιβλιοθήκες που χρησιμοποιούνται. Στη συνέχεια δημιουργείται το dataset από τα δεδομένα που μπορούν να εξαχθούν από τον ΛΑΓΗΕ, αναλύονται και οπτικοποιούνται. Σκοπός λοιπόν της παρούσας διπλωματικής εργασίας είναι η δημιουργία μοντέλου πρόβλεψης της οριακής τιμής συστήματος, (ΟΤΣ ή SMP) δηλαδή της τιμής στην οποία εκκαθαρίζεται η αγορά ενέργειας και αποτελεί την ενιαία τιμή στην οποία προμηθευτές και έμποροι αγοράζουν την ενέργεια που αναμένουν ότι θα απορροφήσουν από το σύστημα οι πελάτες τους και αμείβονται αντίστοιχα οι παραγωγοί και οι εισαγωγείς, με χρήση μεθόδων big data analytics οι οποίες αναλύονται διεξοδικά.

## 1.2 Οργάνωση Διπλωματικής Εργασίας

Στο παρόν 1<sup>ο</sup>Κεφάλαιο αναφέρονται το κίνητρο - στόχος και παρουσιάζεται η οργάνωση της διπλωματικής εργασίας.

Στο 2<sup>ο</sup>Κεφάλαιο παρουσιάζονται οι αρχές του παραδοσιακού συστήματος ηλεκτρικής ενέργειας καθώς και τα κυριότερα μειονεκτήματα του

Στο 3<sup>ο</sup>Κεφάλαιο παρουσιάζονται τα βασικά στοιχεία των ευφυών δικτύων ηλεκτρικής ενέργειας, τα πλεονεκτήματα και τα μειονεκτήματά τους.

Στο Κεφάλαιο 4<sup>ο</sup> Κεφάλαιο παρουσιάζεται το «τοπίο» των μεγάλων δεδομένων, τα κύρια εργαλεία επεξεργασίας μεγάλων δεδομένων, οι τρόποι συλλογής δεδομένων και γίνεται αναφορά στα ανοιχτά δεδομένα και στα εργαλεία που θα χρησιμοποιήσουμε για τη δημιουργία του μοντέλου πρόβλεψης της οριακής τιμής συστήματος.

Στο 5<sup>ο</sup> κεφάλαιο χρησιμοποιούμε τα δεδομένα του ΛΑΓΗΕ και δημιουργούμε τα μοντέλα πρόβλεψης. Αναλύετε το πως και το γιατί επιλέγονται οι συγκεκριμένες ανεξάρτητες μεταβλητές καθώς και το πιο είναι το καταλληλότερο μοντέλο.

Τέλος, στο 6<sup>ο</sup>Κεφάλαιο συνοψίζονται τα βασικά ευρήματα της διπλωματικής εργασίας.

## Κεφάλαιο 2 **ΤΟ ΣΥΣΤΗΜΑ ΗΛΕΚΤΡΙΚΗΣ ΕΝΕΡΓΕΙΑΣ**

### 2.1 Εισαγωγή

Σύστημα Ηλεκτρικής Ενέργειας (Σ.Η.Ε.) είναι το σύνολο των εγκαταστάσεων και των μέσων τα οποία χρησιμοποιούνται για την παροχή ηλεκτρικής ενέργειας στις εξυπηρετούμενες περιοχές. Η παροχή ηλεκτρικής ενέργειας σε κάθε σημείο ζήτησης, με σταθερή τάση και συχνότητα, υψηλή αξιοπιστία τροφοδότησης, το ελάχιστο δυνατό κόστος και την ελάχιστη δυνατή περιβαλλοντική επιβάρυνση, καθιστούν τη λειτουργία ενός Σ.Η.Ε. επιτυχή. [1].

Το Σ.Η.Ε. αποτελείται από τα ακόλουθα συστήματα:

- Σύστημα παραγωγής: Αποτελείται από τους σταθμούς παραγωγής ηλεκτρικού ρεύματος και τους μετασχηματιστές ανύψωσης τάσης για τη μεταφορά ρεύματος από υπερυψηλή και υψηλή τάση. Η μεγαλύτερη ποσότητα ηλεκτρικής ενέργειας προέρχεται σήμερα από τη μετατροπή της θερμικής ενέργειας των ορυκτών καυσίμων και της μηχανικής ενέργειας των υδάτινων ροών και υδατοπτώσεων σε ηλεκτρική ενέργεια, μέσω των θερμικών και υδροηλεκτρικών σταθμών αντίστοιχα.
- Σύστημα μεταφοράς: Αποτελεί τη βάση του Σ.Η.Ε.. Λειτουργεί στη μέγιστη δυνατή τάση και συνδέει τους μεγάλους σταθμούς παραγωγής και τα διαφορετικά συστήματα μεταξύ τους και μεταφέρει μεγάλα μεγέθη ισχύος σε μεγάλες αποστάσεις προς τα μεγάλα κέντρα κατανάλωσης. Αποτελείται από τα δίκτυα των γραμμών υπερυψηλής και υψηλής τάσης, τους υποσταθμούς ζεύξης των συγκεκριμένων δικτύων και τους υποσταθμούς μετασχηματισμού τάσης.
- Σύστημα υπομεταφοράς: Μεταφέρει ισχύ σε μικρότερα μεγέθη και αποστάσεις υπό χαμηλότερη τάση, από υποσταθμούς μεταφοράς σε υποσταθμούς διανομής μικρότερων κέντρων κατανάλωσης. Από αυτό το σύστημα τροφοδοτούνται, συνήθως απευθείας, οι μεγάλοι βιομηχανικοί καταναλωτές. Δεδομένου ότι οι παλαιότερες γραμμές μεταφοράς λειτουργούν κάτω από χαμηλότερα επίπεδα τάσης, λόγω των αναγκών επέκτασης επέκταση του Σ.Η.Ε. και της αναγκαιότητας που προκύπτει για μεταφορά από υψηλότερα επίπεδα τάσης, τα δίκτυα υπομεταφοράς και μεταφοράς δεν είναι δυνατόν να διακριθούν με ευχέρεια.
- Σύστημα διανομής: Αποτελείται από τα δίκτυα διανομής ηλεκτρικής ενέργειας μέσης και χαμηλής τάσης, περιλαμβανομένων των υποσταθμών διανομής, όπου η μέση τάση υποβιβάζεται σε χαμηλή και εξυπηρετούνται οι καταναλωτές μέσης και χαμηλής τάσης.

Τα συστήματα παραγωγής και μεταφοράς είναι δυνατόν να λειτουργούν είτε απομονωμένα, είτε συνδεδεμένα με ένα ή περισσότερα γειτονικά συστήματα, συνήθως σε εθνικό επίπεδο, που αποτελεί τη βέλτιστη από τεχνικής και την πλέον συμφέρουσα από οικονομικής απόψεως λύση.

Οι σημαντικότερες παράμετροι για το σχεδιασμό ενός ηλεκτρικού δικτύου είναι [2]

- Η τάση του δικτύου, δηλ. η μέγιστη τάση λειτουργίας των ηλεκτρικών γραμμών.
- Η ισχύς βραχυκύκλωσης του δικτύου, ήτοι η συμβατική ισχύς που αντιστοιχεί στη μέγιστη αποδιδόμενη στο δίκτυο ισχύ, σε περίπτωση που συμβεί τριφασικό βραχυκύκλωμα σε κάποιο σημείο του.
- Η στάθμη μόνωσης του δικτύου, δηλ. η διηλεκτρική αντοχής της μόνωσης του εξοπλισμού των υποσταθμών σε κρουστικές υπερτάσεις τυποποιημένης μορφής.

## **2.2 Το Ελληνικό Σ.Η.Ε.**

Το ελληνικό σύστημα μεταφοράς αποτελείται από το διασυνδεδεμένο σύστημα της ηπειρωτικής Ελλάδας και των συνδεδεμένων με αυτό νησιών, στα επίπεδα υψηλής τάσης 150 kV και 66 kV και υπερυψηλής τάσης 400 kV. Ο κορμός του αποτελείται από τρεις γραμμές διπλού κυκλώματος 400 kV, οι οποίες αναλαμβάνουν τη μεταφορά ηλεκτρικής ενέργειας, κατά κύριο λόγο από το σύστημα παραγωγής της Δυτικής Μακεδονίας, όπου παράγεται το μεγαλύτερο μέρος της ηλεκτρικής ενέργειας (περίπου 70%). Η ηλεκτρική ενέργεια μεταφέρεται από εκεί στα σημαντικότερα κέντρα κατανάλωσης της κεντρικής και νοτίου Ελλάδος, όπου και καταγράφεται το 65% της συνολικής κατανάλωσης ηλεκτρικής ενέργειας της χώρας. Παλαιότερα, το σύστημα μεταφοράς παρουσίαζε σημαντικά προβλήματα, ιδιαίτερα τη θερινή περίοδο, λόγω της άνισης γεωγραφικής κατανομής μεταξύ της παραγωγής και της κατανάλωσης. Τα προβλήματα αυτά έχουν πλέον περιοριστεί, λόγω της εισαγωγής νέων μονάδων παραγωγής στο νότιο σύστημα, την αξιοποίηση των πυκνωτών αντιστάθμισης και τη μείωση των φορτίων. Παρ' όλα αυτά, η Αττική και η Πελοπόννησος εξακολουθούν να αποτελούν τις κρισιμότερες περιοχές του συστήματος, όσον αφορά την ευστάθεια της τάσης [3].

Οι συνιστώσες του Ελληνικού Συστήματος είναι [4]:

Υποσταθμοί 150 kV/MT:

Μέχρι το τέλος του 2015 ήταν συνδεδεμένοι στο σύστημα:

- 202 υποσταθμοί υποβιβασμού 150 kV/MT, που εξυπηρετούν τις ανάγκες των πελατών του δικτύου διανομής, από τους οποίους:
  - 188 περιλαμβάνουν τμήματα τα οποία βρίσκονται υπό την κυριότητα και διαχείριση του ΑΔΜΗΕ. Στους συγκεκριμένους περιλαμβάνονται και 21 υποσταθμοί στους οποίους είναι συνδεδεμένοι και μετασχηματιστές ανυψώσεως (16 συμβατικών σταθμών παραγωγής και 6 σταθμών ΑΠΕ) και
  - 14 υποσταθμοί συνδεδεμένοι στην πλευρά 150 kV των κέντρων υπερυψηλής τάσης.
    - 14 υποσταθμοί οι οποίοι εξυπηρετούν τις ανάγκες του δικτύου διανομής στην Αττική και ανήκουν αποκλειστικά στην αρμοδιότητα του ΔΕΔΔΗΕ.
- 16 υποσταθμοί υποβιβασμού 150 kV/MT της ΔΕΗ Α.Ε., από τους οποίους:
  - 4 εξυπηρετούν την τροφοδότηση των φορτίων ορυχείων. Ανάγκες ορυχείων εξυπηρετούνται και από τον Υ/Σ 1 της Πτολεμαΐδας, ο οποίος ανήκει στους 188 προαναφερθέντες Υ/Σ οι οποίοι εξυπηρετούν ανάγκες διανομής.
  - 1 Υ/Σ εξυπηρετεί τις ανάγκες άντλησης του ΥΗΣ Πολυφύτου.
- 48 υποσταθμοί για την υποδοχή της ισχύος μονάδων ΑΠΕ, από τους οποίους οι Υ/Σ Καρύστου, Λιβαδίου, Αργυρού, Λούρου, Σκάλας και Πύλου εξυπηρετούν παράλληλα και φορτία διανομής (συμπεριλαμβάνονται στους προαναφερθέντες 48 υποσταθμούς ΑΠΕ).
- 3 υποσταθμοί ανύψωσης σε θερμικούς σταθμούς παραγωγής ανεξάρτητων παραγωγών. Οι μονάδες παραγωγής των συγκεκριμένων σταθμών συνδέονται στα 150 kV μέσω μετασχηματιστή ανύψωσης MT/150 kV.
- 38 υποσταθμοί υποβιβασμού 150 kV/MT οι οποίοι εξυπηρετούν τις εγκαταστάσεις πελατών υψηλής τάσης, από τους οποίους οι Υ/Σ Αλουμινίου και ΜΟΤΟΡ ΟΪΛ εξυπηρετούν ταυτόχρονα και τη σύνδεση σταθμών παραγωγής (ο πρώτος συμπεριλαμβάνεται στους προαναφερθέντες 3 Υ/Σ ανυψώσεων συμβατικών σταθμών παραγωγής και ο δεύτερος στους 48 Υ/Σ ΑΠΕ).

#### Κέντρα υπερυψηλής τάσης:

Τα Κέντρα υπερυψηλής τάσης λειτουργούν ως σημεία διασύνδεσης των συστημάτων 400 kV και 150 kV και εξυπηρετούν ανάγκες σχετικές με την απομάστευση ισχύος από το σύστημα των 400 kV προς το σύστημα των 150 kV. Υπάρχουν 14 κέντρα υπερυψηλής τάσης



που περιλαμβάνουν έναν ή περισσότερους αυτομετασχηματιστές τριών τυλιγμάτων 400 kV/150kV/30kV. Επίσης, υπάρχουν 11 κέντρα υπερυψηλής τάσης (εκτός από τα 14 που προαναφέρθηκαν), τα οποία είναι εγκατεστημένα δίπλα από τους ομόνυμους σταθμούς παραγωγής και εξυπηρετούν ανάγκες ανύψωσης τάσης από τις μονάδες παραγωγής προς το σύστημα 400 kV.

#### Γραμμές μεταφοράς:

Υπάρχουν γραμμές μεταφοράς υψηλής (66 και 150 kV) και υπερυψηλής (400 kV) τάσης διαφόρων ειδών. Επιπλέον, υπάρχουν εγκατεστημένα υπόγεια καλώδια 150 kV τα οποία μεταφέρουν ισχύ στις πυκνοκατοικημένες περιοχές της Αθήνας και ανήκουν στο δίκτυο 150 kV. Τα σχετικά στοιχεία παρατίθενται στον πίνακα που ακολουθεί.

<b>ΕΠΙΠΕΔΟ ΤΑΣΗΣ (kV)</b>	<b>ΕΙΔΟΣ Γ.Μ.</b>	<b>ΣΥΝΟΛΙΚΟ ΜΗΚΟΣ (km)</b>
66	Εναέριες	39
	Υποβρύχιες	15
150	Εναέριες	8157
	Υπόγειες	161
	Υποβρύχιες	208
400	Εναέριες	2647
	Υπόγειες	31
	Εναέριες Σ.Ρ.	107
	Υποβρύχιες Σ.Ρ.	160 <sup>6</sup>

*Πίνακας 1- Συνολικά μήκη γραμμών μεταφοράς του Ελληνικού Σ.Η.Ε. [4]*

#### Συσκευές αντιστάθμισης άεργης ισχύος

Οι ανάγκες της αντιστάθμισης άεργης ισχύος καλύπτονται από στατούς πυκνωτές και πηνία. Για την τοπική στήριξη των τάσεων στους υποσταθμούς 150 kV/MT χρησιμοποιούνται στατοί πυκνωτές που εγκαθίστανται κατά βάση στους ζυγούς μέσης τάσης των υποσταθμών (συνολικής ισχύος περίπου 4150 MVar). Επίσης, έχουν εγκατασταθεί συστοιχίες πυκνωτών 150 kV, συνολικής ισχύος 450 MVar σε υποσταθμούς και κέντρα υψηλής τάσης του συστήματος. Οι κατασκευαστικές υπηρεσίες του ΑΔΜΗΕ έχουν

εγκαταστήσει και συστήματα αυτόματης ένταξης/απένταξης πυκνωτών σε βαθμίδες (3x4 MVar) σε υποσταθμούς μέσης και υψηλής τάσης.

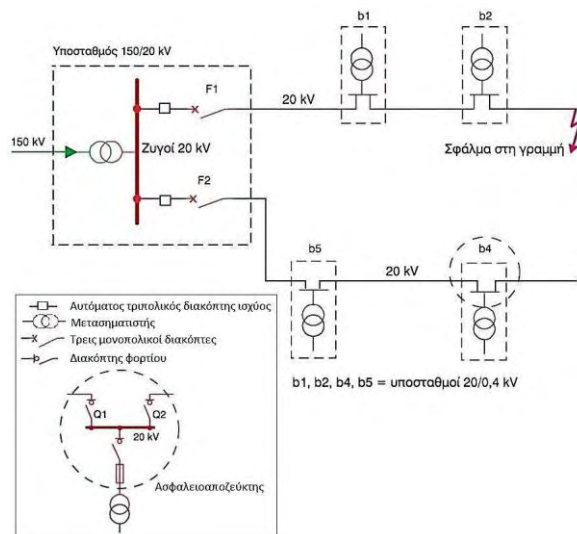
Επιπλέον, έχουν εγκατασταθεί πηνία στην πλευρά 150 kV σε υποσταθμούς 150 kV/MT (εκείνους στους οποίους συνδέονται υποβρύχια καλώδια), όπως και στο τριτεύον τύλιγμα (πλευρά 30 kV) των ΑΜ/Σ των κέντρων υψηλής τάσης, έτσι ώστε να αντιμετωπίζονται προβλήματα που σχετίζονται με την εμφάνιση υψηλών τάσεων στις ώρες χαμηλού φορτίου.

Το δίκτυο διανομής ηλεκτρικής ενέργειας αποτελείται από το δίκτυο μέσης τάσης και χαμηλής τάσης από τα κέντρα υπερευψηλής τάσης μέχρι και τον τελικό καταναλωτή. Αποτελείται από τους υποσταθμούς μέσης τάσης των 20 kV/0,4 kV, τις εναέριες γραμμές, τα καλώδια μέσης και χαμηλής τάσης και τον εξοπλισμό προστασίας και ελέγχου. Στο δίκτυο διανομής συνδέονται καταναλωτές μέσης (στα 20 kV) και χαμηλής (400 V – 230 V) τάσης.

Υπάρχουν δύο υποκατηγορίες συστημάτων μέσης τάσης, ανάλογα με τη δομή ή τον τρόπο αξιοποίησης τους:

#### 1. Ακτινικό Δίκτυο Διανομής (Radial Main Distribution System)

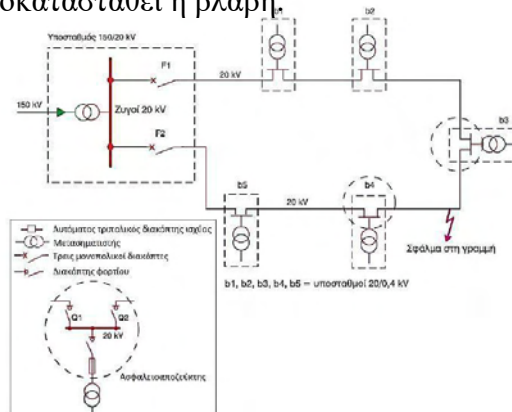
Στα ακτινικά δίκτυα διανομής οι γραμμές των 20 kV (συνήθως εναέριες) ξεκινούν από τον κεντρικό υποσταθμό 150/20 kV και αναπτύσσονται ακτινικά (σχ. 1). Κατά μήκος κάθε γραμμής συνδέονται μετασχηματιστές 20/0,4 kV των δικτύων χαμηλής τάσης ή καταναλωτές μέσης τάσης. Σημαντικό μειονέκτημα της κατηγορίας αυτής, είναι το γεγονός ότι σε περίπτωση που προκύψει σφάλμα κατά μήκος της γραμμής ο διακόπτης ισχύος F1 ανοίγει, το οποίο έχει ως αποτέλεσμα να παραμείνουν χωρίς τάση όλοι οι μετασχηματιστές που είναι τοποθετημένοι κατά μήκος της γραμμής. Όταν υπάρχουν καταναλωτές (όπως νοσοκομεία, στρατιωτικές εγκαταστάσεις, ψυκτικοί θάλαμοι, κ.λ.π.) που εξυπηρετούνται από τα δίκτυα αυτά και δεν πρέπει να μείνουν εκτός λειτουργίας για μεγάλο χρονικό διάστημα, η απώλεια της μέσης τάσης αντιμετωπίζεται με την τοποθέτηση τοπικού ηλεκτροπαραγωγού ζεύγους, μέχρι να αποκατασταθεί η βλάβη. Τα συγκεκριμένα δίκτυα σπάνια συναντώνται στη μέση τάση, ενώ αποτελούν το σύνολο των δικτύων χαμηλής τάσης.



Σχήμα 1 - Ακτινικό δίκτυο διανομής

## 2. Βροχοειδές δίκτυο διανομής (Ring Main Distribution System)

Σε αυτά τα δίκτυα διανομής οι γραμμές των 20 kV (εναέριες ή υπόγειες) που αναχωρούν από τον κεντρικό υποσταθμό 150/20 kV σχηματίζουν ένα κλειστό βρόχο που καταλήγει σε ζυγούς των 20 kV του ίδιου ή άλλου υποσταθμού. Κατά μήκος του βρόχου συνδέονται οι καταναλωτές (b1, b2, b3, b4, b5 στο παρακάτω σχήμα 2). Οι δύο άκρες του βρόχου προστατεύονται με διακόπτες ισχύος F1, F2. Στην περίπτωση κατά την οποία προκύψει σφάλμα σε κάποιο σημείο του βρόχου, ενεργοποιούνται οι προστασίες των διακοπών, οι διακόπτες ανοίγουν και δεν υπάρχει τάση στον βρόχο. Μετά τον εντοπισμό του σημείου του σφάλματος, ανοίγουν έπειτα από σχετική εντολή οι διακόπτες φορτίων (Q1 και Q2 στο επόμενο σχήμα 2) και απομονώνεται το τμήμα b3, b4. Στη συνέχεια οι διακόπτες κλείνουν ξανά και επανέρχεται η μέση τάση στο δίκτυο. Το δίκτυο λειτουργεί ως δύο ακτινικά δίκτυα έως ότου αποκατασταθεί η βλάβη.



Σχήμα 2 - Βροχοειδές δίκτυο διανομής

### 2.3 Τα μειονεκτήματα του παραδοσιακού ηλεκτρικού δικτύου

Τα παραδοσιακά δίκτυα διανομής είναι ιδιαίτερα εκτεταμένα, αλλά σχεδόν στην πλειοψηφία τους παθητικά, χωρίς να διαθέτουν ιδιαίτερες δυνατότητες επικοινωνίας, ενώ και ο έλεγχος τους είναι ελάχιστος και τοπικά περιορισμένος. Εκτός από τους πολύ μεγάλους καταναλωτές (όπως οι μεταλλουργικές βιομηχανίες) δεν υπάρχει καμία δυνατότητα εποπτείας της τάσης ή του ρεύματος ενός φορτίου σε πραγματικό χρόνο. Η όλη αλληλεπίδραση μεταξύ φορτίων και συστήματος ηλεκτρικής ενέργειας περιορίζεται απλά στη διασφάλιση της κάλυψης των απαιτήσεων ηλεκτρικής ενέργειας των καταναλωτών.

Η λειτουργία του παραδοσιακού συστήματος θεωρείται μη αποδοτική, ανελαστική και δαπανηρή. Οι σύγχρονες τεχνολογικές εξελίξεις των τηλεπικοινωνιών και του διαδικτύου παρέχουν τις δυνατότητες για εξαιρετικά αποτελεσματικότερη εποπτεία και πολύ καλύτερο έλεγχο των συστημάτων ηλεκτρικής ενέργειας, συμβάλλοντας στη βελτίωση της αποδοτικότητας και της ευελιξίας τους, καθώς και στην εξοικονόμηση οικονομικών πόρων [5].

Τα περισσότερα παραδοσιακά συστήματα υιοθετούν το μοντέλο της συγκεντρωτικής παραγωγής ηλεκτρικής ενέργειας. Σύμφωνα με αυτή την προσέγγιση, η ηλεκτρική ενέργεια παράγεται από μεγάλες μονάδες και μεταφέρεται στους καταναλωτές μέσα από τα δίκτυα μεταφοράς και διανομής.

Η συγκεκριμένη προσέγγιση παρουσιάζει εξαιρετικά υψηλά λειτουργικά έξοδα. Οι μονάδες της συγκεντρωτικής παραγωγής πρέπει ανά πάσα στιγμή να είναι σε θέση να καλύψουν τη συνολική ενεργειακή ζήτηση, γεγονός το οποίο απαιτεί [1], [2] πολύ μεγάλες δαπάνες για τη δημιουργία νέων μονάδων, ή αναβάθμιση των υπαρχουσών. Η εγκατάσταση μίας μεγάλης μονάδας (άνω των 500 MW) πρέπει να είναι εντεταγμένη στο μακροχρόνιο δημόσιο ενεργειακό προγραμματισμό, δεδομένου ότι αφ' ενός μεν απαιτεί σημαντικού ύψους κεφάλαια, αφ' ετέρου δε ο χρόνος υλοποίησης του έργου είναι της τάξης της πενταετίας, ενώ ταυτόχρονα οι μονάδες λειτουργούν σε χαμηλότερη ισχύ από την ονομαστική τους (περίπου στο 90%). Η υψηλή εφεδρεία οφείλεται στην κάλυψη πιθανής στιγμιαίας αύξησης της ζήτησης ή στην απότομη μείωση της παραγωγής ηλεκτρικής ενέργειας από τις ΑΠΕ. Με αυτό το πλαίσιο λειτουργίας, παρουσιάζεται σημαντική αύξηση της δαπάνης σε καύσιμα, δεδομένου ότι η

βέλτιστη απόδοση της ηλεκτρικής ισχύος εξόδου, σε σχέση με την κατανάλωση καυσίμου, σημειώνεται στο επίπεδο της ονομαστικής ισχύος των γεννητριών.

Το σύστημα μεταφοράς εκτείνεται σε μεγάλες εκτάσεις και απαιτεί γραμμές μήκους πολλών χιλιομέτρων, γεγονός το οποίο έχει ως συνέπεια [1],[2]

- Υψηλές δαπάνες σε εξοπλισμό (γραμμές, μετασχηματιστές, μονωτήρες, ηλεκτρονικά ισχύος).υψηλής και υπερυψηλής τάσης, προκειμένου να ελαχιστοποιούνται οι απώλειες μεταφοράς.
- Αυξημένες δαπάνες επέκτασης και συντήρησης του δικτύου και αποκατάστασης βλαβών. Λόγω της διέλευσης των γραμμών του συστήματος μεταφοράς στην πλειοψηφία τους από δυσπρόσιτες περιοχές όπου δεν υπάρχει οδικό δίκτυο. Επιπλέον, η επέκταση του δικτύου απαιτεί το σχεδιασμό της στα πλαίσια μακροχρόνιου ενεργειακού σχεδιασμού, ενώ η κατασκευή και η συντήρηση του απαιτούν ειδικό εξοπλισμό αυξημένες δαπάνες. Λόγω δε, των περιορισμένων δυνατοτήτων εποπτείας, είναι δυσχερής ο ακριβής εντοπισμός και δαπανηρές οι αποκαταστάσεις των βλαβών.
- Υψηλές θερμικές απώλειες, επειδή οι απώλειες ηλεκτρικής ενέργειας σε τόσο μεγάλες γραμμές μεταφοράς δεν μπορεί εκλείψουν, παρά τη λειτουργία σε επίπεδα υψηλής και υπερυψηλής τάσης. Στην περίπτωση κατά την οποία παρουσιάζεται αυξημένη ροή ηλεκτρικής ενέργειας, οι θερμικές απώλειες στις γραμμές μεταφοράς, λόγω της αυξημένης έντασης του ρεύματος, αυξάνονται σημαντικά

Για τη μελέτη των συστημάτων ηλεκτρικής ενέργειας χρησιμοποιείται η μέθοδος της Ανάλυσης Ροής Φορτίου. Η μέθοδος αυτή υπολογίζει [2], [6] (κατά μέτρο και γωνία) τις άγνωστες τάσεις των ζυγών και τις άγνωστες ροές ισχύος στις γραμμές του συστήματος ούτως, με σκοπό την επιλογή του οικονομικότερου σημείου λειτουργίας των γεννητριών του συστήματος, τη διατήρηση των τάσεων των ζυγών και των ροών στις γραμμές εντός των προκαθορισμένων ορίων λειτουργίας τους

Σημαντικό μειονέκτημα στην παραπάνω προσέγγιση, αποτελεί το γεγονός του ότι για την ανάλυση ροής φορτίου των υποδικτύων του συστήματος ηλεκτρικής ενέργειας ορίζεται ένας ζυγός αναφοράς, ο οποίος αντιπροσωπεύει ουσιαστικά το σημείο σύνδεσης του εξεταζόμενου υποδικτύου με το υπόλοιπο σύστημα, ενώ λειτουργεί ως ρυθμιστής. Απαιτείται δε, να διατηρεί το μέτρο και τη γωνία της τάσης σταθερά, καθώς επίσης και να παρέχει ή να

απορροφά την απαιτούμενη ισχύ για την παραμονή του υποδικτύου εντός των ορίων ευσταθείας. Οι ζυγοί αναφοράς των υποδικτύων συνδέονται συνήθως με το σύστημα μεταφοράς, ή τα ανώτερα επίπεδα των δικτύων διανομής. Το γεγονός αυτό έχει ως αποτέλεσμα το σύστημα μεταφοράς και ένα μέρος των δικτύων διανομής, να πρέπει να ανταποκρίνονται διαρκώς στις απαιτήσεις των υποδικτύων (τα οποία παρουσιάζουν στοχαστική συμπεριφορά) καθώς επίσης και να διατηρούν τα επίπεδα τάσης και συχνότητας εντός των προδιαγραφόμενων ορίων.

Η στοχαστική και ανελαστική συμπεριφορά της ζήτησης, σε συνδυασμό με την ανεξέλεγκτη παραγωγή εντός των δικτύων διανομής (όπως βιομηχανικές γεννήτριες, ΑΠΕ), καθιστούν εξόχως σημαντική τη λειτουργία του συστήματος μεταφοράς και των μεγάλων μονάδων παραγωγής ως καθοριστικών παραγόντων διατήρησης της ευστάθειας του συστήματος ηλεκτρικής ενέργειας, η οποία για να διασφαλισθεί, απαιτεί συνεχώς νέες επενδύσεις.

Σύμφωνα με τη θεωρία της Διεσπαρμένης Παραγωγής, η οποία διατυπώθηκε στις αρχές της δεκαετίας του 1990, προτείνεται η σύνδεση μικρών μονάδων παραγωγής απευθείας στο δίκτυο διανομής για την αντιμετώπιση των μειονεκτημάτων της συγκεντρωτικής παραγωγής. Η ισχύς της κυμαίνεται μεταξύ 1 KW και 100 MW ανά εγκατάσταση. Οι εγκαταστάσεις της περιλαμβάνουν μονάδες ΑΠΕ και μονάδες συμπαραγωγής ηλεκτρισμού – θερμότητας [7].

Το μοντέλο αυτό παρουσιάζει τα εξής πλεονεκτήματα:

- Ελαχιστοποίηση των απωλειών και του κόστους μεταφοράς, με αποτέλεσμα τη μείωση του συνολικού κόστους παροχής σε ποσοστό πάνω από 30%, δεδομένου ότι η ηλεκτρική ενέργεια παράγεται στην περιοχή κατανάλωσης
- Αποσυμφόρηση των υφισταμένων δικτύων.
- Δυνατότητα αξιοποίησης της θερμικής ενέργειας σε εφαρμογές συμπαραγωγής, και ως εκ τούτου βελτίωση της συνολικής απόδοσης του συστήματος.
- Ευνοεί τους καταναλωτές οι οποίοι διαθέτουν θερμικά φορτία μέσω των εφαρμογών συμπαραγωγής, εκείνους οι οποίοι έχουν πρόσβαση σε οικονομικά καύσιμα (όπως είναι το φυσικό αέριο) και εκείνους οι οποίοι ζουν σε περιοχές στις οποίες παρέχεται η δυνατότητα αξιοποίησης των ΑΠΕ.

□ Αξιοποιεί σε μεγάλο βαθμό τις ΑΠΕ, με αποτέλεσμα τη μείωση της κατανάλωσης των ορυκτών καυσίμων και των εκπομπών των αερίων του θερμοκηπίου και άλλων επιβλαβών αερίων (όπως οξείδια του θείου και του αζώτου), συμβάλλοντας στην προστασία του περιβάλλοντος.

□ Παρέχει τη δυνατότητα ανάπτυξης νέων τεχνολογιών, μεταξύ των οποίων ανανεώσιμες τεχνολογίες, οι οποίες παρέχουν ηλεκτρική ισχύ μικρής κλίμακας θέσεις κοντά στην κατανάλωση. Μέσω των τεχνολογιών αυτών, δημιουργούνται νέες προοπτικές στην ενεργειακή αγορά και αύξηση του βιομηχανικού ανταγωνισμού.

□ Καθίσταται ευκολότερη η επιλογή και εύρεση θέσεων εγκατάστασης για ΑΠΕ και μονάδες διεσπαρμένης παραγωγής, σε σχέση με τα μεγάλα εργοστάσια παραγωγής ηλεκτρικής ενέργειας, ενώ επιπλέον οι συγκεκριμένες μονάδες συνδέονται ευκολότερα και ταχύτερα στο δίκτυο.

Η Διεσπαρμένη Παραγωγή στα συστήματα ηλεκτρικής ενέργειας δεν είναι σήμερα ευρέως διαδεδομένη, λόγω περιορισμών σχετικά με τις προδιαγραφές της ευστάθειας των συστημάτων και την ποιότητα της παρεχόμενης ενέργειας.

Οι προδιαγραφές για την ομαλή λειτουργία του δικτύου, κατά τη σύνδεση μιας μονάδας παραγωγής ηλεκτρικής ενέργειας σε αυτό, είναι οι ακόλουθες[8]:

- **Επάρκεια των στοιχείων του δικτύου:** Τα στοιχεία του δικτύου πρέπει να βρίσκονται σε θέση να καλύψουν τις ανάγκες λειτουργίας και προστασίας του δικτύου και της εγκατάστασης. Τα βασικά στοιχεία είναι ο υποσταθμός της εγκατάστασης, οι μετασχηματιστές, οι γραμμές του δικτύου και τα μέσα ζεύξης και προστασίας.
- **Συμβολή στο ρεύμα βραχυκύκλωσης του σημείου σύνδεσης:** Η ισχύς μιας μονάδας παραγωγής επιδρά στην ισχύ βραχυκύκλωσης του δικτύου και επομένως στα ρεύματα βραχυκύκλωσης τα οποία προκύπτουν σε ενδεχόμενες βλάβες του δικτύου.
- **Αργές μεταβολές της τάσης:** Υπό τον όρο αυτό εννοούνται οι μεταβολές της τάσης μόνιμης κατάστασης του δικτύου οι οποίες υπολογίζονται από τη διακύμανση του μέσου όρου της τάσης σε διάστημα 10 λεπτών από την ονομαστική τιμή της τάσης. Οι συγκεκριμένες μεταβολές προκύπτουν είτε λόγω αντίστοιχων διακυμάνσεων της ισχύος εξόδου των εγκαταστάσεων παραγωγής είτε λόγω μεταβολών του φορτίου του δικτύου.
- **Ταχείες μεταβολές της τάσης:** Υπό τον όρο αυτό εννοούνται οι διάφορες μεταβολές της τάσης στις οποίες παρατηρούνται χρονικές κλίμακες έως και ορισμένα δευτερόλεπτα.

Οφείλονται σε εργασίες που διεξάγονται στις εγκαταστάσεις παραγωγής (για παράδειγμα, ζεύξη – απόζευξη, αλλαγή γεννητριών), αλλά και στη μεταβλητότητα της ισχύος εξόδου.

- **Εκπομπές flicker:** Οφείλονται στη διακύμανση της φωτεινότητας στους λαμπτήρες πυράκτωσης. Δημιουργούνται λόγω διακυμάνσεων στην τάση οι οποίες προκαλούνται από τις εγκαταστάσεις παραγωγής ηλεκτρικής ενέργειας. Η κυριότερη πηγή των εκπομπών αυτών είναι οι ανεμογεννήτριες.
- **Αρμονικές συνιστώσες τάσης και ρεύματος:** Η τάση και το ρεύμα του δικτύου σημειώνουν μία απόκλιση σε σχέση με την ιδανική καθαρά ημιτονοειδή κυματομορφή. Οι εξάρσεις και βυθίσεις που σημειώνονται στις κυματομορφές προσδίδουν πριονωτή μορφή, ενώ αποτελούν μικρές κυματομορφές με αρμονικές συχνότητες πολλαπλάσιες της θεμελιώδους συχνότητας της τάσης και του ρεύματος και των οποίων είναι δυνατός ο εντοπισμός και η ανάλυση. Οι εγκαταστάσεις παραγωγής και τα μη γραμμικά κυκλώματα δημιουργούν αρμονικές συνιστώσες οι οποίες πρέπει να καταπιέζονται.

Ακόμη όμως και εάν τηρούνται οι συγκεκριμένες προδιαγραφές, ενδέχεται να δημιουργηθούν σημαντικά προβλήματα από τη σύνδεση των μονάδων παραγωγής στο δίκτυο διανομής τα οποία οφείλονται στη μορφή του δικτύου, το οποίο συνήθως είναι ακτινικό ή βροχοειδές αλλά λειτουργεί ως ακτινικό με συνέπεια η ροή ισχύος να είναι προς μία κατεύθυνση. Ένα επίσης σημαντικό ζήτημα είναι η αδυναμία του διαχειριστή του δικτύου να παρακολουθεί τη ροή ενέργειας στο δίκτυο διανομής σε πραγματικό χρόνο.

Τα κυριότερα προβλήματα που προκύπτουν είναι [5], [8]:

- Υπερτάσεις σε περιπτώσεις χαμηλών εξυπηρετούμενων φορτίων, για την αντιμετώπιση του οποίου απαιτείται συντονισμένη λειτουργία των μονάδων και ειδικός εξοπλισμός, όπως είναι οι μετασχηματιστές μεταβλητού λόγου σχηματισμού για το συνεχή έλεγχο της τάσης.
- Μεταβολές της συχνότητας του συστήματος ηλεκτρικής ενέργειας. Η συγκεκριμένη συχνότητα εξαρτάται από το ισοζύγιο παραγωγής – κατανάλωσης. Οποιαδήποτε απόκλιση της από τα συγκεκριμένα όρια, οδηγεί είτε σε απόκλιση από την επιθυμητή συχνότητα του ευρωπαϊκού δικτύου (50 Hz), είτε σε υψηλή ροή ισχύος στη γραμμή σύνδεσης του τοπικού δικτύου με τα γειτονικά του, η οποία προξενεί καταπόνηση του εξοπλισμού του δικτύου διανομής.



Το σημαντικότερο μέρος της Διεσπαρμένης Παραγωγής σχετίζεται με τις ΑΠΕ. Η διείδυση όμως των ΑΠΕ περιορίζεται σε χαμηλά επίπεδα λόγω:

- Της μειωμένης δυνατότητας βραχυπρόθεσμης πρόβλεψης της παραγωγής των ανεμογεννητριών και των φωτοβολταϊκών. Λόγω της στοχαστικότητας των καιρικών φαινομένων σημειώνεται και μεταβαλλόμενη ισχύς εξόδου στις εγκαταστάσεις. Κατά συνέπεια είναι δύσκολο να προγραμματιστεί από τον διαχειριστή η παραγωγή σε βραχυπρόθεσμη βάση.
- Της έλλειψης ευελιξίας, όσον αφορά την παραγωγή, των μεγάλων κεντρικών εγκαταστάσεων, επειδή παρουσιάζουν χρονικό διάστημα έναρξης της τάξεως των 5 ωρών, ενώ μετά την εκκίνηση τους δεν είναι δυνατή η λειτουργία τους σε ποσοστό χαμηλότερο του 50% – 65% της ονομαστικής τους ισχύος.
- Της εκ προτέρων συμφωνίας ορισμένων παραγωγών για εγγυημένη απορρόφηση της παραγωγής τους (π.χ. μονάδες φυσικού αερίου), γεγονός το οποίο λειτουργεί ανασταλτικά στη διείδυση των ΑΠΕ στη συνολική παραγωγή.

Προκειμένου η Διεσπαρμένη Παραγωγή στο σύστημα ηλεκτρικής ενέργειας να αυξηθεί, είναι αναγκαία η εισαγωγή εποπτείας και ελέγχου στο δίκτυο διανομής, ώστε να αντιμετωπιστούν τα προβλήματα ευστάθειας που προαναφέρθηκαν. Τα δίκτυα διανομής χαρακτηρίζονται σήμερα από παθητικό τρόπο λειτουργίας, χαμηλό επίπεδο αυτοματισμού και μειωμένες δυνατότητες κεντρικής διαχείρισης. Επίσης, ιδιαίτερη σημασία έχει η αύξηση της ευελιξίας στην παραγωγή, αλλά και η εισαγωγή ευελιξίας στη ζήτηση ούτως ώστε να καταστεί δυνατός ο δυναμικός έλεγχος του ισοζυγίου παραγωγής και ζήτησης, παράγοντας ιδιαίτερης σημασία για την ευστάθεια και την αξιοπιστία του δικτύου. Η αύξηση της ευελιξίας της παραγωγής θα μπορούσε να πραγματοποιηθεί με την εισαγωγή συστημάτων αποθήκευσης ενέργειας και μικρών μονάδων παραγωγής οι οποίες έχουν τη δυνατότητα να παρέχουν σταθερή ισχύ, όπως οι γεννήτριες βιοκαυσίμων, ώστε να υπάρχει η δυνατότητα άμεσης ανταπόκρισης στη ζήτηση. Οι συγκεκριμένες γεννήτριες θα πρέπει να λειτουργούν στο 70% - 80% της ονομαστικής τους ισχύος, ώστε να βρίσκονται σε θέση να παρέχουν άμεσα την επιπρόσθετη ισχύ.

Οι παραπάνω λύσεις όμως είναι, από οικονομικής απόψεως, ασύμφορες. Η σύγχρονη προσέγγιση των διαχειριστών των συστημάτων ηλεκτρικής ενέργειας περιλαμβάνει την αξιοποίηση μεθόδων μέσω των οποίων, εκτός από την παραγωγή, ελέγχεται και η κατανάλωση ηλεκτρικής ενέργειας. Οι σύγχρονες τάσεις εξηλεκτρισμού ακόμα

περισσότερων τομέων κατανάλωσης ενέργειας, όπως οι τομείς της θέρμανσης και των μεταφορών, οδηγούν σε αύξηση της ευελιξίας των φορτίων και ταυτόχρονα καθιστούν δυνατό τον έλεγχο ενός μεγαλύτερου ποσοστού της ζήτησης. Με τη μετατροπή τμήματος της ζήτησης από ανελαστική σε ευέλικτη, καθίσταται εφικτή η διατήρηση της ευστάθειας του συστήματος ηλεκτρικής ενέργειας, χωρίς να απαιτείται η ενίσχυση του εξοπλισμού των δικτύων και χωρίς η εφεδρεία μέσω γεννητριών αποτελεί αναγκαιότητα.

## Κεφάλαιο 3 **ΕΥΦΥΗ ΔΙΚΤΥΑ ΗΛΕΚΤΡΙΚΗΣ ΕΝΕΡΓΕΙΑΣ**

### 3.1 Ορισμός

Για τον όρο «Ευφύες Δίκτυο», ο οποίος χρησιμοποιείται σήμερα περισσότερο ως όρος marketing, παρά ως τεχνικός όρος, δεν υπάρχει συγκεκριμένος, αυστηρά λεπτομερής και καθολικά αποδεκτός ορισμός

- Σύμφωνα με τον ορισμό της EU'sTP (European Union's Technology Platform):

*«είναι το ηλεκτρικό δίκτυο, το οποίο ευφύως ενσωματώνει, κάνοντας χρήση σύγχρονων τεχνολογιών και συστημάτων μέτρησης, ελέγχου και παρακολούθησης, όλες τις ενέργειες των διασυνδεδεμένων χρηστών –εκείνων που παράγουν, καταναλώνουν ή και κάνουν και τα δύο –με σκοπό η ΗΕ να παράγεται, να μεταφέρεται και να διανέμεται με αξιοπιστία, οικονομία και ασφάλεια.»*

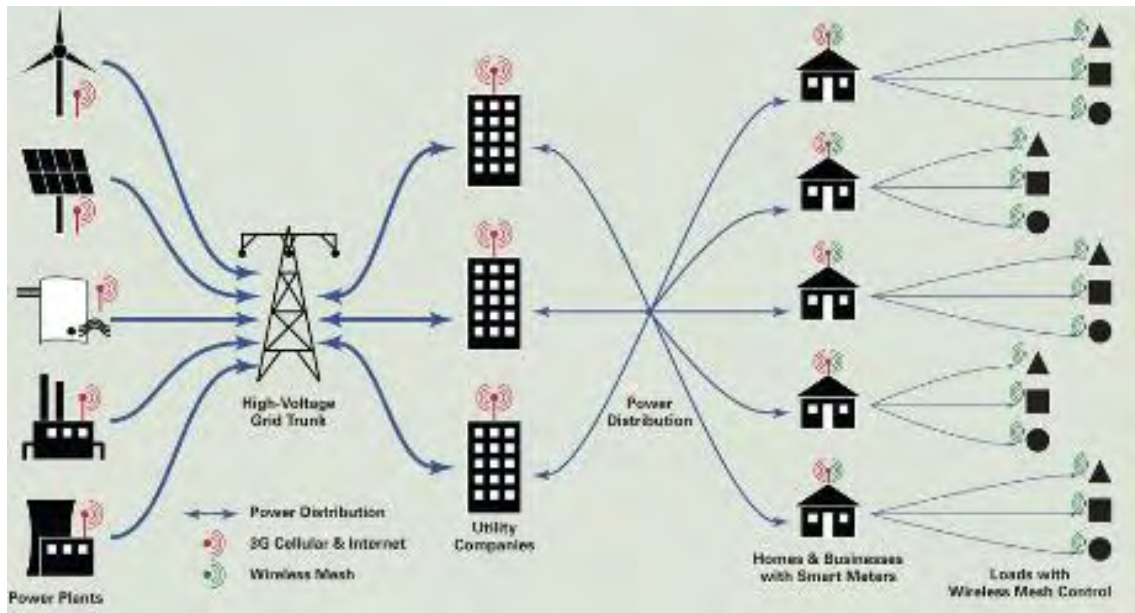
- Σύμφωνα με το αμερικανικό NIST (National Institute of Standards and Technology):

*«ο όρος αναφέρεται στον εκσυγχρονισμό του παραδοσιακού ηλεκτρικού δικτύου ενσωματώνοντας την παρακολούθηση, την προστασία και την αυτόματη βελτιστοποίηση των λειτουργιών των διασυνδεδεμένων σε αυτό στοιχείων. Χαρακτηρίζεται από μια αμφίδρομη ροή ενέργειας και πληροφορίας, εισάγοντας στο δίκτυο υπολογιστικό και τηλεπικοινωνιακό εξοπλισμό που παρέχει πληροφορία σε πραγματικό χρόνο, καθιστώντας δυνατή τη σχεδόν άμεση εξισορρόπηση προσφοράς-ζήτησης σε ηλεκτρική ενέργεια.»*

- Τέλος, σύμφωνα με την Διεθνή Ηλεκτροτεχνική Επιτροπή (IEC):

*“το ευφύες δίκτυο εμπεριέχει οτιδήποτε σχετικό με το ηλεκτρικό σύστημα από το σημείο της παραγωγής έως το σημείο της κατανάλωσης, με τη διαφορά πως ενσωματώνει τεχνολογίες που καθιστούν το δίκτυο πιο ευέλικτο, διαδραστικό και ικανό να παρέχει ανταπόκριση σε πραγματικό χρόνο.»*

Χάριν συντομίας, το ΕΗΔ μπορεί να χαρακτηριστεί ως ένα δίκτυο που ενσωματώνει τα παραδοσιακά χαρακτηριστικά του ΗΔ, παράλληλα με ένα πολυεπίπεδο στρώμα πληροφοριών και επικοινωνιών και στοχεύει στην αυτοματοποίηση των λειτουργιών του δικτύου



Εικόνα 1 - Δομή του Έξυπνου Ηλεκτρικού Δικτύου

Ουσιαστικά, το έξυπνο δίκτυο αποτελεί την εξέλιξη των παραδοσιακών ηλεκτρικών δικτύων σε δίκτυα όπου επικρατούν οι τεχνολογίες πληροφοριών. Μέσω των συγκεκριμένων τεχνολογιών επιτρέπεται ο απομακρυσμένος έλεγχος όλων των σταδίων από την παραγωγή της ενέργειας μέχρι και την κατανάλωση της, η αμφίδρομη επικοινωνία μεταξύ παραγωγής και κατανάλωσης, με δυνατότητα συμμετοχής του καταναλωτή στην παραγωγή, η διασφάλιση της βιωσιμότητας και ποιότητας των υπηρεσιών, η κατανεμημένη παραγωγή ηλεκτρικής ενέργειας, η τοπική επεξεργασία των πληροφοριών, η αποθήκευση της παραγόμενης ενέργειας, και η έξυπνη μέτρηση των καταναλώσεων. Βασικό στόχο των σχετικών υλοποιήσεων αποτελούν η διασφάλιση της αξιοπιστίας, της αποδοτικότητας και της ασφάλειας του δικτύου και ειδικότερα:

- ❖ **Η αξιοπιστία** του δικτύου επιτυγχάνεται μέσω ειδικού σχεδιασμού ο οποίος επιτρέπει την ανίχνευση των αιτιών των βλαβών και την επίλυση τους, ενώ παράλληλα μπορεί να εντοπίσει εναλλακτικές πηγές τροφοδοσίας στην περίπτωση κατά την οποία οι υπάρχουσες αδυνατούν να ικανοποιήσουν τη ζήτηση, μειώνοντας με τον τρόπο αυτό τις πιθανότητες των blackouts.

- ❖ **Η αποδοτικότητα** επιτυγχάνεται μέσω της αξιοποίησης εναλλακτικών μορφών ενεργειακών για τη βελτιστοποίηση της διαδικασίας παραγωγής – μεταφοράς και μετάδοσης ενέργειας, καθώς και μέσω της συμμετοχής του καταναλωτή στη διαδικασία εξοικονόμησης

ενέργειας, γεγονός το οποίο επιτυγχάνεται με τις ευέλικτες υλοποιήσεις ζήτησης και απόκρισης.

❖ **Η ασφάλεια** επιτυγχάνεται μέσω της υλοποίησης ενός αξιόπιστου συστήματος ελέγχου και παρακολούθησης της διαδικασίας παροχής ηλεκτρικής ενέργειας [10].

### 3.2 Χαρακτηριστικά του ΕΗΔ

Το ΕΗΔ παρουσιάζει τα παρακάτω σημαντικά χαρακτηριστικά:

**1.** Καθιστά δυνατή την ενημέρωση και τη συμμετοχή του καταναλωτή.

Η διαρκής και αμφίδρομη ροή πληροφοριών είναι σχεδόν σε πραγματικό χρόνο αξιοποιήσιμη τόσο από τους διαχειριστές του δικτύου όσο και από τους καταναλωτές. Έτσι ο καταναλωτής είναι σε θέση να έχει άμεση εποπτεία των καταναλώσεών του και δυνατότητα επέμβασης στη χρήση που κάνει σε ηλεκτρική ενέργεια και κατά συνέπεια μπορεί να προσαρμόζει τη χρήση αυτή ανάλογα με τις ανάγκες και το κόστος.

**2.** Βελτιστοποιεί την αξιοποίηση του εξοπλισμού και τη λειτουργική αποδοτικότητα του Σ.Η.Ε.

Το ΕΗΔ αναβαθμίζει το υπάρχον δίκτυο με εξοπλισμό που εξασφαλίζει την αποτελεσματική εξυπηρέτηση των χρηστών, τη συνεχή και ασφαλή παροχή ενέργειας, τη μείωση απωλειών και την αποφυγή συμφορήσεων σε ζήτηση. Η κατάσταση του εξοπλισμού είναι ανά πάσα στιγμή ορατή μέσω ψηφιακών ενδείξεων και κατά συνέπεια η συντήρησή του καθίσταται πολύ πιο εύκολη. Η ανάγκη επέκτασης των υποδομών του δικτύου παύει να υφίσταται καθώς οι ήδη υπάρχουσες υποδομές αξιοποιούνται βέλτιστα.

**3.** Επιτρέπει τη μεγαλύτερη διείσδυση ΑΠΕ και αυτόνομων παραγωγών στο δίκτυο.

Τα εξελιγμένα υπολογιστικά συστήματα του ΕΗΔ είναι σε θέση να εξασφαλίσουν την αξιοπιστία και την ευστάθεια του δικτύου και δημιουργούν ένα πιο ευέλικτο και προσαρμοστικό δίκτυο που γνωρίζει τις ανάγκες ζήτησης ανά πάσα στιγμή. Σε συνθήκες αιχμής φορτίου, η δυνατότητα αποθήκευσης ηλεκτρικής ενέργειας στο ΕΗΔ και πλήρους ελέγχου της ζήτησης, μειώνουν την ανάγκη αύξησης της ηλεκτροπαραγωγής από τις μεγάλες μονάδες και επιτρέπουν την εξυπηρέτηση των αναγκών σε ενέργεια από αλλού. Με τον τρόπο αυτό, ενθαρρύνεται η διείσδυση των ΑΠΕ καθώς και των ιδιωτών παραγωγών.

**4.** Διαθέτει ικανότητα αυτοϊασης και ανθεκτικότητα απέναντι σε βλάβες.

Το ΕΗΔ είναι σε θέση να αντιλαμβάνεται, να εντοπίζει και να αντιμετωπίζει τα σφάλματα τη στιγμή που αυτά εμφανίζονται, κάνοντας χρήση αισθητήρων και ευφών συσκευών. Το προβληματικό στοιχείο απομονώνεται από το υπόλοιπο σύστημα και η κανονική λειτουργία του δικτύου αποκαθίσταται, χωρίς να είναι καν απαραίτητη η ανάγκη ανθρώπινης παρέμβασης. Με τον τρόπο αυτό, μειώνονται ή και εξαφανίζονται φαινόμενα διακοπής της παροχής υπηρεσιών στους καταναλωτές συμβάλλοντας σε ένα πιο αξιόπιστο δίκτυο. Το πολλά υποσχόμενο ΕΗΔ, θέτει πολύ υψηλές προσδοκίες και καθιστά αναπόφευκτη τη σύγκριση με το υπάρχον παραδοσιακό ηλεκτρικό δίκτυο.

Στον πίνακα 2 παρατίθενται συγκεντρωτικά τα χαρακτηριστικά του ΕΗΔ σε αντιπαραβολή με εκείνα του συμβατικού δικτύου [12]

<i><b>Συμβατικό δίκτυο ΗΕ</b></i>	<i><b>Ευφύες δίκτυο ΗΕ</b></i>
Ηλεκτρομηχανολογικός εξοπλισμός	Ψηφιακός εξοπλισμός
Μονόδρομη ροή πληροφορίας και ισχύος	Αμφίδρομη ροή πληροφορίας και ισχύος
Ενιαία και προσεγγιστική τιμολόγηση	Ευέλικτη και ακριβής τιμολόγηση
Μηχανικοί μετρητές	Έξυπνοι μετρητές
Κεντριοποιημένη παραγωγή	Διεσπαρμένη παραγωγή
Χειροκίνητη αποκατάσταση βλαβών	Αυτόματη αποκατάσταση βλαβών
Ανενημέρωτος/αδρανής καταναλωτής	Ενημερωμένος/ενεργός καταναλωτής
Φαινόμενα συμφόρησης/υπερφόρτωσης	Αξιοπιστία-Ασφάλεια-Βιωσιμότητα
Περιορισμένη διείσδυση ΑΠΕ	Ενσωμάτωση ΑΠΕ και αυτοπαραγωγών
Διαφορετική λειτουργία δικτύων ανά κράτος	Διεθνώς διασυνδεδεμένα δίκτυα
	Δυνατότητα αποθήκευσης ενέργειας

Πίνακας 2 - Χαρακτηριστικά συμβατικού δικτύου Η.Ε. – ευφύους δικτύου Η.Ε.

### 3.3 Οφέλη και πιθανοί κίνδυνοι των ευφών δικτύων ηλεκτρικής ενέργειας

Την τελευταία δεκαετία περίπου, η συζήτηση για τη μετάβαση από το συμβατικό δίκτυο ΗΕ στο ευφύες δίκτυο γίνεται διαρκώς εντονότερη και η αναβάθμιση του «γερασμένου» και ανεπαρκούς σύγχρονου δικτύου αποτελεί αδήριτη αναγκαιότητα. Εξετάζοντας το θέμα υπό ευρύτερη έννοια και λαμβάνοντας υπόψη σχετικές έρευνες, προκύπτει, με σχετική ασφάλεια, το συμπέρασμα, ότι τα οφέλη αυτής της μετάβασης υπερτερούν σημαντικά έναντι του κόστους του εγχειρήματος. Τα οφέλη αυτά, ανάλογα με τον τομέα που αφορούν, μπορούν να συνοψισθούν σε έξι βασικές κατηγορίες [9] :

- **Αξιοπιστία (Reliability)**
  - Η χρήση έξυπνων μετρητών καθιστά δυνατό τον άμεσο εντοπισμό θεμάτων ποιότητας ισχύος ή απώλειας ισχύος, διευκολύνοντας τους διαχειριστές του συστήματος στην άμεση διάγνωση και αντιμετώπιση.

- Η δυνατότητα ελέγχου της ζήτησης διευκολύνει τη διαχείριση του συστήματος τις ώρες αιχμής, μειώνοντας την πιθανότητα γενικής διακοπής (blackout) ή άλλων διαταραχών.
- Το δίκτυο διανομής αποκτά τη δική του ευφυΐα, καθώς εξελιγμένοι αισθητήρες και συσκευές διευκολύνουν τον εντοπισμό ενός προβληματικού στοιχείου, είτε ενημερώνοντας και προλαμβάνοντας τη βλάβη, είτε απομονώνοντάς το από το υπόλοιπο σύστημα, μετά από την εμφάνιση της βλάβης.
- **Οικονομικά (Economics)**
  - Οι προηγμένες υποδομές μέτρησης με έξυπνους μετρητές και δυνατότητα αμφίδρομης επικοινωνίας, παρέχουν στους καταναλωτές μετρητικές ενδείξεις τιμών σε σχεδόν πραγματικό χρόνο. Το γεγονός αυτό δημιουργεί κίνητρο στον καταναλωτή να ενδιαφερθεί για τις καταναλώσεις του, αντιμετωπίζοντάς τις σαν ένα ακόμα προϊόν που αγοράζει και ως εκ τούτου αναμένεται να οδηγήσει σε μείωση της ζήτησης και κατ'επέκταση των τιμών της ηλεκτρικής ενέργειας.
  - Περιορισμός των ρευματοκλοπών.
  - Η αναμενόμενη μείωση βλαβών θα οδηγήσει και σε μείωση του κόστους για την αποκατάστασή τους, συνεπώς σε εξοικονόμηση κεφαλαίων που μπορούν να διατεθούν σε άλλους σκοπούς.
  - Η διεύθυνση διεσπαρμένης παραγωγής στο δίκτυο, σημαίνει και τη δυνατότητα τοπικής παραγωγής και αποθήκευσης ενέργειας που εκτός από τη στήριξη του δικτύου στον έλεγχο της ζήτησης, συνεπάγεται και το «άνοιγμα» των ενδιαφερόμενων παραγωγών στην αγορά ηλεκτρισμού.
- **Αποδοτικότητα (Efficiency)**
  - Οι καταναλωτές είναι πλέον σε θέση να κατανοήσουν και να αλληλεπιδράσουν με τις λειτουργίες του δικτύου, πραγματοποιώντας μια πιο αποδοτική διαχείριση της καταναλισκόμενης ενέργειας στις κατοικίες και τις επιχειρήσεις τους, με προφανή οφέλη και για το ίδιο το δίκτυο.
  - Ο εξαιρετικά προηγμένος εξοπλισμός δίνει στους διαχειριστές του δικτύου τη δυνατότητα να βελτιστοποιήσουν τη λειτουργία του, αξιοποιώντας στο μέγιστο τα υφιστάμενα διασυνδεδεμένα στοιχεία (Improved Asset Utilization)
  - Η δυνατότητα ελέγχου της ροής ισχύος μέσω προηγμένων συστημάτων, μειώνει τόσο τις απώλειες ισχύος όσο και τη ζήτηση αιχμής.

ο Το έξυπνο δίκτυο καθιστά τις ανάγκες σε εξοπλισμό εγκαίρως προβλέψιμες, διευκολύνοντας τη συντήρηση και την αντικατάσταση του, συνεισφέροντας παράλληλα στην επέκταση της διάρκειας ζωής του

- **Περιβάλλον (Environment)**

ο Η αυξημένη διείσδυση ΑΠΕ στο δίκτυο έχει προφανή περιβαλλοντικά οφέλη δεδομένου ότι μορφές ενέργειας όπως η ηλιακή ή η αιολική δεν επιβαρύνουν το περιβάλλον. Το γεγονός αυτό συνεπάγεται αισθητή μείωση των εκπεμπόμενων αέριων ρύπων καθώς και των ζημιωγόνων επεμβάσεων στα φυσικά οικοσυστήματα.

ο Πολλές απρόσμενες βλάβες του δικτύου οι οποίες μέχρι πρότινος οδηγούσαν σε εξαιρετικά ρυπογόνες καταστροφές μετασχηματιστών και εμφανίσεις πετρελαιοκηλίδων, θα μπορούν πλέον να προβλεφθούν και να αποφευχθούν.

ο Η αλληλεπίδραση των καταναλωτών με έννοιες όπως η «πράσινη ενέργεια», πιθανότατα θα προσφέρει κίνητρα ευαισθητοποίησης και ενασχόλησης με την προστασία του περιβάλλοντος.



Εικόνα 2 - Τα αλυσιδωτά οφέλη του Ευφυούς Ηλεκτρικού Δικτύου

- **Προστασία και ασφάλεια (Security and Safety)**

ο Οι διαχειριστές του δικτύου έχουν πρόσβαση στην κατάσταση κάθε καταναλωτή ξεχωριστά και είναι σε θέση σε πραγματικό χρόνο να επέμβουν συνδέοντας ή αποσυνδέοντας φορτία, εάν αυτό κριθεί απαραίτητο, προκειμένου να διασφαλιστεί η ασφάλεια του καταναλωτή.



- Πολλές φορές, μια διακοπή ρεύματος ή μια βλάβη, μπορεί να θέσει σε κίνδυνο την υγεία ή την ασφάλεια ενός ατόμου. Το ΕΗΔ εξασφαλίζει την αποφυγή τέτοιων καταστάσεων.
- Οι διαχειριστές του δικτύου, αξιοποιώντας τους σύγχρονους αισθητήρες, τα αμφίδρομα τηλεπικοινωνιακά συστήματα και τις προηγμένες συσκευές ελέγχου, διαθέτουν τα απαραίτητα εργαλεία για να ανιχνεύσουν, να εντοπίσουν, να επέμβουν και να διορθώσουν οποιοδήποτε σφάλμα προκύψει. Με τον τρόπο αυτό μειώνεται η πιθανότητα να προκύψουν επείγουσες καταστάσεις που θα έχουν αλυσιδωτές αντιδράσεις σε ένα εκτενές μέρος του δικτύου με εξαιρετικά επιζήμιες συνέπειες.
- Το γεγονός ότι οι Μονάδες Διεσπαρμένης Παραγωγής ενός δικτύου λειτουργούν σε μεγάλο βαθμό αυτόνομα, αποτελεί δικλείδα ασφαλείας, καθώς είναι σπάνιο, έως αδύνατο, να τεθούν όλες ταυτόχρονα εκτός λειτουργίας.

Συμπερασματικά, από όλα τα παραπάνω, προκύπτει ότι η αναβάθμιση του συμβατικού δικτύου στο ευφυές δίκτυο θα έχει ευεργετικά αποτελέσματα για όλες τις ενδιαφερόμενες ομάδες.

Οι πάροχοι ηλεκτρικής ενέργειας έχουν την ευκαιρία να εξοικονομήσουν τεράστια ποσά σε λειτουργικά κόστη και κόστη συντήρησης, ενώ η προοπτική ανοίγματος της παραγωγής ΑΠΕ σε νέες αγορές διαφαίνεται πιο έντονη από ποτέ. Από την άλλη, ο κίνδυνος που καλούνται να αναλάβουν σε ένα τέτοιο εγχείρημα είναι η γρήγορη και αποτελεσματική απόσβεση των επενδεδυμένων κεφαλαίων.

Οι καταναλωτές είναι και αυτοί κερδισμένοι καθώς τους δίνεται η ευκαιρία να μειώσουν τους λογαριασμούς τους, έχοντας τον έλεγχο και την εποπτεία των καταναλώσεών τους. Επιπλέον, καθίσταται ευκολότερο για εκείνους να πάψουν να αποτελούν μόνο καταναλωτές του δικτύου και να γίνουν και παραγωγοί, συνεισφέροντας αμφίδρομα στη λειτουργία του. Ο ενδεχόμενος κίνδυνος για τον καταναλωτή αφορά στην πιθανότητα εφαρμογής του ευφυούς δικτύου χωρίς τα οφέλη του να γίνουν άμεσα αντιληπτά από εκείνον.

Η κοινωνία γενικότερα, θα ωφεληθεί από μια αναζωογονημένη οικονομία, που ενισχύεται με θέσεις εργασίας καθώς και μια πτωτική τάση στην τιμή της ηλεκτρικής ενέργειας, ενώ σημαντικό όφελος για το κοινωνικό σύνολο αποτελούν οι βελτιωμένες περιβαλλοντικές συνθήκες. Πρακτικά, το εγχείρημα αυτό δεν ενέχει κάποιο ουσιαστικό κίνδυνο.

Παράλληλα όμως με τα πολλαπλά οφέλη που αναμένονται από τη χρήση της τεχνολογίας, των ΕΗΔ, προκύπτουν, όπως είναι φυσικό και ορισμένα ζητήματα τα οποία χρήζουν προσοχής και διερεύνησης.

Τα ζητήματα αυτά αφορούν σε κινδύνους που σχετίζονται με [11]:

- Τη διασφάλιση της σωστής εξυπηρέτησης των καταναλωτών.

Τα συστήματα αυτά δίνουν στους προμηθευτές σαφώς περισσότερες δυνατότητες παρέμβασης έναντι των καταναλωτών και ως εκ τούτου τη δυνατότητα αυθαιρεσίας σε ζητήματα χρεώσεων, διακοπών παροχής. Απαραίτητη κρίνεται λοιπόν η ύπαρξη κατάλληλης νομοθεσίας και όρων στη σύμβαση μεταξύ καταναλωτή και παρόχου, ώστε να προστατεύεται ο καταναλωτής.

- Την προστασία των προσωπικών δεδομένων των καταναλωτών.

Η ύπαρξη ιδιωτικών πληροφοριών σε ψηφιακά δεδομένα, εκτός της ευκολίας που παρέχει στον εντοπισμό και τη μεταφορά, ενέχει και τον κίνδυνο της υποκλοπής ή της παράνομης δημοσίευσης. Για τον περιορισμό του εν λόγω κινδύνου, μέτρα μπορεί να λάβει τόσο ο πάροχος (εξουσιοδότηση πρόσβασης σε προσωπικά δεδομένα σε συγκεκριμένα άτομα υπό συγκεκριμένες συνθήκες, χρήση κρυπτογραφίας στα υπολογιστικά συστήματα) όσο και η πολιτεία (νομοθεσία).

- Την προστασία και αντιμετώπιση ηλεκτρονικών επιθέσεων.

Όπως όλα τα συστήματα που χρησιμοποιούν υπολογιστές, τηλεπικοινωνίες και το διαδίκτυο, έτσι και το έξυπνο μετρητικό σύστημα είναι ευάλωτο σε κακόβουλες απομακρυσμένες ηλεκτρονικές επιθέσεις. Σε περίπτωση μιας τέτοιας επίθεσης, βρίσκονται εκτεθειμένοι τόσο οι καταναλωτές, όσο και οι πάροχοι και το ίδιο το δίκτυο και ο δράστης της επίθεσης αποκτά πρόσβαση σε απόρρητες πληροφορίες και μερικό ή ολικό έλεγχο του συστήματος. Ως εκ τούτου, επιτακτική είναι η διασφάλιση της αποτελεσματικής αντιμετώπισης τέτοιων επιθέσεων με αμυντικά ηλεκτρονικά συστήματα και τείχη προστασίας, κρυπτογραφία, δυνατότητα χειροκίνητης απομακρυσμένης διακοπής λειτουργίας του δικτύου που έχει πληγεί κ.α.

- Τη διασφάλιση της υγείας των καταναλωτών.

Οι έξυπνοι μετρητές ως γνωστόν είναι ηλεκτρονικές συσκευές που επικοινωνούν ασύρματα κάνοντας χρήση τηλεπικοινωνιών και ως εκ τούτου εκπέμπουν σήματα και ακτινοβολίες. Οι εκπομπές αυτές είναι στα ελάχιστα δυνατά επίπεδα παρόλαυτα ένα κομμάτι της παγκόσμιας

κοινότητας είναι σκεπτική απέναντι στη χρήση ασύρματης τεχνολογίας και επισημαίνει κινδύνους για άτομα «ηλεκτρομαγνητικά υπερευαίσθητα» των οποίων η έκθεση σε ασύρματες τεχνολογίες προκαλεί ζαλάδα, υπερκόπωση, προβλήματα ύπνου κ.α. Προφανώς, προτού γίνει καθολική εφαρμογή των έξυπνων μετρητών, προαπαιτείται η διασφάλιση πως τέτοιου είδους κίνδυνοι δεν είναι υπαρκτοί.

### **3.4 Η δομή των ευφών δικτύων ηλεκτρικής ενέργειας**

Ένα ευφές δίκτυο ηλεκτρικής ενέργειας μπορεί να διαχωριστεί σε επτά τομείς οι οποίοι αλληλεπιδρούν μεταξύ τους και καθένας εκ των οποίων περιλαμβάνει μία ή περισσότερες συσκευές, συστήματα ή προγράμματα (για παράδειγμα, έξυπνους μετρητές, συστήματα SCADA κλπ.) τα οποία διεκπεραιώνουν διαδικασίες ανταλλαγής πληροφοριών και λήψης αποφάσεων ούτως ώστε το δίκτυο να λειτουργεί εύρυθμα. Οι συγκεκριμένοι τομείς είναι οι πελάτες, οι αγορές, οι πάροχοι υπηρεσιών, οι λειτουργίες, η παραγωγή, η μεταφορά και η διανομή. Η όλη δομή παρουσιάζεται στην εικόνα 3, ενώ τα συστατικά της στοιχεία αναλύονται παρακάτω [13].

**Πελάτες:** Αποτελούνται από τους καταναλωτές και τις συσκευές παραγωγής, αποθήκευσης και διαχείρισης ενέργειας που βρίσκονται υπό την κατοχή τους. Υπάρχουν τρεις κύριες κατηγορίες πελατών: οι πελάτες του δικτύου, οι οποίοι επιθυμούν να πραγματοποιήσουν οικιακή χρήση της ηλεκτρικής ενέργειας, εκείνοι που επιθυμούν να πραγματοποιήσουν εμπορική χρήση, καθώς και εκείνοι που πραγματοποιούν βιομηχανική χρήση.

**Αγορές:** Αποτελούνται από τους λειτουργούς και τους συμμετέχοντες στην ενεργειακή αγορά.

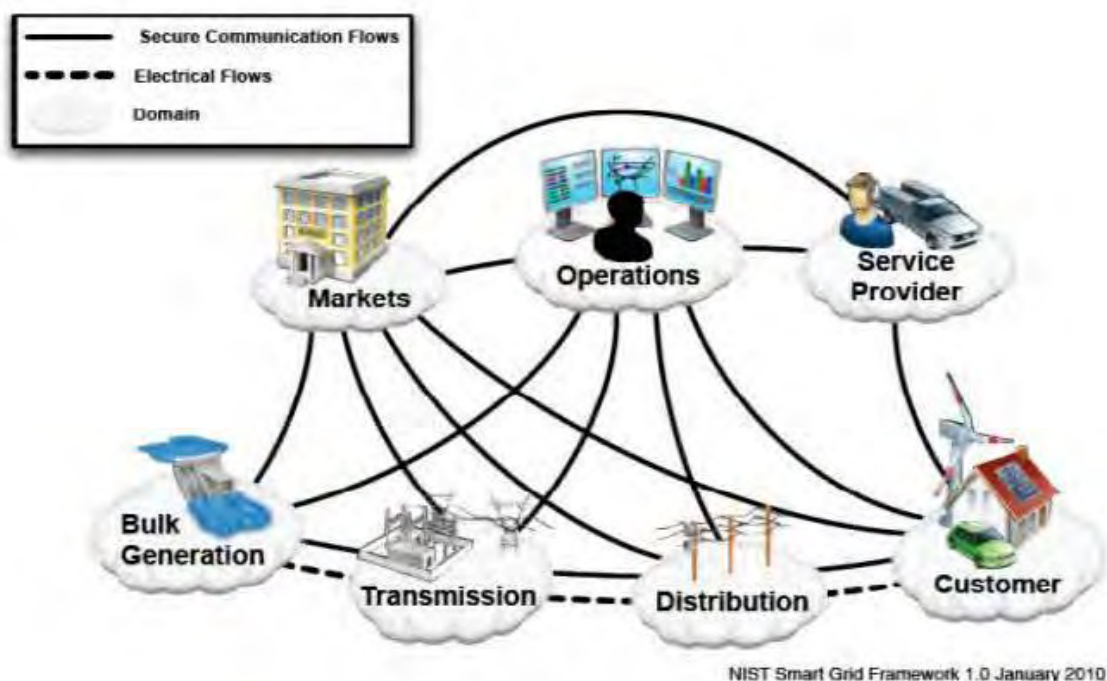
**Πάροχοι υπηρεσιών:** Αποτελούνται από οργανισμούς οι οποίοι έχουν ως αντικείμενο τους την παροχή υπηρεσιών ηλεκτρικής ενέργειας τόσο σε πελάτες όσο και σε παρόχους άλλων υπηρεσιών.

**Λειτουργίες:** Αφορά τους διαχειριστές της διακίνησης της ηλεκτρικής ενέργειας μεταξύ των δικτύων.

**Παραγωγή:** Αποτελείται από τις γεννήτριες και τους σταθμούς παραγωγής ηλεκτρικής ενέργειας οι οποίοι παράγουν μεγάλες ποσότητες ενέργειας καθώς και τις μονάδες αποθήκευσης της ενέργειας.

**Μεταφορά:** Αποτελείται από τις υποδομές μεταφοράς της ενέργειας σε μακρινές αποστάσεις. Ενδέχεται να περιλαμβάνει και τοπικά μέσα παραγωγής ή αποθήκευσης ενέργειας.

**Διανομή:** Αποτελείται από τις υποδομές της διανομής ηλεκτρικής ενέργειας από και προς τους πελάτες, ενώ ενδέχεται εντός του πεδίου της να εμπίπτει και η αποθήκευση και παραγωγή ενέργειας.



Εικόνα 3 - Η αρχιτεκτονική ενός ευφυούς δικτύου ηλεκτρικής ενέργειας

Τα συστατικά στοιχεία κάθε τομέα, αποσκοπούν στην εκπλήρωση ορισμένων κοινών στόχων και συνεργάζονται τόσο μεταξύ τους όσο και με συστατικά στοιχεία άλλων τομέων ούτως ώστε να τους επιτύχουν. Η δομή που παρουσιάστηκε παραπάνω δεν έχει απόλυτη μορφή, καθώς συχνά ένα συστατικό στοιχείο ενός τομέα ενδέχεται να επιτελεί λειτουργίες οι οποίες βασίζονται σε πληροφορίες που προέρχονται από άλλους τομείς (για παράδειγμα, το δίκτυο διανομής σχετίζεται τόσο με τον τομέα διανομής όσο και με τον τομέα λειτουργιών, στον οποίο ανήκουν τα συστήματα διαχείρισης δικτύων διανομής).

### 3.5 Μετάδοση δεδομένων στο ευφυές δίκτυο

Τα τηλεπικοινωνιακά δίκτυα ενός ΕΗΔ μπορούν να διαχωριστούν σε τρία κύρια επίπεδα, αναλόγως με τη γεωγραφική τους κάλυψη και των αναγκών που εξυπηρετούν. Τα δίκτυα αυτά παρουσιάζονται παρακάτω [14]:

#### □ Home Area Network (HAN):

Τα συγκεκριμένα δίκτυα βρίσκονται στο χαμηλότερο επίπεδο επικοινωνίας, ενώ αφορούν τη σύνδεση ενεργειακών συσκευών εντός της οικίας. Καθιστούν δυνατή την παρακολούθηση και τον έλεγχο των οικιακών ενεργειακών καταναλώσεων, καθώς και την εφαρμογή σύγχρονων λειτουργιών, όπως είναι η απόκριση στη ζήτηση, καθώς και την υλοποίηση σύγχρονων υποδομών, όπως είναι η Advanced Metering Infrastructure (AMI). Οι αντίστοιχες εφαρμογές για επιχειρήσεις είναι τα Business Area Networks (BANs) και για τη βιομηχανία τα Industrial Area Networks (IANs). Η επικοινωνία στα συγκεκριμένα δίκτυα διεξάγεται σε χαμηλό εύρος ζώνης, αλλά πρέπει να είναι τακτική και αδιάλειπτη μεταξύ των συσκευών και του συστήματος ηλεκτρικής ενέργειας. Διαχωρίζονται σε ασύρματα, ενσύρματα χωρίς πρόσθετη καλωδίωση και ενσύρματα με πρόσθετη καλωδίωση, ενώ είναι δυνατοί και συνδυασμοί των παραπάνω. Τα δύο επικρατέστερα πρωτόκολλα ασύρματης επικοινωνίας είναι το WiFi και το ZigBee. Όσον αφορά την ενσύρματη επικοινωνία με νέα καλωδίωση, το πλέον διαδεδομένο πρωτόκολλο είναι το Ethernet, ενώ στην ενσύρματη επικοινωνία η οποία χρησιμοποιεί την υπάρχουσα καλωδίωση του ηλεκτρικού δικτύου συνήθως χρησιμοποιείται το πρωτόκολλο PLC [14].

,

#### □ Neighborhood Area Network (NAN):

Το συγκεκριμένο δίκτυο επιτελεί τον ρόλο της διεκπεραίωσης της ενδιάμεσης επικοινωνίας μεταξύ των έξυπνων μετρητών κατανάλωσης ενέργειας και των μονάδων διεσπαρμένης παραγωγής ηλεκτρικής ενέργειας, από και προς τους προμηθευτές. Επίσης διεκπεραιώνει τις λειτουργίες επικοινωνίας μεταξύ των εφαρμογών εποπτείας και ελέγχου του δικτύου διανομής. Τα συγκεκριμένα δίκτυα ξεκινούν από τις θύρες εξόδου των HANs και καταλήγουν στους συλλέκτες και τους απομαστευτές δεδομένων οι οποίοι βρίσκονται στους υποσταθμούς μέσης – χαμηλής τάσης του δικτύου διανομής. Για την επικοινωνία στο συγκεκριμένο επίπεδο αξιοποιούνται διάφορα πρωτόκολλα, είτε ασύρματα είτε ενσύρματα. Στην ασύρματη επικοινωνία τα πλέον διαδεδομένα είναι τα πρωτόκολλα IEEE 802 (για παράδειγμα, WiMAX), καθώς και οι εφαρμογές της κινητής τηλεφωνίας (για παράδειγμα,

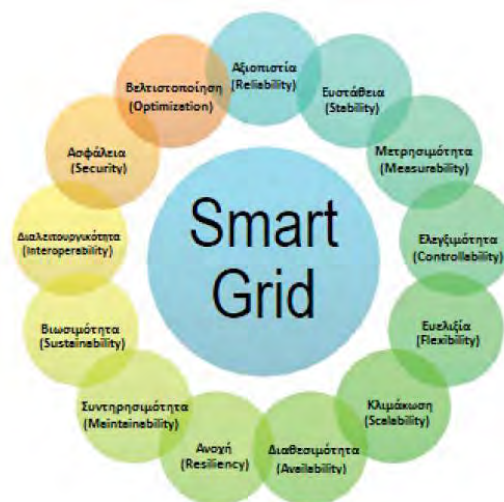
GPRS, EDGE, UMTS, LTE), ενώ στην ενσύρματη επικοινωνία χρησιμοποιούνται κυρίως ευρυζωνικές τεχνολογίες μέσω τηλεφωνικών γραμμών (DSL), καθώς και τεχνολογίες PLC [15].

□ Wide Area Network (WAN):

Τα συγκεκριμένα δίκτυα αφορούν την επικοινωνία στο επίπεδο του κορμού. Καλύπτουν μία ευρεία γεωγραφική περιοχή, ενώ φιλοξενούν τερματικούς σταθμούς και τοπικά δίκτυα. Αποτελούν το υπόβαθρο της τηλεπικοινωνιακής υποδομής ενός ευφούς δικτύου ηλεκτρικής ενέργειας. Διακινούν τις πληροφορίες στο σύστημα μεταφοράς και το δίκτυο διανομής, συνδέοντας τους υποσταθμούς, τις μονάδες παραγωγής ηλεκτρικής ενέργειας, τις διατάξεις αποθήκευσης ενέργειας, καθώς και τον αυτοματοποιημένο ηλεκτρομηχανολογικό εξοπλισμό του δικτύου διανομής, όπως είναι οι μετασχηματιστές, οι αντισταθμιστές αέργου ισχύος και οι διακόπτες ισχύος. Λόγω των δυνατοτήτων αμφίδρομης επικοινωνίας που έχουν, είναι δυνατή, η εφαρμογή αυτοματισμού της διανομής (Distribution Automation) καθώς και απεικόνισης της ποιότητας ισχύος (Power Quality Monitoring) από τους διαχειριστές των δικτύων [15].

### 3.6 Χαρακτηριστικά των ευφών δικτύων ηλεκτρικής ενέργειας

Στην παρούσα ενότητα θα αναλυθούν τα κυριότερα χαρακτηριστικά ενός ευφούς δικτύου ηλεκτρικής ενέργειας. Τα συγκεκριμένα χαρακτηριστικά διασυνδέονται μεταξύ τους με μία στενή σχέση αιτίας και αποτελέσματος, ενώ πρέπει να λαμβάνονται υπόψη κατά τη διαδικασία του σχεδιασμού οποιουδήποτε ευφούς δικτύου [17].



Εικόνα 4 - Χαρακτηριστικά ενός ευφούς δικτύου ηλεκτρικής ενέργειας

**Αξιοπιστία και ευστάθεια:** Υπό τον όρο της αξιοπιστίας εννοείται η δυνατότητα του δικτύου και των συνιστωσών του να διεκπεραιώνουν τις απαιτούμενες εργασίες υπό δεδομένες συνθήκες σε ένα συγκεκριμένο χρονικό διάστημα. Έχει ένα χαρακτηριστικό ανθεκτικότητας. Γενικά, αποτελεί ένα μέσο ερμηνείας της λειτουργικής υγείας και του βαθμού μεταβλητότητας του συστήματος. Επίσης, παρουσιάζει τη συνοχή, επαναληψιμότητα και φερεγγυότητα που εκτιμάται ότι θα διατηρήσει το δίκτυο. Απαιτούνται χαμηλές πιθανότητες βλαβών, οι οποίες όταν συμβαίνουν θα έχουν τη χαμηλότερη δυνατή επίπτωση στο συνολικό σύστημα, ενώ το στοιχείο που έχει πρόβλημα θα αντικαθίσταται ή θα επιδιορθώνεται το συντομότερο δυνατό. Η αξιοπιστία εξαρτάται από τα χαρακτηριστικά που περιγράφονται στη συνέχεια της συγκεκριμένης παραγράφου. Η ευστάθεια καθορίζει το επίπεδο αξιοπιστίας του δικτύου. Τα ευφυή δίκτυα πρέπει να παρέχουν εγγυήσεις σχετικά με τη σταθερότητα της τάσης και του ρεύματος, να περιορίζουν τη ζήτηση αιχμής και τη μεταβλητότητα του φορτίου, αξιοποιώντας την κατανομημένη παραγωγή ηλεκτρικής ενέργειας και την αποθήκευση της σε μεγάλες εκτάσεις και αποτρέποντας διάφορα ανεπιθύμητα συμβάντα.

**Μετρησιμότητα και ελεγχιμότητα:** Η διακοπή των υπηρεσιών ή οι βλάβες είναι συμβάντα με σοβαρές επιπτώσεις, τα οποία παρουσιάζουν υψηλές πιθανότητες να συμβούν. Πρέπει να είναι μετρήσιμες και ελεγχίμες, ούτως ώστε να είναι δυνατή η διεξαγωγή εκτιμήσεων και αξιολογήσεων. Το ευφυές δίκτυο έχει τη δυνατότητα παρακολούθησης και επιδιόρθωσης λειτουργικών ζητημάτων μέσα από δυναμικές μετρήσεις και παρακολούθηση πραγματικού χρόνου. Επίσης πρέπει να υπάρχει ορισμένη παρατηρησιμότητα και διαφάνεια ούτως ώστε να διεξάγεται μία αποτελεσματική ανάλυση και διαχείριση καθώς και πρόβλεψη και αντίδραση των μεταβαλλόμενων καταστάσεων του δικτύου. Τα δεδομένα του δικτύου πρέπει επίσης να είναι μετρήσιμα, παρατηρήσιμα και διαχειρίσιμα.

**Ευελιξία και κλιμάκωση:** Το δίκτυο κινείται από μία κεντρική δομή σε διάφορα αποκεντρωμένα μικροδίκτυα (microgrids). Η κλιμάκωση πρέπει να είναι καλά ορισμένη. Μέσα από τη νησιδοποίηση (islanding), τα μικροδίκτυα αποσκοπούν στην ενσωμάτωση της κατανομημένης παραγωγής και της αποθήκευσης ενέργειας ούτως ώστε να συνεισφέρουν ενέργεια στους παρόχους στις περιόδους υψηλής ζήτησης.

Η νησίδα εισάγει την προσέγγιση ενός γιγαντιαίου έξυπνου δικτύου το οποίο αποτελείται από διάφορα έξυπνα μικροδίκτυα. Κάθε μικροδίκτυο έχει τη δυνατότητα αυτόνομης λειτουργίας αναφορικά με τη διαχείριση ζήτησης, το μοντέλο ποιότητας και αξιοπιστίας, τη διαχείριση

προβλημάτων και τη διαχείριση ασφαλείας. Η ευελιξία δίδει τη δυνατότητα παροχής πολλών εναλλακτικών διαδρομών για τη μεταφορά της ενέργειας και των δεδομένων, ενώ καθιστά δυνατούς τον έλεγχο και τη λειτουργία όποτε είναι αναγκαίοι. Παρουσιάζει τέσσερις πτυχές: **επεκτασιμότητα**, για την εισαγωγή καινοτόμων τεχνολογιών παραγωγής κατά το μέλλον, **προσαρμοστικότητα** στις διάφορες γεωγραφικές τοποθεσίες και κλίματα, **πολλαπλές στρατηγικές ελέγχου**, ούτως ώστε να συντονίζονται τα αποκεντρωμένα συστήματα ελέγχου ανάμεσα στους υποσταθμούς και τα κέντρα ελέγχου, **συμβατότητα** με τα διάφορα πλαίσια λειτουργίας των αγορών ηλεκτρικής ενέργειας και δυνατότητες σταδιακής αναβάθμισης εν λειτουργία του υλικού και του λογισμικού που χρησιμοποιείται. Η ευελιξία είναι εφαρμόσιμη σε διάφορα πρότυπα που χρησιμοποιούνται στο δίκτυο (ανάμεσα τους τα ANSI, IEC, PLC, wireless M-Bus, ZigBee) τα οποία έτσι είναι διαθέσιμα και αναβαθμίσιμα παγκοσμίως.

**Διαθεσιμότητα:** Είναι ένα ουσιώδες χαρακτηριστικό για την παροχή ενέργειας και πληροφοριών στους καταναλωτές, ενώ στηρίζεται στη διαθεσιμότητα των δεδομένων που ανταλλάσσονται στο δίκτυο. Απαιτείται μία υψηλή διαθεσιμότητα πόρων, ειδικά για θέματα τα οποία άπτονται της καθυστέρησης (latency) ή της ασφάλειας. Για παράδειγμα, στα συστήματα προστασίας και ελέγχου της γραμμής η καθυστέρηση πρέπει να είναι της τάξεως των ms, αλλά ένα φαινόμενο άρνησης υπηρεσίας (Denial of Service) ενδέχεται να δημιουργήσει προβλήματα στις επιδόσεις του δικτύου καθιστώντας τους servers ή τις υπηρεσίες μη διαθέσιμες προσωρινά. Το πρόβλημα θα μπορούσε να επιλυθεί μέσω του πλεονασμού (redundancy). Η αποτελεσματικότητα του όμως εξαρτάται από τον σχεδιασμό του συστήματος ούτως ώστε να αποφευχθεί το κόστος της υψηλής πολυπλοκότητας του δικτύου καθώς και η κλιμάκωση.

**Ανθεκτικότητα:** Καθορίζει την αξιοπιστία του δικτύου όταν προκύπτουν διάφορα συμβάντα. Το δίκτυο θα πρέπει να παρέχει ασφαλώς ηλεκτρική ενέργεια στους καταναλωτές ασχέτως των οποιωνδήποτε εσωτερικών ή εξωτερικών κινδύνων. Αναφορικά με την ασφάλεια, η ανθεκτικότητα αντιπροσωπεύει τη δυνατότητα ανάκτησης και αποκατάστασης μετά από την οποιαδήποτε διαταραχή ή δυσλειτουργία, μέσα από μία εύρωστη διαδικασία γρήγορης απόκρισης. Το δίκτυο με τον τρόπο αυτό αποκτά τη δυνατότητα δυναμικού επαναπροσδιορισμού ούτως ώστε να ανακάμπτει από επιθέσεις, διακοπές ρεύματος, φυσικές καταστροφές, κακόβουλες ενέργειες και βλάβες των στοιχείων του. Ευάλωτα στοιχεία είναι οι γραμμές μεταφοράς και οι σταθμοί, οι μεγάλες μονάδες παραγωγής ηλεκτρικής ενέργειας



και οι πυρηνικοί σταθμοί με διαρροή. Για την αντιμετώπιση των συγκεκριμένων περιστατικών απαιτείται η εκπόνηση σχεδίων έκτακτης ανάγκης.

**Συντηρησιμότητα:** Επί της ουσίας απεικονίζει την αντοχή στον χρόνο και την αξιοπιστία του δικτύου. Αποτελεί μίας ένδειξης των δυνατοτήτων αποτελεσματικής και αποδοτικής εκτέλεσης διαφόρων εργασιών συντήρησης. Οι σχετικές διαδικασίες αποτελούνται από την επιθεώρηση, την αντιμετώπιση προβλημάτων και την αντικατάσταση. Το δίκτυο πρέπει να σχεδιάζεται έτσι ώστε να διευκολύνεται η συντήρηση, ούτως ώστε οι διάφορες συνιστώσες ενέργειας και επικοινωνιών (για παράδειγμα, εγκαταστάσεις, εξοπλισμός, συστήματα, υποσυστήματα, ασφάλεια του δικτύου, διαχείριση) να επιδιορθώνονται γρήγορα και οικονομικά. Ένα ακόμα σημαντικό στοιχείο του συστήματος συντήρησης του δικτύου είναι η υψηλή αποδοτικότητα εργατοώρας, εργαλείων και εξοπλισμού.

**Βιωσιμότητα:** Σχετίζεται με την επάρκεια, την αποδοτικότητα και τη φιλικότητα του δικτύου προς το περιβάλλον. Η αύξηση της ζήτησης ηλεκτρικής ενέργειας πρέπει να καλυφθεί μέσα από την αξιοποίηση προσιτών εναλλακτικών ενεργειακών πηγών, την αύξηση της εξοικονόμησης ενέργειας μέσα από την αξιοποίηση της τεχνολογίας στη λειτουργία του συστήματος παροχής και τον περιορισμό της συμφόρησης του δικτύου. Οι νέες τεχνολογίες που θα αξιοποιηθούν πρέπει να προκαλούν χαμηλότερη μόλυνση ή εκπομπές και να μην εξαρτώνται από τον άνθρακα, λαμβάνοντας υπόψη τις περιβαλλοντικές και κλιματικές αλλαγές.

**Διαλειτουργικότητα:** Εξαρτάται σε σημαντικό βαθμό από αυτήν η αποδοτικότητα και αποτελεσματικότητα των συνολικών επιδόσεων του δικτύου. Είναι απαραίτητη η ύπαρξη ενός συνόλου από κοινά και διαλειτουργικά πρότυπα για τη διασύνδεση της ενέργειας και των επικοινωνιών. Το συγκεκριμένο χαρακτηριστικό είναι αναγκαίο κατά την ενσωμάτωση και σύγκλιση διαφόρων τεχνολογιών και τηλεπικοινωνιακών πρωτοκόλλων, ούτως ώστε να γίνονται κατανοητά μεταξύ τους και να παρέχουν αδιάλειπτη μεταφορά ενέργειας και δεδομένων. Μία μη αποτελεσματική διασύνδεση των διαφόρων μερών του δικτύου θα είχε ως αποτέλεσμα την επιβράδυνση του χρόνου απόκρισης και την υποβάθμιση της λειτουργίας και της αποδοτικότητας του συνολικού συστήματος.

**Ασφάλεια:** Έχει ως αντικείμενο της προβλήματα του δικτύου τα οποία έχουν τη ρίζα τους σε ανθρώπινες παρεμβάσεις, όπως είναι οι κακόβουλες επιθέσεις και οι μη εξουσιοδοτημένες τροποποιήσεις. Μία ασφαλής σύνδεση παρόχου και καταναλωτή παρέχει προστασία για τις σημαντικές εφαρμογές και τα δεδομένα καθώς και άμυνες ενάντια σε παραβιάσεις της ασφάλειας. Διάφορα μέτρα και εργαλεία που σχετίζονται με την ασφάλεια αποτελούν ουσιώδη συστατικά στοιχεία ενός ευφυούς δικτύου. Ορισμένα σχετικά παραδείγματα είναι τα συστήματα firewall, τα συστήματα ανίχνευσης και αποτροπής εισβολών (IDS/IPS), τα εικονικά ιδιωτικά δίκτυα (Virtual Private Networks – VPNs), τα εικονικά τοπικά δίκτυα (Virtual Local Area Networks – VLANs), καθώς και ο έλεγχος πρόσβασης.

**Βελτιστοποίηση:** Αποτελεί ένα ουσιώδες χαρακτηριστικό του οποιουδήποτε ευφυούς δικτύου. Επιτυγχάνεται μέσα από την αξιοποίηση των πλέον σύγχρονων τεχνολογιών και των ευφύων ηλεκτρικών συσκευών, καθώς και του ευφυούς ελέγχου και των αυτοματισμών, με την ταυτόχρονη εξισορρόπηση ενός εύρους μεταβλητών. Η βελτιστοποίηση του ευφυούς δικτύου νοείται στους τομείς της αξιοπιστίας της παροχής ηλεκτρικής ενέργειας, της αποδοτικότητας της μετατροπής και χρήσης της, της ποιότητας παραγωγής και διανομής, της διαθεσιμότητας για τη μεταφορά ενέργειας και δεδομένων, της αποτελεσματικότητας και ακρίβειας των δεδομένων και επικοινωνιών, της χρονικής απόκρισης και διαχείρισης των σφαλμάτων, καθώς και των οικονομικών κερδών. Επίσης η μείωση των εξόδων, η πολυπλοκότητα του δικτύου και η αξιοποίηση των πόρων είναι ιδιαίτερα σημαντικοί παράγοντες του ευφυούς δικτύου το οποίο θα υλοποιηθεί επί του πρακτέου.

Ορισμένα επιπρόσθετα χαρακτηριστικά των ευφύων δικτύων ηλεκτρικής ενέργειας του μέλλοντος θα μπορούσαν να είναι και τα παρακάτω:

**Ψηφιοποίηση:** Το ευφύες δίκτυο θα αξιοποιεί μία μοναδική, ψηφιακή πλατφόρμα για την ταχεία και αξιόπιστη ανίχνευση, μέτρηση, επικοινωνία, υπολογισμό, έλεγχο, προστασία, απεικόνιση και συντήρηση όλου του συστήματος μεταφοράς. Το συγκεκριμένο χαρακτηριστικό θα καταστήσει εύκολη και την ανάπτυξη άλλων ευφύων λειτουργιών. Η συγκεκριμένη πλατφόρμα θα χαρακτηρίζεται από ένα περιβάλλον φιλικό προς τον χρήστη για την ενημέρωση σημαντικών καταστάσεων καθώς και υψηλή ανοχή προς τα ανθρώπινα λάθη.

**Ευφυΐα:** Στο δίκτυο μεταφοράς θα ενσωματωθούν ευφυείς τεχνολογίες και ανθρώπινη τεχνογνωσία. Θα είναι διαθέσιμη η αυτο-επίγνωση της κατάστασης λειτουργίας του δικτύου

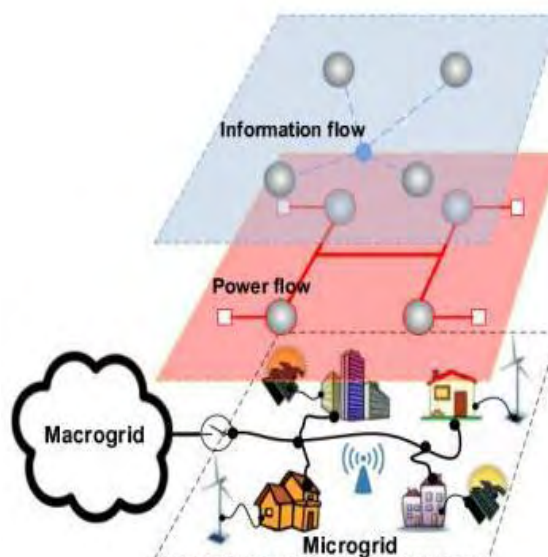
μέσα από την online ανάλυση στο πεδίο του χρόνου, όπως είναι η ανάλυση της σταθερότητας τάσης - γωνίας και της ασφάλειας. Επίσης θα υπάρχει η δυνατότητα αυτοθεραπείας για την ενίσχυση της ασφάλειας του δικτύου μεταφοράς μέσα από συντονισμένα σχήματα προστασίας και ελέγχου.

**Προσαρμογή:** Το ευφυές δίκτυο θα έχει σχεδιασθεί προσαρμοσμένο προς τον πελάτη, ούτως ώστε να διευκολυνθούν οι φορείς εκμετάλλευσής του, χωρίς να χάνει τις λειτουργίες και τη διαλειτουργικότητά του. Θα προσφέρει στους καταναλωτές περισσότερες επιλογές κατανάλωσης ενέργειας, ούτως ώστε να αυξηθεί ο λόγος ποιότητας – τιμής. Η αγορά ενέργειας θα απελευθερωθεί περισσότερο χάρη στα ευφυή δίκτυα μέσα από την αύξηση της διαφάνειας και τη βελτίωση του ανταγωνισμού μεταξύ των παρόχων ηλεκτρικής ενέργειας.

### 3.7 Παράδειγμα δικτύων στο έξυπνο δίκτυο

#### 3.7.1 Μικροδίκτυο

Το Μικροδίκτυο, που δημιουργείται λόγω της κατακεκομμένης παραγωγής ηλεκτρικής ενέργειας, αποτελεί σημαντικό τμήμα των μελλοντικών έξυπνων δικτύων. Περιλαμβάνει ηλεκτρογεννήτριες, αποθηκευτικές διατάξεις και ηλεκτρικά φορτία και είναι συνδεδεμένο με το κλασικό δίκτυο ηλεκτρικής ενέργειας (μακροδίκτυο), με δυνατότητα αποσύνδεσής του από αυτό. Μετά την αποσύνδεση του έχει την ικανότητα να λειτουργεί απομονωμένα (islanded mode). Οι ηλεκτρικές γεννήτριες κατακεκομμένης παραγωγής συνεχίζουν να τροφοδοτούν τους πελάτες του δικτύου με ενέργεια χωρίς να είναι απαραίτητες οι εγκαταστάσεις παραγωγής του μακροδικτύου. Αξιοσημείωτο είναι το γεγονός ότι οι χρήστες, παρ' όλο που δεν είναι συνδεδεμένοι στο μακροδίκτυο, ανταλλάσσουν πληροφορίες με αυτό. Αυτό γίνεται διότι πρέπει να είναι γνωστή η κατάσταση του μακροδικτύου έτσι ώστε αν υπάρξει ανάγκη να επανασυνδεθεί το μικροδίκτυο. Στην εικόνα 5 φαίνεται ένα παράδειγμα μακροδικτύου.

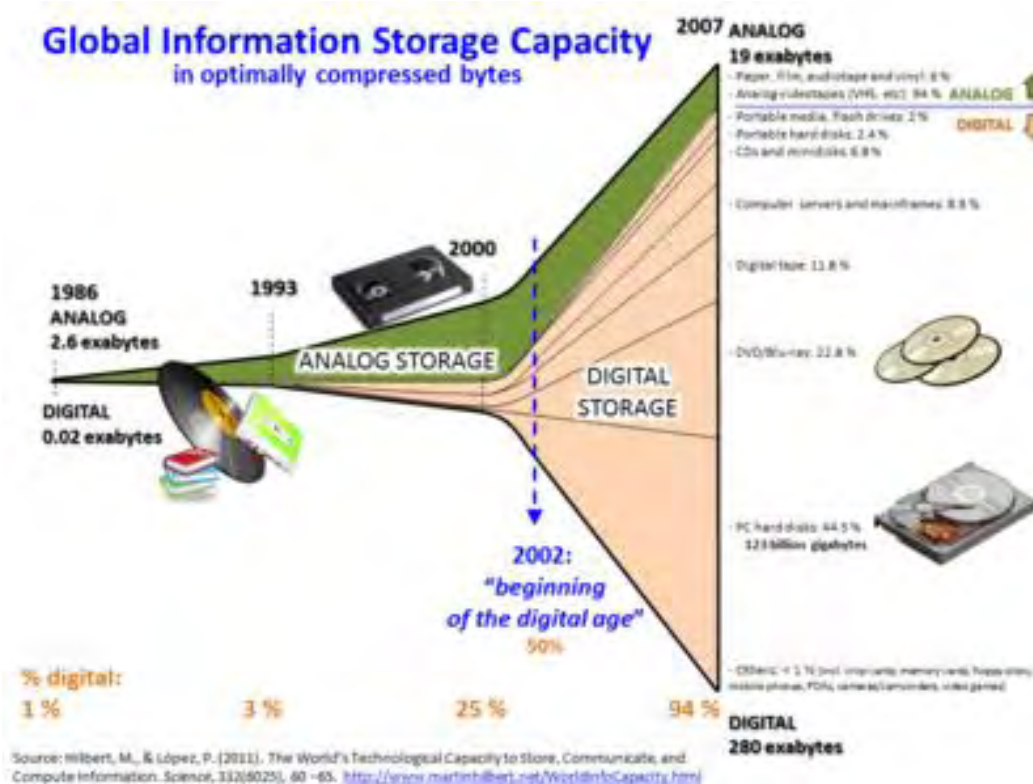


Εικόνα 5 - Στο κατώτερο επίπεδο φαίνεται ένα απλό μικροδίκτυο. Στο μεσαίο η ροή της ενέργειας σε αυτό. Στο πάνω επίπεδο η ροή της πληροφορίας.

# Κεφάλαιο 4 ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ (BIGDATA) - ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΟΥΣ.

## 4.1 Γενικά

Ο όρος μεγάλα δεδομένα (bigdata) αναφέρεται στη μελέτη και τις εφαρμογές για datasets τα οποία είναι τόσο μεγάλα και σύνθετα που δεν είναι δυνατόν να αποθηκευτούν και να αντιμετωπισθούν από τα κλασικά λογισμικά επεξεργασίας δεδομένων.



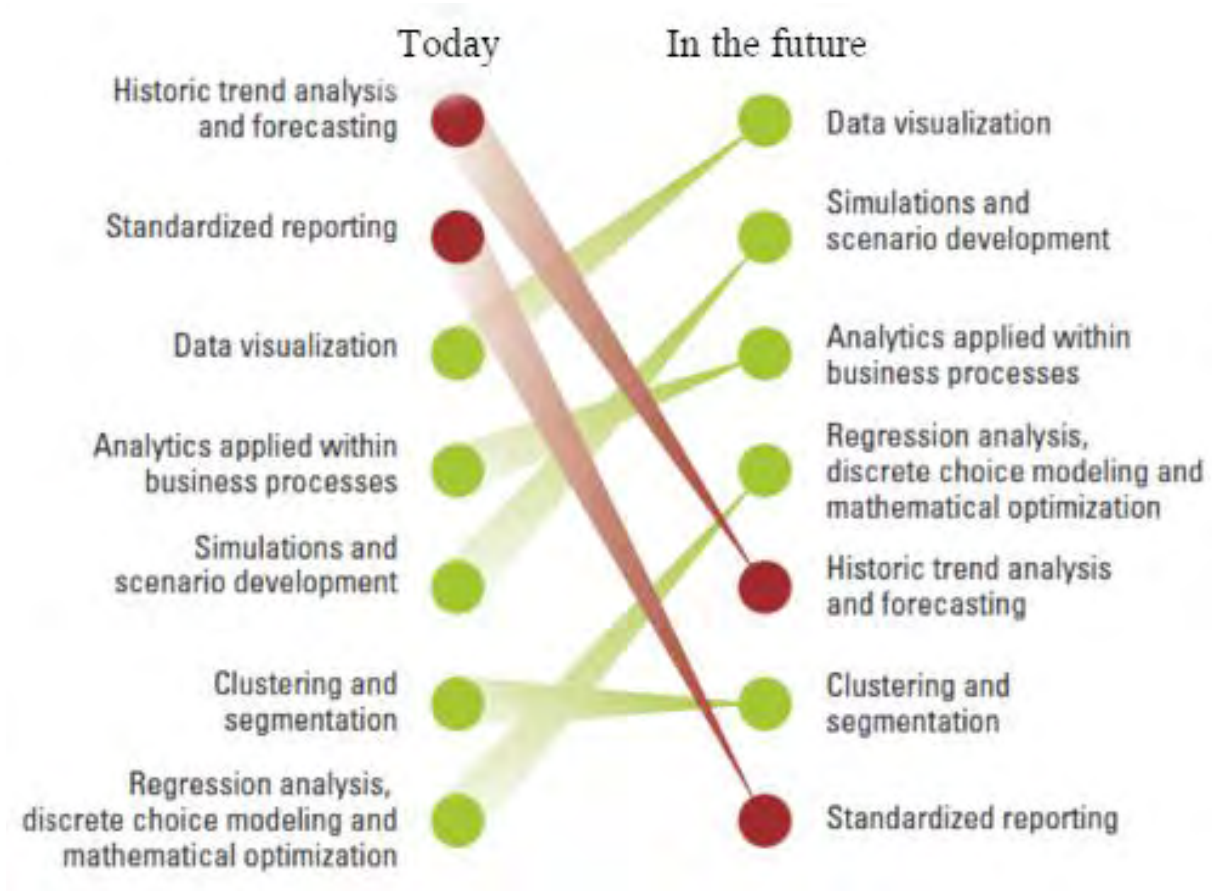
Διάγραμμα1 - Growth of and digitization of global information-storage capacity

Η ανάπτυξη της τεχνολογίας τα τελευταία χρόνια (αύξηση αποθηκευτικού χώρου, φθηνή τεχνολογία για πρόσβαση στο Internet διαθέσιμη για μεγαλύτερο μέρος του πληθυσμού, φθηνοί αισθητήρες) και του Internet of Things, έχει προκαλέσει έκρηξη από δεδομένα. Σήμερα, τα μεγάλα δεδομένα θεωρούνται από πληθώρα αναλυτών παγκοσμίως, αλλά και από περιοδικά όπως το "The Economist" η πιο πολύτιμη ύλη για τις επιχειρήσεις. Έτσι, όλο και περισσότερες εταιρίες χρησιμοποιούν την ανάλυση δεδομένων για την επίτευξη των στόχων τους, με αποτέλεσμα αυτή η μέθοδος να έχει καταστεί απαραίτητη και τα εργαλεία που μπορούν να χρησιμοποιηθούν σε αυτή, προαπαιτούμενη γνώση.

## 4.2 Εξέλιξη των μεγάλων δεδομένων

Από την οπτική γωνία της πληροφορίας και της τεχνολογίας της επικοινωνίας, τα μεγάλα δεδομένα αποτελούν μία ισχυρή ώθηση για τη νέα γενιά βιομηχανιών της τεχνολογίας της πληροφορίας (information technology) οι οποίες βασίζονται κυρίως στα μεγάλα δεδομένα, το cloud computing, το Internet of Things (IoT) και το social business.

Με την εισαγωγή πλατφορμών μεγάλων δεδομένων, υπήρξε μια αλλαγή στις τεχνικές ανάλυσης των οργανισμών. Το επίκεντρο των οργανισμών έχει μεταφερθεί από ορθόδοξες μεθόδους, όπως η ανάλυση τάσεων και η πρόβλεψη, χρησιμοποιώντας ιστορικά δεδομένα, στις συμπληρωματικές και πολύ καλύτερες τεχνικές οπτικοποίησης (visualization - παρουσίαση των δεδομένων σε εικονογραφική ή γραφική μορφή) δεδομένων. Η προσομοίωση και ανάπτυξη σεναρίων είναι πιο σημαντική σε σχέση με τις τυποποιημένες τεχνικές αναφοράς και η ανάλυση είναι ένας παράγοντας κλειδί για την ενίσχυση των επιχειρηματικών διαδικασιών. [16]



Εικόνα 6 - Μετάβαση τεχνολογίας (Πηγή: <http://www.ttivanguard.com/realtime/bigdata.pdf>)

Τα μεγάλα δεδομένα χρησιμοποιούνται σε πολλούς τομείς όπως η ενέργεια, οι μεταφορές, η υγεία, η δημόσια διοίκηση, το εμπόριο, η βιοχημεία και άλλες διεπιστημονικές έρευνες. Οι εφαρμογές που βασίζονται στο διαδίκτυο συναντούν συχνά μεγάλα δεδομένα, όπως είναι τα μέσα κοινωνικής δικτύωσης, οι μηχανές αναζήτησης κ.α.

Στον πίνακα 3. που ακολουθεί, παρουσιάζονται στοιχεία της ποσότητας παραγωγής και της ποικιλομορφίας των μεγάλων δεδομένων που παράγονται από κοινωνικά δίκτυα.

Δεδομένα social media	Παραγωγή δεδομένων
<b>YouTube</b> (Youtube, 2018)	<ul style="list-style-type: none"> <li>· Οι χρήστες ανεβάζουν 100 ώρες από καινούργια βίντεο κάθε λεπτό</li> <li>· Πάνω από 1 δις ξεχωριστοί χρήστες ανοίγουν το YouTube κάθε μήνα</li> <li>· Πάνω από 6 δισεκατομμύρια ώρες αφιερώνονται στην παρακολούθηση βίντεο κάθε μήνα. Δηλαδή σχεδόν μια ώρα για κάθε άτομο στη Γη και 50% περισσότερο από πέρυσι</li> <li>· Πάνω από τις μισές προβολές στο YouTube προέρχονται από κινητές συσκευές</li> </ul>
<b>Facebook</b> (Facebook, 2018)	<ul style="list-style-type: none"> <li>· Λαμβάνει 34.722 "μου αρέσει" κάθε λεπτό</li> <li>· 100 στοιχεία terabyte δεδομένων που μεταφορτώνονται καθημερινά</li> <li>· 1,45 δισεκατομμύρια καθημερινά ενεργοί χρήστες κατά μέσο όρο για το Μάρτιο του 2018</li> <li>· Χρησιμοποιεί 70 γλώσσες</li> <li>· 27.742 εργαζόμενοι στις 31 Μαρτίου 2018</li> </ul>
<b>Twitter</b> (Twitter, 2018)	<ul style="list-style-type: none"> <li>· 336 εκατομμύρια ενεργοί χρήστες στο πρώτο τρίμηνο του 2018</li> <li>· Πάνω από 500 εκατομμύρια tweets στέλνονται καθημερινά</li> </ul>
<b>Google+</b> (plus, 2014)	<ul style="list-style-type: none"> <li>· 1 δισεκατομμύριο χρήστες</li> </ul>

Δεδομένα social media	Παραγωγή δεδομένων
<b>Google</b> (Google, 2014)	Λαμβάνει πάνω από 2 εκατομμύρια ερωτήματα αναζήτησης ανά λεπτό Επεξεργάζεται 25 petabytes δεδομένων κάθε μέρα
<b>Pinterest</b> (Pinterest, 2018)	175 εκατομμύρια μηνιαίοι ενεργοί χρήστες
<b>Apple</b> (Apple, 2018)	Λαμβάνει περίπου 47.000 λήψεις εφαρμογών ανά λεπτό 1,3 δισεκατομμύρια ενεργές συσκευές Apple
<b>Tumblr</b> (Tumblr, 2014)	Οι ιδιοκτήτες blog δημοσιεύουν 27.000 νέες δημοσιεύσεις ανά λεπτό
<b>Instagram</b> (Instagram, 2018)	800 εκατομμύρια ενεργοί μηνιαίοι χρήστες Πάνω από 40 δισεκατομμύρια φωτογραφίες έχουν δημοσιευτεί μέχρι στιγμής
<b>Snapchat</b> (Snapchat, 2018)	Πάνω από 300 εκατομμύρια ενεργοί μηνιαίοι χρήστες Το 71% των χρηστών Snapchat είναι κάτω των 34 ετών
<b>LinkedIn</b> (LinkedIn, 2018)	Πάνω από 250 εκατομμύρια ενεργοί μηνιαίοι χρήστες
<b>Foursquare</b> (Foursquare, 2014)	8 εκατομμύρια ενεργά “check-ins” με την αντίστοιχη εφαρμογή Swarm

Πίνακας 3 - Δεδομένα που παράγονται από κοινωνικά δίκτυα. (Πηγή: <https://www.omnicoreagency.com/>)

### 4.3 State of the art εργαλεία ανάλυσης μεγάλων δεδομένων

Ο όγκος των μεγάλων δεδομένων και η πληροφορία που αυτά περιέχουν αυξάνεται συνεχώς. Για το λόγο αυτό πρέπει να βρεθούν λύσεις, ώστε να τα διαχειριστούμε κατάλληλα, να κατανοήσουμε την πληροφορία που περιέχουν και την αξία της και να εξάγουμε τη γνώση αυτή προς όφελός μας. Επιπλέον, οι υπεύθυνοι λήψης αποφάσεων πρέπει να είναι σε θέση να λάβουν αποφάσεις βασισμένοι σε αυτά τα ποικίλα και ταχέως μεταβαλλόμενα δεδομένα, από real time μεταβολές στην παραγωγή και ζήτηση ενέργειας και πως αυτή επηρεάζει την τιμή της kWh έως και δεδομένα κοινωνικού δικτύου. Η χρήση αναλυτικών εργαλείων μεγάλων δεδομένων αναδεικνύει την αξία της πληροφορίας που περιέχουν.

#### 4.4 Batch based εργαλεία ανάλυσης

Με τα batch based εργαλεία ανάλυσης πραγματοποιείται επεξεργασία δεδομένων που έχουν ήδη αποθηκευτεί σε μια συγκεκριμένη παρελθοντική χρονική περίοδο. Για παράδειγμα, τη ζήτηση σε kwh που είχαμε κάθε ώρα της ημέρας για ένα χρόνο, ή πόσες kwh εισάγαμε από τις γειτονικές χώρες και από τι είδους μονάδες παραγωγής ενέργειας. Αυτά τα δεδομένα μπορούν να αποθηκευτούν ως αρχείο, ή μέτρηση, κ.λ.π. Το αρχείο αυτό θα υποβληθεί σε επεξεργασία για διάφορες αναλύσεις που επιθυμεί να κάνει ο εκάστοτε φορέας. Είναι προφανές ότι θα απαιτηθεί μεγάλο χρονικό διάστημα για την επεξεργασία αυτού του αρχείου.

Στον πίνακα 4 παρουσιάζονται βασικά batch based εργαλεία επεξεργασίας, περιγράφονται βασικές λειτουργίες του και παρουσιάζονται τα πλεονεκτήματα και τα μειονεκτήματά τους.

Batch-based εργαλεία επεξεργασίας	Περιγραφή	Πλεονεκτήματα	Μειονεκτήματα
<b>Hadoop</b>	Εκτέλεση επεξεργασίας εφαρμογών που απαιτούν μεγάλα δεδομένα	- Κατανεμημένα δεδομένα - Επεξεργασία - Ανεξάρτητα καθήκοντα - Εύκολη διαχείριση μερικής αποτυχίας - Γραμμική κλιμάκωση σε ιδανικές περιπτώσεις - Απλό μοντέλο προγραμματισμού	- Περιοριστικός προγραμματιστικός μοντέλου - Συνδέεται με πολλαπλά σύνολα δεδομένων που το καθιστούν δύσκολο και αργό - Δύσκολη διαχείριση συμπλεγμάτων - Ενιαίος κύριος κόμβος - Απρόσμενη διαμόρφωση των κόμβων
<b>Skytree Server</b>	Επεξεργασία μεγάλης ποσότητας δεδομένων σε υψηλή ταχύτητα	- Γρήγορη επεξεργασία μεγάλης ποσότητας δεδομένων με ακριβή τρόπο - Σύνθετη ανάλυση - Υψηλής απόδοσης μηχανών εκμάθησης	- Υψηλή πολυπλοκότητα



<b>Batch-based εργαλεία επεξεργασίας</b>	<b>Περιγραφή</b>	<b>- Πλεονεκτήματα</b>	<b>- Μειονεκτήματα</b>
<b>Talend Open Studio</b>	Να παρέχει ένα γραφικό περιβάλλον για τη διεξαγωγή ανάλυσης για εφαρμογές μεγάλων δεδομένων	- Πλούσια σύνολα συνιστωσών - Μετατροπή κώδικα - Συνδεσιμότητα με όλες τις βάσεις δεδομένων - Σχεδιασμός υψηλού επιπέδου	- Το σύστημα καθυστερεί μετά την εγκατάσταση του εργαλείου - Μικρός παραλληλισμός
<b>Jaspersoft</b>	Για την δημιουργία αναφορών από βάσεις δεδομένων σε στήλες	- Χαμηλή τιμή - Εύκολη εγκατάσταση - Μεγάλη λειτουργικότητα και αποτελεσματικότητα	- Το Jaspersoft υποστηρίζει σφάλματα τεκμηρίωσης - Η εξυπηρέτηση πελατών δεν ισχύει μετά την επέκταση συγκεκριμένων λειτουργιών του προγράμματος
<b>Tableau</b>	Επεξεργασία μεγάλες ποσότητες βάσεων δεδομένων	- Καταπληκτική οπτικοποίηση δεδομένων - Λύση χαμηλού κόστους για αναβάθμιση - Εξαιρετική υποστήριξη κινητής τηλεφωνίας	- Έλλειψη δυνατοτήτων πρόβλεψης - Επικίνδυνη ασφάλεια - Θέματα διαχείρισης αλλαγής

Πίνακας 4 - Σύγκριση των batch based εργαλείων επεξεργασίας (Πηγή: Yaqoob, et al., 2016).

Τα batch based δεδομένα λειτουργούν καλά σε καταστάσεις όπου δεν χρειαζόμαστε αποτελέσματα σε πραγματικό χρόνο, αλλά όταν είναι σημαντικό να γίνει επεξεργασία μεγάλων όγκων δεδομένων για πιο λεπτομερή κι χρονοβόρα ανάλυση.

## 4.5 Stream based εργαλεία ανάλυσης

Τα stream based εργαλεία ανάλυσης παρέχουν αποτελέσματα σε πραγματικό χρόνο. Επιτρέπουν την επεξεργασία δεδομένων σε πραγματικό χρόνο, σε μικρό χρονικό διάστημα από το σημείο λήψης των δεδομένων και παρέχουν τη δυνατότητα να αποκτηθούν άμεσα αποτελέσματα ανάλυσης για γρήγορη λήψη αποφάσεων και δράσεων.

Στον πίνακα 5 παρουσιάζονται βασικά stream based εργαλεία επεξεργασίας, περιγράφονται βασικές λειτουργίες του και παρατίθενται τα πλεονεκτήματα και τα μειονεκτήματά τους.

Stream-based εργαλεία ανάλυσης	Περιγραφή	Πλεονεκτήματα	Μειονεκτήματα
<b>Storm</b>	Επεξεργασία μεγάλου όγκου δεδομένων σε πραγματικό χρόνο	Εύκολο στην χρήση Λειτουργεί με οποιαδήποτε γλώσσα προγραμματισμού Δυνατότητα κλιμάκωσης Ανεκτικό σε λάθη	Πολλά μειονεκτήματα όσον αφορά την αξιοπιστία, την απόδοση, την αποδοτικότητα και τη δυνατότητα διαχείρισης
<b>Splunk</b>	Συλλογή ευρετηρίων και συσχέτιση δεδομένων σε πραγματικό χρόνο με στόχο την δημιουργία αναφορών, ειδοποιήσεων και οπτικοποιήσεων	Πολλά πλεονεκτήματα από την ασφάλεια των επιχειρηματικών αναλύσεων μέχρι παρακολούθηση της υποδομής	Υψηλό κόστος εγκατάστασης από άποψη χρημάτων Υψηλή πολυπλοκότητα
<b>SAP Hana</b>	Να παρέχει ανάλυση των επιχειρηματικών διαδικασιών σε πραγματικό χρόνο	Ανάλυση υψηλής απόδοσης Γρήγορη επεξεργασία Επεξεργασία μνήμης	Έλλειψη υποστήριξης για όλα τα προϊόντα ERP Υψηλό κόστος Δύσκολη συντήρηση της βάσης SAP Hana
<b>SQLstream s-Server</b>	Ανάλυση μεγάλου όγκου δεδομένων και αρχείων καταγραφής σε πραγματικό χρόνο	Χαμηλό κόστος Με δυνατότητα κλιμάκωσης για μεγάλου όγκου και δεδομένα υψηλής ταχύτητας Χαμηλή	- Υψηλή πολυπλοκότητα

		καθυστέρηση	
<b>Apache Kafka</b>	Διαχείριση μεγάλων δεδομένων μέσω αναλυτικών στοιχείων μνήμης για την λήψη αποφάσεων	Υψηλής απόδοσης Σταθερό Με δυνατότητα κλιμάκωσης Ανεκτικό σε λάθη	API υψηλού επιπέδου

Πίνακας 5–Σύγκριση stream based εργαλείων ανάλυσης (Πηγή: Yaqoob, et al., 2016).

Τα stream based δεδομένα είναι χρήσιμα για εργασίες όπως είναι η ανίχνευση σφάλματος. Εάν, για παράδειγμα γίνει επεξεργασία δεδομένων ροής φορτίου, μπορεί να ανιχνευτούν ανωμαλίες που υποδηλώνουν προβλήματα, συμβάντα, ατυχήματα σε πραγματικό χρόνο, ώστε στη συνέχεια οι υπεύθυνοι να προβούν στις κατάλληλες ενέργειες.

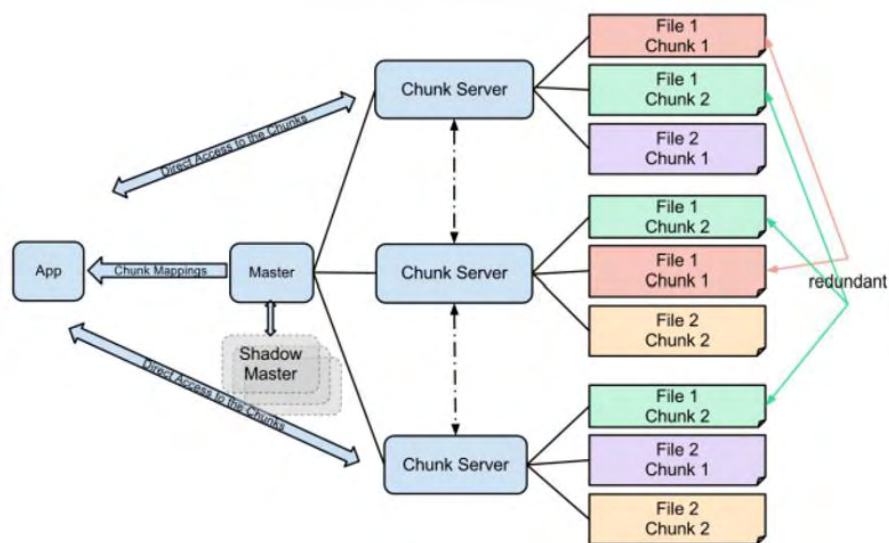
### Apache Hadoop

Για την αντιμετώπιση της αυξημένης ζήτησης στις απαιτήσεις αποθήκευσης και υπολογισμών, τα παλαιά συστήματα βασίζονταν είτε σε λύσεις scale up είτε σε λύσεις scale out. Οι λύσεις scale up χρησιμοποιούν μια κλασσική μη παράλληλη αρχιτεκτονική με βελτιωμένους πόρους, αλλά είναι πολύ ακριβές (κόστος σε σχέση με απόδοση) και περιορίζονται από ένα τεχνικό εμπόδιο. Οι λύσεις scale out χρησιμοποιούν παράλληλη αρχιτεκτονική για να βελτιώσουν τους πόρους υπολογισμού με χαμηλότερο κόστος, αλλά με πολύ υψηλότερη τεχνική προσπάθεια. Από την άλλη πλευρά οι αναλύσεις μεγάλων δεδομένων, όπως το Apache Hadoop, βασίζονται σε ένα πλαίσιο με το μεγαλύτερο μέρος της μηχανικής προσπάθειας να προκαλείται από παράλληλες αρχιτεκτονικές.

Το Apache Hadoop είναι ένα πλαίσιο ανοιχτού κώδικα, γραμμένο σε γλώσσα προγραμματισμού Java. Έχει σχεδιαστεί για να ασχολείται με μεγάλα σύνολα δεδομένων χρησιμοποιώντας συστοιχίες υπολογιστών υλικού. Έχει δύο βασικά μέρη, ένα καταναμημένο τμήμα αποθήκευσης, το **Hadoop** Distributed File System (HDFS) και ένα τμήμα επεξεργασίας, το μοντέλο προγραμματισμού **MapReduce**. Ο σχεδιαστής λογισμικών Doug Cutting ανέπτυξε το Hadoop το 2005 με βάση το Google File System (GFS) και το Google MapReduce.

Η αρχιτεκτονική του **Hadoop** αποτελείται από δύο κύρια τμήματα. Ένα τμήμα αποθήκευσης που διαχειρίζεται το HDFS και ένα τμήμα επεξεργασίας που διαχειρίζεται το μοντέλο προγραμματισμού MapReduce ή υψηλότερες γλώσσες προγραμματισμού. Ένα τυπικό σύστημα Hadoop αποτελείται από ένα κύριο διακομιστή (με ένα ή δύο backup mirrors) και πολλά χαμηλού κόστους μηχανήματα που τρέχουν σε περιβάλλον linux. Ο κύριος διακομιστής έχει συστατικά «Όνομα κόμβου» και «Εργαλείο παρακολούθησης εργασιών» που διαχειρίζονται αντίστοιχα το «Data node» (storage task) και το «task tracker» σε άλλες μηχανές.

Όσον αφορά το Κατανεμημένο Σύστημα Αρχείων Hadoop (HDFS), του οποίου η αρχιτεκτονική παρουσιάζεται στο διάγραμμα 2, πρόκειται για ένα πλεονάζον κατανεμημένο σύστημα αποθήκευσης που ονομάζεται HDFS το οποίο αποθηκεύει αρχεία σε μπλοκ που πολλαπλασιάζονται σε πολλαπλές μηχανές. Ένας κύριος διακομιστής (master node) διαχειρίζεται τη διάσπαση δεδομένων και την αναπαραγωγή στους άλλους διακομιστές μπλοκ (worker nodes) που χρησιμοποιούνται τόσο για την αποθήκευση, όσο και για την επεξεργασία δεδομένων.



Διάγραμμα 2–Αρχιτεκτονική HDFS. (Πηγή: [http://en.wikipedia.org/wiki/Google\\_File\\_System](http://en.wikipedia.org/wiki/Google_File_System))

### Μοντέλο προγραμματισμού MapReduce:

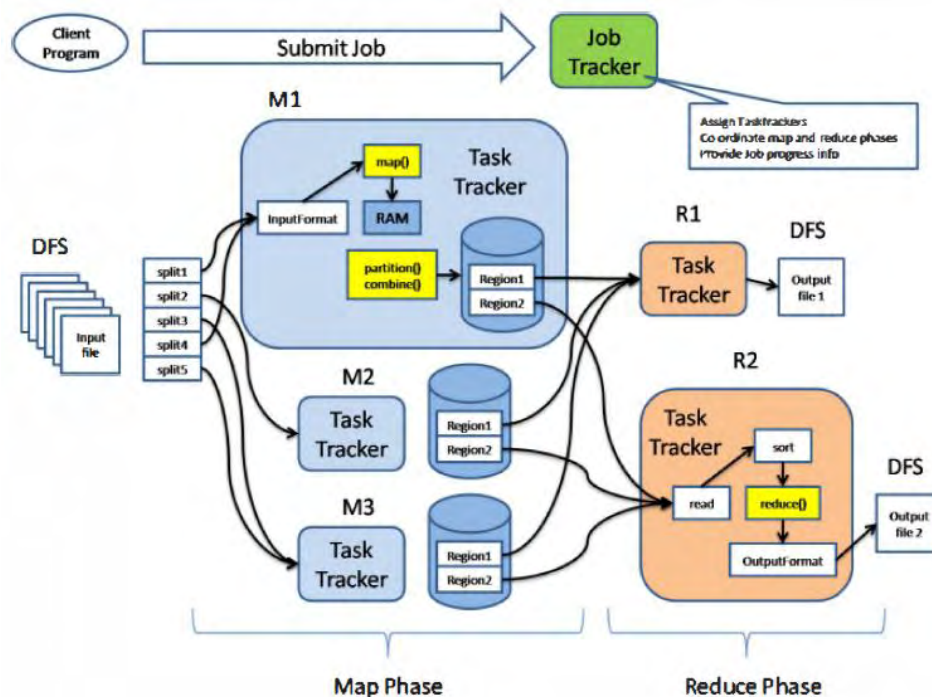
Το Hadoop χρησιμοποιεί κυρίως το MapReduce ως μοντέλο προγραμματισμού για την επεξεργασία μεγάλων συνόλων δεδομένων. Το MapReduce αποτελείται από δύο λειτουργίες:

το «Map» το οποίο χωρίζει τα προβλήματα σε μικρότερα και

το «Reduce» το οποίο συνδυάζει τα αποτελέσματα.

Οι λειτουργίες Map και Reduce πρέπει να γραφούν από το χρήστη. Το MapReduce φροντίζει για όλες τις λεπτομέρειες του κατανεμημένου υπολογισμού. Ο κύριος διακομιστής (master node) δεν είναι υπερφορτωμένος με υπολογισμούς, αλλά είναι υπεύθυνος μόνο για την επικοινωνία με την εφαρμογή των χρηστών και τη διαχείριση των άλλων κόμβων. Οι εργασίες αποστέλλονται στα δεδομένα (όχι στα δεδομένα που αποστέλλονται στη μηχανή των κόμβων) τα οποία βελτιώνουν την απόδοση του συστήματος και κυρίως το bandwidth (μέγιστο ρυθμό μεταφοράς δεδομένων σε μια δεδομένη διαδρομή).

Η αρχιτεκτονική του παρουσιάζεται στο διάγραμμα 3.



Διάγραμμα3–ΑρχιτεκτονικήMapReduce. (Πηγή: <http://architects.dzone.com/articles/how-hadoop-map-reduce-works>)

Εκτός από το μοντέλο προγραμματισμού MapReduce, οι χρήστες μπορούν να αναπτύξουν τους κώδικές τους σε άλλες ευκολότερες γλώσσες προγραμματισμού υψηλού επιπέδου οι οποίες θα μεταφραστούν αυτόματα στις λειτουργίες Map και Reduce, εργασίες όπως:

- Hive: γλώσσα αποθήκης δεδομένων χρησιμοποιώντας ερωτήματα της SQL-92.
- Pig: Γλώσσα προσανατολισμένη προς τις ροές δεδομένων με τη χρήση γλώσσας προγραμματισμού Pig Latin.
- Hbase: Μια sparse βάση δεδομένων για την αποθήκευση μεγάλων ποσοτήτων δεδομένων.

## 4.6 Τεχνικές ανάλυσης δεδομένων

Η ανάλυση δεδομένων είναι ιδιαίτερα σημαντική στη διαδικασία απόκτησης νέας γνώσης καθώς και στη λήψη αποφάσεων. Γενικότερα, από τη μελέτη ενός αντικείμενου που εντάσσεται σε ένα ευρύτερο σύστημα, με τη συμβολή της ανάλυσης των δεδομένων που έχουν συλλεχθεί και με την αρχική γνώση του αντικείμενου, εξάγονται χρήσιμα συμπεράσματα και που δύνανται να οδηγήσουν στη λήψη αποφάσεων για τη βελτιστοποίηση του συστήματος. Στον πίνακα 6, περιγράφονται ορισμένες τεχνικές ανάλυσης μεγάλων δεδομένων, η χρήση τους σε διεπιστημονικές εφαρμογές π.χ. υγεία, εξυπηρέτηση πελατών κ.α., καθώς και αλγόριθμοι, τεχνικές και διαθέσιμα εργαλεία που τα απαρτίζουν.

Τεχνικές ανάλυσης μεγάλων δεδομένων	Περιγραφή	Χρήση σε μερικές διεπιστημονικές εφαρμογές	Αλγόριθμοι / Τεχνικές	Διαθέσιμα εργαλεία
<b>Data mining</b>	Εύρεση σταθερών προτύπων ή / και συστηματικών σχέσεων μεταξύ των μεταβλητών	-Βιοϊατρική -Υγεία	-K-Mean -Fuzzy C-Mean -CLARA -CLARANS -BIRCH	-Excel -Rapid-I -Rapidminer-R -KNMINE -Weka/Pentaho
<b>Social network analysis</b>	Για να δείτε τις κοινωνικές σχέσεις από την άποψη της θεωρίας του δικτύου	-Ανθρωπολογία -Μέσα κοινωνικής δικτύωσης	-PCA -LTSA -LLE -Autoencoder	-Cytoscape -Gephi -Cuttlefish -MeerKat
<b>Web mining</b>	Ανακάλυψη προτύπων χρήσης από μεγάλα αποθετήρια ιστού	-Ηλεκτρονική μάθηση -Ψηφιακές βιβλιοθήκες -Ηλεκτρονική διακυβέρνηση	-LOGML -Apriori	-KXEN -LIONsolver -Dataiku
Τεχνικές ανάλυσης μεγάλων δεδομένων	Περιγραφή	Χρήση σε μερικές διεπιστημονικές εφαρμογές	Αλγόριθμοι / Τεχνικές	Διαθέσιμα εργαλεία
<b>Machine learning</b>	Επιτρέπει στους υπολογιστές να εξελίσσουν συμπεριφορές βάσει εμπειρικών δεδομένων	-Υγεία -Εξυπηρέτηση πελατών	-Pattern recognition -Artificial neural Networks	-Weka -Scikit-Learn -PyMc -Shogun
<b>Visualization approaches</b>	Αναπαράσταση τη γνώσης μέσω της χρήσης	-Τραπεζικά -Κατασκευαστικά βοηθήματα	-FLOT -GGPLOT2	-Data wrapper -Highcharts JS -MAPBox

	γραφημάτων			
<b>Optimization Methods</b>	Επίλυση ποσοτικών προβλημάτων	-Επιστήμη του κοινωνικού δικτύου -Υπολογιστική βιολογία	[–]reduction –Parallelization –Simulated annealing –Quantum annealing – Swarm optimization	–Matlab

Πίνακας 6–Σύγκριση διαφορετικών τεχνικών ανάλυσης δεδομένων. (Πηγή: Yaqoob, et al., 2016)

## 4.7 Ανάλυση δεδομένων με την γλώσσα Python

Στην παρούσα εργασία, η ανάλυση των δεδομένων γίνεται με τη γλώσσα προγραμματισμού Python στο περιβάλλον του Jupyter Notebook. Οι βιβλιοθήκες που θα χρησιμοποιηθούν παρουσιάζονται αναλυτικά παρακάτω.

### 4.7.1 Ιστορική αναδρομή

Η γλώσσα προγραμματισμού Python δημιουργήθηκε το 1990 από τον Guido van Rossum στο Centrum Wiskunde & Informatica (CWI).

Είναι γλώσσα προγραμματισμού γενικής χρήσης, πάρα πολύ υψηλού επιπέδου, απλή και εύκολη στην εκμάθηση, ισχυρή, δυναμική, αποδοτική, παραγωγική και επεκτάσιμη.

Διαθέτει αποδοτικές δομές δεδομένων υψηλού επιπέδου και υποστηρίζει, χωρίς να αναγκάζει, μια απλή αλλά ταυτόχρονα αρκετά αποτελεσματική προσέγγιση στον αντικειμενοστραφή προγραμματισμό.

Υποστηρίζει και άλλες γνωστές προγραμματιστικές προσεγγίσεις, όπως είναι ο διαδικαστικός και ο συναρτησιακός προγραμματισμός.

Είναι διερμηνευόμενη (interpreted) γλώσσα προγραμματισμού και μπορεί να χρησιμοποιηθεί τόσο για τη δημιουργία σεναρίων εντολών όσο και για τη γρήγορη ανάπτυξη ολοκληρωμένων εφαρμογών σε διάφορες περιοχές ενδιαφέροντος (διαχείριση συστήματος υπολογιστή, ανάπτυξη εφαρμογών Διαδικτύου, επεξεργασία αρχείων κειμένου, επιστημονικές εφαρμογές, εκπαίδευση, ανάπτυξη παιχνιδιών, κ.λπ.) και στις περισσότερες πλατφόρμες υλικού υπολογιστών και Λειτουργικών Συστημάτων (Windows, Unix, Linux, MacOS X, κ.λπ.)

Η ονομασία της οφείλεται στη δημοφιλή κωμική σειρά Monty Python's Flying Circus του Β.Β.С. της Μεγάλης Βρετανίας (δεκαετία του '70).

#### 4.7.2 Python και ανάλυση δεδομένων

Για ένα σημαντικό αριθμό εργαζομένων, η ανάλυση δεδομένων αποτελεί το βασικότερο αντικείμενο της εργασίας τους. Η αυξημένη διαθεσιμότητα δεδομένων, η δυναμικότερη επεξεργασία και η αποδιδόμενη έμφαση σε επιχειρηματικές αποφάσεις βασιζόμενες στην ανάλυση των δεδομένων, έχει φέρει στο προσκήνιο την επιστήμη των δεδομένων. Σύμφωνα με έκθεση της IBM, το 2015 υπήρχαν 2,35 εκατομμύρια κενές θέσεις εργασίας σχετικά με την ανάλυση δεδομένων στις ΗΠΑ και εκτιμάται ότι ο αριθμός αυτός θα αυξηθεί σε 2,72 εκατομμύρια μέχρι το 2020.

Ένα σημαντικό ποσοστό των ανθρώπων για τους οποίους η χρήση αριθμών δεσπόζει στη φύση της εργασίας τους χρησιμοποιούν το Microsoft Excel ή άλλα προγράμματα υπολογιστικών φύλλων, όπως το Google Sheets. Άλλοι χρησιμοποιούν στατιστικό λογισμικό όπως το SAS, το Stata ή το SPSS που αρχικά έμαθαν στο σχολείο τους.

Παρότι το Excel και το SAS είναι ισχυρά εργαλεία, έχουν σοβαρούς περιορισμούς. Το Excel δε μπορεί να χειριστεί σύνολα δεδομένων πάνω από ένα συγκεκριμένο μέγεθος και δεν επιτρέπει εύκολα την αναπαραγωγή προηγούμενων αναλύσεων σε νέα σύνολα δεδομένων. Η κύρια αδυναμία των προγραμμάτων όπως το SAS είναι ότι αναπτύχθηκαν για πολύ συγκεκριμένες εργασίες επίσης και δεν υπάρχει μεγάλος αριθμός συνεργατών που να του προσθέτει συνεχώς νέα εργαλεία.

Πέρα από τα όρια αυτών των εφαρμογών, υπάρχουν οι R και η Python. Οι γλώσσες R και Python είναι οι δύο πιο δημοφιλείς γλώσσες προγραμματισμού που χρησιμοποιούνται από τους αναλυτές δεδομένων και τους επιστήμονες δεδομένων. Και οι δύο είναι ελεύθερες και ανοιχτές πηγές λογισμικού και αναπτύχθηκαν στις αρχές της δεκαετίας του 1990 – η R για στατιστική ανάλυση και η Python ως γλώσσα προγραμματισμού γενικής χρήσης. Είναι ιδανικές για όσους ενδιαφέρονται να επεξεργασθούν μεγάλα σύνολα ή να δημιουργήσουν σύνθετες απεικονίσεις δεδομένων.

Η Python είναι κατάλληλη για χειρισμό δεδομένων και επαναλαμβανόμενες διεργασίες, ενώ η R είναι κατάλληλη ειδικά για ανάλυση και διερεύνηση συνόλων δεδομένων.

Σημαντικό πλεονέκτημα της Python είναι ότι είναι μια γενικότερη γλώσσα προγραμματισμού που αντικατοπτρίζει τον τρόπο σκέψης των προγραμματιστών



ηλεκτρονικών υπολογιστών και είναι κατάλληλη για χειρισμό δεδομένων και επαναλαμβανόμενες διεργασίες.

Το Facebook, σύμφωνα με άρθρο του περιοδικού Fast Company για το 2014, επέλεξε να χρησιμοποιήσει την Python για ανάλυση δεδομένων, επειδή ήδη χρησιμοποιήθηκε τόσο ευρέως σε άλλα μέρη της εταιρείας. Για το λόγο αυτό και για άλλους, η Python είναι πολύ αγαπητή από τους προγραμματιστές.

### 4.7.3 Βιβλιοθήκες της Python

Όπως και για πολλές άλλες γλώσσες προγραμματισμού, οι διαθέσιμες βιβλιοθήκες αποτελούν για την Python πολύ σημαντικό πλεονέκτημα. Υπάρχουν περίπου 72.000 περίπου στο Python Package Index (PyPI) και αυξάνονται συνεχώς.

Επιγραμματικά, οι βασικές βιβλιοθήκες της είναι:

- **NumPy**

Αποτελεί τη βάση για τη δημιουργία όλων των εργαλείων υψηλότερου επιπέδου για την επιστημονική Python. Ορισμένες από τις λειτουργίες της είναι:

1. N- Dimensional array, ένας πολυδιάστατος πίνακας με γρήγορη και αποδοτική μνήμη που παρέχει αριθμητικές πράξεις.
2. Μπορούν να εφαρμοστούν τυποποιημένες μαθηματικές πράξεις σε πίνακες ολόκληρων δεδομένων χωρίς να γραφούν βρόχοι.
3. Είναι πολύ εύκολο να μεταφερθούν δεδομένα σε εξωτερικές βιβλιοθήκες γραμμένες σε γλώσσα χαμηλού επιπέδου και επίσης για εξωτερικές βιβλιοθήκες να επιστρέφουν δεδομένα σε Python ως Numpy arrays.

Η NumPy δεν παρέχει λειτουργικότητα ανάλυσης δεδομένων υψηλού επιπέδου.

- **Pandas**

Το ακρωνύμιο Pandas προέρχεται από τον οικονομετρικό όρο «panel data» και «Python and data analysis».

Το Pandas είναι μια βιβλιοθήκη λογισμικού γραμμένη για τη γλώσσα προγραμματισμού Python. Χρησιμοποιείται για τον χειρισμό και την ανάλυση δεδομένων. Παρέχει ειδικές δομές δεδομένων και λειτουργίες για τον χειρισμό αριθμητικών πινάκων και χρονοσειρών.

Στοχεύει σε πέντε τυπικά βήματα στην επεξεργασία και ανάλυση δεδομένων, ανεξάρτητα από την προέλευση των δεδομένων: φόρτωση, προετοιμασία, χειρισμό, μοντελοποίηση και ανάλυση.

- **Scipy**

Η βιβλιοθήκη αυτή εξαρτάται από τη NumPy, η οποία παρέχει εύκολο και γρήγορο χειρισμό πίνακα N-διαστάσεων(N- Dimensional array). Είναι δημιουργημένη ώστε να λειτουργεί με τους αριθμητικούς πίνακες NumPy και παρέχει πολλές φιλικές προς το χρήστη και αποτελεσματικές αριθμητικές ρουτίνες, όπως ρουτίνες για αριθμητική ενσωμάτωση και βελτιστοποίηση. Διαθέτει ενότητες βελτιστοποίησης, γραμμικής άλγεβρας, ολοκλήρωσης και άλλων κοινών διεργασιών στην επιστήμη των δεδομένων.

- **Scikit-learn**

Είναι βιβλιοθήκη της Python ιδανική για μηχανική μάθηση (Machine Learning) που βασίζεται στη NumPy και τη SciPy. Η Scikit-learn είναι ταχύτερη στην εφαρμογή γνωστών αλγορίθμων μηχανικής μάθησης στο σύνολο δεδομένων. Περιλαμβάνει επίσης εργαλεία για πολλές τυπικές εργασίες εκμάθησης μηχανών και εξόρυξης δεδομένων (όπως ομαδοποίηση, ταξινόμηση, παλινδρόμηση κ.λπ.).

- **Ipython**

Η Ipython επεκτείνει τη λειτουργικότητα του διαδραστικού διερμηνέα της Python με ένα αλληλεπιδραστικό κελί που προσθέτει ενδοσκόπηση, εμπλουτισμένα μέσα, σύνταξη κελιού, ολοκλήρωση καρτελών και ανάκτηση ιστορικού εντολών. Λειτουργεί επίσης ως ενσωματωμένος διερμηνέας για τα προγράμματα, ο οποίος μπορεί να είναι πραγματικά χρήσιμος για τον εντοπισμό σφαλμάτων.

- **Matplotlib**

Η Matplotlib είναι μια βιβλιοθήκη της Python για απεικόνιση (visualization). Η Matplotlib επιτρέπει να γίνονται εύκολα γραφήματα γραμμών, διάγραμμα πίτας, ιστόγραμμα και άλλα επαγγελματικά στοιχεία. Όταν χρησιμοποιείται μέσα στην Ipython, η Matplotlib έχει διαδραστικά χαρακτηριστικά όπως το ζουμ και το panning. Υποστηρίζει διαφορετικά GUI

back ends σε όλα τα λειτουργικά συστήματα και μπορεί επίσης να εξάγει γραφικά σε κοινές μορφές διανύσματος και γραφικών: PDF, SVG, JPG, PNG, BMP, GIF κλπ.

Άλλες γνωστές βιβλιοθήκες της Python είναι η Theano και η TensorFlow για μηχανική μάθηση, η Scrapy, η NLTK και η Pattern για εξόρυξη δεδομένων και επεξεργασία της γλώσσας και η Bokeh, η Basemap και η NetworkX για κατασκευή διαγράμματος και απεικόνιση.

- **Seaborn**

Η Seaborn είναι μια βιβλιοθήκη της Python για απεικόνιση (visualization). Είναι βασισμένη στην Matplotlib και συνδεδεμένη με την pandas. Στοχεύει να κάνει την οπτικοποίηση των δεδομένων το βασικό εργαλείο εξερεύνησης και κατανόησής τους. Έχει προσανατολισμένο στα dataset API για τον έλεγχο των σχέσεων μεταξύ πολλαπλών μεταβλητών και αυτοματοποιημένη εκτίμηση και σχεδίαση μοντέλων γραμμικής παλινδρόμησης για διαφορετικά είδη εξαρτημένων μεταβλητών.

#### **4.7.4 Το μοντέλο StatsModel**

Το StatsModel είναι ένα «module» της Python που περιέχει «classes» και συναρτήσεις για την προσέγγιση πολλών διαφορετικών στατιστικών μοντέλων τόσο για την διεξαγωγή στατιστικών δοκιμών όσο και για τη στατιστική εξερεύνηση των δεδομένων. Περιλαμβάνει εκτενή κατάλογο αποτελεσμάτων για την κάθε προσέγγιση. Ακόμη τα αποτελέσματα αντιπαραβάλλονται με υπάρχοντα στατιστικά πακέτα για να βεβαιωθεί πως είναι τα σωστά. Το μοντέλο είναι «ανοιχτού κώδικα» με ModifiedBSD (3-clause) άδεια.

# Κεφάλαιο 5 ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ

## 5.1 Γενικά

Στο κεφάλαιο αυτό θα δημιουργηθούν μοντέλα πρόβλεψης της οριακής τιμής συστήματος SMP βασισμένα στα ανοιχτά μεγάλα δεδομένα που κοινοποιεί ο ΛΑΓΗΕ στην ιστοσελίδα του.

Τα δεδομένα σε raw μορφή, δηλαδή όπως τα διαθέτει ο ΛΑΓΗΕ στην ιστοσελίδα του (Αγορά ->Ιστορικά στοιχεία->Αποτελέσματα ΗΕΠ) έχουν ως εξής:

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	TOTAL
<b>LOAD DECLARATIONS + LOSSES</b>	5.662	5.195	5.209	5.023	4.884	4.943	5.274	5.708	6.165	6.656	6.953	7.115	7.206	7.183	6.927	7.015	7.039	7.253	7.759	7.898	7.821	7.559	7.079	6.648	156.1
<b>SOUTH ZONE LOAD DECLARATIONS</b>	3.278	2.980	3.036	2.931	2.884	3.002	3.365	3.811	4.000	4.261	4.415	4.437	4.379	4.352	4.199	4.253	4.293	4.518	4.814	4.947	4.812	4.609	4.227	3.887	95.6
<b>SYSTEM LOSSES</b>	92	87	88	83	79	80	89	105	111	129	137	140	132	133	121	139	128	133	147	152	150	139	142	134	2.8
<b>SMP</b>	51,098	49,545	49,541	49,164	47,060	48,938	51,061	51,098	51,540	51,251	51,249	51,235	51,153	49,804	51,145	51,283	55,290	66,000	70,000	70,000	64,000	56,413	51,312	51,510	53.7
<b>NORTH PMP</b>	51,098	49,545	49,541	49,164	47,060	48,938	51,061	51,098	51,540	51,251	51,249	51,235	51,153	49,804	51,145	51,283	55,290	66,000	70,000	70,000	64,000	56,413	51,312	51,510	53.7
<b>SOUTH PMP</b>	51,098	49,545	49,541	49,164	47,060	48,938	51,061	51,098	51,540	51,251	51,249	51,235	51,153	49,804	51,145	51,283	55,290	66,000	70,000	70,000	64,000	56,413	51,312	51,510	53.7
<b>MANDATORY HYDRO INJECTIONS</b>	0	0	0	0	0	0	26	114	156	251	152	64	25	81	46	0	0	30	415	741	530	110	85	5	2.8
<b>BORDER ADJUSTMENT SCHEDULES</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>RENEWABLES</b>	343	353	383	384	392	392	392	392	625	1024	1415	1673	1782	1767	1594	1233	829	593	592	612	642	652	673	674	19.4
<b>NORTH BORDER SCHEDULES</b>	-713	-930	-950	-925	-950	-950	-729	-815	-387	-323	-303	-338	-338	-255	-254	-334	-286	158.2	-66.8	-186	-176	-355	-458	-517	-11.3
<b>ITALY SCHEDULES</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>TOTAL SCHEDULES</b>	-713	-930	-950	-925	-950	-950	-729	-815	-387	-323	-303	-338	-338	-255	-254	-334	-286	158.2	-66.8	-186	-176	-355	-458	-517	-11.3
<b>UNIT / HOUR</b>	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	TOTAL
<b>AG_DIMITRIOS1</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>AG_DIMITRIOS2</b>	247	247	247	247	219	247	247	247	247	247	247	247	247	247	247	247	247	247	247	247	247	247	247	247	5.9
<b>AG_DIMITRIOS3</b>	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	6.7
<b>AG_DIMITRIOS4</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<b>AG_DIMITRIOS5</b>	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333	333	7.9
<b>KARDIA1</b>	229	229	229	151	151	151	229	229	229	229	229	229	229	229	229	229	229	229	229	229	229	229	229	229	5.2
<b>KARDIA2</b>	261	261	261	208.5	151	151	261	261	261	261	261	261	261	261	261	261	261	261	261	261	261	261	261	261	5.9
<b>KARDIA3</b>	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	6.7
<b>KARDIA4</b>	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	280	6.7
<b>MEGALOPOL13</b>	230	230	230	230	230	230	230	230	230	230	230	230	230	230	230	230	230	230	230	230	230	230	230	230	5.5
<b>MEGALOPOL14</b>	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	5.0
<b>AMYNDEO1</b>	237	237	237	237	190	181	237	237	237	237	237	237	237	237	237	237	237	237	237	237	237	237	237	237	5.5
<b>AMYNDEO2</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Στιγμιότυπο οθόνης 1-Μορφή δεδομένων ΛΑΓΗΕ-1



Στιγμιότυπο οθόνης 4–Μορφή δεδομένων ΛΑΓΗΕ– 4

Το συγκεκριμένο αρχείο excel διαθέτει φραγή επεξεργασίας και δεν είναι δυνατόν να αντιγραφεί το περιεχόμενό του για τη δημιουργία των μοντέλων μας. Για τον λόγο αυτό δημιουργήθηκε ένα νέο αρχείο excel. Στο αρχείο του ΛΑΓΗΕ παρουσιάζονται τα δεδομένα για κάθε ώρα της ημέρας, ενώ στο νέο αρχείο συγχωνεύθηκαν σε ένα κελί, για κάθε μέρα, όλο το εικοσιτετράωρο. Επίσης, κατηγοριοποιήθηκαν οι μονάδες παραγωγής ενέργειας, ανάλογα με την καύσιμη ύλη και συγχωνεύθηκε σε ένα κελί το άθροισμα εισαγωγών-εξαγωγών. Το αρχείο excel το οποίο αποτελεί και το dataset μας, έχει την ακόλουθη μορφή:

Στιγμιότυπο οθόνης 5–Μορφή Dataset– 1

DAY	DEMAND	SYSTEM LOSSES	SMP	THERMAL	RENEWABLE	HYDRO	NET BORDER	UNIT COAL	UNIT GAS	THERMAL	RENEWABLE	HYDRO	NET BO
25/1/2017	174705.9221		91.09	129235.09	22601.00	17496.14	-5374.00	104703.00	24532.00	73.97%	12.94%	10.01%	-3.08
26/1/2017	180815.1459		102.34	135465.40	23217.00	15002.38	-7835.00	104868.00	30598.00	74.92%	12.84%	8.30%	-4.33
27/1/2017	177549.4299		108.76	137069.60	22418.00	10482.60	-8409.00	107605.00	29464.00	77.20%	12.63%	5.90%	-4.74
28/1/2017	165605.6323		76.95	136626.39	15260.00	6908.83	-7140.00	109997.00	26630.00	82.50%	9.21%	4.17%	-4.31
29/1/2017	160058.9377		80.24	133250.80	16822.00	7308.48	-2678.00	107303.00	25948.00	83.25%	10.51%	4.57%	-1.67
30/1/2017	176121.1455		87.76	139507.26	26289.00	6137.15	-4893.00	108644.00	30864.00	79.21%	14.93%	3.48%	-2.78
31/1/2017	177728.6712		105.84	143235.93	24377.00	9511.09	-1920.00	106390.00	36846.00	80.59%	13.72%	5.35%	-1.08
1/2/2017	171304.9303		94.93	147389.77	14114.00	6825.74	-3555.00	112689.00	34701.00	86.04%	8.24%	3.98%	-2.08
2/2/2017	171042.6783		88.79	143833.07	16609.00	6116.51	-4814.00	109212.00	34621.00	84.09%	9.71%	3.58%	-2.81
3/2/2017	166612.1484		78.04	130286.51	24093.00	6662.56	-6760.00	102892.00	27394.00	78.20%	14.46%	4.00%	-4.06
4/2/2017	149433.0925		50.12	112229.09	27087.00	4641.00	-5476.00	89584.00	22645.00	75.10%	18.13%	3.11%	-3.66
5/2/2017	138339.5603		51.39	97927.02	30872.00	4211.00	-5330.00	85627.00	12300.00	70.79%	22.32%	3.04%	-3.85
6/2/2017	148732.0967		51.48	110293.00	33375.00	4949.47	-115.00	92515.00	17778.00	74.16%	22.44%	3.33%	-0.08
7/2/2017	149952.5581		54.80	120940.21	23228.00	5069.00	-1420.00	95435.00	25505.00	80.65%	15.49%	3.38%	-0.95
8/2/2017	151877.1327		67.28	123347.77	17770.00	6090.93	-4668.00	90233.00	33115.00	81.22%	11.70%	4.01%	-3.07
9/2/2017	155245.8149		60.81	117081.17	22461.00	7160.27	-8918.00	94606.00	22475.00	75.42%	14.47%	4.61%	-5.74
10/2/2017	157666.7052		52.86	112768.98	31492.00	6983.00	-7613.00	94630.00	18139.00	71.52%	19.97%	4.43%	-4.83
11/2/2017	148965.6674		54.22	102307.12	36082.00	5710.75	-5571.00	85860.00	16447.00	68.68%	24.22%	3.83%	-3.74
12/2/2017	147927.4534		49.68	93719.50	34636.00	5265.65	-15246.00	82261.00	11458.00	63.36%	23.41%	3.56%	-10.31
13/2/2017	162249.3448		50.61	115971.51	33042.00	6084.00	-8217.00	92039.00	23932.00	71.48%	20.36%	3.75%	-5.06
14/2/2017	169849.2336		56.75	117209.11	32195.00	6230.76	-15294.00	85134.00	32076.00	69.01%	18.96%	3.67%	-9.00
15/2/2017	161789.8397		56.70	108436.02	23301.00	5904.00	-24964.00	82057.00	26379.00	67.02%	14.40%	3.65%	-15.43
16/2/2017	165718.913		57.19	112250.74	27679.00	6553.00	-20536.00	80323.00	31927.00	67.74%	16.70%	3.95%	-12.39

Στιγμιότυπο οθόνης 6–Μορφή Dataset– 2

## 5.2 Εισαγωγή του dataset στο Jupyter Notebook

Πραγματοποιείται αρχικά εισαγωγή των αλγορίθμων και των δεδομένων που θα χρησιμοποιηθούν στο Jupyter Notebook.

```

In [152]: import numpy as np
from numpy import arange
import matplotlib.pyplot as plt
from pandas import read_csv
from pandas import set_option
from pandas.plotting import scatter_matrix
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.linear_model import ElasticNet
from sklearn.tree import DecisionTreeRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.svm import SVR
from sklearn.pipeline import Pipeline
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.ensemble import ExtraTreesRegressor
from sklearn.ensemble import AdaBoostRegressor
from sklearn.metrics import mean_squared_error
import pandas as pd

```

Στιγμιότυπο οθόνης 7–Εισαγωγή του dataset στο Jupyter Notebook

### 5.3 Προεπεξεργασία του dataset

Πριν από τη χρήση κάθε dataset απαιτείται προεργασία. Συνήθως πρέπει να εξαιρεθούν στήλες ή γραμμές με λανθασμένες ή καθόλου τιμές. Για το σκοπό αυτό πραγματοποιείται σύντομη οπτικοποίηση των δεδομένων με τις εντολές `data.head()` και `data.shape()`. Παρατηρούμε ότι το dataset αποτελείται από 20 στήλες των 365 σειρών

```
In [154]: data=pd.read_excel('LAGHE_2017.xlsx', sheet_name='Φύλλο1')
```

```
In [155]: data.head()
```

```
Out[155]:
```

	DAY	DEMAND	SYSTEM LOSSES	SMP	THERMAL	RENEWABLE	HYDRO	NE BORDE
0	2017-01-01	153013.63	2835.95	49.57	113397.63	16430	3443.00	-20579
1	2017-01-02	156173.44	NaN	53.78	116700.08	19411	8681.79	-11381
2	2017-01-03	162803.98	NaN	59.17	125328.48	20715	16649.28	-111
3	2017-01-04	156143.82	NaN	58.61	124492.49	19318	10294.34	-2039
4	2017-01-05	162458.67	NaN	53.78	130153.61	31450	4859.06	4004

```
In [156]: data.shape
```

```
Out[156]: (365, 20)
```

*Στιγμιότυπο οθόνης 8–Σύντομη οπτικοποίηση των δεδομένων του datasheet*

Με τη χρήση της εντολής

```
data = data.dropna(how = 'any',axis=1)
```

διαγράφονται όλες οι στήλες που έχουν τιμές n/a, προφανώς για να μη ληφθούν υπόψη στον υπολογισμό κελιά χωρίς τιμές.



```
In [54]: data = data.dropna(how = 'any',axis=1)
data.shape
```

```
Out[54]: (368, 14)
```

```
In [55]: data.head()
```

```
Out[55]:
```

	DEMAND	SMP	THERMAL	RENEWABLE	HYDRO	NET BORDER	UNIT COAL	UI G
0	153013.63	49.57	113397.63	16430.0	3443.00	-20579.0	85161.0	2823
1	156173.44	53.78	116700.08	19411.0	8681.79	-11381.0	84033.0	3266
2	162803.98	59.17	125328.48	20715.0	16649.28	-111.0	89869.0	3545
3	156143.82	58.61	124492.49	19318.0	10294.34	-2039.0	90277.0	3421
4	162458.67	53.78	130153.61	31450.0	4859.06	4004.0	94021.0	3613

*Στιγμιότυπο οθόνης 9–Χρήση εντολής data = data.dropna(how = 'any',axis=1)*

Με την εντολή

```
data1=data.drop(data.columns[8:], axis=1)
```

διαγράφονται οι στήλες από την 8<sup>η</sup> έως το τέλος, οι οποίες δεν θα χρησιμοποιηθούν τελικά στο μοντέλο μας.

```
Data1.shape
```

```
data1.head()
```

```
In [56]: data1=data.drop(data.columns[8:], axis=1)
```

```
In [57]: data1.shape
```

```
Out[57]: (368, 8)
```

```
In [58]: data1.head()
```

```
Out[58]:
```

	DEMAND	SMP	THERMAL	RENEWABLE	HYDRO	NET BORDER	UNIT COAL	UI G
0	153013.63	49.57	113397.63	16430.0	3443.00	-20579.0	85161.0	2823
1	156173.44	53.78	116700.08	19411.0	8681.79	-11381.0	84033.0	3266
2	162803.98	59.17	125328.48	20715.0	16649.28	-111.0	89869.0	3545
3	156143.82	58.61	124492.49	19318.0	10294.34	-2039.0	90277.0	3421
4	162458.67	53.78	130153.61	31450.0	4859.06	4004.0	94021.0	3613

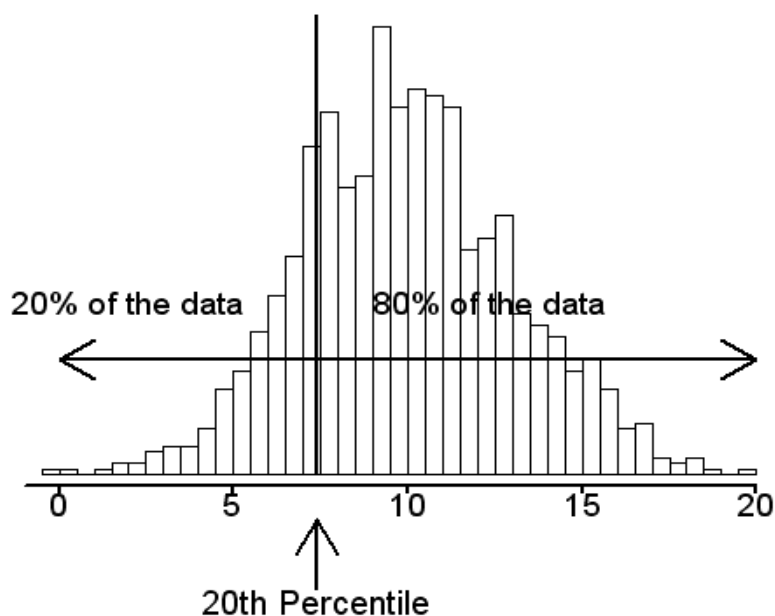
*Στιγμιότυπο οθόνης 10–Χρήση εντολής data1=data.drop(data.columns[8:], axis=1)*

Στη συνέχεια, χρησιμοποιώντας την εντολή

```
data1.describe()
```

προκύπτει πόσες τιμές έχει η κάθε στήλη, ο μέσος όρος τους, η σχετική απόκλιση από τον μέσο όρο τους, η τιμή του μέγιστου και ελάχιστου στοιχείου καθώς οι τιμές που παίρνει η κάθε στήλη στο 25% στο 50% και στο 75% του δείγματος των στοιχείων.

Με τον τρόπο αυτό αποκτάται μια διαισθητική εικόνα των στοιχείων.



Σχήμα 3–Σχηματική επεξήγηση για σημασία ποσοστού 20% του δείγματος. Image: UPRM.edu

### 5.3.1 Συσχέτιση δεδομένων

Επιπλέον στοιχείο το οποίο θα καθορίσει το είδος των μεταβλητών οι οποίες θα χρησιμοποιηθούν στο μοντέλο, είναι ο βαθμός συσχέτισης μεταξύ των τυχαίων μεταβλητών και ο συντελεστής συσχέτισης (correlation coefficient)  $\rho$  που ορίζεται ως εξής:

$$\rho = \frac{\sigma_{xy}}{\sigma_x * \sigma_y}$$

Όπου,

$x, y$  δύο τυχαίες μεταβλητές με διασπορά  $\sigma^2_x$  και  $\sigma^2_y$  και συνδιασπορά  $\sigma_{xy} = \text{Cov}(X, Y)$ .

Ο συντελεστής συσχέτισης  $\rho$ , όπως και η συνδιασπορά  $\sigma_{xy}$ , εκφράζουν το βαθμό που οι δύο μεταβλητές συσχετίζονται, δηλαδή πώς η μία τυχαία μεταβλητή μεταβάλλεται ως προς την άλλη. Η  $\sigma_{xy}$  παίρνει τιμές που εξαρτώνται από το πεδίο τιμών των  $X$  και  $Y$  ενώ ο συντελεστής  $\rho$  παίρνει τιμές στο διάστημα  $[-1, 1]$ .

Οι χαρακτηριστικές τιμές του  $\rho$  ερμηνεύονται ως εξής:

- $\rho = 1$ : υπάρχει τέλεια θετική συσχέτιση μεταξύ των  $X$  και  $Y$
- $\rho = 0$ : δεν υπάρχει καμία (γραμμική) συσχέτιση μεταξύ των  $X$  και  $Y$
- $\rho = -1$ : υπάρχει τέλεια αρνητική συσχέτιση μεταξύ των  $X$  και  $Y$

## 5.4 Οπτικοποίηση δεδομένων

Έχοντας υπόψη μας τα παραπάνω, το επόμενο βήμα είναι η οπτικοποίηση των δεδομένων. Με την οπτικοποίηση γίνεται εύκολη η διάκριση της συσχέτισης μεταξύ μεταβλητών, η κατανομή της κάθε μεταβλητής, των τιμών γύρω από τις οποίες κυμαίνεται η πλειοψηφία των δειγμάτων μας κ.α.

### 5.4.1 Smp timeseries

Αρχικά απεικονίζεται το πώς κυμάνθηκε η οριακή τιμή συστήματος SMP το 2017.

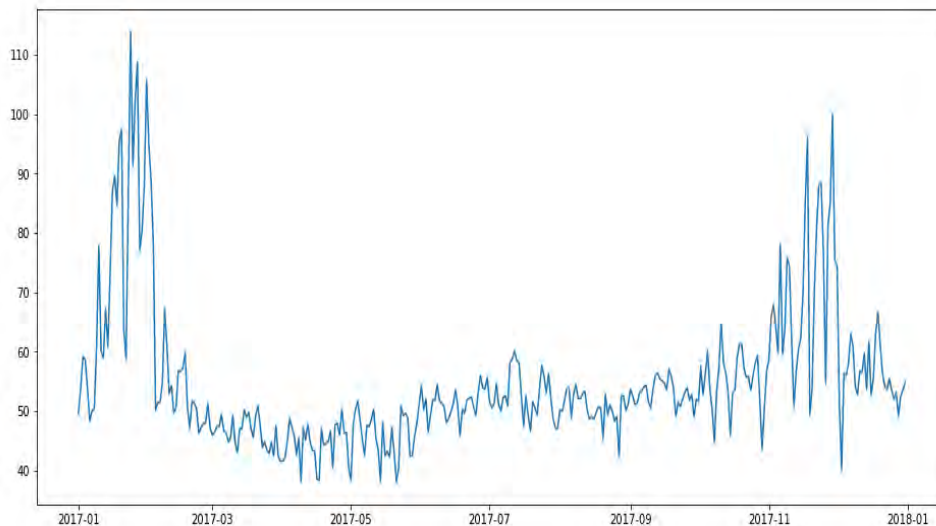
```
In [157]: data.DAY = pd.to_datetime(data.DAY,unit='s')
```

```
In [159]: fig = plt.figure(figsize = (16,7))
ax = plt.axes()

ax.plot(np.array(data.DAY),np.array(data.SMP));

plt.show()
```

Στιγμιότυπο οθόνης 11–Smp timeseries



Σχήμα 4–Οριακή τιμή συστήματος SMP το 2017

Παρατηρούμε ότι η SMP αυξάνεται τον Ιανουάριο και λαμβάνει τη μέγιστη τιμή της στα τέλη Ιανουαρίου. Επίσης, την άνοιξη παρουσιάζει πτωτική πορεία πριν αυξηθεί ξανά κατά τους

καλοκαιρινούς μήνες. Στη συνέχεια, έχουμε μια μικρή πτώση και σταθεροποίησή της το φθινόπωρο πριν αυξηθεί ξανά στα τέλη Νοεμβρίου και τις αρχές Δεκεμβρίου.

Οι διακυμάνσεις της οριακής τιμής που παρατηρείται στο διάγραμμα είναι απολύτως λογική. Η Ελλάδα είναι μεσογειακή χώρα με εύκρατο κλίμα. Επομένως, αναμένεται να υπάρχει μεγαλύτερη κατανάλωση ρεύματος το χειμώνα και το καλοκαίρι, με αντίστοιχη αύξηση της οριακής τιμής συστήματος, ενώ την άνοιξη και το φθινόπωρο, που οι θερμοκρασίες είναι κοντά στους 25 βαθμούς Κελσίου, μικρότερη κατανάλωση με αντίστοιχη μείωση της οριακής τιμής συστήματος.

Από την Εθνική Μετεωρολογική Υπηρεσία (Ε.Μ.Υ.) και το site του μετεωρολογικού σταθμού Αχαρνών <http://www.meteoacharnes.gr/statistika/datasummary.htm> παίρνουμε τη μέση θερμοκρασία για κάθε μέρα της χρονιάς.

Μέσος όρος θερμοκρασίας (°C)												
2017 ▼	Ιαν	Φεβ	Μαρ	Απρ	Μαΐ	Ιουν	Ιουλ	Αυγ	Σεπ	Οκτ	Νοε	Δεκ
1	4,2	5,8	13,8	14,6	21,8	24,1	35,4	27,7	24,0	16,5	13,1	16,9
2	5,8	8,9	12,8	13,7	21,3	25,8	35,9	27,8	26,0	17,1	11,8	17,6
3	8,2	10,5	12,2	13,4	21,4	24,9	31,0	30,1	27,1	17,0	13,5	17,1
4	9,3	11,5	12,6	16,0	21,3	22,9	25,2	31,0	26,9	17,8	13,4	12,7
5	11,4	12,7	13,0	15,3	22,5	23,2	25,1	31,3	24,2	19,4	13,8	9,4
6	9,4	13,1	13,8	15,1	22,0	23,4	26,9	32,5	23,2	20,7	12,9	6,9
7	1,0	10,9	12,5	14,2	21,5	24,1	27,8	32,6	23,8	21,0	15,5	11,0
8	-1,2	10,0	11,1	13,7	20,9	22,9	28,4	32,2	24,6	16,6	13,9	9,7
9	-1,1	8,7	10,5	14,1	20,3	23,7	29,6	31,6	25,3	16,9	14,8	11,6
10	0,7	7,9	11,1	13,4	19,7	23,6	29,7	31,6	25,4	17,3	14,2	13,1
11	5,7	7,0	9,6	14,5	20,0	20,0	30,5	31,4	25,7	17,7	15,9	9,4
12	8,2	5,3	11,0	15,2	20,7	22,1	30,8	31,2	26,2	18,1	17,0	11,4
13	6,7	4,4	11,8	17,3	25,1	24,8	31,7	29,3	26,2	18,9	14,5	12,9
14	9,8	5,0	10,4	17,8	26,3	26,4	30,2	27,7	26,8	19,8	15,5	14,0
15	8,9	6,9	9,9	18,2	23,0	26,6	27,7	27,5	26,0	19,2	17,4	12,7
16	8,6	7,5	9,6	17,5	20,0	24,4	25,8	28,4	26,1	20,4	15,4	13,8
17	8,9	7,7	10,4	15,4	16,6	23,4	22,3	28,7	26,8	20,8	13,5	14,2
18	9,4	10,1	11,0	14,4	15,5	18,8	22,8	29,4	28,5	19,6	13,8	10,0
19	9,6	11,6	13,4	16,5	18,2	20,6	25,3	29,8	27,4	19,2	12,6	7,3
20	6,4	12,6	16,1	16,9	18,8	23,3	26,8	29,5	26,4	19,7	11,8	7,1
21	6,5	10,9	15,9	12,4	19,9	24,7	27,8	26,3	23,2	19,8	8,6	7,1
22	5,4	11,3	15,5	11,4	19,5	26,2	29,0	24,5	21,4	19,7	10,1	5,5
23	6,0	12,1	16,2	13,4	21,0	27,9	30,7	24,0	20,4	18,7	12,0	5,8
24	7,3	12,7	16,2	14,8	20,5	28,9	31,0	24,3	20,5	17,2	12,6	6,4
25	6,3	13,4	16,8	15,7	19,9	29,4	30,8	25,8	21,1	15,8	12,7	10,1
26	4,0	13,3	16,9	17,0	19,3	29,3	29,5	26,3	19,6	14,8	13,6	9,7
27	2,8	12,7	11,9	18,8	18,9	29,2	26,7	26,2	20,1	16,0	12,8	10,2
28	4,5	12,8	10,8	20,1	16,4	29,9	26,2	27,2	19,3	15,2	10,3	12,2
29	5,6	13,4	21,4	19,4	31,2	27,6	28,2	17,6	15,0	10,7	11,1	
30	5,1	15,3	21,5	20,9	33,9	28,2	23,1	17,3	16,0	14,9	9,5	
31	4,5	15,6	21,9	28,1	23,3	14,3	8,4					

Πίνακας 7–Ημερήσια μέση θερμοκρασία έτους 2007 (Ε.Μ.Υ., σταθμός Αχαρνών)

Πράγματι, σύμφωνα με τα παραπάνω στοιχεία, εκείνη τη χρονιά στην Ελλάδα οι θερμοκρασίες ήταν τέτοιες που δικαιολογούν απόλυτα τις διακυμάνσεις αυτές, επαληθεύοντας την ορθότητα των δεδομένων μας.

Έχει πλέον προκύψει μια διαισθητική εικόνα των δεδομένων μας και έχει επαληθευθεί λογικά η ορθότητά τους. Πριν δημιουργηθεί το μοντέλο μας, δημιουργούνται ορισμένα επιπλέον διαγράμματα χρήσιμα για την εξαγωγή τελικών συμπερασμάτων.

## 5.4.2 Histogramms

Χρησιμοποιώντας τη βιβλιοθήκη `seaborn` και τις παρακάτω εντολές δημιουργούμε ιστογράμματα.

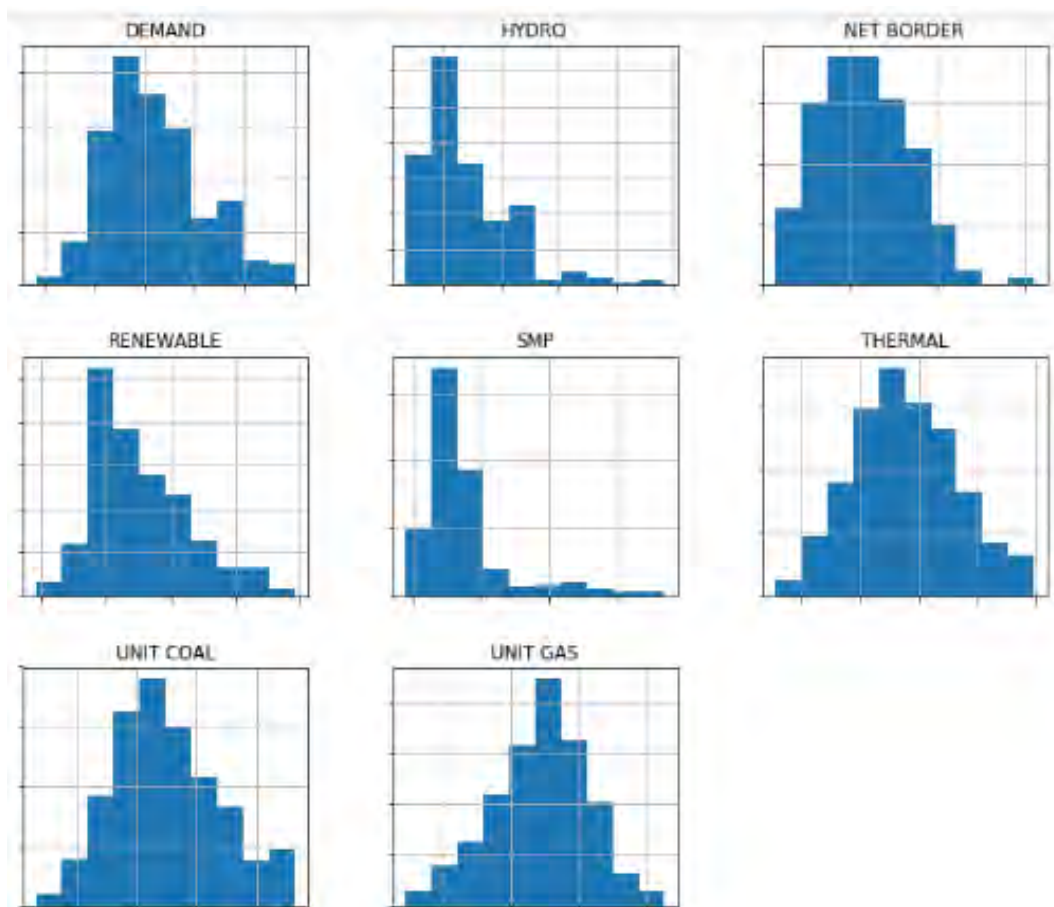
```
In [61]: import seaborn as sns

# histograms
data1.hist(sharex=False, sharey=False, xlabelsize=1, ylabelsize=1, figsize=(12,10))
plt.show()
```

*Στιγμιότυπο οθόνης 12–Χρήση βιβλιοθήκης `seaborn` για τη δημιουργία ιστογραμμάτων.*

Το ιστόγραμμα είναι μια ακριβής γραφική αναπαράσταση της κατανομής αριθμητικών δεδομένων. Είναι μία εκτίμηση της κατανομής της πιθανότητας μιας συνεχούς μεταβλητής (ποσοτικής μεταβλητής). Τα δεδομένα για κάθε μεταβλητή χωρίζονται σε διαστήματα και με το ιστόγραμμα παρατηρείται πόσες τιμές αντιστοιχούν σε κάθε διάστημα. Για το παράδειγμά μας βλέπουμε ότι οι μεταβλητές `Demand`, `NetBorder`, `Renewable`, `Thermal`, `UnitCoal` και `UnitGas` είναι ή μοιάζουν αρκετά με την κατανομή `Gauss`, ενώ οι μεταβλητές `Hydro` και `SMP` είναι κατανομές τύπου `Γάμμα`.

Χωρίς περαιτέρω αναφορά στο είδος των κατανομών, σημειώνεται ότι κατηγοριοποιώντας τις μεταβλητές σε κάποια κατανομή, είναι δυνατόν, χρησιμοποιώντας γνωστά θεωρητικά εργαλεία και τύπους να εξαχθούν χρήσιμα συμπεράσματα.



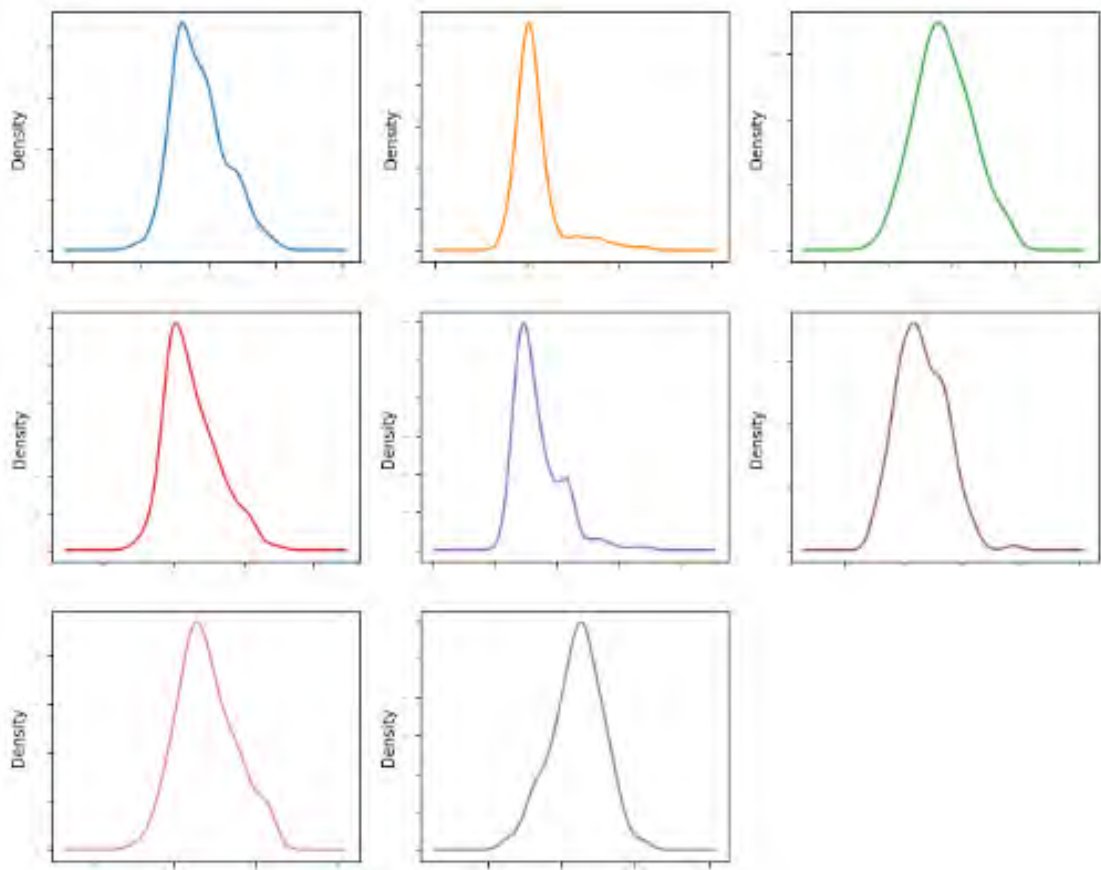
Στιγμιότυπο οθόνης 13–Ιστογράμματα

### 5.4.3 density plots

Ένας ακόμα τρόπος ακριβούς γραφικής αναπαράστασης της κατανομής αριθμητικών δεδομένων είναι τα densityplots. Τα densityplots, σε αντίθεση με τα histograms, μπορούν στο ίδιο διάγραμμα να περιλαμβάνουν ευδιάκριτα περισσότερες από μία μεταβλητές, καθιστώντας ακόμη πιο ευδιάκριτη τη σύγκριση μεταξύ των κατανομών των μεταβλητών.

Χρησιμοποιώντας τη βιβλιοθήκη seaborn και τις παρακάτω εντολές δημιουργούνται τα densityplots των μεταβλητών μας.

```
In [62]: # density
data1.plot(kind='density', subplots=True, layout=(3,3), sharex=False, legend=False, fontsize=1, figsize=(12,10))
plt.show()
```



Στιγμιότυπο οθόνης 14–Χρήση βιβλιοθήκης *seaborn* για τη δημιουργία *densityplots*

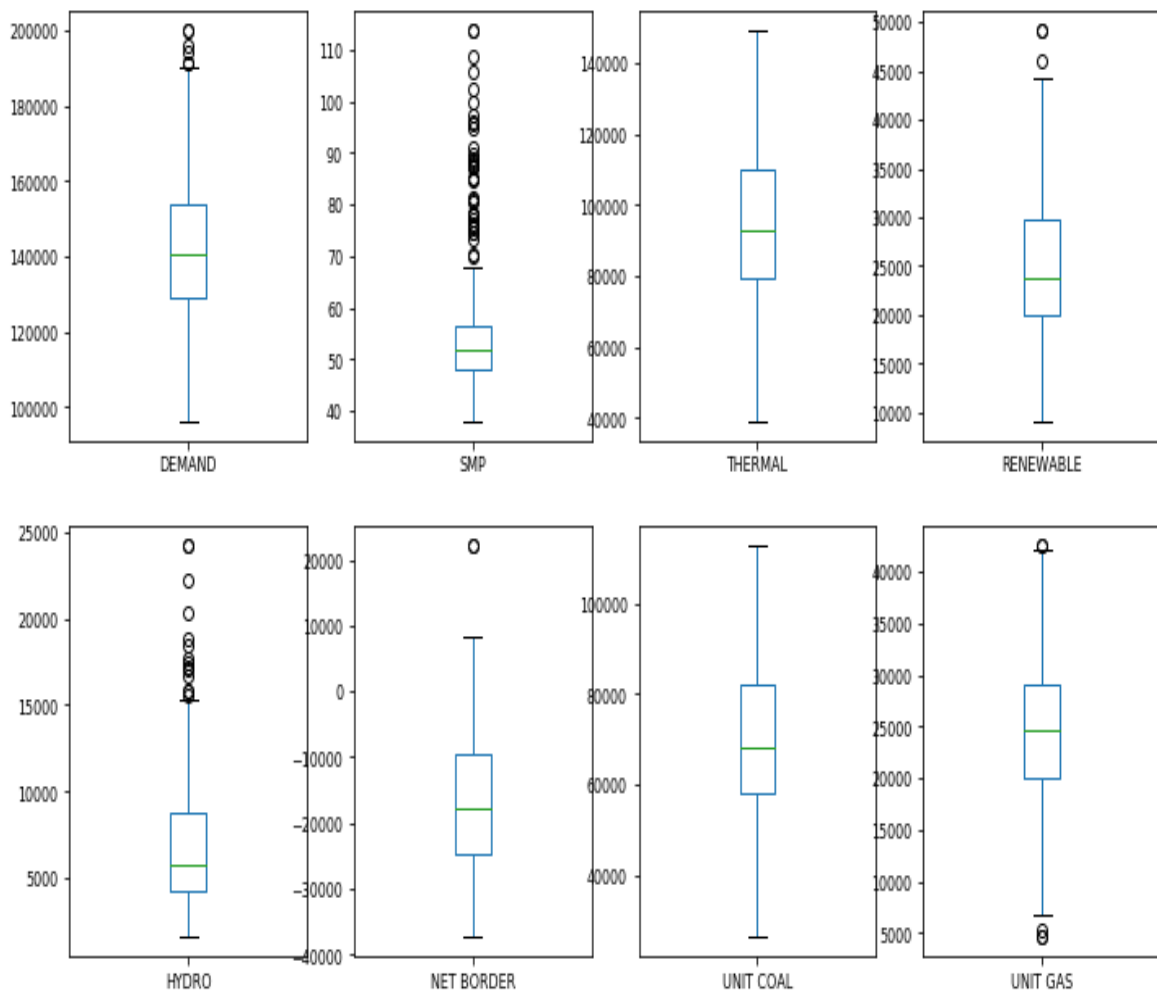
#### 5.4.4 Box and whisker plots

Με τα *box and whisker* διαγράμματα παρατηρείται το εύρος των τιμών της κάθε μεταβλητής. Η γραμμή στο κέντρο υποδεικνύει τη μέση τιμή, ενώ οι κουκίδες έξω από το «κουτί» υποδεικνύουν τις ακραίες τιμές.

Χρησιμοποιώντας την βιβλιοθήκη *matplotlib* και τις παρακάτω εντολές δημιουργούμε τα *box and whiskerplots* των μεταβλητών μας.

```
In [63]: # box and whisker plots
data1.plot(kind='box', subplots=True, layout=(2,4), sharex=False, sharey=False, fontsize=8, figsize = (12,8))
plt.show()
```





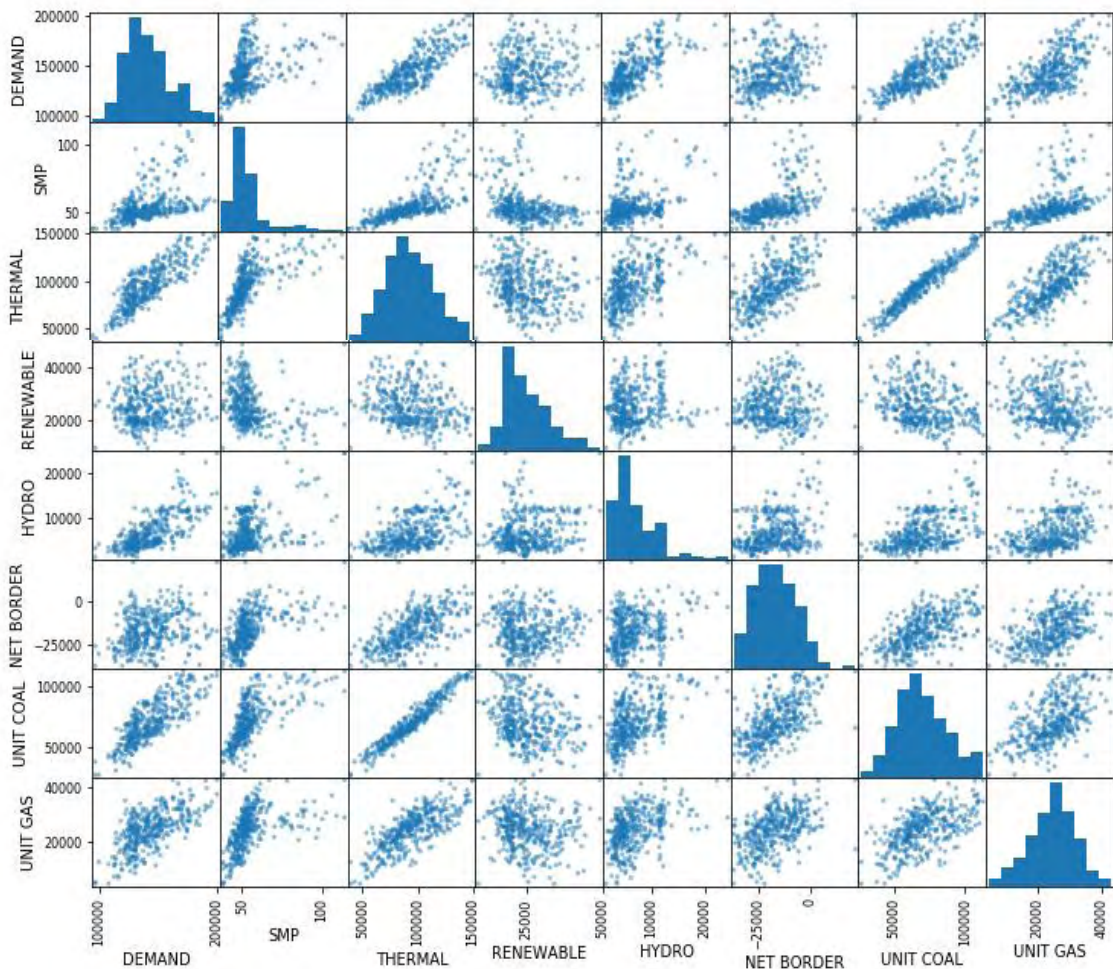
Στιγμιότυπο οθόνης 15 – Χρήση βιβλιοθήκης matplotlib για τη δημιουργία.Box and wiskerplots

### 5.4.5 Scatter plot and correlation matrix

Στην παράγραφο 4.2 αναφέρθηκε η συσχέτιση μεταξύ των μεταβλητών. Εδώ οπτικοποιείται με δύο τρόπους.

Αρχικά δημιουργείται το scatter plot.

```
In [64]: # scatter plot matrix
scatter_matrix(data1, figsize = (12,10))
plt.show()
```



Στιγμιότυπο οθόνης 16–Scatterplot

Το συγκεκριμένο διάγραμμα αποτελεί έναν από τους καλύτερους τρόπους οπτικοποίησης της συσχέτισης μεταξύ των μεταβλητών. Έτσι, μπορεί εύκολα να διακρίνει κανείς αν υπάρχει συσχέτιση και αν είναι θετική ή αρνητική. Ακόμα, στη διαγώνιο παρατηρείται η κατανομή της κάθε μεταβλητής.

Τέλος, πριν αποφασίσουμε ποιες μεταβλητές θα χρησιμοποιηθούν τελικά στο μοντέλο μας, θα δημιουργηθεί ένας πίνακας συσχέτισης (correlationmatrix) των μεταβλητών.

Με τη χρήση των βιβλιοθηκών matplotlib και seaborn και τις παρακάτω εντολές δημιουργούμε τον πίνακα συσχέτισης.

```

In [50]: import numpy as np

# Compute the correlation matrix
corr = data1.corr()

# Generate a mask for the upper triangle
mask = np.zeros_like(corr, dtype=np.bool)
mask[np.triu_indices_from(mask)] = True

# Set up the matplotlib figure
f, ax = plt.subplots(figsize=(11, 9))

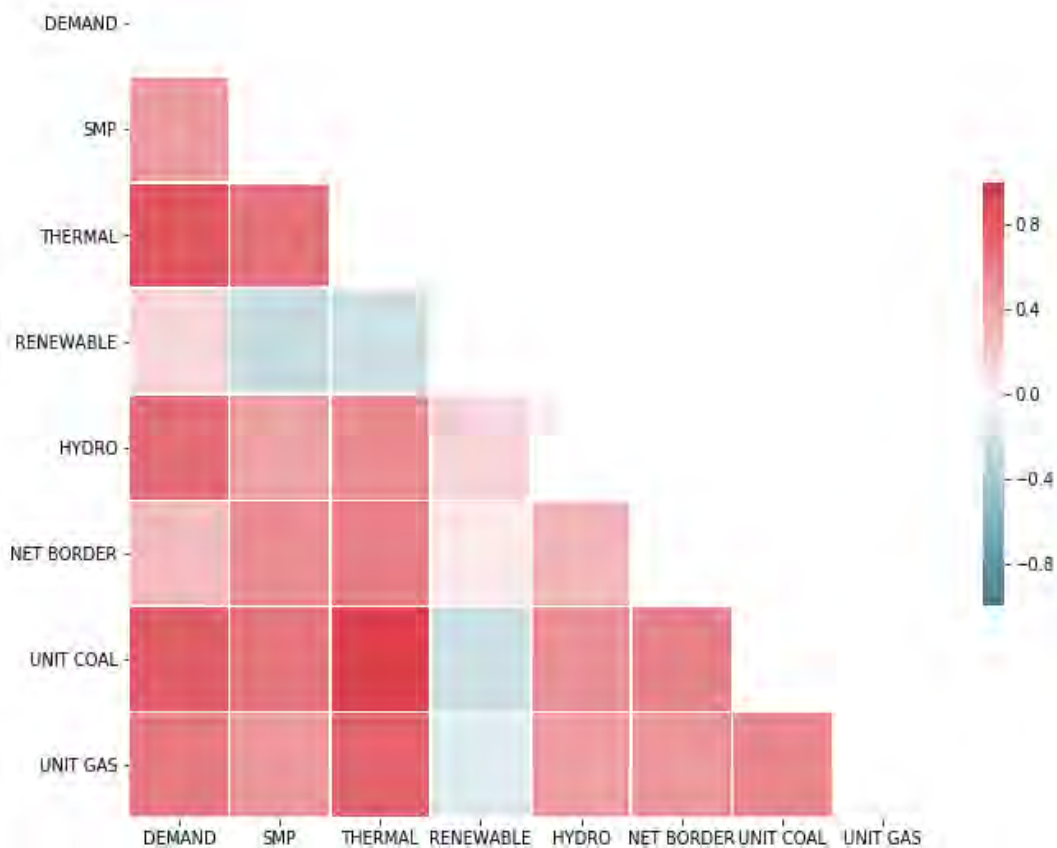
# Generate a custom diverging colormap
cmap = sns.diverging_palette(220, 10, as_cmap=True)

# Draw the heatmap with the mask and correct aspect ratio
sns.heatmap(corr, mask=mask, cmap=cmap, vmin=-1, vmax= 1, center=0,
            square=True, linewidths=.5, cbar_kws={"shrink": .5})

plt.show()

```

Στιγμιότυπο οθόνης 17–Χρήσιμων βιβλιοθηκών matplotlib και seaborn για τη δημιουργία του πίνακα συσχέτισης



Πίνακας 8–Πίνακας συσχέτισης μεταβλητών

Τα τετραγωνίδια με κόκκινο χρώμα δείχνουν θετική συσχέτιση μεταξύ των μεταβλητών, ενώ τα μπλε δείχνουν αρνητική συσχέτιση. Όσο εντονότερο είναι το χρώμα, τόσο πιο έντονη είναι και η συσχέτιση. Παρατηρείται ότι η μεταβλητή THERMAL έχει ισχυρή συσχέτιση με τη μεταβλητή UNITCOAL. Επομένως, από αυτές τις δύο μεταβλητές παίρνουμε σε μεγάλο βαθμό την ίδια πληροφορία και πιθανότατα για το μοντέλο μας αρκεί να χρησιμοποιηθεί η μια από τις δύο.

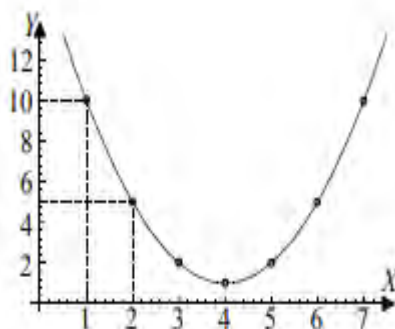
## 5.5 Regression Models (Παλινδρόμηση)

Με την ανάλυση παλινδρόμησης (regression analysis) εξετάζεται η σχέση μεταξύ δύο ή περισσότερων μεταβλητών με σκοπό την πρόβλεψη των τιμών της μιας, μέσω των τιμών της άλλης (ή των άλλων). Σε κάθε πρόβλημα παλινδρόμησης διακρίνουμε δύο είδη μεταβλητών: τις ανεξάρτητες ή ελεγχόμενες ή επεξηγηματικές (independent, predictor, casual, input, explanatory variables) και τις εξαρτημένες ή απόκρισης (dependent, response variables). Σε πειραματικές έρευνες, ανεξάρτητη μεταβλητή  $X$  είναι εκείνη την οποία μπορούμε να ελέγξουμε, δηλαδή, να καθορίσουμε τις τιμές της (π.χ. το ύψος της διαφημιστικής δαπάνης ενός προϊόντος, ο αριθμός των λειτουργούντων ταμείων σε ένα υποκατάστημα τράπεζας, η ποσότητα λιπάσματος, η θερμοκρασία επεξεργασίας ενός προϊόντος, κ.α.). Εξαρτημένη μεταβλητή  $Y$  είναι εκείνη στην οποία αντανακλάται το αποτέλεσμα των μεταβολών στις ανεξάρτητες μεταβλητές (π.χ. η ζήτηση ενός προϊόντος, ο χρόνος αναμονής των πελατών ενός υποκαταστήματος τράπεζας, η απόδοση μιας καλλιέργειας, η αντοχή ενός υλικού, κ.α.). Σε μη πειραματικές έρευνες (δειγματοληψίες) η διάκριση μεταξύ ανεξάρτητων και εξαρτημένων μεταβλητών δεν είναι πάντοτε σαφής, επειδή καμία μεταβλητή δεν είναι ελεγχόμενη, αλλά όλες είναι τυχαίες (π.χ. το ύψος και το βάρος των φοιτητών, οι ώρες μελέτης των φοιτητών ενός πανεπιστημιακού τμήματος και η απόδοση τους σε ένα τεστ, οι εβδομάδες εμπειρίας ενός εργάτη σε μια επιχείρηση και ο αριθμός των ελαττωματικών προϊόντων που παράγει, η κατάταξη δέκα προϊόντων από έναν κριτή και η κατάταξη των ιδίων προϊόντων από έναν άλλο κριτή, ο αριθμός των πωλήσεων μουσικών CD σε μια περιοχή και ο αριθμός των νέων στην ίδια περιοχή).

Ας θεωρήσουμε δύο μεταβλητές  $X$ ,  $Y$ . Αν οι μεταβλητές αυτές συνδέονται με μια σχέση της μορφής  $Y = f(X)$  μέσω της οποίας για κάθε τιμή της  $X$  μπορούμε να προβλέψουμε ακριβώς την τιμή της  $Y$ , δηλαδή, αν οι τιμές της  $Y$  δεν υπόκεινται σε σφάλματα, τότε λέμε ότι οι δύο μεταβλητές συνδέονται με τη συναρτησιακή προσδιοριστική (deterministic) σχέση  $Y =$

$f(X)$ . Για παράδειγμα, το ποσό που καταθέτει κάποιος στην τράπεζα και ο τόκος που παίρνει για το ποσό αυτό, συνδέονται με συναρτησιακή-προσδιοριστική σχέση. Σε αυτές τις περιπτώσεις τα σημεία του διαγράμματος διασποράς βρίσκονται όλα πάνω στην καμπύλη που έχει εξίσωση  $Y = f(X)$  και όσες φορές και αν επαναλάβουμε το πείραμα, θέτοντας το  $X$  στο ίδιο επίπεδο,  $X = x_i$ , θα παίρνουμε πάντα την ίδια τιμή για το  $Y$ . Για παράδειγμα, η εξίσωση  $Y = (X-4)^2 + 1$  (που παριστάνει μια παραβολή) περιγράφει προσδιοριστικά τη σχέση μεταξύ των  $X$  και  $Y$  του παρακάτω πίνακα:

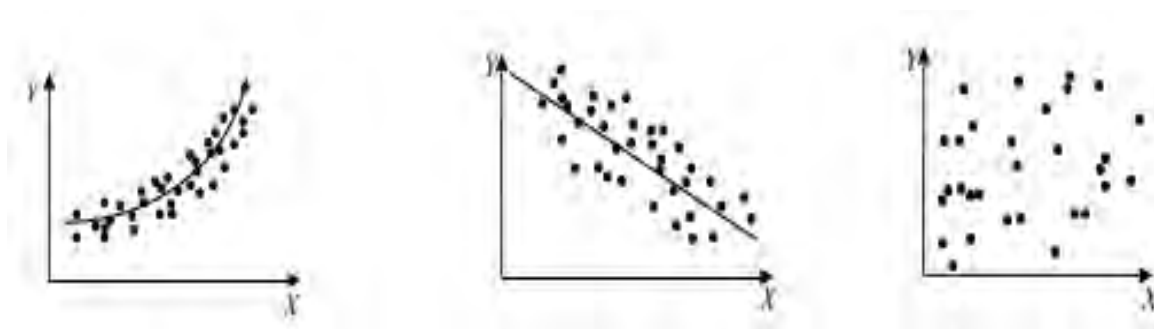
$x_i$	$y_i$
1	10
2	5
3	2
4	1
5	2
6	5
7	10



Οι μη προσδιοριστικές σχέσεις μεταξύ μεταβλητών ονομάζονται στοχαστικές – στατιστικές (stochastic, probabilistic) σχέσεις. Στην περίπτωση αυτή, αν επαναλάβουμε το πείραμα πολλές φορές θέτοντας το  $X$  στο ίδιο επίπεδο  $X = x_i$ , τότε στην τιμή  $x_i$  της  $X$  δεν αντιστοιχεί μια μόνο τιμή  $y_i$  της  $Y$  αλλά, γενικά, αντιστοιχεί ένα πλήθος διαφορετικών τιμών της  $Y$ . Για παράδειγμα, αν  $X$  είναι η τιμή ενός προϊόντος και  $Y$  είναι η ζήτησή του, η  $Y$  βρίσκεται σε στοχαστική σχέση-εξάρτηση από τη  $X$ , γιατί η ζήτηση ενός προϊόντος επηρεάζεται και από άλλους παράγοντες, όπως είναι το ύψος του εισοδήματος των καταναλωτών, οι τιμές ομοειδών προϊόντων, οι καταναλωτικές συνήθειες, κ.ά.

Σε μια στοχαστική σχέση το διάγραμμα διασποράς είναι, γενικά, ένα νέφος σημείων το οποίο πολλές φορές καθορίζει μια ιδεατή γραμμή η οποία δίνει μια πρώτη εικόνα της σχέσης που συνδέει τις δύο μεταβλητές. Η σχέση μάλιστα μεταξύ των δύο μεταβλητών είναι τόσο ισχυρότερη, όσο πιο κοντά στην ιδεατή γραμμή βρίσκονται τα σημεία του διαγράμματος διασποράς. Στο πρώτο από τα παρακάτω σχήματα έχουμε το διάγραμμα διασποράς μιας ισχυρής σχέσης στην οποία όταν αυξάνουν οι τιμές της  $X$ , αυξάνουν γενικά και οι τιμές της  $Y$ , ενώ στο δεύτερο σχήμα έχουμε μια λιγότερο ισχυρή σχέση στην οποία όταν αυξάνουν οι

τιμές της  $X$ , ελαττώνονται γενικά και οι τιμές της  $Y$ . Τέλος, στην περίπτωση του τρίτου σχήματος δε φαίνεται να υπάρχει κάποια σχέση μεταξύ των  $X$  και  $Y$ .



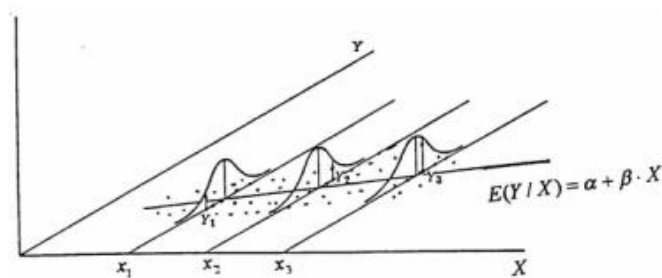
Γενικά, δύο μεταβλητές που συνδέονται είτε με συναρτησιακή-προσδιοριστική σχέση, είτε με στοχαστική σχέση, λέγονται «εξαρτημένες». Αν υπάρχει εξάρτηση μεταξύ δύο μεταβλητών, τότε μπορούμε τη μια από αυτές να τη χαρακτηρίσουμε ως «αιτία» και την άλλη ως «αποτέλεσμα». Αυτό όμως, μόνο στην περίπτωση που η εξάρτηση οφείλεται σε σχέση αιτιότητας των δύο μεταβλητών και όχι σε μια απλή συμμεταβολή, η οποία μπορεί να οφείλεται σε εξάρτηση των δύο μεταβλητών από μια Τρίτη μεταβλητή. Αν, για παράδειγμα,  $X$  είναι το ετήσιο εισόδημα μιας οικογένειας και  $Y, Z$  είναι τα ποσά που ξοδεύει η οικογένεια αυτή σε ένα έτος για κρέας και για αγορά λογοτεχνικών βιβλίων, τότε: αν διαπιστώσουμε σε ένα σύνολο οικογενειών σχέση μεταξύ των  $X$  και  $Y$  (ή μεταξύ των  $X$  και  $Z$ ), δεχόμαστε ότι υπάρχει εξάρτηση μεταξύ των δύο μεταβλητών και τότε μπορούμε να χαρακτηρίσουμε τη  $X$  ως «αιτία» και την  $Y$  (ή τη  $Z$ ) ως «αποτέλεσμα». Αν όμως διαπιστωθεί σχέση μεταξύ των  $Y$  και  $Z$  (που είναι πολύ πιθανό, αφού και οι δύο μεταβάλλονται με το ετήσιο εισόδημα  $X$ ), ασφαλώς θα πρόκειται για «νόθα» εξάρτηση.

Για να περιγράψουμε τη στοχαστική εξάρτηση δύο μεταβλητών  $X$  και  $Y$  προσπαθούμε να βρούμε, όπως και στην προσδιοριστική εξάρτηση, μια σχέση μεταξύ των  $X$  και  $Y$  η οποία όμως τώρα δε θα δίνει ακριβή, αλλά προσεγγιστική μόνο εικόνα της εξάρτησης των  $X$  και  $Y$  και τα σημεία του διαγράμματος διασποράς των  $X$  και  $Y$  δε θα βρίσκονται πάνω, αλλά, γύρω από μια καμπύλη. Μια μέθοδος που χρησιμοποιείται για την περιγραφή της στοχαστικής εξάρτησης δύο μεταβλητών είναι η μέθοδος των ελαχίστων τετραγώνων και αυτή θα εφαρμόσουμε στη συνέχεια για να μελετήσουμε την πιο απλή μορφή στοχαστικής εξάρτησης, τη γραμμική.

### 5.5.1 Απλή Γραμμική Παλινδρόμηση

Αν το διάγραμμα διασποράς δύο μεταβλητών  $X$  και  $Y$  έχει μορφή επιμήκους κεκλιμένης έλλειψης ή πλατυσμένου  $J$ , η σχέση των  $X$  και  $Y$  είναι κατά προσέγγιση γραμμική. Στην περίπτωση αυτή έχουμε την απλούστερη μορφή παλινδρόμησης, την απλή γραμμική παλινδρόμηση, όπου υπάρχει μόνο μια ανεξάρτητη μεταβλητή  $X$  και η εξαρτημένη μεταβλητή  $Y$  μπορεί να προσεγγισθεί ικανοποιητικά από μια γραμμική συνάρτηση του  $X$ .

Η γραμμική σχέση  $Y = \alpha + \beta \cdot X$  δε μπορεί, ασφαλώς, να περιγράψει τη γραμμική στοχαστική εξάρτηση των μεταβλητών  $X$  και  $Y$  αφού αν, για παράδειγμα,  $X$  είναι η τιμή ενός προϊόντος και  $Y$  είναι η ζήτηση του προϊόντος αυτού και διατηρήσουμε τη  $X$  στο ίδιο επίπεδο  $X = x_1$ , τότε οι αντίστοιχες τιμές του  $Y$  θα είναι φυσικά διαφορετικές στις διάφορες επαναλήψεις. Επίσης, αν  $X$  είναι η ποσότητα λιπάσματος και  $Y$  είναι η απόδοση μιας καλλιέργειας και διατηρήσουμε τη  $X$  στο ίδιο επίπεδο  $X = x_1$ , τότε οι αντίστοιχες τιμές του  $Y$  θα είναι φυσικά διαφορετικές στις διάφορες επαναλήψεις, αφού παράγοντες, όπως η θερμοκρασία, οι βροχοπτώσεις, η ποιότητα του εδάφους, θα επηρεάζουν, επίσης, την παραγωγή. Επιπλέον, συμβαίνει να παρατηρούνται και σφάλματα μέτρησης των τιμών της  $Y$  (λόγω οργάνων ή ελλιπούς πληροφόρησης). Έτσι, για  $X = x_1$  το αντίστοιχο  $Y$  είναι μια τυχαία μεταβλητή  $Y_1$  που ακολουθεί κάποια κατανομή. Ομοίως, για  $X = x_2$  θα έχουμε κάποια άλλη κατανομή  $Y_2$  κ.ο.κ..



Επομένως, στην εξίσωση  $Y = \alpha + \beta \cdot X$ , πρέπει να προσθέσουμε έναν ακόμη όρο  $\varepsilon$  ο οποίος, για δεδομένη τιμή της  $X$ , να περιγράφει τη διαφορά της παρατηρούμενης από τη θεωρητική  $(\alpha + \beta \cdot X)$  τιμή της  $Y$ . Δηλαδή,  $\varepsilon = Y - (\alpha + \beta \cdot X)$ . Προκύπτει, επομένως, το στοχαστικό μοντέλο

$$Y = \alpha + \beta \cdot X + \varepsilon.$$

Για λόγους απλούστευσης των υπολογισμών και εφικτότητας λύσης του προβλήματος, κάνουμε κάποιες υποθέσεις, όπως  $E(\varepsilon) = 0$  και  $E(Y|X) = \alpha + \beta \cdot X$ . Δηλαδή, υποθέτουμε ότι τα σφάλματα έχουν μέση τιμή μηδέν και ότι για τις διάφορες τιμές

της  $X$ , οι αντίστοιχες μέσες τιμές της  $Y$  βρίσκονται πάνω σε μια ευθεία. Η ευθεία αυτή ( $E(Y/X) = \alpha + \beta \cdot X$ ), ονομάζεται πληθυσμιακή ευθεία παλινδρόμησης.

Με τη μέθοδο των ελαχίστων τετραγώνων θα προσδιορίσουμε στη συνέχεια μια εκτίμηση

$\hat{Y} = \hat{\alpha} + \hat{\beta} \cdot X$  της ευθείας  $E(Y/X) = \alpha + \beta \cdot X$  όπου  $\hat{\alpha}$  και  $\hat{\beta}$  εκτιμήτριες των  $\alpha$  και  $\beta$  αντίστοιχα. Η εκτίμηση  $\hat{Y} = \hat{\alpha} + \hat{\beta} \cdot X$  της πληθυσμιακής ευθείας παλινδρόμησης  $E(Y/X) = \alpha + \beta \cdot X$ , ονομάζεται ευθεία ελαχίστων τετραγώνων από τη μέθοδο υπολογισμού των παραμέτρων της.

### Μέθοδος ελαχίστων τετραγώνων

Θεωρούμε  $n$  ζεύγη παρατηρήσεων  $(x_i, y_i)$ ,  $i = 1, 2, 3, \dots, n$ . Αναζητούμε προσέγγιση της μορφής:  $y_i = \alpha + \beta \cdot x_i + \varepsilon_i$ , όπου τα  $\varepsilon_i$  παριστάνουν τις αποκλίσεις της πραγματικής τιμής  $y_i$  από την προσαρμοσμένη (θεωρητική)  $\alpha + \beta \cdot x_i$ . Δηλαδή,  $\varepsilon_i = y_i - (\alpha + \beta \cdot x_i)$ .

Είναι φανερό, ότι η εκλογή (εκτίμηση) των  $\alpha$  και  $\beta$  θα πρέπει να γίνει έτσι, ώστε να ελαχιστοποιηθούν οι ποσότητες  $\varepsilon_i$ . Για το σκοπό αυτό, θα αναζητήσουμε τις τιμές των  $\alpha$  και  $\beta$  για τις οποίες ελαχιστοποιείται το άθροισμα των τετραγώνων των  $\varepsilon_i$ . Δηλαδή, η ποσότητα

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \alpha - \beta \cdot x_i)^2 \quad (1)$$

(Η ελαχιστοποίηση του αθροίσματος  $\sum \varepsilon_i$  δεν αποτελεί ασφαλές κριτήριο επιλογής, διότι κάποια αρνητικά  $\varepsilon_i$  θα αναιρούν αντίστοιχες θετικές ποσότητες του αθροίσματος). Παραγωγίζοντας την (1) ως προς  $\alpha$  και  $\beta$  και εξισώνοντας με μηδέν παίρνουμε τις ακόλουθες δύο εξισώσεις που ονομάζονται κανονικές εξισώσεις:

$$\begin{aligned} \sum_{i=1}^n y_i &= n \cdot \alpha + \beta \cdot \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i &= \alpha \cdot \sum_{i=1}^n x_i + \beta \cdot \sum_{i=1}^n x_i^2 \end{aligned}$$

Λύνοντας το σύστημα των κανονικών εξισώσεων, παίρνουμε:

$$\hat{\beta} = \frac{n \cdot \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \cdot \left( \sum_{i=1}^n y_i \right)}{n \cdot \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$$

ή

$$\boxed{\hat{\beta} = \frac{s_{xy}}{s_x^2} \text{ και } \hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}}$$



**Η εκτίμηση ελαχίστων τετράγωνων  $\hat{Y} = \hat{\alpha} + \hat{\beta} \cdot X$  της ευθείας παλινδρόμησης από το δείγμα των  $n$  ζευγών παρατηρήσεων είναι, επομένως, η**

$$\hat{Y} = \hat{\alpha} + \hat{\beta} \cdot X = \bar{y} - \hat{\beta} \cdot \bar{x} + \hat{\beta} \cdot X = \bar{y} + \hat{\beta} \cdot (X - \bar{x})$$

ή

$$\hat{Y} = \bar{y} + \frac{s_{xy}}{s_x^2} \cdot (X - \bar{x})$$

Προφανώς, η ευθεία ελαχίστων τετραγώνων, διέρχεται από το σημείο  $(\bar{x} + \bar{y})$ . Επισημαίνουμε ότι πρέπει να γίνεται διάκριση μεταξύ της παρατηρούμενης τιμής του  $Y$  και της  $\hat{Y}$  που εκτιμάμε. Η παρατηρούμενη τιμή  $y_i$  είναι η πραγματική τιμή της  $Y$ , ενώ η τιμή  $\hat{y}_i$  της  $\hat{Y}$ , είναι εκτίμηση της μέσης τιμής  $E(Y/X = x_i)$ .

### 5.5.2 Πολλαπλή γραμμική παλινδρόμηση

Στην παρούσα εργασία, όπως και συχνά στην πράξη συμβαίνει, το μοντέλο της απλής γραμμικής παλινδρόμησης είναι ανεπαρκές για την περιγραφή της μεταβλητότητας που υπάρχει στην εξαρτημένη μεταβλητή  $Y$  με συνέπεια οι προβλέψεις να έχουν μεγάλη ανακρίβεια. Στην περίπτωση αυτή, χρειάζεται, ενδεχομένως, να λάβουμε υπ' όψη μας  $k \geq 2$  ανεξάρτητες μεταβλητές για την πρόβλεψη της εξαρτημένης μεταβλητής. Το γενικό γραμμικό μοντέλο με  $k$  ανεξάρτητες μεταβλητές  $X_1, X_2, \dots, X_k$ , λέγεται πολλαπλή γραμμική παλινδρόμηση και έχει την μορφή

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

όπου  $\beta_0, \beta_1, \dots, \beta_k$  είναι οι (άγνωστες) παράμετροι παλινδρόμησης και  $\varepsilon$  είναι το τυχαίο σφάλμα. Αυτό αντιστοιχεί στη μεταβλητότητα του  $Y$  που δεν μπορεί να περιγράψει η πολλαπλή παλινδρόμηση, για το οποίο υποθέτουμε ότι προέρχεται από την κανονική κατανομή με μέση τιμή 0 και (άγνωστη) διακύμανση  $\sigma^2$ .

Κατά την ανάλυση της πολλαπλής παλινδρόμησης είναι σημαντικό να γνωρίζουμε τόσο τη μορφή, όσο και το βαθμό της σχέσης που υπάρχει σε κάθε ζεύγος μεταξύ των μεταβλητών. Επιθυμητή είναι η ύπαρξη συσχέτισης μεταξύ των  $Y$  και  $X_i$ , αλλά όχι και η ύπαρξη συσχέτισης μεταξύ των  $X_i$ ,  $i = 1, 2, \dots, k$ , η οποία είναι γνωστή σαν πολλαπλή συγγραμμικότητα.

### 5.5.3 Κριτήρια καταλληλότερου μοντέλου

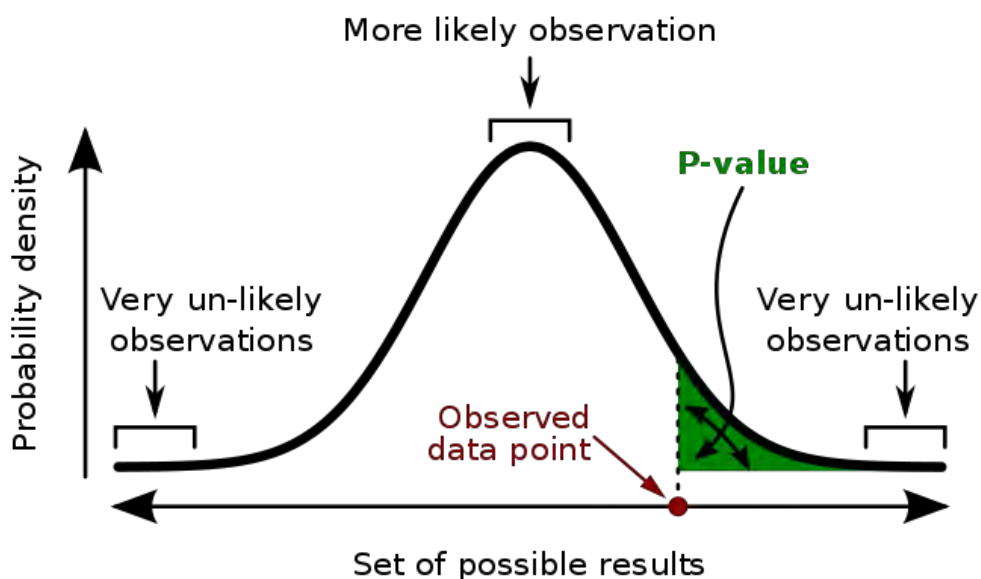
#### Στατιστική Σημαντικότητα (p-values)

Important:

**$\Pr(\text{observation} \mid \text{hypothesis}) \neq \Pr(\text{hypothesis} \mid \text{observation})$**

The probability of observing a result given that some hypothesis is true is *not equivalent* to the probability that a hypothesis is true given that some result has been observed.

Using the p-value as a “score” is committing an egregious logical error:  
**the transposed conditional fallacy.**



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

Η στατιστική σημαντικότητα (ή ένα στατιστικά σημαντικό αποτέλεσμα) είναι η πιθανότητα για ένα στατιστικό μοντέλο όπου, όταν η μηδενική υπόθεση είναι αληθής, η στατιστική σύνοψη (όπως η διαφορά των μέσων τιμών δύο συγκρινόμενων δειγμάτων) θα ήταν ίση ή μεγαλύτερη με τα ακριβή παρατηρούμενα αποτελέσματα. Η στατιστική σημαντικότητα είναι θεμελιώδης για κάθε σημαντικό πείραμα ή παρατήρηση που περιλαμβάνει τη συλλογή ενός δείγματος από έναν πληθυσμό καθώς υπάρχει η πιθανότητα ότι ένα παρατηρούμενο αποτέλεσμα συμβαίνει μόνο λόγω σφάλματος δειγματοληψίας και μόνο. Αν όμως η τιμή

είναι μικρότερη από ( $p < 0,05$ ) τότε συμπεραίνεται ότι η παρατηρούμενη επίδραση του πειράματος αντανακλά στην πραγματικότητα τα χαρακτηριστικά του πληθυσμού και όχι μόνο ένα δειγματοληπτικό σφάλμα.

### **R- squared ή coefficient of determination**

Είναι ένα είδος στατιστικής μέτρησης που χρησιμοποιείται για να εκτιμηθεί το πόσο καλά προσεγγίζει τα δεδομένα το μοντέλο της γραμμικής παλινδρόμησης που έχει κατασκευαστεί.

Στο R-squared έχουμε ένα βασικό μοντέλο, το οποίο είναι το χειρότερο και με αυτό συγκρίνουμε κάθε ένα που κατασκευάζουμε. Το χειρότερο μοντέλο δεν χρησιμοποιεί καθόλου τις ανεξάρτητες μεταβλητές  $X$  για να προβλέψει την τιμή της εξαρτημένης μεταβλητής  $Y$ . Αντ' αυτού, χρησιμοποιεί το μέσο όρο των παρατηρούμενων τιμών της  $Y$  και πάντα προβλέπει από αυτό το μέσο όρο την εκτιμώμενη τιμή του  $Y$ . Κάθε μοντέλο παλινδρόμησης που κατασκευάζεται συγκρίνεται με αυτό το βασικό μοντέλο για να γίνει κατανοητό το πόσο καλά προσεγγίζει αυτό το μοντέλο τα δεδομένα, πόσο καλύτερο είναι δηλαδή από το χειρότερο.

Οι τιμές του R-squared είναι από μηδέν μέχρι ένα, όπου το μηδέν παριστά μία ελλειπή γραμμή παλινδρόμησης ενώ το ένα μία τέλεια γραμμή παλινδρόμησης.

Το κυριότερο πρόβλημα του R-squared είναι πως μπορεί τεχνητά να αυξηθεί η τιμή του απλώς προσθέτοντας περισσότερες ανεξάρτητες μεταβλητές  $X$ . Για το λόγο αυτό χρησιμοποιούμε ως είδος μέτρησης της καταλληλότητας του μοντέλου μας το adjusted R-squared.

### **Μαθηματική αναπαράσταση του R-squared**

$$R^2 = 1 - \frac{SSE}{SST}$$

Όπου το SSE είναι το άθροισμα των ελαχίστων τετραγώνων του μοντέλου μας

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

και SST το άθροισμα των ελαχίστων τετραγώνων του βασικού, χειρότερου μοντέλου.

$$SST = \sum_{i=1}^n (y_i - \bar{y}_i)^2.$$

### **AdjustedR-squared.**

Το adjustedR-squared είναι παρόμοιο με το R-squared, αλλά μειώνεται όταν προσθέτουμε ανεξάρτητες μεταβλητές που δεν προσφέρουν ιδιαίτερη πληροφορία στο μοντέλο μας, ενώ αυξάνεται σε αντίθετη περίπτωση.

Επομένως, παρατηρώντας τις p-values θα απορριφθούν μεταβλητές που μπορεί να έχουν κάποιο σφάλμα δειγματοληψίας. Με το R-squared θα γίνει κατανοητό το πόσο καλό είναι το κάθε μοντέλο, ενώ με το adjustedR-squared θα γίνει εμφανές το πόσο σημαντικές είναι αυτές οι ανεξάρτητες μεταβλητές για το κάθε μοντέλο.

### **Μαθηματική αναπαράσταση του adjustedR-squared**

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}$$

n=no of data points

p=no of independent variables used in the model

## 5.6 Τα μοντέλα παλινδρόμησης

### Μοντέλο πρώτο

Στο πρώτο μας μοντέλο επιλέγονται όλες οι ανεξάρτητες μεταβλητές που έχουν παραμείνει μετά την αρχική εκκαθάριση και έχουμε:

```
import statsmodels.formula.api as smf

results = smf.ols('SMP ~ DEMAND + THERMAL + RENEWABLE + HYDRO + NET_BORDER + UNIT_COAL + UNIT_GAS', data=data1).fit()

print (results.summary())
```

```

=====
                    OLS Regression Results
=====
Dep. Variable:          SMP      R-squared:                0.564
Model:                  OLS      Adj. R-squared:           0.555
Method:                 Least Squares   F-statistic:              66.51
Date:                   Mon, 10 Sep 2018   Prob (F-statistic):       4.26e-61
Time:                   19:39:37         Log-Likelihood:           -1297.7
No. Observations:      368             AIC:                      2611.
Df Residuals:          360             BIC:                      2643.
Df Model:               7
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	51.0186	4.402	11.591	0.000	42.362	59.675
DEMAND	-0.0006	0.000	-3.294	0.001	-0.001	-0.000
THERMAL	-0.0012	0.001	-1.284	0.200	-0.003	0.001
RENEWABLE	0.0002	0.000	1.003	0.317	-0.000	0.001
HYDRO	0.0012	0.000	4.867	0.000	0.001	0.002
NET_BORDER	-0.0003	0.000	-1.399	0.163	-0.001	0.000
UNIT_COAL	0.0020	0.001	2.288	0.023	0.000	0.004
UNIT_GAS	0.0019	0.001	2.095	0.037	0.000	0.004

```
=====
Omnibus:                157.579   Durbin-Watson:           0.564
Prob(Omnibus):          0.000     Jarque-Bera (JB):        647.112
Skew:                   1.873     Prob(JB):                3.03e-141
Kurtosis:               8.307     Cond. No.                1.95e+06
=====
```

Παρατηρείται πως το R-squared και το adjust R-squared έχουν κοντινές τιμές. Το μοντέλο μας κάνει fit κατά 56.4% και οι μεταβλητές που έχουν επιλεγεί είναι αντιπροσωπευτικές κατά 55%. Παρατηρείται επίσης πως οι p-values της μεταβλητής RENEWABLE και της μεταβλητής THERMAL είναι μεγαλύτερες από 0.05. Επομένως στο επόμενο μοντέλο θα αφαιρεθούν.

## Μοντέλο δεύτερο

```
import statsmodels.formula.api as smf

results = smf.ols('SMP ~ DEMAND + HYDRO + UNIT_COAL + UNIT_GAS', data=data1).fit()

print (results.summary())
```

```
=====
                        OLS Regression Results
=====
Dep. Variable:          SMP      R-squared:                0.558
Model:                  OLS      Adj. R-squared:           0.553
Method:                 Least Squares   F-statistic:             114.4
Date:                   Mon, 10 Sep 2018   Prob (F-statistic):      5.43e-63
Time:                   19:46:51         Log-Likelihood:          -1300.4
No. Observations:      368             AIC:                     2611.
Df Residuals:          363             BIC:                     2630.
Df Model:               4
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	50.2592	4.402	11.418	0.000	41.603	58.916
DEMAND	-0.0004	5.02e-05	-7.783	0.000	-0.000	-0.000
HYDRO	0.0010	0.000	5.939	0.000	0.001	0.001
UNIT_COAL	0.0006	4.24e-05	14.397	0.000	0.001	0.001
UNIT_GAS	0.0004	8.29e-05	5.363	0.000	0.000	0.001

```
=====
Omnibus:                157.003   Durbin-Watson:           0.580
Prob(Omnibus):          0.000   Jarque-Bera (JB):        652.543
Skew:                   1.859   Prob(JB):                 2.00e-142
Kurtosis:               8.360   Cond. No.                  1.65e+06
=====
```

Παρατηρείται πως το R-squared και το adjustR-squared έχουν σχεδόν την ίδια τιμή. Το μοντέλο μας κάνει fit κατά 55,8% και οι μεταβλητές που έχουν επιλεγεί είναι αντιπροσωπευτικές κατά 55,3%. Παρατηρείται επίσης πως όλες οι p-values των μεταβλητών είναι μικρότερες από 0.05.

Μετά από πειράματα που έγιναν και με άλλες μεθόδους και αλγορίθμους πολλαπλής ή γραμμικής παλινδρόμησης όπως KNN, Lasso, LR, DecisionTreeRegressor κ.α δεν κατέστη δυνατό να αυξηθεί αυτό το ποσοστό. Η καλύτερη μέθοδος αποδείχτηκε η χρησιμοποίηση του statmodels. Για το λόγο αυτό το επόμενο βήμα είναι η δημιουργία μοντέλων με τις ίδιες ανεξάρτητες μεταβλητές όπως στα παραπάνω μοντέλα των οποίων όπως οι τιμές θα υποστούν κανονικοποίηση ώστε να αποκλειστεί το ενδεχόμενο τα διαφορετικά εύρη τιμών της κάθε μεταβλητής να επηρεάζουν το τελικό fit του μοντέλου.

## Κανονικοποίηση Δεδομένων

Επειδή έχουμε και αρνητικές τιμές δεν κανονικοποιούνται τα δεδομένα στο διάστημα (0, 1) αλλά στο (-1, 1). Χρησιμοποιώντας τον MinMaxScaler της sklearn έχουμε:

```
#normalize the data in (-1,1)

from sklearn import preprocessing

scaler = preprocessing.MinMaxScaler(feature_range=(-1, 1))
scaled_df = scaler.fit_transform(data1)

scaled_df = pd.DataFrame(scaled_df)
scaled_df.head()
```

	0	1	2	3	4	5	6	7
0	0.09	-0.70	0.35	-0.64	-0.84	-0.44	0.36	0.24
1	0.15	-0.58	0.41	-0.49	-0.38	-0.13	0.33	0.48
2	0.28	-0.44	0.57	-0.42	0.33	0.25	0.47	0.63
3	0.15	-0.46	0.55	-0.49	-0.23	0.18	0.48	0.56
4	0.28	-0.58	0.65	0.11	-0.72	0.39	0.57	0.66

Μετονομάζεται ο πίνακας με τα αρχικά ονόματα των ανεξάρτητων μεταβλητών και γίνεται ως εξής:

```
scaled_df = scaled_df.rename(index =int, columns = {0:"DEMAND",1:"SMP",2:"THERMAL",
3:"RENEWABLE",4:"HYDRO",5:"NET_BORDER",6:"UNIT_COAL",7:"UNIT_GAS"})

scaled_df.head()
```

	DEMAND	SMP	THERMAL	RENEWABLE	HYDRO	NET_BORDER	UNIT_COAL	UNIT_GAS
0	0.09	-0.70	0.35	-0.64	-0.84	-0.44	0.36	0.24
1	0.15	-0.58	0.41	-0.49	-0.38	-0.13	0.33	0.48
2	0.28	-0.44	0.57	-0.42	0.33	0.25	0.47	0.63
3	0.15	-0.46	0.55	-0.49	-0.23	0.18	0.48	0.56
4	0.28	-0.58	0.65	0.11	-0.72	0.39	0.57	0.66

```
pd.DataFrame(scaler.inverse_transform(scaled_df)).head()
```

	0	1	2	3	4	5	6	7
0	153013.63	49.57	113397.63	16430.0	3443.00	-20579.0	85161.0	28236.0
1	156173.44	53.78	116700.08	19411.0	8681.79	-11381.0	84033.0	32667.0
2	162803.98	59.17	125328.48	20715.0	16649.28	-111.0	89869.0	35459.0
3	156143.82	58.61	124492.49	19318.0	10294.34	-2039.0	90277.0	34215.0
4	162458.67	53.78	130153.61	31450.0	4859.06	4004.0	94021.0	36133.0

Ενώ γίνεται και έλεγχος για το αν έγινε σωστά η κανονικοποίηση, κάνοντας αντίστροφη κανονικοποίηση και ελέγχοντας αν οι τιμές αυτού του πίνακα είναι ίδιες με το αρχικό μας dataset.

## Μοντέλο τρίτο

```
results = smf.ols('SMP ~ DEMAND + THERMAL + RENEWABLE + HYDRO + NET_BORDER + UNIT_COAL + UNIT_GAS', data=scaled_df).fit()
print (results.summary())
```

```
=====
                        OLS Regression Results
=====
Dep. Variable:          SMP      R-squared:                0.564
Model:                 OLS      Adj. R-squared:           0.555
Method:                Least Squares      F-statistic:              66.51
Date:                  Mon, 10 Sep 2018    Prob (F-statistic):       4.26e-61
Time:                  20:02:24          Log-Likelihood:           39.669
No. Observations:     368              AIC:                     -63.34
Df Residuals:          360              BIC:                     -32.07
Df Model:              7
Covariance Type:      nonrobust
=====
                    coef    std err          t      P>|t|      [0.025    0.975]
-----
Intercept          -0.5324      0.041    -13.099     0.000     -0.612     -0.452
DEMAND             -0.8400      0.255     -3.294     0.001     -1.341     -0.338
THERMAL            -1.7463      1.360     -1.284     0.200     -4.420     0.927
RENEWABLE           0.1002      0.100     1.003     0.317     -0.096     0.297
HYDRO               0.3706      0.076     4.867     0.000     0.221     0.520
NET_BORDER         -0.1970      0.141     -1.399     0.163     -0.474     0.080
UNIT_COAL           2.3119      1.010     2.288     0.023     0.325     4.299
UNIT_GAS            0.9293      0.444     2.095     0.037     0.057     1.802
=====
Omnibus:              157.579    Durbin-Watson:           0.564
Prob(Omnibus):        0.000    Jarque-Bera (JB):        647.112
Skew:                 1.873    Prob(JB):                 3.03e-141
Kurtosis:              8.307    Cond. No.                 187.
=====
```

Παρατηρείται πως το R-squared και το adjustR-squared έχουν κοντινές τιμές. Το μοντέλο μας κάνει fit κατά 56.4% και οι μεταβλητές που έχουν επιλεγεί είναι αντιπροσωπευτικές κατά 55%. Παρατηρείται επίσης, πως οι p-values της μεταβλητής RENEWABLE και της μεταβλητής THERMAL είναι μεγαλύτερες από 0.05. Επομένως, στο επόμενο μοντέλο θα αφαιρεθούν. Είναι εμφανές πως εν τέλει το μοντέλο αυτό δεν διαφοροποιείται από το πρώτο αν και κανονικοποιήσαμε τις τιμές των μεταβλητών μας.



### Μοντέλο τέταρτο

Παρατηρείται πως το R-squared και το adjustR-squared έχουν σχεδόν την ίδια τιμή. Το μοντέλο μας κάνει fit κατά 55,8% και οι μεταβλητές που έχουν επιλεγεί είναι αντιπροσωπευτικές κατά 55,3%. Παρατηρείται επίσης πως όλες οι p-values των μεταβλητών είναι μικρότερες από 0.05. Είναι εμφανές πως εν τέλει το μοντέλο αυτό δεν διαφοροποιείται από το δεύτερο αν και κανονικοποιήσαμε τις τιμές των μεταβλητών μας.

```
results = smf.ols('SMP ~ DEMAND + HYDRO + UNIT_COAL + UNIT_GAS', data=scaled_df).fit()
print (results.summary())
```

```

=====
                        OLS Regression Results
=====
Dep. Variable:          SMP      R-squared:                0.558
Model:                  OLS      Adj. R-squared:           0.553
Method:                 Least Squares  F-statistic:             114.4
Date:                   Mon, 10 Sep 2018  Prob (F-statistic):      5.43e-63
Time:                   20:03:23      Log-Likelihood:          36.991
No. Observations:      368          AIC:                     -63.98
Df Residuals:          363          BIC:                     -44.44
Df Model:               4
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-0.4621	0.027	-17.263	0.000	-0.515	-0.409
DEMAND	-0.5336	0.069	-7.783	0.000	-0.668	-0.399
HYDRO	0.3008	0.051	5.939	0.000	0.201	0.400
UNIT_COAL	0.6930	0.048	14.397	0.000	0.598	0.788
UNIT_GAS	0.2222	0.041	5.363	0.000	0.141	0.304

```
=====
Omnibus:                157.003  Durbin-Watson:           0.580
Prob(Omnibus):          0.000    Jarque-Bera (JB):        652.543
Skew:                   1.859    Prob(JB):                 2.00e-142
Kurtosis:               8.360    Cond. No.                 8.41
=====
```

### Μοντέλο πέμπτο

Για το πέμπτο και τελευταίο μοντέλο αποφασίζεται να δημιουργήσουμε ένα νέο dataset διαγράφοντας τις μεταβλητές με υψηλή συσχέτιση.

```
corr_matrix = data1.corr().abs()

# Select upper triangle of correlation matrix
upper = corr_matrix.where(np.triu(np.ones(corr_matrix.shape), k=1).astype(np.bool))

# Find index of feature columns with correlation greater than 0.95
to_drop = [column for column in upper.columns if any(upper[column] > 0.75)]

print ("Highly Correlated Features :\t",to_drop)
```

Highly Correlated Features : ['THERMAL', 'UNIT\_COAL', 'UNIT\_GAS']

```
data2 = data1.drop(data1[to_drop], axis=1)

data2.head()
```

	DEMAND	SMP	RENEWABLE	HYDRO	NET_BORDER
0	153013.63	49.57	16430.0	3443.00	-20579.0
1	156173.44	53.78	19411.0	8681.79	-11381.0
2	162803.98	59.17	20715.0	16649.28	-111.0
3	156143.82	58.61	19318.0	10294.34	-2039.0
4	162458.67	53.78	31450.0	4859.06	4004.0

Το μοντέλο λοιπόν με τα στατιστικά του αποτελέσματα είναι το εξής:

```
results = smf.ols('SMP ~ DEMAND + HYDRO + RENEWABLE + NET_BORDER', data=data2).fit()
print (results.summary())
```

```

=====
                        OLS Regression Results
=====
Dep. Variable:          SMP      R-squared:                0.527
Model:                 OLS      Adj. R-squared:           0.521
Method:                Least Squares  F-statistic:              100.9
Date:                  Mon, 10 Sep 2018  Prob (F-statistic):       1.11e-57
Time:                  20:15:08      Log-Likelihood:           -1312.8
No. Observations:      368          AIC:                      2636.
Df Residuals:          363          BIC:                      2655.
Df Model:               4
Covariance Type:      nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	47.3158	4.514	10.482	0.000	38.439	56.193
DEMAND	0.0002	3.41e-05	5.705	0.000	0.000	0.000
HYDRO	0.0004	0.000	2.432	0.016	8.25e-05	0.001
RENEWABLE	-0.0006	6.16e-05	-9.243	0.000	-0.001	-0.000
NET_BORDER	0.0005	4.65e-05	11.361	0.000	0.000	0.001

```

=====
Omnibus:                160.657      Durbin-Watson:           0.608
Prob(Omnibus):          0.000      Jarque-Bera (JB):        704.638
Skew:                   1.885      Prob(JB):                 9.77e-154
Kurtosis:               8.634      Cond. No.                 1.48e+06
=====

```

Παρατηρείται πως το R-squared και το adjustR-squared έχουν σχεδόν την ίδια τιμή. Το μοντέλο μας κάνει fit κατά 52,7% και οι μεταβλητές που έχουν επιλεγεί είναι αντιπροσωπευτικές κατά 52,1%. Παρατηρείται επίσης πως όλες οι p-values των μεταβλητών είναι μικρότερες από 0.05. Είναι εμφανές πως εν τέλει το μοντέλο αυτό, αν και διαγράφηκαν όσες ανεξάρτητες μεταβλητές είχαν υψηλή συσχέτιση και τα δεδομένα μας έχουν κανονικοποιηθεί, έχει το μικρότερο ποσοστό επιτυχίας σε σχέση με τα προηγούμενα.

## Κεφάλαιο 6 ΣΥΜΠΕΡΑΣΜΑΤΑ

Το υπάρχον σύστημα ηλεκτρικής ενέργειας της Ελλάδος είναι παρωχημένο.

Όπως, διαπιστώθηκε στο 2<sup>ο</sup> κεφάλαιο, στηρίζεται κυρίως σε λιγνιτικές μονάδες παραγωγής ενέργειας και εκμεταλλεύεται σε μικρό βαθμό τις εξελίξεις της τεχνολογίας στον τομέα.

Στο τρίτο κεφάλαιο αναλύθηκε το έξυπνο ηλεκτρικό δίκτυο, δηλαδή ένα δίκτυο που ενσωματώνει τα παραδοσιακά χαρακτηριστικά του ηλεκτρικού δικτύου, παράλληλα με ένα πολυεπίπεδο στρώμα πληροφοριών και επικοινωνιών και στοχεύει στην αυτοματοποίηση των λειτουργιών του δικτύου.

Στο 4<sup>ο</sup> κεφάλαιο έγινε μια ανάλυση των μεγάλων δεδομένων και της επιστήμης των δεδομένων, καθώς και των εργαλείων που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία για τη δημιουργία μοντέλων πρόβλεψης της οριακής τιμής συστήματος.

Στο 5<sup>ο</sup> κεφάλαιο χρησιμοποιήθηκαν τα ανοιχτά μεγάλα δεδομένα από τον ΛΑΓΗΕ, τα οποία περιείχαν πληροφορίες για την ωριαία προσφορά και ζήτηση της ενέργειας, αλλά και τον τύπο του καυσίμου και τις μονάδες από τις οποίες παράγεται η ενέργεια και επεξεργάστηκαν. Στη συνέχεια οπτικοποιήθηκαν και αναλύθηκε διεξοδικά η μεθοδολογία που ακολουθήθηκε, ώστε να δημιουργηθεί ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης που προσεγγίζει τα δεδομένα σε ποσοστό 55,8%. Προφανώς αυτό το ποσοστό χρήζει περαιτέρω βελτίωσης. Είναι εμφανές ότι η οριακή τιμή συστήματος δεν επηρεάζεται μόνο από τα δεδομένα που είχαμε διαθέσιμα. Προτείνεται μελλοντική δημιουργία ενός πιο ακριβούς μοντέλου πρόβλεψης βασιζόμενο σε πιο πλούσια σε πληροφορίες δεδομένα. Ακόμη θα μπορούσαν να χρησιμοποιηθούν δεδομένα και από προηγούμενες χρονιές καθώς έχουν χρησιμοποιηθεί μόνο τα δεδομένα του 2017. Τέλος, θα μπορούσε να δημιουργηθεί και να ελεγχεί ως προς την επάρκειά του κάποιο μοντέλο πρόβλεψης με τη χρήση νευρωνικών δικτύων.

## Βιβλιογραφία

- [1] Κ. Βουρνάς, Γ. Κονταξής, Εισαγωγή στα Συστήματα Ηλεκτρικής Ενέργειας, Αθήνα: Εθνικό Μετσόβιο Πολυτεχνείο, 2001
- [2] Β.Κ. Παπαδιάς, Ανάλυση Συστήματος Ηλεκτρικής Ενέργειας, Τόμος Ι, Αθήνα: Εθνικό Μετσόβιο Πολυτεχνείο, 1985
- [3] ΑΔΜΗΕ, Διαθέσιμο στο link: <http://www.admie.gr/>
- [4] ΑΔΜΗΕ, Δεκαετής Πρόγραμμα Ανάπτυξης Συστήματος Μεταφοράς 2017-2026, Αθήνα: ΑΔΜΗΕ, 2016
- [5] J. Ekanayake, K. Liyanage, J. Wu, Smart Grid Technology and Applications, Chichester, UK: John Wiley & Sons, 2012
- [6] J. Grainger, W. Stevenson Jr., Power System Analysis, New York: McGraw – Hill, 1994
- [7] Η. Κυριακίδης, Διεσπαρμένη Παραγωγή, Διαθέσιμο στο link: <http://www.eng.ucy.ac.cy/elias/courses/ECE681/presentations/Distributed%20Generation.pdf>
- [8] Σ. Παπαθανασίου, Παραγωγή Ηλεκτρικής Ενέργειας από Ανανεώσιμες Πηγές, Αθήνα: Εθνικό Μετσόβιο Πολυτεχνείο
- [9] U.S. Department of Energy . «Understanding the Benefits of the Smart Grid.», June 2010.
- [10] X. Li, I. Lille, Securing Smart Grid: Cyber Attacks, Countermeasures, and Challenges, IEEE Communications Magazine, 50(8), pp. 38-45, 2012
- [11] EEI-AEIC-UTC. Smart Meters and Smart Meter Systems: A Metering Industry Perspective, EEI, March 2011.
- [12] Xi Fang, Satyajayant Misra, Guoliang Xue, Dejun Yang. Smart Grid – The New and Improved Power Grid: A Survey. IEEE, 2011
- [13] S.Russell, P.Norvig, “Τεχνητή νοημοσύνη, μια σύγχρονη προσέγγιση”, Εκδόσεις ΚΛΕΙΔΑΡΙΘΜΟΣ, 2<sup>nd</sup> edition
- [14] L. Zadeh, “Fuzzy sets. Information and Control”, 8:338–353, 1965
- [15] Γ. Θεοδώρου, «Εισαγωγή στην ασαφή λογική, Βασικές αρχές της ασαφούς λογικής με εφαρμογές στην Τεχνολογία», Θεσσαλονίκη 2010, Εκδόσεις ΤΖΙΟΛΑ

[16] Stavros, Samaras-Kamilarakis & Petros-Angelos, Vogiatzakis & Nathanail, Eftihia & Mitropoulos, Lambros. (2019). The Contribution of Open Big Data Sources and Analytics Tools to Sustainable Urban Mobility: Proceedings of 4th Conference on Sustainable Urban Mobility (CSUM2018), 24 - 25 May, Skiathos Island, Greece. 10.1007/978-3-030-02305-8\_85.

[17] S.Dash, G.Mohanty, A. Mohanty, “Intelligent Air Conditioning System using Fuzzy Logic”,International Journal of Scientific & Engineering Research,vol 3, Issue 12, Dec-2012