



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΟΡΥΚΤΩΝ ΠΟΡΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΑΝΑΛΥΣΗΣ ΡΕΥΣΤΩΝ ΚΑΙ ΠΥΡΗΝΩΝ



Ανάπτυξη μοντέλου ANN για τη πρόβλεψη των ογκομετρικών συντελεστών V_o και R_s πετρελαίων σε συνθήκες ταμειωτήρα



Διπλωματική Εργασία

ΒΑΝΟΣ ΠΑΝΤΕΛΗΣ

Εξεταστική επιτροπή:
ΝΙΚΟΛΑΟΣ ΒΑΡΟΤΣΗΣ (επιβλέπων)
ΒΑΣΙΛΕΙΟΣ ΓΑΓΑΝΗΣ
ΜΙΧΑΗΛ ΓΑΛΕΤΑΚΗΣ

ΝΟΕΜΒΡΙΟΣ 2015

ΧΑΝΙΑ

Περίληψη

Στη βιομηχανία πετρελαίου υπάρχει η απαίτηση για όσο το δυνατόν πιο σωστές τιμές των ιδιοτήτων των ρευστών του ταμιευτήρα ώστε να μπορούν να πραγματοποιηθούν ακριβείς υπολογισμοί της Μηχανικής Ταμιευτήρων και της Μηχανικής Παραγωγής όπως π.χ. να καταστεί δυνατή μια ακριβής εκτίμηση των αποθεμάτων, της παραγωγικής ικανότητας και της απόδοσης ανάκτησης ενός ταμιευτήρα. Επίσης μπορούν να χρησιμοποιηθούν για την ανάλυση των δεδομένων της *δοκιμής γεώτρησης* (well test) και των δεδομένων παραγωγής, για το σχεδιασμό των γεωτρήσεων παραγωγής καθώς και για περαιτέρω μελέτες της μηχανικής πετρελαίου όπως η ανάπτυξη μοντέλων ισοζυγίου μάζας, μέτρηση ροής, βελτιστοποίηση συστημάτων παραγωγής, το σχεδιασμό των επιφανειακών εγκαταστάσεων παραγωγής, τη μαθηματική προσομοίωση ταμιευτήρα, την παραμετροποίηση των καταστατικών εξισώσεων κλπ.

Εφόσον έχει πραγματοποιηθεί δειγματοληψία ρευστού του ταμιευτήρα εν συνεχεία μετρώνται οι τιμές των ιδιοτήτων εργαστηριακά από τις **PVT** (πίεση-όγκος-θερμοκρασία) μελέτες. Στη περίπτωση που τα δεδομένα αυτά πρέπει να προσδιοριστούν άμεσα, γιατί η πραγματοποίηση των μελετών PVT χρειάζονται χρόνο εκτέλεσης πολλών εβδομάδων αλλά και σημαντικό κόστος, συχνά οι μηχανικοί εφαρμόζουν εμπειρικές συσχετίσεις ή μαθηματικούς αλγορίθμους οι οποίοι χρησιμοποιούν άμεσα διαθέσιμα δεδομένα παραγωγής ώστε να εκτιμήσουν τις τιμές των ιδιοτήτων με ένα σχετικά αποδεκτό σφάλμα.

Η παρούσα διπλωματική εργασία εστιάζει στην ανάπτυξη μοντέλων πρόβλεψης των θερμοδυναμικών ιδιοτήτων πετρελαϊκών ρευστών με χρήση αλγορίθμων από το χώρο της Μηχανικής Εκμάθησης (machine learning) γνωστών και ως Τεχνητά Νευρωνικά Δίκτυα (*artificial neural network*) **ANN**. Τα μοντέλα χρησιμοποιούν δεδομένα παραγωγής όπως αυτά της πυκνότητας πετρελαίου (**API°**), της πυκνότητας αερίου (**γg**), του **GOR** και της θερμοκρασία ταμιευτήρα (**T**), ώστε να επιτευχθεί ελαχιστοποίηση του σφάλματος πρόβλεψης σε ό,τι αφορά τις παρακάτω ιδιότητες του πετρελαϊκού ρευστού:

1. Ογκομετρικός συντελεστής σχηματισμού υγρής φάσης **B_o**
2. Του λόγου αερίου εν διαλύσει/πετρελαίου **R_s**
3. Πίεση στο σημείο φυσαλίδας-κορεσμού **p_b**

Η ανάπτυξη των νευρωνικών έγινε με την επίλυση του προβλήματος εποπτευόμενης παλινδρόμησης (supervised regression problem) η οποία είναι μια ευρέως χρησιμοποιημένη τεχνική

μοντελοποίησης για την έρευνα της συσχέτισης μεταξύ μίας ή περισσότερων εξαρτώμενων μεταβλητών και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Για τα συγκεκριμένα προβλήματα που επιλύθηκαν, ως εξαρτημένες μεταβλητές θεωρήθηκαν οι ιδιότητες του ρευστού (R_s , B_o , ρ_b) οι οποίες αποτέλεσαν και τις εξόδους των νευρωνικών ενώ ως ανεξάρτητες μεταβλητές-είσοδοι θεωρήθηκαν τα δεδομένα της βάσης δεδομένων που χρησιμοποιήθηκαν για την εκπαίδευση τους.

Χρησιμοποιήθηκε η βάση δεδομένων του Εργαστηρίου Ανάλυσης Ρευστών και Πυρήνων του Πολυτεχνείου Κρήτης η οποία περιέχει μεγάλο αριθμό δεδομένων PVT πετρελαίων ταμειυτήρα με εκτεταμένη γεωγραφική διασπορά. Η βάση δεδομένων ελέγχθηκε ποιοτικά και μετασχηματίστηκε ανάλογα με τις ανάγκες του κάθε νευρωνικού δικτύου.

Για την ανάπτυξη των νευρωνικών δικτύων έγινε χρήση της δομής «Perceptron πολλών επιπέδων» υιοθετώντας ως μέθοδο σύγκλισης την επιβλεπόμενη “Levenberg-Marquardt”.

Τέλος, παρουσιάζονται αναλυτικά τα αποτελέσματα της ακρίβειας των προβλέψεων των νευρωνικών δικτύων που αναπτύχθηκαν και εν συνεχεία οι επιδόσεις τους συγκρίθηκαν με τις επιδόσεις των ήδη εφαρμοζόμενων από τη βιομηχανία πετρελαίου εμπειρικών συσχετίσεων όπου και διαπιστώθηκε ότι τα νευρωνικά που αναπτύχθηκαν παρουσίασαν σημαντικά μεγαλύτερη ακρίβεια επί όλων των μεθόδων για ένα σύνολο δεδομένων τα οποία δεν χρησιμοποιήθηκαν για την εκπαίδευσή τους.

Abstract

In the oil industry there is a requirement for access to correct values of the properties of the reservoir fluids at real time during well testing as well as before the delivery of the PVT report in order to conduct as accurately as possible the necessary Reservoir Engineering and Production Engineering calculations. Such calculations include, among others, the accurate assessment of stocks, the production capability and performance recovery of the reservoir. These early predicted PVT data can also be used for data analysis of well test, for the design of the production drilling and for further petroleum engineering studies such as mass-balance models, multiphase flow measurements, optimizing production systems, design of surface production facilities, preliminary reservoir simulation, proper tuning of the equation of state models, etc.

Once the sampling of the fluid reservoir has been performed, the values of properties are later on measured by laboratory PVT (pressure-volume-temperature) studies. This can take several weeks for the PVT report to be delivered and therefore engineers often apply empirical correlations or mathematical algorithms which use readily available production data to estimate the property values with an acceptable accuracy.

During this thesis predictive models of thermodynamic properties of petroleum fluids were developed using algorithms from the field of machine learning known as Artificial Neural Networks - ANN. These models use production data as the oil density (API°), the gas density (SG), the GOR and the reservoir temperature (T), in order to accurately predict:

1. Oil formation volume factor B_o
2. Solution gas-oil ratio R_s
3. Bubble point pressure-saturation point p_b

The development of neural networks was done by solving the problem of supervised regression. Supervised regression investigates the correlation between one or more dependent variables with one or more independent variables.

A rich database was used which contains a large number of reservoir oil PVT data from a broad geographic origin.

The accuracy of the predictions of the developed neural networks is presented in detail. Their precision was compared against the accuracy of the empirical correlations widely used by the oil industry. It was found that the neural networks outperformed all other models.

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

I.	ΕΙΣΑΓΩΓΗ	12
1.	Πετρέλαιο	12
2.	Βιομηχανία πετρελαίου	15
II.	ΑΛΛΕΣ ΧΡΗΣΕΙΣ ΤΟΥ ΠΕΤΡΕΛΑΙΟΥ	20
1.	Γεωργία	20
2.	Πλαστικά	20
3.	Ελαστικά.....	21
4.	Στη φαρμακευτική.....	21
5.	Χρώματα, απορρυπαντικά, και άλλα	21
III.	ΧΑΡΑΚΤΗΡΙΣΜΟΙ ΚΑΙ ΤΑΞΙΝΟΜΗΣΗ ΤΟΥ ΠΕΤΡΕΛΑΙΟΥ	22
1.	API Πυκνότητα	22
2.	Χαρακτηρισμός πετρελαίου σε ‘γλυκό’ και ‘ξινό’ (Sweet vs Sour).....	23
3.	Πετρέλαια Αναφοράς	24

ΚΕΦΑΛΑΙΟ 2

ΕΚΜΕΤΑΛΕΥΣΗ ΚΟΙΤΑΣΜΑΤΩΝ ΚΑΙ ΠΑΡΑΓΩΓΗ ΥΔΡΟΓΟΝΑΝΘΡΑΚΩΝ

I.	ΕΙΣΑΓΩΓΗ.....	27
II.	DRILL STEM TEST ΚΑΙ ΔΕΙΓΜΑΤΟΛΗΨΙΕΣ	27
III.	ΜΗΧΑΝΙΣΜΟΙ ΠΑΡΑΓΩΓΗΣ ΤΑΜΙΕΥΤΗΡΑ ΠΕΤΡΕΛΑΙΟΥ	28
IV.	ΙΣΟΡΡΟΠΙΑ ΦΑΣΕΩΝ.....	31
V.	ΟΡΙΣΜΟΣ ΤΩΝ ΟΓΚΟΜΕΤΡΙΚΩΝ ΣΥΝΤΕΛΕΣΤΩΝ ΥΓΡΗΣ ΚΑΙ ΑΕ-ΡΙΑΣ ΦΑΣΗΣ	34
1.	Ογκομετρικός συντελεστής σχηματισμού υγρής φάσεως B_o	34
2.	Λόγος αερίου εν διαλύσει/πετρελαίου R_s	36
3.	Λόγος αερίου/πετρελαίου GOR	37
4.	Ογκομετρικός συντελεστής σχηματισμού αέριας φάσης B_g	37
VI.	ΣΗΜΑΣΙΑ ΤΩΝ PVT ΜΕΛΕΤΩΝ	41

ΚΕΦΑΛΑΙΟ 3

ΕΡΓΑΣΤΗΡΙΑΚΕΣ ΜΕΛΕΤΕΣ ΠΕΤΡΕΛΑΙΩΝ

I.	PVT ΜΕΛΕΤΕΣ ΠΕΤΡΕΛΑΙΩΝ.....	43
1.	Μελέτη σταθερής μάζας (constant mass).....	43
2.	Μελέτη Διαφορικής Εξάτμισης.....	45
3.	Μελέτη αριστοποίησης των συνθηκών Διαχωρισμού (Separation test).....	47
II.	ΑΠΕΥΘΕΙΑΣ ΕΚΤΟΝΩΣΗΣ ΡΕΥΣΤΟΥ ΤΟΥ ΤΑΜΙΕΥΤΗΡΑ ΣΕ ΕΠΙΦΑΝΕΙΑΚΕΣ ΣΥΝΘΗΚΕΣ (DIRECT FLASH LIBERATION)	49
III.	ΣΥΝΘΕΤΗ ΕΞΑΤΜΙΣΗ (COMPOSITE FLASH).....	49

ΚΕΦΑΛΑΙΟ 4

ΕΚΤΙΜΗΣΗ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΟΓΚΟΜΕΤΡΙΚΩΝ ΣΥΝΤΕΛΕΣΤΩΝ B_o , R_s ΚΑΙ ΤΟΥ ΣΗΜΕΙΟΥ ΠΙΕΣΗΣ ΦΥΣΑΛΙΔΑΣ-ΚΟΡΕΣΜΟΥ ΑΠΟ ΤΗΝ ΒΙΟΜΗΧΑΝΙΑ ΠΕΤΡΕΛΑΙΟΥ

I.	ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΙΔΙΟΤΗΤΩΝ B_o , R_s ΚΑΙ p_b	53
II.	ΕΜΠΕΙΡΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΟΓΚΟΜΕΤΡΙΚΩΝ ΣΥΝΤΕΛΕΣΤΩΝ B_o , R_s ΚΑΙ ΤΟΥ ΣΗΜΕΙΟΥ ΠΙΕΣΗΣ ΦΥΣΑΛΙΔΑΣ-ΚΟΡΕΣΜΟΥ	54
1.	Κατηγορίες εμπειρικών συσχετίσεων	55
2.	Συσχέτιση των άμεσα διαθέσιμων δεδομένων παραγωγής με τις ζητούμενες ιδιότητες	56
3.	Παρουσίαση των επιλεγμένων εμπειρικών συσχετίσεων ανά ιδιότητα.....	58
a.	Εμπειρικές συσχετίσεις πρόβλεψης B_o	59
b.	Εμπειρικές συσχετίσεις πρόβλεψης p_b	65

c. Εμπειρικές συσχετίσεις πρόβλεψης R_s	70
III. ΑΝΤΙΚΕΙΜΕΝΟ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ	74

ΚΕΦΑΛΑΙΟ 5

ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

I. ΕΙΣΑΓΩΓΗ.....	77
II. ΕΦΑΡΜΟΓΕΣ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ.....	81
III. ΟΡΙΣΜΟΣ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΠΡΟΣ ΕΠΙΛΥΣΗ ΤΗΣ ΠΑΡΟΥΣΑΣ ΔΙΠΛΩΜΑΤΙΚΗΣ	82
IV. PERCEPTRON ΠΟΛΛΩΝ ΕΠΙΠΕΔΩΝ.....	84
1. ΠΡΟΒΛΗΜΑΤΑ ΣΤΗΝ ΕΚΠΑΙΔΕΥΣΗ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ..	87
a. Ο στόχος της γενίκευσης και το πρόβλημα της υπερπροσαρμο- γής (overfitting).....	87
b. Το πρόβλημα των τοπικών ελαχίστων.....	89
2. ΜΕΘΟΔΟΣ LEVENBERG-MARQUARDT	90

ΚΕΦΑΛΑΙΟ 6

ΑΝΑΠΤΥΞΗ ΜΟΝΤΕΛΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΓΙΑ ΤΗ ΠΡΟ- ΒΛΕΨΗ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΙΔΙΟΤΗΤΩΝ B_o , R_s , p_b

I. ΟΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗΝ ΑΝΑΠΤΥΞΗ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ	96
II. ΣΤΑΤΙΣΤΙΚΑ ΚΡΙΤΗΡΙΑ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΗΣ ΑΚΡΙΒΕΙΑΣ ΤΩΝ ΠΡΟΒΛΕΨΕΩΝ ΤΩΝ ΜΟΝΤΕΛΩΝ ΕΚΤΙΜΗΣΗΣ ΙΔΙΟΤΗΤΩΝ	99
III. ΑΝΑΠΤΥΞΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΕ ΠΕΡΙΒΑΛΛΟΝ MATLAB	104
IV. ΑΝΑΠΤΥΞΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΟΥ B_o 1. Δημιουργία Υπό-Βάσης δεδομένων για την εκπαίδευση του νευρωνι- κού	112

2. Παρουσίαση της αρχής λειτουργίας του νευρωνικού δικτύου	118
3. Τρόπος επιλογής του βέλτιστου νευρωνικού	133
4. Παρουσίαση του βέλτιστου νευρωνικού	135
V. ΑΝΑΠΤΥΞΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΟΥ p_b	
1. Βάση δεδομένων για την ανάπτυξη του νευρωνικού	138
2. Παρουσίαση της αρχής λειτουργίας του νευρωνικού	138
3. Τρόπος επιλογής του βέλτιστου νευρωνικού	138
4. Παρουσίαση του βέλτιστου νευρωνικού	139
VI. ΑΝΑΠΤΥΞΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΟΥ R_s	
1. Βάση δεδομένων για την εκπαίδευση του νευρωνικού	142
2. Παρουσίαση της αρχής λειτουργίας του νευρωνικού	144
3. Ανηγμένο ή κανονικοποιημένο σχετικό ποσοστιαίο σφάλμα και από- λυτο σχετικό ποσοστιαίο σφάλμα για την αξιολόγηση της ιδιότητας R_s	144
4. Τρόπος επιλογής του βέλτιστου νευρωνικού	146
5. Παρουσίαση του βέλτιστου νευρωνικού	147

ΚΕΦΑΛΑΙΟ 7

ΣΥΓΚΕΝΤΡΩΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΕΠΙ ΤΩΝ ΕΚΤΙΜΗΣΕΩΝ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΙΔΙΟΤΗΤΩΝ B_o , R_s , p_b

I. ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΣΦΑΛΜΑΤΩΝ ΕΠΙ ΤΩΝ ΕΚΤΙΜΗΣΕΩΝ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΙΔΙΟΤΗΤΩΝ B_o, p_b, R_s ΤΩΝ ΕΜΠΕΙΡΙΚΩΝ ΣΥΣΧΕ- ΤΙΣΕΩΝ	153
II. ΣΥΓΚΡΙΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΠΡΟΒΛΕΨΕΩΝ ΕΠΙ ΤΩΝ ΠΡΑΓΜΑ- ΤΙΚΩΝ ΤΙΜΩΝ ΤΩΝ ΙΔΙΟΤΗΤΩΝ B_o , R_s , p_b ΤΩΝ ΑΝΑΠΤΥΧΘΕΝΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΣΕ ΣΧΕΣΗ ΜΕ ΤΙΣ ΑΝΤΙΣΤΟΙΧΕΣ ΠΡΟΒΛΕΨΕΙΣ ΤΩΝ ΕΜΠΕΙΡΙΚΩΝ ΣΥΣΧΕΤΙΣΕΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΑΠΟ ΤΗ ΒΙΟΜΗΧΑΝΙΑ ΠΕΤΡΕΛΑΙΟΥ	176

ΠΑΡΑΡΤΗΜΑ	182
ΣΥΜΠΕΡΑΣΜΑΤΑ	185
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	189

ΚΕΦΑΛΑΙΟ 1

I. Εισαγωγή

i. Πετρέλαιο

Το πετρέλαιο (από τα Ελληνικά πέτρα και έλαιο, «λάδι της πέτρας» / λατινικά Petroleum) μαζί με το κάρβουνο (άνθρακας) και το φυσικό αέριο χαρακτηρίζονται ως ορυκτά καύσιμα. Τα ορυκτά καύσιμα δημιουργήθηκαν από την οργανική ύλη φυτικών και ζωικών οργανισμών. Τα απανθρακωμένα δένδρα, τα αποτυπώματα φύλλων σε κοιτάσματα άνθρακα, οι βιοδείκτες στο πετρέλαιο είναι μερικές από τις ενδείξεις της προέλευσης των ορυκτών καυσίμων. Στη σημερινή μορφή τους έφθασαν μέσω της αποδόμησης της οργανικής ύλης, η οποία όμως στα συνήθη οξειδωτικά περιβάλλοντα οδηγεί στη δημιουργία CO₂. Μόνο ένα μικρό μέρος της οργανικής ύλης διατηρήθηκε σε συνθήκες έλλειψης οξυγόνου (αναερόβιες) και μετασχηματίστηκε σε ορυκτά καύσιμα. Η ηλικία των νεκρών οργανισμών που με την εναπόθεσή τους σχημάτισαν τα ορυκτά καύσιμα κυμαίνεται από μερικά εκατομμύρια μέχρι 650 εκατομμύρια χρόνια.

Το πετρέλαιο είναι ένα εξαιρετικά σύνθετο μίγμα κυρίως υδρογονανθράκων που περιέχουν επιπλέον άζωτο, οξυγόνο, θείο και ίχνη μετάλλων. Το σημείο βρασμού των περισσότερων συστατικών του πετρελαίου ξεπερνάει τους 1000°C, ενώ ο αριθμός των ατόμων άνθρακα μπορεί να φθάσει και τα 80-100 στα μόρια των συστατικών του. Η σύσταση του πετρελαίου παρουσιάζει έντονες διαφοροποιήσεις ανάλογα με την προέλευση και την ηλικία του. Ακόμη και στα όρια του ίδιου ταμιευτήρα ή και στην ίδια γεώτρηση, δείγματα του παραγόμενου πετρελαίου μπορεί να διαφέρουν μεταξύ τους. Το πετρέλαιο στην ακατέργαστη μορφή του πριν τη διύλιση ονομάζεται συνήθως αργό πετρέλαιο, και μπορεί να έχει ένα διακριτό πράσινο ή μαύρο χρώμα και μπορεί να είναι είτε λεπτόρρευστο, όπως η βενζίνη ή παχύρρευστο σαν πίσσα.

Το φυσικό αέριο είναι ένα μίγμα ελαφρών υδρογονανθράκων. Συναντάται υπό πίεση σε πορώδεις σχηματισμούς. Η ύπαρξη και κοιτάσματος πετρελαίου κάτω από το αέριο δεν είναι απαραίτητη. Τα κύρια συστατικά του φυσικού αερίου είναι τα πρώτα μέλη της σειράς των αλκανίων – μεθάνιο, αιθάνιο, προπάνιο, κανονικό(η)βουτάνιο, ισο(ι)βουτάνιο, n-πεντάνιο, i-πεντάνιο και εξάνιο. Επίσης περιέχονται σε μικρότερες συνήθως συγκεντρώσεις άζωτο, διοξείδιο του άνθρακα, υδρόθειο και ευγενή αέρια. Συνήθως το μεθάνιο αποτελεί το 70-95% του φυσικού αερίου. Όταν το κύριο συστατικό είναι το μεθάνιο, το αέριο ονομάζεται ξηρό αέριο (dry gas), ενώ όταν περιέχονται σε μεγάλες συγκεντρώσεις βαρύτεροι υδρογονάνθρακες που είναι υγροποιήσιμοι σε κανονικές συνθήκες ονομάζεται υγρό αέριο (wet gas).

Υπάρχουν αρκετές περιοχές που διαθέτουν κολοσσιαία κοιτάσματα πετρελαίου σε όλο τον κόσμο. Τα πεδία αργού πετρελαίου του Κουβέιτ και της Σαουδικής Αραβίας είναι τα μεγαλύτερα, αν και στη περιοχή της Μέση Ανατολής υπάρχουν και άλλες χώρες όπως το Ιράν και το Ιράκ που επίσης διαθέτουν μεγάλα κοιτάσματα και όλες μαζί συνεισφέρουν το μεγαλύτερο μέρος της παγκόσμιας παραγωγής.

Τα πεδία της Βόρειας Θάλασσας διαθέτουν μεγάλες ανεκμετάλλευτες ποσότητες αργού πετρελαίου, και ασκούν αναμφισβήτητη τη δεύτερη μεγαλύτερη επιρροή από οικονομικής άποψης στη βιομηχανία πετρελαίου. Τα πεδία του Τέξας, κάποτε αποτελούσαν τις μεγαλύτερες περιοχές πετρελαίου παγκοσμίως αλλά τώρα έχουν σχεδόν εξαντληθεί. Το πεδίο του Κόλπου του Μεξικού, τόσο στην ξηρά όσο και στη θάλασσα, είναι μία από τις πιο σημαντικές περιοχές για τους ενεργειακούς της πόρους και τις υποδομές της παγκοσμίως. Από τον κόλπο του Μεξικού προέρχεται το 17% της συνολικής παραγωγής αργού πετρελαίου των ΗΠΑ και το 5% της συνολικής παραγωγής ξηρού αερίου (dry gas). Πάνω από το 45% της συνολικής διύλισης του πετρελαίου των ΗΠΑ πραγματοποιείται κατά μήκος της ακτής του Κόλπου, καθώς και το 51% της συνολικής επεξεργασίας φυσικού αερίου.

Το πετρέλαιο, αν και ήταν γνωστό στον άνθρωπο από την 4^η π.Χ. χιλιετία, μόνο στα τέλη του 19^{ου} αιώνα απέκτησε εμπορικό ενδιαφέρον ως καύσιμο. Για αιώνες και κυρίως στις πλούσιες πετρελαιοφόρες περιοχές (Μεσοποταμία, Κίνα) χρησιμοποιείτο σαν δομικό υλικό, για την μόνωση και διατήρηση του ξύλου και σαν φαρμακευτική πρώτη ύλη. Η πρώτη γεώτρηση για παραγωγή πετρελαίου έγινε στην Πενσυλβανία το 1859 από τον Edwin Drake. Το παραγόμενο πετρέλαιο χρησιμοποιείτο αποκλειστικά για την παραγωγή κηροζίνης (φωτιστικό πετρέλαιο), ενώ τα ελαφρύτερα κλάσματα καίγονταν και τα βαρύτερα έμεναν αχρησιμοποίητα. Όταν ο Edison άνοιξε το πρώτο εργοστάσιο παραγωγής ηλεκτρικής ενέργειας παγκοσμίως το 1882, η ζήτηση για κηροζίνη άρχισε να μειώνεται. Μόνο με την δημιουργία των μηχανών εσωτερικής καύσης από τον Henry Ford που έδειξε στον κόσμο ότι το αυτοκίνητο θα είναι το καλύτερο μέσο μεταφοράς για τις επόμενες δεκαετίες, η βενζίνη άρχισε να είναι ένα προϊόν υψηλής ζήτησης.

Ο πρώτος Παγκόσμιος Πόλεμος ήταν ο πραγματικός καταλύτης για τη συστηματική βιομηχανοποίηση της παραγωγής πετρελαίου λόγω του ότι κατά την διάρκεια αυτής της περιόδου παράχθηκε περισσότερο πετρέλαιο από ό, τι είχε ποτέ παραχθεί στο παρελθόν. Στη σύγχρονη εποχή, το πετρέλαιο θεωρείται ως ένα πολύτιμο αγαθό και διαπραγματεύεται σε όλο τον κόσμο με τον ίδιο τρόπο όπως ο χρυσός και τα διαμάντια.

Πίνακας 1.1 Απεικόνιση των βεβαιωμένων κοιτασμάτων πετρελαίου σε χιλιάδες εκατομμύρια βαρέλια ανά ήπειρο.(πηγή www.bp.com)

Oil: Total proved reserves	at end 1994 Thousand million barrels	at end 2004 Thousand million barrels	at end 2014 Thousand million barrels	Share of total	R/P ratio
Total North America	127.6	223.7	232.5	13.7%	34.0
Total S. & Cent. America	81.5	103.4	330.2	19.4%	*
Total Europe & Eurasia	141.2	140.8	154.8	9.1%	24.7
Total Middle East	663.6	750.1	810.7	47.7%	77.8
Total Africa	65.0	107.6	129.2	7.6%	42.8
Total Asia Pacific	39.2	40.6	42.7	2.5%	14.1
Total World	1118.0	1366.2	1700.1	100.0%	52.5
of which: OECD	148.4	244.4	248.6	14.6%	30.3
Non-OECD	969.5	1121.8	1451.5	85.4%	60.1
OPEC	778.9	918.8	1216.5	71.6%	91.1
Non-OPEC £	216.0	324.1	341.7	20.1%	24.5
European Union #	8.0	7.3	5.8	0.3%	11.2
Former Soviet Union	123.1	123.3	141.9	8.3%	28.2

Πίνακας 1.2 Απεικόνιση των βεβαιωμένων κοιτασμάτων φυσικού αερίου σε τρισεκατομμύρια κυβικά μέτρα ανά ήπειρο. (πηγή www.bp.com)

Natural Gas: Total proved reserves	at end 1994 Trillion cubic metres	at end 2004 Trillion cubic metres	at end 2014 Trillion cubic feet	at end 2014 Trillion cubic metres	Share of total	R/P ratio
Total North America	8.5	7.5	429.0	12.1	6.5%	12.8
Total S. & Cent. America	5.7	7.0	270.6	7.7	4.1%	43.8
Total Europe & Eurasia	40.6	42.7	2049.5	58.0	31.0%	57.9
Total Middle East	45.5	72.2	2818.6	79.8	42.7%	*
Total Africa	9.1	14.2	499.8	14.2	7.6%	69.8
Total Asia Pacific	9.7	13.0	539.0	15.3	8.2%	28.7
Total World	119.1	156.5	6606.4	187.1	100.0%	54.1
of which: OECD	14.5	14.8	688.2	19.5	10.4%	15.6
Non-OECD	104.6	141.7	5918.3	167.6	89.6%	75.8
European Union	3.6	2.8	52.7	1.5	0.8%	11.3
Former Soviet Union	35.5	37.5	1926.9	54.6	29.2%	71.8

Σημείωση : (R/P) ratio = Reserves-to-production δηλαδή εάν τα αποθέματα που υπάρχουν στο τέλος κάθε έτους διαιρεθούν με τη παραγωγή κατά έτος αυτό, το αποτέλεσμα δίνει τα έτη για τα οποία τα αποθέματα θα διαρκέσουν αν η παραγωγή επρόκειτο να συνεχίσει στον ίδιο ρυθμό.

** πάνω από 100 χρόνια*

ii. Βιομηχανία πετρελαίου

Η βιομηχανία πετρελαίου είναι αρκετά περίπλοκη. Ένα από τα στοιχεία που τη κάνει τόσο περίπλοκη είναι το γεγονός ότι το μεγαλύτερο μέρος των βεβαιωμένων αποθεμάτων πετρελαίου παγκοσμίως ελέγχονται από εταιρείες δημοσίου συμφέροντος (state agencies) και όχι από ιδιωτικές εταιρείες. Στην πραγματικότητα, πάνω από τα μισά συνολικά παγκόσμια πετρελαιοκά αποθέματα ελέγχονται από κρατικές εταιρείες χωρών της Μέσης Ανατολής. Αυτή η περίπλοκη και αλληλεξαρτώμενη λειτουργία αυτών των μεγάλων παικτών της βιομηχανίας και σε συνδυασμό με ισχυρά γεωπολιτικά συμφέροντα για την ενέργεια και την οικονομία της κάθε περιοχής δικαιολογούν την πολυπλοκότητα της λειτουργίας αυτής της βιομηχανίας. Για να καταστεί ευκολότερη η κατανόηση της βιομηχανίας πετρελαίου την υποδιαιρούμε σε δύο μεγάλες κατηγορίες: τις εθνικές εταιρείες πετρελαίου (NOCs) και τις Διεθνείς Εταιρείες Πετρελαίου (IOCs).

Διεθνείς Εταιρείες Πετρελαίου (IOCs)

Στις διεθνείς εταιρείες πετρελαίου περιλαμβάνονται γνωστά ονόματα όπως η ExxonMobil και η Royal Dutch Shell. Κατά κανόνα είναι εταιρείες εισηγμένες στο χρηματιστήριο και λειτουργούν όπως και κάθε άλλη εταιρεία με τη μόνη διαφορά ότι το προϊόν το οποίο εμπορεύονται είναι το πετρέλαιο. Οι πλειοψηφία των IOCs έχουν μακρά ιστορία και γενικά χρονολογούνται από τα τέλη του 19ου αιώνα όταν ιδρύθηκαν. Οι περισσότερες IOCs στις Ηνωμένες Πολιτείες προέκυψαν από τη διάλυση της Standard Oil η οποία ήταν η κυρίαρχη εταιρεία πετρελαίου μέχρι το 1911.

Ένας όρος ο οποίος χρησιμοποιείται συχνά και σχετίζεται με τις IOCs είναι ο όρος "Supermajor" και αναφέρεται στις 6 μεγαλύτερες εισηγμένες εταιρείες πετρελαίου στον κόσμο. Οι 'Supermajors' έχουν περάσει από πολλές αλλαγές από το 1990 και μετά ως αποτέλεσμα των συγχωνεύσεων και των εξαγορών στο δευτερογενή τομέα της αγοράς, τον ανταγωνισμό με τις νέο-ιδρυθέντες NOCs και την μεγάλη ύφεση της τιμής του πετρελαίου στις αρχές του 1990. Ως ομάδα οι supermajors ελέγχουν το 6% του πετρελαίου παγκοσμίως συγκριτικά με τις NOCs που ελέγχουν το 88% του πετρελαίου παγκοσμίως. Οι έξι supermajors έχουν ως εξής :

Πίνακας 1.3 Αναφορά των 6 μεγαλύτερων εισηγμένων εταιρειών πετρελαίου

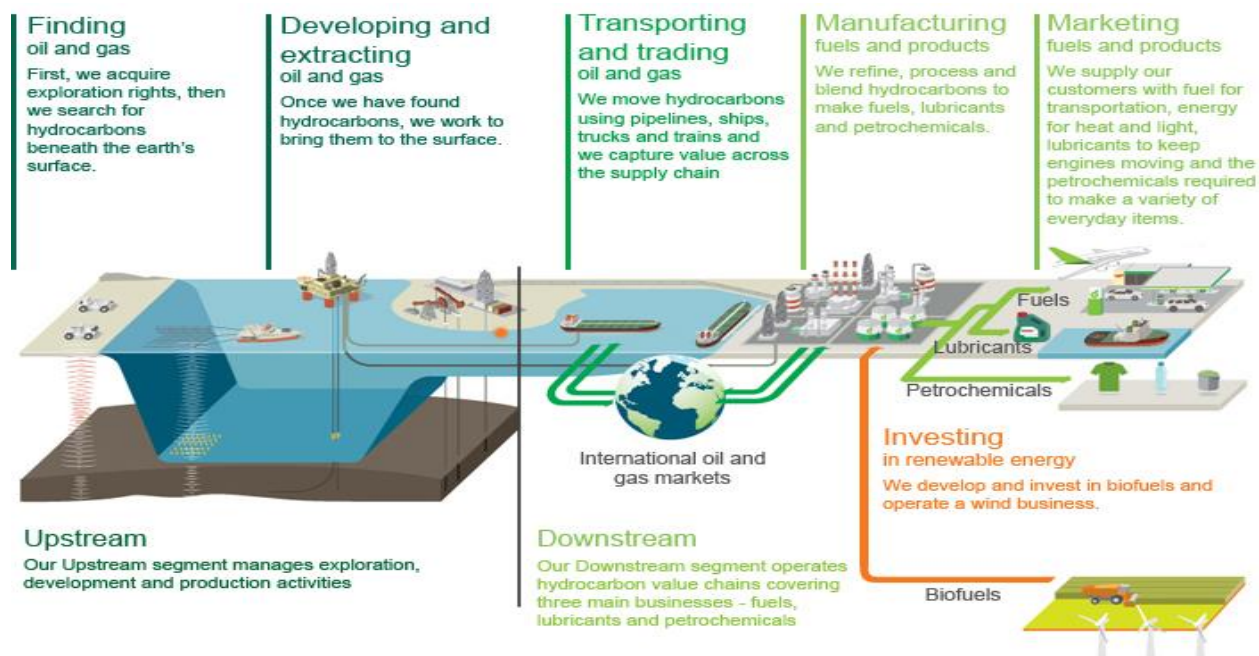
Supermajors	Τοποθεσία	Έσοδα σε δισεκατομμύρια δολάρια	Πιστοποιημένα αποθέματα σε δισεκατομμύρια βαρέλια
ExxonMobil	Texas-United States	383	72
Royal Dutch Shell-Netherlands	The Hague-Netherlands	368	20
BP/Amoco-United Kingdom	London-United Kingdom	308	18
Total SA-France	Paris-France	229	10.5
Chevron-USA	California-United States	204	10.5
ConocoPhillips-USA	Texas-United States	198	8.3

Το μέγεθος των αποθεμάτων δεν είναι ο μόνος τρόπος για να κατηγοριοποιηθεί μια εταιρεία ως προς το μέγεθος επίδρασης της πάνω στη βιομηχανία πετρελαίου. Συνήθως, το μέγεθος αποθεμάτων χρησιμοποιείται ποιο συχνά ως αναφορά από τις κρατικές εταιρίες NOCs ενώ το μέγεθος των αποθεμάτων και οι επιχειρηματικοί κλάδοι που δραστηριοποιείται η κάθε εταιρία χρησιμοποιούνται για να περιγράψουν τις IOCs. Το Αμερικανικό Ινστιτούτο Πετρελαίου έχει κατηγοριοποιήσει τη βιομηχανία σε πέντε κατηγορίες με βάση το πεδίο δραστηριότητά τους. Αυτό μπορεί να δώσει απάντηση στην ερώτηση γιατί η κατοχή μεγάλων αποθεμάτων πετρελαίου δεν συνεπάγεται αυτόματα και μεγάλη εισροή εσόδων και γιατί οι “supermajors” εταιρείες παρά το σχετικά μικρό μέγεθος των αποθεμάτων τους σε σύγκριση με τις NOCs κυριαρχούν στην αγορά. Οι κλάδοι της βιομηχανίας πετρελαίου είναι οι εξής:

Κατηγορία κλάδου	Αντικείμενο κλάδου
Upstream	Αναζήτηση και εκμετάλλευση πετρελαίου
Downstream	τάνκερς, διυλιστήρια, καταναλωτές
Pipeline	αγωγοί πετρελαίου, φυσικού αερίου, CO2
Marine	θαλάσσιες μεταφορές πετρελαίου
Service and Supply(Γενικά)	Κατασκευαστές εξοπλισμού, εταιρείες συμβούλων, κ.λπ.

Οι περισσότερες εταιρείες “supermajors” έχουν καθετοποιημένη παραγωγή. Αυτό σημαίνει ότι μπορεί να υπάρχουν διαφορετικά τμήματα της εταιρείας που ειδικεύονται σε διάφορους κλάδους της βιομηχανίας όπως Upstream, Downstream και Marine. Ενώ όλες οι “supermajors” εταιρείες δραστηριοποιούνται στους κλάδους Upstream και Downstream, λίγες συμμετέχουν στο κλάδο του Pipeline και Marine. Οι περισσότερες έχουν κάποια συμμετοχή στο κλάδο του Service and Supply.

Ο Upstream κλάδος για τις περισσότερες “supermajors” είναι η βασική πηγή των εσόδων. Για παράδειγμα, στη Royal Dutch Shell τα 2/3 των κερδών της προέρχονται από την αναζήτηση και εκμετάλλευση του πετρελαίου. Επειδή οι “supermajors” έχουν αναπτύξει μεγάλη τεχνογνωσία στην αναζήτηση και εκμετάλλευση πετρελαίου λόγω του ότι δραστηριοποιούνται επί πολλές δεκαετίες έχουν αποκτήσει ένα συγκριτικό πλεονέκτημα έναντι των NOCs. Ως αποτέλεσμα της δεσπόμενης θέσης στην αγορά σε αυτό τον τομέα οι “supermajors” πραγματοποιούν το μεγαλύτερο μέρος των Upstream εργασιών στη βιομηχανία και έτσι αντλούν το μεγαλύτερο μέρος των εσόδων τους από την παροχή αυτών των υπηρεσιών, τόσο για τα δικά τους αποθέματα πετρελαίου όσο και για τα αποθέματα άλλων εταιρειών.



Εικόνα 1.1 Απεικονίζεται η καθετοποιημένη παραγωγή μιας supermajor εταιρίας της BP από την αναζήτηση του πετρελαίου μέχρι το τελικό προϊόν στον καταναλωτή. (Πηγή : www.bp.com)

Κρατικές εταιρείες πετρελαίου

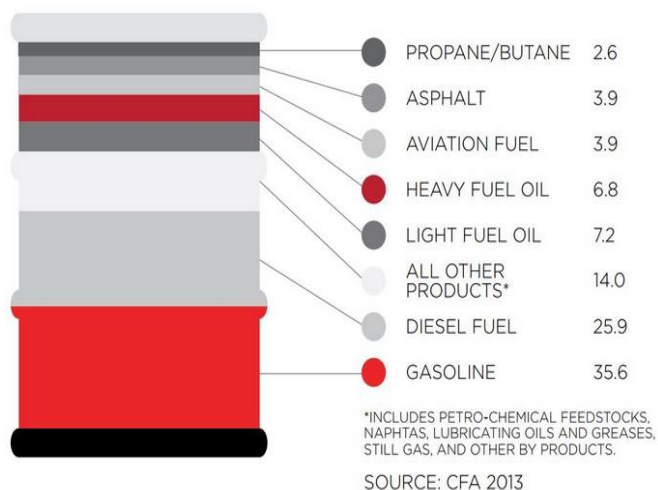
Οι κρατικές εταιρείες πετρελαίου (NOCs) έχουν συσταθεί στα πρότυπα λειτουργίας των Διεθνών Εταιρειών Πετρελαίου IOCs. Η κύρια διαφορά είναι ότι οι IOCs δημοσιεύουν τους ισολογισμούς τους και τις οικονομοτεχνικές τους μελέτες και έχουν μετόχους. Στις αρχές του 20^{ου} αιώνα οι IOCs ήταν οι μεγαλύτεροι παραγωγοί. Τις τελευταίες δεκαετίες ιδρύθηκαν NOCs στις περισσότερες χώρες με μεγάλα αποθέματα πετρελαίου για να τα εκμεταλλευτούν οι ίδιες. Αυτή η τάση δημιουργήθηκε για δύο λόγους.

Ο πρώτος λόγος για την άνοδο των NOCs είναι οι πολιτικές αλλαγές. Στις περισσότερες χώρες που υπήρχαν μεγάλα αποθέματα πετρελαίου άρχισε μια προσπάθεια επανελέγχου των δικαιωμάτων εκμετάλλευσης που κατείχαν αρχικά οι IOCs. Για παράδειγμα πολλοί στρατιωτικοί δικτάτορες στη Μέση Ανατολή κατέλαβαν την εξουσία υποσχόμενοι την αναδιανομή των εσόδων από το πετρέλαιο στη κοινωνία ιδρύοντας NOCs ώστε να κερδίσουν μεγαλύτερη υποστήριξη. Φυσικά, σε πολλές περιπτώσεις, αυτές οι υποσχέσεις δεν τηρήθηκαν.

Ο άλλος λόγος για την άνοδο των NOCs είναι η βιομηχανική πρόοδος. Πολλά από τα πλούσια σε πετρέλαιο κράτη μόχλευσαν τους τεράστιους φυσικούς πόρους τους ώστε να διαπραγματευτούν κερδοφόρα συμβόλαια με IOCs για την αναζήτηση και εκμετάλλευση πετρελαίου. Η δημιουργία του ΟΠΕΚ ήρθε ως αντιστάθμισμα της μεγάλης διαπραγματευτικής δύναμης των IOCs. Σαν ένα γιγάντιο συνδικάτο ο ΟΠΕΚ επέτρεψε στις πλούσιες σε πετρέλαιο χώρες να ασκήσουν μεγαλύτερη πίεση στις IOCs ώστε να προσφέρουν χαμηλότερες τιμές στα συμβόλαια τους. Η ανάπτυξη των δικών τους μέσων για την εξόρυξη και διύλιση του πετρελαίου έδωσε τη δυνατότητα στις NOCs να περιορίσουν την εξάρτησή τους από τις IOCs. Ακολουθεί πίνακας με τις δέκα μεγαλύτερες κρατικές εταιρείες σε σχέση με τα βεβαιωμένα αποθέματα πετρελαίου.

Πίνακας 1.4 Οι δέκα μεγαλύτερες κρατικές εταιρείες σε σχέση με τα βεβαιωμένα αποθέματα(www.petroleum.co.uk)

Ονομασία	Τοποθεσία	Πιστοποιημένα αποθέματα σε δισεκατομμύρια βαρέλια
Saudi Aramco	Saudi Arabia-Middle East	303
National Iranian Oil Company	Iran-Middle East	300
Qatar Petroleum	Qatar-Middle East	170
Iraq National Oil Company	Iraq-Middle East	134
Petroleos de Venezuela	Venezuela-South America	129
Abu Dhabi National Oil Company	Abu Dhabi-Middle East	126
Kuwait Petroleum Corporation	Kuwait-Middle East	111
Nigerian National Petroleum Corporation	Nigeria-Africa	68
Libya NOC	Libya-Africa	50
Sonatrach	Algeria-Africa	39



Εικόνα 1.2 ποσοστιαία κατανομή της χρήσης ενός βαρελιού πετρελαίου

Οι περισσότεροι τείνουν να πιστεύουν ότι το πετρέλαιο χρησιμοποιείται κυρίως για την τροφοδοσία των κινητήρων εσωτερικής καύσης, με τη μορφή της βενζίνης ή του πετρελαίου. Παρά το γεγονός ότι τα μέσα μεταφοράς καταναλώνουν τη μεγαλύτερη ποσότητα του πετρελαίου ένα σημαντικό ποσοστό χρησιμοποιείται για την κατασκευή ενός μεγάλου εύρους προϊόντων.

II. Άλλες χρήσεις του Πετρελαίου

Είναι γνωστό ότι τα σημαντικότερα παράγωγα του πετρελαίου είναι η βενζίνη και το ντίζελ. Παρόλα αυτά μια μεγάλη ποσότητα υποπροϊόντων του πετρελαίου σχετίζονται με τη καθημερινότητα της σύγχρονης ζωής βρίσκοντας εφαρμογή σε μια πληθώρα προϊόντων. Επειδή το αργό πετρέλαιο περιέχει έναν τεράστιο αριθμό διαφορετικών υδρογονανθράκων, διάφορα προϊόντα διύλισης έχουν βρει χρηστική αξία σχεδόν σε όλα, από πλαστικά μέχρι φαρμακευτικά προϊόντα και τρόφιμα. Η βιομηχανία που χρησιμοποιεί το πετρέλαιο για την παραγωγή άλλων χημικών ουσιών αναφέρεται ως πετροχημική βιομηχανία. Εκτιμάται ότι στις βιομηχανοποιημένες χώρες ένας μέσος καταναλωτής χρησιμοποιεί πετροχημικά προϊόντα που αντιστοιχούν σε ημερήσια κατανάλωση τρεισήμισι γαλονιών πετρελαίου και 1.200 λίτρων το χρόνο.

i. Γεωργία

Μια από τις σημαντικότερες χρήσεις του πετρελαίου είναι η παραγωγή αμμωνίας για να χρησιμοποιηθεί ως πηγή αζώτου στα γεωργικά λιπάσματα. Στις αρχές του 20ου αιώνα, ο Fritz Haber εφηύρε μια μέθοδο που επέτρεψε την παραγωγής αμμωνίας σε βιομηχανική κλίμακα. Πριν από αυτό, η αμμωνία για λίπασμα προέρχονταν μόνο από την κοπριά και άλλες βιολογικές διεργασίες. Η γεωργία εξαρτάται επίσης από τη χρήση των φυτοφαρμάκων για να διασφαλιστεί η συνεπής και υγιείς απόδοση των καλλιεργειών. Σχεδόν όλα τα φυτοφάρμακα παράγονται από το πετρέλαιο. Η γεωργία είναι ένας από τους μεγαλύτερους χρήστες των προϊόντων με βάση το πετρέλαιο από τη κίνηση των γεωργικών μηχανημάτων μέχρι τη παραγωγή και χρήση φυτοφαρμάκου.

ii. Πλαστικά

Το πλαστικό είναι ένα βασικό είδος της σύγχρονης βιομηχανίας. Από τις οθόνες των υπολογιστών μέχρι το νάιλον και το φελιζόλ, τα πλαστικά, αποτελούν αναπόσπαστο στοιχείο πολλών μεταποιημένων προϊόντων. Το πολυστυρένιο, από το οποίο είναι κατασκευασμένο το φελιζόλ, και το πολυβινυλοχλωρίδιο (PVC) ήταν και τα δύο δημιουργήματα της εκβιομηχάνισης μετά τον Β Παγκόσμιο Πόλεμο. Το νάιλον το οποίο χρησιμοποιείται κατά κόρον από κάλτσες και μηχανικά γρανάζια μέχρι και σε κινητήρες αυτοκινήτων, είναι το πιο επιτυχημένο πετροχημικό προϊόν με μέχρι σήμερα. Τα περισσότερα πλαστικά προέρχονται από ολεφίνες, οι οποίες περιλαμβάνουν αιθυλένιο και προπυλένιο.

iii. Ελαστικά

Τα ελαστικά είναι κατασκευασμένα από καουτσούκ . Μέχρι το 1910 όλα τα ελαστικά παράγονταν από φυσικά ελαστομερή που προέρχονταν από φυτά. Η ανάγκη για συνθετικό καουτσούκ ήταν σχετικά μικρή μέχρι τον Β Παγκόσμιο Πόλεμο. Ως αποτέλεσμα του εμπάργκο του φυσικού καουτσούκ από τη Νότια Αμερική δημιουργήθηκε η ανάγκη για την παραγωγή συνθετικού ελαστικού σε μεγάλη κλίμακα. Καουτσούκ είναι κατά κύριο λόγο ένα προϊόν του βουταδιενίου.

iv. Στη φαρμακευτική

Ορυκτέλαιο και βαζελίνη είναι υποπροϊόντα πετρελαίου που χρησιμοποιούνται σε πολλές κρέμες και τοπικά φαρμακευτικά προϊόντα. Η λιθανθρακόπισσα που χρησιμοποιείται για την ψύξη και τη πιτυρίδα παράγεται επίσης από το πετρέλαιο. Τα περισσότερα φαρμακευτικά προϊόντα περιέχουν σύνθετα οργανικά μόρια, τα οποία έχουν τη βάση τους σε μικρότερα, απλούστερα οργανικά μόρια. Οι περισσότερες από αυτές τις πρόδρομες ουσίες είναι υποπροϊόντα πετρελαίου.

v. Χρώματα, απορρυπαντικά, και άλλα

Τα αποστάγματα του πετρελαίου όπως βενζόλιο, τολουόλιο, ξυλόλιο και άλλα παρέχουν την πρώτη ύλη για προϊόντα που περιλαμβάνουν βαφές, συνθετικά απορρυπαντικά, και υφάσματα. Βενζόλιο και τολουόλιο είναι οι πρώτες ύλες που χρησιμοποιούνται για τη παραγωγή πολυουρεθανών τα οποία χρησιμοποιούνται σε επιφανειοδραστικές ουσίες, έλαια, ακόμη και στο βερνίκι ξύλου.

Λίστα με μερικά απροσδόκητα προϊόντα που περιέχουν παράγωγα του πετρελαίου

- | | | |
|-------------------------------|------------------------|---------------------------|
| • αντιισταμινικά | • αποσμητικά | • Οι μαλακοί φακοί επαφής |
| • αντισηπτικά | • Ψηφιακά Ρολόγια | • Κρέμα ξυρίσματος |
| • τεχνητές καρδιακές Βαλβίδες | • Συντηρητικά τροφίμων | • σύριγγες |
| • ασπιρίνη | • μελάνι | • οδοντόκρεμα |
| • Κάψουλες βιταμινών | • εντομοκτόνα | • Παιχνίδια |
| • Κορτιζόνη | • Αναισθητικά | • Τηλέφωνα |
| • κραγιόν | • Ιατρικός εξοπλισμός | • Σαμπουάν |
| • πιστωτικές κάρτες | • ομπρέλες | • Κεριά |
| • τεχνητές οδοντοστοιχίες | • βηματοδότες | • φωτογραφικό φιλμ |
| | • αρώματα | |

III. Χαρακτηρισμοί και ταξινόμηση του πετρελαίου

i. API Πυκνότητα

Το API (American Petroleum Institute) είναι συντομογραφία του Αμερικανικού Ινστιτούτου Πετρελαίου και αναφέρεται στην κύρια εμπορική συλλογικότητα της βιομηχανίας πετρελαίου και φυσικού αερίου των Ηνωμένων Πολιτειών. Το API αντιπροσωπεύει περίπου 400 εταιρείες της βιομηχανίας πετρελαίου και βοηθά ώστε να καθοριστούν τα πρότυπα για την παραγωγή, διύλιση και τη διανομή των προϊόντων πετρελαίου. Ένα από τα πιο σημαντικά πρότυπα που το API έχει θέσει, είναι η μέθοδος που χρησιμοποιείται για τη μέτρηση της πυκνότητας του πετρελαίου. Το πρότυπο αυτό ονομάζεται πυκνότητα API.

Η σχετική πυκνότητα, γ_6 , είναι ο λόγος της πυκνότητας μίας ουσίας προς την πυκνότητα μιας ουσίας αναφοράς, συνήθως νερό. Η πυκνότητα API δεν είναι τίποτα περισσότερο από τη κανονική σχετική πυκνότητα που χρησιμοποιείται από τη βιομηχανία πετρελαίου, η οποία συγκρίνει την πυκνότητα του πετρελαίου με εκείνη του νερού. Ένα χαμηλής πυκνότητας αργό πετρέλαιο ή «ελαφρύ πετρέλαιο» είναι προτιμότερο από ότι ένα υψηλής πυκνότητας ή «βαρύ πετρέλαιο» καθώς περιέχει μεγαλύτερες ποσότητες υδρογονανθράκων που μπορούν να μετατραπούν σε βενζίνη. Όπως αναφέρθηκε και στη προηγούμενη ενότητα, το μεγαλύτερο ποσοστό πετρελαίου χρησιμοποιείται για την παραγωγή καυσίμου οπότε τα κοιτάσματα με ελαφρύτερο πετρέλαιο έχουν μεγαλύτερη οικονομική απόδοση.

Η API πυκνότητα υπολογίζεται ως εξής: **$API = (141.5 / \text{σχετική πυκνότητα σε } 60^\circ F) - 131.5$**

Αν και οι τιμές API δεν έχουν μονάδες, συχνά αναφέρονται ως βαθμοί. Έτσι η πυκνότητα API West Texas Intermediate λέγεται ότι είναι 39.6 βαθμοί. Η API πυκνότητα κινείται αντιστρόφως ανάλογα με την σχετική πυκνότητα (γ_6), το οποίο σημαίνει ότι όσο πυκνότερο είναι ένα πετρέλαιο, τόσο χαμηλότερη θα είναι η τιμή της πυκνότητας API. Η τιμή του API χρησιμοποιείται για να χαρακτηρίσει το πετρέλαιο ως ελαφρό, μεσαίο, βαρύ ή πολύ βαρύ. Καθώς το «βάρος» του πετρελαίου είναι σημαντικός καθοριστικός παράγοντας της αγοραίας αξίας του, η εύρεση της API πυκνότητας είναι εξαιρετικά σημαντική. Οι τιμές API για κάθε "βάρος" έχουν ως εξής:

❖ **Ελαφρύ = API > 31.1**

❖ **Μεσαίο = API μεταξύ 22.3 και 31.1**

❖ **Βαρύ = API < 22.3**

❖ **Πολύ βαρύ = API < 10.0** (Στις συγκεκριμένες περιπτώσεις το πετρέλαιο έχει μεγαλύτερη πυκνότητα από το νερό)

Επειδή η πυκνότητα ισούται με μάζα/όγκο, η API μπορεί να χρησιμοποιηθεί για να υπολογιστούν πόσα βαρέλια αργού μπορούν να εξαχθούν από ένα μετρικό τόνο ενός συγκεκριμένου τύπου πετρελαίου. Ένας μετρικός τόνος του (West Texas Intermediate) με API = 39.6 θα παράγει 7,6 βαρέλια (42 γαλόνια κάθε βαρέλι). Η εξίσωση υπολογισμού είναι:

$$\text{Βαρέλια ανά μετρικό τόνο} = 1 / [(141.5 / (API + 131.5)) \times 0.159]$$

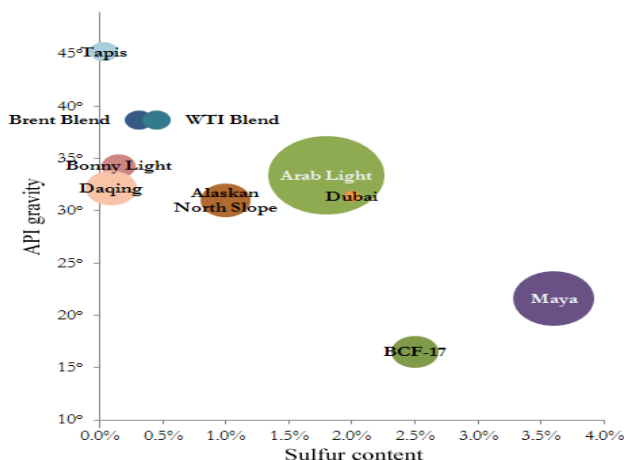
ii. Χαρακτηρισμός πετρελαίου σε ‘γλυκό’ και ‘ξινό’ (Sweet vs Sour)

Οι όροι ‘γλυκό’ και ‘ξινό’ σχετίζονται με την περιεκτικότητα του πετρελαίου ταμιευτήρα (reservoir oil) σε θείο. Οι πρώιμοι παραγωγοί πετρελαίου δοκίμαζαν το πετρέλαιο για να καθορίσουν την ποιότητα του όπου το χαμηλής περιεκτικότητας πετρέλαιο σε θείο είχε στην πραγματικότητα γλυκιά γεύση. Σήμερα γλυκό πετρέλαιο ταμιευτήρα χαρακτηρίζεται αυτό που έχει περιεκτικότητα σε θείο λιγότερη από 0.5%. Το ‘γλυκό’ πετρέλαιο είναι ευκολότερο στη διύλιση και πιο ασφαλές στην εξαγωγή και μεταφορά του από ότι το “ξινό” πετρέλαιο. Επειδή το θείο δρα σαν διαβρωτικό, το γλυκό πετρέλαιο προκαλεί λιγότερη ζημιά στα διυλιστήρια που αυτό συνεπάγεται σε χαμηλότερο κόστος συντήρησης στη πάροδο του χρόνου. Λόγω αυτού του γεγονότος, το γλυκό πετρέλαιο έχει ανώτερη τιμή, της τάξης περίπου των 15\$ ανά βαρέλι, σε σχέση με το ξινό.

Σαν ξινό πετρέλαιο ταμιευτήρα χαρακτηρίζεται το πετρέλαιο το οποίο έχει περιεκτικότητα σε θείο πάνω από 0.5%. Το περισσότερο θείο στο πετρέλαιο συναντάται συνδεδεμένο με άτομα του άνθρακα. Για αυτό το λόγο πριν μεταφερθεί το ξινό πετρέλαιο με δεξαμενόπλοια χρειάζεται κάποιας μορφής αποθείωσης (απομάκρυνση ιχνών υδροθείου για λόγους ασφαλείας). Το πετρέλαιο που παράγεται από τα κράτη του ΟΠΕΚ είναι σχετικά “ξινό” με μέση περιεκτικότητα 1.77%.

iii. Πετρέλαια Αναφοράς

Τα πετρέλαια αναφοράς(**benchmark**) χρησιμοποιούνται ως σημεία αναφοράς κατά την τιμολόγηση του εκάστοτε πετρελαίου. Υπάρχουν περίπου 161 διαφορετικά πετρέλαια αναφοράς από τα οποία τα τρία σημαντικότερα είναι West Texas Intermediate, Brent Crude και το Dubai Crude. Το αργό πετρέλαιο έχει την πιο έντονη συναλλακτική δραστηριότητα από τα εμπορεύσιμα



Διάγραμμα 1.1 Πετρέλαια αναφοράς

αγαθά και αγοράζεται και πωλείται σε συμβόλαια. Ένα συμβόλαιο εμπορεύεται 1.000 βαρέλια πετρελαίου και τα πετρέλαια αναφοράς βοηθούν να καθοριστεί τη τιμή του βαρελιού του εκάστοτε πετρελαίου σε μια εμπορική συμφωνία. Δίπλα παρουσιάζεται Διάγραμμα 1.1 απεικονίζει τα πετρέλαια αναφοράς σε σχέση με τους βαθμούς της API πυκνότητας και το ποσοστό περιεκτικότητας σε θείο.

West Texas Intermediate

WTI είναι ίσως το πιο γνωστό από τα πετρέλαια αναφοράς. Είναι ένα ελαφρύ και γλυκό πετρέλαιο με πυκνότητα API 39.6 βαθμούς που αντιστοιχεί σε σχετική πυκνότητα ίση με 0.827 που σημαίνει ότι σε 60 βαθμούς Fahrenheit το WTI αναλογεί στα 8/10 του βάρους του νερού. Περιέχει 0.24% θείο και προέρχεται από τη Νοτιοδυτικές Ηνωμένες Πολιτείες Αμερικής.

Brent Crude

Το πετρέλαιο Brent πήρε το όνομά του από το κοίτασμα Brent που αυτό με τη σειρά πήρε το όνομα του από μια αγριόχην που προέρχεται από τη Βόρεια Θάλασσα. Είναι ένα ελαφρύ και γλυκό πετρέλαιο με πυκνότητα API 38.06 και σχετική πυκνότητα 0.835 που το καθιστά ελαφρώς "βαρύτερο" από το West Texas Intermediate. Η περιεκτικότητά του σε θείο είναι 0.37%. Η τιμή του πετρελαίου τύπου Brent χρησιμοποιείται για τον καθορισμό της τιμής περίπου των 2/3 των πετρελαίων παγκοσμίως. Ως επί το πλείστον διυλίζεται στην Βορειοδυτική Ευρώπη και μπορούμε να το συναντήσουμε και με τις εξής ονομασίες Brent Blend, London Brent και Brent Petroleum. Το πεδίο Brent βρίσκεται ανατολική λεκάνη της περιοχή Shetland στα μισά της απόστασης μεταξύ της Σκωτίας και της Νορβηγίας.

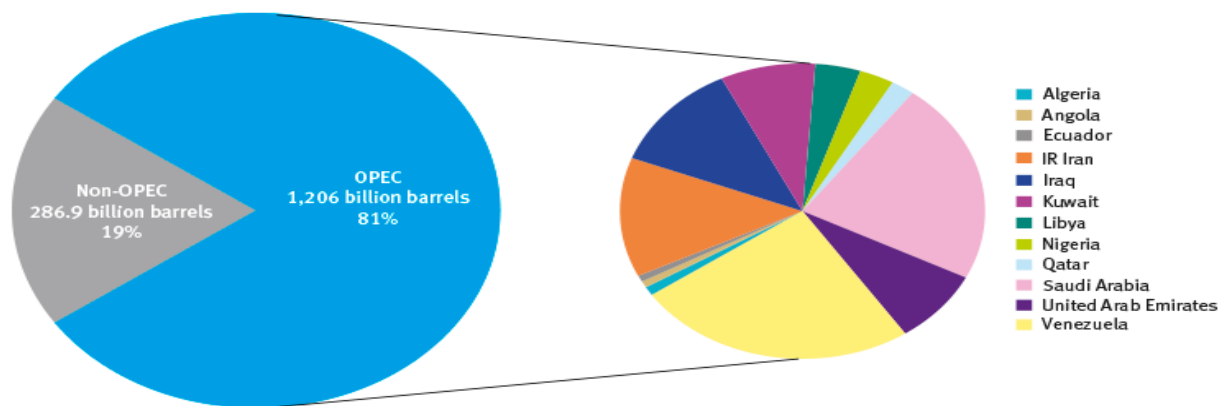
Dubai Crude

Το πετρέλαιο του Ντουμπάι είναι ένα ελαφρύ και 'ξινό' πετρέλαιο με πυκνότητα API 31 βαθμούς και σχετική πυκνότητα 0.871. Η περιεκτικότητά του σε θείο είναι 2% που το καθιστά 6 φορές περισσότερο ξινό από το Brent Crude και 8 φορές περισσότερο ξινό από West Texas Intermediate. Γενικά χρησιμοποιείται για την τιμολόγηση του πετρελαίου που προέρχεται από τον Περσικό Κόλπο. Επίσης είναι γνωστό και ως "Fateh" που σημαίνει καλή μοίρα στα Αραβικά. Η σημασία του δεν εξαρτάται μόνο από την ποιότητά του αλλά και από το γεγονός ότι μέχρι πρόσφατα ήταν το μόνο πετρέλαιο της Μέσης Ανατολής που συναλλασσόταν ελεύθερα σαν χρηματιστηριακό προϊόν.

OPEC Reference Basket (ORB)

Αυτό δεν είναι ένα συγκεκριμένο πετρέλαιο αλλά ένας σταθμισμένος μέσος όρος των πετρελαίων που προέρχονται από τις χώρες του ΟΠΕΚ. Υπάρχουν επί του παρόντος 12 διαφορετικά πετρέλαια που συνδυάζονται για να παραγάγουν το ORB. Η μέση πυκνότητα API για τον παραπάνω συνδυασμό ισούται με 32.7 βαθμούς και η περιεκτικότητά του σε θείο είναι 1.77%. Αυτές οι τιμές άλλαξαν πρόσφατα έτσι ώστε να αντικατοπτρίζουν τη μέση ποιότητα του αργού πετρελαίου των χωρών μελών του ΟΠΕΚ. Η αλλαγή αυτή μείωσε τη τιμή της πυκνότητας API και αύξησε την περιεκτικότητα σε θείο.

OPEC share of world crude oil reserves, 2014



OPEC proven crude oil reserves, at end of 2014 (billion barrels, OPEC share)

Venezuela	299.95	24.9%	Iraq	143.07	11.9%	Libya	48.36	4.0%	Algeria	12.20	1.0%
Saudi Arabia	266.58	22.1%	Kuwait	101.50	8.4%	Nigeria	37.07	3.1%	Angola	8.42	0.7%
IR Iran	157.53	13.1%	UAE	97.80	8.1%	Qatar	25.24	2.1%	Ecuador	8.27	0.7%

Source: OPEC Annual Statistical Bulletin 2015.

Διάγραμμα 1.2 Απεικόνιση των χωρών μελών του ΟΠΕΚ και των βεβαιωμένων αποθεμάτων πετρελαίου της κάθε χώρας όπως και την ποσοστιαία συμμετοχή της(Πηγή :www.opec.org)

ΚΕΦΑΛΑΙΟ 2

ΕΚΜΕΤΑΛΛΕΥΣΗ ΚΟΙΤΑΣΜΑΤΩΝ ΚΑΙ ΠΑΡΑΓΩΓΗ ΥΔΡΟΓΟ- ΝΑΝΘΡΑΚΩΝ

I. Εισαγωγή

Ένα από τα σημαντικά δεδομένα που από την αρχή είναι απαραίτητα στην αξιολόγηση ενός νέου ταμιευτήρα πετρελαίου, ώστε αυτός να χαρακτηριστεί εκμεταλλεύσιμος, είναι ο όγκος των υδρογονανθράκων που μπορούν να παραχθούν από το συγκεκριμένο σχηματισμό. Η μελέτη του ταμιευτήρα, που ξεκινά με τα γεωφυσικά, τις διαγραφίες και την γεωλογική έρευνα της περιοχής συνεχίζεται με την πραγματοποίηση της πρώτης ερευνητικής γεώτρησης.

Οι ερευνητικές γεωτρήσεις έχουν ως σκοπό την επιβεβαίωση της λιθολογικής δομής του υπεδάφους, καθώς και την τελική απόδειξη ύπαρξης ή μη, υδρογονανθράκων. Τα σημεία των γεωτρήσεων επιλέγονται πολύ προσεκτικά έτσι ώστε να μπορεί να συλλεγεί η μέγιστη δυνατή πληροφορία με τις λιγότερες δυνατές γεωτρήσεις. Τα δείγματα των πετρωμάτων που λαμβάνονται μελετώνται με τις μεθόδους της κλασικής ανάλυσης των πυρήνων, τα αποτελέσματα της οποίας οδηγούν στην εκ νέου ερμηνεία των διαγραφιών. Εν συνεχεία πραγματοποιούνται τα τεστ ροής (well tests) που στόχο έχουν την πληροφόρηση των μηχανικών πετρελαίου για τις παραμέτρους του ταμιευτήρα, των διαστάσεων του, και της παραγωγικής του ικανότητας ώστε να γίνει μια πρώτη εκτίμηση του κεφαλαίου που απαιτείται για την ανάπτυξη του κοιτάσματος.

II. Drill stem test και δειγματοληψίες

Το drill stem test (DST) δίνει την ευκαιρία διεξαγωγής των τεστ ροής ενώ ακόμα το φρέαρ δεν έχει συμπληρωθεί. Η συσκευή του DST πραγματοποιεί μια προσωρινή συμπλήρωση χρησιμοποιώντας το στέλεχος του γεωτρύπανου και επανακτίσιμα packers. Αντιπροσωπευτικό δείγμα υδρογονανθράκων μπορεί να ληφθεί κατά τη διάρκεια αυτού του τεστ καθώς επίσης και να αντληθούν πληροφορίες σχετικά με τον ταμιευτήρα συμπεριλαμβανομένων και της στατικής πίεσης του σχηματισμού καθώς και της πίεσης ροής. Για τον προσδιορισμό των εμπορικών ρυθμών παραγωγής του φρέατος, το DST, ανάλογα με την περίπτωση, μπορεί να διαρκέσει αρκετές ώρες ακόμα και μέρες. Το δείγμα ρευστού που λαμβάνεται κατά την διεξαγωγή του DST από την πρώτη ερευνητική γεώτρηση εν συνεχεία αποστέλλεται σε εξειδικευμένα εργαστήρια ώστε να μελετηθούν.

Οι δειγματοληψίες του ρευστού που πραγματοποιούνται κατά την διάρκεια του τεστ ροής μπορούν να κατηγοριοποιηθούν ως εξής:

1. Υπόγειες δειγματοληψίες όπου το δείγμα λαμβάνεται στο βάθος της γεώτρησης (subsurface or bottom-hole sampling).
2. Επιφανειακές δειγματοληψίες όπου δείγματα αερίου και υγρού λαμβάνονται σε συνθήκες διαχωριστήρα στην επιφάνεια (surface sampling).

Κατά την πρώτη μέθοδο, το δείγμα λαμβάνεται στις συνθήκες του ταμιευτήρα και χρησιμοποιείται ως έχει για την εργαστηριακή μελέτη της θερμοδυναμικής του συμπεριφοράς. Κατά την δεύτερη μέθοδο, τα δυο επιφανειακά δείγματα αναμειγνύονται στο εργαστήριο στην αναλογία του σταθεροποιημένου GOR που κατεγράφη κατά τη διάρκεια του τεστ, ώστε να προκύψει ένα αντιπροσωπευτικό δείγμα του ρευστού του ταμιευτήρα.

Η επιλογή του είδους δειγματοληψίας επηρεάζεται από:

1. Τον όγκο του ρευστού που απαιτείται για τις αναλύσεις
2. Τον τύπο του κοιτάσματος

III. Μηχανισμοί παραγωγής ταμιευτήρα πετρελαίου

Οι περισσότεροι ταμιευτήρες οι οποίοι ανιχνεύονται, αρχικά είναι μονοφασικοί και ακόρεστοι. Όταν η πίεση του ταμιευτήρα πετρελαίου πέσει κάτω από το σημείο φυσαλίδας-κορεσμού το απελευθερούμενο αέριο διαχωρίζεται από το αρχικό μονοφασικό ρευστό και αρχικώς παραμένει στους πόρους σε επαφή με το πετρέλαιο (με την υγρή φάση) με το οποίο έχει επέλθει ισορροπία με μια διεργασία του τύπου της στιγμιαίας εκτόνωσης (flash expansion process). Στα αρχικά στάδια η ροή πραγματοποιείται μόνο από την υγρή φάση.

Όταν αρκετό αέριο συσσωρευτεί αρχίζει να ρέει. Με κάθε περαιτέρω μείωση της πίεσης περισσότερο αέριο θα απομακρύνεται από την υγρή φάση. Επειδή το αέριο έχει μικρότερο ιξώδες και ρέει γρηγορότερα από την υγρή φάση η εξάτμιση γίνεται περισσότερο διαφορετική. Μια τέτοια διεργασία καταλήγει στο προοδευτικό εμπλουτισμό των βαρέων συστατικών της υγρής φάσης. Η εκτόνωση η οποία πραγματοποιείται καθώς το ρευστό ωθείται (λόγω διαφοράς πίεσης, ΔP) εντός της σωλήνωσης της γεώτρησης μπορεί να χαρακτηριστεί ως απευθείας εκτόνωση διότι η ανατάραξη της ροής διατηρεί τις δυο φάσεις σε επαφή μεταξύ τους. Επίσης, σαν απευθείας μπορεί να χαρακτηριστεί η εκτόνωση που πραγματοποιείται στον διαχωριστή όπου η απότομη πτώση της

πίεσης απελευθερώνει περισσότερο αέριο από το ρευστό καθώς οι δύο φάσεις παραμένουν σε επαφή.

Το σημείο φυσαλίδας-κορεσμού είναι σημαντικό για την κατανόηση της συμπεριφοράς του ρευστού του ταμιευτήρα και των μηχανισμών που το διέπουν. Για παράδειγμα:

- Εάν η αρχική πίεση του ταμιευτήρα είναι στο σημείο φυσαλίδας τότε υπάρχει ισχυρή πιθανότητα να υπάρχει υπερκείμενο στρώμα αερίου το οποίο κατά την εκμετάλλευση θα διασταλεί και αυτή η διαστολή θα οδηγήσει το πετρέλαιο προς την γεώτρηση αυξάνοντας την αποληψιμότητα ακόμα και κοντά στο 40%.
- Σε ένα ταμιευτήρα ο οποίος παράγει κάτω από το σημείο φυσαλίδας, το αέριο που εξέρχεται από το ρευστό είτε πάει στη γεώτρηση παραγωγής είτε μπορεί να υποβοηθήσει την παραγωγή πετρελαίου, εφόσον υπό την προϋπόθεση ύπαρξης κατακόρυφης διαπερατότητας, μπορεί να σχηματίσει ένα δευτερογενές υπερκείμενο στρώμα αερίου το οποίο βοηθάει στην ώθηση του πετρελαίου προς τις γεωτρήσεις παραγωγής. Αυτός ο μηχανισμός ονομάζεται διαχωρισμός βαρύτητας και μπορεί να αυξήσει την αποληψιμότητα κοντά στο 70%.
- Εάν η αρχική πίεση ενός ταμιευτήρα είναι πολύ μεγαλύτερη από το σημείο φυσαλίδας που σημαίνει ότι το πετρέλαιο είναι μάλλον χαμηλής συρρίκνωσης(χαμηλή περιεκτικότητα σε εν διαλύσει αέρια) τότε η μόνη δύναμη που επιδρά στην ώθηση του ρευστού είναι η μονοφασική διαστολή υγρού που έχει σαν αποτέλεσμα πολύ χαμηλή αποληψιμότητα. Για να βελτιωθεί αυτό χρησιμοποιούνται γεωτρήσεις εισπίεσης θαλασσινού νερού το οποίο δρα ως μηχανισμός ώθησης του ρευστού και μπορεί να οδηγήσει μέχρι 60% αποληψιμότητα.

Σημειώνεται ότι έχει γίνει θεώρηση ότι δεν υπάρχει μηχανισμός υποκείμενου υδροφόρου ορίζοντα.

Ο διαχωρισμός πετρελαίου και αερίου που λαμβάνει χώρα στον ταμιευτήρα αλλά και στις εγκαταστάσεις επεξεργασίας στην επιφάνεια είναι στην πράξη μια μη αντιστρεπτή (irreversible) διεργασία υπό την έννοια της θερμοδυναμικής. Κατά συνέπεια το τελικό αποτέλεσμα εξαρτάται σε ένα βαθμό από τις ταχύτητες από τις οποίες οι διεργασίες πραγματοποιούνται καθώς και από την επιλεγείσα πορεία (path) έστω και αν οι αρχικές και τελικές συνθήκες του συστήματος είναι οι ίδιες στην περίπτωση που απομακρύνεται ρευστό σε κάποιο από τα στάδια του διαχωρισμού.

The diagram illustrates the process of oil and gas separation and storage. It is divided into two main sections: a surface facility and a subsurface reservoir.

Surface Facility:

- Wellhead:** Oil produced from the reservoir enters the surface system through a wellhead with valves.
- Separator:** The oil enters a horizontal separator. The text "Oil Produced to Surface Undergoes Pressure and Temperature Reduction" is associated with this stage. The separator has a gas outlet at the top and an oil outlet at the bottom. The conditions at the separator are labeled p_{sep} and T_{sep} .
- Oil Storage:** Oil from the separator is collected in a vertical storage tank labeled "Oil".
- Stock Tank:** Oil from the storage tank flows into a larger "Stock Tank". The conditions at the stock tank are labeled p_{st} and T_{st} .
- Gas Release:** From the stock tank, "Released Gas" is vented to the atmosphere, and "Oil" is collected in another storage tank.
- Released and Greatly Expanded Gases:** A separate box labeled "Released and Greatly Expanded Gases" is shown above the oil storage tank, indicating the state of the gas after release.

Subsurface Reservoir:

- Reservoir:** The bottom section shows a cross-section of the reservoir. A well is shown with a "Reservoir Pressure Declines with Production" label.
- Oil and Gas Flow:** An arrow labeled Δp indicates the flow of "Oil" from a reservoir cylinder (labeled p_{res} and T_{res}) towards a "Released Gas" cylinder, representing the expansion of gas as it moves to the surface.

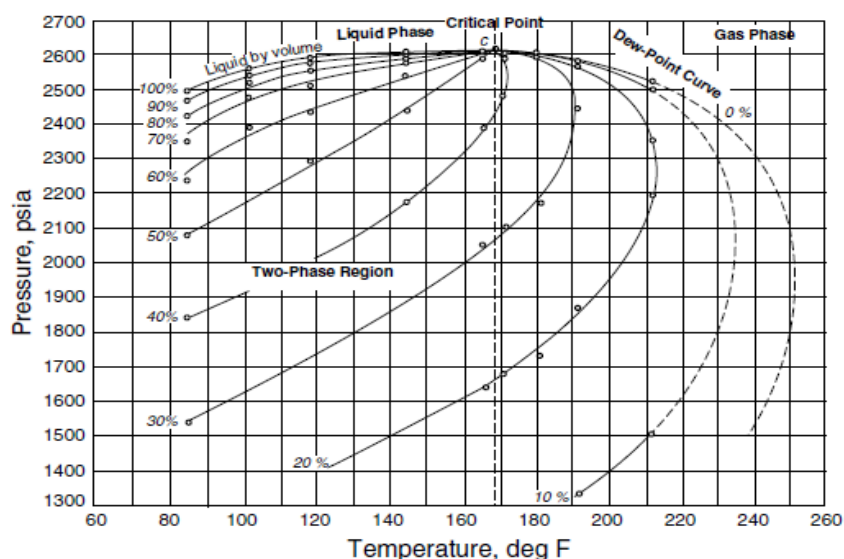
1. Τους ρυθμούς διαχωρισμού του αερίου
2. Το είδος των διεργασιών που λαμβάνουν χώρα στον ταμιευτήρα και στο φρέαρ
3. Τις συνθήκες και τον αριθμό των διαχωριστήρων

IV. Ισορροπία φάσεων

Τα μείγματα υδρογονανθράκων, και εν γένει οποιοδήποτε ρευστό μείγμα πολλών συστατικών, τείνουν να αλλάζουν την ισορροπία φάσεων στην οποία βρίσκονται όταν μεταβληθούν οι συνθήκες πίεσης και θερμοκρασίας. Η ισορροπία φάσεων ελέγχεται από δυο είδη διαμοριακών δυνάμεων, των ελκτικών και των απωστικών που ασκούνται επί των μορίων του δείγματος. Η ισορροπία αυτή καθορίζεται τόσο ποιοτικά όσο και ποσοτικά από τον συνολικό συσχετισμό όλων των αναπτυσσόμενων διαμοριακών δυνάμεων σε ένα σύστημα.

Για τη μελέτη της ισορροπίας φάσεων χρησιμοποιούνται ευρέως δισδιάστατα διαγράμματα 2 μεταβλητών. Ο ευρύτερα χρησιμοποιούμενος τύπος διαγράμματος στη Μηχανική Πετρελαίου είναι το διάγραμμα $P - T$ (πίεση – θερμοκρασία). Τα διαγράμματα αυτά χρησιμοποιούνται για την κατανόηση της συμπεριφοράς των θερμοδυναμικών συστημάτων κατά την μεταβολή των συνθηκών τους. Επίσης, τα διαγράμματα $P - T$ και οι φάκελοι φάσεων σε αυτά αποτελούν χρήσιμα εργαλεία για τον μηχανικό ταμειυτήρων.

Εικόνα 2.1 Απεικόνιση φακέλου φάσεων ενός μίγματος πολλών συστατικών μιας συγκεκριμένης συνολικής σύστασης.



Στην Εικόνα 2.1 παρουσιάζεται ένα τυπικό διάγραμμα πίεσης-θερμοκρασίας ενός μίγματος πολλών συστατικών μιας συγκεκριμένης συνολικής σύστασης. Παρά το γεγονός ότι κάθε διαφορετικό μίγμα υδρογονανθράκων έχει διαφορετικό διάγραμμα φάσης, η γενική θεώρηση είναι παρόμοια. Αυτά τα σύνθετα διαγράμματα πίεσης-θερμοκρασίας χρησιμοποιούνται για:

- Την ταξινόμηση των ταμειυτήρων
- Την περιγραφή της θερμοδυναμικής συμπεριφοράς του ρευστού του ταμειυτήρα

Για να γίνει κατανοητή πλήρως η σημασία των διαγραμμάτων πίεσης-θερμοκρασίας, είναι απαραίτητο να προσδιοριστούν και να καθοριστούν τα ακόλουθα βασικά χαρακτηριστικά τους:

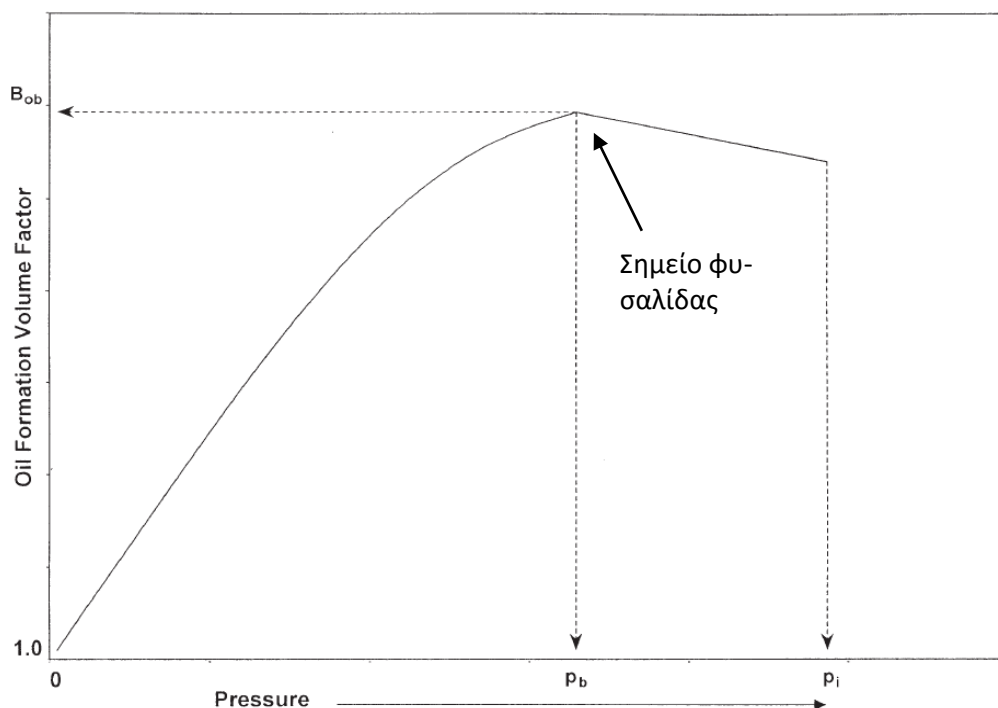
- **Cricondentherm** : ορίζεται σαν η πιο υψηλή θερμοκρασία για την οποία δύναται να υπάρξει διφασική ισορροπία.
- **Cricondenbar** : ορίζεται σαν η μεγαλύτερη πίεση κάτω από την οποία υγρή και αέρια φάση μπορούν να συνυπάρχουν.
- **Κρίσιμο σημείο (critical point)**: Ως κρίσιμο σημείο αναφέρεται η πίεση και η θερμοκρασία υπό τις οποίες όλες οι εντατικές ιδιότητες της αέριας και υγρής φάσης είναι ίσες. Είναι επίσης το σημείο συνάντησης των δυο γεωγραφικών τόπων : των σημείων φουσαλίδας – δρόσου.
- **Φάκελος φάσεων** (διφασική περιοχή) : Αναφέρεται στη περιοχή που περικλείεται από τις καμπύλες των σημείων φουσαλίδας και δρόσου. Σε αυτή τη περιοχή, η αέρια φάση με την υγρή φάση συνυπάρχουν σε ισορροπία και αποτελούν τον φάκελο φάσεως του μίγματος υδρογονανθράκων.
- **Γραμμές ποιότητας (quality lines)**: Οι γραμμές ποιότητας προσδιορίζουν τους γεωμετρικούς τόπους συγκεκριμένων τιμών όγκου της υγρής φάσης σε διάφορες θερμοκρασίες και πιέσεις.
- **Καμπύλη σημείων φουσαλίδας (κορεσμού) (bubble-point curve)**: Από θερμοδυναμικής απόψεως, το σημείο φουσαλίδας αναφέρεται στη πίεση (σε δεδομένη θερμοκρασία), όπου η πρώτη φουσαλίδα αερίου σχηματίζεται στο μίγμα του υγρού. Το σύνολο των σημείων φουσαλίδας για κάθε πίεση και θερμοκρασία αποτελεί την καμπύλη των σημείων φουσαλίδας η οποία χωρίζει την περιοχή της υγρής φάσης από την διφασική περιοχή.
- **Καμπύλη σημείων δρόσου (dew-point curve)** : Από θερμοδυναμικής απόψεως, το σημείο δρόσου αναφέρεται στη πίεση (σε δεδομένη θερμοκρασία), όπου η πρώτη σταγόνα υγρού σχηματίζεται στο μίγμα αερίου. Το σύνολο των σημείων δρόσου για κάθε πίεση και θερμοκρασία αποτελεί την καμπύλη των σημείων δρόσου η οποία χωρίζει την περιοχή της αέριας φάσης από την διφασική περιοχή.

V. Ορισμός των ογκομετρικών συντελεστών υγρής και αέριας φάσης

Χρησιμοποιώντας τους ογκομετρικούς συντελεστές υγρής και αέριας φάσης οι οποίοι μετρούνται στο εργαστήριο ή προσδιορίζονται υπολογιστικά, ο μηχανικός πετρελαίου μπορεί να περιγράψει την ποσοτική μεταβολή του όγκου που οι διάφορες φάσεις των ρευστών του ταμειυτήρα υφίστανται κατά την διάρκεια την παραγωγής. Επίσης, χρησιμοποιούνται και για την εκτίμηση αποθεμάτων υδρογονανθράκων του εκάστοτε ταμειυτήρα. Εν συνεχεία, θα παρουσιαστούν οι ογκομετρικοί συντελεστές και οι ορισμοί τους.

I. Ογκομετρικός συντελεστής σχηματισμού υγρής φάσεως B_o

Σαν ογκομετρικός συντελεστής σχηματισμού υγρής φάσεως (oil formation volume factor) ορίζεται ο λόγος του όγκου που ένα υγρό μίγμα (συμπεριλαμβανομένων και των αερίων εν διάλυση) καταλαμβάνει στις συνθήκες του σχηματισμού προς τον όγκο του υγρού που θα απομείνει αφού το μίγμα ισορροπήσει σε κανονικές συνθήκες στην επιφάνεια (τα εν διάλυση αέρια έχουν απελευθερωθεί). Ο συντελεστής αυτός συμβολίζεται με B_o ή FVF και λαμβάνει τιμές πάντα μεγαλύτερες από την μονάδα.

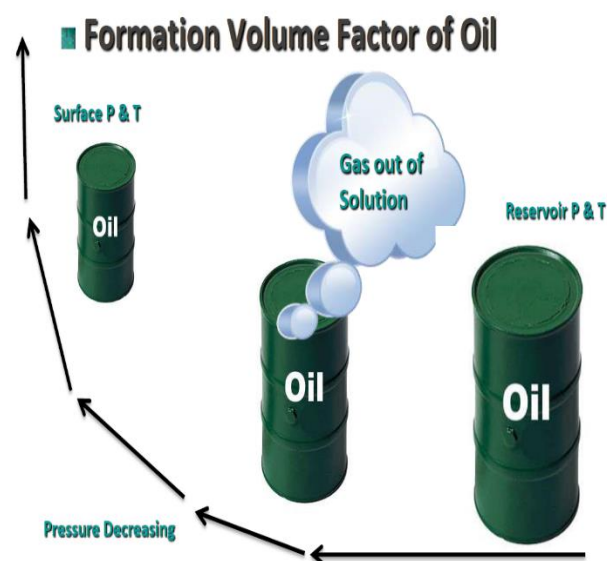


Διάγραμμα 2.1 Ογκομετρικός συντελεστής υγρής φάσεως σε συνάρτηση της πίεσης. Δεξιά του σημείου φυσαλίδας μειούμενης της πίεσης του ταμειυτήρα αυξάνεται η τιμή του συντελεστή γιατί ακολουθεί την συμπίεστικότητα της υγρής φάσης. Αριστερά του σημείου φυσαλίδας έχουμε μείωση της τιμής του συντελεστή λόγω της απελευθέρωσης των εν διάλυση αερίων.

Σαν συνθήκες επιφανείας ορίζονται τα 14.7 psi και 60⁰ F ή 1 ατμόσφαιρα(atmosphere) και 15.5⁰ C. Σαν μονάδες μέτρησης το Bo έχει [res bbl/STB or res m³/stock-tank m³] που σημαίνουν, βαρέλι ταμειυτήρα/βαρέλι επιφανείας ή κυβικό μέτρο ταμειυτήρα ανά κυβικό μέτρο επιφανείας.

$$Bo = \frac{\text{όγκος πετρελαίου ταμειυτήρα}}{\text{όγκος πετρελαίου επιφανείας}}$$

$$Bo = \frac{Vo_{res}}{Vo_{sc}}$$



Εικόνα 2.4 Γραφιστική απεικόνιση της μεταβολής του όγκου ενός βαρελιού από τον ταμειυτήρα στην επιφάνεια.

Για να γίνει αντιληπτή η ογκομετρική μεταβολή χρησιμοποιείται το εξής παράδειγμα. Έστω ότι δίνεται η τιμή FVF ή Bo = 1.310 τότε για κάθε 1.310 βαρέλι ρευστού σε πίεση και θερμοκρασία ταμειυτήρα παράγεται 1 βαρέλι στην επιφάνεια. Αυτό σημαίνει ότι μόνο το 76.3% του αρχικού ρευστού θα φθάσει στο διαχωριστήρα ή στο τανκ. Αυτό το ποσοστό εκφρασμένο σε κλάσμα 0.763 καλείται συντελεστής συρρίκνωσης(*shrinkage factor*) του πετρελαίου και ισούται με 1/Bo.

Για την εκτίμηση αποθεμάτων υδρογονανθράκων με χρήση της τιμής του Bo χρησιμοποιείται μια ογκομετρική προσέγγιση όπου οι τιμές λαμβάνονται ύστερα από γεωλογικές αναλύσεις, διαγραφές και εργαστηριακές αναλύσεις πυρήνων ταμειυτήρα. Η εξίσωση είναι :

$$STOIIP = \frac{V\Phi(1-S_w)}{Bo}$$

Όπου : V = ο όγκος ταμειυτήρα

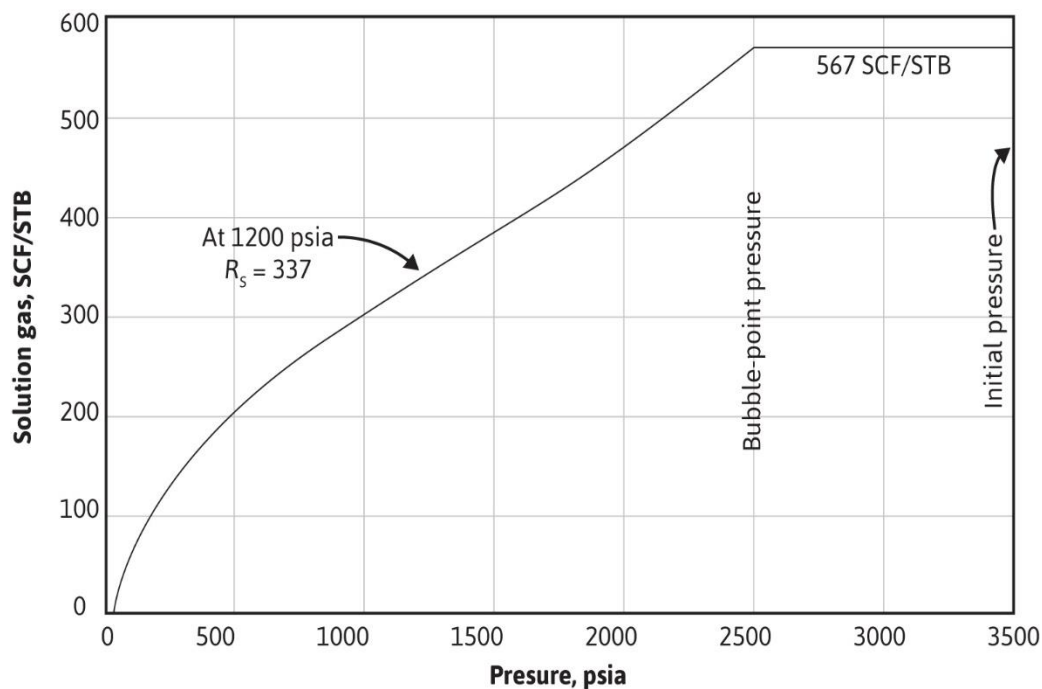
Φ = το μέσο πορώδες του πετρώματος

Sw = κορεσμός σε νερό

II. Λόγος αερίου εν διαλύσει/πετρελαίου R_s

Σαν λόγος αερίου εν διαλύσει/πετρελαίου (solution gas-oil ratio) R_s , σε συνθήκες πίεσης και θερμοκρασίας, ορίζεται το κλάσμα το οποίο έχει αριθμητή τον ολικό όγκο του αερίου που παραμένει διαλελυμένο στο πετρέλαιο στις συνθήκες p, T , υπολογισμένο στις κανονικές συνθήκες, δια του όγκου του υγρού το οποίο θα απομείνει όταν το πετρέλαιο θα παραχθεί σε συνθήκες επιφανείας (STO). Η τιμή R_s δίνει το μέτρο της πηχτικότητας ενός πετρελαίου.

Σαν συνθήκες επιφανείας ορίζονται 14.7 psi και 60° F ή 1 ατμόσφαιρα και 15.5° C. Σαν μονάδες μέτρησης το R_s έχει [scf/STB or $\text{sm}^3/\text{stock-tank m}^3$] που σημαίνουν, κυβικά πόδια/βαρέλι επιφανείας ή κυβικό μέτρο ανά κυβικό μέτρο επιφανείας.



Διάγραμμα 2.2 Λόγος αερίου εν διαλύσει/υγρού R_s σαν συνάρτηση της πίεσης. Δεξιά του σημείου φυσαλίδας έχουμε ακόρεστο κοίτασμα (μονοφασικό) και όλο το αέριο είναι εν διαλύσει. Αριστερά του σημείου φυσαλίδας το R_s βαίνει μειούμενο γιατί το αέριο εν διαλύσει μειώνεται λόγω της αύξησης της εξάτμισης. (Σημείο φυσαλίδας = bubble-point pressure)

$$R_s = \frac{\text{όγκος εν διαλύσει αερίου στην επιφάνεια}}{\text{όγκος πετρελαίου επιφάνειας}}$$

$$R_s = \frac{V_{gsc}}{V_{oSTO}}$$

Για την εκτίμηση της ποσότητας του εν διαλύσει αερίου που μπορεί να παραχθεί στην επιφάνεια, μετρημένου σε scf(κυβικά πόδια), γίνεται χρήση της τιμής του Rs για την οποία χρησιμοποιείται μια ογκομετρική προσέγγιση όπου οι τιμές λαμβάνονται ύστερα από γεωλογικές αναλύσεις, διαγραφίες και εργαστηριακές αναλύσεις πυρήνων ταμιευτήρα η εξίσωση είναι :

$$STGIIP_{SOLUTION} = \frac{V\phi(1-S_w)Rs}{Bo}$$

Όπου : V = ο όγκος ταμιευτήρα

Φ = το μέσο πορώδες του πετρώματος

Sw = κορεσμός σε νερό

III. Λόγος αερίου/πετρελαίου GOR

Σαν λόγος αερίου/πετρελαίου GOR ορίζεται το κλάσμα που έχει αριθμητή τον όγκο του παραχθέντος αερίου μετρημένο υπό κανονικές συνθήκες και παρονομαστή τον όγκο του παραχθέντος πετρελαίου μετρημένο και αυτόν υπό κανονικές συνθήκες. Σημειώνεται ότι στην περίπτωση αυτή σαν αέριο εκλαμβάνεται το σύνολο του εκλυθέντος αερίου τόσο το εν διαλύσει σε συνθήκες κοιτάσματος όσο και αυτό που ευρίσκεται σε ελεύθερη μορφή.

$$GOR = \frac{\text{όγκος συνολικού αερίου στην επιφάνεια}}{\text{όγκος πετρελαίου επιφάνειας}} \quad GOR = \frac{Vg sc}{Vo sto}$$

Μερικές φορές και ιδίως στις αναφορές των τεστ ροής αναφέρεται και ο λόγος αερίου/υγρού (GLR). Στις περιπτώσεις αυτές θα πρέπει να ελέγχονται οι μονάδες διότι συνήθως το GLR υπονοεί τον λόγο του όγκου του αερίου, μετρούμενο όπως προηγουμένως, ως προς τον όγκο του υγρού αυτή τη φορά μετρούμενο όμως στις συνθήκες που επικράτησαν κατά το διαχωρισμό.

- Τιμές πάνω από 5000 GOR συναντάμε στα αέρια
- Τιμές κάτω από 5000 GOR συναντάμε στο υγρό πετρέλαιο
- Σε μονοφασικά κοιτάσματα πάνω από το σημείο φουσαλίδας GOR=RS

IV. Ογκομετρικός συντελεστής σχηματισμού αέριας φάσης Bg

Ο ογκομετρικός συντελεστής σχηματισμού αέριας φάσης (gas formation volume factor) ή αλλιώς Bg ορίζεται σαν τον όγκο σε συνθήκες ταμιευτήρα που καταλαμβάνεται από μια μονάδα όγκου αερίου που βρίσκεται υπό κανονικές συνθήκες.

Παρόλο που εξορισμού το Bg είναι αδιάστατο, είναι σύνηθες να χρησιμοποιείται το Πρακτικό Σύστημα μονάδων και να εκφράζονται οι δυο όγκοι με διαφορετικές μονάδες και πιο συγκεκριμένα με bbl τον όγκο του αερίου σε συνθήκες ταμιευτήρα και με scf τον όγκο του υπό κανονικές συνθήκες.

$$\mathbf{Bg} = \frac{\text{όγκος αερίου συνθήκες ταμιευτήρα}}{\text{όγκος αερίου συνθήκες επιφανείας}} \quad \mathbf{Bg} = \frac{Vg_{res}}{Vg_{sto}}$$

$Bg = f(p, T)$, Συνάρτηση δηλαδή της πίεσης και της Θερμοκρασίας

Χρησιμοποιώντας δυο καταστατικές εξισώσεις πραγματικών αερίων, μια για συνθήκες ταμιευτήρα και μια για συνθήκες επιφανείας και αφού διαιρεθούν μεταξύ τους για να παραχθεί ο τύπος του Bg σε συνάρτηση P,T εύκολα αποδεικνύονται τα παρακάτω:

- $P Vg = n z R T$ συνθήκες ταμιευτήρα
- $P_{sc} Vg_{sto} = n z R T_{sc}$ συνθήκες επιφανείας ($z \cong 1$)

$$\frac{P}{P_{sc}} \frac{Vg_{res}}{Vg_{sto}} = \frac{\cancel{n} z \cancel{R} T}{\cancel{n} \cancel{R} T_{sc}} \Rightarrow Bg = \frac{P_{sc}}{T_{sc}} Z \frac{T}{P} \quad \text{Άρα :}$$

$$Bg = 0.00504 * z * \frac{P}{T} \text{ σε res.bbl/scf} \quad (p = \text{psia} , T = \text{Rankin}) \quad \text{ή}$$

$$Bg = 0.0282 * z * \frac{P}{T} \text{ σε ft}^3/\text{scf} \quad (p = \text{psia} , T = \text{Rankin}) \quad \text{ή}$$

$$Bg = 0.3495 * z * \frac{P}{T} \text{ σε m}^3/\text{sm}^3 \quad (p = \text{Kpa} , T = \text{Kelvin}) \quad \text{ή}$$

$$Bg = 0.0346 * z * \frac{P}{T} \text{ σε m}^3/\text{sm}^3 \quad (p = \text{atm} , T = \text{Kelvin})$$

Προσοχή πρέπει να δίδεται επίσης α , σαν αποτέλεσμα συμπύκνωσης μέρος του αερίου εντός του φρέατος της γεώτρησης ή στο σύστημα διαχωρισμού, η σύσταση του αερίου στην επιφάνεια είναι διαφορετική από αυτήν στον ταμιευτήρα. Στην περίπτωση αυτή ο όγκος του συμπυκνώματος θα πρέπει να ληφθεί υπόψιν για τον υπολογισμό του όγκο σε συνθήκες κοιτάσματος.

Όταν έχουμε υγρό αέριο δεν μπορούμε να απλοποιήσουμε στις καταστατικές τα mol για το λόγω του ότι ένα μέρος των mol του αερίου του ταμιευτήρα μετατρέπονται σε mol αερίου συμπυκνώματος.

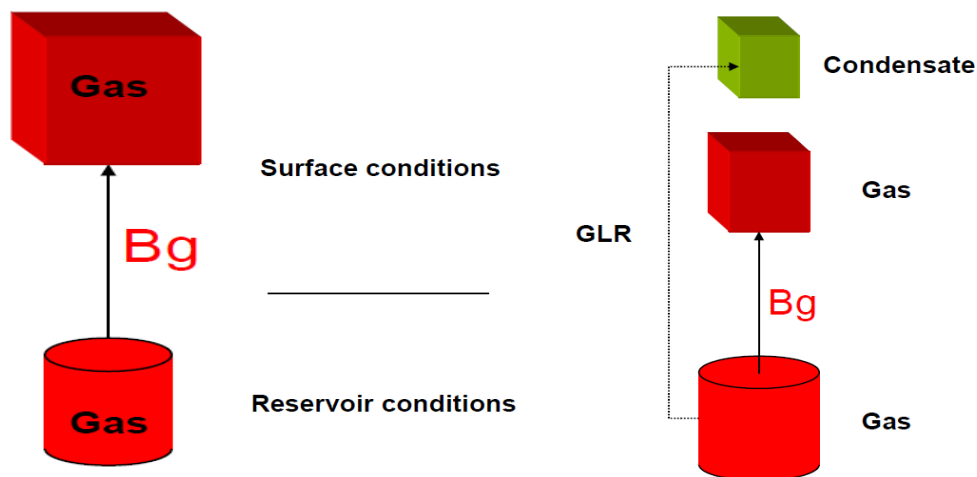
Για την εκτίμηση του όγκου του φυσικού αέριου σε συνθήκες επιφανείας του ταμιευτήρα γίνεται χρήση της τιμής του B_g χρησιμοποιείται η κατωτέρω ογκομετρική προσέγγιση όπου οι τιμές λαμβάνονται ύστερα από γεωλογικές αναλύσεις, διαγραφίες και εργαστηριακές αναλύσεις πυρήνων ταμιευτήρα. ο τύπος είναι :

$$STGIIP = \frac{V\phi(1-S_w)}{B_g}$$

Όπου : V = ο όγκος ταμιευτήρα

ϕ = το μέσο πορώδες του πετρώματος

S_w = κορεσμός σε νερό



Σχήμα 1.16 Το αριστερό σχήμα απεικονίζει το B_g γραφικά ενός ταμιευτήρα(χωρίς αέριο συμπύκνωμα) και δεξιά έχουμε την γραφική απεικόνιση του B_g ενός ταμιευτήρα αέριου συμπυκνώματος(gas Condensate) που θα πρέπει να ληφθεί υπόψιν για τον υπολογισμό του όγκο σε συνθήκες κοιτάσματος.

Ολικός ογκομετρικός συντελεστής σχηματισμού B_t

Σαν ολικός ογκομετρικός συντελεστής σχηματισμού (total formation volume factor) B_t ορίζεται ο λόγος που έχει αριθμητή τον ολικό όγκο του ρευστού, δηλαδή τον όγκο του υγρού στις συνθήκες ταμιευτήρα p και T συν τον όγκο που καταλαμβάνει στις ίδιες συνθήκες σαν αέρια φάση το αέριο το οποίο όταν το κοίτασμα βρισκόταν σε συνθήκες κορεσμού ήταν διαλελυμένο στο υγρό και παρονομαστή τον όγκο που το ρευστό σε συνθήκες p και T θα αποδώσει όταν παραχθεί σε συνθήκες επιφανείας.

$$B_t = \frac{\text{όγκος πετρελαίου} + \text{αερίου σε συνθήκες ταμειυτήρα}}{\text{όγκος πετρελαίου συνθήκες επιφανείας}}$$

$$B_t = \frac{V_o \text{ res} + V_g \text{ res}}{V_o \text{ sto}}$$

Επίσης, ο ολικός ογκομετρικός συντελεστής σε συσχετισμό με B_o και B_g ισούται με :

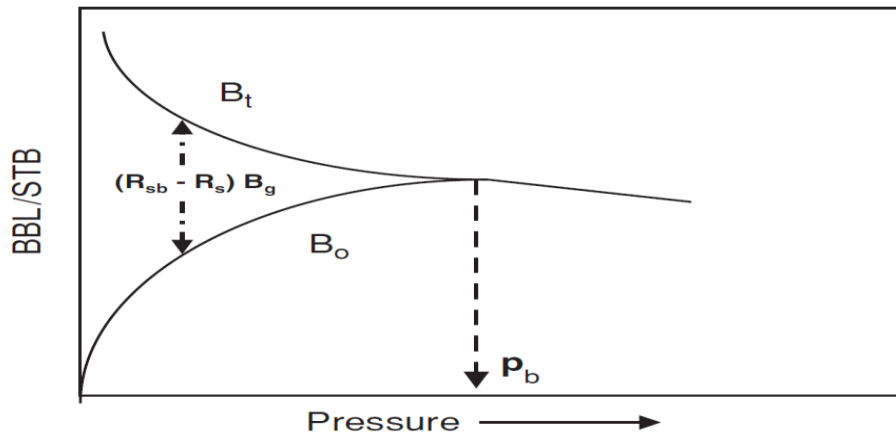
$$B_t = \frac{V_o \text{ res} + V_g \text{ res}}{V_o \text{ sto}} = \frac{V_o \text{ res}}{V_o \text{ sto}} + \frac{V_g \text{ res}}{V_o \text{ sto}} = B_o + B_g * \frac{V_g \text{ sto}}{V_o \text{ sto}} = B_o + \frac{B_g * (R_{sb} - R_s) * V_o \text{ sto}}{V_o \text{ sto}}$$

$$\Rightarrow B_t = B_o + B_g * (R_{sb} - R_s)$$

Όπου R_s = ο λόγος αερίου εν διαλύσει/πετρελαίου μετά το σημείο φυσαλίδας

R_{sb} = ο λόγος αερίου εν διαλύσει/πετρελαίου πριν τον σημείο φυσαλίδας

Αυτός ο τύπος ισχύει για ταμειυτήρες πετρελαίου που έχουν διφασική ροή στις συνθήκες παραγωγής και άντλησης.



Σχήμα 1.17 Διάγραμμα το οποίο απεικονίζει την μεταβολή του B_o και B_t σε συνάρτηση της πίεσης. Μέχρι και πριν το σημείο φυσαλίδας από το δεξιά προς αριστερά το $B_t = B_o$ και ακολουθεί την συμπίεσιτότητα της υγρής φάσης. Αυτό συμβαίνει γιατί δεν υπάρχει παραγωγή αερίου εν διαλύσει οπότε με μείωση της πίεσης αυξάνει σχετικά ο όγκος του υγρού. Μετά το σημείο φυσαλίδας το B_t αυξάνεται μειούμενης της πίεσης γιατί υπάρχει αύξηση του όγκου της αέριας φάσης.

Πηγή διαγραμμάτων (TaReK AHMED "Reservoir Engineering Handbook")

VI. Σημασία των PVT μελετών

Για των υπολογισμό των ογκομετρικών συντελεστών απαιτείται μια σειρά πειραματικών μετρήσεων οι οποίες και ονομάζονται PVT (πίεση-όγκος-θερμοκρασίας) μελέτες. Οι μηχανικοί που ειδικεύονται στη διαχείριση των ταμιευτήρων (reservoir engineers) και της παραγωγής (production engineers) αξιοποιώντας τα δεδομένα που προκύπτουν από αυτές τις μελέτες είναι σε θέση να απαντήσουν ερωτήματα που σχετίζονται μεταξύ άλλων με :

- ❖ με το μέγεθος του ταμιευτήρα
- ❖ με τον όγκο της υγρής και αέριας φάσης που θα παραχθεί στην επιφάνεια
- ❖ με την χημική σύσταση
- ❖ με την ποιότητα του πετρελαίου
- ❖ με την πίεση στην οποία ο διαχωριστήρας βελτιστοποιεί την ανάκτηση της υγρής φάσης
- ❖ με την αθροιστική ανάκτηση
- ❖ με το αν περιέχει ή όχι το ρευστό μη επιθυμητά στοιχεία όπως θείο το οποίο προκαλεί προβλήματα διάβρωσης
- ❖ με το ποιά θα μπορούσε να είναι η διαχείριση του παραγόμενου αερίου: επανεισαγωγή για τεχνητή ανύψωση ρευστού (artificial lift), καύση ή πώληση
- ❖ με πρόβλημα κατακάθισης ασφατενίων
- ❖ με πρόβλημα δημιουργίας υδριτών

Ο στόχος της απάντησης αυτών των ερωτημάτων αλλά και της περαιτέρω αξιοποίησης των πληροφοριών είναι η μεγιστοποίηση της αποληψιμότητας υδρογονανθράκων στον ελάχιστο δυνατό χρόνο και με το ελάχιστο κόστος. Εν συνεχεία, στο επόμενο κεφάλαιο θα παρουσιαστούν αναλυτικά οι PVT μελέτες.

Κεφάλαιο 3

ΕΡΓΑΣΤΗΡΙΑΚΕΣ ΜΕΛΕΤΕΣ ΠΕΤΡΕΛΑΙΩΝ

I. PVT ΜΕΛΕΤΕΣ ΠΕΤΡΕΛΑΙΩΝ

Οι εργαστηριακές μελέτες για τη μέτρηση των τιμών των ιδιοτήτων των ρευστών των ταμιευτήρων πραγματοποιούνται είτε σε εξειδικευμένα εργαστήρια της εταιρείας πετρελαίου είτε ανατίθενται εργολαβικά σε εταιρείες του κλάδου παροχής υπηρεσιών ή πανεπιστήμια. Λόγω του ότι για τη πραγματοποίηση των μετρήσεων απαιτείται χρήση υψηλής εξειδίκευσης προσωπικού και εξοπλισμού μεγάλου κόστους συνήθως από την βιομηχανία πετρελαίου επιλέγεται το δεύτερο.

Οι βασικές μελέτες PVT για τα πετρέλαια είναι :

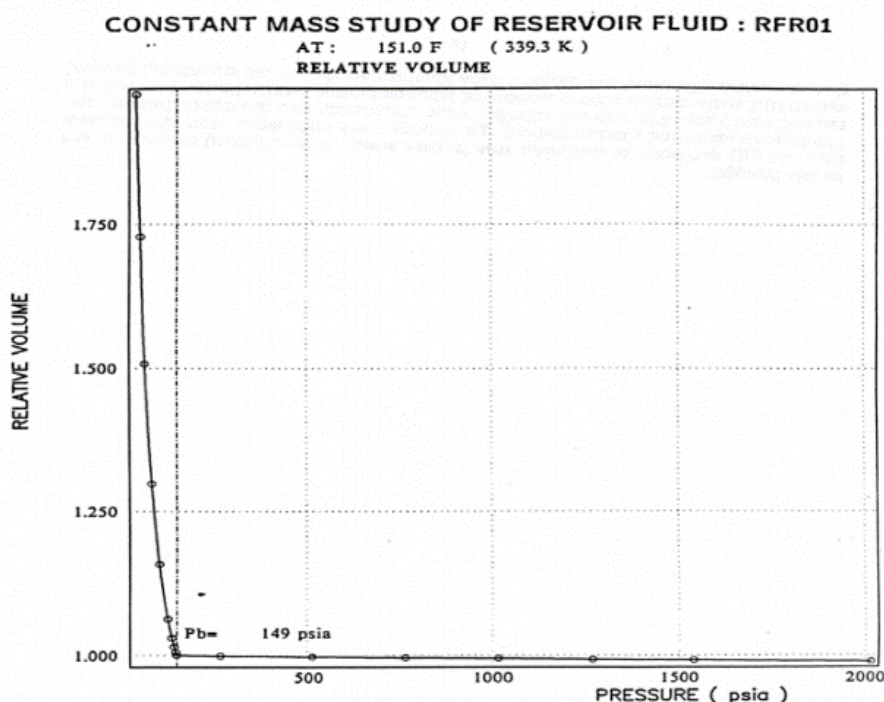
- Μελέτη Σταθερής Μάζας (Constant Mass Study)
- Μελέτη Διαφορικής Εξάτμισης (Differential Vaporization)
- Μελέτη αριστοποίησης των συνθηκών Διαχωρισμού (Separation test)

Μελέτη σταθερής μάζας (constant mass)

Η μελέτη σταθερής μάζας χρησιμοποιείται για τον προσδιορισμό της πίεσης στο σημείο φυσαλίδας-κορεσμού όπου το ρευστό μεταβαίνει από τη μονοφασική περιοχή στη διφασική περιοχή λόγω της πτώσης πίεσης, όπως επίσης και για το καθορισμό της ισοθερμοκρασιακής συμπεριφοράς της υγρής φάσης και του σχετικού της όγκου σε θερμοκρασία ταμιευτήρα. Επιπρόσθετως, είναι δυνατή και η μελέτη της ογκομετρικής συμπεριφοράς των δύο φάσεων κάτω από το σημείο φυσαλίδας-κορεσμού.

Για την εκτέλεση της μελέτης, εισάγεται το δείγμα σε ένα οπτικό κελί με προδιαγραφές υψηλών πιέσεων και εν συνεχεία θερμαίνεται σε θερμοκρασία ταμιευτήρα. Το δείγμα ύστερα συμπιέζεται από ένα εμβολοφόρο κύλινδρο ο οποίος ωθείται από μια αντλία νερού μέχρι την αρχική πίεση του ταμιευτήρα όπου και μετρείται ο όγκος του. Ακολούθως, η πίεση μειώνεται σταδιακά μέχρι το σημείο φυσαλίδας-κορεσμού το οποίο και υπολογίζεται από το παραγόμενο διάγραμμα σχετικού όγκου - πίεσης της μελέτης. Επίσης, σε αυτό το στάδιο της μελέτης υπολογίζεται η συμπεριφορά του ρευστού στη μονοφασική περιοχή.

Αρχικά, το μονοφασικό ρευστό του πετρελαίου του δείγματος δεν παρουσιάζει μεγάλες μεταβολές μειούμενης της πίεσης στη μέτρηση του σχετικού του όγκου. Όταν εισέλθει στη διφασική περιοχή και υπάρξει απελευθέρωση των εν διαλύσει αερίων οι μεταβολές, μειούμενης της πίεσης, του σχετικού όγκου είναι σημαντικές. Το σημείο της πίεσης στο οποίο παρατηρείται αυτή η σημαντική μεταβολή στο διάγραμμα $p - T$ αποτελεί και το σημείο φυσαλίδας-κορεσμού.



Διάγραμμα 3.1 όπου απεικονίζεται το διάγραμμα $p - T$ μια μελέτης σταθερής μάζας για το προσδιορισμό του σημείου φυσαλίδας-κορεσμού. Το σημείο καμπής από το σημείο. Πηγή(Σημειώσεις μηχανικής Ταμειυτήρων)

Η μελέτη σταθερής μάζας συνεχίζεται και κάτω από το σημείο φυσαλίδας-κορεσμού έως ότου η πίεση μειωθεί κατά στάδια μέχρι μια τιμή που εκτιμάται ότι θα είναι χαμηλότερη από την τελική πίεση που θα επικρατεί στο ταμειυτήρα την στιγμή που θα σταματήσει η παραγωγή. Σε αυτό το στάδιο της μελέτης, μετριέται η συμπιεστότητα του διφασικού ρευστού σε πιέσεις εντός του φακέλου φάσεων χωρίς να έχει πραγματοποιηθεί καμία απομάκρυνση της αέριας ή υγρής φάσης. Το δείγμα αναταράσσεται αρκετά κατά τη διάρκεια του πειράματος, ώστε να εξασφαλιστεί η θερμοδυναμική ισορροπία. Μετά την ισορροπία του διφασικού πλέον μείγματος μετριέται ο συνολικός όγκος του υγρού και του αερίου. Με αυτό τον τρόπο μπορεί να γίνει σύγκριση των όγκων που προκύπτουν σε κάθε πίεση ως προς τον όγκο του υγρού στην πίεση του σημείου φυσαλίδας-κορεσμού που εκφράζεται από το σχετικό όγκο.

$$RV = \frac{V_i}{V_{pb}}$$

Όπου:

- RV = ο σχετικός όγκος
- V_i = ο όγκος που προκύπτει σε κάθε πίεση
- V_{pb} ο όγκος του υγρού στη πίεση του σημείο φυσαλίδας-κορεσμού

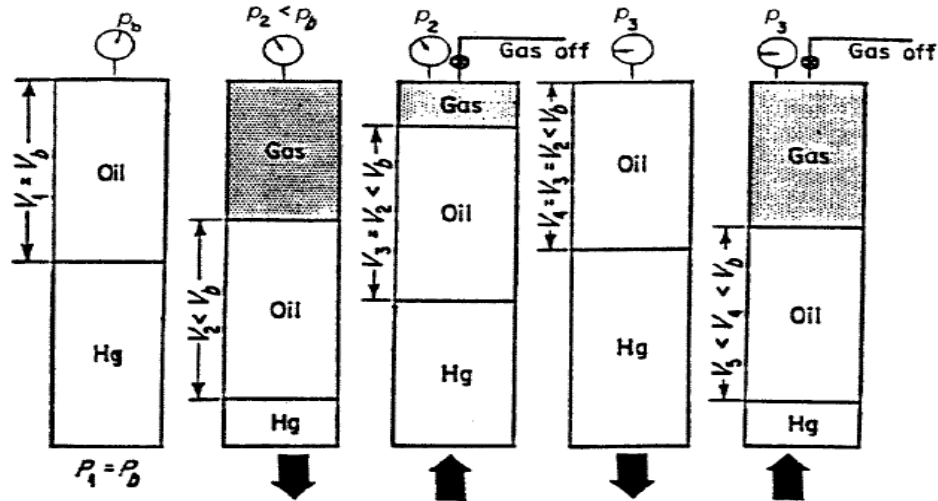
Μελέτη Διαφορικής Εξάτμισης

Η μελέτη της διαφορικής εξάτμισης θεωρείται ότι περιγράφει καλύτερα τις μεταβολές της θερμοδυναμικής ισορροπίας που λαμβάνουν χώρα μέσα στον ταμιευτήρα κατά την διάρκεια της παραγωγής κάτω από το σημείο φυσαλίδας-κορεσμού του ρευστού, καθώς για λόγους διαφορετικής κινητικότητας η αέρια φάση συνεχώς απομακρύνεται από την υγρή φάση με την οποία έχει αποκαταστήσει προηγουμένως ισορροπία. Κατά συνέπεια, αυτή η συμπεριφορά ακολουθεί τα χαρακτηριστικά της διαφορικής εξάτμισης.

Για τη μελέτη της διαφορικής εξάτμισης, χρησιμοποιείται το δείγμα πετρελαίου που χρησιμοποιήθηκε για τη μελέτη σταθερής μάζας όπου και επανασυμπιέζεται στην αρχική πίεση και θερμοκρασία του ταμιευτήρα και εν συνεχεία αναταράσσεται ώστε να εξασφαλιστεί η πλήρης επαναδιάλυση του αερίου στο υγρό. Εφόσον έχει εξασφαλιστεί η επανδιάλυση του αερίου, πραγματοποιείται μείωση της πίεσης μέχρι το σημείο φυσαλίδας-κορεσμού που έχει υπολογιστεί από την μελέτη σταθερής μάζας έτσι ώστε να αρχίσει η μελέτη της διαφορικής εξάτμισης.

Όπως παρουσιάζεται γραφικά στην εικόνα 3.1, η πίεση από το σημείο φυσαλίδας-κορεσμού μειώνεται σταδιακά σε βήματα μέχρι τις ατμοσφαιρικές συνθήκες, συνήθως σε 10 με 15 βήματα, και το αέριο που απελευθερώνεται σε κάθε βήμα μετρίεται τόσο υπό τις επικρατούσες συνθήκες πίεσης και θερμοκρασίας του εκάστοτε βήματος όσο και σε συνθήκες επιφανείας. Επίσης ένα μικρό δείγμα του κάθε απελευθερωμένου αερίου εισάγεται στο χρωματογράφο για τη μέτρηση της σύστασής του.

Ο όγκος του υγρού που παραμένει, V_L , επίσης μετρίεται σε κάθε βήμα της πίεσης. Πρέπει να σημειωθεί ότι η υγρή φάση που απομένει σε κάθε βήμα υπόκειται σε συνεχείς αλλαγές ως προς τη σύνθεση της και προοδευτικά γίνεται πλουσιότερη σε πιο βαριά συστατικά.



Εικόνα 3.1 Γραφική απεικόνιση της διαφορικής εξάτμισης

Κατά τη μελέτη της **διαφορικής εξάτμισης**, θα πρέπει να σημειωθεί ότι το εν διαλύσει αέριο που απελευθερώνεται από το δείγμα του ρευστού λόγω της μειούμενης πίεσης απομακρύνεται συνεχώς από το αυτόκλειστο κελί μετά την αποκατάσταση ισορροπίας με την υγρή φάση. Αυτού του τύπου η εξάτμιση χαρακτηρίζεται από την συνεχή μεταβολή της σύστασης των υγρών υδρογονανθράκων. Σημειώνεται ότι η μελέτη της διαφορικής εξάτμισης λαμβάνει χώρα για όλα τα βήματα σε θερμοκρασία ταμειυτήρα. Στο τελευταίο βήμα μετριέται όπου ο όγκος της υγρής φάσης που έχει απομείνει σε ατμοσφαιρική πίεση και 60 ° F.

Η σύσταση του τελικού υγρού εξαρτάται από τον αριθμό των σταδίων που ακολουθήθηκαν, καθότι εφόσον απομακρύνεται στα διάφορα βήματα μάζα από το σύστημα, η διεργασία καθίσταται μη αντιστρεπτή.

Τα πειραματικά δεδομένα που λαμβάνονται από αυτή τη πειραματική διαδικασία περιλαμβάνουν:

- ❖ Τους ογκομετρικούς συντελεστές σχηματισμού της υγρής και αέριας φάσης B_o και B_g

Οι ογκομετρικοί συντελεστές υγρής φάσης B_{od} σε όλα τα βήματα πίεσης υπολογίζονται από τον τύπο :

$$B_{od} = \frac{V_L}{V_{SC}}$$

Όπου:

- B_{od} = ο ογκομετρικός συντελεστής της υγρής φάσης για το συγκεκριμένο βήμα της διαφορικής μελέτης
- V_L = ο εναπομείναν όγκος στις συνθήκες του βήματος της διαφορικής εξάτμισης
- V_{sc} = ο όγκος που υπολογίστηκε σε συνθήκες επιφανείας.

❖ Τους λόγους εν διαλύσει αερίου / όγκου του πετρελαίου (R_s) σε κάθε βήμα

Οι λόγοι αερίου εν διαλύσει προς υγρό R_{sd} υπολογίζονται για κάθε στάδιο πίεσης αν διαιρεθεί το άθροισμα των όγκων του αερίου που παρήχθησαν αθροιστικά στα στάδια των χαμηλότερων πιέσεων δια του όγκου V_{sc}^d

- R_{sd} = ο λόγος του αερίου εν διαλύσει προς υγρό για τις επικρατούσες συνθήκες του βήματος της διαφορικής εξάτμισης
- V_{scd} = όγκος του υγρού της διαφορικής εξάτμισης σε συνθήκες επιφανείας.

❖ Τις πυκνότητες της υγρής φάσης ως συνάρτηση της πίεσης σε κάθε βήμα

Μελέτη αριστοποίησης των συνθηκών Διαχωρισμού (Separation test)

Ο σκοπός του τεστ Διαχωριστήρα είναι να προσδιοριστεί ο αριθμός των απαιτούμενων διαχωριστήρων καθώς και οι συνθήκες λειτουργίας τους ώστε να βελτιστοποιηθεί η παραγωγή με το να μειωθεί η παραγωγή αερίου και να μεγιστοποιηθεί η παραγωγή του πετρελαίου στο τανκ το οποίο έχει και τη μεγαλύτερη εμπορική αξία. Συνήθως, πραγματοποιούνται 2-3 στάδια διαχωρισμού με το τελευταίο να πραγματοποιείται σε ατμοσφαιρική πίεση και θερμοκρασία.

Ένα αντιπροσωπευτικό δείγμα του ρευστού του ταμιευτήρα υφίσταται στο εργαστήριο μια διαδοχική αλληλουχία εκτονώσεων ίδια με αυτές που υφίσταται το ρευστό του κοιτάσματος όταν διέλθει από το σύστημα παραγωγής στην επιφάνεια. Στη συνέχεια, το δείγμα εκτονώνεται στην ενδεχόμενη πίεση και θερμοκρασία του πρώτου διαχωριστήρα που έχει σχεδιαστεί. Ο όγκος του υγρού και του αερίου μετρώνται και το αέριο μετά απομακρύνεται ενώ ένα μέρος του αερίου πηγαίνει για ανάλυση χρωματογραφίας για την εύρεση των συστατικών του. Το υγρο που απέμεινε

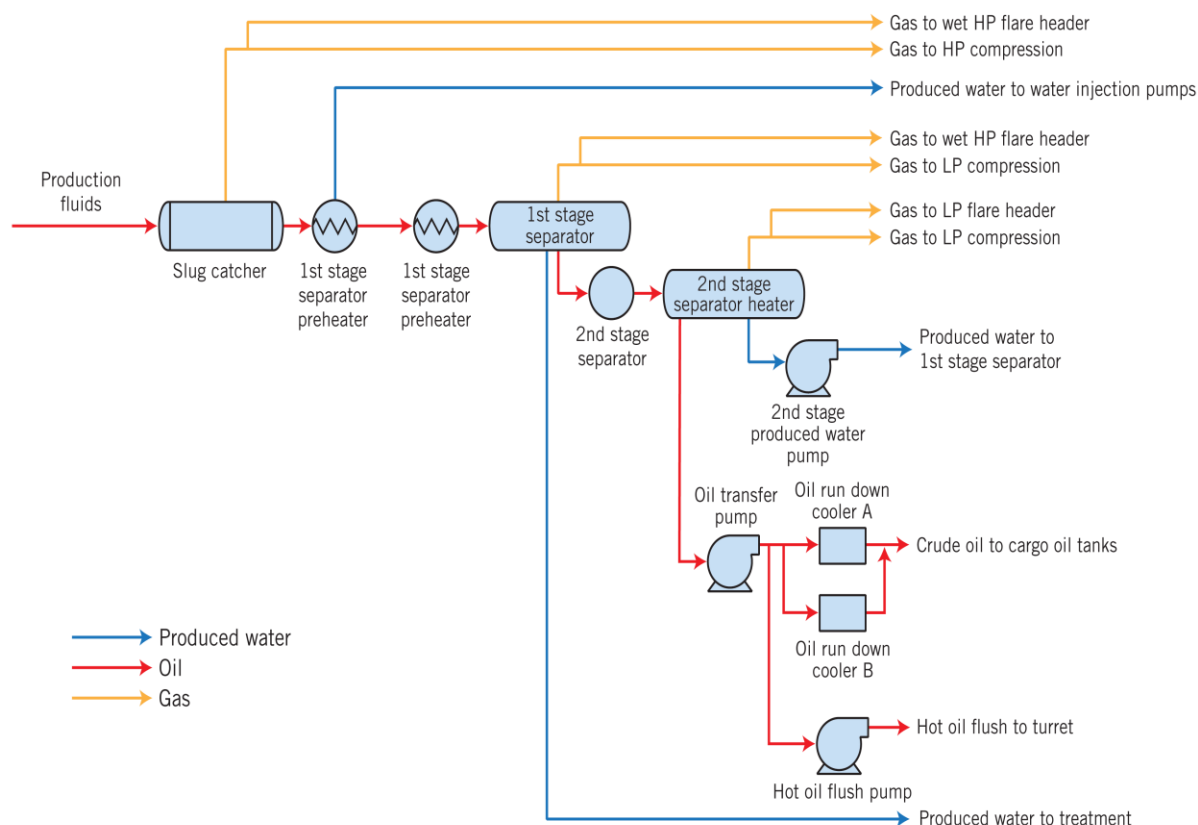
μετά την απομάκρυνση του αερίου εκτονώνεται στις αναμενόμενες συνθήκες του δεύτερου διαχωριστήρα εάν έχει σχεδιαστεί από τους μηχανικούς και ακολουθείται η ίδια διαδικασία μέχρι τις επιφανειακές συνθήκες, όπου μετριέται ο τελικός όγκος του υγρού και η σχετική πυκνότητα.

Η θερμοκρασία διαχωρισμού που επιβάλλεται στο σύστημα συνήθως είναι ίδια με τη θερμοκρασία με την οποία το παραγόμενο υγρό φθάνει στην επιφάνεια και επομένως πρόκειται βασικά για βελτιστοποίηση μιας μεταβλητής δηλαδή της πίεσης διαχωρισμού.

Το ισοζύγιο μάζας μεταξύ των moles του αρχικού υγρού που εισέρχεται και των προϊόντων που εξέρχονται από το σύστημα, χρησιμοποιείται για τον υπολογισμό των μεγεθών όπως του ογκομετρικού συντελεστή σχηματισμού υγρής φάσης Bo^{SEP} διαχωριστήρα, του συντελεστή SF του διαχωριστήρα και του GOR^{SEP} του διαχωριστήρα.

PROCESSING

FIG. 4



Source: BP Quad204 project environmental statement, November 2010

Εικόνα 3.2 Στη οποία απεικονίζεται ένα διάγραμμα ροής των σταδίων διαχωρισμού που υφίσταται το πετρέλαιο που έρχεται από τη γεώτρηση παραγωγής μέχρι το να φτάσει στο τανκ όπου και μεταφερθεί μέσω αγωγών ή πλοίων. Η πηγή αναφέρεται στην εικόνα.

II. ΑΠΕΥΘΕΙΑΣ ΕΚΤΟΝΩΣΗΣ ΡΕΥΣΤΟΥ ΤΟΥ ΤΑΜΙΕΥΤΗΡΑ ΣΕ ΕΠΙΦΑΝΕΙΑΚΕΣ ΣΥΝΘΗΚΕΣ (DIRECT FLASH LIBERATION)

Ο σκοπός της πραγματοποίησης της απευθείας εκτόνωσης του ρευστού του ταμειυτήρα σε συνθήκες επιφανείας είναι ο προσδιορισμός της σύστασης ενός υπόγειου δείγματος (bottom hole sample). Επίσης, αυτού του είδους η εκτόνωση θα μπορούσε να αντιστοιχεί στην παραγωγή του πετρελαίου απ' ευθείας στο τανκ χωρίς την παρεμβολή διαχωριστήρα.

Η διεργασία της απευθείας εκτόνωσης του ρευστού του ταμειυτήρα σε συνθήκες επιφανείας πειραματικά έχει την ίδια μεθοδολογία με το τεστ Διαχωριστήρα που αναφέρθηκε στις PVT μελέτες με τη διαφορά ότι στο εργαστήριο η εκτόνωση που υφίσταται το αντιπροσωπευτικό δείγμα του ρευστού του ταμειυτήρα σε συνθήκες ταμειυτήρα πραγματοποιείται για ένα στάδιο αυτό των επιφανειακών συνθηκών.

III. ΣΥΝΘΕΤΗ ΕΞΑΤΜΙΣΗ (COMPOSITE FLASH)

Οι τιμές οι οποίες παράγονται από τη μελέτη διαφορικής εξάτμισης παρόλο που είναι αξιόλογες για την περιγραφή των μεταβολών της θερμοδυναμικής ισορροπίας που λαμβάνει χώρα μέσα στο ταμειυτήρα, δεν μπορούν να χρησιμοποιηθούν για εφαρμογές που απαιτούν μετατροπές όγκων μεταξύ των συνθηκών του ταμειυτήρα και των συνθηκών επιφανείας. Αυτό συμβαίνει γιατί οι τιμές της διαφορικής εξάτμισης έχουν προκύψει από μια αλληλουχία εξατμίσεων που έλαβαν χώρα σε θερμοκρασία ταμειυτήρα και όχι σε θερμοκρασία επιφανείας που σαν αποτέλεσμα είχε την περαιτέρω απογύμνωση του υγρού από τα πτητικά συστατικά του. Αυτός ο διαχωρισμός είναι πιο εντατικός από ότι συμβαίνει στην πράξη κατά τη διαδικασία της παραγωγής με αποτέλεσμα οι τιμές της διαφορικής εξάτμισης να είναι μεγαλύτερες από τις αντίστοιχες τιμές μιας απευθείας εκτόνωσης του ρευστού σε ατμοσφαιρικές συνθήκες που περιγράφουν καλύτερα τις συνθήκες παραγωγής.

Η σύνθετη εξάτμιση είναι μια προσομοίωση στο εργαστήριο του τρόπου παραγωγής των ρευστών από τον ταμειυτήρα στην επιφάνεια. Η μέθοδος πειραματικά μελετά τον τρόπο διαχωρισμού του ρευστού στον ταμειυτήρα ο οποίος προσεγγίζει τη διεργασία διαφορικής εξάτμισης και κατόπιν στο διαχωριστήρα όπου προσεγγίζει τη διεργασία της στιγμιαίας εκτόνωσης. Επειδή η πραγματοποίηση της μελέτης της σύνθετης εξάτμισης απαιτεί μεγάλη ποσότητα αρχικού δείγματος ρευστού ώστε να πραγματοποιηθούν οι απευθείας εκτονώσεις για κάθε στάδιο της διαφορικής αλλά και λόγω του μεγάλου κόστους και χρόνου συνήθως χρησιμοποιούνται υπολογιστικοί μέθο-

δοι οι οποίες μετατρέπουν τις τιμές προερχόμενες από διαφορική εξάτμιση σε σύνθετη. Στη παρούσα διπλωματική υιοθετήθηκε η μέθοδος που αναφέρεται στο εγχειρίδιο του L.Dake “Fundamentals of Reservoir Engineering “ και παρουσιάζεται εν συνεχεία:

Για να προκύψουν προσεγγιστικά οι τιμές Rs_i της σύνθετης εκτόνωσης για κάθε στάδιο πίεσης της μελέτης διαφορικής εξάτμισης που αντιστοιχούν στις συνθήκες παραγωγής εφαρμόστηκε υπολογιστική μέθοδος η οποία συνδύαζε τα δεδομένα από διαφορική εξάτμιση και του τεστ διαχωριστήρα. Για το σκοπό αυτό

Για $p < p_b$ η καμπύλη του Bo είναι μετατοπισμένη κατά τον σταθερό συντελεστή

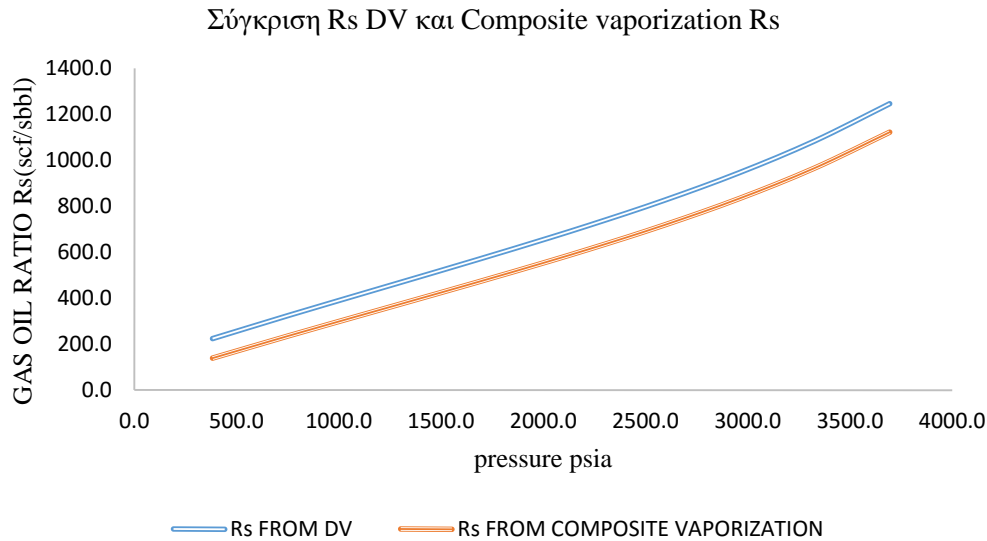
$$\frac{Bob_{flash}}{Bob_{differential}}$$

$$\text{Επομένως: } Rs = Rs_f - (Rs_{di} - Rs_d) * \left(\frac{Bob_f}{Bob_d} \right)$$

Όπου :

- **Rs_f** = Το μετρημένο Rs από εκτόνωση διαχωριστήρα
- **Rs_{di}** = Το μετρημένο Rs από διαφορική εξάτμιση του αρχικού ρευστού προς μελέτη στο αρχικό σημείο κορεσμού
- **Rs_d** = Το μετρημένο Rs από σε κάθε βήμα-στάδιο της διαφορικής εξάτμισης
- **Bob_f** = ο ογκομετρικός συντελεστής σχηματισμού της υγρής φάσης υπολογισμένος από από δεδομένα που προήλθαν από εκτόνωση του ρευστού του ταμιευτήρα σε συνθήκες επιφανείας.
- **Bob_d** = ο ογκομετρικός συντελεστής σχηματισμού της υγρής φάσης υπολογισμένος από διαφορική εξάτμιση.

Κάτωθι παρατίθεται και το διάγραμμα 3.2 όπου γίνεται μια γραφική απεικόνιση της αντι-παραβολής μιας καμπύλης σύνθετης εκτόνωσης (composite flash) σε σχέση με μια καμπύλη διαφορικής εξάτμισης για ένα τυχαίο ρευστό όπου εφαρμόστηκε η υπολογιστική μέθοδος για την ιδιότητα του Rs . Όπως είναι αναμενόμενο οι τιμές της καμπύλης της σύνθετης εκτόνωσης είναι εμφανώς μικρότερες από τις τιμές της καμπύλης της διαφορικής εξάτμισης.



Διάγραμμα 3.2 Απεικόνιση των καμπυλών της μελέτης διαφορικής εξάτμισης και της σύνθετης εξάτμισης

Για να προκύψουν προσεγγιστικά οι τιμές Bo_i της σύνθετης εξάτμισης για κάθε στάδιο πίεσης της μελέτης διαφορικής εξάτμισης που αντιστοιχούν στις συνθήκες παραγωγής εφαρμόστηκε υπολογιστική μέθοδος η οποία συνδυάζει τα δεδομένα από διαφορική εξάτμιση και του τεστ διαχωριστήρα.

$$Bo = Bod_i \frac{Bob_{flash}}{Bob_{differential}}$$

- **Bodi** = ο ογκομετρικός συντελεστής σχηματισμού της υγρής φάσης υπολογισμένος από κάθε στάδιο πίεσης της μελέτης διαφορικής εξάτμισης
- **Bob_f** = ο ογκομετρικός συντελεστής σχηματισμού της υγρής φάσης υπολογισμένος από από δεδομένα που προήλθαν από εκτόνωση του ρευστού του ταμιευτήρα σε συνθήκες επιφανείας.
- **Bob_d** = ο ογκομετρικός συντελεστής σχηματισμού της υγρής φάσης υπολογισμένος από διαφορική εξάτμιση.

ΚΕΦΑΛΑΙΟ 4

ΕΚΤΙΜΗΣΗ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΟΓΚΟΜΕΤΡΙΚΩΝ ΣΥΝΤΕΛΕΣΤΩΝ B_o , R_s ΚΑΙ ΤΟΥ ΣΗΜΕΙΟΥ ΠΙΕΣΗΣ ΦΥΣΑΛΙΔΑΣ-ΚΟΡΕΣΜΟΥ ΑΠΟ ΤΗΝ ΒΙΟΜΗΧΑΝΙΑ ΠΕΤΡΕΛΑΙΟΥ

I. ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΙΔΙΟΤΗΤΩΝ B_o , R_s ΚΑΙ p_b

Οι τιμές των ιδιοτήτων του ρευστού του ταμιευτήρα υπολογίζονται πειραματικά με τη χρήση PVT μελετών και επιπρόσθετων μετρήσεων όπως περιεγράφηκε αναλυτικά στο προηγούμενο κεφάλαιο. Στη περίπτωση που τα δεδομένα αυτά πρέπει να προσδιοριστούν άμεσα, γιατί η πραγματοποίηση των μελετών PVT χρειάζονται χρόνο εκτέλεσης πολλών εβδομάδων αλλά και σημαντικό κόστος, οι μηχανικοί πετρελαίου μπορούν να στραφούν σε μοντέλα πρόβλεψης των θερμοδυναμικών ιδιοτήτων πετρελαϊκών ρευστών τα οποία μπορούν να κατηγοριοποιηθούν ως εξής:

- A. Εάν είναι γνωστή η αναλυτική σύσταση του ρευστού μπορούν να χρησιμοποιηθούν αλγόριθμοι προσομοίωσης (κώδικες) οι οποίοι χρησιμοποιούν καταστατικές εξισώσεις, συνήθως κυβικής μορφής (Equations of State) για να περιγράψουν τη θερμοδυναμική συμπεριφορά του ρευστού. Πιο αναλυτικά, οι κυβικές καταστατικές εξισώσεις (EoS) είναι εξισώσεις που συσχετίζουν την πίεση, τον όγκο και τη θερμοκρασία. Περιγράφουν με ακρίβεια την ογκομετρική συμπεριφορά και τις φάσεις των καθαρών ενώσεων και μειγμάτων, απαιτώντας μόνο τις κρίσιμες ιδιότητες και τον ακεντρικό παράγοντα του κάθε συστατικού. Οι κυβικές καταστατικές εξισώσεις χρησιμοποιούνται για τον υπολογισμό των ιδιοτήτων όλων των φάσεων.
- B. Εάν η αναλυτική σύσταση του ρευστού δεν είναι γνωστή, τότε επιλέγονται εμπειρικές συσχετίσεις ή αλγόριθμοι από το χώρο της Μηχανικής Εκμάθησης (machine learning) γνωστών και ως Τεχνητά Νευρωνικά Δίκτυα (*artificial neural network*) ANN που χρησιμοποιούν άμεσα διαθέσιμα δεδομένα παραγωγής όπως αυτά της πυκνότητας πετρελαίου (API°), της πυκνότητας αερίου (γ_g), του GOR και της θερμοκρασίας ταμιευτήρα (T), ώστε να εκτιμηθούν οι τιμές των παραπάνω ιδιοτήτων.

Το αντικείμενο της παρούσας διπλωματικής είναι τα μοντέλα τα οποία μπορούν να χρησιμοποιηθούν μόνο με τα δεδομένα παραγωγής πριν η μελέτη PVT καταστεί διαθέσιμη οπότε και για τα μοντέλα της κατηγορίας A (EoS) δεν θα υπάρξουν περαιτέρω αναφορές.

II. ΕΜΠΕΙΡΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΟΓΚΟΜΕΤΡΙΚΩΝ ΣΥΝΤΕΛΕΣΤΩΝ B_o , R_s ΚΑΙ ΤΟΥ ΣΗΜΕΙΟΥ ΠΙΕΣΗΣ ΦΥΣΑΛΙΔΑΣ-ΚΟΡΕΣΜΟΥ

Εμπειρικές συσχετίσεις ή εξισώσεις ονομάζονται οι εξισώσεις που προέρχονται από τη στατιστική εξομάλυνση πειραματικών δεδομένων. Αυτές υποτίθεται ότι δεν έχουν γενική ισχύ, αλλά ισχύουν μόνο για περιπτώσεις που ανήκουν στα όρια του συγκεκριμένου συστήματος από το οποίο προέκυψαν πειραματικά. Δεν αποτελούν θεωρητική εξήγηση της υποκρυπτούμενης μαθηματικής σχέσης αλλά απλή καταγραφή των πειραματικών δεδομένων. Ο όρος χρησιμοποιείται σε αντιδιαστολή με τον όρο "θεωρητικές συσχετίσεις" οι οποίες έχουν αποδειχθεί τόσο θεωρητικά όσο και πειραματικά και αποτελούν τους θεμελιώδεις φυσικού νόμους.

Όταν πειραματικές μετρήσεις δεν είναι δυνατές ή διαθέσιμες, όπως επίσης και όταν δεν είναι διαθέσιμη η αναλυτική σύσταση του ρευστού, τότε δύναται να χρησιμοποιούνται εμπειρικές συσχετίσεις οι οποίες έχουν προκύψει από την παραμετροποίηση σταθερών μαθηματικών μοντέλων π.χ. πολυωνυμικών, λογαριθμικών, εκθετικών ώστε να εκφράσουν τη υποκρυπτούμενη μαθηματική σχέση που περιγράφει την αλληλεξάρτηση των άμεσα διαθέσιμων χρησιμοποιούμενων δεδομένων παραγωγής όπως αυτά της πυκνότητας πετρελαίου (API°), της πυκνότητας αερίου (γ_g), του GOR και της θερμοκρασία ταμειυτήρα (T) κ.λ. και των εκτιμώμενων τιμών των ζητούμενων ιδιοτήτων.

Αξίζει να αναφερθεί ότι η διεθνής βιβλιογραφία για τις εμπειρικές συσχετίσεις που χρησιμοποιούνται για την εκτίμηση των τιμών των ιδιοτήτων PVT είναι αρκετά εκτεταμένη. Αυτό οφείλεται στο γεγονός ότι οι μηχανικοί στηρίζονται σε αυτές τις προβλέψεις των τιμών για να κατανοήσουν συγκεκριμένα χαρακτηριστικά του ρευστού, όπως το ποσοστό της συρρίκνωσης του πετρελαίου κατά την αποληψιμότητα ή την ποσότητα του εν διαλύσει αερίου στο ταμειυτήρα που περιεγράφηκαν αναλυτικά στο κεφάλαιο 2 στα πρώτα στάδια και πολύ προτού καταστούν διαθέσιμες οι μελέτες PVT. Επίσης, αυτές οι συσχετίσεις είναι σχετικά απλές και οι τιμές των ανεξάρτητων μεταβλητών που χρησιμοποιούν είναι εύκολα μετρήσιμες.

Οι ιδιότητες του πετρελαίου που έχουν μελετηθεί περισσότερο από τους μηχανικούς σε σχέση με άλλες ιδιότητες όπως το ιξώδες και η πυκνότητα ώστε να αναπτυχθούν εμπειρικές συσχετίσεις για την εκτίμηση των τιμών τους είναι του B_o και του R_s . Αυτό συνέβη γιατί οι εκτιμώμενες τιμές των συντελεστών αυτών χρησιμοποιούνται από τους μηχανικούς πετρελαίου για να εισάγουν τα ογκομετρικά δεδομένα επιφάνειας στις εξισώσεις ισοζυγίου μάζας ώστε να γίνει μια πρώτη εκτίμηση των αποθεμάτων. Επίσης ιδιαίτερη σημασία έχουν και οι εμπειρικές συσχετίσεις που προβλέπουν την τιμή της πίεσης στο σημείο φυσαλίδας-κορεσμού (bubble point). Πρώτον γιατί η ιδιότητα αυτή λειτουργεί σαν ακρογωνιαίος λίθος για τον καθορισμό των άλλων PVT ιδιοτήτων και κατά δεύτερον γιατί υποδηλώνει το σημείο μετάβασης από ακόρεστο σε κορεσμένο μίγμα ρευστών κατά την διάρκεια της παραγωγής.

Κατηγορίες εμπειρικών συσχετίσεων

Γενικά, μπορεί να θεωρηθεί ότι οι εμπειρικές συσχετίσεις της διεθνούς βιβλιογραφίας μπορούν να χωριστούν σε δύο κατηγορίες:

- Η πρώτη κατηγορία αναφέρεται στις εμπειρικές συσχετίσεις που αναπτύχθηκαν από βάσεις δεδομένων που εμπεριείχαν αναφορές τιμών ιδιοτήτων PVT πετρελαϊκών ρευστών με μεγάλη γεωγραφική διασπορά.
- Η δεύτερη κατηγορία αναφέρεται στις εμπειρικές συσχετίσεις που αναπτύχθηκαν από βάσεις δεδομένων που εμπεριείχαν αναφορές τιμών ιδιοτήτων PVT πετρελαϊκών ρευστών που εστιάζουν σε συγκεκριμένη γεωγραφική περιοχή.

Η πρώτη πλεονεκτεί έναντι της δεύτερης κατηγορίας στο γεγονός ότι μπορεί να δώσει μια ικανοποιητική εκτίμηση για τις τιμές των ιδιοτήτων PVT για ένα μεγάλο αριθμό πετρελαϊκών ρευστών με διαφορετική σύσταση. Ενώ η δεύτερη πλεονεκτεί έναντι της πρώτης στο γεγονός ότι μπορεί να δώσει σημαντικά καλύτερη ποσοτική αλλά και ποιοτική εκτίμηση των τιμών των PVT ιδιοτήτων για τα πετρελαϊκά ρευστά της συγκεκριμένης γεωγραφικής περιοχής όπου η σύσταση δεν παρουσιάζει σημαντικές μεταβολές.

Εδώ θα πρέπει να σημειωθεί ότι παρόλο που οι συγγραφείς των εμπειρικών συσχετίσεων της πρώτης κατηγορίας είχαν ως αντικειμενικό σκοπό τη δυνατότητα εκτίμησης τιμών των ιδιοτήτων PVT για μεγάλη γκάμα πετρελαϊκών ρευστών εμπλουτίζοντας τις βάσεις δεδομένων που χρησιμοποιήθηκαν για την ανάπτυξη των συσχετίσεων τους με πολλές διαφορετικές συστάσεις πετρελαϊκών ρευστών έχει παρατηρηθεί ότι κατά την εφαρμογή τους υπάρχει ετερογένεια από άποψη ακρίβειας των εκτιμήσεων μεταξύ των πετρελαϊκών ρευστών που δυνητικά εμπεριέχονται

στο προσφερόμενο εύρος του εκάστοτε συγγραφέα. Αυτό παρατηρείται γιατί οι εμπειρικές συσχετίσεις έχουν γενικευτεί καλύτερα στις αναφορές των τιμών των ιδιοτήτων PVT των πετρελαϊκών ρευστών των γεωγραφικών περιοχών με μεγαλύτερη συχνότητα αναφοράς.

Συσχέτιση των άμεσα διαθέσιμων δεδομένων παραγωγής με τις ζητούμενες ιδιότητες

Οι τιμές του p_b , του R_{sb} και του B_{ob} εξαρτώνται βεβαίως από τη σύσταση του ρευστού του ταμιευτήρα. Όμως η αναλυτική σύσταση του ρευστού του ταμιευτήρα δεν είναι εύκολο να προσδιοριστεί χωρίς πειραματική διαδικασία και χρωματογραφική ανάλυση όπως έχει ήδη ειπωθεί, παρόλο που τα τελευταία χρόνια έχει γίνει πρόοδος για την εκτίμηση της αναλυτικής σύστασης του με μια μέση ακρίβεια κάνοντας χρήση νευρωνικών δικτύων που αξιοποιούν τις πληροφορίες που μπορούν να ληφθούν από ένα οπτικό αναλυτή ρευστών (OFA με μέτρηση στο βάθος των διατρήσεων).

Οι εμπειρικές συσχετίσεις έχουν αναπτυχθεί με τέτοιο τρόπο ώστε η πληροφορία της σύστασης του ρευστού του ταμιευτήρα να μην απαιτείται αναλυτικά για την εκτίμηση των τιμών των ζητούμενων ιδιοτήτων αλλά να εκφράζεται εμμέσως και προσεγγιστικά χρησιμοποιώντας τα άμεσα διαθέσιμα δεδομένα παραγωγής.

- ❖ Για την εκτίμηση της ιδιότητας του Bob στη πίεση κορεσμού στις εμπειρικές συσχετίσεις η πληροφορία Zi (αναλυτική σύσταση) αντικαθίσταται έμμεσα και προσεγγιστικά από τις τιμές των πυκνοτήτων αερίου (γ_g) και υγρού (API°) σε συνδυασμό με τη τιμή του GOR. Χρησιμοποιώντας τις τιμές των άμεσων διαθέσιμων δεδομένων παραγωγής, οι εμπειρικές συσχετίσεις αλλά και τα νευρωνικά δίκτυα είναι σε θέση να εκτιμήσουν την ιδιότητα του Bob χωρίς να χρειάζεται κάποια περαιτέρω πληροφορία από τον ταμιευτήρα πέραν της θερμοκρασίας. Η μαθηματική σχέση που εκφράζει την συσχέτιση της ιδιότητας του ογκομετρικού συντελεστή υγρής φάσεως Bob με τις ανεξάρτητες μεταβλητές είναι η ακόλουθη:

$$Bob = f(R_{sb}, \gamma_g, API, T)$$

- ❖ Για την εκτίμηση της ιδιότητας του p_b στις εμπειρικές συσχετίσεις η πληροφορία Zi (αναλυτική σύσταση) αντικαθίσταται έμμεσα και προσεγγιστικά από τις τιμές GOR, γ_g , API

και σε αυτή τη περίπτωση. Επίσης όπως και στην ιδιότητα του Bob χρειάζεται και η πληροφορία της θερμοκρασίας του ταμιευτήρα. Η μαθηματική σχέση που εκφράζει την συσχέτιση της ιδιότητας του p_b (της πίεσης του σημείου φουσαλίδας) με τις ανεξάρτητες μεταβλητές είναι η ακόλουθη:

$$p_b = f(R_{sb}, \gamma_g, API, T)$$

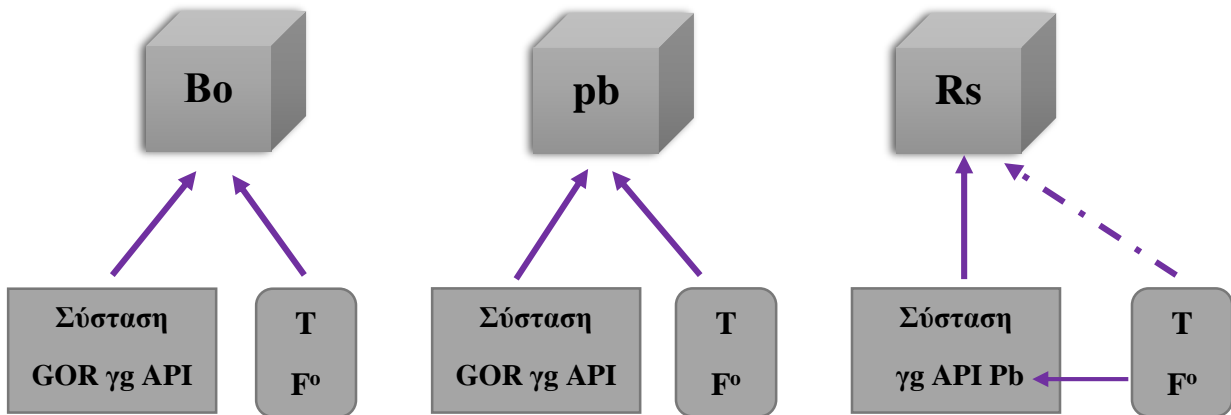
- ❖ Για την εκτίμηση της ιδιότητας του R_s στις εμπειρικές συσχετίσεις η πληροφορία Z_i (αναλυτική σύσταση) αντικαθίσταται έμμεσα και προσεγγιστικά από τις τιμές p_b , γ_g , API . Σε αυτή τη περίπτωση, για τον υπολογισμό της ιδιότητας R_s από τις εμπειρικές συσχετίσεις και τα νευρωνικά δίκτυα εκτός των τιμών των άμεσα διαθέσιμων δεδομένων παραγωγής χρειάζεται και η πληροφορία του σημείου φουσαλίδας-κορεσμού του ταμιευτήρα. Όταν η πληροφορία αυτή δεν είναι διαθέσιμη μπορεί να προσδιοριστεί είτε μέσω μιας εκτιμητικής μεθόδου είτε μέσω μιας γρήγορης μελέτης σταθερής μάζας που διεξάγεται στο πεδίο. Επειδή η τιμή του p_b παράγεται συναρτήσει της θερμοκρασίας χρησιμοποιείται και η θερμοκρασία για την εκτίμηση της τιμής του R_{sb} παρόλο που συσχετίζεται εμμέσως μαζί της. Η μαθηματική σχέση που εκφράζει την συσχέτιση της ιδιότητας του όγκου εν διαλύσει αερίου / όγκο πετρελαίου R_{sb} στη πίεση κορεσμού με τις ανεξάρτητες μεταβλητές είναι η ακόλουθη:

$$R_{sb} = f(p_b, \gamma_g, API, T)$$

Θα πρέπει να σημειωθεί ότι για την ανάπτυξη των εμπειρικών συσχετίσεων έγιναν οι εξής παραδοχές από τους συγγραφείς:

1. Το πετρέλαιο του ταμιευτήρα έχει ανασυσταθεί από πετρέλαιο και αέριο επιφανείας
2. Το αέριο της επιφανείας που έχει προέλθει από ρευστό του ταμιευτήρα έχει τις ίδιες ιδιότητες με το εν διαλύσει αέριο του ταμιευτήρα
3. Οι ιδιότητες του πετρελαίου και του αερίου επιφανείας δεν αλλάζουν κατά την εκμετάλλευση

Κάτωθι παρουσιάζεται μια γραφική απεικόνιση των συσχετισμών των ιδιοτήτων Bo, Rs και pb με τις ανεξάρτητες μεταβλητές που χρησιμοποιούνται για τον υπολογισμό τους από τις εμπειρικές συσχετίσεις και τα νευρωνικά δίκτυα.



Εικόνα 1.7 Γραφική απεικόνιση συσχετισμών Bo,pb,Rs

Παρουσίαση των επιλεγμένων εμπειρικών συσχετίσεων ανά ιδιότητα

Ύστερα από έρευνα της διεθνούς βιβλιογραφίας επιλέχθηκαν οι κάτωθι εμπειρικές συσχετίσεις ώστε να συγκριθούν ως προς την ακρίβεια των εκτιμήσεων σε επόμενο κεφάλαιο με τα νευρωνικά δίκτυα που αναπτύχθηκαν για τις ιδιότητες Bob , Rsb , pb.

- Για την ιδιότητα του Bob επιλέχθηκαν 15 εμπειρικές συσχετίσεις
- Για την ιδιότητα του Rsb επιλέχθηκαν 9 εμπειρικές συσχετίσεις
- Για την ιδιότητα του pb επιλέχθηκαν 12 εμπειρικές συσχετίσεις

Εν συνεχεία, παρατίθενται οι τύποι των εμπειρικών συσχετίσεων ανά ιδιότητα και ένας ενημερωτικός πίνακας που περιγράφει το εύρος τιμών των συντελεστών των εξισώσεων το οποίο συνοδεύει τις εξισώσεις καθώς και τον αριθμό των πειραματικών δεδομένων, την γεωγραφική περιοχή από την οποία προέρχονται τα δείγματα που χρησιμοποιήθηκαν και το έτος δημοσίευσης.

Εμπειρικές συσχετίσεις πρόβλεψης Bo

1. Standing Correlation

Τύπος εξίσωσης : $B_{ob} = a_1 + a_2 [R_s (\gamma_g / \gamma_o)^{a_3} + a_4 T]^{a_5}$

$$a_1 = 0.975, a_2 = 0.00012, a_3 = 0.5, a_4 = 1.25, a_5 = 1.2$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1947	Βάσης : 105	Καλιφόρνια

2. Vazquez and Beggs

Τύπος εξίσωσης : $B_{ob} = 1 + C_1 R_s + C_2 (T - 60) (API / \gamma_g) + C_3 R_s (T - 60) (API / \gamma_g)$

$$\text{Για } API \leq 30 \quad C_1 = 0.0004677 \quad C_2 = 0.00001751 \quad C_3 = -1.811E-08$$

$$\text{Για } API > 30 \quad C_1 = 0.000467 \quad C_2 = 0.000011 \quad C_3 = 1.337E-09$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1980	Βάσης : 6004	Παγκόσμια δειγματοληψία

3. Glaso Correlation

Τύπος εξίσωσης : $B_{ob} = 1 + 10^{[a_1 + a_2 (\log G) - a_3 (\log G)^2]}$

$$G = R_s (\gamma_g / \gamma_o)^{a_4 + a_5 T}$$

$$a_1 = -6.58511, a_2 = 2.91329, a_3 = 0.27683, a_4 = 0.526, a_5 = 0.968$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1980	Βάσης : 41	Βόρεια Θάλασσα

4. Al-Marhoun Correlation

Τύπος εξίσωσης : $Bob = a1 + a2 * (T + 460) + a3 * M + a4 * M^2$

$$M = Rs^{a5} * \gamma g^{a6} * \gamma o^{a7}$$

$a1 = 0.497069$, $a2 = 0.000862963$, $a3 = 0.00182594$, $a4 = 3.18099E-06$
 $a5 = 0.74239$, $a6 = 0.323294$, $a7 = -1.20204$

Έτος δημοσίευσης :

1988

Αριθμός Δεδομένων

Βάσης : 160

Γεωγραφική περιοχή :

Μέση Ανατολή

5. Majeed and Salman Correlation

Τύπος εξίσωσης : $Bob = a1 + (a3 * M - a4 * M^2) + a2 * T$

$$M = Rs^{a5} * \gamma g^{a6} * \gamma o^{a7}$$

$a1 = 0.9657876$, $a2 = 0.000773$, $a3 = 0.000048141$, $a4 = -6.8987E-10$
 $a5 = 1.2$, $a6 = -0.147$, $a7 = -5.222$

Έτος δημοσίευσης :

1988

Αριθμός Δεδομένων

Βάσης : 420

Γεωγραφική περιοχή :

Δεν αναφέρεται

6. Dokla and Osman Correlation

Τύπος εξίσωσης : $Bob = a1 + a2 * (T + 460) + a3 * M + a4 * M^2$

$$M = Rs^{a5} * \gamma g^{a6} * \gamma o^{a7}$$

$a1 = 0.0431935$, $a2 = 0.00156667$, $a3 = 0.00139775$, $a4 = 3.80525E-06$, $a5 = 0.773572$,
 $a6 = 0.40402$, $a7 = -0.882605$

Έτος δημοσίευσης :

1992

Αριθμός Δεδομένων

Βάσης : 51

Γεωγραφική περιοχή :

H.A.E

7. Farshad, Leblance, Garber and Osorio Correlation

Τύπος εξίσωσης : $B_o = 1 + a_7 * R_{sc}^{a_8} + A$

$$A = (a_1 + a_2 * R_{sc}^{a_3}) * (T - 60)^{a_4} * API^{0.72} * \gamma_g^{a_6}$$

$a_1 = 0.000292$, $a_2 = 0.0000459$, $a_3 = 0.4595$, $a_4 = 0.5657$, $a_5 = 0.72$, $a_6 = -0.215$
 $a_7 = 0.00003$, $a_8 = 1.3267$

Έτος δημοσίευσης :

1992

Αριθμός Δεδομένων

Βάσης : 107

Γεωγραφική περιοχή :

Κολομβία

8. Macary and El-Batanony Correlation

Τύπος εξίσωσης : $B_o = [a_1 + a_2 * T] * N$

$$N = EXP [a_3 * R_s + a_4 (\gamma_o / \gamma_g)]$$

$a_1 = 1.0031$, $a_2 = 0.0008$, $a_3 = 0.0004$, $a_4 = 0.0006$

Έτος δημοσίευσης :

1992

Αριθμός Δεδομένων

Βάσης : 90

Γεωγραφική περιοχή :

Κόλπος του Σουέζ

9. Al-Marhoun Correlation

Τύπος εξίσωσης :

$$B_o = 1 + (a_1 * R_s) + (a_2 * R_s * (\gamma_g / \gamma_o)) + a_3 * R_s * (1 - \gamma_o) (T - 60) + a_4 * (T - 60)$$

$a_1 = 0.000177342$, $a_2 = 0.00022016$, $a_3 = 4.29258E-06$,
 $a_4 = 0.000528707$

Έτος δημοσίευσης :

1992

Αριθμός Δεδομένων

Βάσης : 4012

Γεωγραφική περιοχή :

Παγκόσμια Δειγματοληψία

10. Petrosky and Farshad Correlation

Τύπος εξίσωσης : $B_{ob} = a_1 + a_2 [R_s^{a_3} (\gamma_g^{a_4} / \gamma_o^{a_5}) + a_6 * T^{a_7}]^{a_8}$

$a1 = 1.0113, a2 = 0.000072046, a3 = 0.3738, a4 = 0.2914,$
 $a5 = 0.6265, a6 = 0.24626, a7 = 0.5371, a8 = 3.0936$

Έτος δημοσίευσης :

1993

Αριθμός Δεδομένων

Βάσης : 90

Γεωγραφική περιοχή :

Κόλπος του Μεξικού

11. Omar and Todd Correlation

Τύπος εξίσωσης : $Bob = a1 + a2 [Rs (\gamma g / \gamma o)^{a3} + a4 T]^{a5}$

$$X = b1 + b2 * (API / \gamma g) + b3 * \gamma g$$

$a1 = 0.972, a2 = 0.00147, a3 = 0.5, a4 = 1.25$

$b1 = 1.1663, b2 = 0.000762, b3 = -0.0399$

Έτος δημοσίευσης :

1993

Αριθμός Δεδομένων

Βάσης : 93

Γεωγραφική περιοχή :

Μαλαισία

12. Kartoatmodjo and Schmidt Correlation

Τύπος εξίσωσης : $Bob = a1 + a2 [Rs^{a3} (\gamma g^{a4} / \gamma o^{a5}) + a6 * T]^{a7}$

$a1 = 0.98496, a2 = 0.0001, a3 = 0.755, a4 = 0.25, a5 = 1.5,$
 $a6 = 0.45, a7 = 1.5$

Έτος δημοσίευσης :

1994

Γεωγραφική περιοχή :

Παγκόσμια Δειγματοληψία

Αριθμός Δεδομένων

Βάσης : 5392

13. Elsharkawy Correlation

Τύπος εξίσωσης : $Bob = 1 + a1 * Rs + a2 * (T - 60) + a3 * (Rs * (T - 60) * (\gamma g / \gamma o))$

$a_1 = 0.00040428, a_2 = 0.00063802, a_3 = 0.0000078$

Έτος δημοσίευσης :

1996

Αριθμός Δεδομένων

Βάσης : 175

Γεωγραφική περιοχή :

Κουβέιτ

14. Almehaideb Correlation

Τύπος εξίσωσης : $Bo = a_1 + a_2(Rs(T/\gamma_o)^2)$

$a_1 = 1.122018, a_2 = 0.00000141$

Έτος δημοσίευσης :

1997

Αριθμός Δεδομένων

Βάσης : 62

Γεωγραφική περιοχή :

H.A.E

15. Birol Dindoruk and Peter G.Christman Correlation

Τύπος εξίσωσης : $Bob = a_{11} + a_{12} * A + a_{13} * A^2 + a_{14} * (T-60) * (API/\gamma_g)$

$A = ((Rs^{a_1} \gamma_g^{a_2}) / \gamma_o^{a_3 + a_4 (T-60)^{a_5 + a_6 Rs}})^{a_7} / (a_8 + (2Rs^{a_9} / \gamma_g^{a_{10}}) * (T-60))^{a_{11}}$

$a_1 = 2.510755, a_2 = -4.852538, a_3 = 11.835, a_4 = 136542.8,$

$a_5 = 2.25288, a_6 = 10.0719, a_7 = 0.4450849, a_8 = 5.352624,$

$a_9 = -0.6309052, a_{10} = 0.9000749, a_{11} = 0.9871766,$

$a_{12} = 0.000786515, a_{13} = 2.68917E-06, a_{14} = 1.1E-05$

Έτος δημοσίευσης :

2001

Αριθμός Δεδομένων

Βάσης : 100

Γεωγραφική περιοχή :

Κόλπος του Μεξικού

Συσχετίσεις	Bo		T		Rs		API		γg	
	“bbl/STB”		“F”		“scf/STB”		“°API°”		“air = 1”	
	Από	Μέχρι	Από	Μέχρι	Από	Μέχρι	from	Μέχρι	Από	Μέχρι
Standing	1.024	2.150	100	258	20	1425	16.5	63.8	0.59	0.95
Vazquez and Beggs	1.028	2.226	75	294	0	2199	15.3	59.3	0.51	1.35
Glaso	1.032	2.588	80	280	90	2637	22.3	48.1	0.65	1.28
Al-Marhoun	1.032	1.997	74	240	26	1602	19.4	44.6	0.75	1.37
Majeed & Salman	1.028	2.042	75	290	0	1664	9.5	59.5	0.51	1.35
Dokla & Osman	1.216	2.493	190	275	181	2266	28.2	40.3	0.80	1.29
Farshad, Leblance, Garber & Osorio	1.060	2.064	95	260	6	1645	18.0	44.9	0.66	1.70
Macary & El-Batanony	1.200	2.000	130	290	200	1200	25	40	0.70	1.00
Al-Marhoun	1.010	2.960	75	300	0	3265	9.5	55.9	0.57	2.52
Petrosky & Farshad	1.118	1.623	114	288	217	1406	16.3	45	0.58	0.85
Omar & Todd	1.085	1.954	125	280	142	1440	26.6	53.2	0.61	1.32
Kartoatmodjo & Schmidt	1.007	2.144	75	320	0	2890	14.4	58.9	0.38	1.71
Elsharkawy	1.076	1.969	130	250	39	1586	19.9	42.8	0.66	1.27
Almehaideb	1.142	3.562	190	306	128	3871	30.9	48.6	0.75	1.12
Birol Dindoruk and Peter G. Christman	1.084	2.898	117	276	133	3050	14.7	40.0	0.60	1.03

Πίνακας 4.1 Στο παραπάνω πίνακα απεικονίζεται το εύρος τιμών που χρησιμοποιήθηκε για την ανάπτυξη των εμπειρικών συσχετίσεων. Αυτά είναι κατά σειρά τα όρια του Bo των πετρελαίων που χρησιμοποιήθηκαν, τα όρια των θερμοκρασιών (T) των πετρελαίων που χρησιμοποιήθηκαν, τα όρια του Rs των πετρελαίων που χρησιμοποιήθηκαν, τα όρια της πυκνότητας (API) των πετρελαίων που χρησιμοποιήθηκαν και τα όρια της πυκνότητας των αερίων (γg) που χρησιμοποιήθηκαν.

Εμπειρικές συσχετίσεις πρόβλεψης ρh

1. Standing Correlation

Τύπος εξίσωσης : $p_b = a_1 * [(R_s/\gamma_g)^{a_2}] * (10)^X - a_3$

$$X = a_4 * T - (a_5 * (API))$$

$$a_1 = 18.2, a_2 = 0.83, a_3 = 1.4, a_4 = 0.00091, a_5 = 0.0125$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1947	Βάσης : 105	Καλιφόρνια

2. Vazquez and Beggs

Τύπος εξίσωσης : $p_b = [(C_1 * R_s / \gamma_g) * (10)^X]^{C_2}$

$$X = -C_3 * (API / (T + 460))$$

$$\text{Για } API \leq 30 \quad C_1 = 27.624 \quad C_2 = 0.914328 \quad C_3 = 11.172$$

$$\text{Για } API > 30 \quad C_1 = 56.18 \quad C_2 = 0.84246 \quad C_3 = 10.393$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1980	Βάσης : 6004	Παγκόσμια δειγματοληψία

3. Glaso Correlation

Τύπος εξίσωσης : $p_b = 10^{(1.7669 + 1.7447 * \log(G) - 0.30218 * [\log(G)]^2)}$

$$G = ((R_s / \gamma_g)^{a_1}) * (T^{a_2}) * (API^{a_3})$$

$$a_1 = 0.816, a_2 = 0.172, a_3 = -0.989$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1980	Βάσης : 41	Βόρεια Θάλασσα

4. Al-Marhoun Correlation

Τύπος εξίσωσης : $p_b = a_1 * (R_s^{a_2}) * (\gamma_g^{a_3}) * (\gamma_o^{a_4}) * ((T + 460)^{a_5})$

$a1 = 5.38088E-03$, $a2 = 0.715082$, $a3 = -1.87784$, $a4 = 3.1437$ $a5 = 1.32657$

Έτος δημοσίευσης :

1988

Αριθμός Δεδομένων

Βάσης : 160

Γεωγραφική περιοχή :

Μέση Ανατολή

5. Lasater Correlation

Τύπος εξίσωσης : $p_b = (P_f * (T + 460)) / \gamma_g$

$P_f = ((\exp(1))^{(X_g - 0.15649) / 0.33705}) - 0.15649$

$X_g = (1 + (\gamma_o / ((7.521 * 10^{(-6)}) * GOR * M_o)))^{(-1)}$

If $API \leq 40$ $M_o = 630 - (10 * API)$

If $API > 40$ $M_o = 73110 * API^{(-1.562)}$

Έτος δημοσίευσης :

1958

Αριθμός Δεδομένων

Βάσης : 158

Γεωγραφική περιοχή :

Καναδάς

6. Dokla and Osman Correlation

Τύπος εξίσωσης : $p_b = (a1 * (R_s^{a2}) * (\gamma_o^{a3})) / ((\gamma_g^{a4}) * (T + 460)^{a5})$

$a1 = 8363.86$, $a2 = 0.724047$, $a3 = 0.107991$, $a4 = 1.01049$, $a5 = 0.952584$

Έτος δημοσίευσης :

1992

Αριθμός Δεδομένων

Βάσης : 51

Γεωγραφική περιοχή :

H.A.E

7. Farshad, Leblance, Garber and Osorio Correlation

Τύπος εξίσωσης : $p_b = a1 * ((R_s / \gamma_g)^{a2}) * 10^{((a3 * T) - (a4 * API))}$

$$a_1 = 33.22, a_2 = 0.8283, a_3 = 0.000037, a_4 = 0.0142$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1992	Βάσης : 107	Κολομβία

8. Macary and El-Batanony Correlation

Τύπος εξίσωσης : $pb = a_1 * K * ((Rs^{a_2}) - a_3)$

$$K = EXP((a_4 * T) - (a_5 * API) - (a_6 * \gamma g))$$

$$a_1 = 204.257, a_2 = 0.51, a_3 = 4.7927, a_4 = 0.00077, a_5 = 0.097, a_6 = 0.4003$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1992	Βάσης : 90	Κόλπος του Σουέζ

9. Petrosky and Farshad Correlation

Τύπος εξίσωσης : $pb = a_5 * (((Rs^{a_6}) / (\gamma g^{a_7})) * (10^X)) - a_8$

$$X = (a_1 * (T^{a_2})) - (a_3 * (API^{a_4}))$$

$$a_1 = 4.561E-05, a_2 = 1.3911, a_3 = 7.916E-04, a_4 = 1.541, a_5 = 112.727, a_6 = 0.5774, a_7 = 0.84305, a_8 = 12.34$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1993	Βάσης : 90	Κόλπος του Μεξικού

10. Kartoatmodjo and Schmidt Correlation

Τύπος εξίσωσης : $pb = (Rs / (a_1 * (\gamma g^{a_2}) * 10^{((a_3 * API) / (T + 460))}))^{a_4}$

If API <=30 a1 = 0.05958, a2 = 0.7972, a3 = 13.1405, a4 = 0.9986,
 If API >30 a1 = 0.03150, a2 = 0.7587, a3 = 11.289, a4 = 0.9143

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1994	Βάσης : 5392	Παγκόσμια Δειγματοληψία

11. Birol Dindoruk and Peter G.Christman Correlation

Τύπος εξίσωσης : $pb = a8 * ((Rs^{a9}) / (vg^{a10}) * 10^A + a11)$

$$A = ((a1 * T^{a2}) + a3 * API^{a4}) / (a5 + ((2 * Rs^{a6}) / (vg^{a7})))^2$$

a1 = 1.42828E-10, a2 = 2.84459, a3 = -6.74896e-04, a4 = 1.225226,
 a5 = 0.033383, a6 = -0.272945, a7 = -0.084226, a8 = 1.869979,
 a9 = 1.221486, a10 = 1.370508, a11 = 0.011688,

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
2001	Βάσης : 100	Κόλπος του Μεξικού

12. Velarde Correlation

Τύπος εξίσωσης :

$$pb = 109.47 [(Rs^{0.081465}) * (vg^{-0.161488}) * 10^X - 0.740152]^{5.354891}$$

$$X = (0.013098 * T^{0.282372}) - ((8.2 * 10^{-6}) * API^{2.176124})$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1999	Βάσης : 728	Δεν αναφέρεται

Συσχετίσεις	pb		T		Rs		API		γg	
	“psia”		“F”		“scf/STB”		“°API°”		“air = 1”	
	Από	Μέχρι	Από	Μέχρι	Από	Μέχρι	from	Μέχρι	Από	Μέχρι
Standing	130	7000	100	258	20	1425	16.5	63.8	0.59	0.95
Vazquez and Beggs	15	6055	75	294	0	2199	15.3	59.3	0.51	1.35
Glaso	165	7142	80	280	90	2637	22.3	48.1	0.65	1.28
Al-Marhoun	20	3573	74	240	26	1602	19.4	44.6	0.75	1.37
Lasater	48	5780	82	272	3	2905	17.9	51.1	0.57	1.22
Dokla & Osman	590	4640	190	275	181	2266	28.2	40.3	0.80	1.29
Farshad, Leblance, Garber & Osorio	32	4138	95	260	6	1645	18.0	44.9	0.66	1.70
Macary&El-Batanony	1200	4600	130	290	200	1200	25	40	0.70	1.00
Petrosky & Farshad	1574	6523	114	288	217	1406	16.3	45	0.58	0.85
Kartoatmodjo & Schmidt	15	6055	75	320	0	2890	14.4	58.9	0.38	1.71
Velarde	70	6700	74	327	10	1870	12	55	0.56	1.37
Birol Dindoruk and Peter G.Christman	926	12230	117	276	133	3050	14.7	40.0	0.60	1.03

Πίνακας 4.2 Στο παραπάνω πίνακα απεικονίζεται το εύρος τιμών που χρησιμοποιήθηκε για την ανάπτυξη των εμπειρικών συσχετίσεων. Αυτά είναι κατά σειρά τα όρια του Pb των πετρελαίων που χρησιμοποιήθηκαν, τα όρια των θερμοκρασιών(T) των πετρελαίων που χρησιμοποιήθηκαν, τα όρια του Rs των πετρελαίων που χρησιμοποιήθηκαν, τα όρια της πυκνότητας(API) των πετρελαίων που χρησιμοποιήθηκαν και τα όρια της πυκνότητας των αερίων(γg) που χρησιμοποιήθηκαν.

Εμπειρικές συσχετίσεις πρόβλεψης Rs

1. Standing Correlation

Τύπος εξίσωσης : $R_s = \gamma_g \cdot [10^{x \cdot ((P_b/18.2) + 1.4)}]^{1.2048}$

$$X = 0.0125API - 0.00091Tres$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1947	Βάσης : 105	Καλιφόρνια

2. Lasater Correlation

Τύπος εξίσωσης : $R_s = (132755 * \gamma_o * \gamma_g) / (M_o * (1 - \gamma_g))$

$$M = 6084 / (API - 5.9)$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1958	Βάσης : 137	Καναδάς

3. Glaso Correlation

Τύπος εξίσωσης : $R_s = \gamma_g \left((10^x (API^{0.989}) / (T^{0.172})) \right)^{1.2255}$

$$X = 2.8869 - (14.1811 - 3.3093 * \log P)^{0.5}$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1980	Βάσης : 41	Βόρεια Θάλασσα

4. Vazquez and Beggs Correlation

Τύπος εξίσωσης : Για $API \leq 30$ $R_s = 0.0362 * \gamma_g * P^{1.0937} * \exp(25.724 * (API / (T + 460)))$

$$\text{Για } API > 30 \quad R_s = 0.0178 * \gamma_g * P^{1.187} * \exp(23.931 * (API / (T + 460)))$$

$$\gamma_g = \gamma_g \left(1 + 5.912 * 10^{(-5)} * API * T_{sp} * \log(P_{sp} / 114.7) \right)$$

Έτος δημοσίευσης :	Αριθμός Δεδομένων	Γεωγραφική περιοχή :
1980	Βάσης : 6004	Παγκόσμια δειγματοληψία

5. Al-Marhoun Correlation

Τύπος εξίσωσης : $R_s = A^{1.398441}$

$$A = 185.483208 \gamma_g^{1.87784} \gamma_o^{-3.1437} (T + 460)^{-1.32657} P$$

Έτος δημοσίευσης :	1988
--------------------	------

Αριθμός Δεδομένων

Γεωγραφική περιοχή :

Βάσης : 160

Μέση Ανατολή

6. Petrosky and Farshad Correlation

Τύπος εξίσωσης : $R_s = ((P/112.727) + 12.34) * \gamma_g^{0.8439} * 10^{x^{1.73184}}$

$X = (7.916 * 10^{-4}) * API^{1.541} - (4.561 * 10^{-5}) * T^{1.3911}$

Έτος δημοσίευσης :

1993

Αριθμός Δεδομένων

Βάσης : 90

Γεωγραφική περιοχή :

Κόλπος Μεξικού

7. Farshad, Leblance, Garber and Osorio Correlation

Τύπος εξίσωσης :

$R_s = 0.01456 \gamma_g c * P_b^{1.2073} * 10^{(0.01714 * API - 0.000004467 * T) / 1 - 24.663 * (\gamma_o / T_{sp})}$

Έτος δημοσίευσης :

1992

Αριθμός Δεδομένων

Βάσης : 107

Γεωγραφική περιοχή :

Κολομβία

8. Kartoatmodjo and Schmidt Correlation

Τύπος εξίσωσης : Για $API \leq 30$ $R_{sf} = 0.05958 * \gamma_g^{0.7972} * P^{1.0014} * 10^{13.1405 * API / (T+460)}$

Για $API > 30$ $R_{sf} = 0.03150 * \gamma_g^{0.7587} * P^{1.0937} * 10^{11.289 * API / (T+460)}$

Έτος δημοσίευσης :

1994

Αριθμός Δεδομένων

Βάσης : 5392

Γεωγραφική περιοχή :

Παγκόσμια Δειγματοληψία

9. Birol Dindoruk and Peter G.Christman Correlation

Τύπος εξίσωσης : $R_s = (((P/a_8) + a_9) * \gamma_g^{a_{10}} * 10^A)^{a_{11}}$

$A = (a_1 * API^{a_2} + a_3 * T^{a_4}) / (a_5 + (2 * API^{a_6}) / P^{a_7})^{a_2}$

$a1 = 4.86996E-06$, $a2 = 5.730982539$, $a3 = 0.0099251$, $a4 = 1.776179364$,
 $a5 = 44.2500268$, $a6 = 2.702889206$, $a7 = 0.744335673$, $a8 = 3.35975497$,
 $a9 = 28.10133245$, $a10 = 1.57905016$, $a11 = 0.928131344$

Έτος δημοσίευσης :

2001

Γεωγραφική περιοχή :

Κόλπος του Μεξικού

Αριθμός Δεδομένων

Βάσης : 100

Συσχετίσεις	Pb		T		Rs		API		γg	
	“P _{sia} ”		“F”		“scf/STB”		“°API”		“air = 1”	
	Από	Μέχρι	Από	Μέχρι	Από	Μέχρι	from	Μέχρι	Από	Μέχρι
Standing	130	7000	100	258	20	1425	16.5	63.8	0.59	0.95
Lasater	48	5780	82	272	3	2905	17.9	51.1	0.57	1.22
Glaso	165	7142	80	280	90	2637	22.3	48.1	0.65	1.28
Vazquez and Beggs	15	6055	75	294	0	2199	15.3	59.3	0.51	1.35
Al-Marhoun	130	3570	74	240	26	1602	19.4	44.6	0.75	1.37
Farshad, Leblance, Garber & Osorio	32	4138	95	260	6	1645	18.0	44.9	0.66	1.70
Petrosky & Farshad	1574	6523	114	288	217	1406	16.3	45	0.58	0.85
Kartoatmodjo & Schmidt	15	6055	75	320	0	2890	14.4	58.9	0.38	1.71
Birol Dindoruk and Peter G. Christman	926	12230	117	276	133	3050	14.7	40.0	0.60	1.03

Πίνακας 4.3 Στο παραπάνω πίνακα απεικονίζεται το εύρος τιμών που χρησιμοποιήθηκε για την ανάπτυξη των εμπειρικών συσχετίσεων. Αυτά είναι κατά σειρά τα όρια του p_b των πετρελαίων που χρησιμοποιήθηκαν, τα όρια των θερμοκρασιών (T) των πετρελαίων που χρησιμοποιήθηκαν, τα όρια του R_s των πετρελαίων που χρησιμοποιήθηκαν, τα όρια της πυκνότητας (API) των πετρελαίων που χρησιμοποιήθηκαν και τα όρια της πυκνότητας των αερίων (γ_g) που χρησιμοποιήθηκαν.

III. Αντικείμενο Διπλωματικής Εργασίας

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη μοντέλων πρόβλεψης των θερμοδυναμικών ιδιοτήτων B_o , R_s , p_b πετρελαϊκών ρευστών με χρήση αλγορίθμων από το χώρο της Μηχανικής Εκμάθησης (machine learning) γνωστών και ως Τεχνητά Νευρωνικά Δίκτυα (*artificial neural network*) **ANN** χρησιμοποιώντας μια ευρεία βάση δεδομένων πετρελαϊκών ρευστών ώστε να μπορούν να γίνουν εκτιμήσεις με σχετικά μεγάλη ακρίβεια των τιμών των ιδιοτήτων B_o , R_s και p_b τη στιγμή που είναι διαθέσιμα μόνο δεδομένα παραγωγής.

Εφόσον δεν θα είναι διαθέσιμη η αναλυτική σύσταση του ρευστού του ταμιευτήρα λόγω του ότι δεν θα έχει προηγηθεί κάποια PVT μελέτη του ρευστού ταμιευτήρα προερχόμενο από κάποια δειγματοληψία τα μόνα άμεσα διαθέσιμα δεδομένα ώστε να χρησιμοποιηθούν από τα νευρωνικά δίκτυα για να εκτιμήσουν τις ιδιότητες αυτές θα είναι τα δεδομένα παραγωγής όπως αυτά της πυκνότητας πετρελαίου (**API°**), της πυκνότητας αερίου (**γg**), του **GOR** και της θερμοκρασία ταμιευτήρα (**T**).

Για την εκπαίδευση των νευρωνικών χρησιμοποιήθηκε μια εκτεταμένη βάση δεδομένων αναφορών PVT πετρελαίων ταμιευτήρα με ευρεία γεωγραφική διασπορά που ανήκει στο Εργαστήριο Ανάλυσης Ρευστών και Πυρήνων του Πολυτεχνείου Κρήτης. Η βάση αυτή ελέγχθηκε ποιοτικά και μορφοποιήθηκε ανάλογα με τις ανάγκες του κάθε νευρωνικού για τις συγκεκριμένες ιδιότητες ώστε να θεωρηθεί αξιόπιστη για την εκπαίδευση και παραγωγή της γενίκευσης που εκφράζει τη φυσική μαθηματική σχέση που κρύβεται πίσω από τη μη γραμμική σχέση που συνδέει τα δεδομένα παραγωγής και τις PVT ιδιότητες.

Η παρούσα διπλωματική εργασία χωρίζεται σε δυο μέρη. Το πρώτο αφορά την έρευνα της διεθνούς βιβλιογραφίας ώστε να επιλεγθούν οι κυριότερες και πλέον δημοφιλείς εμπειρικές συσχετίσεις που χρησιμοποιεί η βιομηχανία πετρελαίου και που προβλέπουν τις συγκεκριμένες ιδιότητες και εν συνεχεία να εφαρμοστούν στην επιλεγμένη βάση δεδομένων ώστε να προκύψουν συμπεράσματα με τη χρήση στατιστικών κριτηρίων για την ακρίβεια των εκτιμήσεων τους συγκρίνοντάς τις με τις πραγματικές τιμές των ιδιοτήτων. Το δεύτερο μέρος της παρούσας εργασίας αφορά τη χρήση κώδικα προγραμματισμού σε περιβάλλον Matlab ώστε να παραχθούν νευρωνικά δίκτυα για την πρόβλεψη των PVT ιδιοτήτων που μας αφορούν. Εν συνεχεία με χρήση στατιστικών κριτηρίων γίνεται η επιλογή των βέλτιστων νευρωνικών δικτύων από άποψη ακρίβειας εκτιμήσεων ώστε να συγκριθεί η αποδοτικότητα τους έναντι των εμπειρικών συσχετίσεων σε ένα κοινό υποσύνολο δεδομένων που δεν χρησιμοποιήθηκαν για την εκπαίδευση των νευρωνικών.

Στο πρώτο Κεφάλαιο γίνεται μια γενική αναφορά πάνω στο πετρέλαιο, στη βιομηχανία Πετρελαίου, και στον χαρακτηρισμό των πετρελαίων.

Στο δεύτερο κεφάλαιο γίνεται μια εκτεταμένη παρουσίαση των ογκομετρικών συντελεστών της υγρής και αέριας φάσης και της σημασίας τους όπως επίσης και των μηχανισμών παραγωγής.

Στο τρίτο κεφάλαιο γίνεται παρουσίαση των PVT μελετών

Στο τέταρτο κεφάλαιο παρουσιάζονται οι μέθοδοι εκτίμησης των PVT ιδιοτήτων B_o , R_s , ρ_b και πραγματοποιείται μια αναλυτική αναφορά στις εμπειρικές συσχετίσεις.

Στο πέμπτο κεφάλαιο γίνεται παρουσίαση των νευρωνικών δικτύων.

Στο έκτο κεφάλαιο επεξηγείται αναλυτικά η ανάπτυξη των νευρωνικών όπως επίσης και η παρουσίαση των βέλτιστων νευρωνικών εξ όσων αναπτύχθηκαν για κάθε ζητούμενη ιδιότητα

Στο έβδομο κεφάλαιο γίνεται η σύγκριση της αποδοτικότητας των εκτιμήσεων των βέλτιστων νευρωνικών σε σχέση με την αποδοτικότητα των εμπειρικών συσχετίσεων

Και τέλος παρουσιάζονται τα συμπεράσματα

ΚΕΦΑΛΑΙΟ 5

ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

I. ΕΙΣΑΓΩΓΗ

Ορισμός νευρωνικού δικτύου

Για να οριστεί το νευρωνικό δίκτυο ως προσαρμοστική μηχανή θα χρησιμοποιηθεί ο ορισμός του Simon Haykin που αναφέρεται στο βιβλίο του «Νευρωνικά δίκτυα και μηχανική μάθηση» και διατυπώνεται ως εξής:

«Ένα νευρωνικό δίκτυο είναι ένας τεράστιος παράλληλος επεξεργαστής με κατανεμημένη αρχιτεκτονική, ο οποίος αποτελείται από απλές μονάδες επεξεργασίας και έχει από τη φύση του τη δυνατότητα να αποθηκεύει εμπειρική γνώση και να την καθιστά διαθέσιμη για χρήση. Μοιάζει με τον ανθρώπινο εγκέφαλο σε δύο σημεία:

- 1. Το δίκτυο προσλαμβάνει τη γνώση από το περιβάλλον του, μέσω μιας διαδικασίας μάθησης.*
- 2. Η ισχύς των συνδέσεων μεταξύ των νευρώνων, που αποκαλείται συνοπτικό βάρος, χρησιμοποιείται για την αποθήκευση της γνώσης που αποκτιέται.*

Η διαδικασία μέσω της οποίας επιτυγχάνεται η μάθηση αποκαλείται αλγόριθμος μάθησης και η λειτουργία του είναι να τροποποιεί τα συναπτικά βάρη του δικτύου με τον κατάλληλο τρόπο για την επίτευξη του επιθυμητού στόχου. Η τροποποίηση των συναπτικών βαρών αποτελεί «παράδοσιακή» μέθοδο σχεδιασμού των νευρωνικών. Επίσης ένα νευρωνικό δίκτυο έχει την δυνατότητα να τροποποιεί την τοπολογία του όπως συμβαίνει σε ένα ανθρώπινο εγκέφαλο όταν κάποιοι νευρώνες πεθαίνουν και αναπτύσσονται νέες συναπτικές συνδέσεις.»

Ιστορική Αναδρομή

Το πρώτο μοντέλο νευρωνικού δικτύου το οποίο προτείνει ότι οι νευρώνες είναι η βασική μονάδα του δικτύου παρουσιάσθηκε το 1943 από τους McCulloch και Pitts. Αυτή θεωρείται ιστορικά ότι είναι η πρώτη εικόνα ενός νευρωνικού δικτύου. Ο McCulloch αφιέρωσε περισσότερα από 20 χρόνια έρευνας στην μελέτη του ανθρώπινου εγκεφάλου και νευρικού συστήματος. Όταν λοιπόν συνεργάστηκε με το νεαρό Pitts έκαναν μια δημοσίευση στην οποία αρχικά παρουσίασαν ένα πιθανό σενάριο για τη λειτουργία των νευρώνων και εν συνεχεία ένα πρωταρχικό Νευρωνικό Δίκτυο (ΝΔ), το οποίο είχαν κατασκευάσει χρησιμοποιώντας απλά ηλεκτρικά κυκλώματα.

Ενα άλλο έργο της πρώτης αυτής εποχής που αφήνει ακόμα και σήμερα την επιρροή του είναι το βιβλίο του D. Hebb, "The organization of behavior" (1949), το οποίο εισάγει τον κανόνα μάθησης του Hebb. Το μοντέλο του Hebb έχει ως κεντρική ιδέα τις συνδέσεις μεταξύ μονάδων του συστήματος, δηλαδή τους νευρώνες. Έφτασε στα συμπεράσματα αυτά μετά από σωρεία πειραμάτων νευροφυσιολογίας. Ο κανόνας αυτός λέγει ότι κάθε φορά που το δίκτυο χρησιμοποιεί τις νευρωνικές του συνδέσεις, οι συνδέσεις αυτές ενισχύονται και το δίκτυο πλησιάζει περισσότερο στο να μάθει το πρότυπο το οποίο παρουσιάζεται

Οι εργασίες αυτές πιθανόν να χάνονταν στην βιβλιογραφία αν δεν τις χρησιμοποιούσε ο J. Von Neumann ως παράδειγμα για υπολογιστικές μηχανές την δεκαετία που διαδόθηκε ο ηλεκτρονικός υπολογιστής, δηλ. την δεκαετία του πενήντα. Τότε έγιναν και οι πρώτες προσπάθειες να αντλήσουμε πληροφορίες από τα βιολογικά δίκτυα και να δημιουργηθούν τα πρώτα τεχνητά δίκτυα.

Το μοντέλο του αισθητήρα (perceptron) παρουσιάσθηκε για πρώτη φορά το 1957 από τον F. Rosenblatt, ο οποίος υλοποίησε το πρώτο δίκτυο με αναλογικό εξοπλισμό (π.χ. λυχνίες), και το οποίο μπορούσε να κάνει πολλούς και διάφορους υπολογισμούς. Ήταν ένα πολύ απλό μοντέλο που έχει μόνο δύο επίπεδα, της εισόδου και της εξόδου. Το σήμα προχωρούσε μονόδρομα από την είσοδο στην έξοδο. Το μοντέλο αυτό στην αρχή είχε πολλές επιτυχίες, δημιούργησε μεγάλο ενθουσιασμό, και μάλιστα ήδη άρχιζε να συζητείται η ιδέα ότι πιθανόν τα νευρωνικά δίκτυα να είναι η πιο ανώτερη τεχνική που λύνει όλα τα προβλήματα που μέχρι τότε παρέμεναν άλυτα. Οι πρώτες λοιπόν επιτυχίες μεγαλοποιήθηκαν, αλλά γρήγορα φάνηκε ότι τα μοντέλα αυτά είχαν πολλούς περιορισμούς.

Μια συνολική και εμπειριστατωμένη εικόνα του προτύπου αυτού παρουσιάσθηκε το 1969 στο βιβλίο "Perceptrons" των Minsky και Papert. Στο βιβλίο αυτό γίνεται μία συνολική εκτίμηση της χρησιμότητας του προτύπου του αισθητήρα και όλων των διεργασιών για τα οποία είναι χρήσιμο. Αποδεικνύεται με αναλυτικά μαθηματικά ότι υπάρχουν συγκεκριμένοι περιορισμοί στο πρότυπο αυτό. Έτσι, δεν μπορεί να λύσει, π.χ. το σχετικά απλό πρόβλημα του X-OR (exclusive-or) συνάρτηση της αποκλειστικής διάζευξης. Οι αρχικές προσδοκίες που είχαν δημιουργηθεί ήδη φαίνεται ότι δεν επαληθεύονται, και προς το παρόν τα νευρωνικά δίκτυα χάνουν την δημοτικότητα τους, και ο κόσμος στρέφεται σε μία νέα παρεμφερή περιοχή που τότε άρχισε να γίνεται γνωστή, την Τεχνητή Νοημοσύνη.

Την ίδια περίπου εποχή με την ανάπτυξη του μοντέλου του αισθητήρα, οι Widrow και Hoff ανέπτυξαν το 1959 δύο νέα μοντέλα, το Adaline και το Madaline. Μάλιστα αυτά τα μοντέλα ήταν

τα πρώτα μοντέλα που χρησιμοποιήθηκαν επιτυχώς για πρακτικά προβλήματα: Χρησιμοποιήθηκαν ως φίλτρα για να εξαλείψουν την ηχώ σε τηλεφωνικές γραμμές

Ένα μνημειώδες έργο παρουσιάστηκε το 1982 από τον J. Hopfield, ο οποίος είναι βιολόγος, και το οποίο έδωσε μεγάλη ώθηση στην ανάπτυξη των δικτύων. Σε μία εργασία του μόλις 5 σελίδων, ο Hopfield έδειξε με αυστηρά μαθηματική απόδειξη πως ένα νευρωνικό δίκτυο μπορεί να χρησιμοποιηθεί ως αποθηκευτικός χώρος (storage device), και πως επίσης μπορεί ένα δίκτυο να επανακτήσει όλη την πληροφορία ενός συστήματος έστω και αν του δοθούν μερικά τμήματα μόνο, και όχι ολόκληρο το σύστημα. Αμέσως εκτιμήθηκε η σπουδαιότητα της ιδιότητας αυτής, και ως εκ τούτου η εργασία αυτή έγινε έμπνευση για πολλές άλλες ιδέες που ακολούθησαν.

Το 1986 δημοσιεύεται ένα άλλο σημαντικό έργο από τους McClelland και Rumelhart, το "Parallel Distributed Processing" το οποίο ανοίγει νέους δρόμους στην εκπαίδευση των νευρωνικών δικτύων. Παρουσιάζεται η ιδέα πως ένα νευρωνικό δίκτυο μπορεί να θεωρηθεί και χρησιμοποιηθεί ως παράλληλος επεξεργαστής. Το έργο αυτό κάνει ένα σημαντικό βήμα πέραν από το Perceptron, με το να επιτρέπει την ύπαρξη και άλλων επιπέδων νευρώνων, εκτός από την είσοδο και την έξοδο, που αποτελούν την εσωτερική δομή του δικτύου. Προτείνουν μία νέα διαδικασία εκπαίδευσης, την μέθοδο της οπισθοδιάδοσης (back-propagation), η οποία κατέληξε να είναι η πιο χρήσιμη σήμερα τεχνική εκπαίδευσης δικτύων. Η μέθοδος αυτή είχε συζητηθεί και από άλλους νωρίτερα, αλλά για πρώτη φορά το 1986 παρουσιάστηκε ολοκληρωμένα και με αυστηρό μαθηματικό τρόπο.

Από το 1985 και μετά αρχίζουν τα πρώτα συνέδρια που είναι αφιερωμένα αποκλειστικά σε νευρωνικά δίκτυα, από την IEEE(Institute of Electrical and Electronics Engineers). Ταυτόχρονα δημιουργούνται ειδικές επαγγελματικές εταιρίες νευρωνικών δικτύων με χιλιάδες μέλη, όπως η International Neural Network Society με τρεις πόλους: Αμερική (με διευθυντή τον Grossberg), Ευρώπη (Kohonen), και Ιαπωνία (Amari). Προς τα τέλη της δεκαετίας του ογδόντα παρουσιάζονται τουλάχιστον πέντε (5) νέα περιοδικά αφιερωμένα αποκλειστικά στα νευρωνικά δίκτυα, ενώ πριν λίγα χρόνια δεν υπήρχε ούτε ένα. Πολύ σημαντικό είναι επίσης το γεγονός ότι τα τελευταία δεκαπέντε χρόνια δημιουργήθηκαν και οι πρώτες εμπορικές εταιρίες οι οποίες ασχολούνται αποκλειστικά με νευρωνικά δίκτυα. Βρίσκονται σχεδόν όλες στις ΗΠΑ, συνήθως έχουν μικρό αριθμό εργαζομένων (π.χ. 20 άτομα), και παράγουν εξειδικευμένα προγράμματα για την λύση συγκεκριμένων προβλημάτων. Μερικά από αυτά έχουν επιτυχία, αλλά διαφαίνεται με το πέρασμα του χρόνου ότι οι αρχικές προσδοκίες για ραγδαία αύξηση των εμπορικών εφαρμογών δεν επαληθεύονται. Έχουν φθάσει πλέον σε ένα σταθερό επίπεδο ανάπτυξης, ενώ η ετήσια αύξηση είναι μικρή

Συνοψίζοντας, το ιστορικό ενδιαφέρον για τα νευρωνικά δίκτυα προέρχεται από δύο πηγές:

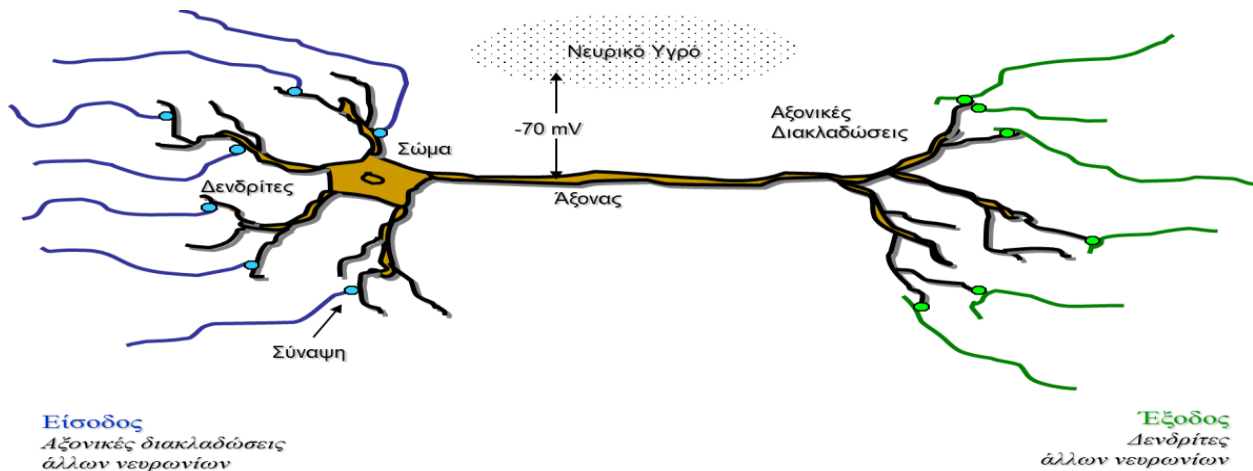
- από την θέληση να γνωρίσουμε τον τρόπο λειτουργίας του εγκεφάλου, και
- από την θέληση να φτιάξουμε μηχανές που είναι ικανές να επιτελούν πολύ περίπλοκες πράξεις και οι οποίες δεν γίνονται με επιτυχία από τον σειριακό τρόπο λειτουργίας του μοντέλου του von Neumann.

Η ανάπτυξη των νευρωνικών δικτύων πέρασε πολλές φάσεις, άλλες από τις οποίες ήταν πολύ ενδιαφέρουσες με μεγάλα επιτεύγματα, και άλλες όχι τόσο. Απαριθμούν μια ιστορία περίπου 40-50 ετών. Το βέβαιο είναι ότι από το 1980 και μετά η περιοχή αυτή έχει πάρει την θέση της ως μία νέα ειδικότητα διαθεματικού χαρακτήρα, όπως φαίνεται από την έρευνα που γίνεται καθημερινά, τα σχετικά δημοσιεύματα και τις εμπορικές εφαρμογές που κυκλοφορούν.

Βιολογικά νευρωνικά δίκτυα

Τα τεχνητά νευρωνικά δίκτυα ανήκουν στον τομέα της **Μη Συμβολικής Τεχνητής Νοημοσύνης** (non symbolic AI), η οποία προσομοιώνει βιολογικές διαδικασίες, όπως τη λειτουργία του εγκεφάλου ή τη διαδικασία εξέλιξης των ειδών, και η οποία διαφέρει από τη **Συμβολική Τεχνητή Νοημοσύνη** (symbolic AI), που προσομοιώνει τον τρόπο σκέψης του ανθρώπου, χρησιμοποιώντας ως δομικές μονάδες τα σύμβολα (ένα σύμβολο μπορεί να αναπαριστά μια έννοια ή μια σχέση ανάμεσα σε έννοιες).

Η ικανότητα του ανθρώπου να σκέφτεται, να θυμάται και να επιλύει προβλήματα εντοπίζεται στον εγκέφαλό του. Όπως είναι γνωστό από τη Βιολογία, η δομική μονάδα του εγκεφάλου είναι ο νευρώνας. Ένας τυπικός βιολογικός νευρώνας αποτελείται από το σώμα, που συνιστά τον πυρήνα του, τους δένδριτες, μέσω των οποίων λαμβάνει σήματα από γειτονικούς νευρώνες (σημεία εισόδου), και τον άξονα, που αποτελεί την έξοδο του νευρώνα και το μέσο σύνδεσής του με τους άλλους νευρώνες. Σε κάθε δένδριτη υπάρχει ένα απειροελάχιστο κενό που ονομάζεται σύναψη. Οι συνάψεις μέσω χημικών διαδικασιών επιταχύνουν ή επιβραδύνουν τη ροή ηλεκτρικών φορτίων προς το σώμα του νευρώνα. Η ικανότητα μάθησης και μνήμης που εμφανίζει ο εγκέφαλος οφείλεται στην ικανότητα των συνάψεων να μεταβάλλουν την αγωγιμότητά τους. Τα ηλεκτρικά σήματα που εισέρχονται στο σώμα μέσω των δένδριτών συνδυάζονται και, εφόσον το αποτέλεσμα ξεπερνά κάποια τιμή κατωφλίου, το σήμα διαδίδεται με τη βοήθεια του άξονα προς άλλους νευρώνες.



Σχήμα 5.1 Γραφική απεικόνιση ενός τυπικού βιολογικού νευρώνα εγκεφάλου.

Ο εγκέφαλος είναι σε θέση να λαμβάνει πολύπλοκες αποφάσεις εκπληκτικά γρήγορα παρά το ότι ο χρόνος απόκρισης των βιολογικών νευρώνων είναι της τάξης των χιλιοστών του δευτερολέπτου. Κατά μια άποψη, αυτό οφείλεται στο ότι η υπολογιστική ικανότητα του εγκεφάλου και η πληροφορία που περιέχει είναι διαμερισμένη σε όλο του τον όγκο. Πρόκειται, δηλαδή, για ένα παράλληλο και κατανεμημένο υπολογιστικό σύστημα. Αυτά τα χαρακτηριστικά διαμορφώνουν το κυριότερο κίνητρο πίσω από την επιθυμία να μοντελοποιηθεί ο ανθρώπινος εγκέφαλος με τη χρήση τεχνητών νευρωνικών δικτύων.

II. ΕΦΑΡΜΟΓΕΣ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Τα τεχνικά νευρωνικά δίκτυα χρησιμοποιούνται ευρέως τα τελευταία χρόνια σε διάφορους τομείς λόγω της υπολογιστικής ταχύτητας, της ικανότητας αντιμετώπισης μη γραμμικών προβλημάτων και της ικανότητας τους να αναγνωρίζουν τις σχέσεις μεταξύ ποσοτήτων που είναι δύσκολο να μοντελοποιηθούν.

Η χρήση των τεχνητών νευρωνικών δικτύων σε διάφορες εφαρμογές προσφέρει ευκολία στην υλοποίηση, σχετικά αξιόπιστη λειτουργία και άμεση απόκριση. Οι εφαρμογές αυτές περιλαμβάνουν αναγνώριση προτύπων, υπολογισμό συναρτήσεων, βελτιστοποίηση, πρόβλεψη τιμών, συστήματα αυτομάτου ελέγχου κ.λ. Κάποιες συγκεκριμένες περιοχές εφαρμογών είναι οι ακόλουθες: Βιολογία, Ιατρική, Βιομηχανία, Επεξεργασία σημάτων, Αεροπλοΐα. Επίσης χρησιμο-

ποιοούνται εντατικά από το χρηματιστηριακό κλάδο όπως τράπεζες, ασφαλιστικές εταιρίες, εταιρίας πιστωτικών καρτών αλλά και χρηματιστηριακές εταιρίες όπως αυτές που είναι εισηγμένες στη Wall street.

Τα τεχνητά νευρωνικά δίκτυα, εφαρμόζονται ευρέως και στην βιομηχανία πετρελαίου από την δεκαετία του 1970. Χρησιμοποιώντας τα νευρωνικά δίκτυα, ο μηχανικός πετρελαίου μπορεί να επιτύχει βελτιστοποίηση της παραγωγής και των λογισμικών προσομοίωσης, να εκτιμήσει τιμές ιδιοτήτων του ρευστού και να επιλύσει προβλήματα ταξινόμησης. Εν συνεχεία αναφέρονται πεδία της μηχανικής πετρελαίου όπου τα νευρωνικά δίκτυα έχουν βρει εφαρμογή:

- Στο χαρακτηρισμό και ταξινόμηση ταμιευτήρα (reservoir characterization)
- Στην συμπλήρωση πηγαδιού (well completion)
- Στους χρόνους που σχετίζονται με τη τσιμέντωση του πηγαδιού (well cementing)
- Στην ερμηνεία δεδομένων από το τεστ παραγωγής (well-test interpretation)
- Στην ερμηνεία διαγραφιών (well logging)
- Στη προσομοίωση ταμιευτήρα (reservoir simulation)
- Στην εκτίμηση των συντελεστών ισορροπίας (prediction of equilibrium k-values)
- Στην εκτίμηση των ογκομετρικών συντελεστών του ταμιευτήρα (prediction of formation oil volume factor of crude oils)
- Στην εκτίμηση της πίεσης φυσαλίδας-κορεσμού (prediction bubble-point pressure)
- Στην πρόβλεψη της καθίζησης των ασφατενίων
- Ως εργαλείο ελέγχου ποιότητας, επί των προβλέψεων προσομοιωτών θερμοδυναμικής συμπεριφοράς των ρευστών υδρογονανθράκων
- Στη πρόβλεψη της θερμοκρασίας σχηματισμού κεριών-παραφινών

III. ΟΡΙΣΜΟΣ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΠΡΟΣ ΕΠΙΛΥΣΗ ΤΗΣ ΠΑΡΟΥΣΑΣ ΔΙΠΛΩΜΑΤΙΚΗΣ

Το πρόβλημα το οποίο καλείται να αντιμετωπιστεί στα πλαίσια της παρούσας διπλωματικής είναι ένα πρόβλημα εποπτευόμενης παλινδρόμησης (supervised regression problem). Η **παλινδρόμηση** είναι μια ευρέως χρησιμοποιημένη στατιστική τεχνική μοντελοποίησης για την έρευνα της συσχέτισης μεταξύ μίας εξαρτώμενης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Στη συγκεκριμένη περίπτωση, ως εξαρτημένες μεταβλητές θεωρήθηκαν οι ιδιότητες του ρευστού (R_s, B_o, P_b) οι οποίες θα αποτελέσουν και τις εξόδους των νευρωνικών και ως ανεξάρτητες μεταβλητές-είσοδοι τα δεδομένα της βάσης δεδομένων που θα επιλεγούν για να χρησιμοποιηθούν για την εκπαίδευση τους. Προϊόν της παλινδρόμησης, όταν χρησιμοποιείται ως

τεχνική εξόρυξης δεδομένων(data mining), αποτελεί το μοντέλο που χρησιμοποιείται αργότερα για να προβλέψει τις τιμές των ζητούμενων ιδιοτήτων για νέα άγνωστα δεδομένα.

Για την επίλυση του προβλήματος παλινδρόμησης υπάρχουν πολλά μοντέλα τα οποία μπορούν να εφαρμοστούν στο λεγόμενο machine learning (μάθηση μηχανής) όπως:

- SUPPORT VECTOR MACHINES
- GAUSSIAN PROCESSES
- **ARTIFICIAL NEURAL NETWORKS = ANN**
- CLASSIFICATION + REGRESSION TREES
- RANDOM FORESTS

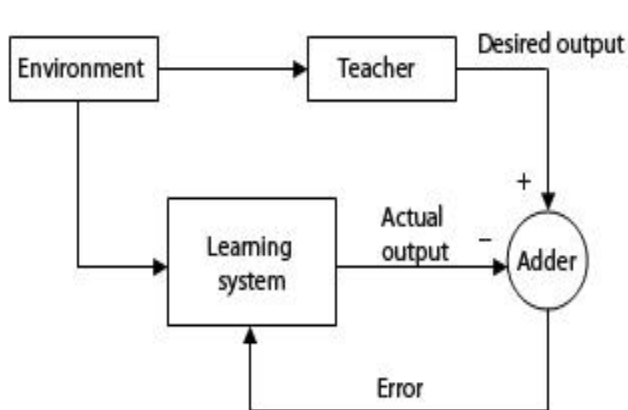
Στη συγκεκριμένη περίπτωση, χρησιμοποιήθηκε το μοντέλο των τεχνητών νευρωνικών δικτύων και ποιά συγκεκριμένα έγινε χρήση της δομής «Perceptron πολλών επιπέδων» υιοθετώντας ως μέθοδο εκπαίδευσης τον αλγόριθμο “back propagation” οπισθοδιάδοσης όπως μεταφράζεται στα ελληνικά. Επίσης, σε αυτό το σημείο θα πρέπει να επισημανθεί, ότι το μοντέλο είναι εποπτευόμενης παλινδρόμησης. Η χρησιμοποίηση του όρου «εποπτευόμενη» αναφέρεται στο γεγονός ότι η μάθηση του νευρωνικού πραγματοποιείται υπό την επίβλεψη ενός εκπαιδευτή ώστε να διασφαλιστεί ότι οι δεδομένες σχέσεις εισόδων-εξόδων έχουν οριστεί σωστά, ανάλογα με το περιβάλλον της συσχέτισης για την εκπλήρωση των αναγκών του προβλήματος παλινδρόμησης.

ΜΑΘΗΣΗ ΜΕ ΕΚΠΑΙΔΕΥΤΗ

Η επιβλεπόμενη μάθηση στην ουσία περιγράφει ότι ο εκπαιδευτής έχει γνώση του περιβάλλοντος και αυτή η γνώση αντιπροσωπεύεται από ένα σύνολο παραδειγμάτων εισόδου-εξόδου, ωστόσο το περιβάλλον είναι άγνωστο για το νευρωνικό. Η συσχέτιση των πληροφοριών εισόδου (y_o, y_g, GOR, p_b, T) και εξόδου (πειραματικές τιμές B_o, p_b, GOR) για κάθε ζητούμενη ιδιότητα ώστε να επιτευχθεί βέλτιστο διάνυσμα εκπαίδευσης πραγματοποιείται από τον εκπαιδευτή. Στη παρούσα διπλωματική, οι συνθήκες του περιβάλλοντος ορίζονται σύμφωνα με τις ανάγκες της Μηχανικής Ταμιευτήρων. Η επιθυμητή απόκριση που ορίζεται από τον εκπαιδευτή αντιπροσωπεύει τη βέλτιστη ενέργεια που πρέπει να εκτελείται από το νευρωνικό δίκτυο. Οι παράμετροι του δικτύου προσαρμόζονται υπό τη συνδυασμένη επιρροή του διανύσματος εκπαίδευσης και του σήματος σφάλματος. Το σήμα σφάλματος ορίζεται ως η διαφορά μεταξύ της πειραματικής τιμής των ιδιοτήτων και της εκτιμώμενης τιμής από το νευρωνικό. Αυτή η προσαρμογή εκτελείται

με επαναληπτικό τρόπο, βήμα προς βήμα, με στόχο να φέρει τελικά το νευρωνικό δίκτυο σε μια κατάσταση όπου θα έχει εκπαιδευτεί ως προς την συμπεριφορά του διατιθέμενου πληθυσμού

Με αυτό τον τρόπο, η γνώση του περιβάλλοντος που είναι διαθέσιμη στην εκπαίδευση, δηλαδή η σχέση μεταξύ των inputs (y_0, y_g, GOR, p_b, T) και των outputs (πειραματικές τιμές B_0, p_b, GOR), μεταφέρεται στο νευρωνικό δίκτυο μέσω της εκπαίδευσης και αποθηκεύεται με τη μορφή σταθερών συναπτικών βαρών, τα οποία αντιπροσωπεύουν μακροπρόθεσμη μνήμη. Όταν επιτευχθεί αυτή η συνθήκη, το νευρωνικό θεωρείται πως περιέχει ικανή γνώση ώστε να μπορεί να προβλέπει με ακρίβεια άγνωστες περιπτώσεις. Η μορφή της επιβλεπόμενης μάθησης που μόλις περιεγράφηκε αποτελεί την βάση της μάθησης μέσω διόρθωσης σφάλματος. Ως δείκτη απόδοσης για το σύστημα, μπορούμε να χρησιμοποιήσουμε το μέσο τετραγωνικό σφάλμα ή το άθροισμα



των τετραγώνων των σφαλμάτων επί του δείγματος εκπαίδευσης, ορισμένο ως συνάρτηση των ελεύθερων παραμέτρων δηλαδή των συναπτικών βαρών του συστήματος.

Σχήμα 5.2 Σχηματικό διάγραμμα της μάθησης με εκπαιδευτή

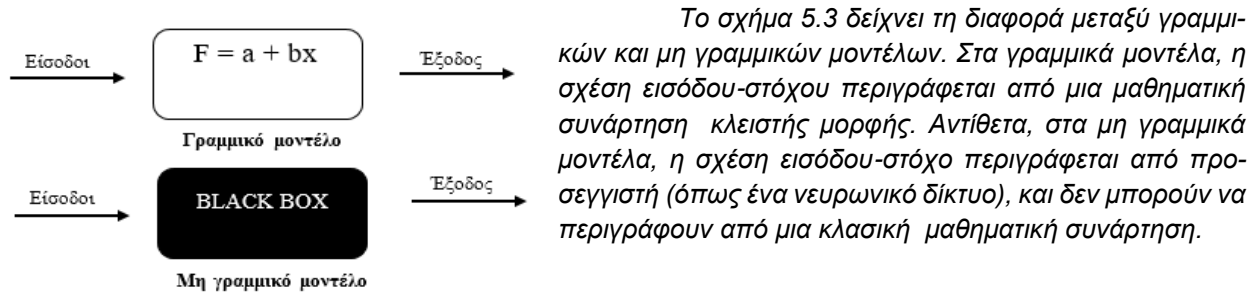
IV. PERCEPTRON ΠΟΛΛΩΝ ΕΠΙΠΕΔΩΝ

Η δομή του νευρωνικού δικτύου Perceptron πολλών επιπέδων μπορεί να χαρακτηριστεί με τις εξής τρεις παρατηρήσεις:

- Το μοντέλο κάθε νευρώνα στο δίκτυο περιλαμβάνει μια μη γραμμική συνάρτηση ενεργοποίησης, η οποία είναι διαφορίσιμη.
- Το δίκτυο περιέχει ένα ή περισσότερα επίπεδα τα οποία παραμένουν κρυφά (hidden) για τους κόμβους των επιπέδων εισόδου και εξόδου.
- Το δίκτυο επιδεικνύει μεγάλη διασυνδεσιμότητα ο βαθμός της οποίας καθορίζεται από τα συναπτικά βάρη του δικτύου.

Αυτά τα χαρακτηριστικά είναι και επίσης υπεύθυνα για την δυσκολία κατανόησης εκ μέρους μας της συμπεριφοράς του δικτύου. Συχνά προσδιορίζεται και ως «black box». Πρώτον

γιατί η μη γραμμικότητα και η υψηλή διασυνδεσιμότητα του το καθιστούν δύσκολο στη θεωρητική ανάλυση και δεύτερον γιατί η χρήση κρυφών νευρωνών καθιστά δυσκολότερη την κατανόηση και σχηματοποίηση της διαδικασίας μάθησης.



Σχήμα 5.3

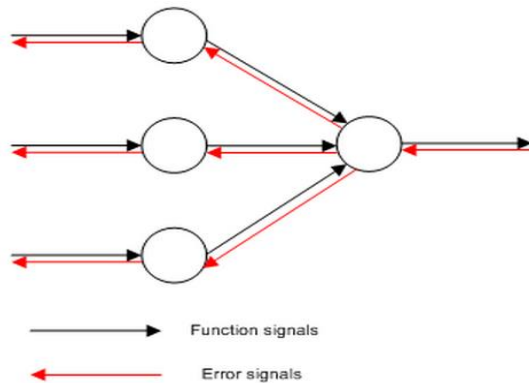
ΑΛΓΟΡΙΘΜΟΣ BACK PROPAGATION

Μια δημοφιλής μέθοδος για την εκπαίδευση των perceptron πολλών επιπέδων είναι ο αλγόριθμος «back-propagation». Η χρήση του όρου εμφανίζεται στο προσκήνιο μετά το 1985 όταν απέκτησε ευρεία απήχηση χάρη στο σημαντικό βιβλίο με τίτλο “Parallel Distributed Processing” (Rumelhart και McClelland, 1986). Η ανάπτυξη του αλγορίθμου «back-propagation» στα μέσα της δεκαετίας το '80 αποτέλεσε ένα ορόσημο στην εξέλιξη του κλάδου των νευρωνικών δικτύων διότι παρείχε μια υπολογιστικής αποτελεσματική μέθοδο για την εκπαίδευση των perceptron πολλών επιπέδων.

Η οπισθοδιάδοση είναι μια συγκεκριμένη τεχνική υλοποίησης βαθμωτής κατάβασης (gradient descent) στο χώρο βαρών για ένα perceptron πολλών επιπέδων. Η βασική ιδέα σχετίζεται με τον αποτελεσματικό υπολογισμό των μερικών παραγώγων μιας συνάρτησης προσέγγισης $F(w, x)$ η οποία υλοποιείται από το δίκτυο, ως προς όλα τα στοιχεία του προσαρμοσίμου διανύσματος βαρών w για μια δεδομένη τιμή του διανύσματος εισόδου x . Σε αυτό οφείλεται η υπολογιστική ισχύς του αλγορίθμου. Εδώ θα πρέπει να σημειωθεί ότι η εκπαίδευση πραγματοποιείται σε δύο φάσεις:

- i. Στη φάση που εξελίσσεται προς τα εμπρός (feedforward), όπου τα συναπτικά βάρη του δικτύου είναι σταθερά και το σήμα εισόδου διαδίδεται διαμέσου του δικτύου, επίπεδο προς επίπεδο, μέχρι να φτάσει στην έξοδο. Συνεπώς, σε αυτή τη φάση οι αλλαγές περιορίζονται στα δυναμικά ενεργοποίησης και στις εξόδους των νευρωνικών του δικτύου.

- ii. Στη φάση που εξελίσσεται προς τα πίσω «Back-propagation», παράγεται ένα σήμα σφάλματος μέσω της σύγκρισης της εξόδου του δικτύου με μια επιθυμητή απόκριση. Το σήμα σφάλματος που προκύπτει, διαδίδεται διαμέσου του δικτύου, ξανά επίπεδο προς επίπεδο, αλλά αυτή τη φορά η διάδοση γίνεται με κατεύθυνση προς τα πίσω. Σε αυτή τη δεύτερη φάση γίνονται διαδοχικές προσαρμογές στα συναπτικά βάρη του δικτύου.



Σχήμα 5.4

Σχήμα 5.4 Οι κατευθύνσεις της ροής των βασικών σημάτων σε ένα *perceptron* πολλών επιπέδων: διάδοση των λειτουργικών σημάτων προς τα εμπρός και διάδοση των σημάτων σφάλματος προς τα πίσω.

Μερικά από τα πλεονεκτήματα της μεθόδου «Back-propagation» :

- Λίγες παράμετροι προς ρύθμιση
- Αλγόριθμος εύκολος σε υλοποίηση
- Μπορεί να εφαρμοστεί σε ευρεία περιοχή δεδομένων
- Η υπολογιστική πολυπλοκότητα του αλγορίθμου είναι γραμμική στο W (συναπτικά βάρη του δικτύου) Στο πέρασμα με κατεύθυνση προς τα εμπρός, οι μόνοι υπολογισμοί που εμπλέκουν τα συναπτικά βάρη είναι εκείνοι που αφορούν τα τοπικά πεδία των διαφόρων νευρώνων του δικτύου. Στο πέρασμα με κατεύθυνση προς τα πίσω, οι μόνοι υπολογισμοί που εμπλέκουν τα συναπτικά βάρη είναι εκείνοι που αφορούν τις τοπικές κλίσεις των κρυφών νευρώνων και την ενημέρωση των ιδίων των συναπτικών βαρών.
- Πλεονεκτεί ο αλγόριθμος στις αναλύσεις ευαισθησίας για την αντιστοίχιση εισόδου-εξόδου
- Ο αλγόριθμος είναι εύρωστος υπό την έννοια ότι διαταραχές με μικρή ενέργεια μπορούν να προκαλέσουν μόνο μικρά σφάλματα εκτίμησης.
- Πολύ δημοφιλής

Μερικά από τα μειονεκτήματα της μεθόδου «Back-propagation»:

- Η εκμάθηση είναι αργή
- Τα νέα στοιχεία θα υπερκαλύπτουν τα παλαιά στοιχεία του προηγούμενου βήματος της εκπαίδευσης και δεν θα είναι δυνατόν να μελετηθούν αναλυτικά τα στοιχεία των βημάτων.
- Δύσκολο να διατηρηθεί το δίκτυο ενημερωμένο για το λόγω του ότι για να προστεθούν καινούριος πληροφορίες θα πρέπει να επανεκπαιδευτεί από την αρχή.
- Το δίκτυο είναι ουσιαστικά black box
- Θα πρέπει να ληφθεί υπόψη το πρόβλημα της «υπερπροσαρμογής»(overfitting)
- Η απόδοση του αλγορίθμου επηρεάζεται με την παρουσία τοπικών ελαχίστων σε συνδυασμό με παρουσία γενικών ελαχίστων.

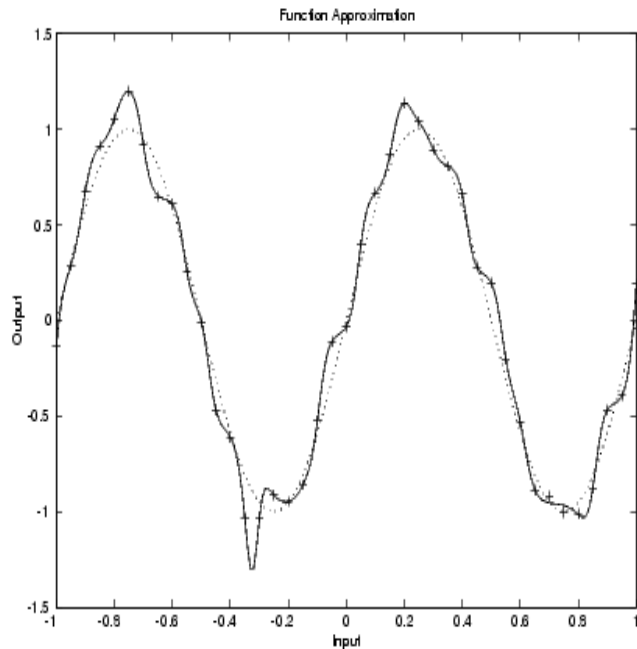
V. ΠΡΟΒΛΗΜΑΤΑ ΣΤΗΝ ΕΚΠΑΙΔΕΥΣΗ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ

A. Ο στόχος της γενίκευσης και το πρόβλημα της υπερπροσαρμογής (overfitting)

Στη μηχανική μάθηση (machine learning), η υπερπροσαρμογή συμβαίνει όταν ένα στατιστικό μοντέλο περιγράφει τυχαίο σφάλμα ή θόρυβο αντί για την υποκείμενη σχέση. Η εκπαίδευση του νευρωνικού έχει ως στόχο την βέλτιστη γενίκευση. Ένα δίκτυο λέγεται ότι γενικεύει επαρκώς όταν η αντιστοίχιση εισόδου-εξόδου που υπολογίζεται από το δίκτυο είναι καλύτερη από ένα όριο που έχει τεθεί για ένα σύνολο δεδομένων ελέγχου που δεν χρησιμοποιήθηκαν κατά την εκπαίδευση του δικτύου. Το σύνολο δεδομένων ελέγχου αντλήθηκε από την συνολική βάση δεδομένων που χρησιμοποιήθηκε για την παραγωγή του συνόλου δεδομένων εκπαίδευσης.

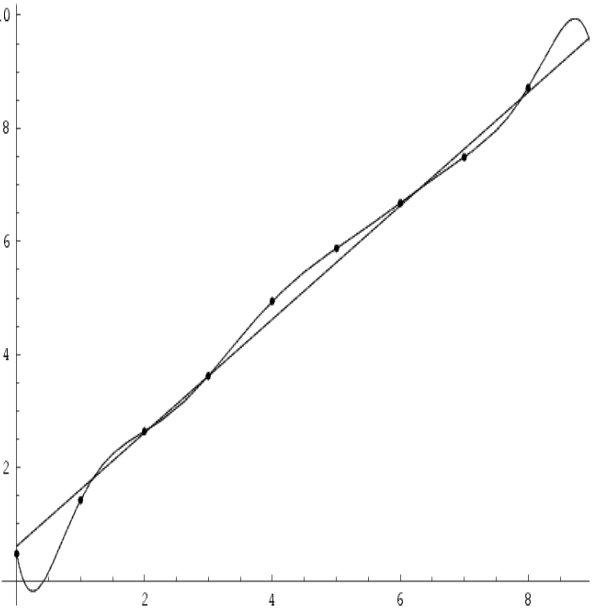
Ένα νευρωνικό δίκτυο το οποίο είναι σχεδιασμένο να γενικεύει καλά, θα παραγάγει σωστή αντιστοίχιση εισόδου-εξόδου ακόμα και όταν η είσοδος του είναι ελαφρώς διαφορετική από τα παραδείγματα που χρησιμοποιήθηκαν για την εκπαίδευσή του. Εδώ θα πρέπει να σημειωθεί όμως ότι ένα νευρωνικό δίκτυο μπορεί αντί να αποκτήσει γενική γνώση από τα παραδείγματα εισόδου-εξόδου, να απομνημονεύσει τα δεδομένα εκπαίδευσης. Αυτό μπορεί να συμβεί εάν λόγω ύπαρξης κάποιων προβλημάτων (π.χ. θόρυβος, εσφαλμένα δεδομένα) στα δεδομένα εκπαίδευσης, η απόκτηση γενικευμένης γνώσης αποτυγχάνει να μειώσει περαιτέρω το σφάλμα εκμάθησης. Αυτό το φαινόμενο αναφέρεται και ως υπερπροσαρμογή(overfitting) ή υπερεκπαίδευση (overtraining) του δικτύου. Όταν το δίκτυο είναι υπερεκπαιδευμένο (δηλαδή εστιάζει αποκλειστικά

στα δεδομένα εκπαίδευσης στερούμενο της γενικότερης εποπτείας της συνάρτησης που συνδέει τις εισόδους με τις εξόδους τότε χάνει την δυνατότητα να γενικεύει βασιζόμενο σε παρόμοια πρότυπα εισόδου-εξόδου. Συνήθως, όταν σε ένα perceptron πολλών επιπέδων γίνεται χρήση περισσότερων κρυφών νευρώνων από όσους είναι πραγματικά αναγκαίοι τότε έχει σαν αποτέλεσμα την αποθήκευση στα συναπτικά βάρη του δικτύου ο «θόρυβος» (γιατί του δίνεται η δυνατότητα να μειώσει περαιτέρω το σφάλμα προσαρμοζόμενο επί των επι μέρους σημείων).



Σχήμα 5.5

Το παραπάνω σχήμα δείχνει την απόκριση ενός 1-20-1 νευρωνικού δικτύου που έχει εκπαιδευτεί για να προσεγγίσει μια θορυβώδη ημιτονοειδή συνάρτηση. Η υποκείμενη ημιτονοειδή συνάρτηση φαίνεται από την διακεκομμένη γραμμή, στις θορυβώδεις μετρήσεις δόθηκαν τα σύμβολα +, και η απόκριση του νευρωνικού δικτύου δίνεται από τη σταθερή γραμμή. Είναι σαφές ότι αυτό το δίκτυο έχει υπερ-εκπαιδεύσει τα δεδομένα και δεν θα γενικεύσει καλά Πηγή (Google)

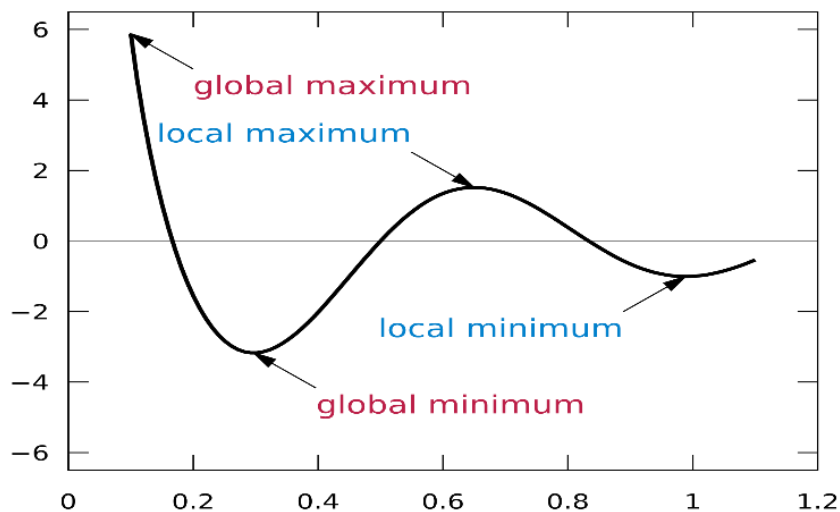


Σχήμα 5.6

Θορυβώδη (σχεδόν γραμμικά) δεδομένα προσομοιώθηκαν από μια γραμμική και μια πολυωνυμική συνάρτηση. Ενώ στην πολυωνυμική είχαμε εξαιρετική προσαρμογή, η γραμμική έχασε τα περισσότερα δεδομένα. Παρόλα αυτά θεωρείται ότι η γραμμική συνάρτηση θα γενικεύσει καλύτερα. Πηγή (Google)

B. Το πρόβλημα των τοπικών ελαχίστων

Μια ιδιομορφία της επιφάνειας σφάλματος που επηρεάζει την απόδοση του αλγορίθμου «back-propagation» είναι η ύπαρξη πολλών διαφορετικών τοπικών ελαχίστων στα οποία μπορεί να εγκλωβιστεί ο αλγόριθμος κατά τη διαδικασία της σύγκλισης. Γενικά, είναι δύσκολο να καθοριστεί το πλήθος των τοπικών ελαχίστων που μπορούν να υπάρχουν στην επιφάνεια σφάλματος. Εφόσον η μάθηση με οπισθοδιάδοση σφάλματος είναι ουσιαστικά μια τεχνική που κάνει μικρά άλματα στη συνάρτηση (hill climbing), διατρέχει τον κίνδυνο να παγιδευτεί σε ένα τοπικό ελάχιστο όπου κάθε μικρή αλλαγή στα συναπτικά βάρη να αυξάνει τη συνάρτηση κόστους. Όμως, κάπου αλλού στο χώρο βαρών υπάρχει ένα άλλο σύνολο συναπτικών βαρών για το οποίο η συνάρτηση κόστους είναι μικρότερη από το τοπικό ελάχιστο στο οποίο έχει εγκλωβιστεί το δίκτυο κατά την εκπαίδευσή του. Προφανώς, το να τερματίζει η διαδικασία μάθησης σε ένα τοπικό ελάχιστο είναι ανεπιθύμητο κυρίως εάν αυτό βρίσκεται πολύ πάνω από το γενικό ελάχιστο της τοπολογίας σφάλματος του προβλήματος.



Σχήμα 5.7 γραφική απεικόνιση της διαφοράς του τοπικού ελαχίστου και μεγίστου σε σχέση με το γενικό ελάχιστο και μέγιστο.

VI. ΜΕΘΟΔΟΣ LEVENBERG-MARQUARDT

Για την εκπαίδευση των νευρωνικών δικτύων σε περιβάλλον Matlab χρησιμοποιήθηκε η μέθοδος Levenberg-Marquardt. Ο αλγόριθμος Levenberg-Marquardt, ο οποίος έχει αναπτυχθεί ανεξάρτητα από τους Kenneth Levenberg και Donald Marquardt, παρέχει λύση στο πρόβλημα της ελαχιστοποίησης μη γραμμικών εξισώσεων. Σαν μέθοδος είναι γρήγορη και έχει σταθερή σύγκλιση. Στο πεδίο των τεχνητών νευρωνικών δικτύων, ο αλγόριθμος αυτός χρησιμοποιείται κυρίως για την εκπαίδευση των μικρής και μεσαίας διαστατικότητας προβλημάτων όπως στη περίπτωση της παρούσας εργασίας.

Πολλές άλλες μέθοδοι έχουν ήδη αναπτυχθεί για την εκπαίδευση των νευρωνικών δικτύων. Ο αλγόριθμος της βαθμωτής κατάβασης γνωστός και ως «back-propagation» (EBP) θεωρείται παρόλα τα προβλήματα του μια από τις πιο σημαντικές ανακαλύψεις στον τομέα της εκπαίδευσης των νευρωνικών δικτύων. Πολλές βελτιώσεις έχουν γίνει με τη πάροδο του χρόνου στον αλγόριθμο «back-propagation» αλλά αυτές οι βελτιώσεις είχαν σχετικά μικρά αποτελέσματα. Όμως, ο συγκεκριμένος αλγόριθμος εξακολουθεί να χρησιμοποιείται ευρέως.

Ο αλγόριθμος αυτός θεωρείται ότι είναι αναποτελεσματικός λόγω της αργής σύγκλισης που πραγματοποιεί και που οφείλεται κυρίως σε δύο λόγους. Ο πρώτος είναι ότι τα μεγέθη των βημάτων της κατάβασης πρέπει να είναι αντίστοιχα των κλίσεων (gradients). Λογικά, θα πρέπει να λαμβάνονται μικρά σε μέγεθος βήματα όπου η κλίση είναι απότομη, έτσι ώστε να μην χαθεί το ζητούμενο ελάχιστο από την ταλάντωση. Έτσι, αν παρομοιαστεί το μέγεθος του βήματος με μια σταθερά, θα πρέπει η σταθερά να είναι μικρή. Στη συνέχεια, όπου η κλίση είναι ομαλή, η διαδικασία της εκπαίδευσης θα πρέπει να είναι πολύ αργή. Ο δεύτερος λόγος είναι ότι η καμπυλότητα της επιφάνειας του σφάλματος μπορεί να μην είναι η ίδια σε όλες τις κατευθύνσεις, οπότε το κλασικό «error valley» πρόβλημα (ref) μπορεί να υπάρχει και μπορεί να οδηγήσει στην αργή σύγκλιση.

Η αργή σύγκλιση της μεθόδου της βαθμωτής κατάβασης μπορεί να βελτιωθεί σε μεγάλο βαθμό από τον αλγόριθμο της μεθόδου Gauss-Newton. Κάνοντας χρήση μερικών παραγώγων δεύτερης τάξης της συνάρτησης κόστους για τη "φυσική" αξιολόγηση της καμπυλότητας της επιφάνειας λάθους, ο αλγόριθμος Gauss-Newton μπορεί να βρει τα κατάλληλα από άποψη μεγέθους βήματα για τη κατάβαση προς κάθε κατεύθυνση ώστε να επιτευχθεί γρήγορη σύγκλιση. Ειδικά, εάν η συνάρτηση κόστους έχει μια τετραγωνική επιφάνεια, μπορεί να συγκλίνει απευθείας κατά τη πρώτη επανάληψη. Αλλά αυτή η βελτίωση θα συμβεί μόνο όταν η τετραγωνική προσέγγιση της συνάρτησης κόστους είναι δυνατή. Σε αντίθετη περίπτωση, ο αλγόριθμος Gauss-Newton θα είναι ως επί το πλείστον αποκλίνων.

Ο αλγόριθμος Levenberg-Marquardt συνδυάζει τη μέθοδο της βαθμωτής κατάβασης και του αλγορίθμου Gauss-Newton. Η Levenberg-Marquardt χρησιμοποιεί το πλεονέκτημα της ταχύτητας του αλγορίθμου Gauss-Newton και τη σταθερότητα της μεθόδου της βαθμωτής κατάβασης. Η Levenberg-Marquardt είναι πιο ισχυρή από τον αλγόριθμο του Gauss-Newton, επειδή σε πολλές περιπτώσεις μπορεί να συγκλίνει καλά, ακόμη και όταν η επιφάνεια σφάλματος είναι πολύ πιο περίπλοκη από τη τετραγωνική. Παρά το γεγονός ότι ο αλγόριθμος Levenberg-Marquardt τείνει να είναι λίγο πιο αργός από ό,τι ο αλγόριθμος Gauss-Newton (όσον αναφορά τη σύγκλιση), συγκλίνει πολύ πιο γρήγορα από ό,τι η μέθοδος βαθμωτής κατάβασης.

Η βασική ιδέα του αλγορίθμου Levenberg-Marquardt είναι ότι εκτελεί μια συνδυασμένη διαδικασία εκπαίδευσης:

Γύρω από μια περιοχή με σύνθετη καμπυλότητα, ο αλγόριθμος Levenberg-Marquardt χρησιμοποιεί τη μέθοδο βαθμωτής κατάβασης έως ότου βρεθεί τοπική καμπυλότητα κατάλληλη για να γίνει τετραγωνική προσέγγιση. Τότε χρησιμοποιεί τον αλγόριθμο Gauss-Newton, ο οποίος μπορεί να επιταχύνει τη σύγκλιση σημαντικά.

Για να βεβαιωθεί η μέθοδος Levenberg-Marquardt ότι η προσέγγιση του πίνακα Hessian $J^T J$ είναι αντιστρέψιμη, εισάγει μια άλλη προσέγγιση στο πίνακα του Hessian

$$H \approx J^T J + \mu I \quad (5.1)$$

Όπου: μ είναι πάντα θετικός, και ονομάζεται συντελεστής συνδυασμού

Το I συμβολίζει τον ταυτοτικό πίνακα

Από την Εξίσωση (3.1) μπορεί κανείς να παρατηρήσει ότι τα στοιχεία της κύριας διαγωνίου της προσέγγισης του πίνακα Hessian θα είναι μεγαλύτερα του μηδενός. Ως εκ τούτου, με την παρούσα προσέγγιση (3.1), μπορεί να θεωρείται σίγουρο ότι ο πίνακας του Hessian είναι πάντα αντιστρέψιμος.

Ο αλγόριθμος Gauss-Newton μπορεί παρουσιαστεί ως :

$$w_{k+1} = w_k - \left(J_k^T J_k \right)^{-1} J_k e_k \quad (5.2)$$

Με το συνδυασμό των Εξισώσεων 5.2 και 5.1 ο αλγόριθμος της Levenberg-Marquardt μπορεί να παρουσιαστεί ως εξής:

$$w_{k+1} = w_k - \left(J_k^T J_k + \mu I \right)^{-1} J_k^T e_k \quad (5.3)$$

Ο αλγόριθμος αυτός αποτελεί ένα συνδυασμό της μεθόδου της βαθμωτής κατάβασης και του αλγορίθμου Gauss-Newton και μπορεί να εναλλάσσεται μεταξύ των δυο μεθόδων κατά την εκπαιδευτική διαδικασία. Όταν ο συντελεστής συνδυασμού μ είναι πολύ μικρός (σχεδόν μηδενικός) η εξίσωση (5.3) πλησιάζει την Εξίσωση (5.2) και ο αλγόριθμος Gauss-Newton χρησιμοποιείται. Όταν ο συντελεστής συνδυασμού μ έχει πολύ μεγάλη τιμή, η εξίσωση 5.3 προσεγγίζει την εξίσωση $w_{k+1} = w_k - \alpha g_k$ (5.4) και η μέθοδος της βαθμωτής κατάβασης χρησιμοποιείται.

$$\text{Όπου : } g = \frac{\partial E(x, w)}{\partial w} = \left[\frac{\partial E}{\partial w_1} \quad \frac{\partial E}{\partial w_2} \quad \frac{\partial E}{\partial w_N} \right]^T$$

Αν ο συντελεστής μ συνδυασμός της εξίσωσης (5.3) είναι πολύ μεγάλος, μπορεί να ερμηνευθεί ως συντελεστής μάθησης στη μέθοδο βαθμωτής κατάβασης (5.4)

$$\text{ : } \alpha = \frac{1}{\mu}$$

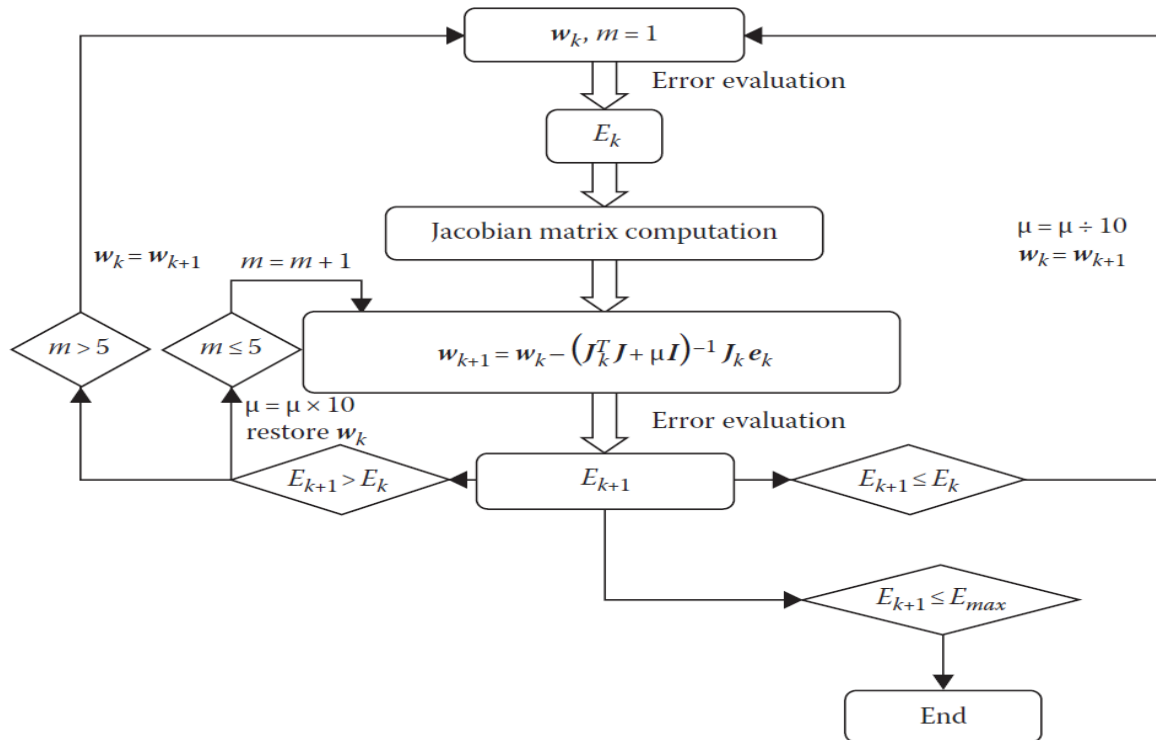
Για την υλοποίηση του αλγορίθμου Levenberg-Marquardt κατά την εκπαιδευτική διαδικασία των νευρωνικών δικτύων, δύο προβλήματα πρέπει να επιλυθούν:

- πώς μπορεί να υπολογιστεί ο Ιακωβιανός πίνακας
- πώς θα οργανωθεί η επαναληπτική διαδικασία κατά τη διάρκεια της εκπαίδευσης για την ενημέρωση των συναπτικών βαρών

$$\text{Σημείωση : } E(x, w) = \frac{1}{2} \sum_{P=1}^P \sum_{M=1}^M e_{P,M}^2$$

Όπου x = είσοδοι , w = συναπτικά βάρη $e_{P,M}$ = Σφάλμα εκπαίδευσης για έξοδο M όταν εφαρμόζεται υπόδειγμα P $e_{P,M} = y - \hat{y}$

ΚΕΦΑΛΑΙΟ 5
ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ



Σχήμα 5.8 Απεικόνιση Διαγράμματος της εκπαίδευσης του αλγορίθμου Levenberg-Marquardt: w_k είναι το τρέχον βάρος, w_{k+1} είναι το επόμενο βάρος, E_{k+1} είναι το τρέχον συνολικό σφάλμα και E_k είναι το προηγούμενο συνολικό σφάλμα.

Η διαδικασία εκπαίδευσης χρησιμοποιώντας τον αλγόριθμο Marquardt-Levenberg θα μπορούσε να σχεδιαστεί ως εξής:

1. Με τα αρχικά βάρη (τυχαία επιλεγμένα), αξιολογείται το συνολικό σφάλμα (SSE).
2. Εν συνεχεία, το νευρωνικό δίκτυο πραγματοποιεί μια ενημέρωση των συναπτικών βαρών w σύμφωνα με την εξίσωση (5.3) που αναφέρεται στο κεφάλαιο ώστε να ρυθμιστούν τα καινούρια βάρη.
3. Αξιολογείται με τα νέα βάρη, το συνολικό σφάλμα.
4. Αν το τρέχον συνολικό σφάλμα αυξηθεί ως αποτέλεσμα της ενημέρωσης, τότε ανακαλείται το βήμα (δηλαδή επαναφέρεται το συναπτικό βάρος στην προηγούμενη του τιμή) και αυξάνεται ο συντελεστής συνδυασμού μ με ένα παράγοντα του 10 ή με κάποιο άλλο παράγοντα που έχει οριστεί. Στη συνέχεια, εκτελείται το βήμα 2 και επιχειρείται νέα ενημέρωση.
5. Εάν το τρέχον συνολικό σφάλμα μειωθεί ως αποτέλεσμα της ενημέρωσης, τότε το βήμα γίνεται αποδεκτό (κρατείται το καινούριο συναπτικό βάρος ως το τρέχον) και μειώνεται ο συντελεστής συνδυασμού μ με ένα παράγοντα του 10 ή τον ίδιο παράγοντα που χρησιμοποιήθηκε στο βήμα 4

6. Η μέθοδος επανέρχεται στο βήμα 2 με το νέο συναπτικό βάρος μέχρι το τρέχον συνολικό σφάλμα να είναι μικρότερο από την απαιτούμενη τιμή
7. Σαν κριτήριο διακοπής της εκπαίδευσης E_{MAX} όπως απεικονίζεται στο Σχήμα 5.8 έχει οριστεί το κριτήριο early stopping του σφάλματος του υποσυνόλου validation(αξιολόγησης). Η επεξήγηση του κριτηρίου αυτού αναφέρεται στο Κεφάλαιο 6.

ΚΕΦΑΛΑΙΟ 6

ΑΝΑΠΤΥΞΗ ΜΟΝΤΕΛΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΓΙΑ ΤΗ ΠΡΟΒΛΕΨΗ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΙΔΙΟΤΗΤΩΝ B_o , R_s , p_b

I. ΟΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗΝ ΑΝΑΠΤΥΞΗ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Οι βάσεις δεδομένων που επιλέχθηκαν για την ανάπτυξη των νευρωνικών αποτελούν υπό-βάσεις της συνολικής βάσης PVT δεδομένων που έχει στη διάθεσή του το Εργαστήριο Ανάλυσης Ρευστών και Πυρήνων του Πολυτεχνείου Κρήτης. Η συνολική βάση δεδομένων αριθμεί 1020 αναφορές διαφορετικών πετρελαϊκών ρευστών και ένα μέρος των πληροφοριών που διαθέτει για κάθε ρευστό προέρχεται από στιγμιαία εκτόνωση σε ατμοσφαιρική πίεση του κάθε ρευστού προερχόμενου από υπόγεια δειγματοληψία ταμιευτήρα (bottom hole) με στόχο τον υπολογισμό της σύστασης του αρχικού ρευστού. Επίσης, εμπεριέχονται στη βάση δεδομένων και οι πληροφορίες που προήλθαν από τις PVT μελέτες του κάθε ρευστού όπως περιεγράφηκαν αναλυτικά στο κεφάλαιο 3. Επιπροσθέτως, θα πρέπει να τονιστεί ότι οι αναφορές των πετρελαϊκών ρευστών παρουσιάζουν μια εκτεταμένη γεωγραφική διασπορά.

Λόγω της εκτεταμένης γεωγραφικής διασποράς των αναφορών των πετρελαϊκών ρευστών, η συγκεκριμένη βάση δεδομένων διαθέτει συγκριτικά πλεονεκτήματα σε σχέση με άλλες βάσεις δεδομένων με πετρελαϊκά ρευστά από συγκεκριμένες και μόνο περιοχές. Επομένως, η χρησιμοποίηση μιας βάσης δεδομένων με μεγάλο πλούτο πετρελαϊκών ρευστών από τις σημαντικότερες πετρελαιοπαραγωγικές περιοχές δίνει τη δυνατότητα στα μοντέλα προσομοίωσης που αναπτύχθηκαν στη παρούσα διπλωματική να έχουν μια καθολική εφαρμογή με ακριβέστερες εν γένει εκτιμήσεις σε σχέση με τις εμπειρικές συσχετίσεις ή τα υπόλοιπα διαθέσιμα νευρωνικά της βιομηχανίας πετρελαίου που έχουν εκπαιδευτεί από βάσεις δεδομένων αναφορών πετρελαϊκών ρευστών από συγκεκριμένες και μόνο γεωγραφικές περιοχές.

Υπενθυμίζεται ότι η αναλυτική σύσταση των πετρελαϊκών ρευστών δεν περιέχεται στις πληροφορίες της βάσης δεδομένων που διατίθενται για την εκπαίδευση των νευρωνικών αλλά δίνεται εμμέσως από τις πληροφορίες των τιμών δεδομένων παραγωγής. Η χρησιμοποίηση μιας πλούσιας σε αναφορές πετρελαϊκών ρευστών βάσης δεδομένων έχει σαν αποτέλεσμα την δημιουργία μιας καλής γενίκευσης της υποκρυπτόμενης σχέσης των δεδομένων παραγωγής και των ζητούμενων ιδιοτήτων από τα αναπτυχθέντα νευρωνικά και επομένως την εξασφάλιση ακρίβειας προβλέψεων ιδιοτήτων άγνωστων ρευστών.

Εν συνεχεία, αναδεικνύεται με γραφικές απεικονίσεις η γεωγραφική διασπορά των αναφορών των πετρελαϊκών ρευστών της υπό-βάσης που χρησιμοποιήθηκε για την εκπαίδευση των νευρωνικών που εκτιμούν την κάθε ιδιότητα ανά ήπειρο αλλά και παρουσίαση με τη χρήση πίνακα της αναλυτικής κατανομής των διαθέσιμων πετρελαϊκών ρευστών ανά χώρα προέλευσης.

Εικόνα 6.1 Απεικόνιση της ποσοστιαίας γεωγραφικής κατανομής ανά ήπειρο.



Πίνακας 6.1 Αναλυτική παρουσίαση της γεωγραφικής κατανομής των αναφορών των πετρελαϊκών ρευστών που εμπεριέχονται στη υπό-βάση δεδομένων.

ΣΥΝΟΛΟ ΧΩΡΩΝ	ΣΥΝΟΛΟ ΑΡΙΘΜΩΝ ΔΕΙΓΜΑΤΩΝ	ΝΟΤΙΑ ΑΦΡΙΚΗ		ΚΕΝΤΡΙΚΗ ΑΜΕΡΙΚΗ	
36	385	ΧΩΡΕΣ	ΑΡΙΘΜΟΣ ΔΕΙΓΜΑΤΩΝ	ΧΩΡΕΣ	ΑΡΙΘΜΟΣ ΔΕΙΓΜΑΤΩΝ
ΕΥΡΩΠΗ		SOUTH AFRICA	3	MEXICO	1
ΧΩΡΕΣ	ΑΡΙΘΜΟΣ ΔΕΙΓΜΑΤΩΝ	ANGOLA	4	ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ	1
NETHERLANDS	2	ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ	7	ΚΕΝΤΡΙΚΗ ΑΣΙΑ	
FRANCE	3	ΝΟΤΙΑ ΑΜΕΡΙΚΗ		ΧΩΡΕΣ	ΑΡΙΘΜΟΣ ΔΕΙΓΜΑΤΩΝ
ITALY	3	ΧΩΡΕΣ	ΑΡΙΘΜΟΣ ΔΕΙΓΜΑΤΩΝ	RUSSIA	7
NORWAY	1	ECUADOR	48	PAKISTAN	1
SPAIN	2	VENEZUELA	41	KAZAKHSTAN	1
DENMARK	1	COLOMBIA	3	ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ	9
UNITED KINGDOM	6	TRINIDAD	1	ΒΟΡΕΙΑ ΑΦΡΙΚΗ	
ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ	18	PERU	1	ΧΩΡΕΣ	ΑΡΙΘΜΟΣ ΔΕΙΓΜΑΤΩΝ
ΜΕΣΗ ΑΝΑΤΟΛΗ		ARGENTINA	1	LIBYA	32
ΧΩΡΕΣ	ΑΡΙΘΜΟΣ ΔΕΙΓΜΑΤΩΝ	BRAZIL	1	NIGERIA	78
REPUBLIC OF YEMEN	9	ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ	96	ALGERIA	9
U.A.E	56	ΚΕΝΤΡΙΚΗ ΑΦΡΙΚΗ		TUNISIA	1
QATAR	1	ΧΩΡΕΣ	ΑΡΙΘΜΟΣ ΔΕΙΓΜΑΤΩΝ	ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ	120
SULTANATE OF OMAN	12	GABON	3	ΑΓΝΩΣΤΑ	
KUWAIT	20	CONGO	4	ΧΩΡΕΣ	ΑΡΙΘΜΟΣ ΔΕΙΓΜΑΤΩΝ
SYRIA	1	EQUATORIAL GUINEA	2	UNKNOWN	19
SAUDI ARABIA	2	CAMEROON	1	ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ	19
IRAN	4	ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ	10		
ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ	105				

Προβολή της προέλευσης των δειγματοληψιών στο χάρτη.



II. ΣΤΑΤΙΣΤΙΚΑ ΚΡΙΤΗΡΙΑ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΗΣ ΑΚΡΙΒΕΙΑΣ ΤΩΝ ΠΡΟΒΛΕΨΕΩΝ ΤΩΝ ΜΟΝΤΕΛΩΝ ΕΚΤΙΜΗΣΗΣ ΙΔΙΟΤΗΤΩΝ

Εισαγωγή

Η στατιστική χρησιμοποιείται ως ένα μαθηματικό εργαλείο για την ποσοτική ανάλυση των δεδομένων. Κάθε προσπάθεια πρόβλεψης μιας φυσικής ιδιότητας όπως π.χ ο ογκομετρικός συντελεστής υγρής φάσης ή ο λόγος του όγκου εν διαλύσει αερίου / όγκου πετρελαίου χαρακτηρίζεται από μία αβεβαιότητα και αναγράφεται με τη μορφή Πραγματική Τιμή \pm αβεβαιότητα = Εκτίμηση. Το σφάλμα στη προκειμένη περίπτωση ορίζεται ως η διαφορά της εκτιμώμενης τιμής από την τιμή της πραγματικής εργαστηριακής μέτρησης.

Η ανάλυση των στατιστικών σφαλμάτων χρησιμοποιείται για να εξαχθούν χρήσιμες πληροφορίες από τα δεδομένα των προβλέψεων που θα οδηγήσουν σε συμπεράσματα για την απόδοση και την ακρίβεια των αναπτυχθέντων νευρωνικών.

Για την παραπάνω ανάλυση, μπορούν να επιλεγούν διάφορα στατιστικά κριτήρια. Εν συνεχεία, παρατίθενται τα κριτήρια που υιοθετήθηκαν στη παρούσα διπλωματική για την αξιολόγηση της ακρίβειας και είναι : το σφάλμα, το απόλυτο σφάλμα, η τυπική απόκλιση, ο συντελεστής συσχέτισης (Pearson), η ελάχιστη-μέγιστη και μέση τιμή του σχετικού ποσοστιαίου σφάλματος, η μέση τιμή του απολύτου σχετικού ποσοστιαίου σφάλματος

Σφάλμα (Error)

Σαν σφάλμα e_i ορίζεται η διαφορά της υπολογισμένης τιμής από την πραγματική τιμή, όπου $y_{i \text{ real}}$ είναι η τιμή που μετρήθηκε πειραματικά και $y_{i \text{ pred}}$ η τιμή που προήλθε από τα μοντέλα παλινδρόμησης (regression models). Στη παρούσα εργασία χρησιμοποιήθηκε το μέσο σφάλμα στο οποίο θα δοθεί το σύμβολο **ME (mean error)**.

$$ME = \frac{1}{n} \sum_{i=1}^n e_i$$

Όπου $e_i = y_{i \text{ pred}} - y_{i \text{ real}}$

Απόλυτο σφάλμα (Absolute error)

Το απόλυτο σφάλμα $|e_i|$ ορίζεται ως η απόλυτη τιμή της διαφοράς της υπολογισμένης τιμής από την πραγματική τιμή : $|e_i| = |y_{i \text{ pred}} - y_{i \text{ real}}|$

Όπου y_i real η τιμή που μετρήθηκε πειραματικά και y_i pred η τιμή που προήλθε από τα μοντέλα παλινδρόμησης (regression models). Θα χρησιμοποιηθεί το μέσο απόλυτο σφάλμα στο οποίο θα δοθεί το σύμβολο **MAE (mean absolute error)**.

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i|$$

Σχετικό ποσοστιαίο σφάλμα (relative percentage error)

Το σχετικό ποσοστιαίο σφάλμα ορίζεται ως το πηλίκο της διαίρεσης του σφάλματος προς την πραγματική τιμή πολλαπλασιασμένο επί 100. $ri = \frac{e_i}{y_{i_{real}}}$. Στη παρούσα διπλωματική, χρησιμοποιήθηκε το μέσο σχετικό ποσοστιαίο σφάλμα και θα δοθεί ο συμβολισμός **Er (mean relative percentage error)** :

$$Er = \frac{1}{n} \sum_{i=1}^n \frac{e_i}{y_{i_{real}}} * 100 .$$

Η εύρεση αυτού του σφάλματος δίνει την δυνατότητα του υπολογισμού της μέσης κατανομής του σφάλματος σε σχέση με το μηδέν που είναι η ιδανική ζητούμενη τιμή του . Αποτελεί ένα ποσοτικό κριτήριο αξιολόγησης της συμπεριφορά της κάθε εμπειρικής συσχέτισης που μελετάται αλλά και των νευρωνικών που παρήχθησαν ώστε να γίνει κατανοητή η κατανομή του σφάλματος γύρω από την πραγματική τιμή και να εκτιμηθεί το εύρος της υπερεκτίμησης ή υποεκτίμησης .

Απόλυτο σχετικό ποσοστιαίο σφάλμα (absolute relative percentage error)

Το απόλυτο σχετικό ποσοστιαίο σφάλμα ορίζεται ως το πηλίκο της διαίρεσης του απόλυτου σφάλματος προς την πραγματική τιμή πολλαπλασιασμένο επί 100. $|ri| = \left| \frac{e_i}{y_{i_{real}}} \right|$. Στην παρούσα εργασία χρησιμοποιείται το μέσο απόλυτο σχετικό ποσοστιαίο σφάλμα και συμβολίζεται ως **Ea (mean absolute relative percentage error)**:

$$Ea = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{y_{i_{real}}} \right| * 100$$

Το Ea είναι ένα σχετικό μέτρο που εκφράζει το σφάλμα ως ποσοστό της πραγματικής τιμής. Αυτό είναι και το μεγαλύτερο πλεονέκτημά του καθώς παρέχει ένα εύκολο και πρακτικό

τρόπο ώστε να κριθεί η έκταση και η σημασία του σφάλματος. Με αυτό το κριτήριο έγινε και η αξιολόγηση των εμπειρικών συσχετίσεων και εκτιμήθηκε η ακρίβεια πρόβλεψης. Χρησιμοποιήθηκε επίσης και ως κριτήριο για την επιλογή του βέλτιστου νευρωνικού εξ όσων αναπτύχθηκαν.

Τυπική Απόκλιση (Standard deviation)

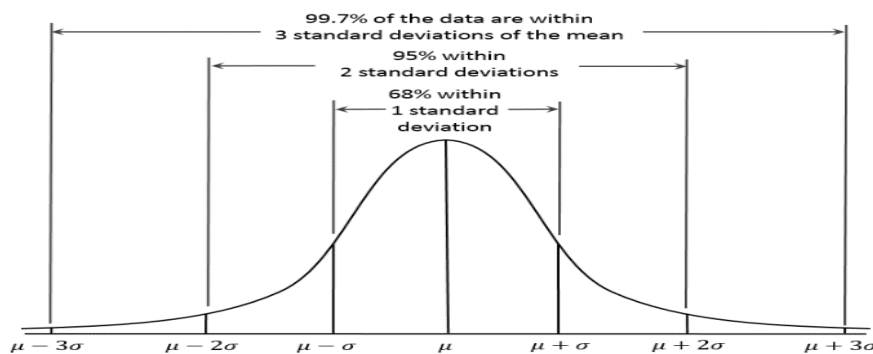
Η διασπορά (variance) ορίζεται ως η μέση τιμή του τετραγώνου των διακυμάνσεων, δηλαδή: $\sigma_X^2 = E[(Xi - \bar{X})^2]$ Η δειγματική διασπορά $\hat{\sigma}_X^2$ δίνεται από την σχέση: $\hat{\sigma}_X^2 =$

$\frac{1}{n} \sum_{i=1}^n (Xi - \bar{X})^2$ Η τυπική απόκλιση (standard deviation) σ_X είναι η τετραγωνική ρίζα της διασποράς

και προσδιορίζει ποσοτικά την απόκλιση την κατανομής από την μέση τιμή :

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^n (Xi - \bar{X})^2}{n - 1}}$$

Το εύρος της συνάρτησης πυκνότητας πιθανότητας, τουλάχιστον για συμμετρικές κατανομές είναι ανάλογο προς την τυπική απόκλιση. Η τυπική απόκλιση είναι στατιστικό μέγεθος που χρησιμοποιείται για να εκφραστεί το πιθανό σφάλμα πειραματικών μετρήσεων και το σφάλμα εκτίμησης των παραμέτρων μιας κατανομής πιθανότητας από ένα πεπερασμένο δείγμα.



Σχήμα 6.2 Κωδωνοειδής καμπύλη Πιθανοτήτων

Στη παρούσα διπλωματική χρησιμοποιείται η τυπική απόκλιση για να δείχθεί το μέτρο της διασποράς των σχετικών ποσοστιαίων σφαλμάτων των εμπειρικών συσχετίσεων και των νευρωνικών δικτύων. Μια χαμηλή τιμή τυπικής απόκλισης δείχνει ότι τα σχετικά ποσοστιαία σφάλματα τείνουν να είναι πολύ κοντά στο μέσο όρο ενώ μια υψηλή τιμή δείχνει ότι απλώνονται σε ένα

μεγάλος εύρος τιμών. Επίσης, η τυπική απόκλιση χρησιμοποιείται μαζί με το κανόνα 68.2 %-95.4%-99.7% της Κωδωνοειδούς Καμπύλης Πιθανοτήτων για να υπολογιστεί η μέγιστη και ελάχιστη τιμή του σχετικού ποσοστιαίου σφάλματος ανάλογα με την πιθανότητα που εξετάζεται κάθε φορά.

Ο κανόνας 68.2%-95.4%-99.7% αναφέρεται στο ότι η κατανομή αυτή έχει συμμετρικότητα με σημείο αναφοράς το σημείο του αριθμητικού μέσου όρου. Στην συνέχεια, αναφέρεται το ποσοστό δεδομένων που υπάρχουν συγκεντρωτικά, δεξιά και αριστερά από τον μέσο όρο, πάντα μέσα στα πλαίσια της κωδωνοειδούς καμπύλης πιθανοτήτων:

- i) Σε απόσταση μιας (1) τυπικής απόκλισης αγγίζουν το 68.2%: $\mu \pm 1\sigma = 68.2\%$
- ii) Σε απόσταση δύο (2) τυπικών αποκλίσεων αγγίζουν το 95.4%: $\mu \pm 2\sigma = 95.4\%$
- iii) Σε απόσταση τριών (3) τυπικών αποκλίσεων αγγίζουν το 99.7%: $\mu \pm 3\sigma = 99.7\%$

Η πιθανότητα μιας τιμής “x” να απέχει 1 ή 2 ή 3 τυπικές αποκλίσεις, τόσο δεξιά όσο και αριστερά από τον μέσο όρο μέσα στη τυπική κατανομή, είναι:

- $\Pr(\mu - 1\sigma \leq x \leq \mu + 1\sigma) = 0.6827$
- $\Pr(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 0.9545$
- $\Pr(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = 0.9973$

Για τις αξιολογήσεις των νευρωνικών που αναπτύχθηκαν θα χρησιμοποιηθεί η τιμή της τυπικής απόκλισης (STDEV) του μέσου σχετικού ποσοστιαίου σφάλματος. επίσης, θα χρησιμοποιηθεί η τυπική απόκλιση σε συνδυασμό με το κανόνα 68.2 %-95.4%-99.7% της Κωδωνοειδούς Καμπύλης Πιθανοτήτων για τον υπολογισμό της ελάχιστης και μέγιστης τιμής του σχετικού ποσοστιαίου σφάλματος που μπορεί να προκύψει από μια εκτίμηση των τιμών των ιδιοτήτων ενός άγνωστου πετρελαϊκού ρευστού, ανάλογα με το ποσοστό βεβαιότητας της εκτίμησης που επιθυμείται (68%,95%,99%).

Συντελεστής συσχέτισης (Pearson) r

Ο συντελεστής συσχέτισης Pearson είναι ένα αδιάστατο μέγεθος το οποίο έχει εύρος τιμών από -1 έως 1 και απεικονίζει το είδος της σχέσης μεταξύ δύο συνόλων δεδομένων και ορίζεται με τη παρακάτω εξίσωση :

$$r = \frac{\sum (y_{real} - \bar{y}_{real})(y_{pred} - \bar{y}_{pred})}{\sqrt{\sum (y_{real} - \bar{y}_{real})^2 \sum (y_{pred} - \bar{y}_{pred})^2}}$$

Εάν $r = 1$, το διάγραμμα διασποράς αντιστοιχεί σε ευθεία γραμμή με θετική κλίση 45° . Εάν $r = -1$, το διάγραμμα διασποράς παριστάνει μια ευθεία γραμμή με αρνητική κλίση -45° . Εάν $|r| < 1$, το διάγραμμα διασποράς μοιάζει με ένα νέφος σημείων, η διασπορά του οποίου αυξάνει όσο η απόλυτη τιμή του συντελεστή συσχέτισης τείνει προς το μηδέν. Ο συντελεστής συσχέτισης χρησιμοποιείται κατά κόρον σε στατιστικές αναλύσεις για την εκτίμηση της εξάρτησης δυο μεταβλητών.

Στη παρούσα εργασία ο συντελεστής συσχέτισης r αποτελεί ένα στατιστικό στοιχείο που δίνει πληροφορίες για την ποιότητα προσαρμογής των μοντέλων παλινδρόμησης (regression models). Με άλλα λόγια αντιπροσωπεύει το βαθμό επιτυχίας στη μείωση της τυπικής απόκλισης. Δίνοντας ένα παράδειγμα, μια τιμή $r = 1$ δείχνει ότι οι προβλέψεις των μοντέλων παλινδρόμησης, των νευρωνικών ή των εμπειρικών εξισώσεων, ταιριάζουν απόλυτα με τα πειραματικά δεδομένα.

III. ΑΝΑΠΤΥΞΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΕ ΠΕΡΙΒΑΛΛΟΝ MATLAB

Εισαγωγή

Η Matlab είναι μια υψηλής απόδοσης γλώσσα για τεχνικό προγραμματισμό. Μπορεί να επιλύσει δύσκολα μαθηματικά προβλήματα, να μοντελοποιήσει συναρτήσεις και σήματα και να προγραμματίζει μέσω ενός εύχρηστου περιβάλλοντος, όπου προβλήματα και λύσεις μπορούν και εκφράζονται μέσω των συνηθών μαθηματικών συμβολισμών. Συνήθεις χρήσεις της Matlab είναι:

- ο υπολογισμός μαθηματικών παραστάσεων
- η ανάπτυξη αλγορίθμων
- η δημιουργία μαθηματικών μοντέλων και η προσομοίωση αυτών
- η ανάλυση δεδομένων και η μοντελοποίηση αυτών
- γραφικές αναπαραστάσεις σύνθετων δεδομένων όπου χρησιμοποιούνται σε επιστημονικές και μηχανολογικές εφαρμογές
- ανάπτυξη εφαρμογών συμπεριλαμβανομένης της δημιουργίας γραφικού περιβάλλοντος χρήστη.

Η Matlab είναι ένα σύστημα αλληλεπιδραστικό βασικό στοιχείο του οποίου είναι ένας **πίνακας** (array) όπου αποθηκεύονται οι πληροφορίες χωρίς αρχικές διαστάσεις. Αυτό της επιτρέπει να μπορεί να διαχειρίζεται άμεσα και εύκολα (σε σχέση με άλλες γλώσσες προγραμματισμού όπως η C ή η Fortran) πολλούς τεχνικούς υπολογισμούς και ειδικότερα εκείνους που έχουν σαν ορίσματα πίνακες (arrays) ή διανύσματα (vectors).

Το όνομα της Matlab προκύπτει από το **Matrix Laboratory**. Σαν γλώσσα, εξελίχθηκε για να φτάσει στην σημερινή της μορφή, με την βοήθεια πολλών χρηστών της. Στα πανεπιστήμια αποτελεί το βασικό εκπαιδευτικό εργαλείο για την εισαγωγή και την εμβάθυνση των μαθηματικών, της μηχανικής και της φυσικής. Στην βιομηχανία δε, είναι ένα βασικό εργαλείο για την έρευνα και την βελτιστοποίηση της παραγωγής, την ανάπτυξη εφαρμογών και την ανάλυση οικονομικών μεγεθών.

Χαρακτηριστικό της Matlab είναι μία συλλογή από ειδικά εργαλεία για συγκεκριμένες εφαρμογές, που ονομάζονται **toolboxes**. Τα toolboxes βοηθούν τους χρήστες της Matlab να μάθουν και να εφαρμόσουν εξειδικευμένες τεχνολογίες, όπως τα Νευρωνικά Δίκτυα, η Fuzzy Logic, κ.α.

Τα toolboxes αποτελούν με άλλα λόγια, περιεκτικές συλλογές από έτοιμες συναρτήσεις της Matlab (M-files) που διευρύνουν την ικανότητά της στο να δίνει ακόμα πιο εύκολα και πιο γρήγορα λύσεις σε πολλά εξειδικευμένα προβλήματα. Επιστημονικές περιοχές που καλύπτουν τα toolboxes της Matlab είναι η επεξεργασία σήματος, τα συστήματα ελέγχου, τα νευρωνικά δίκτυα, η fuzzy logic, η προσομοίωση και πολλά άλλα.

Ανάπτυξη συμπληρωματικού κώδικα

Για τις ανάγκες της παρούσας διπλωματικής χρειάστηκε να γραφτεί συμπληρωματικός κώδικας και να χρησιμοποιηθεί σε συνδυασμό με την εργαλειοθήκη (toolbox) των νευρωνικών της Matlab αλλάζοντας ή προσθέτοντας όποια χαρακτηριστικά κριθούν απαραίτητα για την ανάπτυξη του νευρωνικού. Ένα από τα μεγαλύτερα πλεονεκτήματα που προσέθεσε η ανάπτυξη συμπληρωματικού κώδικα είναι η δυνατότητα εκτέλεσης επαναληπτικής διαδικασίας χωρίς την παρέμβαση του εκπαιδευτή. Έτσι το πρόγραμμα ήταν σε θέση να υλοποιήσει όσες αλληπάλληλες αναπτύξεις είχαν οριστεί στο κώδικα αρχικά από τον εκπαιδευτή και να αποθηκεύσει τα αποτελέσματα που παρήγαγε κάθε νευρωνικό όπως και τα νευρωνικά αυτά καθαυτά σε μια δομή (structure) . Το πρόγραμμα ξεκινούσε την εκπαίδευση των νευρωνικών για μια συγκεκριμένη αρχική δομή και υλοποιούσε αλληπάλληλες αναπτύξεις για αυτή τη δομή νευρωνικών μέχρι τον αριθμό που είχε οριστεί.

Όταν τελείωνε αυτές τις αλληπάλληλες αναπτύξεις για αυτή τη δομή νευρωνικών αύξανε προοδευτικά κατά έναν, τον αριθμό των κρυφών νευρώνων και επαναλάμβανε την διαδικασία των αλληπάλληλων αναπτύξεων για την καινούρια αυτή δομή νευρωνικών. Ο αριθμός του μέγιστου αριθμού κρυφών νευρώνων άρα και τις δομές νευρωνικών που δύναται να εκτελέσει το πρόγραμμα αλληπάλληλες αναπτύξεις είχε οριστεί στο κώδικα μέσω εντολής. Σαν αρχική δομή είχαν οριστεί οι 5 νευρώνες και σαν τελική οι 200 νευρώνες. Για κάθε δομή πραγματοποιούσε 100 αναπτύγματα.

Ένα άλλος λόγος που επιλέχθηκε να γραφεί συμπληρωματικός κώδικας είναι για να υπολογιστούν τα στατιστικά κριτήρια αξιολόγησης που αναφέρθηκαν πιο πάνω έτσι ώστε να αξιολογηθούν οι εκτιμήσεις επί των πραγματικών τιμών όλων των νευρωνικών που παρήχθησαν και εν συνεχεία βάσει αυτών να επιλεγεί το βέλτιστο από άποψη παραγόμενου σφάλματος και γενίκευσης και να συγκριθεί η ακρίβεια του με τις εμπειρικές συσχετίσεις. Θα πρέπει να τονιστεί επίσης

ότι η δόθηκε η δυνατότητα του υπολογισμού των στατιστικών κριτηρίων ανά κατηγορία υποσυνόλου δεδομένων (training, validation, test) υπό-βάσης δεδομένων που χρησιμοποιήθηκε στην εκπαίδευση για κάθε νευρωνικό που παράχθηκε.

Σημείωση : Λόγω του ότι τα προβλήματα παλινδρόμησης που κλήθηκαν να λύσουν τα νευρωνικά δίκτυα είναι χαμηλής διαστατικότητας (4 είσοδοι και μία έξοδος) χρησιμοποιήθηκε ένα κρυφό επίπεδο όπου κρατήθηκε σταθερό για κάθε διαφορετική δομή που πραγματοποιήθηκαν αλληλλάληλα αναπτύγματα.

Διαχωρισμός της βάσης δεδομένων σε υποσύνολα (training, validation, test)

Η υπο-βάση δεδομένων η οποία χρησιμοποιείται για την εκπαίδευση του κάθε νευρωνικού δεν χρησιμοποιείται ως ενιαίο σύνολο δεδομένων αλλά διαχωρίζεται σε τρία υποσύνολα δεδομένων. Το ποσοστό κάθε υποσυνόλου δεδομένων ορίζεται από τον εκπαιδευτή κατά την συγγραφή του κώδικα. Στην παρούσα εργασία τα ποσοστά των υποσυνόλων δεδομένων ορίστηκαν ως εξής:

- Για την εκπαίδευση (training) χρησιμοποιήθηκε το 60% το συνολικών δεδομένων της υπό-βάσης
- Για την επικύρωση (validation) χρησιμοποιήθηκε το 20% το συνολικών δεδομένων της υπό-βάσης
- Για τον έλεγχο αγνώστων στο νευρωνικό αναφορών πετρελαιοικών ρευστών (test) χρησιμοποιήθηκε το 20% το συνολικών δεδομένων της υπό-βάσης

Δεν υπάρχει κάποια συγκεκριμένη ποσόστωση μεταξύ των 3 υποσυνόλων δεδομένων η οποία να θεωρείται ως η ιδανική αλλά επιβάλλεται το ποσοστό του υποσυνόλου δεδομένων για την εκπαίδευση να είναι μεγαλύτερο από το άθροισμα της ποσόστωσης των άλλων 2 υποσυνόλων δεδομένων.

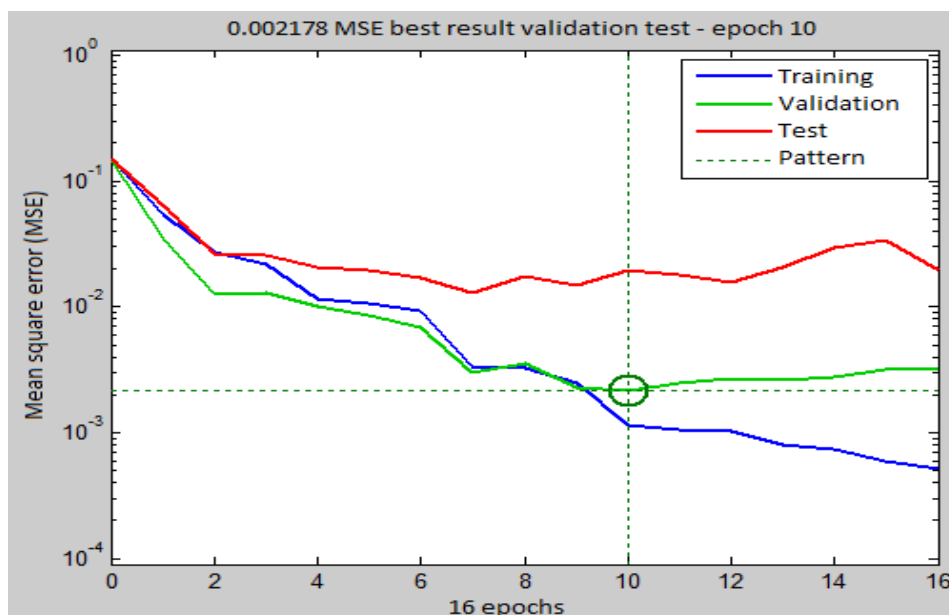
Τα **δεδομένα του υποσυνόλου εκπαίδευσης** χρησιμοποιούνται από το νευρωνικό με σκοπό την ελαχιστοποίηση τη συνάρτησης κόστους μέσω της βαθμωτής κατάβασης (*Levenberg-Marquardt*) έτσι ώστε να τροποποιηθούν τα συναπτικά βάρη w προσεγγίζοντας όσο τον δυνατόν καλύτερα την υποκείμενη συνάρτηση που περιγράφεται από τη σχέση εισόδου-εξόδου.

Τα **δεδομένα του υποσυνόλου επικύρωσης** χρησιμοποιούνται για τον έλεγχο της εκπαίδευσης του νευρωνικού με στόχο την αποφυγή της υπερ-προσαρμογής (overfitting). Τα δε-

δομένα του υποσυνόλου επικύρωσης δεν έχουν χρησιμοποιηθεί από το νευρωνικό για την τροποποίηση των συναπτικών βαρών οπότε ένα μικρό σφάλμα στα δεδομένα του υποσυνόλου επικύρωσης δείχνει ότι προέκυψε μια καλή γενίκευση από την διαδικασία της βαθμωτής κατάβασης και ότι το μοντέλο δεν έχει υπερ-εκπαιδευτεί.

Υπάρχουν διάφορες εκδοχές για την χρησιμοποίηση των δεδομένων του υποσυνόλου επικύρωσης (validation) ως εργαλείο για την αποφυγή της υπερ-εκπαίδευσης. Στην παρούσα εργασία επιλέγεται η επικρατούσα εκδοχή η οποία και ονομάζεται *early stopping* (πρόωρο σταμάτημα). Το νευρωνικό χρησιμοποιώντας αυτή τη τεχνική στην ουσία παρακολουθεί συγχρόνως τα σφάλματα τα οποία παράγονται από τα δεδομένα των υποσυνόλων εκπαίδευσης και επικύρωσης.

Αρχικά το σφάλμα που παράγεται από τα δεδομένα εκπαίδευσης μικραίνει σε κάθε επανάληψη της βαθμωτής κατάβασης (*Levenberg-Marquardt*) και παράλληλα, τουλάχιστον για τις πρώτες επαναλήψεις, μικραίνει και το σφάλμα που παράγεται από τα δεδομένα του υποσυνόλου επικύρωσης. Η εκπαίδευση σταματάει την στιγμή που το σφάλμα που παράγεται από τα δεδομένα του υποσυνόλου επικύρωσης αρχίζει να αυξάνει ενώ ταυτόχρονα το σφάλμα που παράγεται από τα δεδομένα του υποσυνόλου εκπαίδευσης συνεχίζει να μειώνεται. Τα συναπτικά βάρη τα οποία έχουν επιλεγεί εκείνη τη στιγμή, αποτελούν και τα τελικά βάρη του νευρωνικού.



Στο Σχήμα 6.1 γίνεται μια γραφική απεικόνιση της τεχνικής *early stopping* και με το κύκλο προσδιορίζεται η επανάληψη κατά την οποία γίνεται η διακοπή της εκπαίδευσης για την αποφυγή του *overfitting*. Διαπιστώνεται ότι μετά από αυτή την επανάληψη το σφάλμα των δεδομένων επικύρωσης ανεβαίνει.

Επειδή τα δεδομένα του υποσυνόλου επικύρωσης στην πραγματικότητα εμμέσως επηρεάζουν στην διαμόρφωση των συνναπτικών βαρών w δεν μπορεί να θεωρηθεί ότι σαν αναφορές είναι εντελώς άγνωστες για το νευρωνικό οπότε δεν χρησιμοποιούνται για την αξιολόγηση της ακρίβειας των εκτιμήσεων για άγνωστα δείγματα.

Για αυτή τη περίπτωση χρησιμοποιούνται **τα δεδομένα του υποσυνόλου test** με τα οποία δεν έχει έλθει σε καμία επαφή το νευρωνικό κατά την διαδικασία της εκπαίδευσης. Εφόσον το νευρωνικό έχει σταματήσει την διαδικασία της εκπαίδευσης και έχει επιλέξει τα βέλτιστα συνναπτικά βάρη w , εν συνεχεία υπολογίζει το σφάλμα που παράγεται από τα δεδομένα του υποσυνόλου test τα οποία τα εκλαμβάνει ως άγνωστες αναφορές πετρελαϊκών ρευστών. Το σφάλμα το οποίο υπολογίζεται για τα δεδομένα αυτού του υποσυνόλου θεωρείται πως αποτελεί το μέτρο των σφαλμάτων του νευρωνικού για την πρόβλεψη των ιδιοτήτων αγνώστων πετρελαίων

Αντιμετώπιση του προβλήματος της υπερ-προσαρμογής(overfitting)

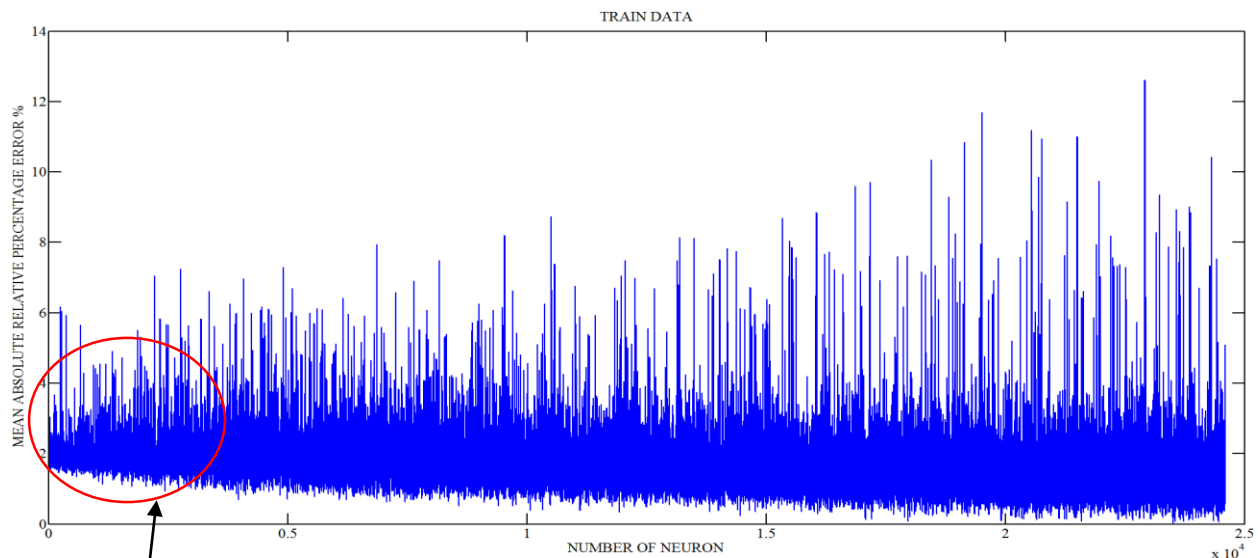
Παρόλο που η τεχνική της πρόωρης διακοπής αφορά την αντιμετώπιση της υπερ-προσαρμογής (overfitting), ένας άλλος παράγοντας ο οποίος επηρεάζει αυτό το πρόβλημα είναι ο αριθμός των κρυφών νευρώνων που επιλέγεται για τη δομή του νευρωνικού. Μάλιστα μια επιλογή αυξημένου αριθμού κρυφών νευρώνων πέραν του δέοντος μπορεί να επηρεάσει πολύ περισσότερο το φαινόμενο της υπερ-προσαρμογής από ότι θα το επηρέαζε μια εσφαλμένη πρόωρη διακοπή της εκπαίδευσης. Ο λόγος για τον οποίο εμφανίζεται αυτό το πρόβλημα είναι διότι αυξάνεται ο αριθμός των συνναπτικών βαρών w μέσω της αύξησης του αριθμού των νευρώνων. Όσο μεγαλύτερος αριθμός κρυφών νευρώνων χρησιμοποιείται, τόσο μεγαλύτερη ευαισθησία διαθέτει το νευρωνικό άρα και μεγαλύτερη προσαρμογή στα δεδομένα εκπαίδευσης αυτά καθ'εαυτά. Για το λόγο αυτό υπάρχει ένας βέλτιστος αριθμός κρυφών νευρώνων για τον οποίο το νευρωνικό δίκτυο γενικεύει καλά και δεν υπερ-προσαρμόζεται.

Υπάρχουν διάφορες προσεγγίσεις στην διεθνή βιβλιογραφία για την εύρεση του βέλτιστου αριθμού κρυφών νευρώνων. Στην παρούσα εργασία χρησιμοποιείται μια εμπειρική προσέγγιση που εκμεταλλεύεται το γεγονός ότι με τη χρήση του συμπληρωματικού κώδικα έγινε δυνατόν να αναπτυχθεί πληθώρα νευρωνικών δικτύων με διαφορετικές δομές, όπως έχει ήδη αναφερθεί, για κάθε ιδιότητα. Κάνοντας χρήση του στατιστικού κριτηρίου του μέσου απολύτου σχετικού ποσοστιαίου σφάλματος στις εκτιμήσεις επί των δεδομένων του υποσυνόλου test για όλα τα νευρωνικά που παρήχθησαν, μπορεί να βρεθούν κατά προσέγγιση οι δομές (ο αριθμός των κρυφών

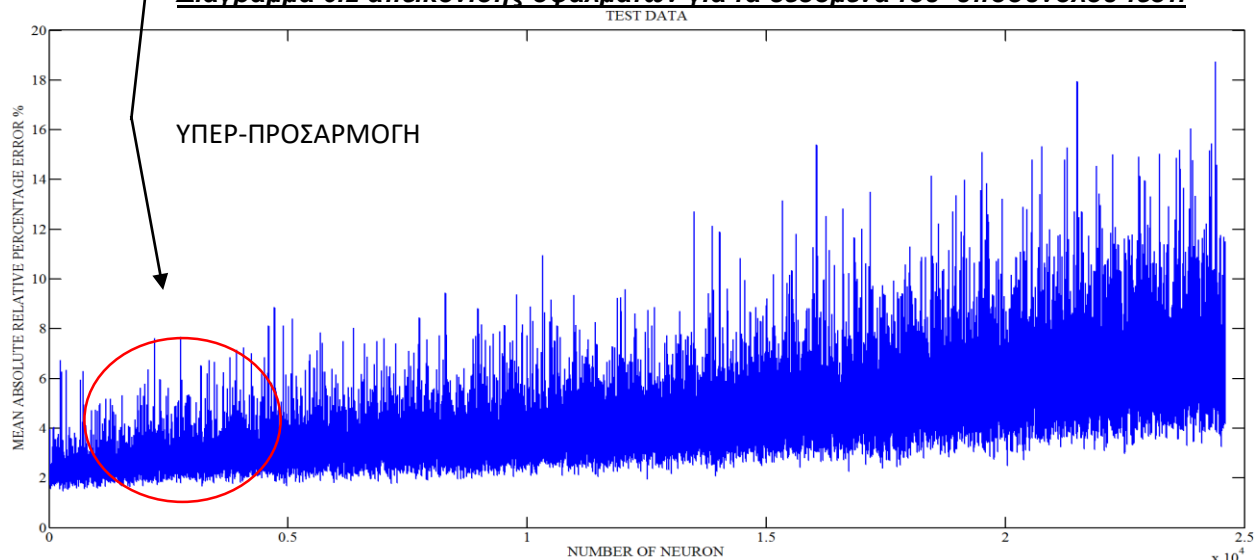
νευρώνων) των νευρωνικών που παραγάγουν υψηλότερο σφάλμα στις εκτιμήσεις των δεδομένων του υποσυνόλου τεστ σε σχέση με εκτιμήσεις προηγούμενων δομών νευρωνικών με μικρότερο αριθμό κρυφών νευρώνων. Αυτό συμβαίνει γιατί έχει χαθεί η γενίκευση της υποκρυπτόμενης σχέσης και το νευρωνικό δίκτυο λόγω της αυξημένης δομής του αρχίζει να υπερ-προσαρμόζεται στα δεδομένα εκπαίδευσης.

Αντίθετα, κάνοντας χρήση του στατιστικού κριτηρίου του μέσου απολύτου σχετικού ποσοστιαίου σφάλματος στις εκτιμήσεις επί των δεδομένων του υποσυνόλου εκπαίδευσης για όλα τα νευρωνικά που παρήχθησαν διαπιστώνεται ότι τα σφάλματα μειώνονται περαιτέρω με την αύξηση του αριθμού των κρυφών νευρώνων στις δομές των νευρωνικών.

Διάγραμμα 6.1 απεικόνισης σφαλμάτων των δεδομένων υποσυνόλου εκπαίδευσης.



Διάγραμμα 6.2 απεικόνισης σφαλμάτων για τα δεδομένα του υποσυνόλου τεστ.



Στην προηγούμενη σελίδα, παρουσιάζονται τα διαγράμματα 6.1 των σφαλμάτων που παρήγαγε κάθε νευρωνικό για τα δεδομένα των υποσυνόλων εκπαίδευσης και τεστ για την εκτίμηση της ιδιότητας B_0 . Στο y άξονα αναφέρεται το μέσο απόλυτο σχετικό ποσοστιαίο σφάλμα ενώ στον x άξονα ο αύξοντας αριθμός του εκάστοτε νευρωνικού από το σύνολο των νευρωνικών που παρήχθησαν για την εκτίμηση αυτής της ιδιότητας B_0 . Με κόκκινο κύκλο αναφέρεται η περιοχή των νευρωνικών όπου αρχίζει να εμφανίζεται υπερ-προσαρμογή. Γνωρίζοντας ότι για κάθε δομή κρυφών νευρώνων πραγματοποιήθηκαν 100 επαναληπτικά αναπτύγματα, μπορούμε να υπολογίσουμε τον αριθμό των κρυφών νευρώνων του εκάστοτε νευρωνικού.

Παρατηρώντας το διάγραμμα των σφαλμάτων που παρήγαγε το κάθε νευρωνικό για τα δεδομένα του υποσυνόλου εκπαίδευσης προκύπτει ότι έχουν παραχθεί νευρωνικά με μεγάλη δομή που εμφανίζουν σημαντικά μεγαλύτερο σφάλμα σε σχέση με μικρές δομές. Αυτό εκ πρώτης όψεως αντιβαίνει το ορισμό της υπερεκπαίδευσης αλλά μπορεί να δικαιολογηθεί ως εξής:

Αυξάνοντας τον αριθμό των κρυφών νευρώνων σε ένα νευρωνικό ουσιαστικά αυξάνεται η δυνατότητα διάδοσης της πληροφορίας που εμπεριέχεται μεταξύ των εισόδων και εξόδων σε περισσότερα συναπτικά βάρη w που έχει σαν αποτέλεσμα την αύξηση του αριθμού των τοπικών ελαχίστων λόγω της αύξησης των διαστάσεων που σχετίζονται με την ελαχιστοποίηση της συνάρτησης κόστους J . Επειδή κάθε νευρωνικό ξεκινάει την επίλυση του προβλήματος βελτιστοποίησης από τυχαίες τιμές συναπτικών βαρών w , κατά τη σύγκλιση για την εύρεση του γενικού ελάχιστου μπορεί να εγκλωβιστεί σε οποιοδήποτε τοπικό ελάχιστο εμφανιστεί πριν το γενικό ελάχιστο λόγω της εφαρμογής του κριτηρίου «πρώου σταματήματος». Επειδή οι διαστάσεις της συνάρτησης κόστους J εξαρτώνται από τον αριθμό των συναπτικών βαρών w που αυτός με τη σειρά του εξαρτάται από τον αριθμό των κρυφών νευρώνων μπορούμε συμπεραίνεται ότι οι δομές των νευρωνικών με χαμηλό αριθμό κρυφών νευρώνων παρουσιάζουν τοπικά ελάχιστα με μικρότερη απόλυτη τιμή σφάλματος από τα αντίστοιχα των μοντέλων με περισσότερους νευρώνες αφού, λόγω χαμηλής διαστατικότητας, τα πρώτα ιχνηλατούν με λιγότερη ακρίβεια την τοπολογία της επιφάνειας σφάλματος.

Αντιμετώπιση του προβλήματος των τοπικών ελαχίστων και του γενικού ελάχιστου

Μια επιλογή για την επίλυση του προβλήματος της εύρεσης ή μη του γενικού ελάχιστου είναι η υιοθέτηση της ορμής (momentum). Η ορμή προσθέτει ένα κλάσμα m (το οποίο το ορίζει ο εκπαιδευτής) του κάθε προηγούμενου επικαιροποιημένου συναπτικού βάρους w στο κάθε τρέχον

συναπτικό βάρος w . Η ορμή στην ουσία αυξάνει το μέγεθος του βήματος κατά την διαδικασία της σύγκλισης με σκοπό την αποφυγή των μικρών τοπικών ελαχίστων. Μια υψηλή τιμή του κλάσματος m έχει ως συνέπεια την αύξηση του μεγέθους των βημάτων σύγκλισης που συνεπάγεται αύξηση της ταχύτητας σύγκλισης του συστήματος. Όμως, μια υψηλή τιμή ορμής μπορεί να οδηγήσει στην υπέρβαση του γενικού ελαχίστου και να οδηγήσει το σύστημα σε αστάθεια. Στον αντίποδα, μια πολύ χαμηλή τιμή ορμής μπορεί πάλι να εγκλωβιστεί σε κάποιο τοπικό ελάχιστο και ταυτόχρονα να καθυστερήσει περισσότερο την εκπαίδευση.

Ο υπολογισμός της βέλτιστης ορμής σε μια σύνθετη εκπαιδευτική διαδικασία όπως είναι η Levenberg Marquardt είναι μια πολύ δύσκολη διαδικασία με αμφίβολα αποτελέσματα και συνήθως δεν υιοθετείται. Στη παρούσα διπλωματική χρησιμοποιήθηκε μια πιο εμπειρική προσέγγιση η οποία στηρίζεται στη πιθανότητα ότι κάποια βαθμωτή κατάβαση ή οποία θα ξεκινήσει από μια τυχαία τιμή θα οδηγήσει στο γενικό ελάχιστο.

Γνωρίζοντας ότι νευρωνικό ξεκινάει τη βαθμωτή κατάβαση από μια τυχαία αρχική τιμή μέχρι να παγιδευτεί σε κάποιο τοπικό ελάχιστο, πραγματοποιώντας πολλά επαναληπτικά αναπτύγματα για τη ίδια δομή θεωρείται ότι κάποιο με βάση τις πιθανότητες θα ξεκινήσει την βαθμωτή κατάβαση από μια τυχαία τιμή που θα οδηγήσει στο γενικό ελάχιστο ή σε ένα τοπικό ελάχιστο πολύ κοντά σε αυτό .

IV. ΑΝΑΠΤΥΞΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΟΥ Βο

A. Δημιουργία Υπό-Βάσης δεδομένων για την εκπαίδευση του νευρωνικού

Πριν την ανάπτυξη του νευρωνικού δικτύου για την εκτίμηση των τιμών της ιδιότητας Βο στο σημείο φυσαλίδας-κορεσμού του ταμιευτήρα με χρήση δεδομένων παραγωγής, θα πρέπει να ελεγχθεί και να μορφοποιηθεί η βάση δεδομένων που θα χρησιμοποιηθεί για την εκπαίδευσή του. Μια από τις πρώτες εργασίες που πρέπει να λάβουν χώρα είναι η μετατροπή των πληροφοριών στη μορφή που απαιτείται για την εκπαίδευση του νευρωνικού. Εν συνεχεία, θα πρέπει να ελεγχθούν μια προς μια οι τιμές των δεδομένων βάσης για να διαπιστωθεί εάν κυμαίνονται στο αναμενόμενο εύρος τιμών ώστε να χρησιμοποιηθούν για την εκπαίδευση του νευρωνικού. Εάν όχι, θα πρέπει να εξαιρεθούν από την βάση δεδομένων ως μη αξιόπιστες.

Οι τιμές των πληροφοριών που εμπεριέχονται στο υποσύνολο της βάσης δεδομένων που επιλέχθηκε και χρησιμοποιήθηκαν για την ανάπτυξη του νευρωνικού προήλθαν από στιγμιαίες εκτονώσεις των πετρελαϊκών ρευστών σε συνθήκες επιφανείας. Το συγκεκριμένο υποσύνολο της βάσης δεδομένων αρχικά, πριν τον έλεγχο, απαριθμούσε 393 σημεία και περιείχε τις εξής πληροφορίες:

Πίνακας 6.2 Πίνακοποίηση πληροφοριών προερχόμενων από τις απευθείας εκτονώσεις των πετρελαϊκών ρευστών σε συνθήκες επιφανείας

			Reservoir fluid properties				Flashed Liquid				Flashed Gas			
RN	Tres (K)	Pb (MPa)	n	C7+ MW	Cn+ MW	MW	No	Cn+ MW	Cn+ ρ	MW	ρ	ng	MW	ρ @ pb

Από τις παραπάνω πίνακα χρησιμοποιήθηκαν οι εξής πληροφορίες:

Tres = Θερμοκρασία στο ταμιευτήρα σε Kelvin (K)

pb (Mpa) = πίεση στο σημείο φυσαλίδας σε megapascal (Mpa)

Reservoir fluid properties: ιδιότητες του ρευστού του ταμιευτήρα:

- MW = το μοριακό βάρος των συνολικών συστατικών του ρευστού του ταμιευτήρα

Flashed liquid: Υγρό το οποίο προήλθε από εκτόνωση του αρχικού ρευστού σε συνθήκες επιφανείας:

- n_o = η αναλογία των mol του υγρού στο ισοζύγιο μάζας $n_o + n_g = 1(z) \leftrightarrow n_o = 1 - n_g$
- MW = Το μοριακό βάρος του υγρού που εκτονώθηκε σε συνθήκες επιφανείας
- ρ = η πυκνότητα της υγρής φάσης σε συνθήκες επιφανείας

Flashed gas: Αέριο το οποίο προήλθε από εκτόνωση του αρχικού ρευστού σε συνθήκες επιφανείας:

- n_g = η αναλογία των mol αερίου στα συνολικά mol που εκτονώθηκαν $n_o + n_g = 1(z) \leftrightarrow n_g = 1 - n_o$
- MW = το μοριακό βάρος του αερίου

$\rho @ p_b$ = πυκνότητα πετρελαίου στο σημείο φουσαλίδας(κορεσμού)

R_N = α αριθμός του συγκεκριμένου πετρελαϊκού ρευστού

Χρησιμοποιώντας τα παραπάνω δεδομένα και ύστερα από μετατροπές μονάδων αλλά και υπολογιστικές πράξεις παρήχθησαν οι ζητούμενες πληροφορίες για την πυκνότητα πετρελαίου – αερίου σε συνθήκες επιφανείας, το GOR και το B_o στο σημείο κορεσμού-φουσαλίδας προκειμένου να εκπαιδευτεί το νευρωνικό:

					sm ³ /sm ³	scf/STB	bbl/STB
Tres (F)	Pb (psi)	γ_o	γ_g (SG)	(API)	GOR	GOR	B_o

Μετατροπές και υπολογιστικές πράξεις :

- $T_{res} (K) \rightarrow T_{res} (F) = (T_{res} (K) * 9 / 5) - 459.67$
- $P_b (Mpa) \rightarrow P_b (psi) = P_b (Mpa) * 145.0377$
- $\gamma_o = \frac{d_o sc}{d_w sc} = \text{Flashed Liquid } \rho / 1000 (\rho_{water})$
- $\gamma_g (SG) = \frac{d_g sc}{d_{air sc}} = \frac{\frac{P_{sc} * MW_g}{z * R * T_{sc}}}{\frac{P_{sc} * MW_{air}}{z * R * T_{sc}}} = \text{Flashed Gas } M_w / 28.97_{M_{wair}}$
- $API = (141.5 / \gamma_o) - 131.5$

$$\begin{array}{l}
 \blacksquare \quad GOR = \frac{V_{gassc}}{V_{oilsc}} \\
 ng = \frac{V_{gassc}}{V_m} \Rightarrow V_{gassc} = ng * V_m \\
 \rho_{oilsc} = \frac{m_{oilsc}}{V_{oilsc}} \Rightarrow V_{oilsc} = \frac{m_{oilsc}}{\rho_{oilsc}} \\
 no = \frac{m_{liquid}}{Mr_{liquid}} \\
 n_{res} = n_{oil} + n_{gas} \\
 V_{oilsc} = \frac{(1-ng)MW_{liquid}}{\rho_{liquid}}
 \end{array}
 \rightarrow
 GOR = \frac{ng}{1-ng} * \frac{23.69(lt / mol) * \rho_{liquid}(kg / m^3)}{MW_{liquid}(g / mol)}$$

- $GOR(sm^3/sm^3) = (23.69) * (ng/no) * [flashed\ liquid(p/Mw)]$
- $GOR(scF/sbbl) = GOR(sm^3/sm^3) * 5.651$

$$\begin{array}{l}
 \bullet \quad Bo = \frac{V_{oilres}}{V_{oilsc}} \\
 no = \frac{m_{liquid}}{Mr_{liquid}} \Rightarrow m_{liquid} = no * Mr_{liquid} \\
 \rho_{oil} = \frac{m_{oil}}{V_{oil}} \Rightarrow V_{oil} = \frac{m_{oil}}{\rho_{oil}} \\
 n_{res} = n_{oil} + n_{gas}
 \end{array}
 \rightarrow
 Bo = \frac{1}{1-ng} * \frac{\rho_{liquidsc}(kg / m^3) * MW_{liquidres}(g / mol)}{\rho_{liquidres}(kg / m^3) * MW_{liquidsc}(g / mol)}$$

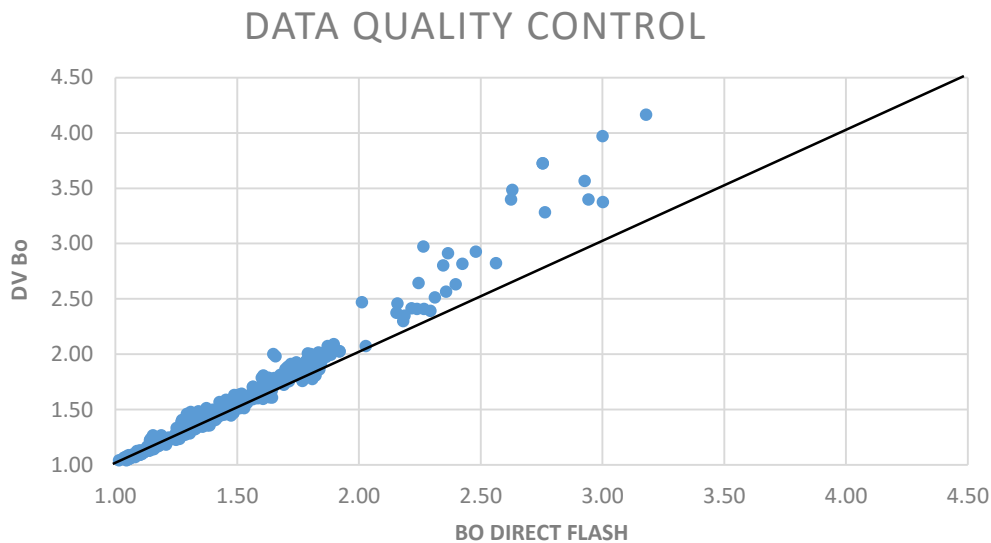
- $Bo(bbl/STB) = (\rho\ flash\ liquid / \rho@pb) * (MW_{res} / MW\ flash\ liquid) * (1/no)$

Ποιοτικός έλεγχος της βάσης δεδομένων

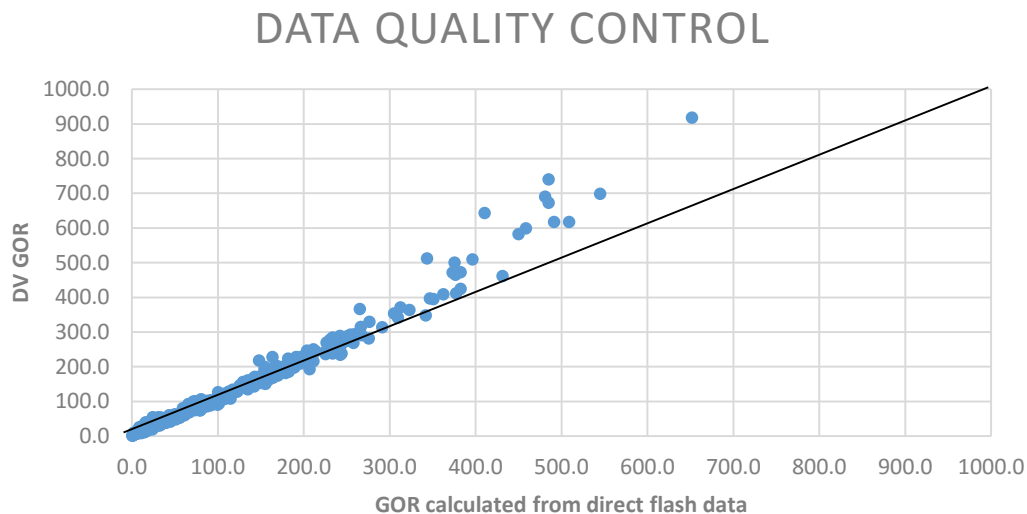
Για τον ποιοτικό έλεγχο της βάσης δεδομένων, πραγματοποιήθηκε σύγκριση των τιμών του ογκομετρικού συντελεστή υγρής φάσης (**Bo**) και του λόγου αερίου/πετρελαίου (**GOR**) που υπολογίστηκαν από την απ' ευθείας εκτόνωση των ρευστών του ταμειυτήρα σε κανονικές συνθήκες με τις αντίστοιχες τιμές των συγκεκριμένων ιδιοτήτων οι οποίες περιέχονται στην συνολική βάση δεδομένων προσδιορισμένες όμως από PVT μελέτη διαφορικής εξάτμισης. Ο στόχος ήταν η επιβεβαίωση των σωστών υπολογισμών των τιμών Bo και Rs που πραγματοποιήθηκαν από τα δεδομένα της απευθείας εκτόνωσης των πετρελαϊκών ρευστών. Ύστερα από ενδελεχή έλεγχο των δεδομένων βάσης διαπιστώθηκε ότι σε πολλά δείγματα υπήρχε σημαντική απόκλιση των υπολογισμένων ιδιοτήτων Bo και GOR από το απευθείας εκτόνωση σε σχέση με τις προσδιορισμένες τιμές αυτών των ιδιοτήτων από την διαφορική εξάτμιση. Οι απόκλίσεις αυτές διαπιστώθηκε ότι οφείλονταν σε κάποιες αντιμεταθέσεις στις εγγραφές των τιμών p_o και p_g από τον δημιουργό της βάσης δεδομένων η οποία και διορθώθηκε. Συνολικά διεγράφησαν από τη βάση 8 αναφορές πετρελαϊκών ρευστών είτε γιατί δεν διέθεταν επαρκείς πληροφορίες για το σύνολο των ιδιοτήτων του ρευστού είτε γιατί κάποιες τιμές ιδιοτήτων τους ήταν υπερβολικά απομακρυσμένες από το γενικό εύρος τιμών και η υιοθέτηση τους θα επηρέαζε την γενίκευση των υπό ανάπτυξη νευρωνικών δικτύων.

Στα διαγράμματα 6.4 & 6.5 παρουσιάζονται οι σύγκρισεις των υπολογισμένων τιμών των ιδιοτήτων Bo και GOR από τα δεδομένα απευθείας εκτόνωσης των πετρελαϊκών ρευστών με τις τιμές των ιδιοτήτων αυτών μετρημένων από μελέτη διαφορική εξάτμισης ώστε να αποδειχθεί η ορθότητα των υπολογισμών των δεδομένων που τελικά χρησιμοποιήθηκαν για την εκπαίδευση του νευρωνικού. Επί της διαγώνιου προφανώς οι τιμές ταυτίζονται.

Αφού χρησιμοποιήθηκε υπόγειο δείγμα για την απευθείας εκτόνωση, προκύπτει αυτόματα ότι $GOR = R_s$. Ύστερα από τους ελέγχους ποιότητας, η τελική υπό-βάση δεδομένων περιορίστηκε σε 385 αναφορές με τις οποίες αναπτύχθηκε το νευρωνικό δίκτυο που προβλέπει τις ιδιότητες του Bo.



Διάγραμμα 6.3 σύγκρισης του Bo απευθείας εκτόνωσης υπολογισμένου από τα δεδομένα βάσης σε σχέση με το Bo μετρημένο στο εργαστήριο από διαφορεική εξάτμιση.



Διάγραμμα 6.4 Απεικόνιση του GOR από απευθείας εκτόνωσης υπολογισμένου από τα δεδομένα βάσης σε σχέση με το GOR μετρημένο στο εργαστήριο από διαφορεική εξάτμιση.

Εν συνεχεία, παρατίθεται πίνακας σχετικά με το εύρος τιμών της βάσης δεδομένων που χρησιμοποιήθηκε για την ανάπτυξη του νευρωνικού.

Πίνακας 6.3 Εύρος PVT ιδιοτήτων 385 αναφορών.

Εύρος PVT ιδιοτήτων της υπό-βάσης δεδομένων του νευρωνικού

Παράμετροι εργαστηριακών μετρήσεων	Ελάχιστο	Μέγιστο	Μέσο	Τυπική Απόκλιση
Oil FVF At Pb ,Bob (bbl/STB)	1.02	3.72	1.47	0.45
Gas-oil Ratio GOR,Rsb (scf/STB)	1.6	4035	729.9	730.31
Tank-oil gravity (API)	10.4	48.5	33.3	9.19
Oil gravity γ _ο	0.786	0.997	0.862	0.05
Surface gas gravity (γ _g)	0.568	2.028	0.974	0.22
Reservoir Temperature,T(Fahrenheit)	85.77	375.98	204.56	47.05
Bubble-point pressure,Pb(psia)	36	6560	2174	1476.22

Κανονικοποίηση της βάσης δεδομένων

Η κανονικοποίηση της βάσης δεδομένων επιβάλλεται να γίνει γιατί τα δεδομένα των μεταβλητών που εμπεριέχονται στη βάση παρουσιάζουν μεγάλη διαφορά στη τάξη μεγέθους μεταξύ των τιμών για τις οποίες ορίζονται. Για παράδειγμα, το εύρος τιμών της πυκνότητας αερίου κυμαίνεται από 0.5 – 2 ενώ το εύρος τιμών της πίεσης είναι της τάξης των μερικών χιλιάδων psia. Για να αντιμετωπιστεί το πρόβλημα της διαφοράς στη τάξη μεγέθους στις τιμές μεταξύ των μεταβλητών, πραγματοποιείται κανονικοποίηση σε κλίμακα $[(-1) - 1]$ έτσι ώστε να αντιμετωπιστούν από το νευρωνικό οι τιμές εισόδου κατά την εκπαίδευσή του ισόβαρα. Στην παρούσα εργασία, η κανονικοποίηση γίνεται με χρήση εντολής «mapminmax» κατά την κατασκευή του κώδικα του νευρωνικού σε περιβάλλον Matlab. Το νευρωνικό κανονικοποιεί τις τιμές εισόδου και εξόδου ώστε να ξεκινήσει την διαδικασία της εκπαίδευσης. Εν συνεχεία παρατίθενται οι εξισώσεις που χρησιμοποιούνται για την κανονικοποίηση των τιμών x των εισόδων και y των εξόδων.

$$\bar{y} = 2 \frac{y - y_{\min}}{y_{\max} - y_{\min}} - 1 \qquad \bar{x} = 2 \frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1$$

B. Παρουσίαση της αρχής λειτουργίας του νευρωνικού δικτύου

Για να γίνει πιο εύκολα αντιληπτή η αρχή λειτουργίας του νευρωνικού δικτύου θα παρουσιαστεί ένα παράδειγμα μοντέλου σχετικού με το νευρωνικό δίκτυο το οποίο αναπτύχθηκε για την πρόβλεψη της ιδιότητας του Bo . Θα χρησιμοποιηθούν 4 είσοδοι αλλά και ένας μικρός αριθμός κρυφών νευρώνων (3) για το πρώτο κρυφό επίπεδο για την ευκολία της επίδειξης της υπολογιστικής μεθόδου. Την ίδια αρχή λειτουργίας ακολουθούν όλα τα νευρωνικά δίκτυα που χρησιμοποιήθηκαν για να αναπτυχθούν τα μοντέλα πρόβλεψης των ιδιοτήτων Bo , pb , Rs .

Πίνακας 6.2 Εισόδων και εξόδων

Είσοδοι X	Εξοδοι Y
$(API_1, SG_1, T_1, GOR_1)$	Bo_1
$(API_2, SG_2, T_2, GOR_2)$	Bo_2
$(API_n, SG_n, T_n, GOR_n)$	Bo_n
$(API_k, SG_k, T_k, GOR_k)$?

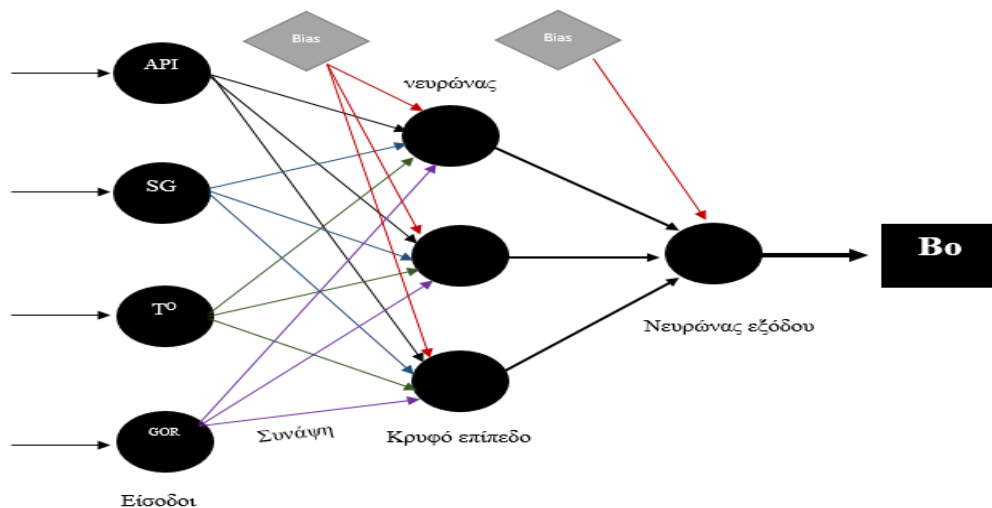
Στο παραπάνω πίνακα παρουσιάζεται η δομή των παραδειγμάτων εισόδου-εξόδου ώστε να επιτευχθεί η εκπαίδευση του νευρωνικού δικτύου μέσω της μάθησης της υποκρυπτόμενης σχέσης που συνδέει τις πληροφορίες των δεδομένων παραγωγής με τη ζητούμενη ιδιότητα Bo που εν συνεχεία εναποθηκεύεται ως γνώση υπό μορφή τιμών στα συναπτικά βάρη w που λειτουργούν ως μνήμη.

Η δομή του νευρωνικού δικτύου από άποψη αριθμού κρυφών επιπέδων (hidden layers) ή αριθμού κρυφών νευρώνων (hidden neurons) είναι προκαθορισμένη από τον εκπαιδευτή και αλλάζει ύστερα από εντολή του εκπαιδευτή όπως έγινε στη συγκεκριμένη περίπτωση με τη χρήση κώδικα και.

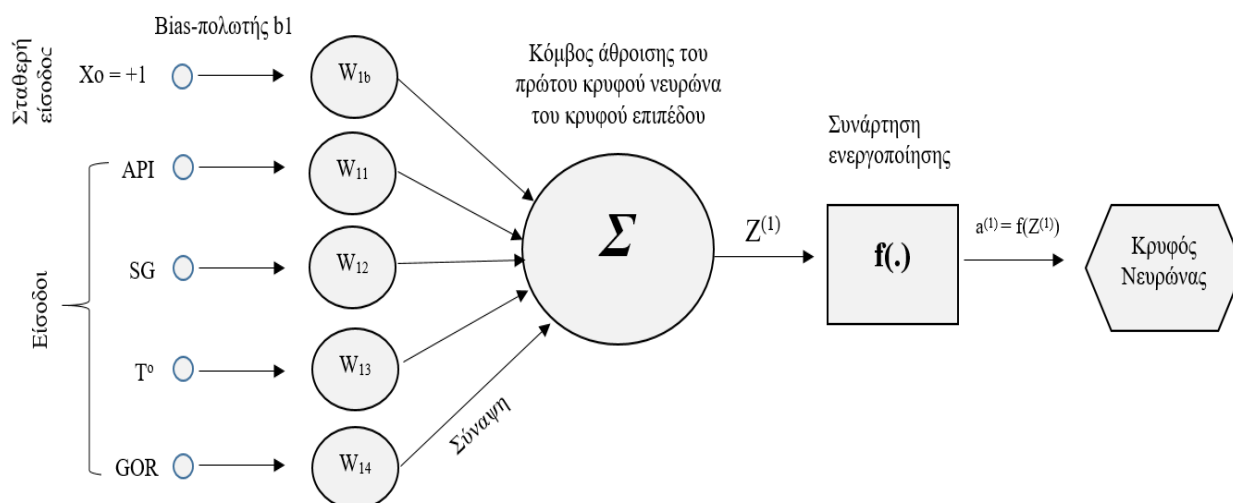
Θα πρέπει να σημειωθεί ότι αντί να εισαχθούν τα δεδομένα στο νευρωνικό δίκτυο ένα προς ένα, χρησιμοποιήθηκαν πίνακες που βοηθούν στη μείωση του χρόνου υπολογισμού κατά την επεξεργασία στο λογισμικό Matlab.

$$\begin{bmatrix} API_1 & SG_1 & T_1 & GOR_1 \\ \vdots & \vdots & \vdots & \vdots \\ API_n & SG_n & T_n & GOR_n \end{bmatrix}_{\substack{nx4 \\ INPUTSIZE}} = \text{δεδομένα εισαγωγής X} \quad \begin{bmatrix} Bo_1 \\ \vdots \\ Bo_n \end{bmatrix}_{\substack{nx1 \\ OUTPUTSIZE}} = \text{δεδομένα εξόδου Y}$$

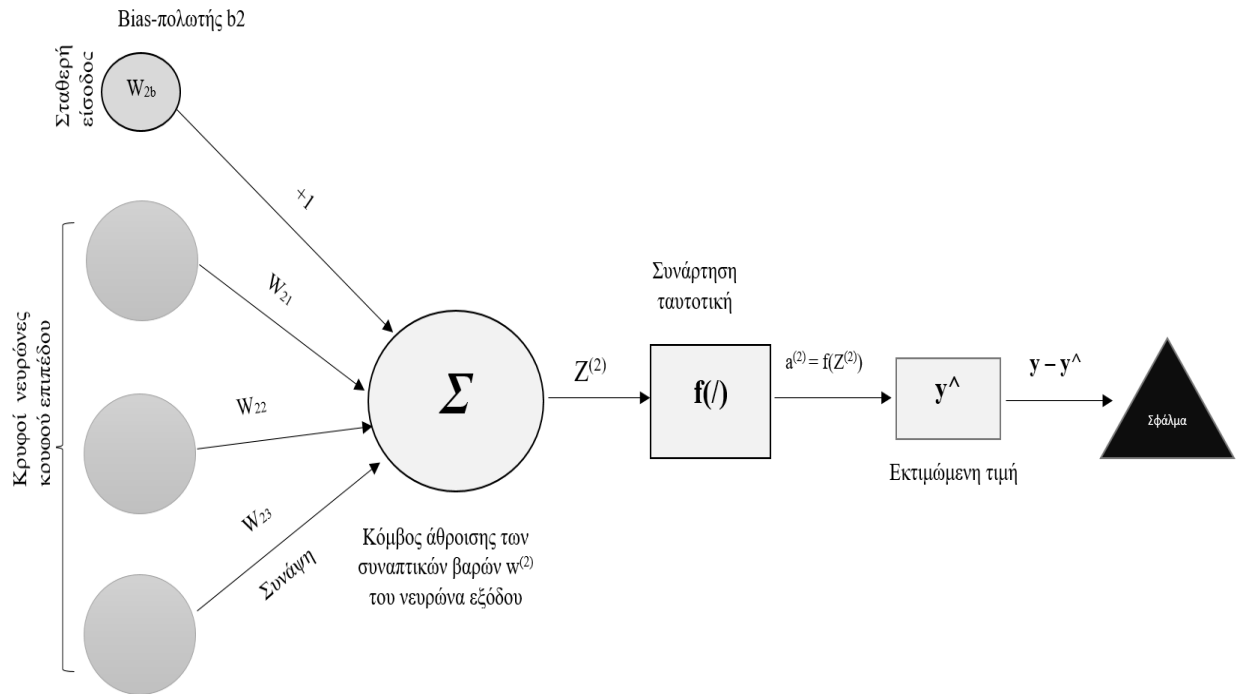
Στα σχήματα 6.4, 6.5 παρουσιάζονται γραφικά το μη γραμμικό μοντέλο νευρωνικού δικτύου για την πρόβλεψη της ιδιότητας B_o για να αναδειχθεί το πώς συσχετίζονται μαθηματικά οι πληροφορίες καθώς καταφθάνουν σε ένα κρυφό νευρώνα (hidden neuron) και σε ένα νευρώνα εξόδου (output neuron).



Σχήμα 6.4 Παρουσίαση της σύνδεσης του δικτύου κατά την πρόσθια τροφοδότηση με ένα κρυφό επίπεδο και ένα επίπεδο εξόδου.



Σχήμα 6.5 Γράφημα ροής σήματος από τις εισόδους στον πρώτο κρυφό νευρώνα και εν συνεχεία στο νευρώνα εξόδου



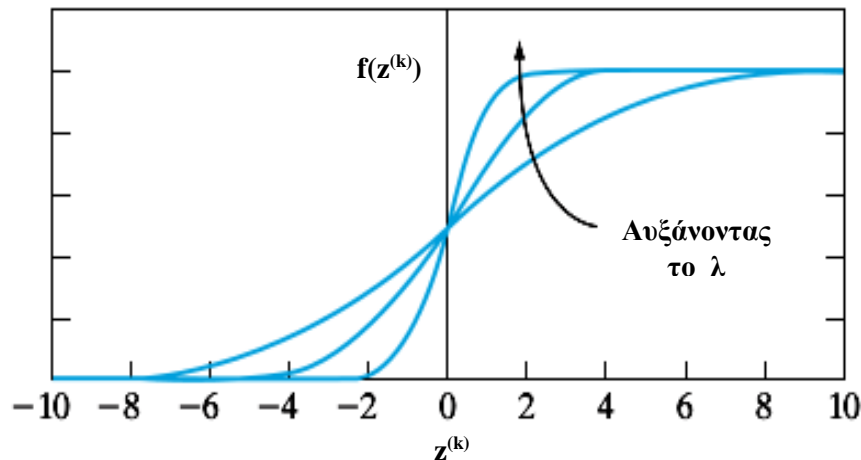
Σχήμα 6.6 Γράφημα ροής σήματος από τους κρυφούς νευρώνες στον νευρώνα εξόδου και εν συνεχεία στην εκτιμώμενη τιμή και τέλος στο σφάλμα

Σε αυτό το σημείο θα πρέπει να επισημανθούν τα βασικά χαρακτηριστικά τα οποία και απαιτούνται στα παραπάνω γραφικά μοντέλα :

1. Τα βέλη τα οποία ενώνουν τις πληροφορίες εισόδου με του κρυφούς νευρώνες και εν συνεχεία τους κρυφούς νευρώνες με τον νευρώνα εξόδου αποτελούν τις συνάψεις. Κάθε μια από αυτές τις διασυνδέσεις χαρακτηρίζεται από το δικό της βάρος ή δύναμη. Συγκεκριμένα ένα σήμα x_n στην είσοδο της σύναψης n που συνδέεται με το νευρώνα k πολλαπλασιάζεται επί το συναπτικό βάρος w_{kn} . Το συναπτικό βάρος ενός τεχνητού νευρώνα μπορεί να λαμβάνει και αρνητικές και θετικές τιμές
2. Οι κόμβοι άθροισης του απεικονίζονται χρησιμοποιούνται για την άθροιση των σημάτων εισόδου, σταθμισμένων από τα αντίστοιχα συναπτικά βάρη του νευρώνα. Οι λειτουργίες που περιγράφονται συνιστούν έναν γραμμικό συνδυαστή (linear combiner).
3. Χρησιμοποιείται επίσης μια συνάρτηση ενεργοποίησης (activation function) για τον περιορισμό του πλάτους του σήματος εξόδου ενός κρυφού νευρώνα. Η συνάρτηση ενεργοποίησης αναφέρεται επίσης ως συνάρτηση περιορισμού (squashing function), επειδή περιορίζει το επιτρεπτό εύρος πλάτους του σήματος εξόδου σε κάποια πεπερασμένη τιμή. Τυπικά το κανονικοποιημένο εύρος τιμών πλάτους της εξόδου ενός νευρώνα γράφεται ως

μοναδιαίο κλειστό διάστημα με τη μορφή $[0,1]$. Στην παρούσα εργασία ως συνάρτηση ενεργοποίησης μεταξύ των κόμβων άθροισης των κρυφών νευρώνων του κρυφού επιπέδου προς τους νευρώνες εξόδου χρησιμοποιήθηκε η σιγμοειδής συνάρτηση της οποίας η γραφική παράσταση θυμίζει το λατινικό γράμμα «S» όπως φαίνεται στο σχήμα 6.7 και είναι η πιο κοινή συνάρτηση που χρησιμοποιείται στα νευρωνικά δίκτυα. Είναι γνησίως αύξουσα συνάρτηση και περιλαμβάνει τμήματα τόσο σχεδόν γραμμικής όσο και μη γραμμικής συμπεριφοράς. Επίσης είναι διαφορίσιμη, ιδιότητα ιδιαίτερα χρήσιμη για τα νευρωνικά δίκτυα. Η σιγμοειδής συνάρτηση λαμβάνει τιμές στο συνεχές διάστημα $[0,1]$. Παράδειγμα σιγμοειδούς συνάρτησης είναι και η *λογιστική συνάρτηση* η οποία ορίζεται ως:

$$f(z^{(k)}) = \frac{1}{1 + \exp(-\lambda z^{(k)})}, \quad \lambda > 0$$



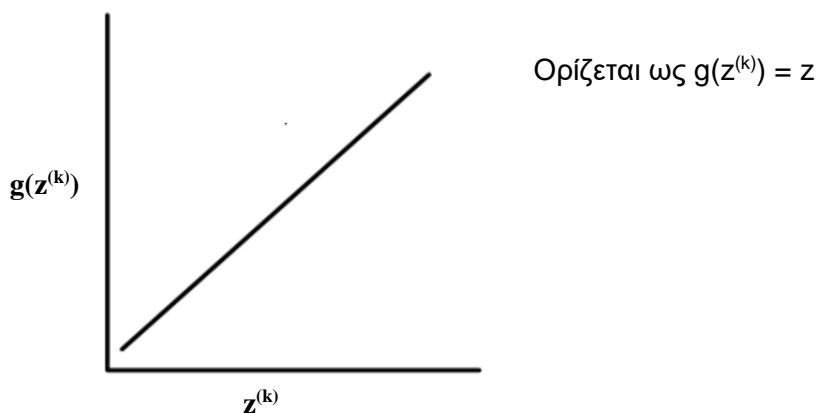
Σχήμα 6.7 σιγμοειδής συνάρτηση για μεταβαλλόμενη παράμετρο κλίσης λ

όπου : λ η παράμετρος **κλίσης** της συνάρτησης.

Μεταβάλλοντας την τιμή της παραμέτρου αυτής μεταβάλλεται και η κλίση της καμπύλης στην αρχή των αξόνων. Καθώς η παράμετρος λ τείνει στο άπειρο η καμπύλη προσεγγίζει τη συνάρτηση κατωφλίου.

4. Μεταξύ των νευρώνων εξόδου και της εκτιμώμενης τιμής δεν χρησιμοποιείται στη παρούσα διπλωματική κάποια συνάρτηση ενεργοποίησης αλλά χρησιμοποιείται μια ταυτοτική συνάρτηση. Όπως αποδεικνύεται μαθηματικά από την κλασική εργασία του

G.Cybenko “Approximation by Super positions of a Sigmoidal function” η χρησιμοποίηση ενός κρυφού επιπέδου και μιας συνάρτησης ενεργοποίησης μόνο για τις εξόδους των κρυφών νευρώνων μπορούν να δημιουργήσουν έναν επαρκώς πυκνό χώρο συναρτήσεων που είναι σε θέση να προσομοιώσουν οποιαδήποτε συνεχή συνάρτηση, όπως αυτές που η παρούσα διπλωματική εργασία προσπαθεί να προσομοιώσει, χωρίς να χρειαστεί η προσθήκη μιας δεύτερης συνάρτησης ενεργοποίησης στους νευρώνες εξόδου. Οπότε δεν χρησιμοποιείται δεύτερη συνάρτηση ενεργοποίησης μεταξύ των νευρώνων εξόδου και της εκτιμώμενης τιμής. Εν κατακλείδι ότι τιμές που βγαίνουν από το κόμβο άθροισης των νευρώνων εξόδου περνούν από μια ταυτοτική συνάρτηση και εν συνεχεία η τελική εκτιμώμενη τιμή αποκανονικοποιείται για να δοθεί στη μορφή που ορίζει ο εκπαιδευτής.



Σχήμα 6.8 απεικόνιση ταυτοτικής συνάρτησης

- Τα μοντέλα των νευρωνικών δικτύων που παρουσιάστηκαν στα σχήματα 6.4 – 6.5 περιλαμβάνουν εξωτερικά βάρη b_1, b_2 . Το βάρος b_k έχει ως αποτέλεσμα την αύξηση ή την μείωση της τιμής που δίνει σαν είσοδο το δίκτυο στην συνάρτηση ενεργοποίησης ανάλογα με το αν είναι αρνητικό ή θετικό. Το βάρος αυτό ονομάζεται μεροληψία (bias).

Με μαθηματικούς όρους, μπορούμε να περιγράψουμε έναν νευρώνα k από το παρακάτω ζεύγος εξισώσεων

$$u_k = \sum_{n=1}^m w_{kn}^{(i)} x_n$$

και

$$a_k = f(u_k + b_k)$$

- όπου x_1, x_2, \dots, x_n είναι τα σήματα εισόδου
- όπου w_1, w_2, \dots, w_n είναι τα αντίστοιχα συναπτικά βάρη του νευρώνα k

- u_k είναι η έξοδος του γραμμικού συνδυαστή που οφείλεται στα σήματα εισόδου
- b_k είναι η μεροληψία-bias
- $f(.)$ είναι η συνάρτηση ενεργοποίησης
- a_k είναι το σήμα εξόδου του νευρώνα

Η χρήση της πόλωσης b_k έχει ως αποτέλεσμα την προώθηση προς τα πάνω ή προς τα κάτω του μετασχηματισμού στην έξοδο u_k του γραμμικού συνδυαστή όπως υποδεικνύεται από την σχέση :

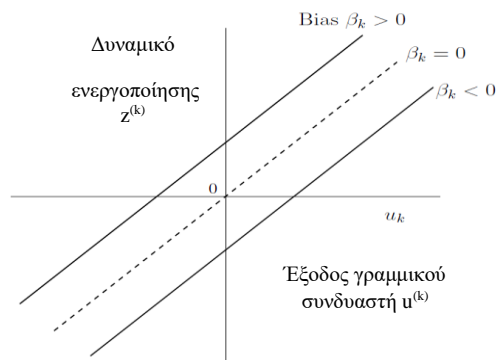
$$Z_k = u_k + b_k$$

Συγκεκριμένα, ανάλογα με το εάν η μεροληψία b_k είναι θετική ή αρνητική, η σχέση μεταξύ του δυναμικού ενεργοποίησης Z_k του νευρώνα k και της εξόδου του γραμμικού συνδυαστή u_k τροποποιείται με τον τρόπο που απεικονίζεται στο σχήμα 6.9 . Σε αυτό το σημείο θα πρέπει να σημειωθεί ότι με την εφαρμογή του μετασχηματισμού αυτού, η γραφική παράσταση του Z_k ως προς το u_k δεν περνάει από την αρχή των αξόνων.

Η μεροληψία-bias b_k είναι μια εξωτερική παράμετρος του τεχνητού νευρώνα k . Οπότε συνυπολογίζεται η ύπαρξη της ως εξής:

$$Z_k = \sum_{n=1}^m w_{kn}^{(i)} x_n \quad \text{και} \quad a_k = f(Z_k)$$

Στην παραπάνω εξίσωση έχει προστεθεί μια νέα σύναψη που έχει σαν είσοδο $X_0 = +1$ και βάρος $w_{k0} = b_k$



Σχήμα 6.8 Μετασχηματισμός της εξόδου u_k λόγω της ύπαρξης Μεροληψίας. Σημείωση : όταν $Z_k = b_k$ τότε $u_k = 0$

Εδώ θα πρέπει να επισημανθεί ότι όπως γίνεται αντιληπτό από τα προαναφερθέντα, το νευρωνικό δίκτυο αλλάζοντας τα συναπτικά βάρη $w_{kn}^{(i)}$ στην ουσία αλλάζει τη κλίση της σιγμοειδούς καμπύλης. Επίσης, δίνοντας του τη δυνατότητα να χρησιμοποιήσει μεροληψία-bias του δίνεται η δυνατότητα να μετατοπίσει ολόκληρη την σιγμοειδή καμπύλη δεξιά ή αριστερά έτσι ώστε να προσμοιώσει καλύτερα την εκτιμώμενη τιμή.

Δυναμικό ενεργοποίησης z_k

Χρησιμοποιώντας το παράδειγμα που αναφέρθηκε στην αρχή του κεφαλαίου γίνεται μια παρουσίαση των πινάκων για τον υπολογισμό του Z_k , εμπεριέχοντας την μεροληψία.

$$z_k = \begin{bmatrix} w_{11}^{(1)} & w_{21}^{(1)} & w_{31}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} & w_{32}^{(1)} \\ w_{13}^{(1)} & w_{23}^{(1)} & w_{33}^{(1)} \\ w_{14}^{(1)} & w_{24}^{(1)} & w_{34}^{(1)} \end{bmatrix} \begin{bmatrix} API_1 & SG_1 & T_1 & GOR_1 \\ \vdots & \vdots & \vdots & \vdots \\ API_n & SG_n & T_n & GOR_n \end{bmatrix} =$$

$$\begin{bmatrix} API_1 w_{11} + SG_1 w_{12} + T_1 w_{13} + GOR_1 w_{14} & API_1 w_{21} + SG_1 w_{22} + T_1 w_{23} + GOR_1 w_{24} & API_1 w_{31} + SG_1 w_{32} + T_1 w_{33} + GOR_1 w_{34} \\ \vdots & \vdots & \vdots \\ API_n w_{11} + SG_n w_{12} + T_n w_{13} + GOR_n w_{14} & API_n w_{21} + SG_n w_{22} + T_n w_{23} + GOR_n w_{24} & API_n w_{31} + SG_n w_{32} + T_n w_{33} + GOR_n w_{34} \\ +b_1^{(1)} & +b_2^{(1)} & +b_3^{(1)} \end{bmatrix}$$

\swarrow Κάθε στήλη αντιστοιχεί σε ένα νευρώνα \swarrow Κάθε γραμμή αντιστοιχεί σε μια τετράδα εισόδων

Η μορφή του πίνακα των συναπτικών βαρών από τους κρυφούς νευρώνες προς τον νευρώνα εξόδου έχει αυτή τη μορφή :

$$w_{kn}^{(2)} = \begin{bmatrix} w_{11}^{(2)} \\ w_{12}^{(2)} \\ w_{13}^{(2)} \\ +b^{(2)} \end{bmatrix}$$

4x1

Έχοντας υπόψη τα παραπάνω προκύπτουν τα εξής:

1. $z_1 = X * w^{(1)}$
2. $a^{(1)} = f(z_1)$
3. $z_2 = a^{(1)} * w^{(2)}$
4. $\hat{y} = g(z_2)$

Όπου \hat{y} είναι η εκτιμώμενη τιμή της ιδιότητας που μελετάται κάθε φορά

Συνάρτηση κόστους J

Για να γίνει κατανοητή η συμπεριφορά της σύγκλισης θα χρησιμοποιηθεί μια συνάρτηση κόστους. Υιοθετήθηκε σε αυτή την περίπτωση μια συνάρτηση κόστους που εκφράζει το άθροισμα των τετραγωνικών σφαλμάτων. Σημείωση ο συντελεστής κλιμάκωσης $\frac{1}{2}$ χρησιμοποιείται για σκοπούς απλοποίησης των τελικών μαθηματικών εκφράσεων.

$$J = \text{κοστος} = \frac{1}{2} e_1^2 + \frac{1}{2} e_2^2 + \frac{1}{2} e_3^2 = \sum \frac{1}{2} e^2 = \sum \frac{1}{2} (y - \hat{y})$$

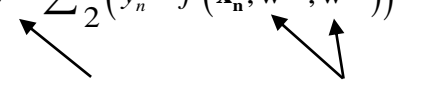
$$\text{Γενικά : } J(w) = \frac{1}{2} \sum_{n=1}^n e^2(n)$$

Όπως γίνεται αντιληπτό, όσο μικρότερος είναι ο αριθμός J τόσο μικρότερο είναι το σφάλμα άρα και τόσο καλύτερη πρόβλεψη της εκτιμώμενης τιμής σε σχέση με τη τιμή στόχο. Θα μπορούσε να πει κανείς ότι όταν αναφέρεται ο όρος εκπαίδευση του νευρωνικού στην ουσία αναφέρεται στην ελαχιστοποίηση της τιμής της συνάρτησης κόστους J. Θεωρείται ότι υπάρχει κάποιος συνδυασμός συναπτικών βαρών που ελαχιστοποιεί βέλτιστα τη συνάρτηση κόστους. Η τιμή του κόστους είναι συναρτήσει δύο πραγμάτων :

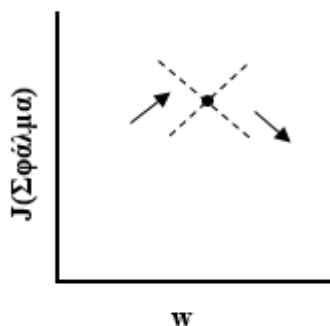
1. Των πληροφοριών εισόδου
2. Των βαρών των συνάψεων.

Επειδή οι πληροφορίες εισόδου δεν αλλάζουν η τιμή του κόστους είναι συνάρτηση μόνο των συναπτικών βαρών και της αμεροληψίας(bias).

$$J = \sum \frac{1}{2} \left(y_n - f(\mathbf{x}_n, \mathbf{w}^{(1)}, \mathbf{w}^{(2)}) \right)^2$$


 Πως αλλάζει αυτή η τιμή αλλάζοντας αυτές τις τιμές

Συναπτικά βάρη w συναρτήσει του



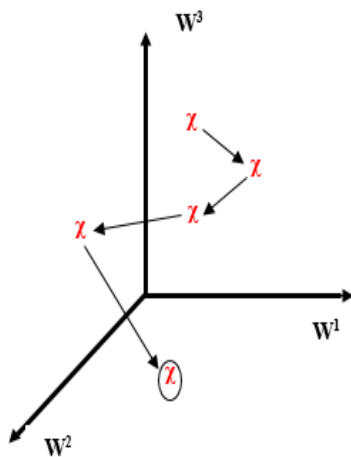
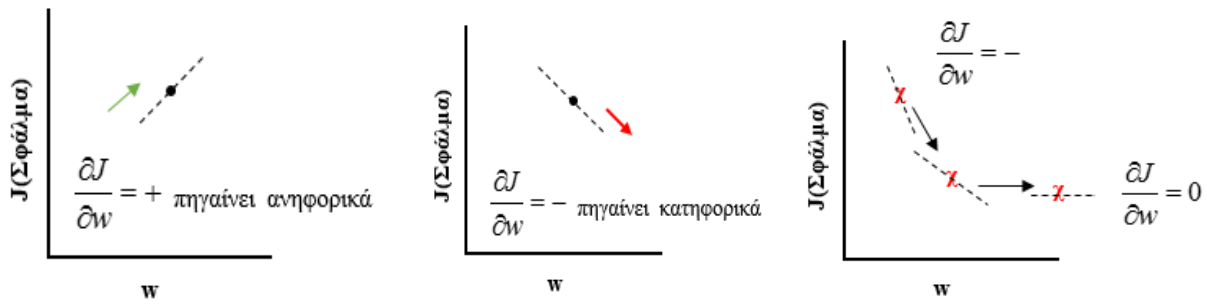
Αναλύοντας το σχήμα 6.9 γίνεται αντιληπτός ο συσχετισμός του σφάλματος με το βάρος w . Το νευρωνικό δίκτυο τροποποιώντας τις τιμές του w προσδοκά να μειώσει το σφάλμα. Θα πρέπει επίσης να σημειωθεί ότι το νευρωνικό δίκτυο ξεκινάει από μια τυχαία τιμή w .

Σχήμα 6.9 Συνάρτηση κόστους σε σχέση με το w

Εάν το νευρωνικό δίκτυο επέλεγε να αλλάζει τις τιμές του w προοδευτικά δηλαδή ξεκινώντας από το 0 έως ότου βρει την βέλτιστη τιμή σε αυτό το πρόβλημα μίας διάστασης (ενός w δηλαδή) δεν θα είχε πρόβλημα. Εάν όμως τα w που τροποποιεί το νευρωνικό δίκτυο ήταν περισσότερα τότε θα εμφανιζόταν το πρόβλημα της λεγόμενης «κατάρας της διαστατικότητας». Το πρόβλημα αυτό αναφέρεται στο γεγονός ότι, όταν μεγαλώνει ο αριθμός των χαρακτηριστικών ή των διαστάσεων, ο αριθμός των δεδομένων που πρέπει να γενικεύσουμε με ακρίβεια αυξάνεται εκθετικά. Για παράδειγμα εάν είχαμε ένα πρόβλημα με 9 διαστάσεις (δηλαδή 9 w) και το εύρος τιμών που έπαιρνε κάθε w ήταν από 0 έως 1000 τότε οι υπολογιστικές πράξεις που θα χρειαζόνταν να γίνουν θα ήταν 10^{27} .

Όπως διαπιστώνεται από το παραπάνω παράδειγμα, είναι αδύνατο να βρεθεί λύση σε λογικά χρονικά πλαίσια ακολουθώντας αυτή τη μέθοδο. Για να υπερκεραστεί το πρόβλημα αυτό, το νευρωνικό χρησιμοποιεί την μερική παραγωγή. Ξεκινώντας λοιπόν από μια τυχαία τιμή, το νευρωνικό χρησιμοποιεί το αποτέλεσμα της μερικής παραγωγού $\frac{\partial J}{\partial w}$ για να διαπιστώσει προς ποιά κατεύθυνση είναι «κατηφορικά» (downhill) προς τη λύση. Γνωρίζοντας αυτή την πληροφορία το νευρωνικό δίκτυο θα είναι σε θέση να γνωρίζει εάν πρέπει να μειώσει ή να αυξήσει την τιμή του w από την αρχική τυχαία τιμή ώστε να μειωθεί η συνάρτηση κόστους J δηλαδή το σφάλμα. Αυτή η μέθοδος ονομάζεται βαθμωτή κατάβαση (gradient decent) και τερματίζεται όταν η τιμή του κόστους δεν μπορεί να γίνει μικρότερη.

Διαγράμματα 6.9



Χρησιμοποιώντας το διάγραμμα 6.10, γίνεται αντιληπτή η βαθμωτή κατάβαση στη περίπτωση που υπάρχουν τρεις διαστάσεις. Κάθε βελάκι αντιπροσωπεύει το άθροισμα των μερικών παραγώγων (partial derivatives) κάθε διάστασης w .

$$\forall k, n, i: \frac{\partial J}{\partial w_{k,n}^{(i)}} \quad \sum \frac{\partial J}{\partial w_{kn}^{(i)}} = \text{Batch gradient descent}$$

Διάγραμμα 6.10

Συντελεστής ευαισθησίας και εύρεση τοπικής κλίσης

Η μερική παράγωγος $\frac{\partial J}{\partial w_{kn}^{(i)}}$ που αναφέρθηκε παραπάνω αντιπροσωπεύει ένα συντελεστή ευαισθησίας ο οποίος καθορίζει την κατεύθυνση της αναζήτησης στο χώρο των βαρών για το συναπτικό βάρος w_{kn} . Ο αλγόριθμος «back-propagation» εφαρμόζει στην ουσία μια διόρθωση Δw_{kn} στο συναπτικό βάρος w_{kn} η οποία είναι ανάλογη με τη μερική παράγωγο $\frac{\partial J}{\partial w_{kn}}$. Για να γίνει πιο εύκολα αντιληπτή η βαθμωτή κατάβαση (gradient descent), θα χρησιμοποιηθεί το παράδειγμα-επίδειξη των 3 νευρώνων το οποίο έχει οριστεί από την αρχή του κεφαλαίου.

$$\begin{array}{c}
 \mathbf{w}^1 = \begin{array}{c} \begin{matrix} & w_{ki}^{(1)} & \\ \begin{matrix} w_{11}^{(1)} & w_{21}^{(1)} & w_{31}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} & w_{32}^{(1)} \\ w_{13}^{(1)} & w_{23}^{(1)} & w_{33}^{(1)} \\ w_{14}^{(1)} & w_{24}^{(1)} & w_{34}^{(1)} \end{matrix} & \end{matrix} \\
 \begin{matrix} 4 \times 3 \end{matrix}
 \end{array}
 \end{array}
 \quad
 \begin{array}{c}
 \text{Μέγεθος κρυφού επιπέδου} \\
 \left. \begin{array}{c} \frac{\partial J}{\partial w_{ki}^{(1)}} \\ \frac{\partial J}{\partial w_{11}^{(1)}} \quad \frac{\partial J}{\partial w_{21}^{(1)}} \quad \frac{\partial J}{\partial w_{31}^{(1)}} \\ \frac{\partial J}{\partial w_{12}^{(1)}} \quad \frac{\partial J}{\partial w_{22}^{(1)}} \quad \frac{\partial J}{\partial w_{32}^{(1)}} \\ \frac{\partial J}{\partial w_{13}^{(1)}} \quad \frac{\partial J}{\partial w_{23}^{(1)}} \quad \frac{\partial J}{\partial w_{33}^{(1)}} \\ \frac{\partial J}{\partial w_{14}^{(1)}} \quad \frac{\partial J}{\partial w_{24}^{(1)}} \quad \frac{\partial J}{\partial w_{34}^{(1)}} \end{array} \right\} \begin{array}{c} \text{Μέγεθος το οποίο εξαρτάται από} \\ \text{τις πληροφορίες εισόδου} \end{array} \\
 \begin{matrix} 4 \times 3 \end{matrix}
 \end{array}
 \quad
 \frac{\partial J}{\partial w^{(1)}} = \begin{array}{c} \frac{\partial J}{\partial w_{ki}^{(1)}} \\ \frac{\partial J}{\partial w_{11}^{(1)}} \quad \frac{\partial J}{\partial w_{21}^{(1)}} \quad \frac{\partial J}{\partial w_{31}^{(1)}} \\ \frac{\partial J}{\partial w_{12}^{(1)}} \quad \frac{\partial J}{\partial w_{22}^{(1)}} \quad \frac{\partial J}{\partial w_{32}^{(1)}} \\ \frac{\partial J}{\partial w_{13}^{(1)}} \quad \frac{\partial J}{\partial w_{23}^{(1)}} \quad \frac{\partial J}{\partial w_{33}^{(1)}} \\ \frac{\partial J}{\partial w_{14}^{(1)}} \quad \frac{\partial J}{\partial w_{24}^{(1)}} \quad \frac{\partial J}{\partial w_{34}^{(1)}} \end{array}
 \end{array}$$

$$\begin{array}{c}
 \mathbf{w}^{(2)} = \begin{array}{c} \begin{matrix} w_{11}^{(2)} \\ w_{12}^{(2)} \\ w_{13}^{(2)} \end{matrix} \\
 \begin{matrix} 3 \times 1 \end{matrix}
 \end{array}
 \quad
 \begin{array}{c}
 \left. \begin{array}{c} \frac{\partial J}{\partial w_{11}^{(2)}} \\ \frac{\partial J}{\partial w_{12}^{(2)}} \\ \frac{\partial J}{\partial w_{13}^{(2)}} \end{array} \right\} \begin{array}{c} \text{Μέγεθος κρυφού επιπέδου} \end{array} \\
 \begin{matrix} 3 \times 1 \end{matrix}
 \end{array}
 \quad
 \frac{\partial J}{\partial w_{ki}^{(2)}} = \begin{array}{c} \frac{\partial J}{\partial w_{11}^{(2)}} \\ \frac{\partial J}{\partial w_{12}^{(2)}} \\ \frac{\partial J}{\partial w_{13}^{(2)}} \end{array}
 \end{array}$$

Θα μελετηθεί πρώτα η διόρθωση που εφαρμόζεται στα συναπτικά βάρη από τον νευρώνα εξόδου στους κρυφούς νευρώνες του κρυφού επιπέδου.

$$J = \sum \frac{1}{2} (y - \hat{y})^2 \rightarrow \frac{\partial J}{\partial w^{(2)}} = \frac{\partial \sum \frac{1}{2} (y - \hat{y})^2}{\partial w^{(2)}} \quad (6.1)$$

Για να εκφραστεί η τοπική κλίση, χρησιμοποιήθηκε ο κανόνας αλυσίδας (chain rule) του διαφορικού λογισμού ο οποίος διατυπώνεται ως εξής : $(f \circ g)' = (f' \circ g) g'$ επίσης ισχύει ότι

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx} \quad \text{Άρα :}$$

Εφαρμόζοντας τον κανόνα της αλυσίδας στην εξίσωση 6.1 και γνωρίζοντας ότι

$$\hat{y} = g(z_2) \text{ προκύπτει ότι : } \frac{\partial J}{\partial w^{(2)}} = -(y - \hat{y}) \frac{\partial \hat{y}}{\partial z_2} \frac{\partial z_2}{\partial w^{(2)}}$$

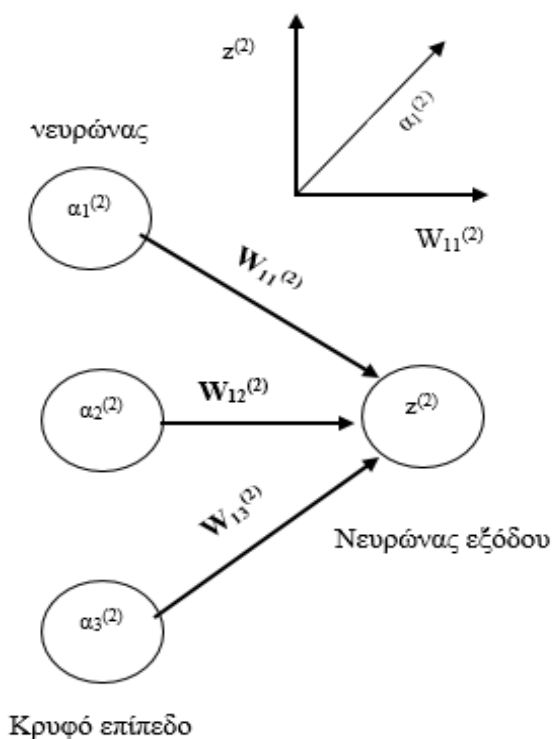
(6.2)

Για να βρεθεί ο ρυθμός μεταβολής της εκτιμώμενης τιμής \hat{y} συναρτήσει του z_2 θα πρέπει να διαφοριστεί το z_2 .

$g(z^{(k)}) = z$ οπότε η (6.2) γίνεται :

$$\frac{\partial J}{\partial w^{(2)}} = -(y - \hat{y}) g'(z_2) \left(\frac{\partial z_2}{\partial w^{(2)}} \right) \quad (6.3)$$

Αυτός ο όρος αντιπροσωπεύει την μεταβολή του z_2 συναρτήσει των βαρών $w^{(2)}$



$$z_2 = \begin{bmatrix} + & w_{11}^{(2)} a_1^{(2)} \\ + & w_{12}^{(2)} a_2^{(2)} \\ + & w_{13}^{(2)} a_3^{(2)} \end{bmatrix}$$

Για κάθε σύναψη υπάρχει μια γραμμική σχέση μεταξύ του z_2 και του $w^{(2)}$ όπου $a^{(2)}$ είναι η κλίση. Οπότε για κάθε σύναψη το $\frac{dz_2}{dw^{(2)}}$ εκφράζει την δραστηριότητα $a^{(2)}$ της σύναψης.

Στην ουσία αυτό που περιγράφεται είναι η οπισθοδιάδοση του σφάλματος σε κάθε βάρος. Πολλαπλασιαζόμενα με τη δραστηριότητα κάθε σύναψης τα βάρη που συνεισφέρουν περισσότερο στο σφάλμα θα έχουν μεγαλύτερες τιμές δραστηριότητας $a^{(2)}$ οπότε και μεγαλύτερες τιμές $w^{(2)}$. Για το λόγο αυτό θα τροποποιούνται περισσότερο κατά τη βαθμωτή κατάβαση.

Βασιζόμενοι στην εξίσωση (6.3) προκύπτουν τα εξής :

$$\begin{bmatrix} -(y_1 - \hat{y}_1) \\ -y_2 - \hat{y}_2 \\ -y_3 - \hat{y}_3 \end{bmatrix} * \begin{bmatrix} f'(z_{12}) \\ f'(z_{22}) \\ f'(z_{32}) \end{bmatrix} \rightarrow \begin{bmatrix} \delta_1^{(2)} \\ \delta_2^{(2)} \\ \delta_3^{(2)} \end{bmatrix} = \boldsymbol{\delta}^{(2)}$$

$$\begin{bmatrix} \alpha_{11}^{(2)} & \alpha_{21}^{(2)} & \alpha_{31}^{(2)} \\ \alpha_{12}^{(2)} & \alpha_{22}^{(2)} & \alpha_{32}^{(2)} \\ \alpha_{13}^{(2)} & \alpha_{23}^{(2)} & \alpha_{33}^{(2)} \end{bmatrix} \begin{bmatrix} \delta_1^{(2)} \\ \delta_2^{(2)} \\ \delta_3^{(2)} \end{bmatrix} = \begin{bmatrix} \alpha_{11}^{(2)} \delta_1^{(2)} + \alpha_{21}^{(2)} \delta_2^{(2)} + \alpha_{31}^{(2)} \delta_3^{(2)} \\ \alpha_{12}^{(2)} \delta_1^{(2)} + \alpha_{22}^{(2)} \delta_2^{(2)} + \alpha_{32}^{(2)} \delta_3^{(2)} \\ \alpha_{13}^{(2)} \delta_1^{(2)} + \alpha_{23}^{(2)} \delta_2^{(2)} + \alpha_{33}^{(2)} \delta_3^{(2)} \end{bmatrix}$$

Μέγεθος επιπέδου εξόδου

Μέγεθος κρυφού επιπέδου

$$\frac{\partial J}{\partial \mathbf{w}^{(2)}} = (\mathbf{a}^{(2)})^T \boldsymbol{\delta}^{(2)}$$

$$\boldsymbol{\delta}^{(2)} = -(y - \hat{y}) \mathbf{f}'(\mathbf{z}_2)$$

Σε αυτό το σημείο θα πρέπει να σημειωθεί ότι κάθε εγγραφή της βάσης δεδομένων που χρησιμοποιείται για την εκπαίδευση του νευρωνικού παράγει ένα δικό της $\frac{\partial J}{\partial w_{kn}^{(i)}}$ συντελεστή ευαισθησίας. Το άθροισμα όλων αυτών των συντελεστών παράγει μια συνισταμένη την οποία και ακολουθεί το νευρωνικό κατά την βαθμωτή κατάβαση.

Για να μελετηθεί η διόρθωση που εφαρμόζεται στα συναπτικά βάρη από τους κρυφούς νευρώνες προς τις πληροφορίες εισόδου ακολουθείται η ίδια λογική. Γίνεται και σε αυτή την περίπτωση χρήση του κανόνα αλυσίδας για την εύρεση της τοπικής κλίσης.

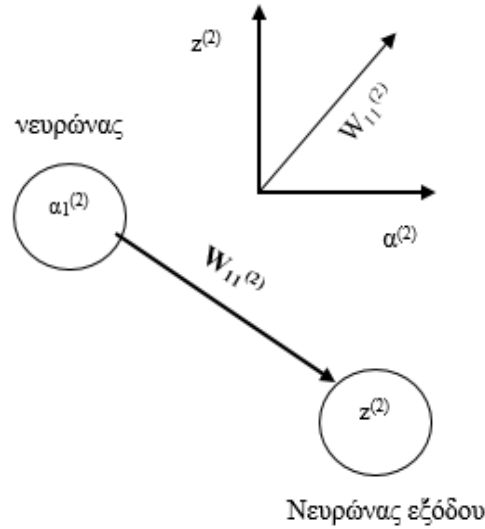
$$J = \sum \frac{1}{2} (y - \hat{y})^2 \rightarrow \frac{\partial J}{\partial w_{kn}^{(1)}} = \frac{\partial \sum \frac{1}{2} (y - \hat{y})^2}{\partial w_{kn}^{(1)}} \quad (6.4)$$

$$\frac{\partial J}{\partial w_{kn}^{(1)}} = -(y - \hat{y}) \frac{\partial \hat{y}}{\partial z_2} \frac{\partial z_2}{\partial w_{kn}^{(1)}} \quad (6.5)$$

$$\frac{\partial J}{\partial w_{kn}^{(1)}} = -(y - \hat{y}) f'(z_2) \frac{\partial z_2}{\partial w_{kn}^{(1)}} \quad (6.6)$$

$$\frac{\partial J}{\partial w_{kn}^{(1)}} = \delta^{(2)} \frac{\partial z_2}{\partial w_{kn}^{(1)}} \quad (6.7)$$

$$\frac{\partial J}{\partial w^{(1)}} = \delta^{(2)} \frac{\partial z_2}{\partial \alpha^{(2)}} \frac{\partial \alpha^{(2)}}{\partial w^{(1)}} \quad (6.8)$$



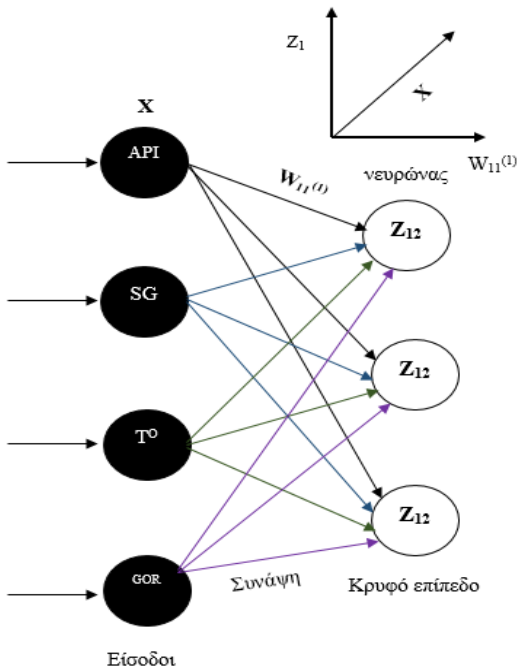
Στο παραπάνω σχήμα παρουσιάζεται μια γραμμική σχέση μεταξύ z_2 και $\alpha^{(2)}$ όπου η κλίση είναι το $w^{(2)}$. Οπότε για κάθε σύναψη το $\frac{dz_2}{d\alpha^{(2)}}$ εκφράζει το συναπτικό βάρος της σύναψης. Χρησιμοποιώντας αυτήν τη πληροφορία η εξίσωση (6.8) μπορεί να γραφτεί και ως εξής :

$$\frac{\partial J}{\partial w^{(1)}} = \delta^{(2)} (w^{(2)})^T \frac{\partial \alpha^{(2)}}{\partial w^{(1)}} \quad (6.9)$$

$$\frac{\partial J}{\partial w^{(1)}} = \delta^{(2)} (w^{(2)})^T \frac{\partial \alpha^{(2)}}{\partial z_1} \frac{\partial z_1}{\partial w^{(1)}} \quad (6.10)$$

$$f(z_1) = \frac{1}{1+e^{-z}} \rightarrow f'(z_1) = \frac{e^{-z}}{(1+e^{-z})^2} \quad (6.11)$$

$$\frac{\partial J}{\partial w^{(1)}} = \delta^{(2)} (w^{(2)})^T f'(z_1) \frac{\partial z_1}{\partial w^{(1)}} \quad (6.12)$$



Στο διπλανό σχήμα παρουσιάζεται μια γραμμική σχέση μεταξύ z_1 και $w^{(1)}$ όπου η κλίση είναι το X . Οπότε για κάθε σύναψη το $\frac{dz_1}{dw^{(1)}}$ εκφράζει την πληροφορία εισόδου X . Χρησιμοποιώντας αυτήν τη πληροφορία η εξίσωση (6.2) μπορεί να γραφτεί και ως εξής :

$$\frac{\partial J}{\partial w^{(1)}} = X^T \delta^{(2)} \overbrace{(w^{(2)})^T f'(z_1)}^{\delta^{(1)}} \quad (2.3)$$

Συγκεντρωτικά, για την διόρθωση των συναπτικών βαρών $w^{(1)}$ και $w^{(2)}$ αλλά και για την εύρεση των τοπικών κλίσεων από τα παραπάνω ισχύουν τα εξής:

$$\frac{\partial J}{\partial w^{(1)}} = X^T \delta^{(1)}$$

$$\delta^{(1)} = \delta^{(2)} (w_{kn}^{(2)})^T f'(z_1)$$

$$\frac{\partial J}{\partial w_{kn}^{(2)}} = (a^{(2)})^T \delta^{(2)}$$

$$\delta^{(2)} = -(y - y^{\wedge}) g'(z_2)$$

Εν κατακλείδει, η εύρεση του συντελεστή ευαισθησίας αλλά και της τοπικής κλίσης κάθε συνάψεως συσχετίζεται με το πρόβλημα της ανάθεσης εμπιστοσύνης και υπαιτιότητας. Το πρόβλημα αυτό είναι ένα πρόβλημα απόδοσης ευσήμων ή υπαιτιότητας για τα συνολικά αποτελέσματα σε κάθε μια από τις εσωτερικές αποφάσεις που λήφθηκαν από κρυφές υπολογιστικές μονάδες του κατανεμημένου συστήματος μάθησης, αναγνωρίζοντας εξαρχής ότι αυτές οι αποφάσεις είναι υπεύθυνες για τα συνολικά αποτελέσματα.

Απλουστευμένα κάθε συναπτικό βάρος w κάθε συνάψεως συμμετέχει με διαφορετικό ποσοστό στο τελικό παραγόμενο σφάλμα. Οπότε κατά την διαδικασία της οπισθοδιάδοσης του σφάλματος το νευρωνικό δίκτυο υπολογίζει τους συντελεστές ευαισθησίας και τις τοπικές κλίσεις κάθε συναπτικού βάρους κάθε συνάψεως και είναι σε θέση να γνωρίζει το ποσοστό συμμετοχής του κάθε w στο τελικό σφάλμα $y - \hat{y}$ και να το τροποποιήσει ανάλογα με το πόσο επηρεάζει το τελικό σφάλμα.

Γ. Τρόπος επιλογής του βέλτιστου νευρωνικού

Για να επιλεγεί το βέλτιστο νευρωνικό δίκτυο χρησιμοποιήθηκαν τα στατιστικά κριτήρια του μέσου απολύτου σχετικού ποσοστιαίου σφάλματος και του μέσου σχετικού ποσοστιαίου σφάλματος που αναφέρθηκαν προηγουμένως. Αξιολογώντας ποσοτικά τα σφάλματα των προαναφερθέν στατιστικών κριτηρίων που παράγονται επί των δεδομένων των υποσυνόλων εκπαίδευσης, επικύρωσης και test για κάθε αναπτυχθέν νευρωνικό γίνεται η επιλογή του βέλτιστου νευρωνικού με την εξής μεθοδολογία:

- Με τη χρήση του στατιστικού κριτηρίου του μέσου απολύτου ποσοστιαίου σφάλματος αναζητείται όχι μόνο το νευρωνικό που παραγάγει το χαμηλότερο στατιστικό λάθος στα δεδομένα του υποσυνόλου ελέγχου (test) αλλά το νευρωνικό που παραγάγει τα χαμηλότερα παραπλήσια στατιστικά λάθη επί των δεδομένων και των τριών υποσυνόλων (εκπαίδευσης, επικύρωσης και test) διασφαλίζοντας με αυτό τον τρόπο ότι το χαμηλό στατιστικό λάθος επί των δεδομένων του υποσυνόλου test δεν οφείλεται μόνο στη τυχαία διαίρεση των δεδομένων της υπό-βάσης στα υποσύνολα αλλά στην γενίκευση του νευρωνικού.

Για το απόλυτο σχετικό ποσοστιαίο σφάλμα, χρησιμοποιήθηκε μια μαθηματική φόρμουλα που περιγράφει την διαφορά της τιμής των σφαλμάτων επί των δεδομένων των υποσυνόλων εκπαίδευσης, επικύρωσης και test.

$$\text{εξίσωση αξιολόγησης} = (\text{train-validation}) + (\text{train-test})$$

Εφόσον επιλεγούν τα νευρωνικά με τα χαμηλότερα σφάλματα επί των δεδομένων των τριών υποσυνόλων εν συνεχεία εφαρμόζεται η παραπάνω εξίσωση ώστε να γίνει αντιληπτό πόσο παραπλήσια είναι τα σφάλματα που παραγάγουν τα βέλτιστα νευρωνικά επί των δεδομένων επικύρωσης και test σε σχέση με τα δεδομένα του υποσυνόλου εκπαίδευσης.

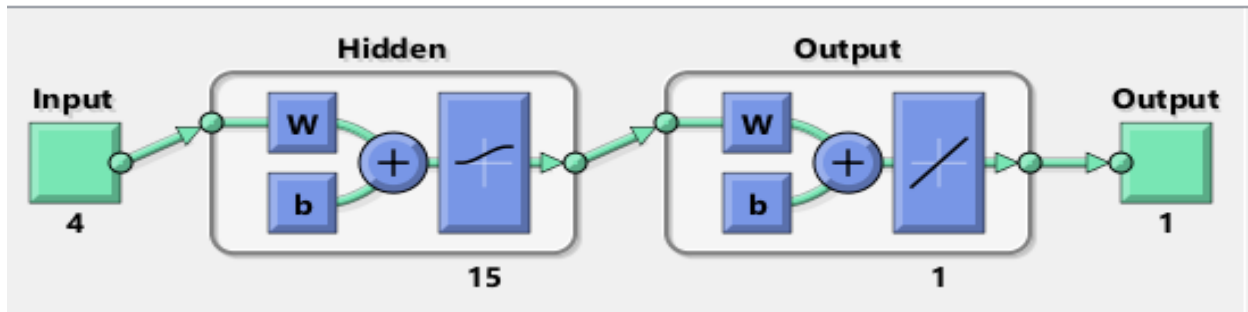
- Εάν τα νευρωνικά με τα χαμηλότερα σφάλματα επί των δεδομένων των τριών υποσυνόλων παρουσιάζουν κοντινές τιμές στην εξίσωση αξιολόγησης εν συνεχεία χρησιμοποιείται το στατιστικό κριτήριο του μέσου σχετικού σφάλματος που προσδιορίζει τη μέση κατανομή του σφάλματος γύρω από τη πραγματική τιμή.

Ως το βέλτιστο από τα νευρωνικά με τα χαμηλότερα σφάλματα επί των δεδομένων των τριών υποσυνόλων θα επιλεγεί αυτό που παρουσιάζει την χαμηλότερη υποεκτίμηση ή υπερκτίμηση σε σχέση με τη πραγματική τιμή της ιδιότητας του B_o .

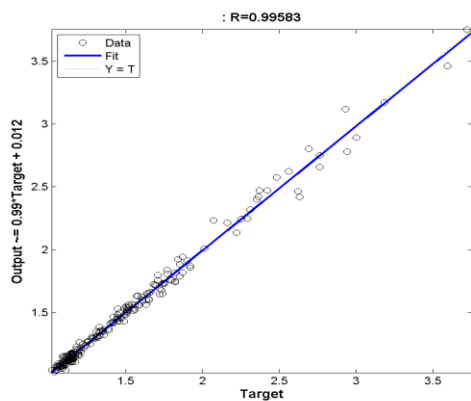
D. Παρουσίαση του βέλτιστου νευρωνικού

$$\text{Γενικός τύπος νευρωνικού: } y_{(1 \times 1)} = b_{2(1 \times 1)} + w_{2(1 \times 15)} f(b_{(15 \times 1)} + w_{1(15 \times 4)} x_{(4 \times 1)})$$

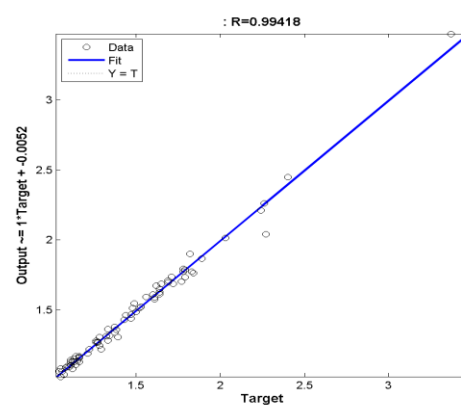
Δομή του νευρωνικού δικτύου για το Bo



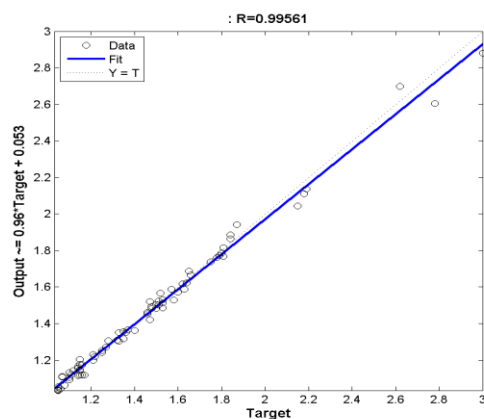
Regression plot of train Bo



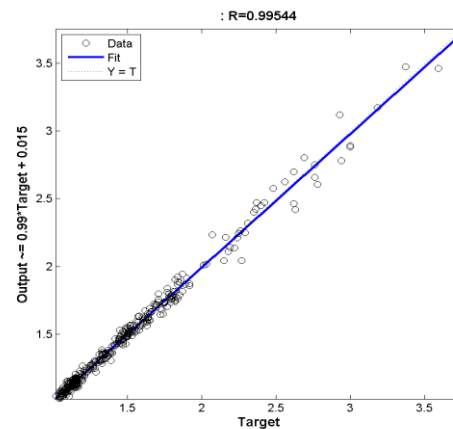
Regression plot of validation Bo



Regression plot of test Bo



Regression plot of all Bo



Στατιστικά κριτήρια	Er % μ	Ermax %	Ermin %	Ea %	STDEV(Er) % σ	68.20%		95.40%		99.70%		r
						$\mu - 1\sigma \leq x \leq \mu + 1\sigma$		$\mu - 2\sigma \leq x \leq \mu + 2\sigma$		$\mu - 3\sigma \leq x \leq \mu + 3\sigma$		
NEURAL-NEYΡΩΝΙΚΟ	0.15	15.49	-8.16	1.75	2.43	-2.28	2.58	-4.72	5.01	-7.15	7.44	0.9958

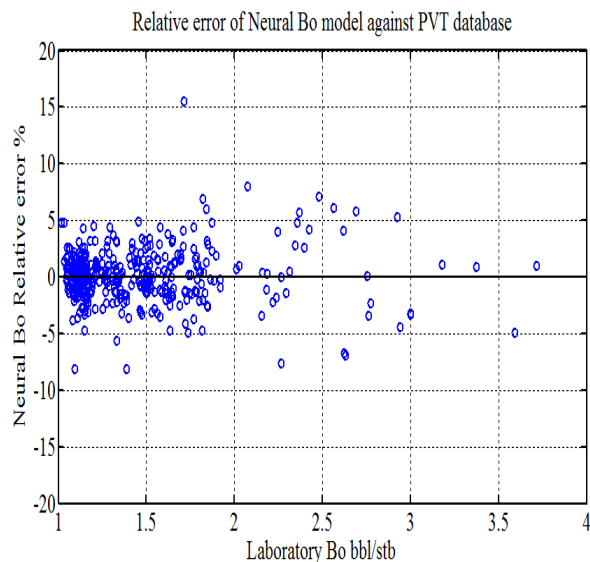
Πίνακας 6.5 Στατιστική ανάλυση του νευρωνικού Bo

Υποσύνολα	Er %	Ea %	r
TRAIN	0.15	1.83	0.9958
VALIDATION	0.26	1.76	0.9941
TEST	0.11	1.59	0.9956

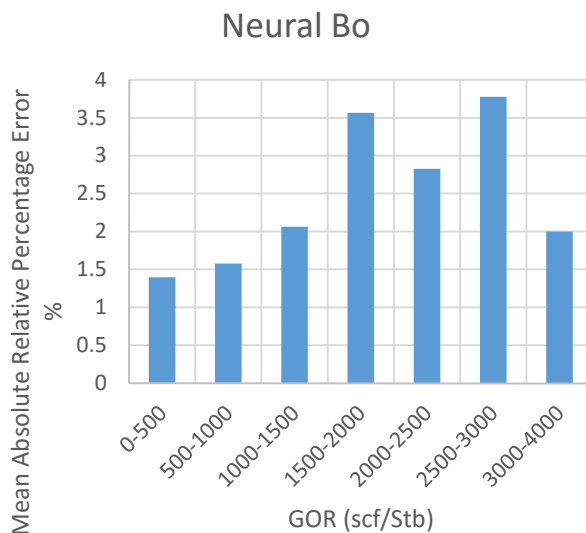
Πίνακας 6.6 των συναπτικών βαρών και των μεροληψιών του νευρωνικού Bo

Συναπτικά βάρη w(1)				Συναπτικά βάρη w ^T (2)	Bias(1)	Bias(2)
-3.28456	3.164886	1.53229	-2.10013	0.426677001	5.890716	-1.1685
4.145699	-0.78539	-3.52075	-0.16317	-0.025154834	-4.86501	
0.981666	-1.06154	2.151329	-4.44987	-0.264084369	-3.73426	
1.642933	1.087099	-2.80656	4.579946	1.221842491	-2.67067	
1.448278	3.479483	-3.40156	1.118448	-0.506571869	-2.56709	
-5.00891	-0.18823	2.452146	-0.27471	-0.033970995	1.533427	
1.214256	1.132841	0.472965	2.801125	1.235658194	-0.25839	
-0.63589	-0.00898	5.082987	-2.27749	0.317473772	0.602897	
2.975175	-0.22926	3.756728	-2.68236	0.534068064	1.092167	
2.606669	-0.83907	4.859496	-1.37152	-0.586589088	2.344483	
-4.97073	-2.48441	1.796151	-0.30943	0.136654024	-1.60361	
2.945932	2.633105	-0.85187	3.077564	0.161908862	3.834658	
-1.60305	-1.44502	-3.05563	3.848315	0.497432997	-4.12172	
-1.74655	1.026799	0.70646	-4.08676	-0.503577155	-4.32762	
3.614602	-0.01565	-3.89065	2.10104	-0.21140585	5.280782	

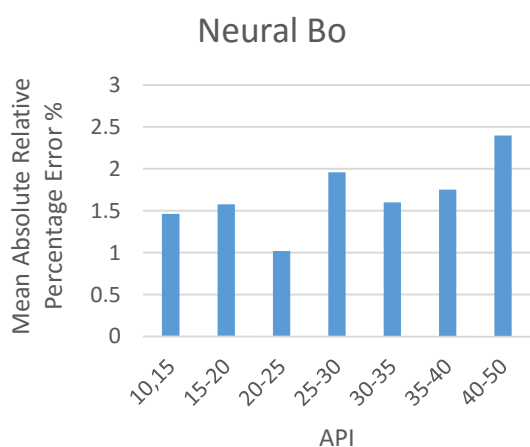
Εύρος PVT ιδιοτήτων του νευρωνικού Bo			
Ιδιότητες	Ελάχιστο	Μέγιστο	Μέσο
Bo (bbl/STB)	1.02	3.72	1.47
GOR (scf/STB)	1.6	4035	729.9
API ^o	10.4	48.5	33.3
SG	0.568	2.028	0.974
T(Fahrenheit)	85.77	375.98	204.56



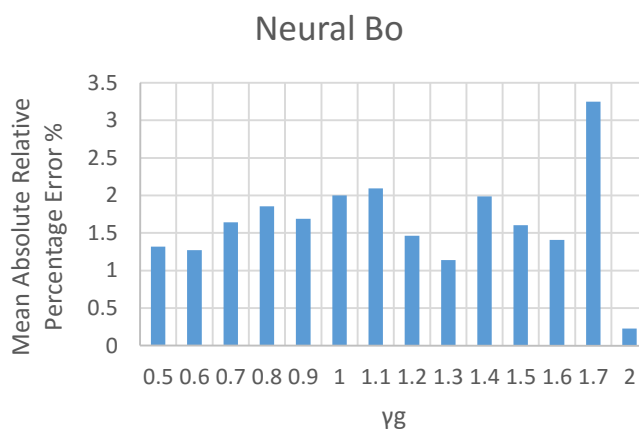
Σχήμα 6.10 Συσχέτιση σχετικού ποσοστιαίου σφάλματος του νευρωνικού-Bo με τις πειραματικές μετρήσεις του Bo



Σχήμα 6.11 Συσχέτιση του μέσου απολύτου σχετικού ποσοστιαίου σφάλματος του νευρωνικού-Bo με το GOR



Σχήμα 6.12 Συσχέτιση του μέσου απολύτου σχετικού ποσοστιαίου σφάλματος του νευρωνικού-Bo με το API



Σχήμα 6.13 Συσχέτιση μέσου του απολύτου σχετικού ποσοστιαίου σφάλματος του νευρωνικού-Bo με το γg

V. ΑΝΑΠΤΥΞΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΟΥ p_b

A. Βάση δεδομένων για την ανάπτυξη του νευρωνικού

Για την ανάπτυξη του νευρωνικού δικτύου για την εκτίμηση των τιμών της ιδιότητας του p_b χρησιμοποιήθηκε η ίδια υπο-βάση που επιλέχτηκε από την συνολική βάση δεδομένων του εργαστηρίου Ανάλυσης Ρευστών και Πυρήνων για την ανάπτυξη του νευρωνικού που εκτιμά τις τιμές της ιδιότητας του Bo . Επειδή οι πληροφορίες των δεδομένων παραγωγής που χρησιμοποιήθηκαν από το νευρωνικό ως πληροφορίες εισόδου με σκοπό την εύρεση της συσχέτισης τους με την ιδιότητα του σημείου φουσαλίδας-κορεσμού είναι ίδιες δεν χρειάστηκε κάποιος περεταίρω έλεγχος ή υπολογισμός για την χρησιμοποίηση της συγκεκριμένης υπό βάσης. Οι πληροφορίες των μετρήσεων των τιμών των πιέσεων του σημείου φουσαλίδας-κορεσμού των πετρελαϊκών ρευστών που χρησιμοποιήθηκαν για την ανάπτυξη του νευρωνικού προήλθαν από τις PVT μελέτες (σταθερής μάζας) που εμπεριέχονταν σαν πληροφορίες στην υπο-βάση πετρελαίων.

B. Παρουσίαση της αρχής λειτουργίας του νευρωνικού

Το νευρωνικό δίκτυο για την εκτίμηση των τιμών της ιδιότητας του p_b επιλύει αντίστοιχο εποπτευόμενο πρόβλημα παλινδρόμησης με αυτό που επιλύθηκε από το αναπτυχθέν νευρωνικό δίκτυο για την εκτίμηση των τιμών του Bo . Οπότε η αρχή λειτουργίας του είναι ακριβώς η ίδια με αυτή του νευρωνικού δικτύου για την εκτίμηση των τιμών του Bo και έχει αναλυθεί εκτεταμένη στην αντίστοιχη υπο-ενότητα του κεφαλαίου.

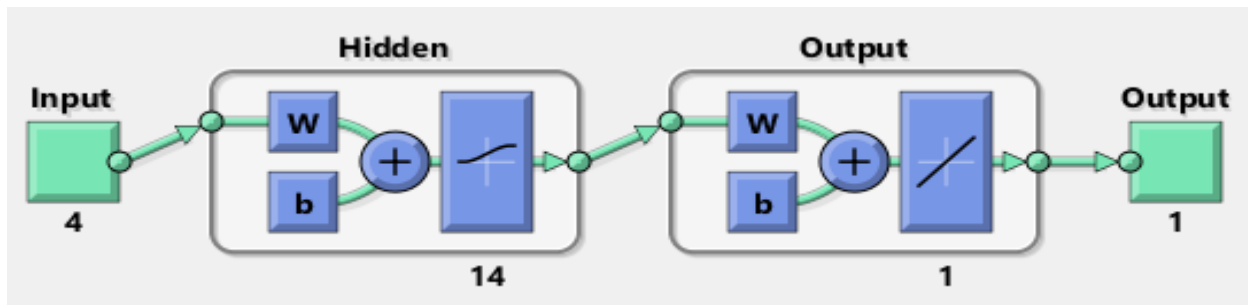
C. Τρόπος επιλογής του βέλτιστου νευρωνικού

Ο τρόπος επιλογής του βέλτιστου νευρωνικού για την εκτίμηση των τιμών της ιδιότητας του p_b από την πληθώρα των νευρωνικών που αναπτύχθηκαν για τη συγκεκριμένη ιδιότητα είναι ακριβώς ο ίδιος με αυτόν που παρουσιάστηκε στο στην αντίστοιχη υπο-ενότητα του κεφαλαίου για τον τρόπο επιλογής του βέλτιστου νευρωνικού δικτύου για την εκτίμηση των τιμών της ιδιότητας του Bo . Χρησιμοποιήθηκαν τα ίδια στατιστικά κριτήρια και η ίδια εξίσωση αξιολόγησης.

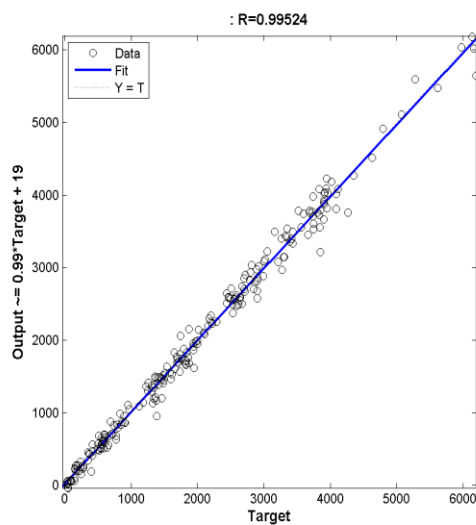
D. Παρουσίαση του βέλτιστου νευρωνικού

Γενικός τύπος νευρωνικού: $y_{(1 \times 1)} = b_{2(1 \times 1)} + w_{2(1 \times 14)} f(b_{(14 \times 1)} + w_{1(14 \times 4)} x_{(4 \times 1)})$

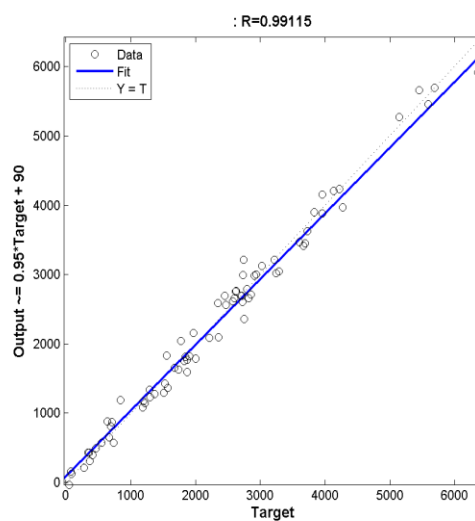
Δομή του νευρωνικού δικτύου pb



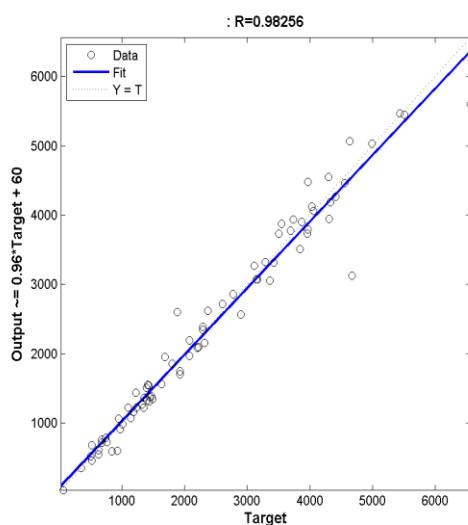
Regression plot of train Pb



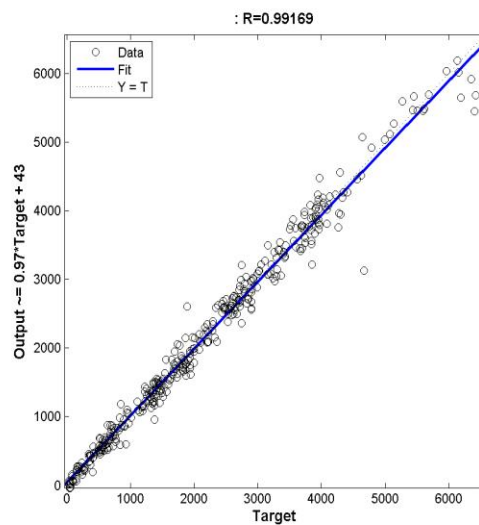
Regression plot of validation Pb



Regression plot of test Pb



Regression plot of all Pb



Πίνακας 6.7 Στατιστική ανάλυση του νευρωνικού δικτύου για το pb

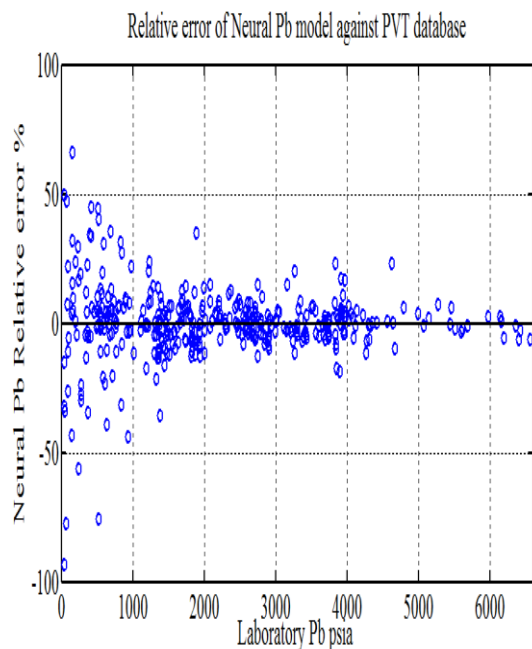
Στατιστικά κριτήρια	Er % μ	Ermax %	Ermin %	Ea %	STDEV(Er) %σ	68.20%		95.40%		99.70%		r
						$\mu - 1\sigma \leq x \leq \mu + 1\sigma$		$\mu - 2\sigma \leq x \leq \mu + 2\sigma$		$\mu - 3\sigma \leq x \leq \mu + 3\sigma$		
Neural-Νευρωνικό	-0.61	66.20	-130.87	9.48	17.30	-17.30	16.69	-35.21	33.99	-52.40	51.29	0.986

Υποσύνολα	Er %	Ea %	r
TRAIN	-0.82	9.46	0.9952
VALIDATION	-1.02	10.97	0.9911
TEST	0.46	8.04	0.9825

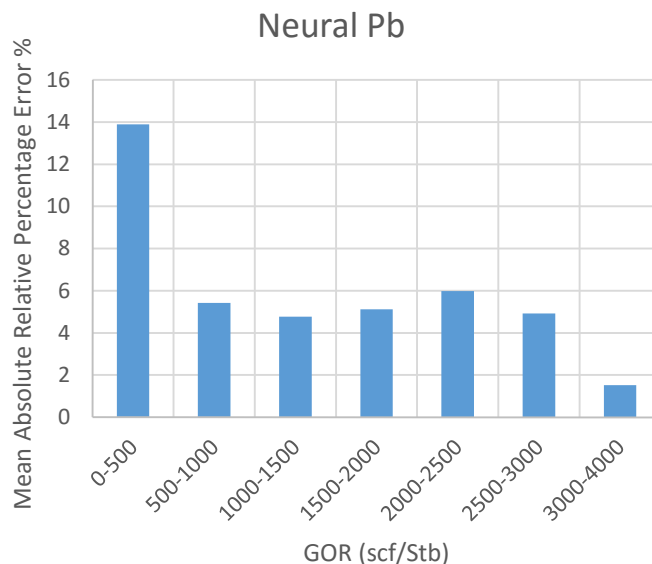
Πίνακας 6.8 των συναπτικών βαρών και των μεροληψιών του νευρωνικού pb

Συναπτικά βάρη $w(1)$				Συναπτικά βάρη $w^T(2)$	Bias(1)	Bias(2)
3.380822	0.210359	-0.61849	7.296407	0.152697243	4.36846	2.5045
-5.55697	0.889565	1.240916	8.545341	-1.79342159	-6.6818	
0.779362	3.415068	0.072544	-2.28479	-4.802447514	2.27138	
-1.12514	-1.50196	-3.77537	-3.92766	-0.941713153	-3.3782	
-0.58759	-1.18297	-1.24378	-1.58332	2.541713217	-1.4722	
-2.69709	-3.18892	5.553405	10.32614	0.491108861	-1.5167	
1.698373	-6.60536	3.058485	3.865867	1.524891222	9.08763	
-2.99586	2.885571	-0.80279	6.543668	-0.627759475	3.92733	
-3.28866	4.742599	-1.41211	4.439746	1.961429628	-0.6228	
3.899967	0.753504	8.765671	0.055221	-0.266416359	-0.1543	
-0.34657	-1.07229	-0.34114	-3.38233	-4.49331911	-4.7907	
-2.6569	-4.72309	0.541096	3.451401	-2.335524276	-2.9066	
-12.4342	-4.92294	1.570678	2.283091	0.454400632	-9.3209	
0.536281	4.839157	6.857698	-12.8539	0.071059905	-8.9243	

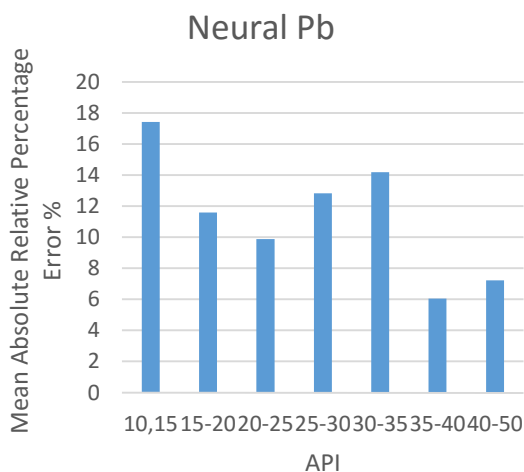
Εύρος PVT ιδιοτήτων του νευρωνικού Pb			
Ιδιότητες	Ελάχιστο	Μέγιστο	Μέσο
Pb (psia)	36	6560	2174
GOR (scf/STB)	1.6	4035	729.9
API ^o	10.4	48.5	33.3
SG	0.568	2.028	0.974
T(Fahrenheit)	85.77	375.98	204.56



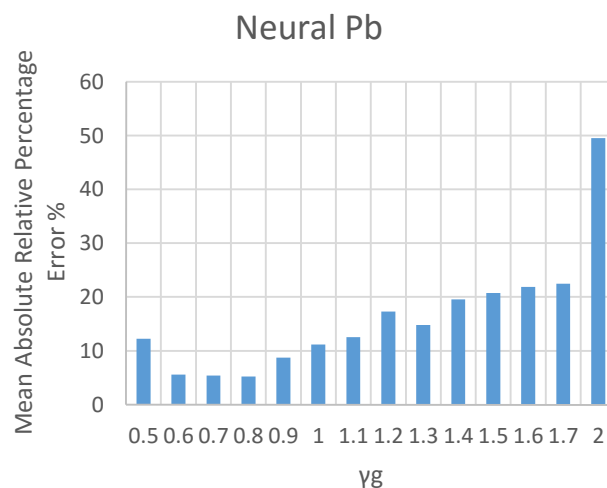
Σχήμα 6.14 Συσχέτιση σχετικού ποσοστιάου σφάλματος του νευρωνικού-Pb με τις πειραματικές μετρήσεις του Pb.



Σχήμα 6.15 Συσχέτιση του μέσου απολύτου σχετικού ποσοστιάου σφάλματος του νευρωνικού-Pb με το GOR



Σχήμα 6.16 Συσχέτιση του μέσου απολύτου σχετικού ποσοστιάου σφάλματος του νευρωνικού-Pb με το API



Σχήμα 6.17 Συσχέτιση του μέσου απολύτου σχετικού ποσοστιάου σφάλματος του νευρωνικού-Pb με τη πυκνότητα αερίου γg

VI. ΑΝΑΠΤΥΞΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΟΥ R_s

A. Βάση δεδομένων για την εκπαίδευση του νευρωνικού

Για την ανάπτυξη του νευρωνικού δικτύου που εκτιμά την ιδιότητα R_s , δημιουργήθηκε μια υπό-βάση δεδομένων η οποία συνδύαζε πληροφορίες από απευθείας εκτονώσεις των πετρελαϊκών ρευστών σε ατμοσφαιρικές συνθήκες καθώς και τις αντίστοιχες πληροφορίες για τα ίδια ρευστά από τα διάφορα βήματα πιέσεων των μελετών διαφορικής εξάτμισης αφού προηγουμένως μετασχηματιστούν ώστε να ανταποκρίνονται στην διεργασία σύνθετης εκτόνωσης. Ο λόγος για τον οποίο επιλέχθηκε να εμπλουτιστεί η υπό-βάση με τις πληροφορίες και των δύο τύπων μελετών ήταν για να δοθεί η δυνατότητα στο νευρωνικό να μπορεί να εκτιμήσει τις τιμές του R_s σε οποιαδήποτε πίεση κάτω από το σημείο φυσαλίδας-κορεσμού του αρχικού μονοφασικού ρευστού του ταμειυτήρα.

Επειδή οι τιμές του R_s από μελέτη διαφορικής εξάτμισης δεν μπορούν να χρησιμοποιηθούν απ' ευθείας για μετατροπές όγκων μεταξύ των συνθηκών του ταμειυτήρα και των συνθηκών επιφανείας όπως επεξηγήθηκε αναλυτικά στο κεφάλαιο 3, εφαρμόστηκε η υπολογιστική μέθοδος του L.Dake για να μετατραπούν οι τιμές R_{si} για κάθε στάδιο πίεσης της μελέτης διαφορικής εξάτμισης σε τιμές σύνθετης εκτόνωσης οι οποίες και εκφράζουν τις συνθήκες παραγωγής.

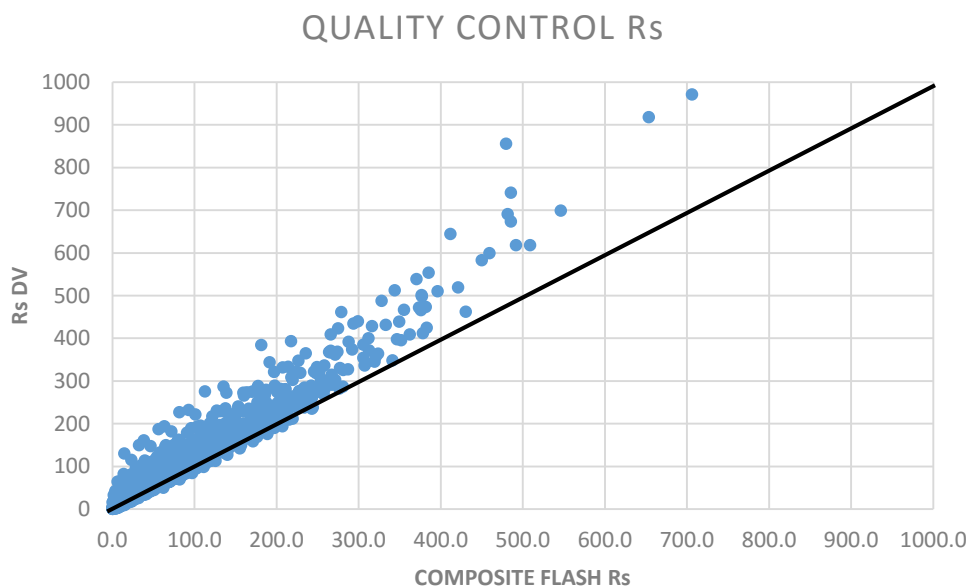
Το πλεονέκτημα αυτής της μεθοδολογίας ήταν ότι ανάγοντας τις τιμές της διαφορικής εξάτμισης σε τιμές σύνθετης εκτόνωσης έγινε εφικτό για κάθε DV καμπύλη ενός πετρελαϊκού ρευστού να ενταχθούν στην βάση δεδομένων 10 «καινούρια» ρευστά (όσα και τα βήματα της κάθε μελέτης διαφορικής εξάτμισης δηλαδή) από τα οποία για το καθένα είναι διαθέσιμη η τιμή R_s στο σημείο κορεσμού του. Με αυτό τον τρόπο εμπλουτίστηκε η υπό-βάση δεδομένων σε αναφορές τιμών R_s καινούριων πετρελαϊκών ρευστών με διαφορετικές συστάσεις.

Ποιοτικός έλεγχος της βάσης δεδομένων R_s

Για τον ποιοτικό έλεγχο αυτής της βάσης δεδομένων, πραγματοποιήθηκε ενδελεχής έλεγχος των υπολογισμένων δεδομένων R_s της σύνθετης εκτόνωσης όπου και διαπιστώθηκε πως σε πολύ χαμηλές πιέσεις ορισμένων μελετών PVT, συνήθως σε πιέσεις που αντιστοιχούν στα δύο τελευταία στάδια της διαφορικής εξάτμισης. Αυτό συνέβαινε όταν συγχρόνως παρατηρείτο σημαντική διαφορά των τιμών \mathbf{Bob}_f από απευθείας εκτόνωση και \mathbf{Bob}_d από διαφορική εξάτμιση διότι ο μαθηματικός τύπος του L.Dake που χρησιμοποιήθηκε για μετασχηματισμό των δεδομένων σε σύνθετης εκτόνωσης δεν εξασφαλίζει αλγεβρικά ότι δεν θα προκύψουν αρνητικές τιμές. Όσα δεδομένα προέκυψαν με αρνητικές τιμές θεωρήθηκαν εσφαλμένα και δεν εντάχθηκαν στη υπο-

βάση δεδομένων. Επίσης, διαγράφηκαν και όσες αναφορές είχαν τιμές **Bob_d** διαφορετικής εξάτμισης πάνω από 5 bbl/STB γιατί θεωρήθηκε ότι οι τιμές ήταν υπερβολικά απομακρυσμένες από το γενικό σύνολο τιμών και η υιοθέτηση τους στην υπό-βάση δεδομένων θα επηρέαζε την γενίκευση των νευρωνικών. Εν κατακλείδι, διαγράφηκαν 73 αναφορές από την αρχική υπό-βάση δεδομένων των 3530 αναφορών και η τελική βάση δεδομένων απαριθμεί 3457 αναφορές με τις οποίες αναπτύχθηκε το νευρωνικό δίκτυο για την πρόβλεψη του R_s .

Στο Διάγραμμα 6.4 πραγματοποιείται σύγκριση των αναφορών R_s από δεδομένα σύνθετης εκτόνωσης (composite flash) με δεδομένα R_s διαφορετικής εξάτμισης



Σχήμα 6.18 Διάγραμμα σύγκρισης των τιμών R_s υπολογισμένες από σύνθετη εκτόνωση (composite flash) σε σχέση με τις τιμές R_s μετρημένες από διαφορεική εξάτμιση (DV)

Εν συνεχεία, παρατίθεται πίνακας με το εύρος τιμών της υπό-βάσης δεδομένων που χρησιμοποιήθηκε για την ανάπτυξη του νευρωνικού του R_s .

Πίνακας 6.9 Εύρος PVT ιδιοτήτων 3457 αναφορών.

Εύρος PVT ιδιοτήτων της βάσης δεδομένων του νευρωνικού για την πρόβλεψη του Rs

Παράμετροι εργαστηριακών μετρήσεων	Ελάχιστο	Μέγιστο	Μέσο	Τυπική Απόκλιση
Oil FVF At Pb ,Bob flash (bbl/STB)	1.05	3.56	1.46	0.41
Oil FVF At Pb ,BobDV (bbl/STB)	1.04	4.63	1.54	0.54
Gas-oil Ratio GORf,Rsb flash (scf/STB)	4.6	3991.1	718.2	664.9
Gas-oil Ratio GOR,Rsb DV (scf/STB)	0.9	5485.1	447.9	490.1
Gas-oil Ratio GO,Rsb composite (scf/STB)	0.5	3991.1	729.9	394.0
Tank-oil gravity (API)	13.3	48.5	33.1	9.3
Oil gravity γ _ο	0.786	0.977	0.862	0.050
Surface gas gravity (γ _g)	0.570	1.719	0.960	0.200
Reservoir Temperature,T(Fahrenheit)	85.8	376.0	204.9	46.7
Bubble-point pressure,Pb(psia)	19.9	6773.0	1235.0	1091.7

B. Παρουσίαση της αρχής λειτουργίας του νευρωνικού

Το νευρωνικό δίκτυο για την εκτίμηση των τιμών της ιδιότητας του Rs και σε αυτή τη περίπτωση επιλύει αντίστοιχο εποπτευόμενο πρόβλημα παλινδρόμησης με αυτό που επιλύθηκε από τα αναπτυχθέντα νευρωνικά δίκτυο για την εκτίμηση των τιμών του Bo, pb. Οπότε η αρχή λειτουργίας του είναι ακριβώς η ίδια με τα υπόλοιπα αναπτυχθέντα νευρωνικά δίκτυα και έχει αναλυθεί εκτεταμένη στην αντίστοιχη υπο-ενότητα του κεφαλαίου που περιγράφει την ανάπτυξη του νευρωνικού δικτύου για την εκτίμηση των τιμών του Bo.

C. Ανηγμένο ή κανονικοποιημένο σχετικό ποσοστιαίο σφάλμα και απόλυτο σχετικό ποσοστιαίο σφάλμα για την αξιολόγηση της ιδιότητας Rs

Τα προηγούμενα στατιστικά κριτήρια για το σχετικό ποσοστιαίο σφάλμα και το απόλυτο σχετικό ποσοστιαίο σφάλμα λειτουργούν ικανοποιητικά στην αξιολόγηση των εκτιμήσεων των νευρωνικών και των εμπειρικών συσχετίσεων για τις ιδιότητες του Pb και του Bo αλλά δεν λειτουργούν ικανοποιητικά στην αξιολόγηση των εκτιμήσεων του νευρωνικού που εκτιμά τις τιμές της ιδιότητας Rs. Αυτό συμβαίνει γιατί το εύρος τιμών του Rs που καλείται να εκτιμηθεί μέσω των μοντέλων κυμαίνεται από πολύ χαμηλές τιμές κοντά στο 0 μέχρι πολύ υψηλές τιμές, πολλαπλάσιες του 1000. Για να γίνει πιο κατανοητή η έκταση του προβλήματος στις πιο χαμηλές τιμές, δίνεται το ακόλουθο παράδειγμα. Στη περίπτωση που ένα πετρελαϊκό ρευστό έχει εργαστηριακά

μετρημένη τιμή $R_s = 10$ scf/sbbl και μια εκτιμώμενη τιμή από τα μοντέλα είναι $R_s = 15$ scf/sbbl, το σχετικό ποσοστιαίο σφάλμα που παράγεται σε αυτή τη περίπτωση είναι της τάξης του 50% όμως αν αυτό το σφάλμα εξεταστεί από την σκοπιά της πρακτικής εφαρμογής των δεδομένων θα μπορούσε να θεωρηθεί σαν μια ανεκτή εκτίμηση.

Αυτό το πρόβλημα της μη αντιπροσωπευτικής προσέγγισης των τιμών των εκτιμήσεων παρατηρείται πιο συχνά στα πετρέλαια χαμηλής συρρίκνωσης γιατί οι τιμές του R_s είναι χαμηλές, της τάξεως των μερικών μόνο δεκάδων scf/sbbl. Όπως γίνεται αντιληπτό, για την αξιολόγηση της ακρίβειας ενός εκπαιδευμένου νευρωνικού, για τις χαμηλές τιμές της ιδιότητας R_s χρειάζεται μια κάποια ανοχή στο παραγόμενο σφάλμα.

Για την επίλυση του ανωτέρω προβλήματος για την ρεαλιστική εκτίμηση της ποιότητας των προβλέψεων της ιδιότητας R_s υιοθετήθηκε μια εμπειρική μαθηματική εξίσωση, η οποία αναπτύχθηκε από το Εργαστήριο Ανάλυσης Ρευστών και Πυρήνων και ως στόχο έχει την κανονικοποίηση του σχετικού ποσοστιαίου σφάλματος και του απολύτου σχετικού ποσοστιαίου σφάλματος ώστε να γίνει όσο το δυνατό πιο αξιόπιστη η αξιολόγηση των εκτιμώμενων αποτελεσμάτων των μοντέλων που αναπτύσσονται.

Έγινε λοιπόν θεώρηση ότι ένα σφάλμα της τάξης του 5% στις μεγάλες τιμές θα μπορούσε να είναι ανεκτό από την βιομηχανία πετρελαίου. Για τις μικρές τιμές της ιδιότητας του R_s , η εμπειρική εξίσωση αυξάνει την ανοχή στο παραγόμενο σφάλμα. Προσδίδει δηλαδή μια ανοχή της τάξης του 100% για τιμές R_s κοντά στο 1 που προοδευτικά μειώνεται όσο αυξάνονται οι τιμές του R_s μέχρις ότου φθάσει σε σταθερή ανοχή της τάξης του 5% που μπορεί να θεωρηθεί ανεκτή για τις μεγάλες τιμές του R_s .

Η φόρμουλα της εμπειρικής εξίσωσης που χρησιμοποιήθηκε είναι η ακόλουθη:

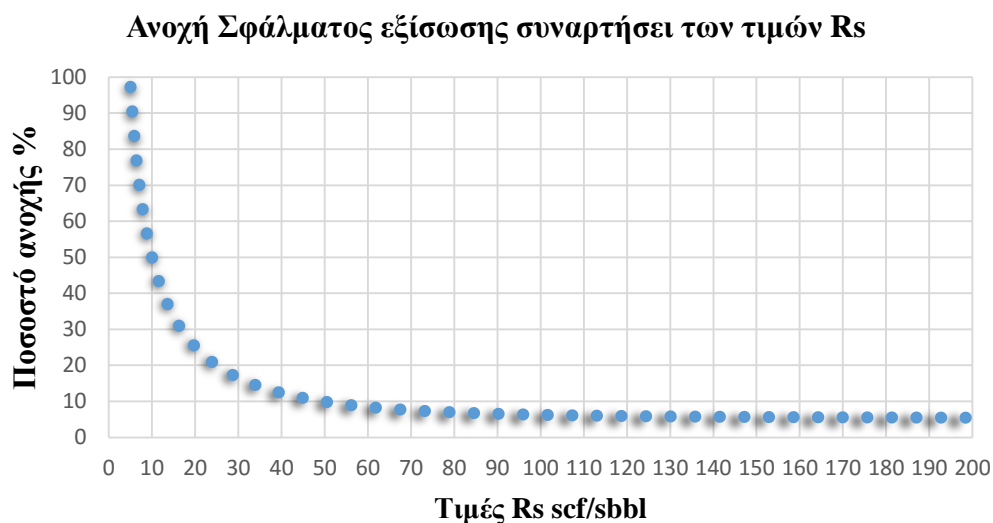
$$T \% = (50.245 + 178.386 * \exp(-0.043 * y_{i_real})) * (0.099 + (2.011 / y_{i_real}))$$

$$Ern = \frac{1}{n} \sum_{i=1}^n \left(\frac{ei}{y_{i_real}} * 100 \right) / T$$

$$Ean = \frac{1}{n} \sum_{i=1}^n \left(\left| \frac{ei}{y_{i_real}} \right| * 100 \right) / T$$

Για την αξιολόγηση του R_s θα χρησιμοποιηθεί το κανονικοποιημένο ή ανηγμένο σχετικό ποσοστιαίο σφάλμα και θα συμβολιστεί ως **Ern (mean normalized relative percentage error)** όπως επίσης και το κανονικοποιημένο ή ανηγμένο απόλυτο σχετικό ποσοστιαίο σφάλμα και θα συμβολιστεί ως **Ean (mean normalized absolute relative percentage error)**.

Στο παρακάτω διάγραμμα γίνεται αντιληπτή η διακύμανση του ποσοστού ανοχής, από τις μικρές τιμές όπου επιτρέπεται αυξημένη ανοχή μέχρι τις μεγαλύτερες τιμές όπου η ανοχή στο σφάλμα επιτρέπεται έως 5%.



Διάγραμμα 6.5 που απεικονίζει τη διακύμανση της ανοχής σφάλματος πρόβλεψης σε σχέση με τις τιμές του Rs.

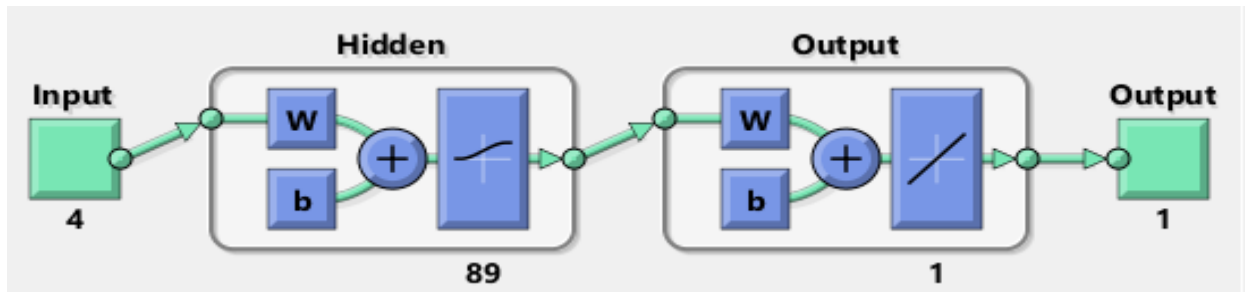
D. Τρόπος επιλογής του βέλτιστου νευρωνικού

Ο τρόπος επιλογής του βέλτιστου νευρωνικού για την εκτίμηση των τιμών της ιδιότητας του Rs από την πληθώρα των νευρωνικών που αναπτύχθηκαν για τη συγκεκριμένη ιδιότητα είναι ακριβώς ο ίδιος με αυτόν που παρουσιάστηκε στην αντίστοιχη υπο-ενότητα του κεφαλαίου για τον τρόπο επιλογής του βέλτιστου νευρωνικού δικτύου για την εκτίμηση των τιμών της ιδιότητας του Bo. Σε αυτή τη περίπτωση όμως χρησιμοποιήθηκαν το κανονικοποιημένο ή ανηγμένο σχετικό ποσοστιαίο σφάλμα και το κανονικοποιημένο ή ανηγμένο απόλυτο σχετικό ποσοστιαίο σφάλμα ως κριτήρια αξιολόγησης για να δοθεί η απαιτούμενη ανοχή στις εκτιμήσεις των μικρών τιμών. Για την εξίσωση αξιολόγησης χρησιμοποιήθηκαν επίσης τα προαναφερθέντα στατιστικά κριτήρια.

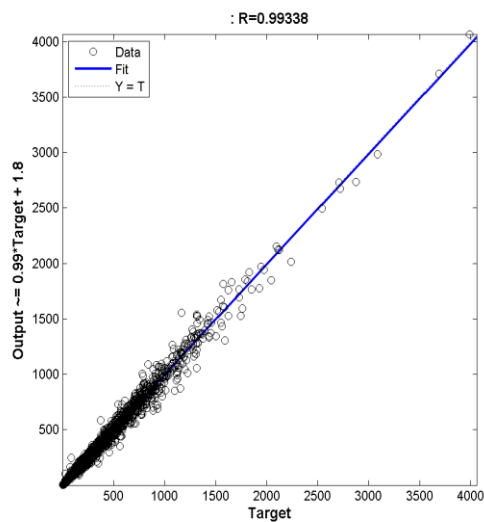
Ε. Παρουσίαση του βέλτιστου νευρωνικού

Γενικός τύπος νευρωνικού: $y_{(1 \times 1)} = b_{2(1 \times 1)} + w_{2(1 \times 89)} f(b_{(89 \times 1)} + w_{1(89 \times 4)} x_{(4 \times 1)})$

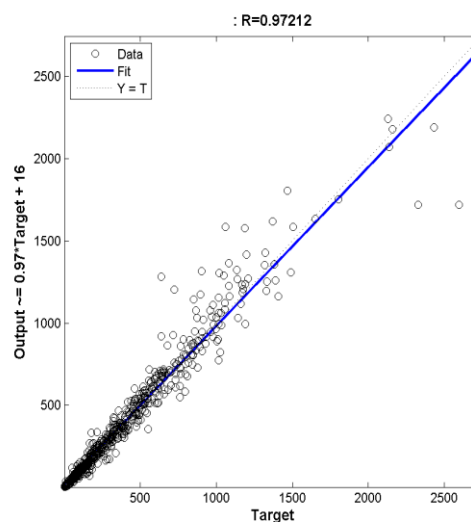
Δομή του νευρωνικού δικτύου για το Rs



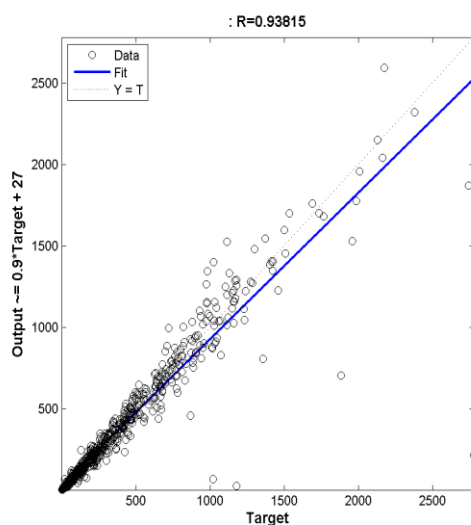
Regression plot of train Rs



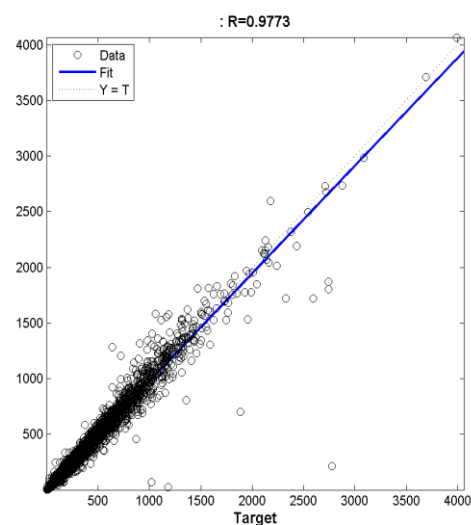
Regression plot of validation Rs



Regression plot of test Rs



Regression plot of all Rs



Πίνακας 6.9 Στατιστική ανάλυση του νευρωνικού Rs με χρήση κανονικοποίησης

Στατιστικά κριτήρια	Ern % μ	Ernmax %	Ernmin %	Ean %	STDEV(Ern) %σ	68.20%	95.40%	99.70%	r
						$\mu - 1\sigma \leq x \leq \mu + 1\sigma$	$\mu - 2\sigma \leq x \leq \mu + 2\sigma$	$\mu - 3\sigma \leq x \leq \mu + 3\sigma$	
Neural-Νευρωνικό	0.067	20.59	-7.47	0.985	1.55	-1.48 1.62	-3.04 3.17	-4.59 4.72	0.990

Πίνακας 6.10 Στατιστική ανάλυση του νευρωνικού Rs χωρίς χρήση κανονικοποίησης

Στατιστικά κριτήρια	Er % μ	Ermax %	Ermin %	Ea %	STDEV(Er) %σ	68.20%	95.40%	99.70%	r
						$\mu - 1\sigma \leq x \leq \mu + 1\sigma$	$\mu - 2\sigma \leq x \leq \mu + 2\sigma$	$\mu - 3\sigma \leq x \leq \mu + 3\sigma$	
Neural-Νευρωνικό	1.31	75.03	-33.47	6.99	11.92	-10.60 13.23	-22.51 25.15	-34.43 37.06	0.990

Υποσύ- νολα	Er %	Ea %	r
TRAIN	0.55	5.79	0.993
VALIDATION	1.67	8.46	0.972
TEST	1.25	9.16	0.938

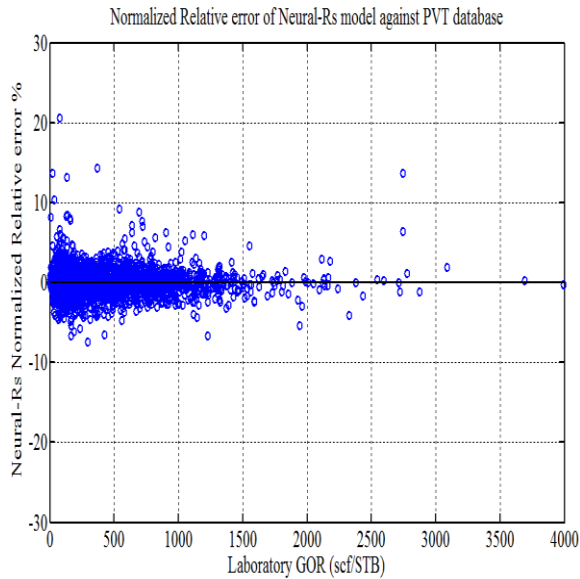
Πίνακας των συναπτικών βαρών και των μεροληψιών του νευρωνικού Rs

Συναπτικά βάρη w(1)				Συναπτικά βάρη w(2)	Bias(1)	Bias(2)
-1.10962	-0.19564	3.117341	-10.2994	-13.74289501	-8.1172	4.8579
-1.48404	11.16398	-7.15256	-4.50476	-15.93277918	20.7252	
9.400384	-0.10577	-4.86095	-2.48901	21.99816911	-9.0301	
-2.84427	-11.619	-7.18643	5.27145	-1.112330682	6.62025	
-5.74767	-1.2054	11.58369	-2.31479	-7.900060489	13.337	
-4.28248	-0.29804	5.21394	-16.6329	8.095631752	-13.56	
-14.9866	-1.5208	6.063288	-3.59418	-8.27300138	13.5009	
-1.65608	0.020743	1.654347	7.319831	-8.696327791	-3.8839	
3.105783	3.833521	0.015864	6.111105	5.096708326	-7.1211	
8.522497	1.092912	-14.3992	12.23234	5.619894927	-16.558	
5.979079	0.618428	-10.0195	8.228656	-14.97186904	-11.43	
3.556989	0.492925	-6.78467	23.04227	0.634868559	10.0186	
14.06169	0.082691	-10.9353	9.783218	-1.135327869	-3.1584	
-2.30081	-4.29514	-1.34936	-6.16893	3.107209865	6.33914	
2.059362	-12.3088	-3.0528	-2.60729	4.849145189	-13.334	
-1.35146	-1.91839	8.195845	-8.09981	-12.79814228	-11.641	
6.107605	0.344802	-8.84516	7.07596	-3.771160893	-7.2577	
11.81062	-0.21042	6.878921	11.3079	17.33386708	-8.5237	
-17.2691	-0.50206	12.35584	10.61936	7.737786363	14.6049	
7.210108	2.12189	2.822756	0.662855	-7.279162903	-6.0418	

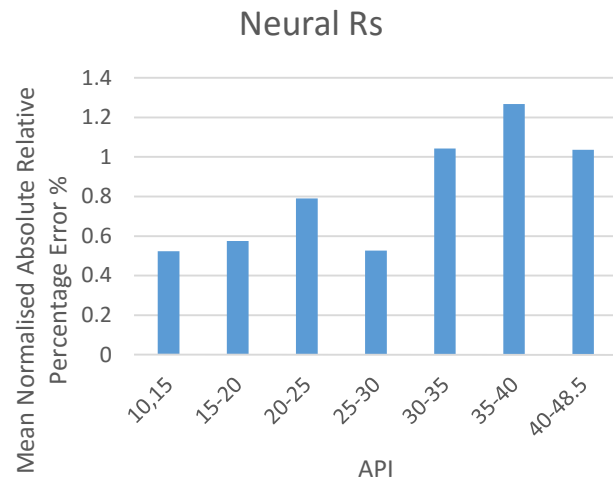
-14.1158	-1.74749	11.12127	-5.76266	-2.306435462	13.7315
-1.55986	0.064144	-11.7816	-5.99271	12.15389615	-6.3957
-1.03086	-0.24451	-8.87057	3.04791	13.74357892	0.95541
9.419176	1.460508	-0.91545	2.346713	-8.873925904	-6.8606
-2.73513	-1.58193	3.700599	0.641963	6.495299523	2.33887
14.55146	-0.19377	10.90713	10.62052	-9.507845213	-7.507
-10.3898	-0.42589	14.57511	9.519866	0.686387875	11.7215
11.5295	0.34623	3.534672	-15.4932	-3.92754579	-8.6654
-1.79847	1.625394	3.113184	-2.57704	-2.650969167	4.59702
0.314607	-2.19137	-1.42377	1.912473	2.681686725	-1.2556
-4.59688	-0.48062	6.845454	-4.18031	-14.61224792	6.36453
3.041719	-0.04003	15.10301	7.989447	6.912307226	8.3527
1.082464	-0.51153	-0.06094	-0.12315	-16.17288772	-0.2335
-7.16509	-0.17136	-2.74099	6.82142	-17.28138455	3.6757
-4.0633	-0.19844	-4.27079	4.816869	-9.4455726	-0.8968
1.468565	0.02066	-15.8479	6.893821	12.0587936	-4.1735
-7.80199	0.28136	-0.92404	0.547117	20.02834467	5.06455
-1.14374	0.277277	-13.7323	-5.4559	-1.728720843	-0.6378
0.715366	0.747956	11.7642	-2.6722	6.072347588	-1.0446
21.4676	0.074351	-19.2857	-25.7778	-0.690625198	-8.9989
12.69761	4.179334	8.75127	0.58316	1.900426069	1.60329
-2.74825	-0.8404	1.63948	-10.9281	3.082057399	1.95609
-0.83544	-0.11171	4.050693	7.238107	6.586538585	-2.6591
-18.4646	0.042296	-4.4697	-0.24946	-7.429676231	0.88131
4.979179	1.045597	-1.62953	14.72026	1.884118413	-2.4252
-18.3475	0.009965	-0.12318	-0.50147	3.600929095	2.94129
0.291396	0.112885	-7.77124	-3.53059	-10.03656726	-3.7934
12.19293	3.774895	7.956425	0.041639	-2.352285375	1.54249
-16.8812	0.06381	1.309863	-10.3267	-0.774924785	-3.0098
18.0463	-0.02387	-25.1856	-4.85067	-9.008544455	8.64918
0.516359	-0.56974	23.91701	-26.7964	-6.363341942	13.1899
18.12109	0.005068	7.247716	0.502012	-5.01994223	0.18667
-9.15714	0.108472	9.744201	-2.81166	-7.089063421	-7.6116
-6.85227	0.091789	-5.11661	2.149633	6.155611903	-3.614
-15.4093	-0.08166	27.72293	6.989057	-7.089266289	-7.5538
-2.74494	-0.08107	21.54358	-9.01083	7.308641287	5.73892
0.81596	-0.56997	27.12744	-30.8084	6.015624465	15.0281
3.644601	0.273977	2.902297	-4.84014	5.906116962	2.86225
0.55061	0.851204	-27.6587	9.387221	0.282036103	-12.951
-7.18973	-14.2037	13.75797	3.08336	0.031769882	-15.046
7.51897	1.698035	-32.0903	-6.11065	0.420084915	-0.3724
9.931727	-0.41743	10.7987	19.37412	0.279001075	-1.4435
-7.88615	0.344166	-4.2785	2.198389	-4.806226105	-3.4749

-20.0314	-0.29448	7.060625	-8.61651	-1.957141466	-11.8
3.506505	0.00715	-13.916	-10.2854	-1.488855329	4.85825
-13.4694	-0.28349	5.146915	-5.04442	5.403464055	-9.2932
4.293843	-0.54485	27.56029	12.75192	1.721735287	-5.2935
1.129212	-29.6956	1.99441	1.905205	-14.09024623	-34.029
1.619591	5.005596	-3.86365	-1.85816	-5.727233704	-8.765
16.2741	0.934775	16.10827	12.72503	-5.245542383	20.5857
-10.1616	0.005054	1.353584	1.242564	1.722167715	-0.7501
-1.98078	11.94615	2.95491	2.475091	5.354809293	12.9816
-3.16856	-2.48574	6.630142	-7.70874	11.17261288	-11.505
-20.6551	-0.2122	5.669964	-1.17421	-4.031087151	-12.597
2.351898	2.420797	-3.9859	-0.34836	3.14169233	-2.3406
23.85804	-0.13359	-11.9663	-9.81879	7.248909333	16.8466
-1.05741	55.72427	-0.02619	-0.18319	40.23894913	57.6088
13.26129	-0.14573	4.29096	14.22342	-6.406093906	-11.209
18.41776	0.959454	-12.2797	11.2919	-0.96628515	7.73194
2.179492	0.239524	-5.84595	2.525108	6.337222184	-3.3222
14.45577	0.015903	-6.00155	-3.15306	-11.12604868	9.82572
-1.46234	8.423085	-1.13291	9.109842	4.519436326	-13.047
-10.5605	-0.60898	-11.5425	-8.65932	-10.69002526	-14.205
-5.10282	0.794589	7.574859	7.359021	4.564505582	-8.1869
-1.15422	60.13577	-0.22776	-0.36699	-9.573842762	60.4054
-12.6539	0.004469	8.571127	21.60325	7.901664846	-21.219
4.248925	0.215587	4.960692	3.804345	-15.4187677	7.04038
5.833325	0.304794	8.405821	4.945745	7.095032646	11.6589
-3.61126	0.172231	2.597087	11.71093	-13.32998286	-11.69

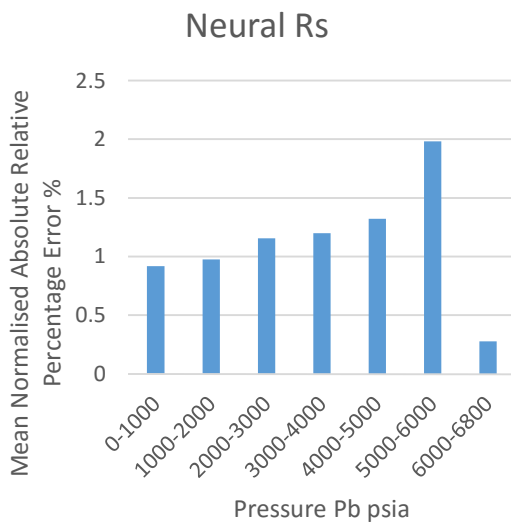
Εύρος PVT ιδιοτήτων του νευρωνικού Rs			
Ιδιότητες	Ελάχιστο	Μέγιστο	Μέσο
Rs (scf/STB)	0.5	3991.1	729.9
Pb (psia)	19.9	6773	1235
API	13.3	48.5	33.1
SG	0.57	1.72	0.96
T(Fahrenheit)	85.8	376	204.9



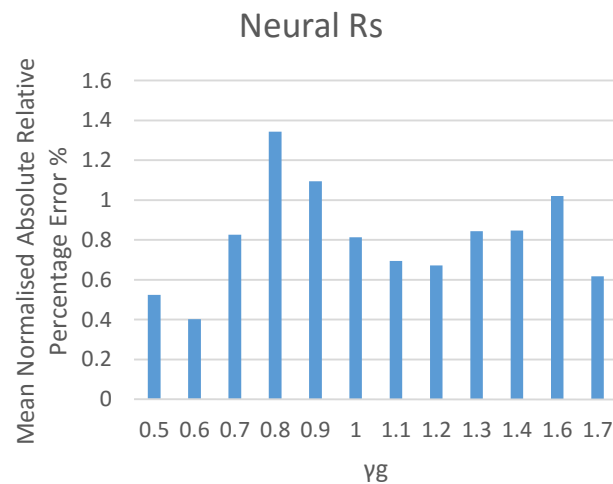
Σχήμα 6.19 Συσχέτιση κανονικοποιημένου σχετικού ποσοστιάου σφάλματος του νευρωνικού-Rs με τις πειραματικές μετρήσεις του composite GOR.



Σχήμα 6.20 Συσχέτιση του μέσου κανονικοποιημένου απολύτου σχετικού ποσοστιάου σφάλματος του νευρωνικού-Rs με το API.



Σχήμα 6.21 Συσχέτιση του κανονικοποιημένου απολύτου σχετικού ποσοστιάου σφάλματος του νευρωνικού-Rs με τη πίση στο σημείο φυσαλίδας(Pb).



Σχήμα 6.22 Συσχέτιση του κανονικοποιημένου απολύτου σχετικού ποσοστιάου σφάλματος του νευρωνικού-Rs με τη πυκνότητα αερίου γg.

ΚΕΦΑΛΑΙΟ 7

ΣΥΓΚΕΝΤΡΩΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΕΠΙ ΤΩΝ ΕΚΤΙΜΗΣΕΩΝ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΙΔΙΟΤΗΤΩΝ B_o , R_s , p_b

I. Στατιστική ανάλυση των σφαλμάτων επί των εκτιμήσεων των τιμών των ιδιοτήτων B_o, P_b, R_s των εμπειρικών συσχετίσεων

Εν συνεχεία, παρουσιάζονται τα αποτελέσματα των στατιστικών κριτηρίων, όπως ορίστηκαν στο κεφάλαιο 6, επί των εκτιμήσεων των τιμών των ιδιοτήτων B_o , p_b , R_s των εμπειρικών συσχετίσεων για τα υποσύνολα των υπό-βάσεων που επιλέχθηκαν από τη συνολική βάση δεδομένων PVT του Εργαστηρίου Ανάλυσης Ρευστών και Πυρήνων του Πολυτεχνείου Κρήτης. Υπενθυμίζεται, ότι σε κάθε εμπειρική συσχέτιση χρησιμοποιήθηκαν μόνο οι αναφορές των πετρελαϊκών ρευστών της υπό-βάσης δεδομένων για κάθε ιδιότητα που τηρούσαν το εύρος τιμών όπως ορίζονταν από το συγγραφέα της κάθε συσχέτισης. Γι αυτό το λόγο στις γραφικές απεικονίσεις υπάρχει διαφορά στον αριθμό των δεδομένων που διατέθηκαν για την δοκιμή κάθε εμπειρικής συσχέτισης, όμως σε όλες τις περιπτώσεις θεωρείται ικανοποιητικός ώστε να χρησιμοποιηθούν για την αξιολόγηση και να εξαχθούν συμπεράσματα.

Παρουσιάζονται κατωτέρω αναλυτικά οι στατιστικοί πίνακες που απεικονίζουν την ακρίβεια των εμπειρικών συσχετίσεων για τις ιδιότητες του **B_o , P_b , R_s** καθώς επίσης και οι γραφικές απεικονίσεις υπό μορφή νέφους σημείων των ζευγών πραγματικής τιμής και εκτιμώμενης τιμής (real value, predicted value) καθώς και του σχετικού σφάλματος συναρτήσεως των πραγματικών τιμών (relative error %, real value) ώστε να γίνει αντιληπτή η συσχέτιση των εκτιμώμενων τιμών με τις πραγματικές καθώς και η κατανομή του σφάλματος.

Σημείωση : Για τη παραγωγή των διαγραμμάτων χρησιμοποιήθηκε το υπολογιστικό πρόγραμμα Matlab

Α. Για τη πρόβλεψη του Βο

Πίνακας 7.1 Στατιστική ανάλυση των αποτελεσμάτων των εμπειρικών εξισώσεων για την πρόβλεψη του Βο

Correlations	Er %	Ermax %	Ermin %	Ea %	STDEV(Er) %	68.20%		95.40%		99.70%		r
	μ					σ						
						$\mu - 1\sigma \leq x \leq \mu + 1\sigma$	$\mu - 2\sigma \leq x \leq \mu + 2\sigma$	$\mu - 3\sigma \leq x \leq \mu + 3\sigma$				
Standing (1947)	2.15	8.94	-2.75	2.50	2.36	-0.20	4.51	-2.56	6.87	-4.91	9.22	0.9935
Vazquez and Beggs(1980)	-2.67	6.59	-14.48	3.47	3.43	-6.10	0.77	-9.54	4.20	-12.97	7.64	0.9876
Glaso(1980)	-0.72	6.38	-9.02	2.23	2.74	-3.46	2.02	-6.21	4.77	-8.95	7.51	0.9997
Al-Marhoun(1988)	-0.30	5.00	-6.61	1.67	2.12	-2.43	1.82	-4.55	3.94	-6.67	6.06	0.9898
Majeed & Salman (1988)	3.33	22.30	-8.77	4.25	4.94	-1.61	8.28	-6.56	13.22	-11.50	18.17	0.9845
Dokla & Osman (1992)	3.65	10.20	-2.27	3.76	2.42	1.23	6.07	-1.19	8.50	-3.61	10.92	0.9986
Farshad,Leblance,Garber & Osorio(1992)	1.23	9.61	-7.80	2.43	2.92	-1.69	4.14	-4.61	7.06	-7.52	9.98	0.9861
Macary&El-Batanony(1992)	9.34	19.02	3.99	9.34	3.27	6.07	12.61	2.80	15.88	-0.46	19.15	0.9714
Al-Marhoun(1992)	0.30	7.41	-9.52	1.85	2.45	-2.15	2.75	-4.60	5.20	-7.05	7.65	0.9939
Petrosky & Farshad (1993)	0.94	5.45	-2.97	1.50	1.73	-0.79	2.67	-2.53	4.41	-4.26	6.14	0.9898
Omar & Todd (1993)	-1.08	7.15	-9.65	2.49	3.03	-4.11	1.95	-7.14	4.98	-10.17	8.00	0.9889
Kartoatmodjo & Schmidt (1994)	0.04	5.89	-10.20	1.75	2.27	-2.23	2.31	-4.51	4.59	-6.78	6.86	0.9979
Almehaideb(1997)	-1.56	12.14	-21.08	3.48	4.66	-6.22	3.10	-10.88	7.76	-15.54	12.43	0.9790
Birol Dindoruk and Peter G.Christman(2001)	-0.33	7.62	-6.89	1.82	2.28	-2.60	1.95	-4.88	4.23	-7.16	6.51	0.9935
Elsharkawy(1996)	50.71	126.89	4.96	50.71	28.85	21.86	79.56	-6.99	108.40	-35.84	137.25	0.9891

Er % = Mean percent relative error – μέσο ποσοστιαίο σχετικό σφάλμα

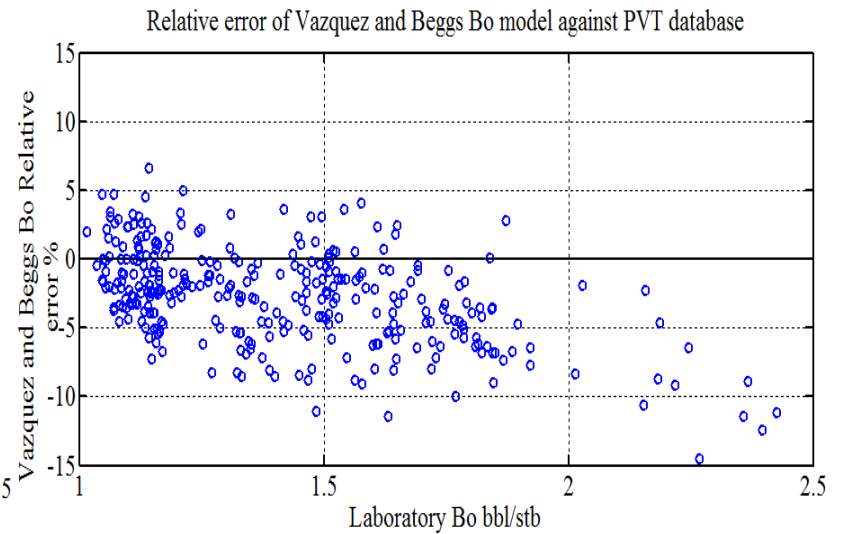
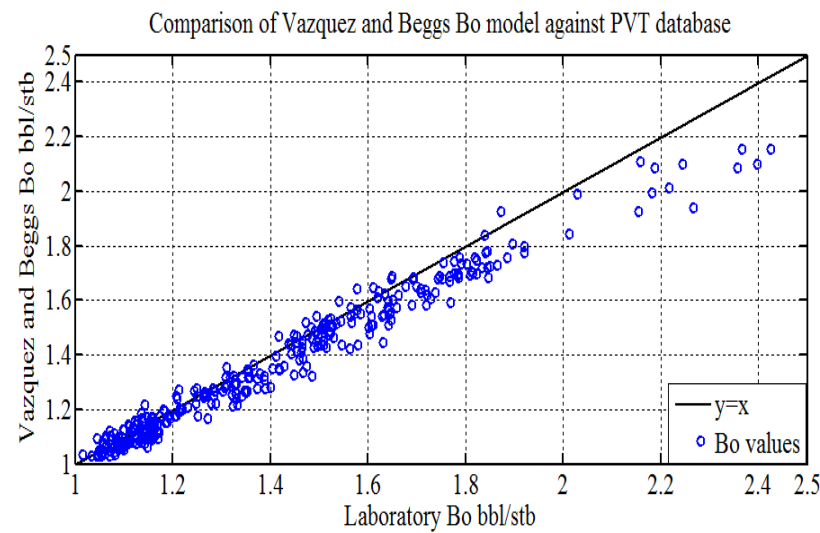
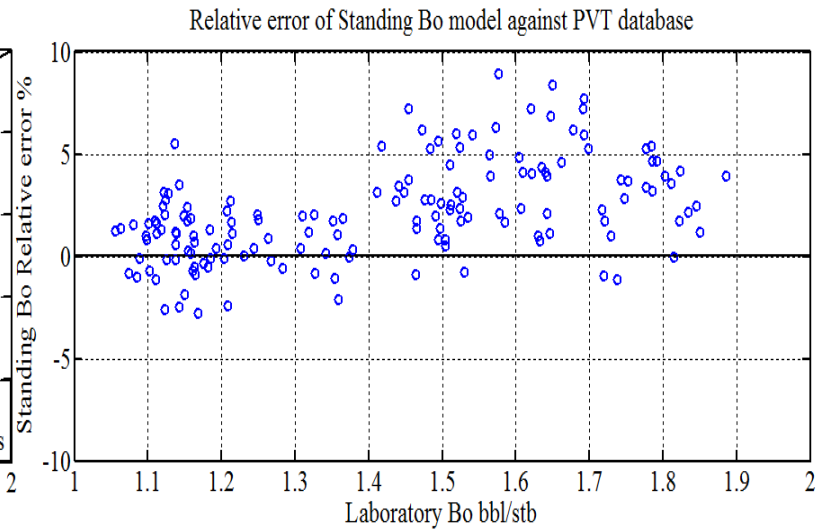
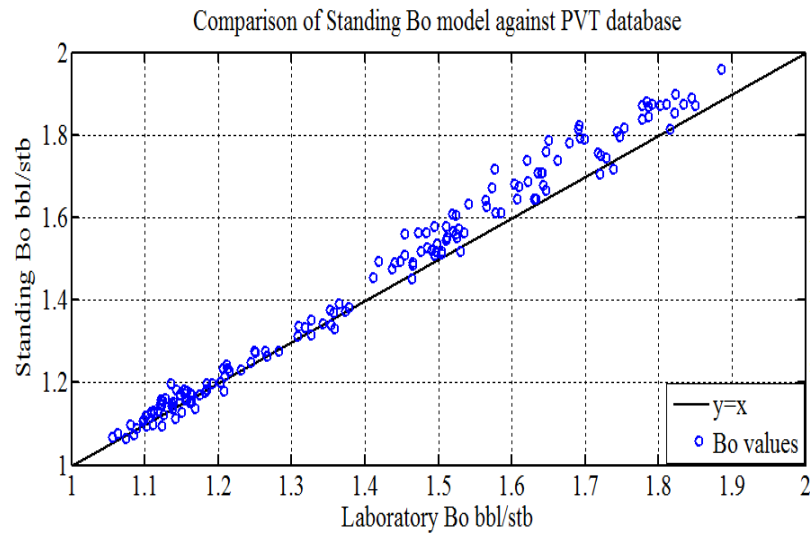
Ea % = Mean absolute percent relative error – μέσο απόλυτο ποσοστιαίο σφάλμα

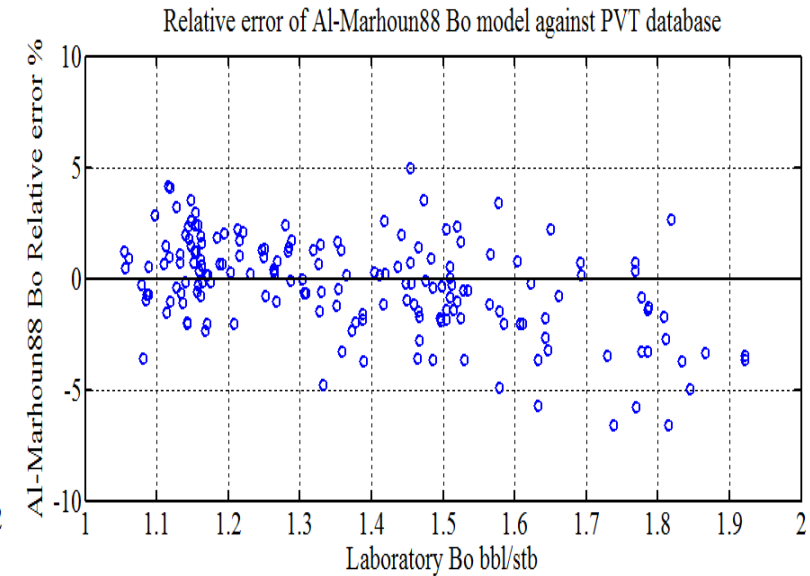
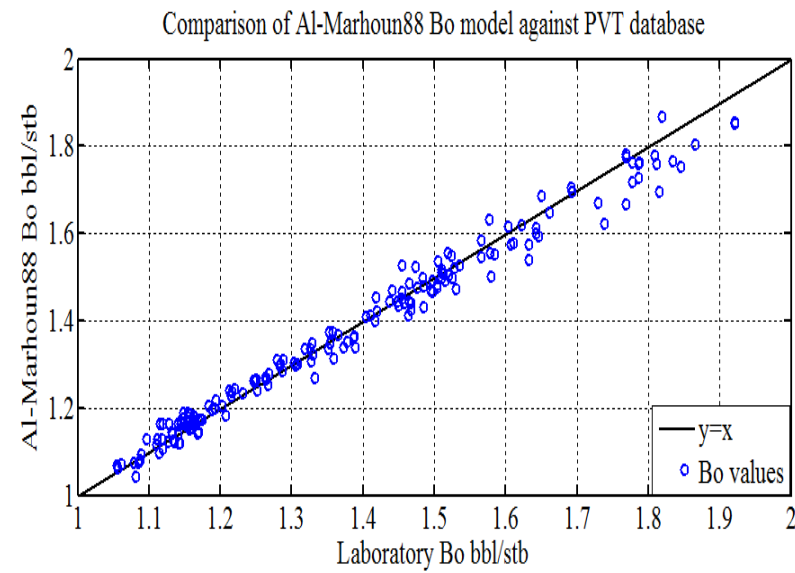
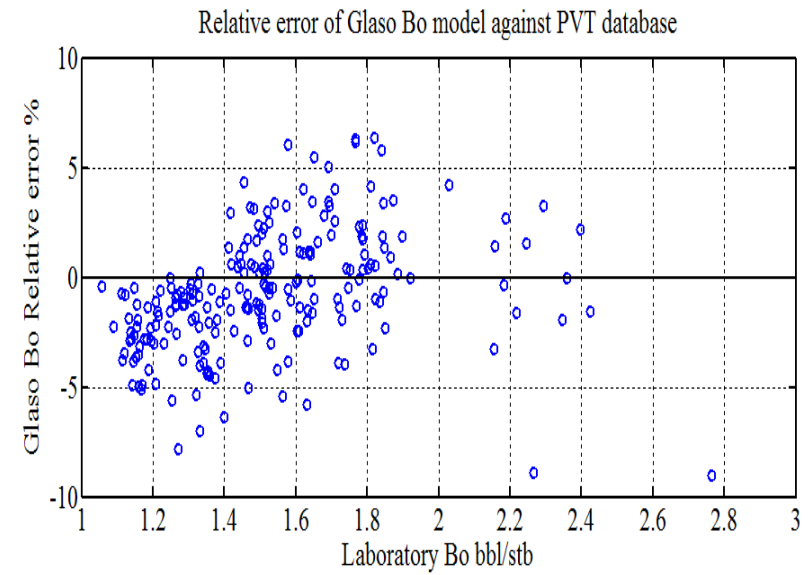
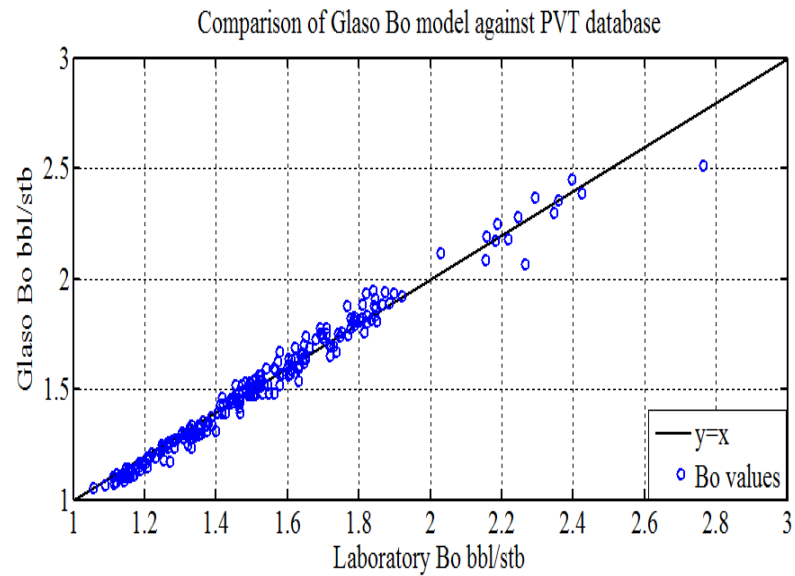
Ermax = Maximum relative percent error – μέγιστο ποσοστιαίο σχετικό σφάλμα

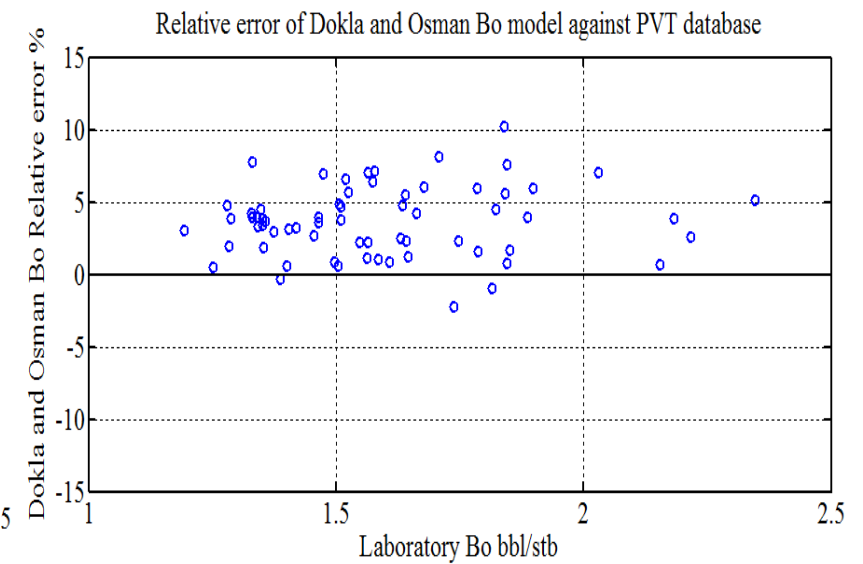
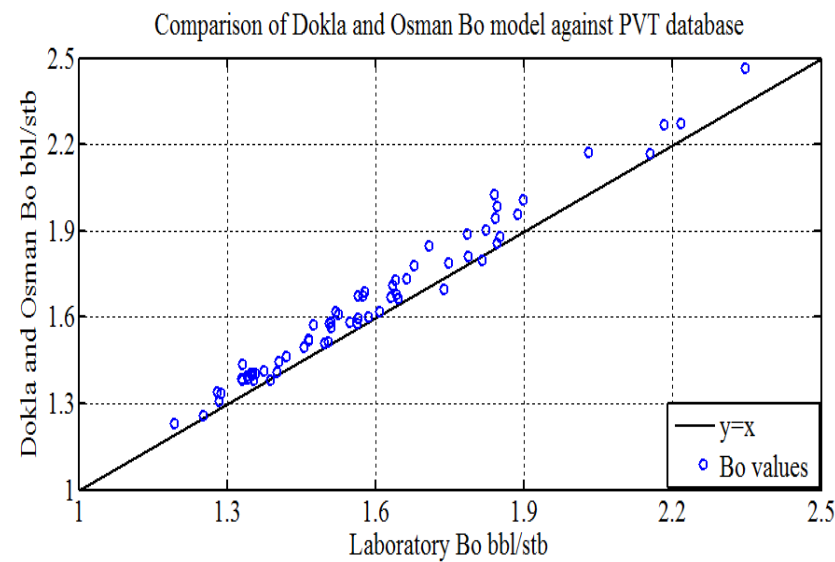
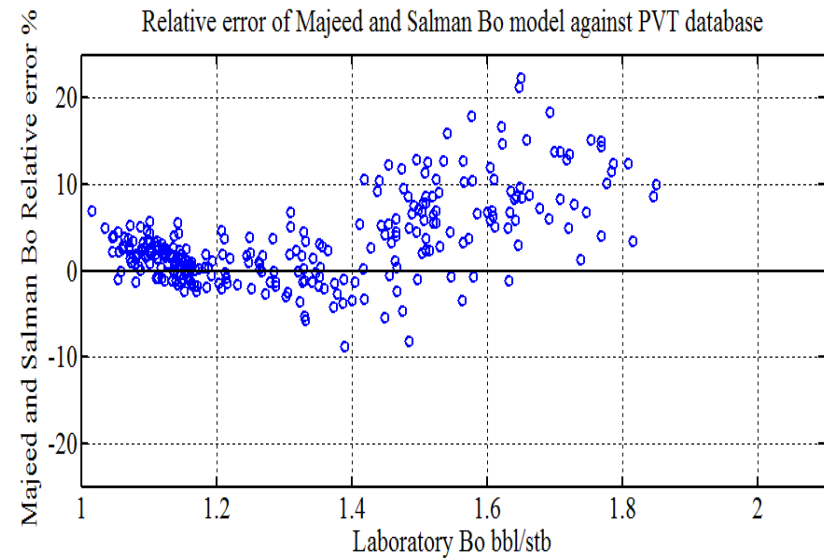
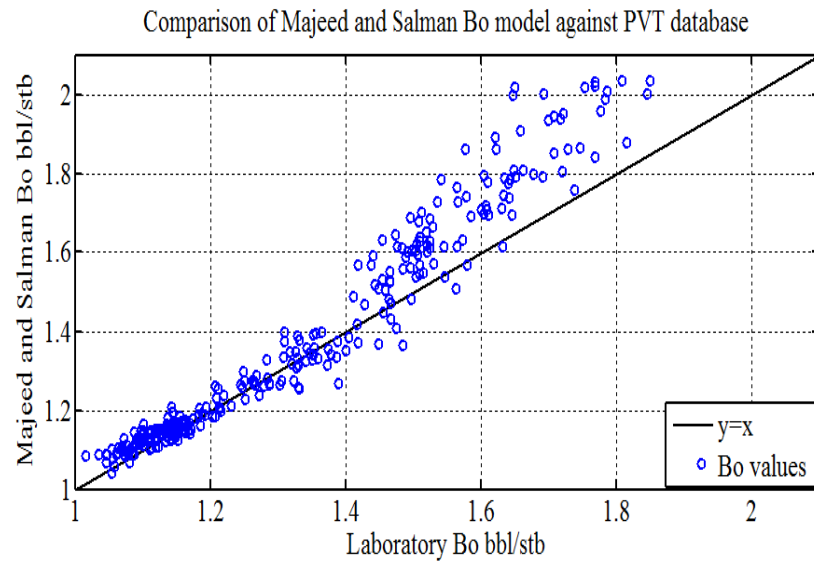
Ermin = Minimum relative percent error – ελάχιστο ποσοστιαίο σχετικό σφάλμα

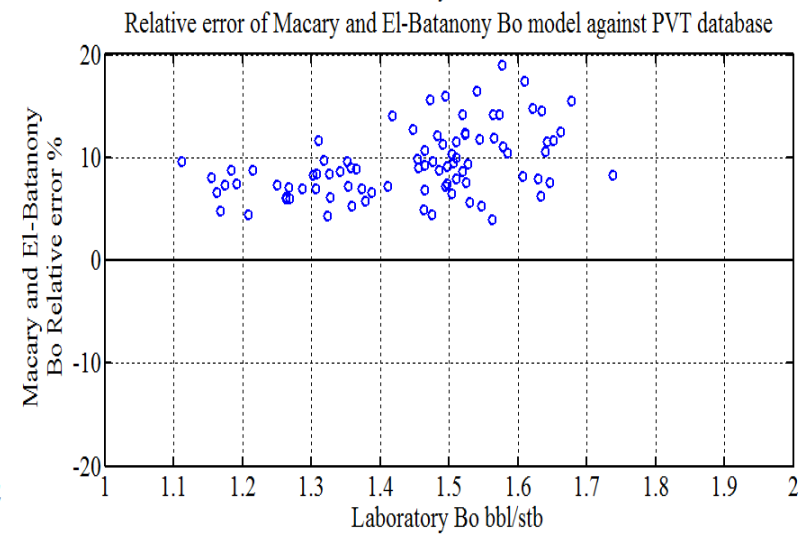
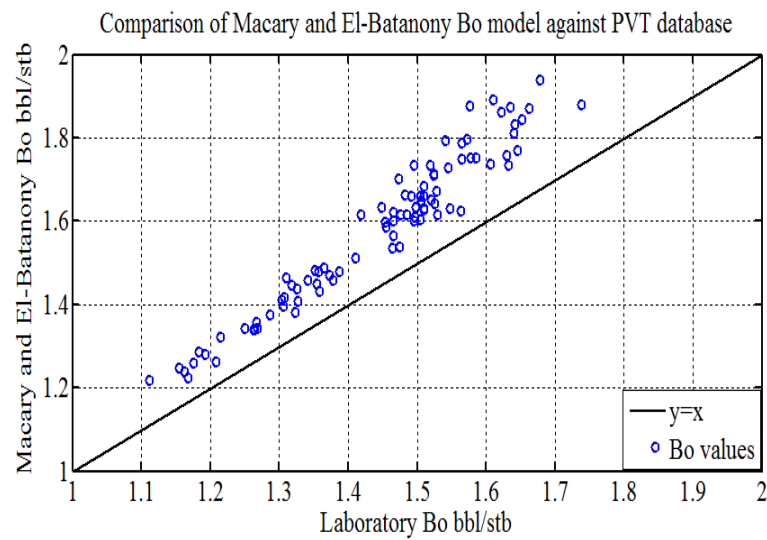
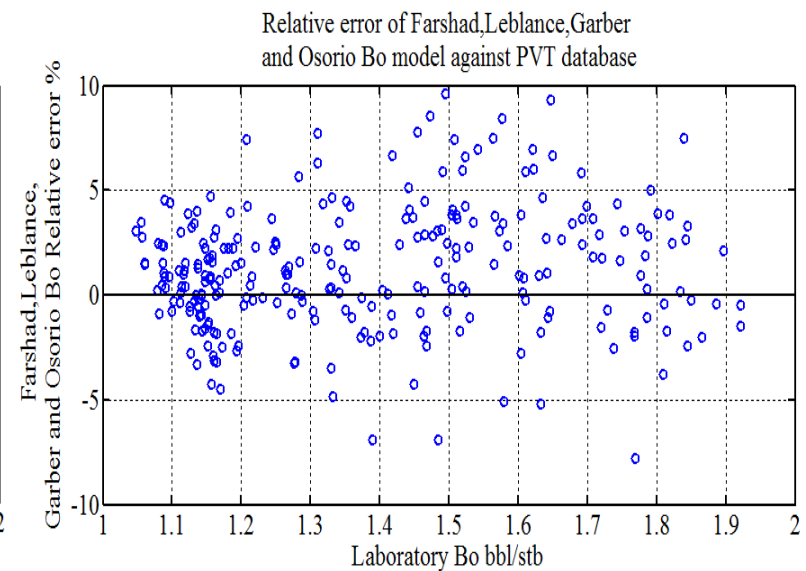
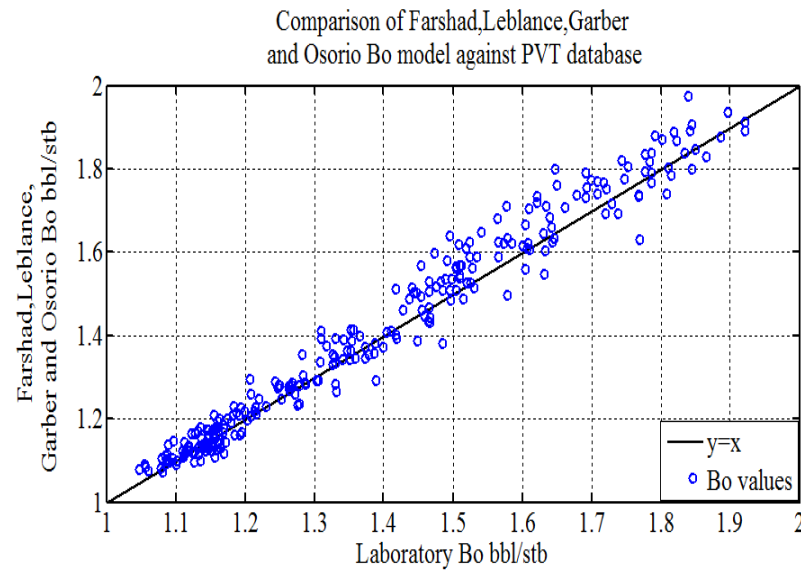
STDEV(Er) = standard Deviation of mean Relative percent error – τυπική απόκλιση του μέσου σχετικού ποσοστιαίου σφάλματος

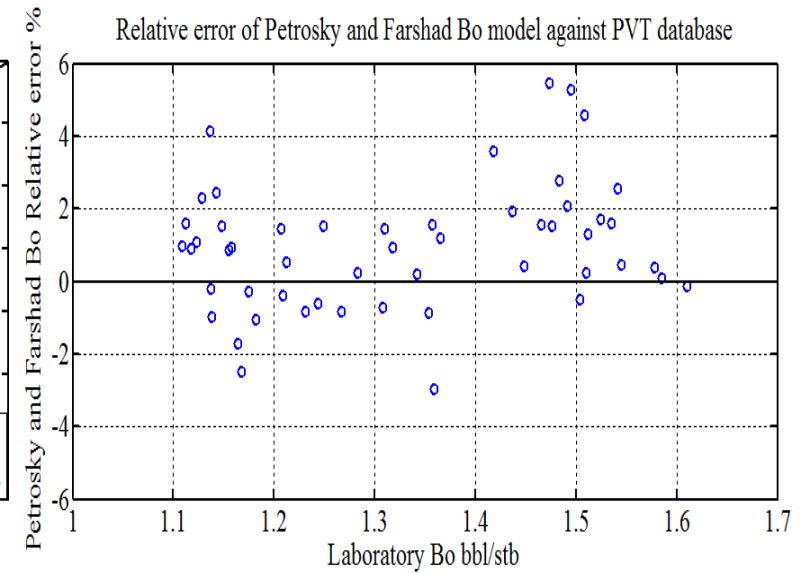
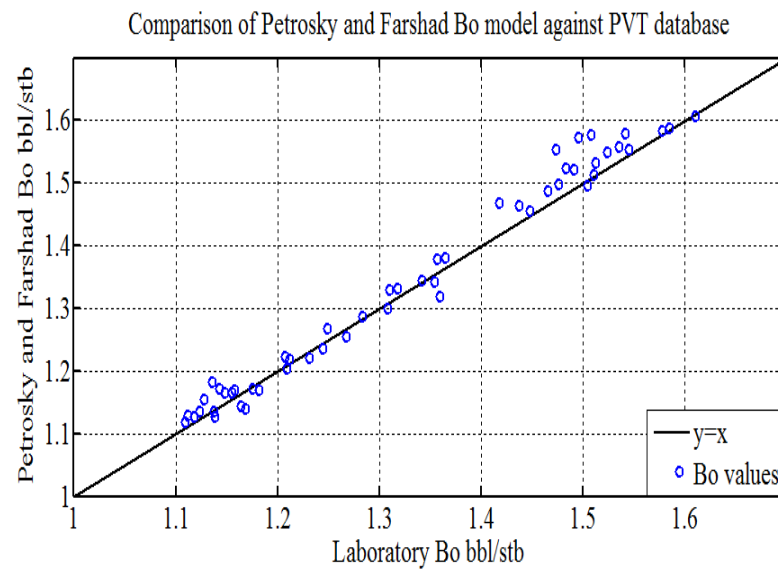
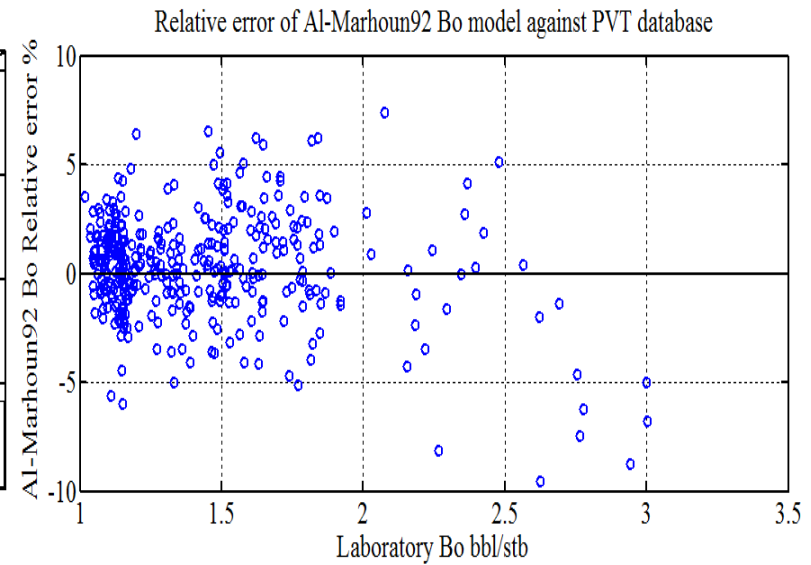
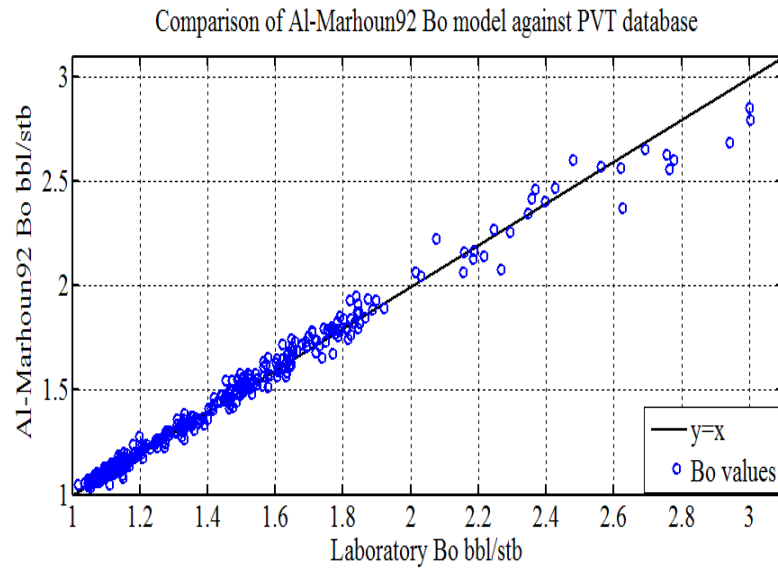
r = correlation coefficients pearson – συντελεστής συσχέτισης κατά pearson

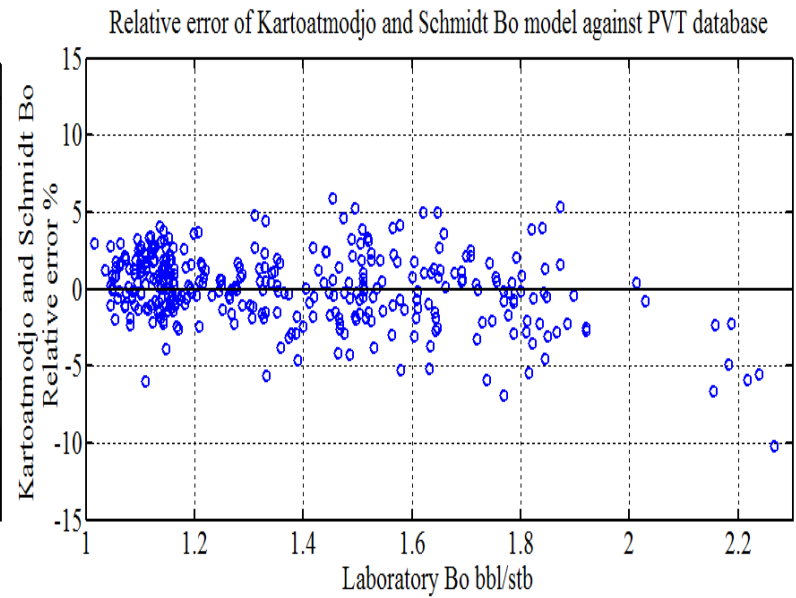
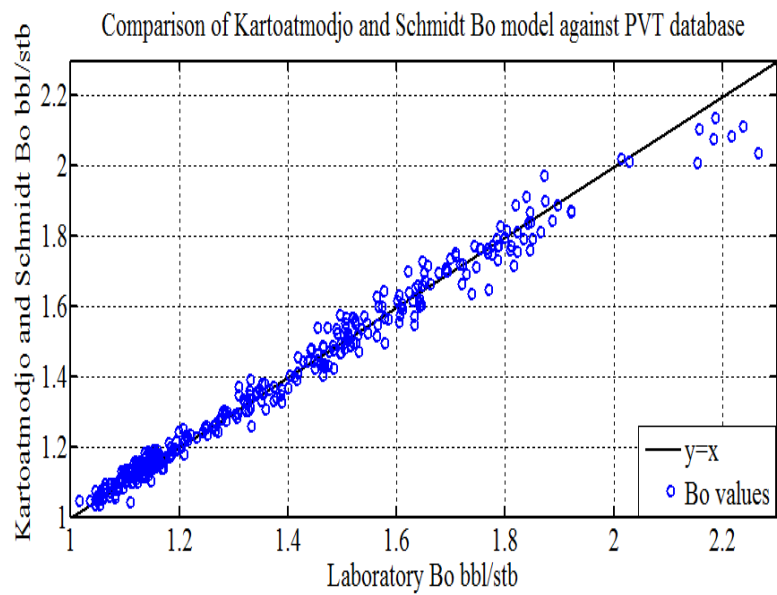
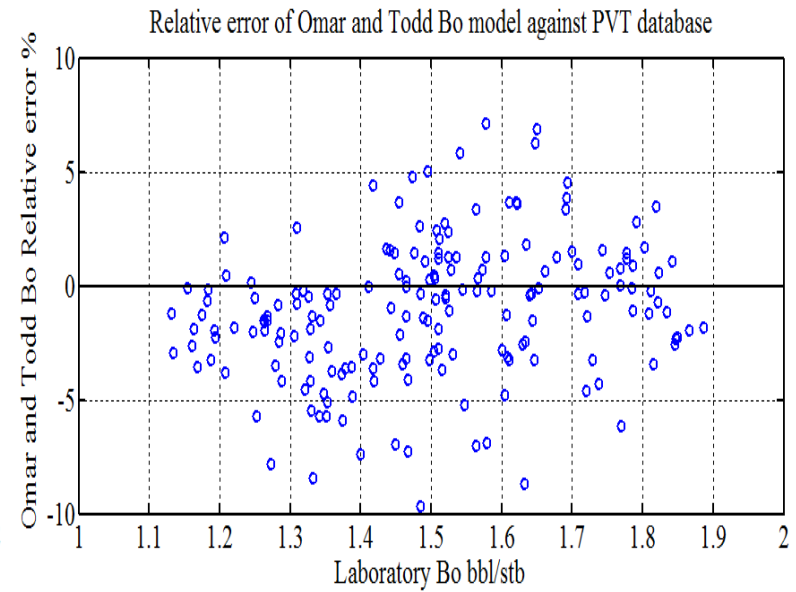
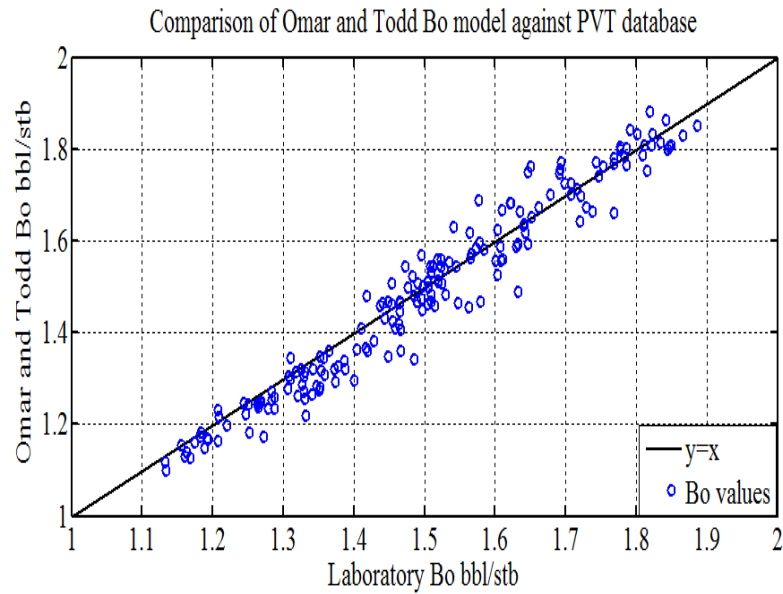


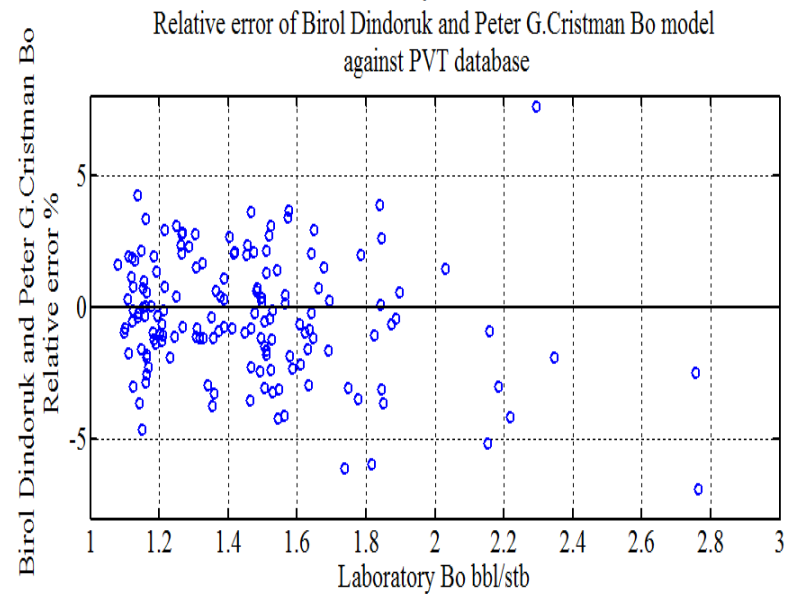
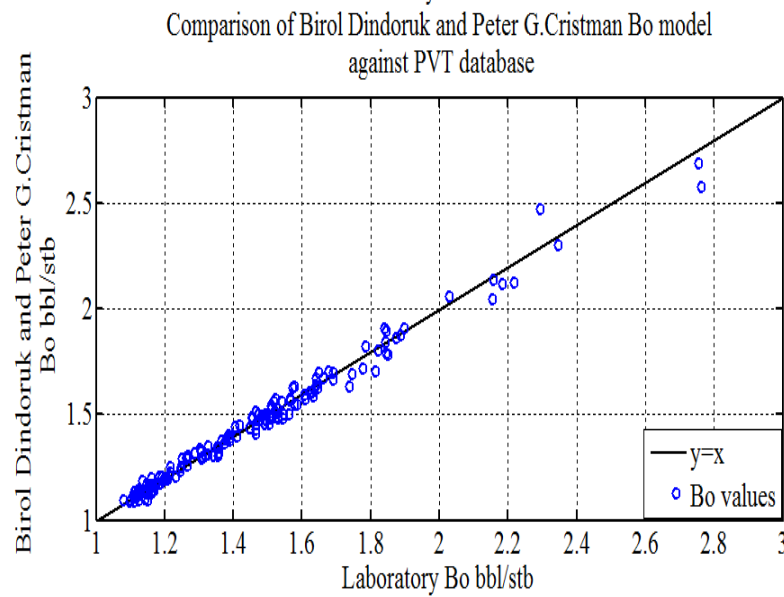
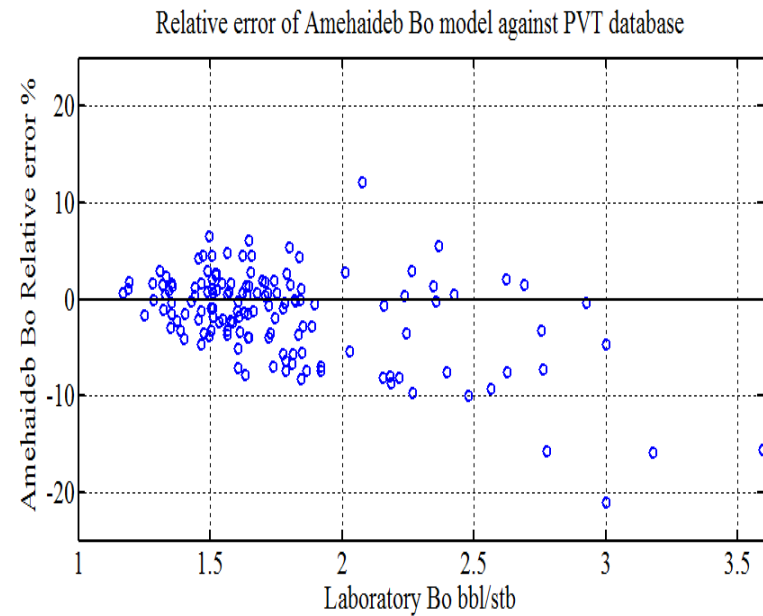
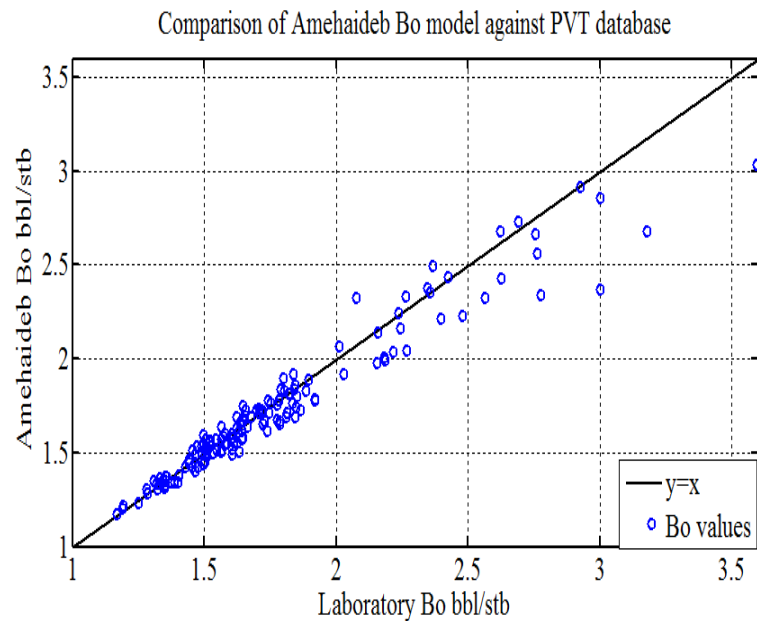


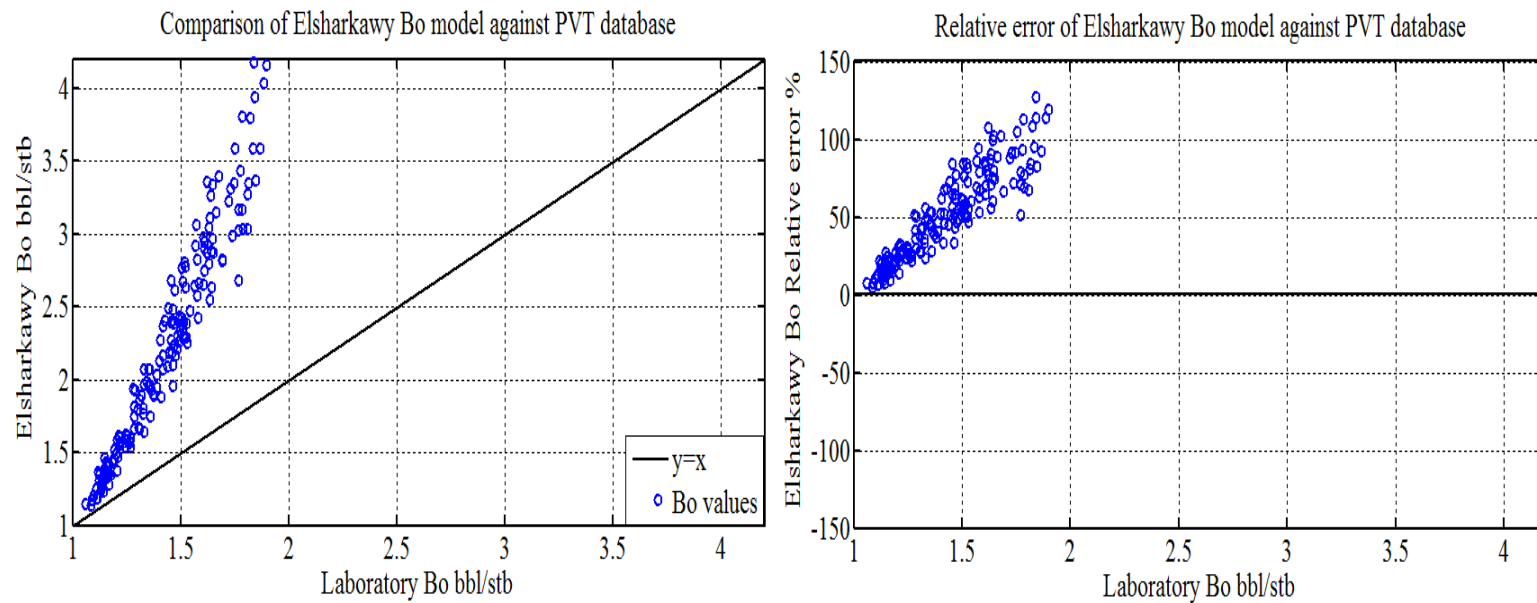












Όπως διαπιστώνεται από τις γραφικές απεικονίσεις αλλά και από τα στατιστικά δεδομένα, οι προβλέψεις των εμπειρικών εξισώσεων των Macary & El-Batanony και του Elsharkawy παρουσιάζουν μεγάλες στατιστικές αποκλίσεις από τα δεδομένα των πετρελαίων της βάσης δεδομένων που χρησιμοποιήθηκε. Για αυτό το λόγο, οι επιδόσεις τους δεν συγκρίθηκαν με αυτές το νευρωνικού που θα αναπτύχθηκε.

B. Για τη πρόβλεψη του p_b

Πίνακας 7.2 Στατιστική ανάλυση των αποτελεσμάτων των εμπειρικών εξισώσεων για την πρόβλεψη του p_b

Correlations	Er %	Ermax %	Ermin %	Ea %	STDEV(Er) %	68.20%		95.40%		99.70%		r
	μ				σ	μ - 1σ ≤ x ≤ μ + 1σ		μ - 2σ ≤ x ≤ μ + 2σ		μ - 3σ ≤ x ≤ μ + 3σ		
	Standing (1947)	-3.48	27.59	-44.44	9.57	12.18	-15.67	8.70	-27.85	20.89	-40.03	
Vazquez and Beggs(1980)	6.74	74.02	-69.11	16.18	20.77	-14.04	27.51	-34.81	48.28	-55.58	69.05	0.999
Glaso(1980)	10.49	86.49	-57.13	17.01	19.94	-9.45	30.42	-29.39	50.36	-49.32	70.29	0.993
Al-Marhoun(1988)	-10.38	17.86	-73.41	12.78	13.85	-24.23	3.47	-38.08	17.31	-51.93	31.16	0.970
Iasater(1958)	-4.66	69.47	-63.77	14.91	19.08	-23.74	14.42	-42.82	33.50	-61.90	52.58	0.981
Dokla & Osman (1992)	-17.57	14.37	-45.54	18.84	12.09	-29.66	-5.48	-41.75	6.60	-53.83	18.69	0.983
Farshad,Leblance,Garber & Osorio(1992)	6.99	90.07	-58.64	17.97	23.42	-16.42	30.41	-39.84	53.83	-63.26	77.25	0.981
Macary &El-Batanony(1992)	3.35	40.38	-21.56	10.66	12.95	-9.60	16.29	-22.55	29.24	-35.50	42.19	0.966
Petrosky & Farshad (1993)	5.36	36.44	-18.48	8.43	9.87	-4.51	15.22	-14.38	25.09	-24.24	34.95	0.978
Kartoatmodjo & Schmidt (1994)	8.03	138.92	-74.57	21.14	28.04	-20.00	36.07	-48.04	64.10	-76.07	92.14	0.984
Birol Dindoruk and Peter G.Christman(2001)	-2.69	33.09	-25.58	8.93	11.05	-13.74	8.36	-24.80	19.41	-35.85	30.47	0.947
Velarde(1999)	-7.08	43.49	-75.36	12.00	14.68	-21.77	7.60	-36.45	22.28	-51.14	36.97	0.979

$Er\%$ = Mean percent relative error – μέσο ποσοστιαίο σχετικό σφάλμα

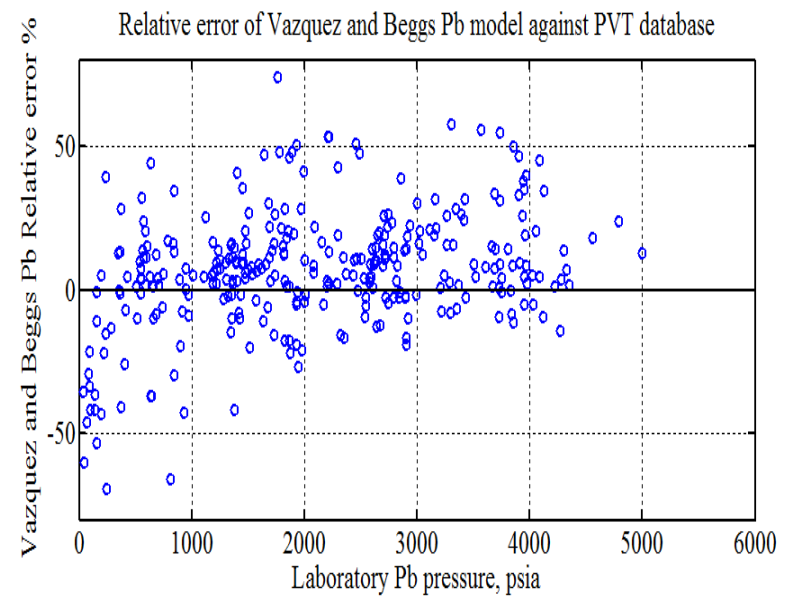
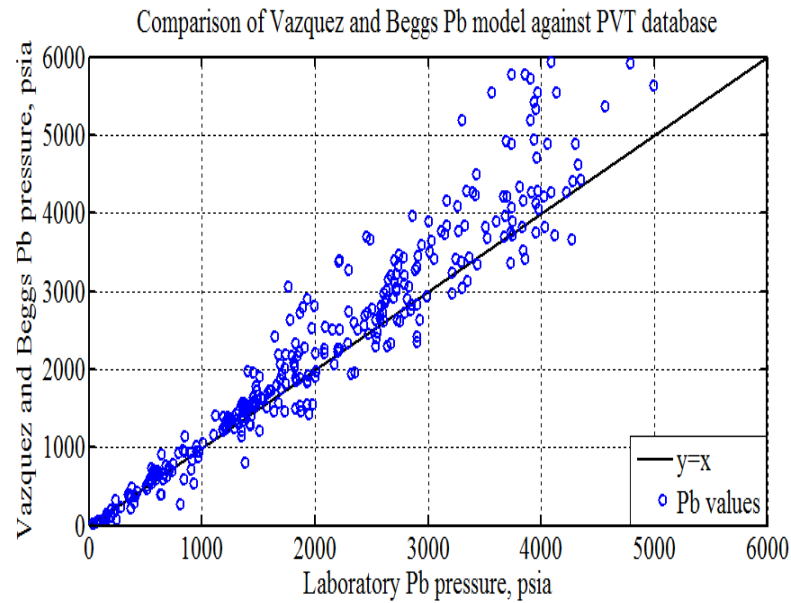
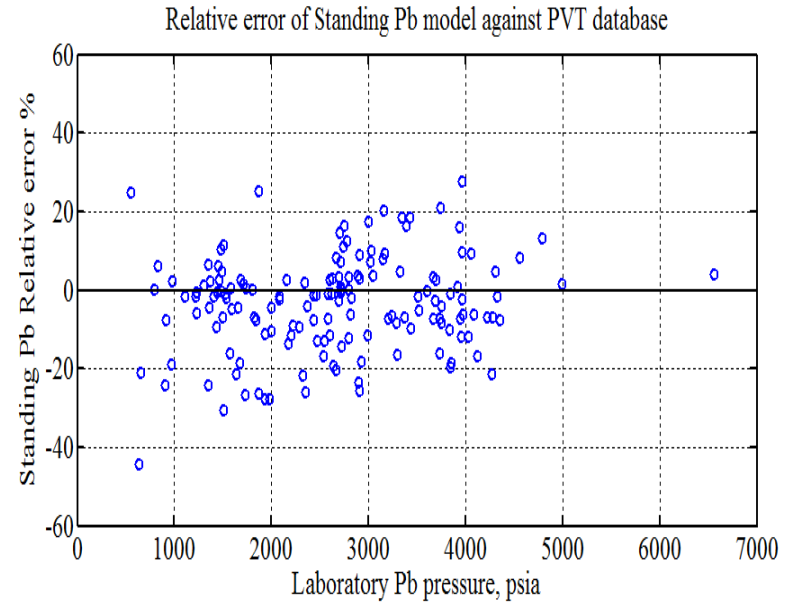
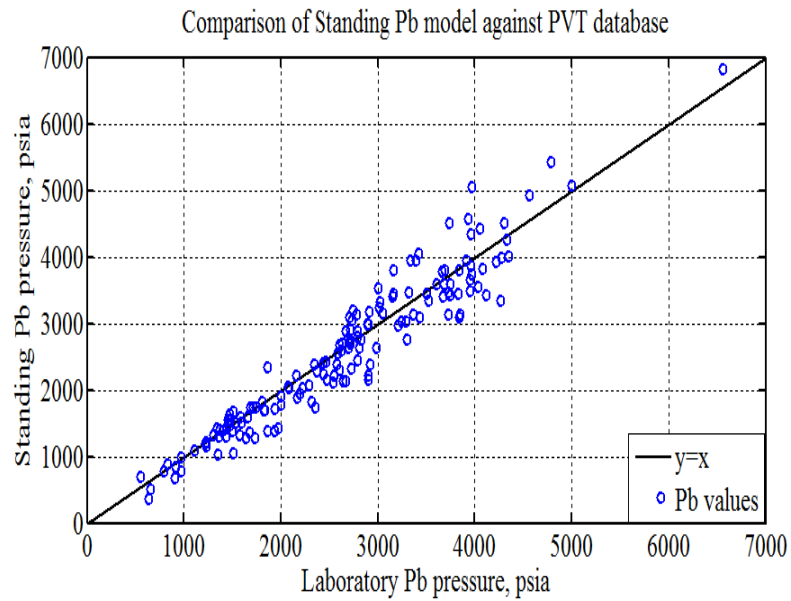
$Ea\%$ = Mean absolute percent relative error – μέσο απόλυτο ποσοστιαίο σφάλμα

$Ermax$ = Maximum relative percent error – μέγιστο ποσοστιαίο σχετικό σφάλμα

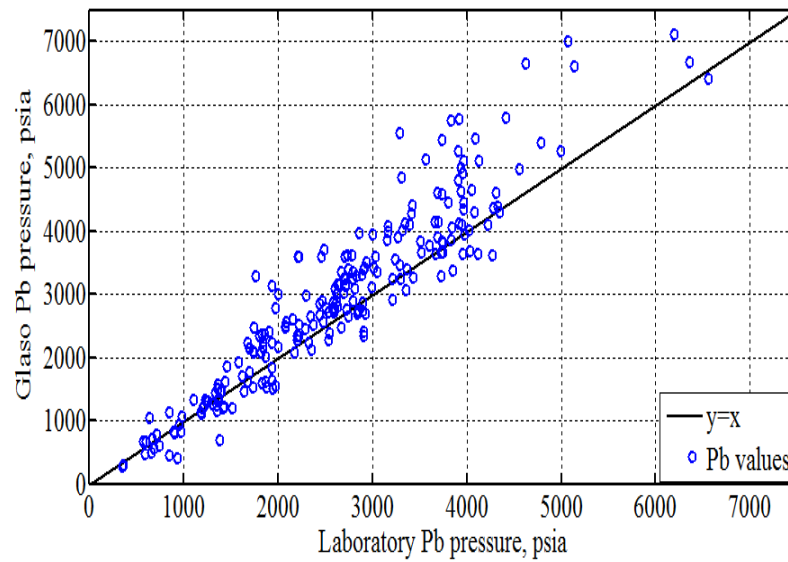
$Ermin$ = Minimum relative percent error – ελάχιστο ποσοστιαίο σχετικό σφάλμα

$STDEV(Er)$ = standard Deviation of mean Relative percent error – τυπική απόκλιση του μέσου σχετικού ποσοστιαίου σφάλματος

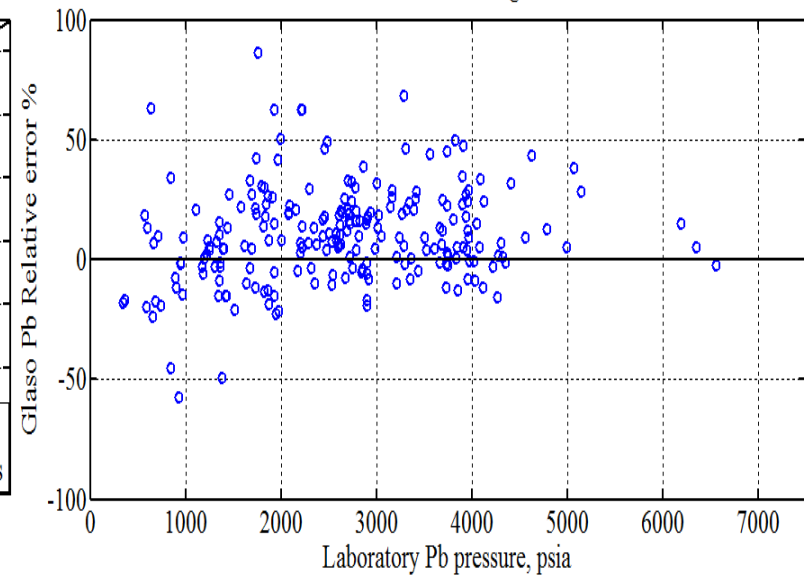
r = correlation coefficients pearson – συντελεστής συσχέτισης κατά pearson



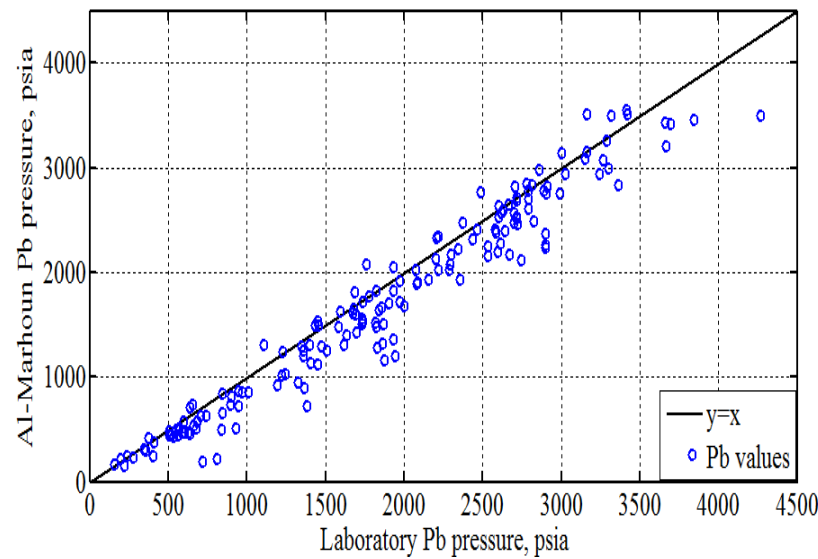
Comparison of Glaso Pb model against PVT database



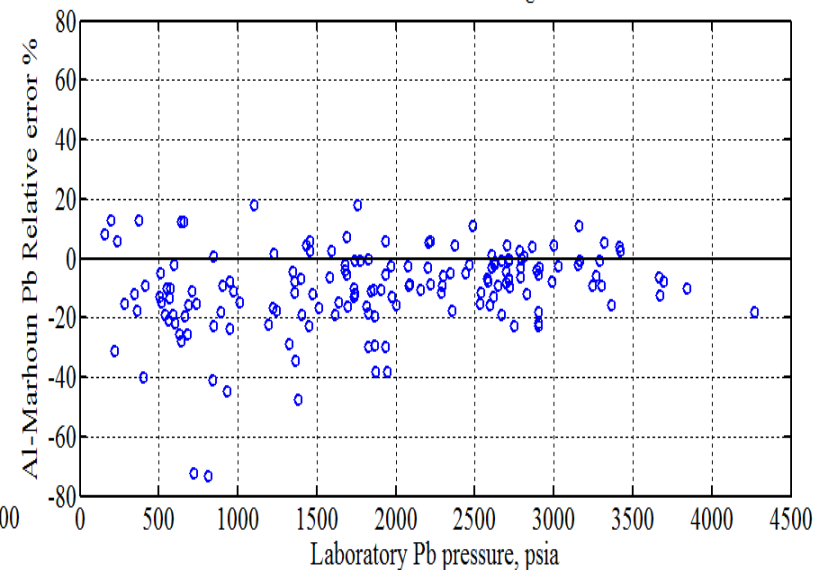
Relative error of Glaso Pb model against PVT database

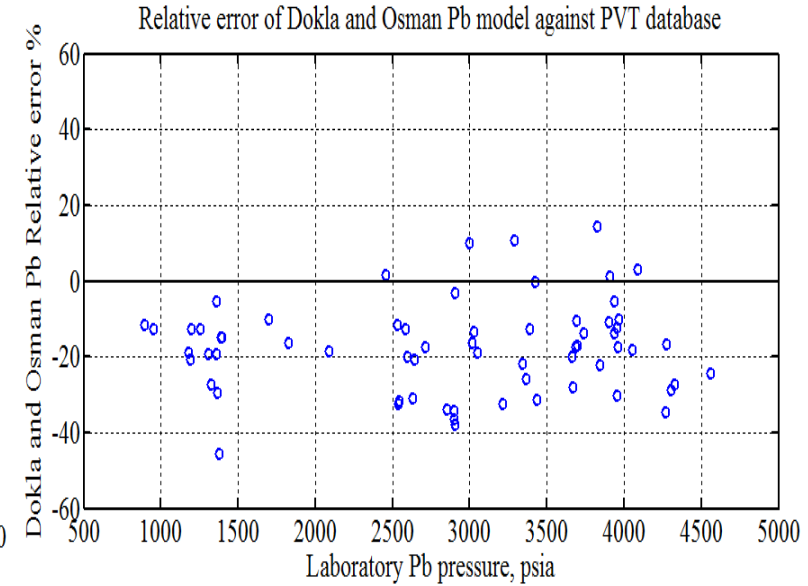
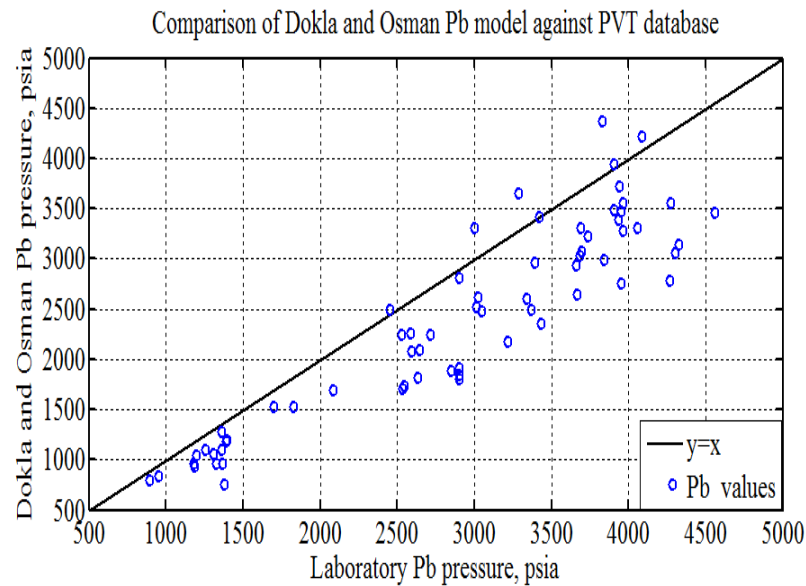
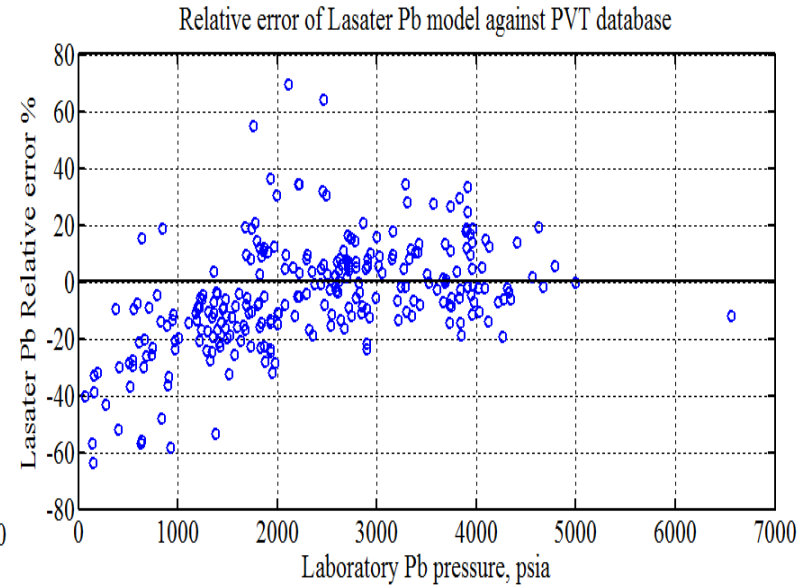
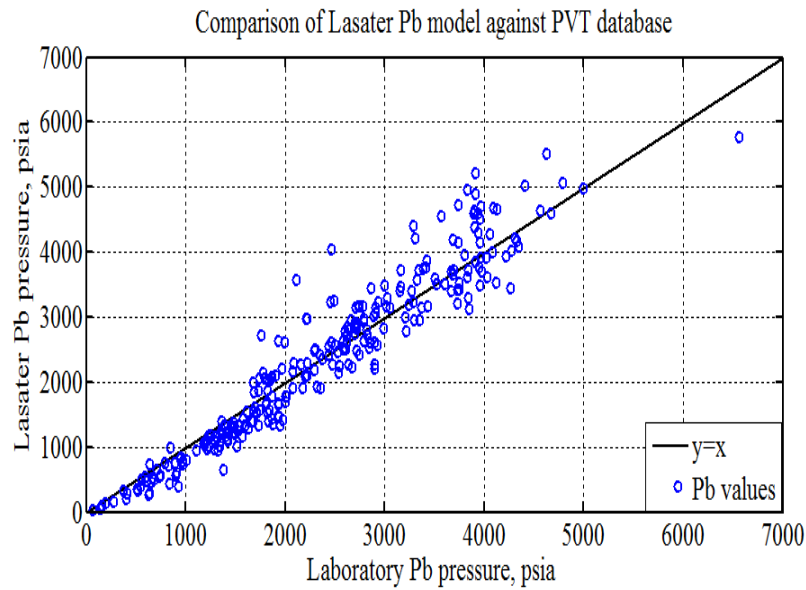


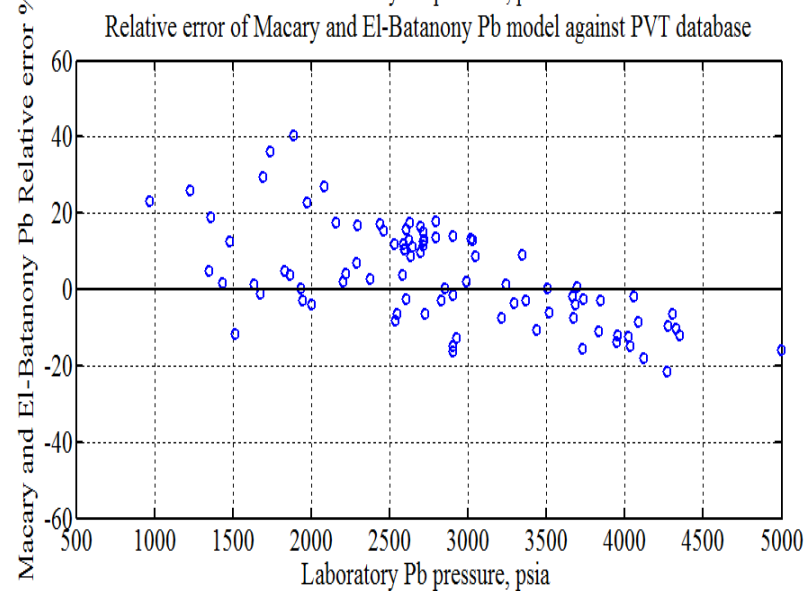
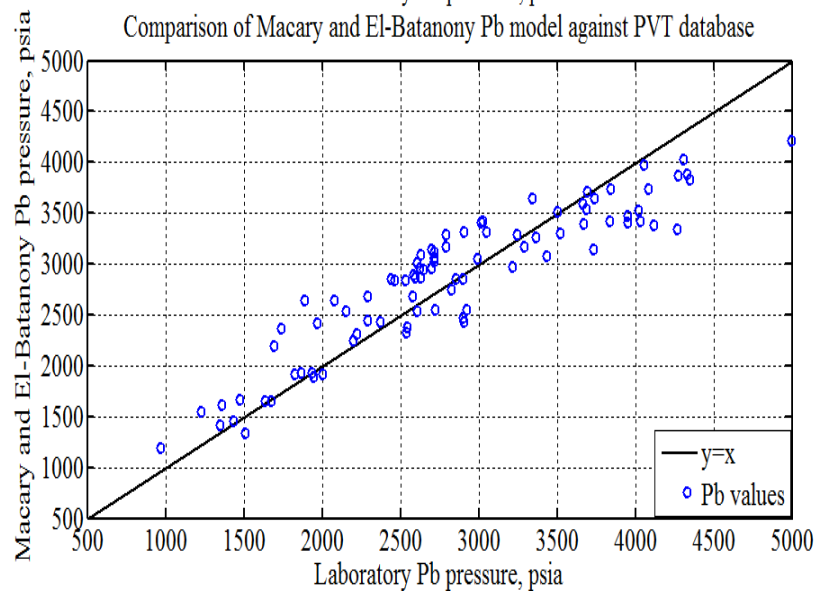
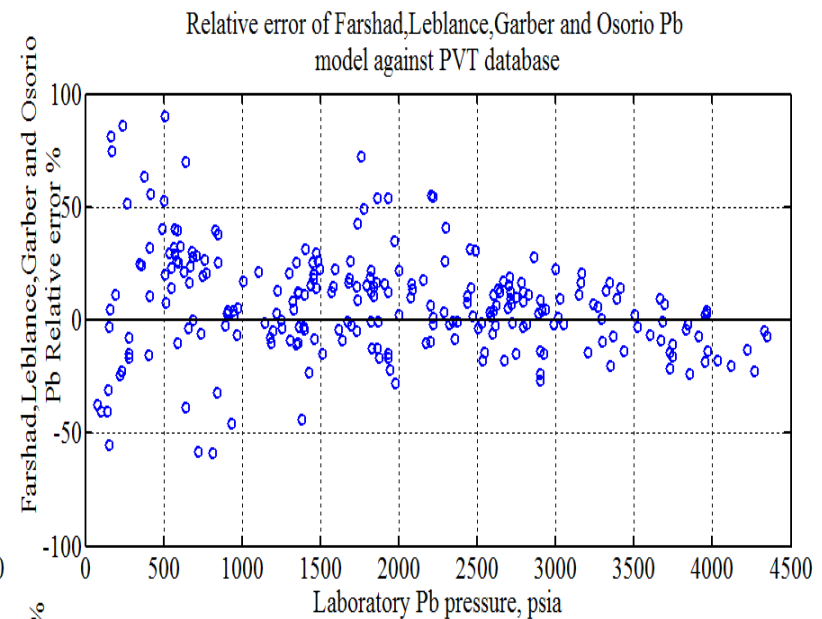
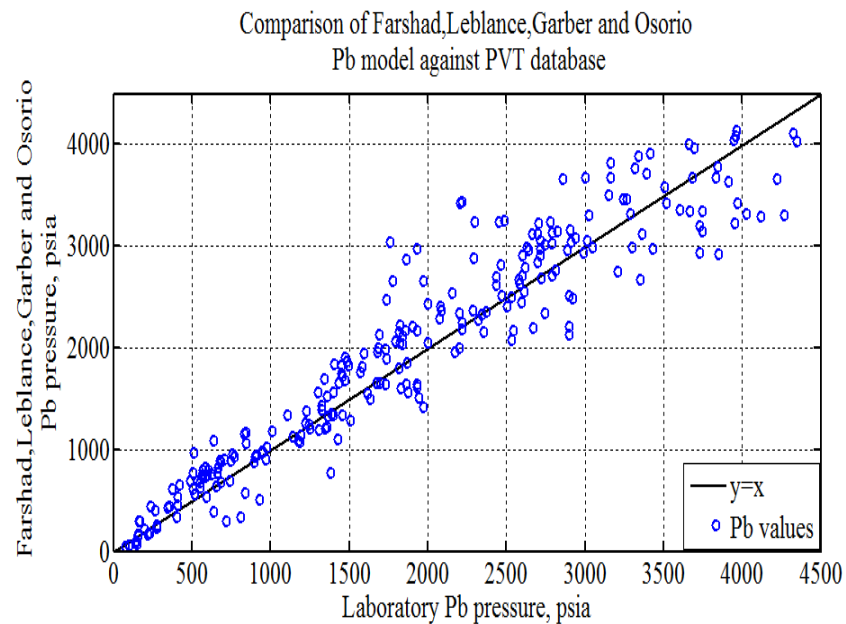
Comparison of Al-Marhoun Pb model against PVT database

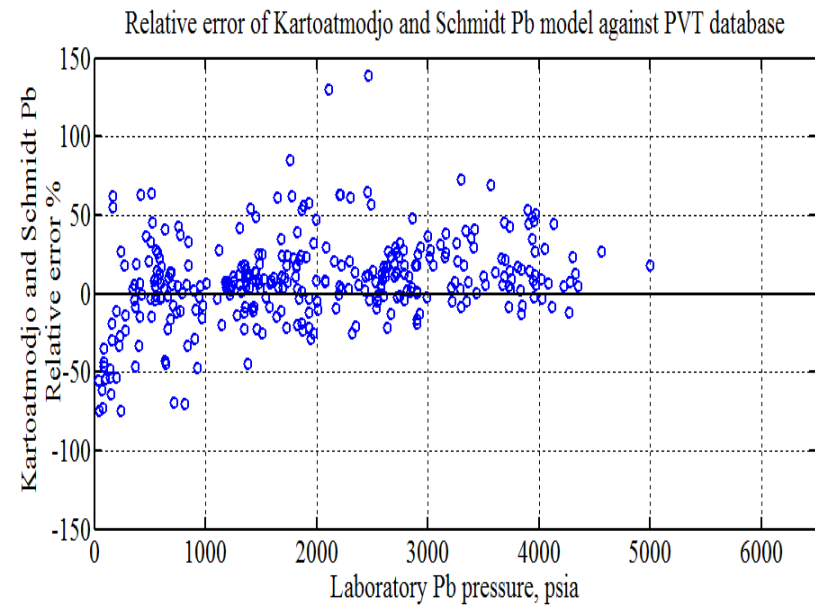
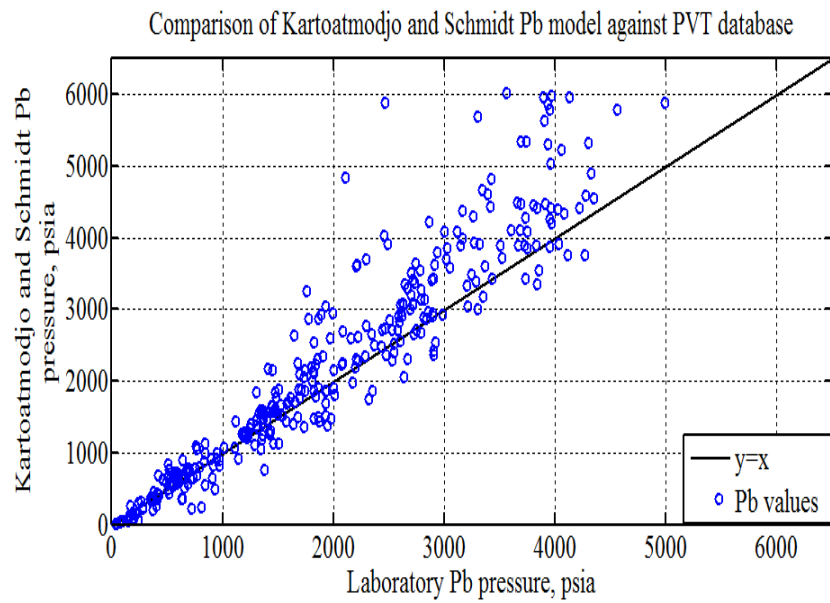
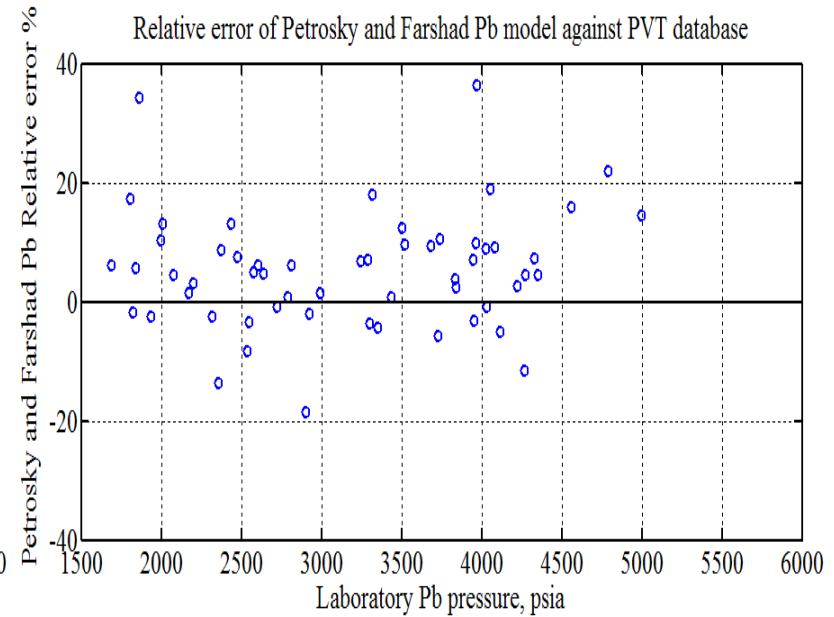
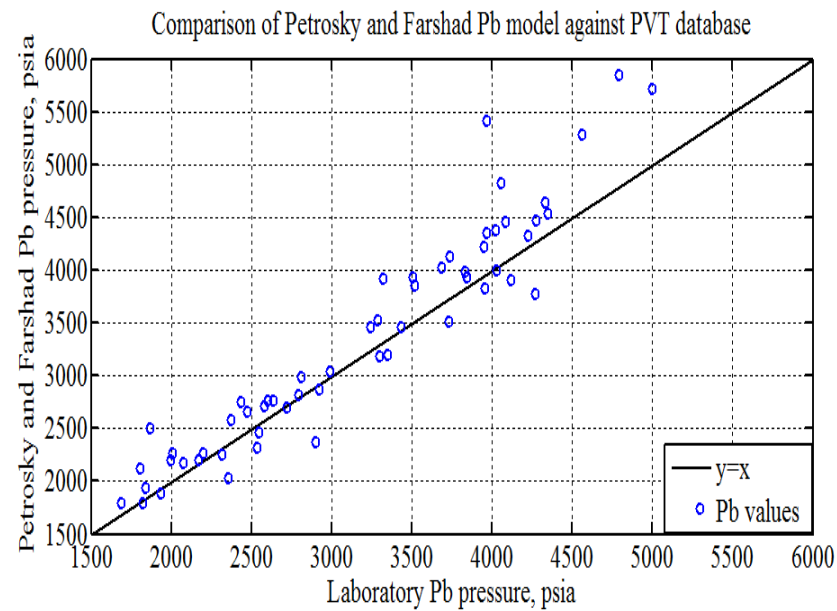


Relative error of Al-Marhoun Pb model against PVT database

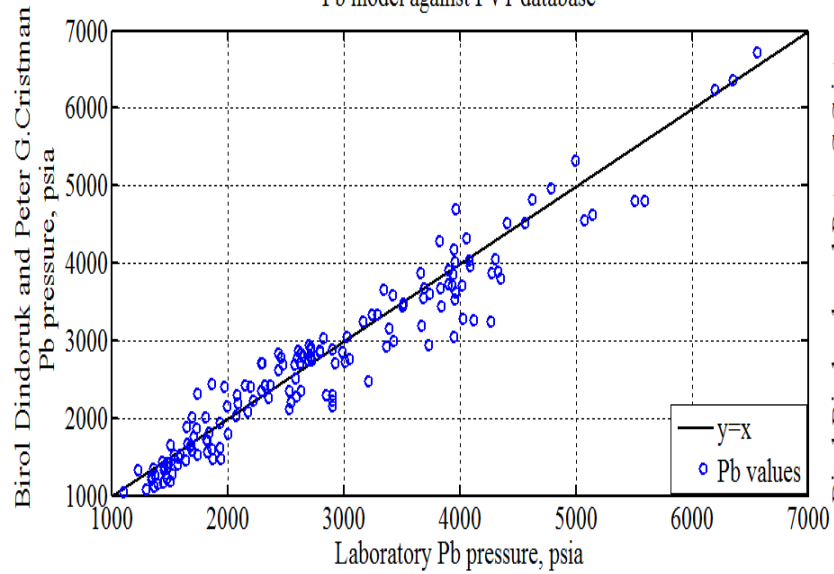




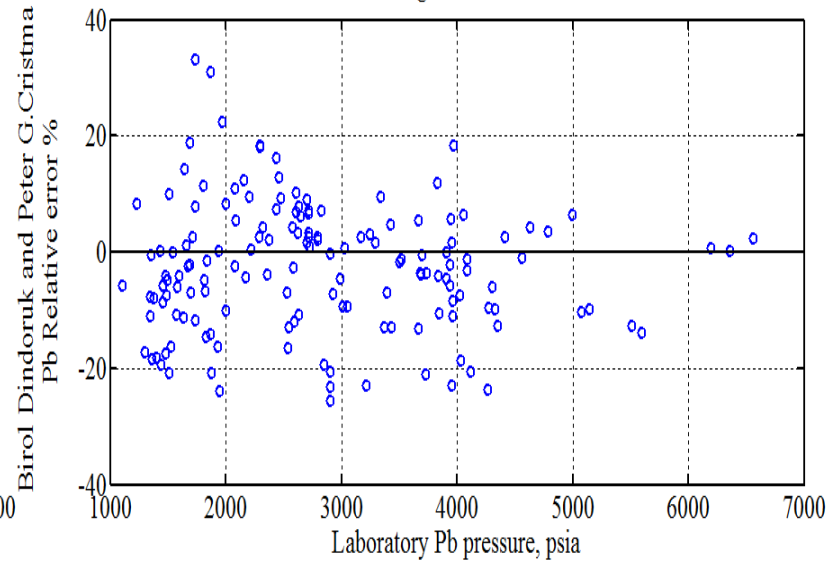




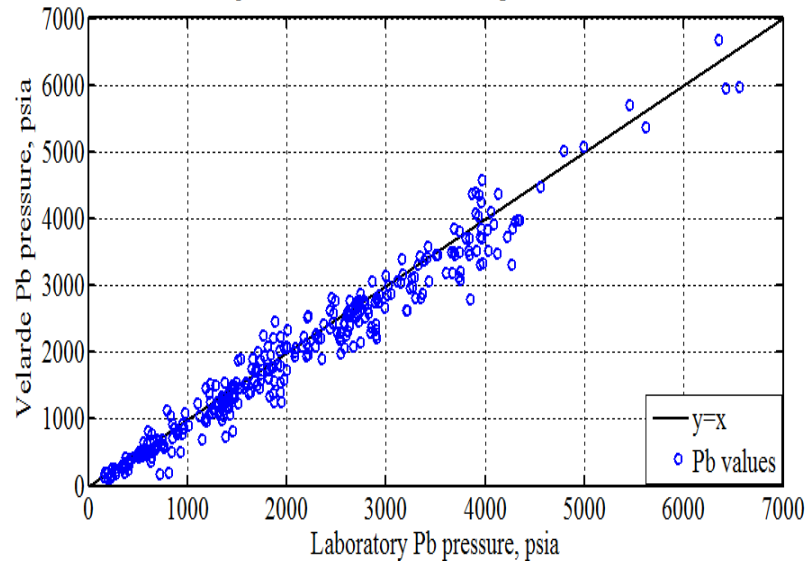
Comparison of Birol Dindoruk and Peter G.Cristman
Pb model against PVT database



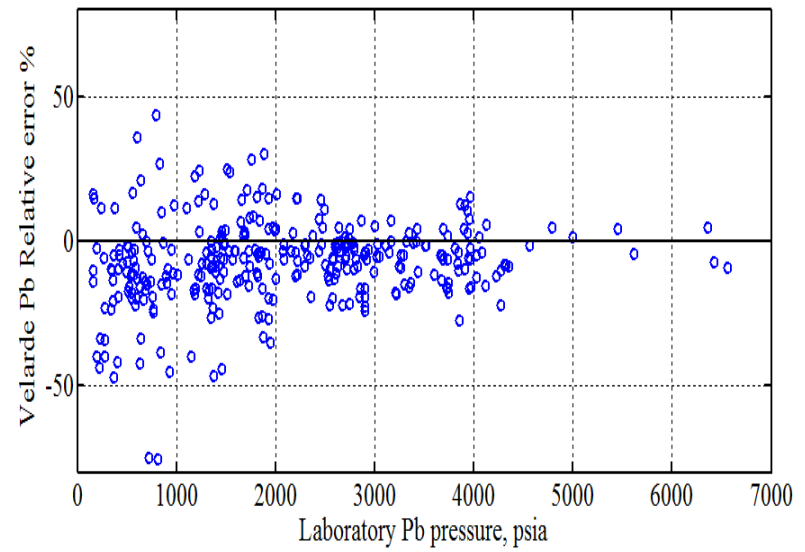
Relative error of Birol Dindoruk and Peter G.Cristman
Pb model against PVT database



Comparison of Velarde Pb model against PVT database



Relative error of Velarde Pb model against PVT database



Πίνακας 7.3 Στατιστική ανάλυση των αποτελεσμάτων των εμπειρικών εξισώσεων για την πρόβλεψη του R_s

Correlations	Ern % μ	Ernmax %	Ernmin %	Ean %	STDEV(Ern) % σ	68.20%		95.40%		99.70%		r
						μ − 1σ ≤ x ≤ μ + 1σ		μ − 2σ ≤ x ≤ μ + 2σ		μ − 3σ ≤ x ≤ μ + 3σ		
Standing (1947)	-1.69	16.98	-12.40	3.36	3.89	-5.58	2.19	-9.47	6.08	-13.36	9.97	0.962
Lasater(1958)	-2.36	31.30	-16.46	4.05	4.67	-7.03	2.32	-11.70	6.99	-16.38	11.66	0.992
Glaso(1980)	-2.80	33.24	-14.11	4.15	4.42	-7.22	1.63	-11.64	6.05	-16.07	10.48	0.990
Vazquez and Beggs(1980)	-3.48	60.50	-16.28	4.45	4.59	-8.07	1.11	-12.66	5.70	-17.25	10.29	0.995
Al-Marhoun(1988)	-2.35	90.57	-16.46	5.41	7.93	-10.29	5.58	-18.22	13.51	-26.15	21.44	0.979
Farshad,Leblance,Garber & Osorio(1992)	-11.90	4.44	-18.21	11.91	3.12	-15.02	-8.78	-18.14	-5.67	-21.26	-2.55	0.901
Petrosky & Farshad (1993)	-0.75	57.55	-9.33	2.66	4.79	-5.55	4.04	-10.34	8.84	-15.14	13.63	0.991
Kartoatmodjo & Schmidt (1994)	-2.73	61.36	-15.69	4.27	5.05	-7.78	2.32	-12.84	7.38	-17.89	12.43	0.965
Birol Dindoruk and Peter G.Christman(2001)	8.29	62.29	-9.94	9.84	11.25	-2.96	19.54	-14.21	30.79	-25.46	42.04	0.868

Ern % = Mean normalized percent relative error – μέσο κανονικοποιημένο ποσοστιαίο σχετικό σφάλμα

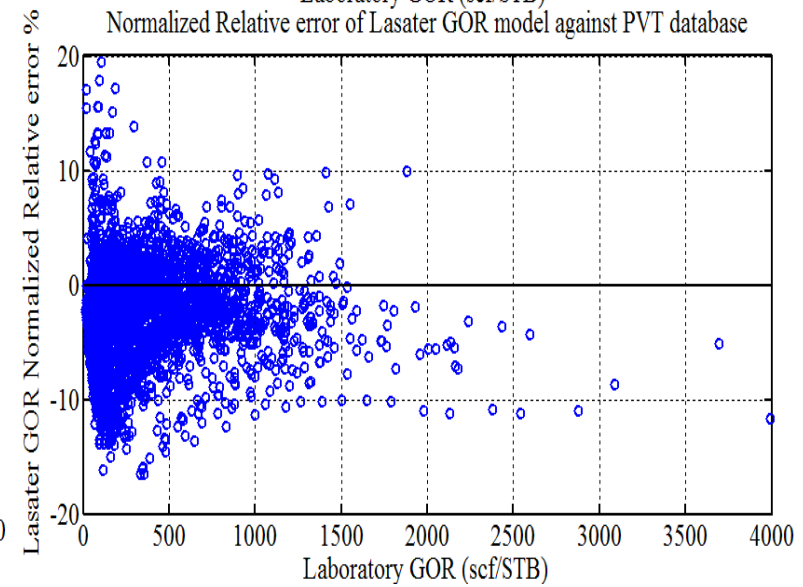
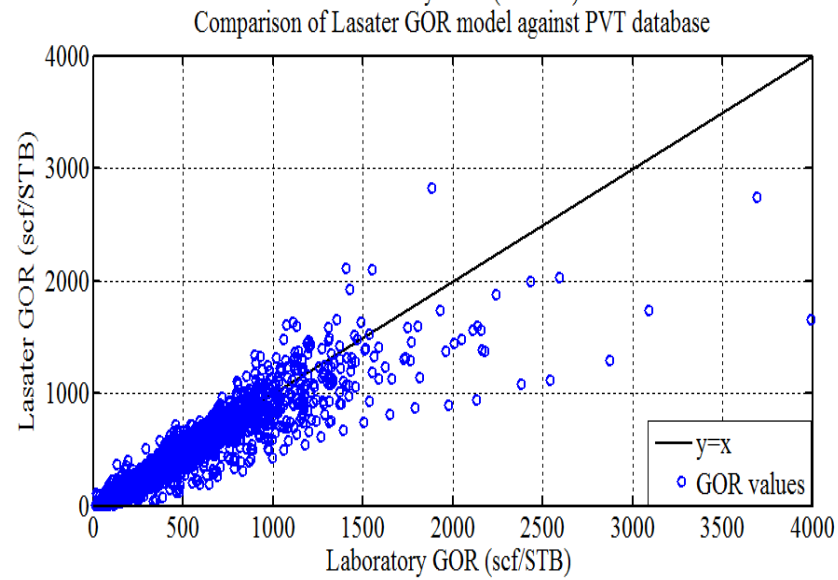
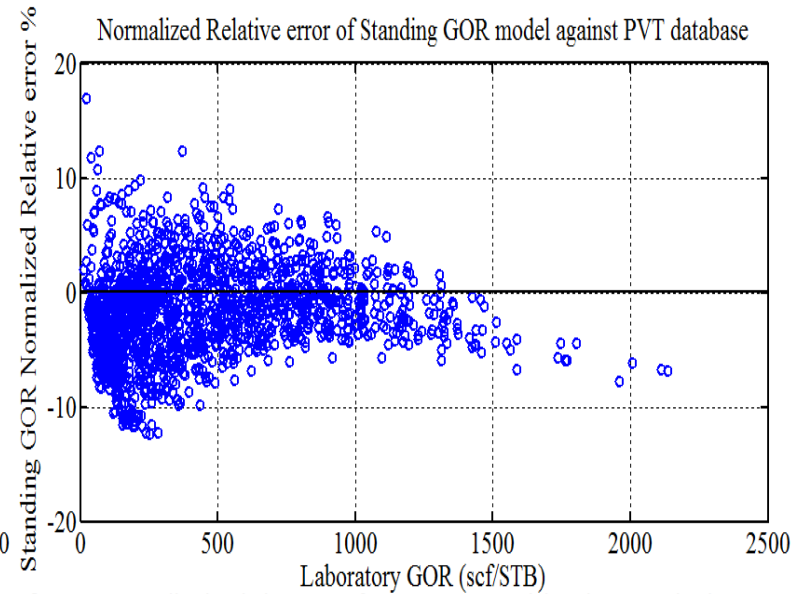
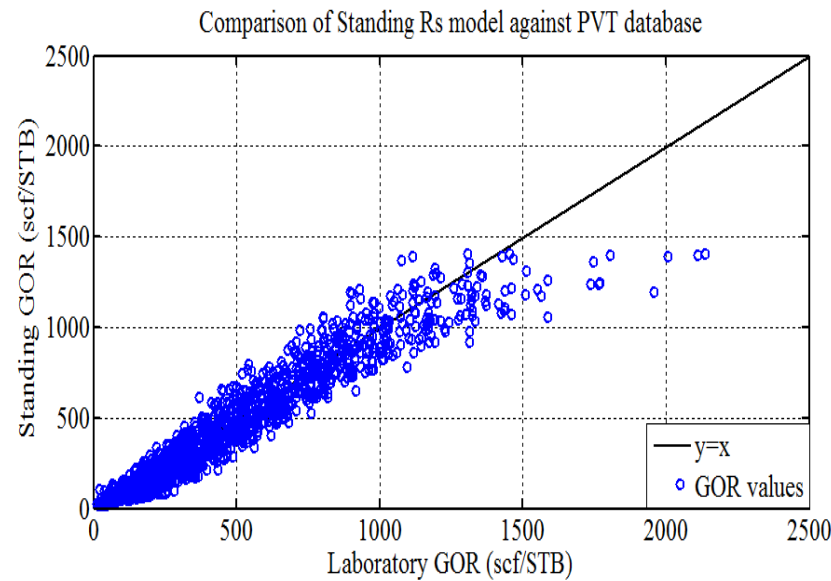
Ean% = Mean normalized absolute percent relative error – μέσο κανονικοποιημένο απόλυτο ποσοστιαίο σφάλμα

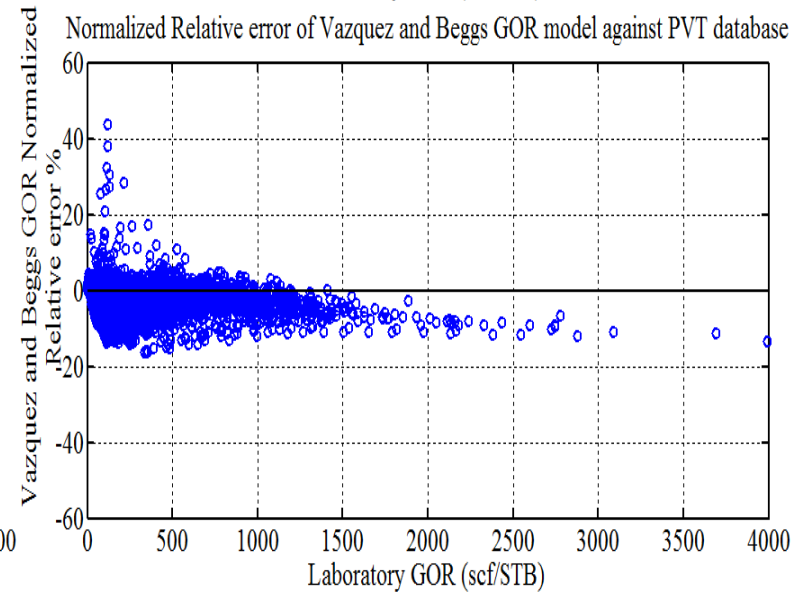
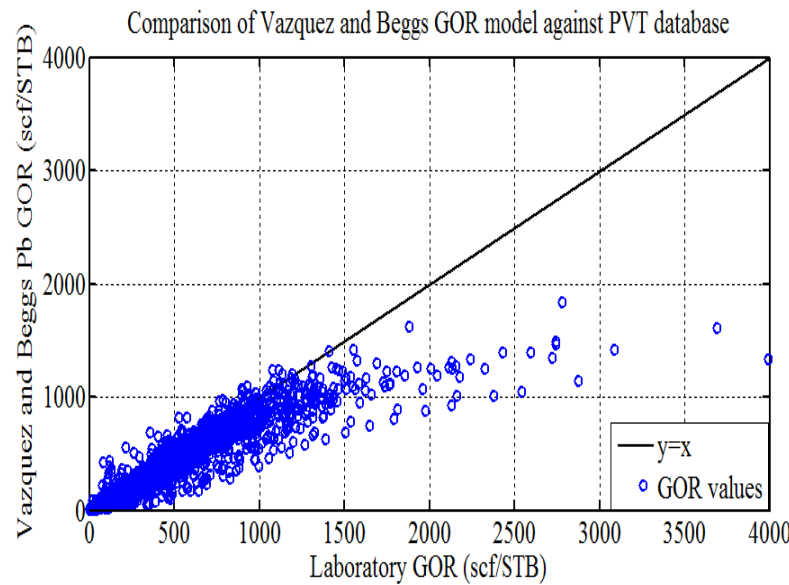
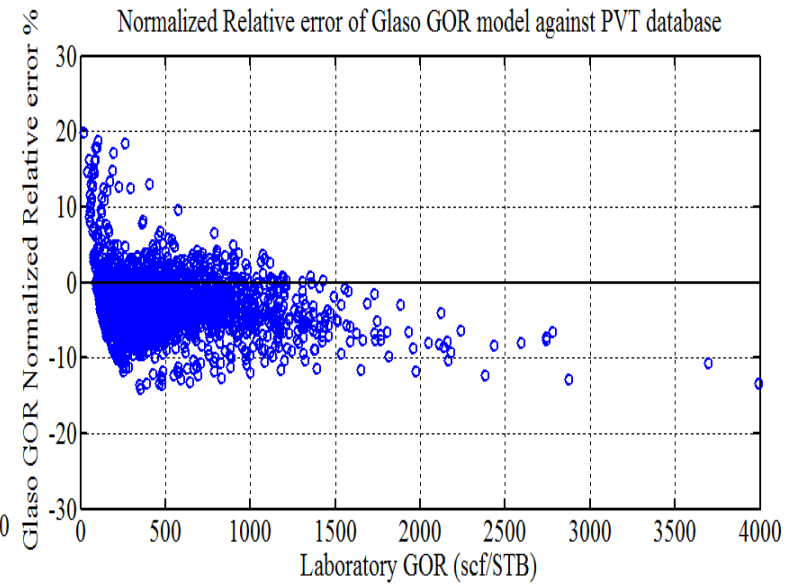
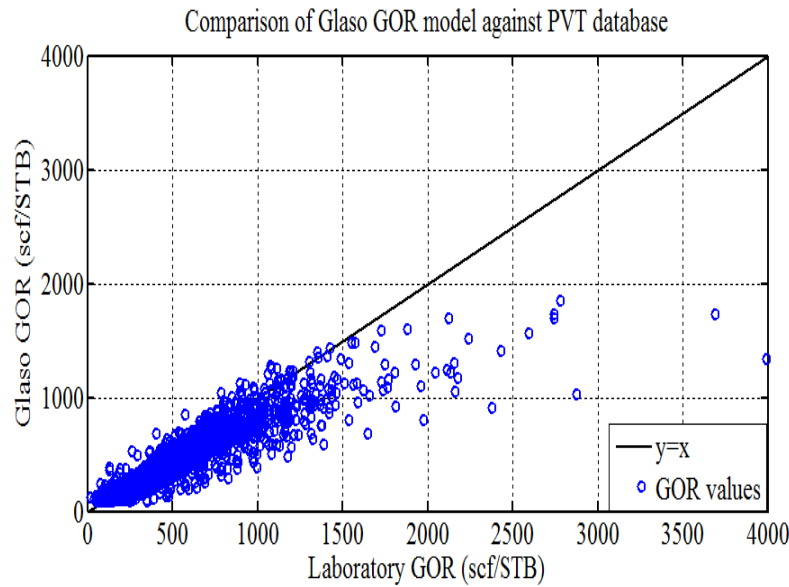
Ernmax = Maximum normalized relative percent error – μέγιστο κανονικοποιημένο ποσοστιαίο σχετικό σφάλμα

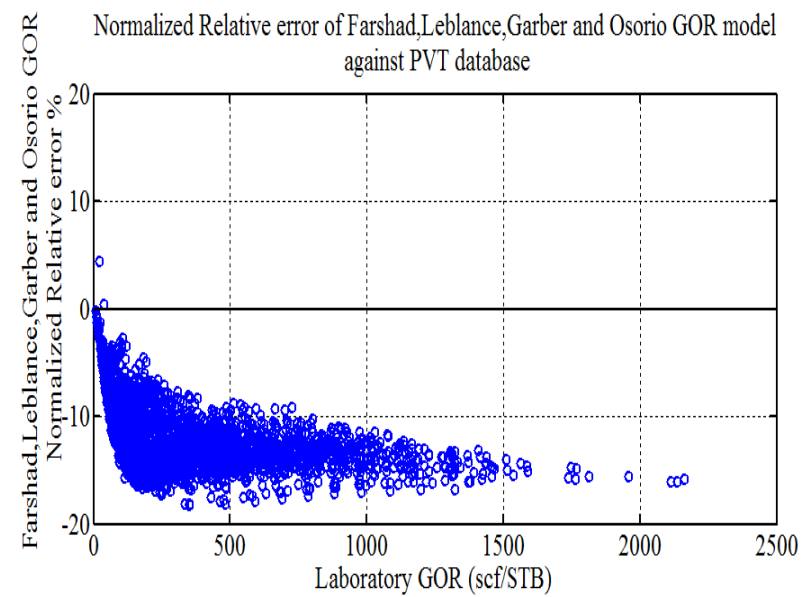
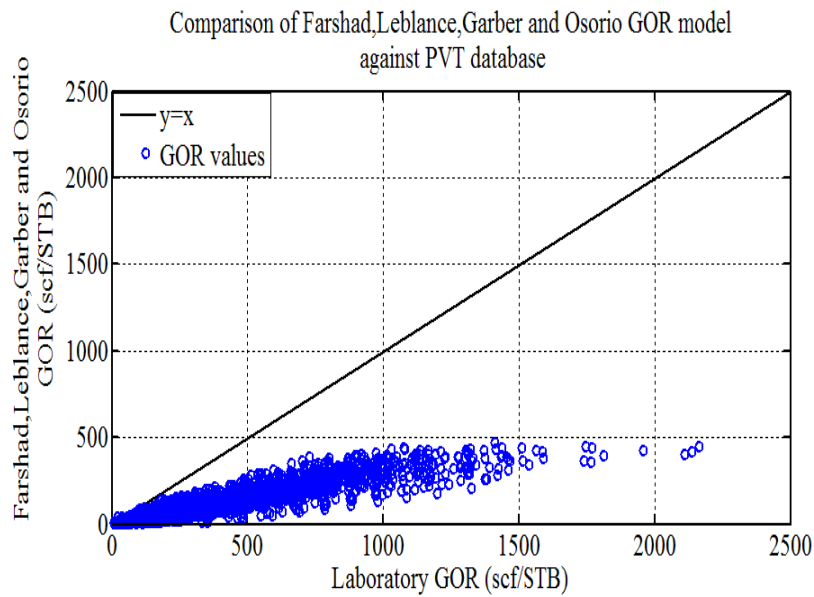
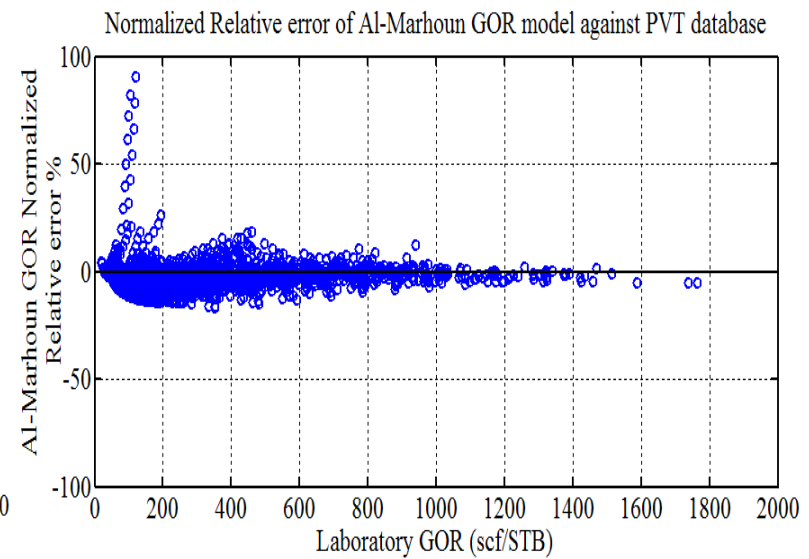
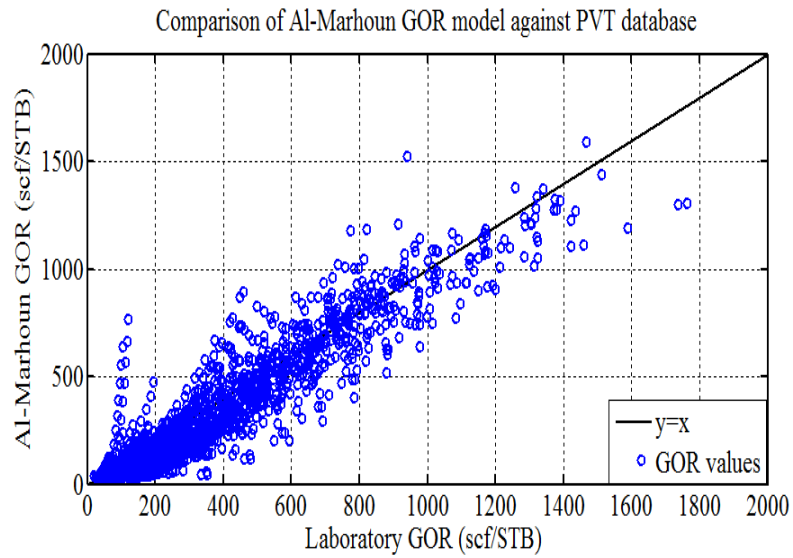
Ernmin = Minimum normalized relative percent error – ελάχιστο κανονικοποιημένο ποσοστιαίο σχετικό σφάλμα

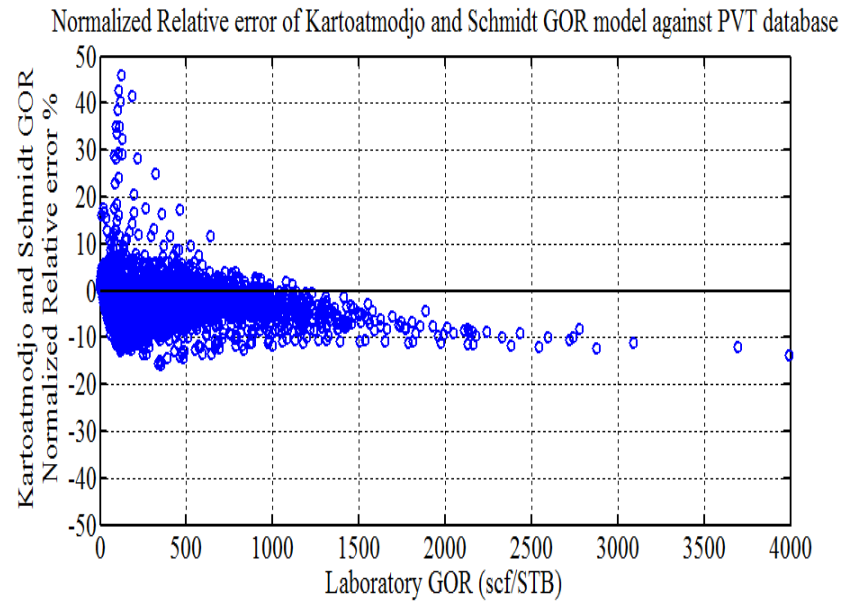
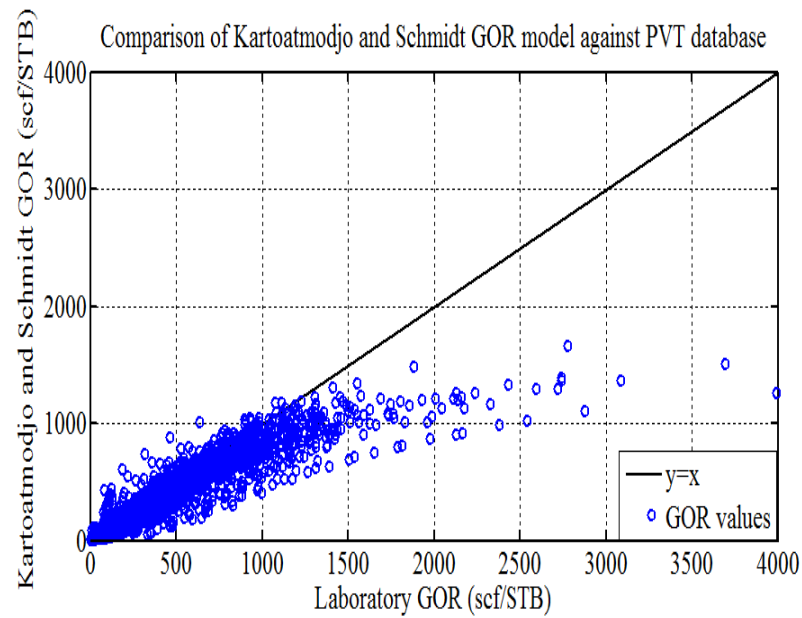
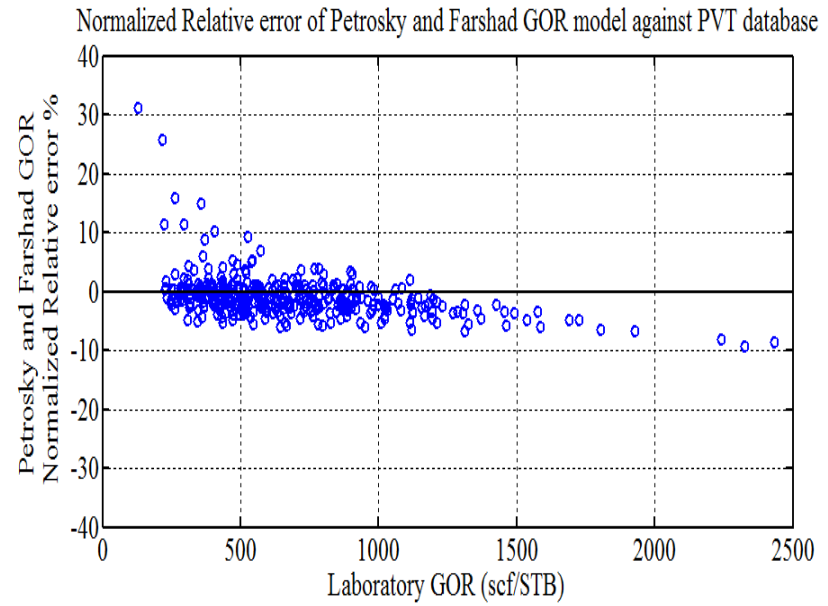
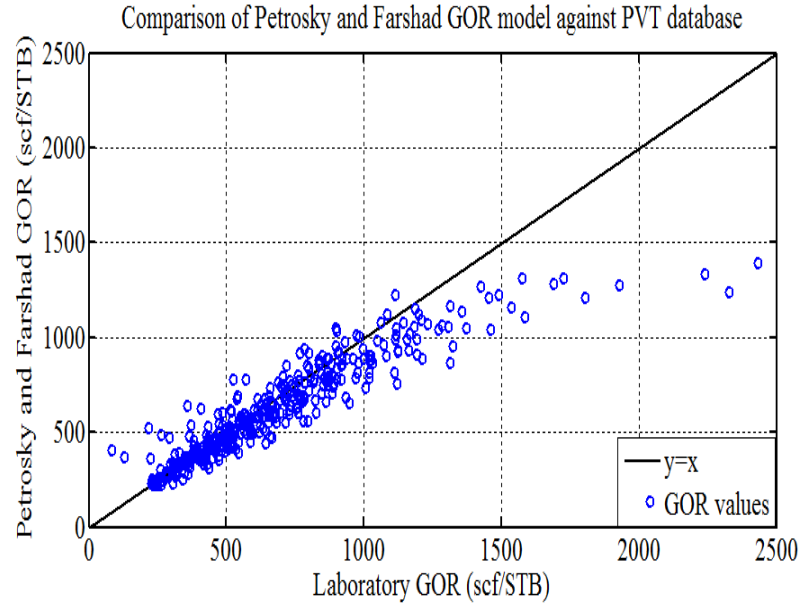
STDEV(Ern) = standard Deviation of mean normalized Relative percent error – τυπική απόκλιση του μέσου κανονικοποιημένου σχετικού ποσοστιαίου σφάλματος

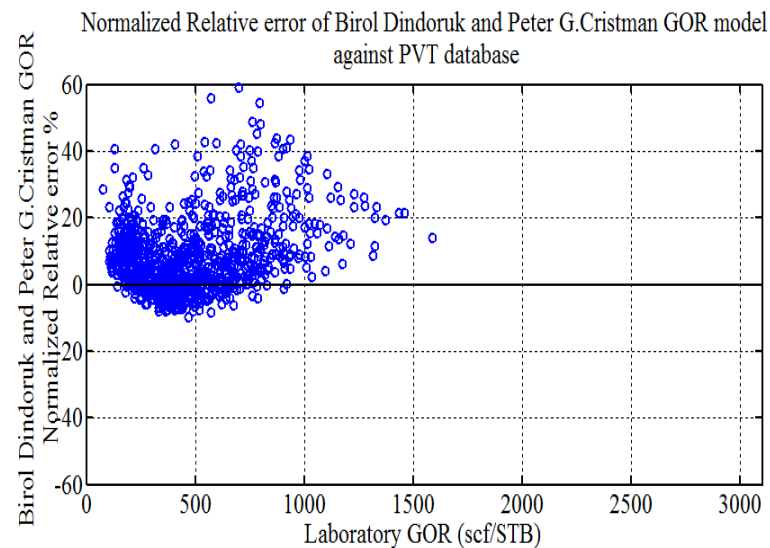
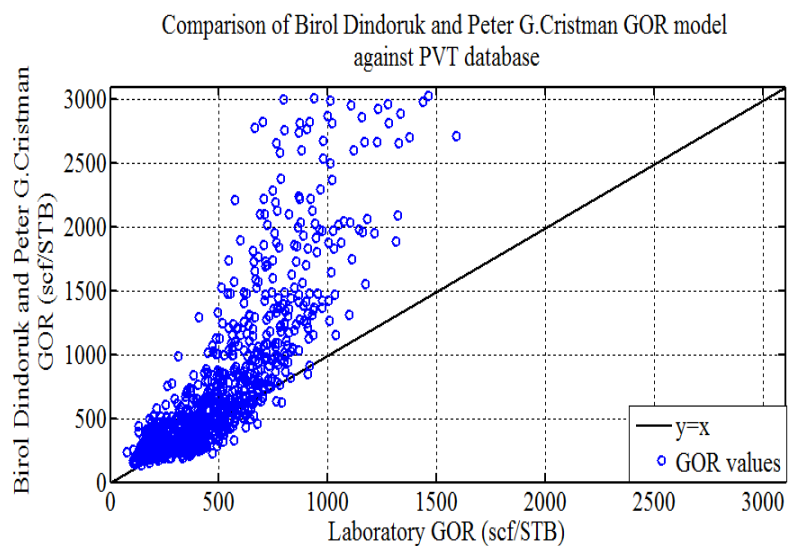
r = correlation coefficients pearson – συντελεστής συσχέτισης κατά pearson











Όπως διαπιστώνεται από τις γραφικές απεικονίσεις αλλά και από τα στατιστικά δεδομένα, οι εμπειρικές εξισώσεις των Farshad, Leblance, Garber & Osorio και των Birol Dindoruk και Peter G.Cristman παρουσιάζουν μεγάλες στατιστικές αποκλίσεις από τα δεδομένα των πετρελαίων της βάσης δεδομένων που χρησιμοποιήθηκε. Για αυτό το λόγο οι επιδόσεις τους δεν συγκρίθηκαν με αυτές το νευρωνικού που αναπτύχθηκε.

ΣΥΓΚΡΙΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΠΡΟΒΛΕΨΕΩΝ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΙΔΙΟΤΗΤΩΝ B_o , R_s , ρ_b ΤΩΝ ΑΝΑΠΤΥΧΘΕΝΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΜΕ ΤΙΣ ΑΝΤΙΣΤΟΙΧΕΣ ΠΡΟΒΛΕΨΕΙΣ ΤΩΝ ΕΜΠΕΙΡΙΚΩΝ ΣΥΣΧΕΤΙΣΕΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΑΠΟ ΤΗ ΒΙΟΜΗΧΑΝΙΑ ΠΕΤΡΕΛΑΙΟΥ

Για την σύγκριση των προβλέψεων των τιμών των ιδιοτήτων B_o , R_s , ρ_b των βέλτιστων νευρωνικών δικτύων που αναπτύχθηκαν με τις προβλέψεις των εμπειρικών συσχετίσεων που εφαρμόζονται στη βιομηχανία πετρελαίου ακολουθήθηκε η εξής μεθοδολογία.

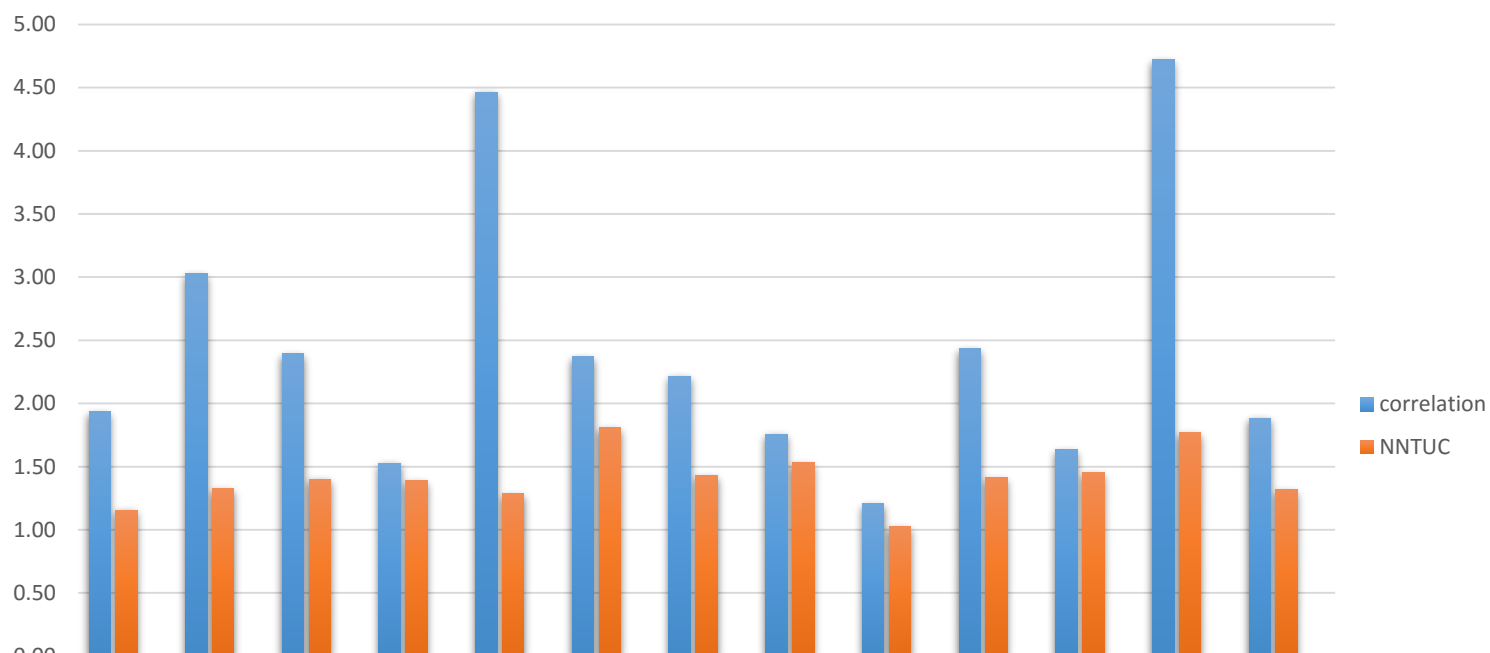
1. Οι εμπειρικές συσχετίσεις μπορούσαν να εφαρμοστούν σε ένα συγκεκριμένο εύρος τιμών για τα δεδομένα παραγωγής αλλά και για τις ζητούμενες ιδιότητες όπως περιεγράφηκαν αναλυτικά στα προηγούμενα κεφάλαια. Για αυτό το λόγο δημιουργήθηκαν υποσύνολα βάσης δεδομένων όπου τηρούνταν το εύρος τιμών που θέτει κάθε εμπειρική συσχέτιση έτσι ώστε να εξαχθούν σωστά συμπεράσματα.
2. Για να πραγματοποιηθεί μια δίκαια σύγκριση θα πρέπει να συγκριθούν τα Τ.Ν.Δ με τις εμπειρικές συσχετίσεις σε σύνολα αναφορών τα οποία να είναι άγνωστα και στις δυο εκτιμητικές μεθόδους. Σαν σύνολα αγνώστων αναφορών στη προκειμένη περίπτωση μπορούν να χρησιμοποιηθούν τα υποσύνολα ελέγχου (test) που είχαν αναφερθεί στο κεφάλαιο 6 για την ανάπτυξη του κάθε νευρωνικού δικτύου και αποτελούν ένα σύνολο αγνώστων αναφορών και για τα νευρωνικά δίκτυα.
3. Θα πρέπει να φτιαχτούν καινούρια υποσύνολα για τη πραγματοποίηση της σύγκρισης των αναπτυχθέντων νευρωνικών με κάθε εμπειρική συσχέτιση για τις ιδιότητες B_o , R_s , ρ_b έτσι ώστε να βρεθεί ένα σύνολο αγνώστων αναφορών προερχόμενων από το υποσύνολο ελέγχου (test) του νευρωνικού δικτύου που τηρούν το εύρος τιμών της κάθε εμπειρικής συσχέτισης ώστε να πραγματοποιηθεί η σύγκριση. Όπως γίνεται αντιληπτό για κάθε εμπειρική συσχέτιση θα χρησιμοποιηθεί διαφορετικό υποσύνολο λόγω του διαφορετικού αριθμού αναφορών που θα είναι σε θέση να τηρούν το εύρος τιμών
4. Σαν κριτήρια αξιολόγησης χρησιμοποιήθηκαν το μέσο απόλυτο σχετικό ποσοστιαίο σφάλμα και το μέσο σχετικό ποσοστιαίο σφάλμα

Εν συνεχεία παρατίθενται οι πίνακες αξιολόγησης

ΑΝΑΠΤΥΧΘΕΝ ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ VS ΕΜΠΕΙΡΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ

Εκτιμήσεις τιμών της ιδιότητας Βο σε σύνολο 77 αγνώστων πετρελαίων

Average absolute percent relative error



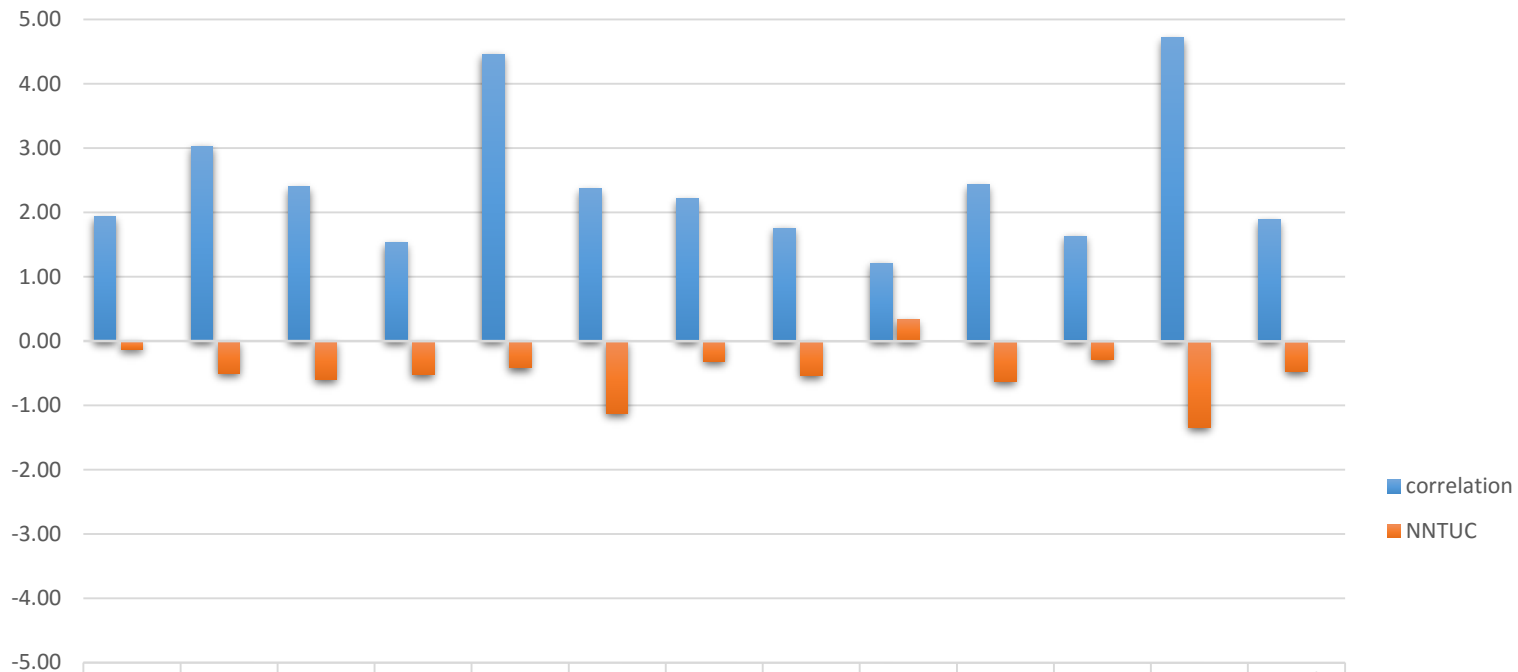
	Standin g	Vazquez and Beggs	Glaso	Al-Marhou n	Majeed & Salman	Dokla & Osman	Farshad, Leblanc e, Garbe r & Osorio	Al-Marhou n	Petrosky & Farshad	Omar & Todd	Kartoat modjo & Schmidt	Almehai deb	Birol Dindoru k and Peter G.Christ man
correlation	1.94	3.03	2.40	1.53	4.47	2.37	2.22	1.76	1.21	2.44	1.63	4.72	1.88
NNTUC	1.15	1.33	1.40	1.39	1.29	1.81	1.43	1.53	1.02	1.41	1.45	1.77	1.32
Numbers of Reports	42	65	57	51	61	19	58	76	25	39	68	28	44

comparison between correlation and NN performance

ΑΝΑΠΤΥΧΘΕΝ ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ VS ΕΜΠΕΙΡΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ

Εκτιμήσεις τιμών της ιδιότητας Bo σε σύνολο 77 αγνώστων πετρελαίων

Average percent relative error

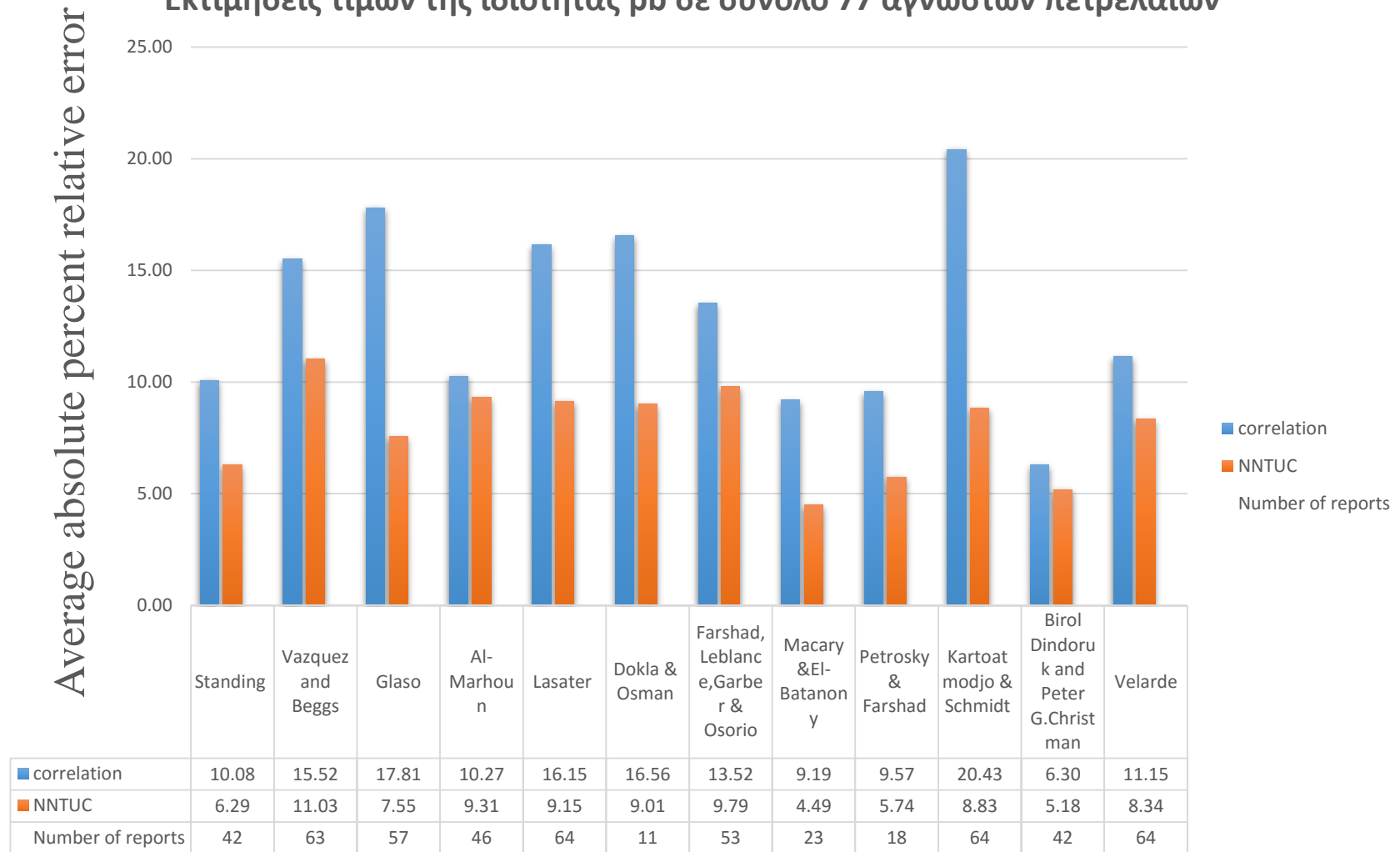


	Standin g	Vazquez and Beggs	Glaso	Al-Marhou n	Majeed & Salman	Dokla & Osman	Farshad, Leblanc e, Garbe r & Osorio	Al-Marhou n	Petrosky & Farshad	Omar & Todd	Kartoat modjo & Schmidt	Almehai deb	Birol Dindoru k and Peter G.Christ man
■ correlation	1.94	3.03	2.40	1.53	4.47	2.37	2.22	1.76	1.21	2.44	1.63	4.72	1.88
■ NNTUC	-0.12	-0.50	-0.59	-0.53	-0.42	-1.12	-0.32	-0.54	0.33	-0.62	-0.29	-1.34	-0.47
Numbers of reports	42	65	57	51	61	19	58	76	25	39	68	28	44

comparison between correlation and NN performance

ΑΝΑΠΤΥΧΘΕΝ ΝΕΥΡΩΝΙΚΟ VS ΕΜΠΕΙΡΙΚΩΝ ΣΥΣΧΕΤΙΣΕΩΝ

Εκτιμήσεις τιμών της ιδιότητας ρ_B σε σύνολο 77 αγνώστων πετρελαίων

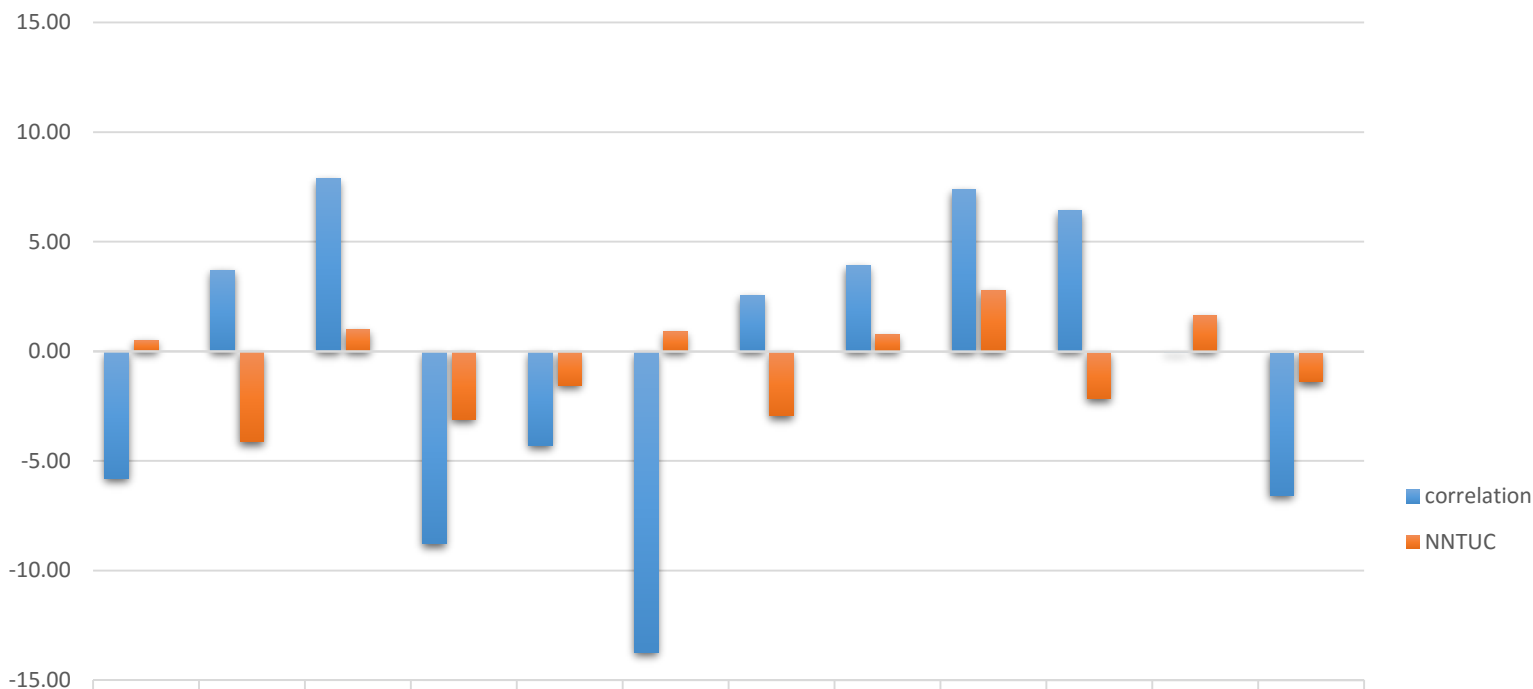


comparison between correlation and NN performance

ΑΝΑΠΤΥΧΘΕΝ ΝΕΥΡΩΝΩΝΙΚΟ VS ΕΜΠΕΙΡΙΚΩΝ ΣΥΣΧΕΤΙΣΕΩΝ

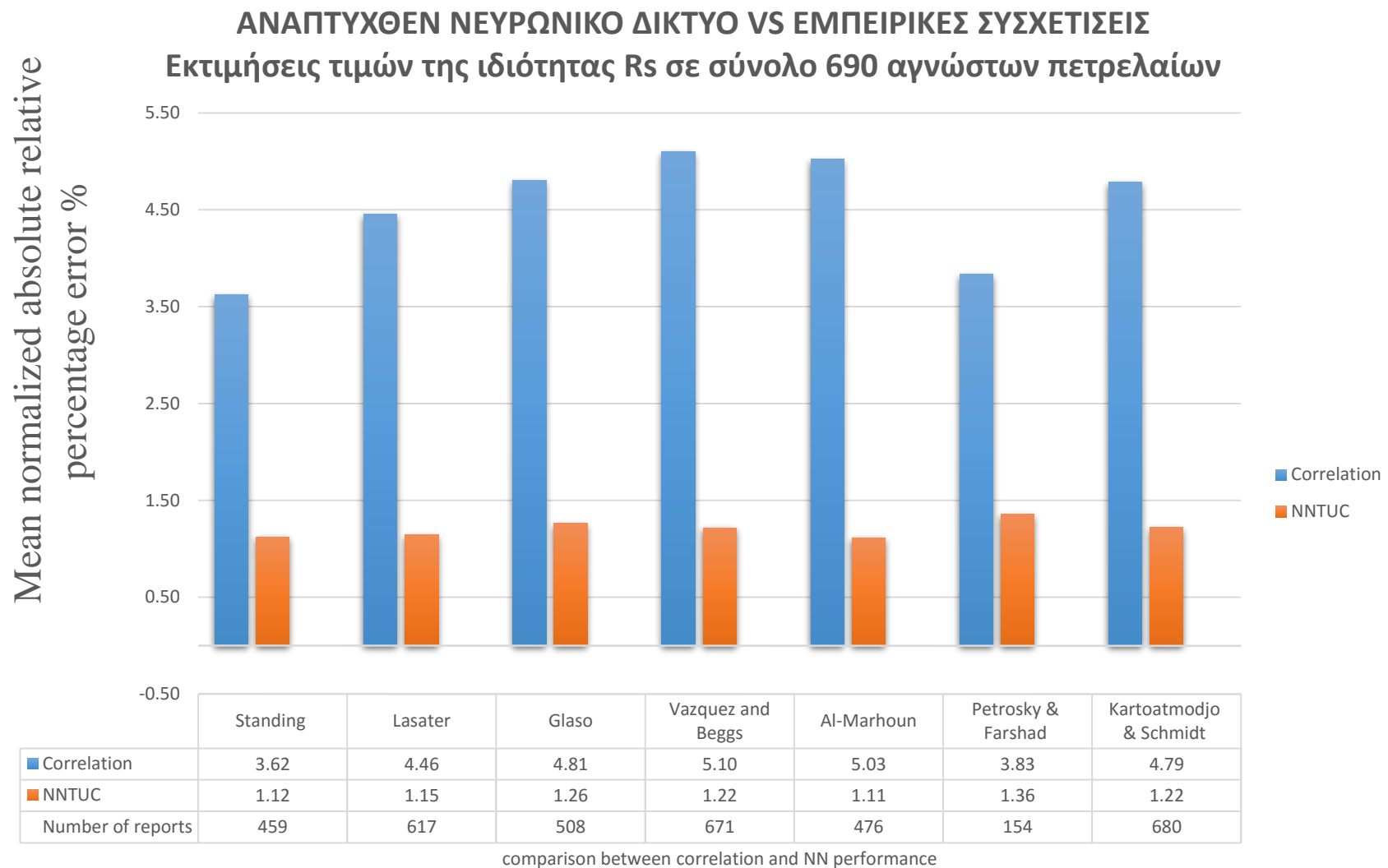
Εκτιμήσεις τιμών της ιδιότητας ρ_b σε σύνολο 77 αγνώστων πετρελαίων

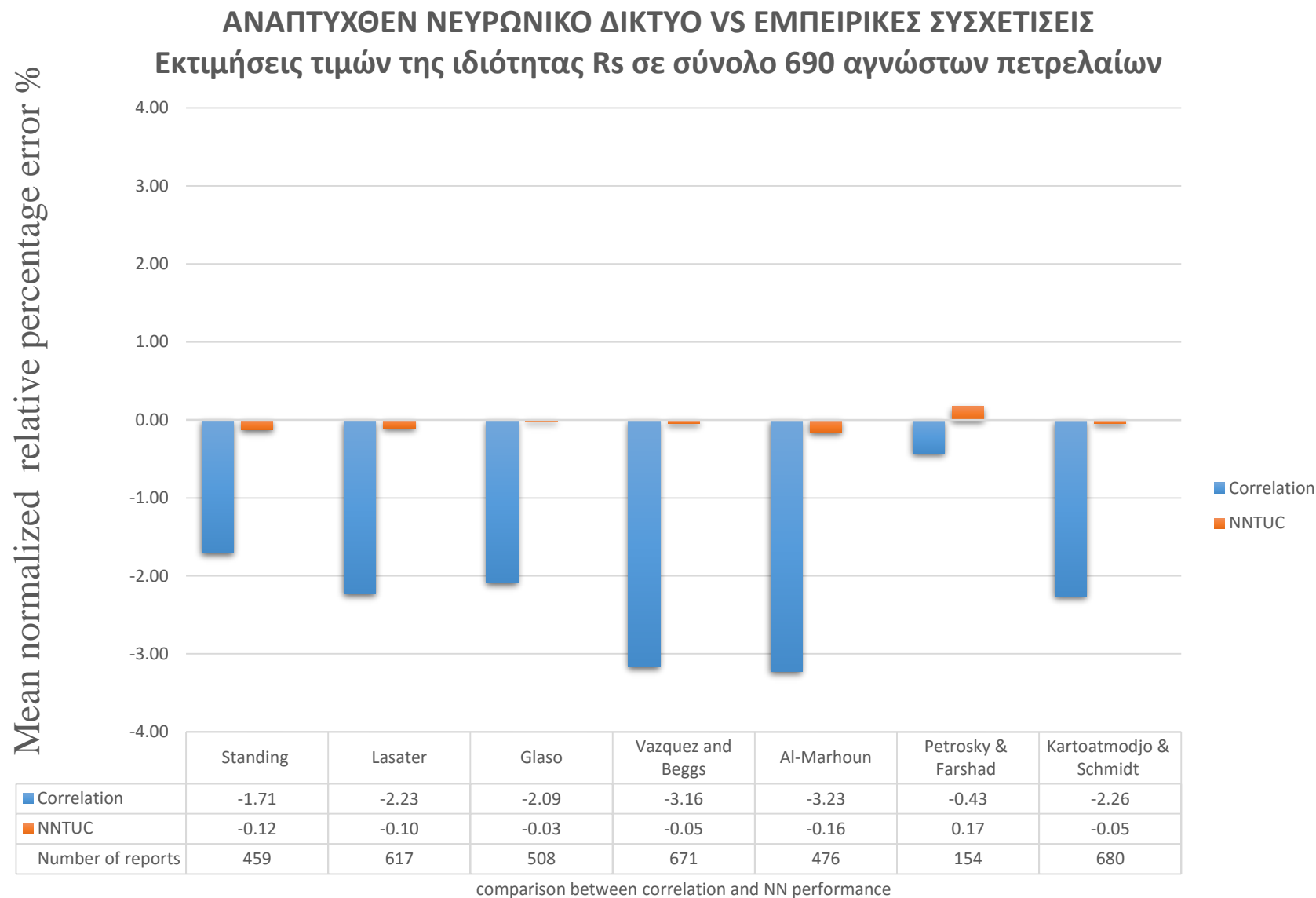
Average percent relative error



	Standing	Vazquez and Beggs	Glaso	Al-Marhoun	Lasater	Dokla & Osman	Farshad, Leblance, Garber & Osorio	Macary & El-Batanony	Petrosky & Farshad	Kartoat modjo & Schmidt	Birol Dindoruk and Peter G.Christman	Velarde
■ correlation	-5.79	3.71	7.88	-8.77	-4.30	-13.73	2.55	3.91	7.38	6.42	-0.06	-6.56
■ NNTUC	0.51	-4.13	0.99	-3.10	-1.55	0.89	-2.94	0.77	2.77	-2.16	1.62	-1.39
Number of reports	42	63	57	46	64	11	53	23	18	64	42	64

comparison between correlation and NN performance





ΓΕΝΙΚΑ ΣΥΜΠΕΡΑΣΜΑΤΑ

- ❖ Εγινε εφικτό με χρήση Τ.Ν.Δ. (ANN) να προβλεφθούν οι τιμές των ιδιοτήτων B_o , R_s , p_b χρησιμοποιώντας μόνο άμεσα διαθέσιμα στο πεδίο δεδομένα παραγωγής, με μέση ακρίβεια της τάξης του **1.75%** για το **B_o** , **6.99%** για το **R_s** , και **9.5%** για το **p_b** για πετρέλαια που προέρχονται από όλες τις γεωγραφικές περιοχές του κόσμου..
- ❖ Απαραίτητη προϋπόθεση για την ανάπτυξη εύρωστων εμπειρικών μοντέλων είναι πρόσβαση σε μεγάλες βάσεις δεδομένων ευρείας γεωγραφικής κατανομής πετρελαϊκών ρευστών από PVT μελέτες. Εν συνεχεία, οι βάσεις δεδομένων θα πρέπει να ελεγχθούν διεξοδικά για την ορθότητα των τιμών που εμπεριέχουν με διαδικασίες ελέγχου (quality control) ώστε να μπορούν να θεωρηθούν αξιόπιστες για τη χρησιμοποίησή τους. Εάν δεν υπάρχει πρόσβαση σε μια βάση δεδομένων με τέτοια χαρακτηριστικά τότε το μοντέλο που θα αναπτυχθεί δεν θα είναι σε θέση να εκτιμήσει ικανοποιητικά τις τιμές των ζητούμενων ιδιοτήτων σε άγνωστα πετρελαϊκά ρευστά με μεγάλη γεωγραφική διασπορά που συνεπάγεται ότι η απόδοση του θα είναι ενδεχομένως υποδεέστερη της των εξειδικευμένων ανά περιοχή εμπειρικών συσχετίσεων.
- ❖ Από την αξιολόγηση των αποτελεσμάτων των εκτιμήσεων που παρέχουν οι εμπειρικές συσχετίσεις για τις τιμές των ιδιοτήτων B_o , p_b , R_s με τη χρησιμοποίηση των επιλεχθέντων στατιστικών κριτηρίων διαπιστώνεται ότι, για τις περισσότερες εξ αυτών που χρησιμοποιούνται μέχρι σήμερα από τη βιομηχανία πετρελαίου, υπάρχει σημαντική απόκλιση στην ακρίβεια των σφαλμάτων που προκύπτουν έναντι άγνωστου πληθυσμού πετρελαϊκών ρευστών σε σχέση με την ακρίβεια που δηλώνουν οι αντίστοιχοι συγγραφείς τους. Για το λόγο αυτό αντενδείκνυται η χρησιμοποίηση εμπειρικών συσχετίσεων (regional) που έχουν παραμετροποιηθεί από αναφορές πετρελαϊκών ρευστά από μια μόνο γεωγραφική περιοχή για την εκτίμηση των τιμών των PVT ιδιοτήτων ρευστών προερχομένων από διαφορετικές γεωγραφικές περιοχές. Αντιθέτως, καλύτερη απόδοση, όπως διαπιστώθηκε, επέδειξαν οι εμπειρικές συσχετίσεις του Standing, Vazquez and Beggs, του Al-Marhoun οι οποίες έχουν αναπτυχθεί από μια ευρεία γκάμα αναφορών πετρελαϊκών ρευστών (global) και είναι σε θέση να δώσουν συγκριτικά πιο αξιόπιστες εκτιμήσεις για τις τιμές των ιδιοτήτων του B_o , p_b , R_s για άγνωστα πετρελαϊκά ρευστά.

- ❖ Όπως διαπιστώνεται από τα παρατεθέντα διαγράμματα σύγκρισης, τα νευρωνικά δίκτυα που αναπτύχθηκαν στα πλαίσια της παρούσας διπλωματικής δοκιμασμένα για την εκτίμηση των τιμών των ιδιοτήτων των Bo, Rs, &rb ενός υποσύνολου αγνώστων αναφορών πετρελαϊκών ρευστών (που δεν χρησιμοποιήθηκαν στην εκπαίδευση), παρουσιάζουν σημαντικά μεγαλύτερη ακρίβεια εκτίμησης σε σχέση με τις όλες τις εμπειρικές συσχετίσεις που έχουν χρησιμοποιούνται ευρέως στη βιομηχανία πετρελαίου. Αυτό οφείλεται κύρια σε δύο λόγους:
 1. Στη πλούσια βάση δεδομένων που χρησιμοποιήθηκε για την ανάπτυξη των νευρωνικών δικτύων από αναφορές πετρελαϊκών ρευστών από διαφορετικές γεωγραφικές περιοχές.
 2. Στη αυξημένη προσαρμοστικότητα που επιδεικνύουν τα νευρωνικά δίκτυα για την επίλυση μη γραμμικών προβλημάτων σε σχέση τα συγκεκριμένης αλγεβρικής μορφής μαθηματικά μοντέλα που χρησιμοποιούν οι εμπειρικές συσχετίσεις.
- ❖ Η χρήση του κριτηρίου πρόωρης διακοπής (early stopping) εφαρμόστηκε κατά τη διαδικασία της εκπαίδευσης και κρίθηκε αποτελεσματική για την αντιμετώπιση του προβλήματος της υπερ-εκπαίδευσης κατά τη διαδικασία της σύγκλισης ώστε να διασφαλιστεί η γενίκευση επί των εκτιμήσεων των τιμών του υποσυνόλου test για τη συγκεκριμένη δομή του κάθε νευρωνικού που αναπτύχθηκε.
- ❖ Όπως διαπιστώθηκε στη παρούσα εργασία, με την ανάπτυξη μεγάλου αριθμού νευρωνικών δικτύων με αυξανόμενη δομή αριθμό κρυφών νευρώνων κατέστη δυνατή η εύρεση της δομής του νευρωνικού στην οποία αρχίζει να εμφανίζεται το πρόβλημα της υπερ-προσαρμογής. Για να επιτευχθεί η εύρεση αυτής της δομής έγινε αξιολόγηση των αποτελεσμάτων των στατιστικών κριτηρίων επί των εκτιμήσεων των τιμών των υποσυνόλων εκπαίδευσης, test.
- ❖ Διαπιστώθηκε επίσης ότι με την ανάπτυξη πολλών επαναληπτικών αναπτύξεων για κάθε δομή νευρωνικού δικτύου επιλύθηκε και το πρόβλημα της σύγκλισης σε ένα τοπικό ελάχιστο σε σχέση με το γενικό. Βάσει των αποτελεσμάτων των στατιστικών κριτηρίων, η βέλτιστη επαναληπτική ανάπτυξη για τη κάθε δομή φαίνεται ότι ξεκίνησε από μια τυχαία τιμή που οδήγησε κατά τη διαδικασία της σύγκλισης στο γενικό ελάχιστο ή σε κάποιο τοπικό ελάχιστο πολύ κοντά σε αυτό.

- ❖ Αποδείχθηκε πως η συγγραφή συμπληρωματικού κώδικα μαζί με τη χρησιμοποίηση της εργαλειοθήκης των τεχνητών νευρωνικών δικτύων του προγράμματος Matlab. Ήταν η καλύτερη λύση για την επίτευξη ανάπτυξης μεγάλου αριθμού νευρωνικών δικτύων χωρίς την παρέμβαση του εκπαιδευτή. Επίσης η χρησιμοποίηση του συμπληρωματικού κώδικα έδωσε τη δυνατότητα τροποποίησης ή προσθήκη χαρακτηριστικών όπως π.χ. η επιλογή των στατιστικών κριτηρίων ώστε να αναπτυχθούν τα νευρωνικά σύμφωνα με τις ανάγκες του προβλήματος.
- ❖ Τέλος διαπιστώθηκε ότι για την επιλογή του βέλτιστου νευρωνικού δικτύου από τα συνολικά αναπτυχθέντα νευρωνικά θα πρέπει να επιλέγεται αυτό με τη υψηλότερη ακρίβεια επί των εκτιμήσεων των τιμών και των τριών υποσυνόλων (εκπαίδευσης, επικύρωσης, τεστ). Με αυτό τον τρόπο διασφαλίζεται η μέση ακρίβεια εκτίμησης για όλο το εύρος τιμών που χρησιμοποιήθηκε για την ανάπτυξη αλλά και ότι στο συγκεκριμένο νευρωνικό έχει επιτευχθεί η καλύτερη δυνατή γενίκευση .

Θέματα προς περαιτέρω μελέτη

- Ανάπτυξη δυο διαφορετικών ANN για την εκτίμηση της ιδιότητας R_s όπου θα μοιράζεται το εύρος τιμών για τιμές κάτω από 1500 scf / STB έτσι ώστε να αυξηθεί η ακρίβεια επί των εκτιμήσεων στις χαμηλές τιμές.
- Έλεγχος της διεθνούς βιβλιογραφίας για να αναζητήσει άλλων εργαλείων της Μηχανικής Μάθησης έτσι ώστε να εξεταστεί εάν μπορεί να επιτευχθεί μεγαλύτερη ακρίβεια στις εκτιμήσεις των PVT ιδιοτήτων

ΠΑΡΑΡΤΗΜΑ

Παρουσίαση του συμπληρωματικού κώδικα

Συμπληρωματικός κώδικας:

```
% x- input data
% y- target data
inputs = x;
targets = y;
numnet = 200;
sizenet = 5 :20;
all_net = cell(1, numnet * size(sizenet, 2));
all_trainPerformance = zeros(1, numnet * size(sizenet, 2));
all_valPerformance = zeros(1, numnet * size(sizenet, 2));
all_testPerformance = zeros(1, numnet * size(sizenet, 2));
all_trainMeanRelError = zeros(1, numnet * size(sizenet, 2));
all_valMeanRelativeError = zeros(1, numnet * size(sizenet, 2));
all_testMeanRelativeError = zeros(1, numnet * size(sizenet, 2));
all_trainMeanAbsRelError = zeros(1, numnet * size(sizenet, 2));
all_valMeanAbsRelError = zeros(1, numnet * size(sizenet, 2));
all_testMeanAbsRelError = zeros(1, numnet * size(sizenet, 2));

current_net = 0;

for Num_of_neurons = sizenet

    disp(['Testing size ', int2str(Num_of_neurons)])

    % create a fitting Network
    for i = 1 : numnet

        current_net = current_net + 1;

        disp(['Training ' num2str(i) '/' int2str(numnet)])

        net = fitnet(Num_of_neurons);

        % Choose Input and Output Pre/Post-Processing Functions
        net.inputs{1}.processFcns = {'removeconstantrows','mapminmax'};
        net.outputs{2}.processFcns = {'removeconstantrows','mapminmax'};

        % Setup Division of Data for Training,Validation,Testing
        net.divideFcn = 'dividerand'; %Divide data randomly
        net.divideMode = 'sample'; %Divide up every sample
        net.divideParam.trainRatio = 60/100;
        net.divideParam.valRatio = 20/100;
        net.divideParam.testRatio = 20/100;
        net.trainFcn = 'trainlm'; %Levenberg-Marquardt
        net.performFcn = 'mse'; %Mean squared error
        net.layers{1}.transferFcn = 'logsig'; %transfer function
        net.plotFcns = {'plotperform','plottrainstate','ploterrhist','plotregres-
sion','plotfit'};

        % Train the Network
        [net, tr] = train(net,inputs,targets);

        %% Evaluate network's performance
        % Calculate network's output and error over ALL pairs
        outputs = net(inputs);
```

```

errors = outputs - targets;
performance = perform(net,targets,outputs);

% Split targets into train, validation and testing
% Set NaN to irrelevant datapoints
trainTargets = targets .* tr.trainMask{1};
valTargets = targets .* tr.valMask{1};
testTargets = targets .* tr.testMask{1};
%Calculate perform
trainPerformance = perform(net,trainTargets,outputs);
valPerformance = perform(net,valTargets,outputs);
testPerformance = perform(net,testTargets,outputs);

% Split errors into train, validation and testing
% Set NaN to irrelevant datapoints
trainErrors = errors .* tr.trainMask{1};
valErrors = errors .* tr.valMask{1};
testErrors = errors .* tr.testMask{1};

% Calculate absolute errors per class
trainAbsErrors = abs(trainErrors);
valAbsErrors = abs(valErrors);
testAbsErrors = abs(testErrors);

% Calculate relative errors per class
trainRelErrors = trainErrors ./ trainTargets * 100;
valRelErrors = valErrors ./ valTargets * 100;
testRelErrors = testErrors ./ testTargets * 100;

% Calculate absolute relative errors per class
trainAbsRelErrors = abs(trainRelErrors);
valAbsRelErrors = abs(valRelErrors);
testAbsRelErrors = abs(testRelErrors);

% Calculate error statistics
trainMeanError = mean(trainErrors(isnan(trainErrors) == 0));
valMeanError = mean(valErrors(isnan(valErrors) == 0));
testMeanError = mean(testErrors(isnan(testErrors) == 0));

trainMeanRelError = mean(trainRelErrors(isnan(trainRelErrors) == 0));
valMeanRelativeError = mean(valRelErrors(isnan(valRelErrors) == 0));
testMeanRelativeError = mean(testRelErrors(isnan(testRelErrors) == 0));

trainMeanAbsError = mean(trainAbsErrors(isnan(trainAbsErrors) == 0));
valMeanAbsError = mean(valAbsErrors(isnan(valAbsErrors) == 0));
testMeanAbsError = mean(testAbsErrors(isnan(testAbsErrors) == 0));

trainMeanAbsRelError = mean(trainAbsRelErrors(isnan(trainAbsRelErrors) == 0));
valMeanAbsRelError = mean(valAbsRelErrors(isnan(valAbsRelErrors) == 0));
testMeanAbsRelError = mean(testAbsRelErrors(isnan(testAbsRelErrors) == 0));

% Store results
all_net{1, current_net} = net;
all_trainPerformance(1,current_net)= trainPerformance;
all_valPerformance(1,current_net)= valPerformance;
all_testPerformance(1,current_net)= testPerformance;
all_trainMeanRelError(1,current_net)= trainMeanRelError;
all_valMeanRelativeError(1,current_net)= valMeanRelativeError;
all_testMeanRelativeError(1,current_net)= testMeanRelativeError;
all_trainMeanAbsRelError(1,current_net)= trainMeanAbsRelError;
all_valMeanAbsRelError(1,current_net)= valMeanAbsRelError;
all_testMeanAbsRelError(1,current_net)= testMeanAbsRelError;

```



```
%View the Network
%view(net)
%plots
% figure,plotperform(tr)
% figure,plottrainstate(tr)
% figure,plotregression(targets,outputs)
% figure,ploterrhist(errors)
end
end
```

ΒΙΒΛΙΟΓΡΑΦΙΑ

N. Βαρότσης. Μηχανική Ταμειωτήρων. Πολυτεχνείο Κρήτης.

site : www.rigzone.com.

site : petrowiki.org.

site : www.slb.com.

site : www.onepetro.com

Peter S.Hegeman (Schlumberger), Chengli Dong (Schlumberger), Nikos Varotsis (Tech. U of Crete), Vassilis Gaganis (Tech. U of Crete). Paper : Application of Artificial Neural Networks to Downhole Fluid Analysis. 123423 –PA SPE Journal Paper 2009.

Albert Tarantola. Inverse Problem Theory and methods for parameter estimation. Institute de Physique du Globe de Paris, Universite de Paris 6, Paris, France.

Διονύσιος Θ. Χρηστόπουλος. Σημειώσεις του Προπτυχιακού Μαθήματος ‘Εισαγωγή στη Γεωστατιστική’. Χανιά Απρίλιος 2008.

Simon Haykin. Νευρωνικά Δίκτυα και Μηχανική Μάθηση. Τρίτη έκδοση. Computer Science. site : cs.stanford.edu.

I. Βλαχάβας, Π. Κεφαλάς, Ν. Βασιλειάδης, Φ. Κόκκορας, Η. Σακελλαρίου. Τεχνητή Νοημοσύνη. Β’ έκδοση.

ARTIFICIAL INTELLIGENCE Comes of Age in Oil and Gas’. JPT (Journal of Petroleum Technology) January 2013.

Ali Abedini, Siavash Ashoori, Yaser Saki. ‘Application of Neural Model for Prediction of Asphaltene Precipitation’. 132760 –MS SPE Conference Paper 2010.

Varotsis N., Gaganis V., Nighswander J., Guieze P., “ Novel Non-Iterative Method for Predicting the PVT Behavior of Reservoir Fluids”, SPE 56745, presented at the 1999 SPE Annual Technical Conference, Oct 3-6, 1999, Houston, USA.

Varotsis N. Schlumberger, Gaganis V. Consultant, Nighswander J.
Schlumberger. 'Quality Assurance Tool for PVT Simulator Predictions'. 81751
PA-SPE Journal Paper – 2002.

Adeyemi B.J., Sulaimon A.A., 'Predicting Wax Formation Using Artificial
Neural Networks'. 163026 –MS SPE Conference Paper 2012.

Standing, M.B. 1947. A Pressure-Volume-Temperature Correlation for Mixtures of California Oils and
Gases. API Drilling and Production Practice (1947): 275-287.

Frick, T.C. 1962. Petroleum Production Handbook, Vol. II, Chap. 18-19. Dallas, Texas: Society of Petro-
leum Engineers

Elam, F.M. 1957. Prediction of Bubble Point Pressures and Formation Volume Factors from Field Data.
MS thesis, University of Texas at Austin, Austin, Texas.

Vazquez, M.E. 1976. Correlations for Fluid Physical Property Prediction. MS thesis, University of Tulsa,
Tulsa, Oklahoma.

Vazquez, M. and Beggs, H.D. 1980. Correlations for Fluid Physical Property Prediction. J Pet Technol 32
(6): 968-970. SPE-6719-PA. <http://dx.doi.org/10.2118/6719-PA>

Glasø, Ø. 1980. Generalized Pressure-Volume-Temperature Correlations. J Pet Technol 32 (5): 785-795.
SPE-8016-PA. <http://dx.doi.org/10.2118/8016-PA>

Labedi, R.M. 1982. PVT Correlations of the African Crudes. PhD thesis. 1982. . PhD thesis, Colorado
School of Mines, Leadville, Colorado (May 1982).

Labedi, R.M. 1990. Use of Production Data to Estimate the Saturation Pressure, Solution Gor, and Chem-
ical Composition of Reservoir Fluids. Presented at the SPE Latin America Petroleum Engineering Con-
ference, Rio de Janeiro, Brazil, 14-19 October. SPE-21164-MS. <http://dx.doi.org/10.2118/21164-MS>

Owolabi, O.O. 1984. Reservoir Fluid Properties of Alaskan Crudes. MS thesis, University of Alaska,
Fairbanks, Alaska (May 1984).

Al-Marhoun, M.A. 1985. Pressure-Volume-Temperature Correlations for Saudi Crude Oils. Presented at
the SPE Middle East Oil Technical Conference and Exhibition, Bahrain, 11–14 March. SPE-13718-MS.

Obomanu, D.A. and Okpobiri, G.A. 1987. Correlating the PVT Properties of Nigerian Crudes. J. Energy
Resour. Technol. 109 (4): 214-217. <http://dx.doi.org/10.1115/1.3231349>

Al-Marhoun, M.A. 1988. PVT Correlations for Middle East Crude Oils. J Pet Technol 40 (5): 650–666.
SPE-13718-PA. <http://dx.doi.org/10.2118/13718-PA>

Asgarpour, S., McLauchlin, L.L., Wong, D. et al. 1989. Pressure-Volume-Temperature Correlations For Western Canadian Gases And Oils. J Can Pet Technol 28 (4): 103. PETSOC-89-04-08. <http://dx.doi.org/10.2118/89-04-08>

Al-Najjar, H.S., Al-Soof, N.B.A., and Al-Khalisy, K.M. 1988. Correlations For Bubble-Point Pressures, Gas Oil Ratios And Formation Volume Factors For Iraqi Crude Oils. Journal of Petroleum Research (June 1988): 13.

Ahmed, T. 1989. Hydrocarbon Phase Behavior, Vol. 7. Tulsa, Oklahoma: Contributions in Petroleum Geology and Engineering, Gulf Publishing Company.

Abdul-Majeed, G.H., Salman, N.H., and Scarth, B.R. 1988. An Empirical Correlation For Oil FVF (Formation Volume Factor) Prediction. J Can Pet Technol 27 (6): 118. PETSOC-88-06-10. <http://dx.doi.org/10.2118/88-06-10>

Dokla, M.E. and Osman, M.E. 1992. Correlation of PVT Properties For UAE (United Arab Emirates) Crudes. SPE Form Eval (March 1992): 41.

Petrosky, G.E. Jr. 1990. PVT Correlations for Gulf of Mexico Crude Oils. MS thesis. 1990. . MS thesis, University of Southwestern Louisiana, Lafayette, Louisiana.

Petrosky, G.E. Jr. and Farshad, F. 1998. Pressure-Volume-Temperature Correlations for Gulf of Mexico Crude Oils. SPE Res Eval & Eng 1 (5): 416-420. SPE-51395-PA. <http://dx.doi.org/10.2118/51395-PA>

Kartoatmodjo, R.S.T. 1990. New Correlations for Estimating Hydrocarbon Liquid Properties. MS thesis, University of Tulsa, Tulsa, Oklahoma.

Kartoatmodjo, T.R.S. and Schmidt, Z. 1991. New Correlations for Crude Oil Physical Properties, Society of Petroleum Engineers, unsolicited paper 23556-MS.

Kartoatmodjo, T. and Z., S. 1994. Large Data Bank Improves Crude Physical Property Correlations. Oil Gas J. 92 (27): 51-55.

Al-Marhoun, M.A. 1992. New Correlations For Formation Volume Factors Of Oil And Gas Mixtures. J Can Pet Technol 31 (3): 22. PETSOC-92-03-02. <http://dx.doi.org/10.2118/92-03-02>

Frashad, F., LeBlanc, J.L., Garber, J.D. et al. 1996. Empirical PVT Correlations For Colombian Crude Oils. Presented at the SPE Latin American and Caribbean Petroleum Engineering Conference, Port of Spain, Trinidad and Tobago, 23-26 April. SPE-36105-MS. <http://dx.doi.org/10.2118/36105-MS>

Macary, S.M. and El-Batanoney, M.H. 1992. Derivation of PVT Correlations for the Gulf of Suez Crude Oils. Proc., 11th EGPC Petroleum Exploration and Production Conference, Cairo, Egypt, Vol. 1, 374.

Omar, M.I. and Todd, A.C. 1993. Development of New Modified Black Oil Correlations for Malaysian Crudes. Presented at the SPE Asia Pacific Oil and Gas Conference, Singapore, 8-10 February. SPE-25338-MS. <http://dx.doi.org/10.2118/25338-MS>

Omar, M.I., Daud, M.E., and Raja, D.M.A. 1993. New Correlation For Determining Bubble Point Oil FVF (Formation Volume Factor). Presented at the 1993 Asian Council Petroleum Conference, Bangkok, Thailand, 2-6 November.

Almehaideb, R.A. 1997. Improved PVT Correlations for UAE Crude Oils. Presented at the Middle East Oil Show and Conference, Bahrain, 15-18 March. SPE-37691-MS. <http://dx.doi.org/10.2118/37691-MS>

Elsharkawy, A.M. and Alikhan, A.A. 1997. Correlations for predicting solution gas/oil ratio, oil formation volume factor, and undersaturated oil compressibility. J. Pet. Sci. Eng. 17 (3-4): 291-302. [http://dx.doi.org/10.1016/S0920-4105\(96\)00075-7](http://dx.doi.org/10.1016/S0920-4105(96)00075-7)

Khairy, M., El-Tayeb, S., and Hamdallah, M. 1998. PVT Correlations Developed for Egyptian Crudes. Oil Gas J. 96 (18): 114.

Al-Shammasi, A.A. 2001. A Review of Bubblepoint Pressure and Oil Formation Volume Factor Correlations. SPE Res Eval & Eng 4 (2): 146-160. SPE-71302-PA. <http://dx.doi.org/10.2118/71302-PA>

Levitan, L.L. and Murtha, M. 1999. New Correlations Estimate Pb, FVF. Oil Gas J. 97 (10): 70.

Velarde, J., Blasingame, T.A., and McCain Jr., W.D. 1997. Correlation of Black Oil Properties At Pressures Below Bubble Point Pressure - A New Approach. Presented at the Annual Technical Meeting of CIM, Calgary, Alberta, 8-11 June. PETSOC-97-93. <http://dx.doi.org/10.2118/97-93>

Dindoruk, B. and Christman, P.G. 2001. PVT Properties and Viscosity Correlations for Gulf of Mexico Oils. Presented at the SPE Annual Technical Conference and Exhibition, New Orleans, 30 September-3 October. SPE-71633-MS. <http://dx.doi.org/10.2118/71633-MS>

Elsharkawy, A.M., Elgibaly, A.A., and Alikhan, A.A. 1995. Assessment of the PVT correlations for predicting the properties of Kuwaiti crude oils. J. Pet. Sci. Eng. 13 (3-4): 219-232. [http://dx.doi.org/10.1016/0920-4105\(95\)00012-7](http://dx.doi.org/10.1016/0920-4105(95)00012-7)

Mahmood, M.A. and Al-Marhoun, M.A. 1996. Evaluation of empirically derived PVT properties for Pakistani crude oils. J. Pet. Sci. Eng. 16 (4): 275-290. [http://dx.doi.org/10.1016/S0920-4105\(96\)00042-3](http://dx.doi.org/10.1016/S0920-4105(96)00042-3)

Robertson, C.J. 1983. Comparison of Revised PVT Properties with Published Correlations. Internal Report, Marathon Oil Company, Houston, Texas (April 1983).

Al-Fattah, S.M. and Al-Marhoun, M.A. 1994. Evaluation of empirical correlations for bubblepoint oil formation volume factor. J. Pet. Sci. Eng. 11 (4): 341-350.

GeoMark Research. 2003. RFDbase (Reservoir Fluid Database), <http://www.RFDbase.co>