

# Ανοικτό Πανεπιστήμιο Κύπρου

Σχολή Θετικών και Εφαρμοσμένων Επιστημών

Μεταπτυχιακό Πρόγραμμα Σπουδών

*Πληροφοριακά Συστήματα*

Μεταπτυχιακή Διατριβή



Εφαρμογή Μοντέλων Ανακάλυψης Γνώσης στη Βάση  
Δεδομένων DACUS, του Υπουργείου Αγροτικής Ανάπτυξης  
και Τροφίμων της Ελλάδας

Ανδρέας Καλογερόπουλος

Επιβλέπων Καθηγητής  
Ιωάννης Κατάκης

**Μάιος 2016**

# **Ανοικτό Πανεπιστήμιο Κύπρου**

**Σχολή Θετικών και Εφαρμοσμένων Επιστημών**

**Μεταπτυχιακό Πρόγραμμα Σπουδών**

***Πληροφοριακά Συστήματα***

## **Μεταπτυχιακή Διατριβή**

**Εφαρμογή Μοντέλων Ανακάλυψης Γνώσης στη Βάση  
Δεδομένων DACUS, του Υπουργείου Αγροτικής Ανάπτυξης  
και Τροφίμων της Ελλάδας**

**Ανδρέας Καλογερόπουλος**

**Επιβλέπων Καθηγητής  
Ιωάννης Κατάκης**

Η παρούσα μεταπτυχιακή διατριβή υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων για απόκτηση μεταπτυχιακού τίτλου σπουδών στα Πληροφοριακά Συστήματα από τη Σχολή Θετικών και Εφαρμοσμένων Επιστημών του Ανοικτού Πανεπιστημίου Κύπρου.

**Μάιος 2016**

ΛΕΥΚΗ ΣΕΛΙΔΑ

## Περίληψη

Στην εν λόγω Διατριβή επεξεργαζόμαστε δεδομένα από το ετήσιο Πρόγραμμα Δακοκτονίας, το οποίο υλοποιείται στην Ελλάδα από τις Διευθύνσεις Αγροτικής Οικονομίας και Κτηνιατρικής (ΔΑΟΚ) των Περιφερειακών Ενοτήτων (πρώην Νομαρχιών). Τα στοιχεία ελήφθησαν από τη Βάση Δεδομένων DACUS, του Υπουργείου Αγροτικής Ανάπτυξης και Τροφίμων, η οποία χρησιμοποιείται για την ηλεκτρονική καταγραφή των στοιχείων και από το φυσικό αρχείο της ΔΑΟΚ Χαλκιδικής.

Ο δάκος (*Bactocera oleae*) είναι ένα είδος μύγας που προσβάλλει τον καρπό της ελιάς και προκαλεί υποβάθμιση της ποιότητάς του. Πρόκειται δε για τον κυριότερο εχθρό των ελαιοκαλλιιεργειών. Η υλοποίηση του προγράμματος δακοκτονίας γίνεται κάθε χρόνο από Μάιο μέχρι Νοέμβριο.

Τα δεδομένα αφορούν σε μετρήσεις του πληθυσμού των ατόμων του δάκου που συλλαμβάνονται σε παγίδες, σε μετρήσεις του μεγέθους της προσβολής και στις επεμβάσεις (ψεκασμοί εδάφους) που πραγματοποιούνται με εντομοκτόνα για τον περιορισμό του πληθυσμού του δάκου. Χρησιμοποιήθηκαν επίσης μετεωρολογικά δεδομένα που ελήφθησαν από το Εθνικό Αστεροσκοπείο Αθηνών.

Τα δεδομένα από φυσικό αρχείο ψηφιοποιήθηκαν και όλα μαζί οργανώθηκαν σε Βάση Δεδομένων που κατασκευάστηκε. Με τη χρήση της γλώσσας στατιστικών υπολογισμών R διαμορφώθηκαν κατάλληλοι πίνακες (data frames) και διερευνήθηκε ο τρόπος αλληλεπίδρασης μεταξύ των διαφόρων παραγόντων (πληθυσμός δάκου, προσβολές, ψεκασμοί, μετεωρολογικά στοιχεία). Επίσης δημιουργήθηκαν, αυτόματα, διαγράμματα που απεικονίζουν τη μεταβολή του δακοπληθυσμού σε συνδυασμό με άλλους παράγοντες, σε κάθε δήμο και έτος. Με τη χρήση του Rattle, μιας εφαρμογής της R για την εξόρυξη δεδομένων, αξιολογήθηκαν δύο μοντέλα κατηγοριοποίησης (Δέντρο Απόφασης και Μηχανή Διανυσμάτων Υποστήριξης) για την πρόγνωση της μεταβολής του δακοπληθυσμού.

Σκοπός της έρευνας είναι η συμβολή σε μία μελλοντική αξιοποίηση των δεδομένων της βάσης DACUS προκειμένου να αναπτυχθούν μοντέλα που θα οδηγήσουν σε μία ορθολογική εφαρμογή των χημικών επεμβάσεων, έτσι ώστε να επιτευχθεί η καλύτερη δυνατή ποιότητα του παραγόμενου προϊόντος με τη μικρότερη δυνατή οικονομική και περιβαλλοντική επιβάρυνση.

## Summary

In this Master Thesis we process data of the annual Greek National Project Against the Olive Fruit Fly which runs by the local Directorates of Rural Economy and Veterinary (DAOK) of the Regional Units (former Prefectures). Data have been obtained by both DACUS, a data base owned by the Hellenic Ministry of Rural Development and Food, which is used for the electronic recording of this kind of data and DAOK of Chalkidiki archive.

Olive fruit fly (*Bactocera [Dacus] oleae*) is an insect pest of olive fruit that impairs its qualities. It is the Number One olive growing pest. Greek National Project Against the Olive Fruit Fly runs annually between May and November.

Data have to do with monitoring of flies' population (using traps), measurement of fruit damage and insecticide treatments for population control. Weather data acquired from National Observatory of Athens were also used.

Data form DAOK of Chalkidiki archive were digitalized and a database was organized to for the whole of data. Using R, a language for statistical computing, necessary tables (data frames) were constructed to investigate how several factors (fly population, fruit damage, insecticide treatments, weather conditions) interact with each other. Plots depicting flies' population fluctuation in combination with other factors were also constructed, automatically, for each year and municipality. Using Rattle, a GUI for data mining based on R, two classifiers (Decision Tree and Support Vector Machine) were evaluated for the prediction of flies' population fluctuation.

Aim of the research is to contribute to a future use of the DACUS database in order to develop models leading to a rational insecticide treatment, in order to achieve the optimal quality of the product obtained, along with the minimum economic and environmental impact.

## Ευχαριστίες

Με την εργασία αυτή ολοκληρώνεται το μεταπτυχιακό μου στα Πληροφοριακά Συστήματα του Ανοικτού Πανεπιστημίου Κύπρου που άρχισε το 2009. Θα ήθελα καταρχήν να την αφιερώσω στη μνήμη του πατέρα μου Θεόδωρου, επειδή το μεταπτυχιακό άρχισε την εποχή αμέσως μετά το θάνατό του και δεν έμαθε ποτέ γι αυτό. Την ίδια εποχή άρχισα να περιμένω και το πρώτο μου παιδί. Την αφιερώνω λοιπόν και στις κόρες μου – πηγή έμπνευσης για ότι κάνω – Θεοδώρα και Αθηνά και στη γυναίκα μου Ιωάννα για την υπομονή που είχαν όποτε απέσπασα την αφοσίωσή μου σε αυτές για να ασχοληθώ με το μεταπτυχιακό. Φυσικά και στη μάνα μου Ευγενία για την οικονομική ενίσχυση του μεταπτυχιακού αλλά κυρίως γιατί έχει πάντα τον τρόπο της να είναι εκεί για να διευκολύνει τις καταστάσεις και στον αδερφό μου το Γιώργο γιατί έχει βάλει άθελά του τις δικές του πλάτες.

Πάμε τώρα στα πιο πρακτικά. Ευχαριστώ ιδιαίτερα τον επιβλέποντα καθηγητή κ. Ιωάννη Κατάκη, ο οποίος δέχτηκε να αναλάβει την επίβλεψη της εργασίας με ένα θέμα που του πρότεινα και αγκάλιασε ζεστά όλη μου την προσπάθεια στα πρώτα στάδια, καθώς και για την πολύτιμη βοήθειά του καθόλη τη διάρκεια της πραγματοποίησης της εργασίας. Τον Ακαδημαϊκό Υπεύθυνο κ. Μιχαήλ Λοΐζο για τη συμβολή του κατά την αναζήτηση των δεδομένων από το ΥΠΑΑΤ. Την κ. Άννα Κατσάρα, από τη Δ/νση Ηλεκτρονικής Διακυβέρνησης του ΥΠΑΑΤ για το χρόνο που αφιέρωσε για να μου δείξει τη Βάση Δεδομένων DACUS και για να εξάγει και να μου στείλει τους πίνακες. Τις συναδέλφους γεωπόνους Δήμητρα Γκιλπάθη και Εμμανουέλα Καπόγια από τη Δ/νση Προστασίας Φυτικής Παραγωγής στην οποία ανήκει η ΒΔ DACUS, το συνάδελφο Γιώργο Κατσικογιάννη από τη ΔΑΟΚ Σάμου και τον κ. Κωνσταντίνο Λαγουβάρδο από το Εθνικό Αστεροσκοπείο Αθηνών. Ιδιαίτερες ευχαριστίες στους συναδέλφους Κώστα Τερτιβανίδη και Μανώλη Χαριζόπουλο από τη ΔΑΟΚ Χαλκιδικής, που με τόση προθυμία δέχτηκαν να μου ταχυδρομήσουν τις Εκθέσεις Δακοκτονίας. Ευχαριστώ επίσης το Δημήτρη Κοντοδήμα, ερευνητή εντομολόγο του Μπενάκειου Φυτοπαθολογικού Ινστιτούτου για το χρόνο που αφιέρωσε και τις κατευθύνσεις που μου έδωσε σε μία περίοδο που δεν ήξερα πώς να χειριστώ τα αποτελέσματα από την επεξεργασία των δεδομένων.

Τέλος ένα μεγάλο ευχαριστώ σε όλους τους συναδέλφους μου, πρώην και νυν, στην Περιφέρεια Αττικής (δεν αναφέρω ονόματα για να μην ξεχάσω κανέναν), για την ηθική υποστήριξη που μου προσέφεραν όλα αυτά τα χρόνια που πραγματοποιούσα το μεταπτυχιακό, η οποία δεν είναι πάντα δεδομένη σε χώρους εργασίας.

# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b> .....	<b>1</b>
1.1	Το Πρόγραμμα Συλλογικής Καταπολέμησης του Δάκου του Υπουργείου Αγροτικής Ανάπτυξης και Τροφίμων .....	1
1.1.1	Η Ελιά και ο Δάκος της Ελιάς .....	1
1.1.2	Η Αντιμετώπιση του Δάκου από το Ελληνικό Κράτος .....	3
1.1.3	Μέτρηση του Πληθυσμού των Δάκων .....	4
1.1.4	Δειγματοληψία ελαιοκάρπου .....	5
1.1.5	Διενέργεια Ψεκασμών .....	5
1.2	Ολοκληρωμένη Διαχείριση Καλλιεργειών .....	6
<b>2</b>	<b>Εξόρυξη Δεδομένων</b> .....	<b>8</b>
2.1	Τι είναι η Εξόρυξη Δεδομένων .....	8
2.2	Δεδομένα .....	10
2.2.1	Τύποι Δεδομένων .....	10
2.2.2	Ποιότητα Δεδομένων .....	11
2.2.3	Προεπεξεργασία Δεδομένων .....	11
2.2.4	Μέτρα Ομοιότητας και Ανομοιότητας .....	12
2.3	Κατηγοριοποίηση (classification) .....	13
2.3.1	Τι είναι και πώς λειτουργούν τα μοντέλα κατηγοριοποίησης .....	13
2.3.2	Υπερπροσαρμογή μοντέλου .....	14
2.3.3	Δένδρα απόφασης (decision trees) .....	15
2.3.4	Μηχανές διανυσμάτων υποστήριξης (support vector machines) .....	16
2.3.5	Άλλοι κατηγοριοποιητές .....	18
2.4	Άλλες τεχνικές Εξόρυξης Δεδομένων .....	19

<b>3</b>	<b>Η Ανάλυση Δεδομένων στη φυτοπροστασία</b> .....	<b>20</b>
3.1	Εφαρμογή τεχνικών Εξόρυξης Δεδομένων στη φυτοπροστασία .....	20
3.2	Η Ανάλυση δεδομένων στην αντιμετώπιση του Δάκου της ελιάς .....	21
<b>4</b>	<b>Περιγραφή Και Συλλογή Δεδομένων</b> .....	<b>23</b>
4.1	Πίνακες από τη Βάση Δεδομένων του ΥΠΑΑΤ .....	23
4.1.1	Χορήγηση Πινάκων από το Υπουργείο .....	23
4.1.1.1	Πίνακας ΔΗΜΟΙ .....	23
4.1.1.2	Πίνακας ΜΕΤΡΗΣΗ_ΔΑΚΩΝ .....	24
4.1.1.3	Πίνακας ΔΕΙΓΜΑΤΟΛΗΨΙΑ .....	25
4.1.1.4	Πίνακας ΨΕΚΑΣΜΟΙ .....	26
4.1.1.5	Άλλοι Βοηθητικοί Πίνακες .....	26
4.1.1.6	Πίνακας ΚΑΡΠΟΦΟΡΙΑ .....	27
4.1.2	Δημιουργία Βάσης Δεδομένων .....	27
4.1.2.1	Πίνακες .....	27
4.1.2.2	Σχέσεις Μεταξύ Πινάκων .....	28
4.1.2.3	Καταχώρηση Δεδομένων .....	28
4.1.2.4	Δημιουργία Ερωτημάτων .....	29
4.2	Αρχείο της Διεύθυνσης Αγροτικής Οικονομίας και Κτηνιατρικής Π.Ε. Χαλκιδικής ...	30
4.2.1	Πρόσβαση στο Αρχείο .....	30
4.2.2	Περιεχόμενο Εκθέσεων Δακοκτονίας .....	31
4.2.2.1	Μετρήσεις Δακοπληθυσμού .....	31
4.2.2.2	Στοιχεία Δειγματοληψιών για τη Μέτρηση της Δακοπροσβολής .....	33
4.2.2.3	Στοιχεία για τους Διενεργηθέντες Ψεκασμούς .....	33
4.2.3	Καταγραφή Στοιχείων για Καταχώρησή τους στη Βάση Δεδομένων .....	34
4.2.3.1	Μετρήσεις Δακοπληθυσμού .....	34
4.2.3.2	Στοιχεία Δειγματοληψιών για τη Μέτρηση της Δακοπροσβολής .....	35
4.2.3.3	Στοιχεία για τους Διενεργηθέντες Ψεκασμούς .....	36
4.3	Μετεωρολογικά Δεδομένα .....	37
4.3.1	Στοιχεία από το Εθνικό Αστεροσκοπείο Αθηνών .....	37
4.3.2	Καταχώρηση Στοιχείων στη Βάση Δεδομένων .....	39
4.3.3	Συσχέτιση του Πίνακα ΜΕΤΕΩΡΟΛΟΓΙΚΑ ΣΤΟΙΧΕΙΑ με άλλους Πίνακες της Βάσης Δεδομένων .....	40
4.3.4	Δημιουργία νέου ερωτήματος και συμπλήρωση πεδίων στα ήδη υπάρχοντα .....	42



<b>5</b>	<b>Προεπεξεργασία, Ανάλυση Δεδομένων και Συμπεράσματα</b> .....	<b>43</b>
5.1	Επεξεργασία Πινάκων .....	43
5.1.1	Δημιουργία και Επεξεργασία των data frames .....	43
5.1.1.1	Μετατροπές και προσθήκες σχετικές με τις ημερομηνίες .....	44
5.1.1.2	Προσθήκες στηλών με τους μέσους όρους δάκων ανά παγίδα .....	44
5.1.1.3	Συναρθροίσεις (aggregations) δεδομένων .....	45
5.2.	Κατασκευή διαγραμμάτων (plots) που απεικονίζουν τις μεταβολές του δακοπληθυσμού στο χρόνο .....	46
5.2.1.	Πληθυσμός δάκου σε συνδυασμό με τη δακοπροσβολή και τις ημερομηνίες των ψεκασμών .....	47
5.2.1.1.	Παρατηρήσεις επί των διαγραμμάτων .....	48
5.2.2.	Πληθυσμός δάκου σε συνδυασμό με τους τύπους της δακοπροσβολής αναλυτικά .	48
5.2.2.1.	Παρατηρήσεις επί των διαγραμμάτων .....	50
5.2.3.	Πολλαπλά διαγράμματα με πληθυσμό δάκου (σύνολο, θηλυκά, αρσενικά, ποσοστό θηλυκών) μαζί με ψεκασμούς, γόνιμες προσβολές και μετεωρολογικά στοιχεία .....	51
5.2.3.1.	Παρατηρήσεις επί των διαγραμμάτων .....	53
5.3.	Εκτίμηση της συσχέτισης μεταξύ πληθυσμού δάκων και προσβολών στους καρπούς .....	53
5.3.1.	Εκτίμηση της συσχέτισης χωρίς τη χρήση των μετεωρολογικών στοιχείων .....	54
5.3.1.1.	Παρατηρήσεις επί των αποτελεσμάτων .....	57
5.3.2.	Εκτίμηση της συσχέτισης συνεκτιμώντας και τα μετεωρολογικά στοιχεία .....	58
5.3.2.1.	Παρατηρήσεις επί των αποτελεσμάτων .....	59
5.4.	Δημιουργία Ενιαίου Πίνακα Μετρήσεων Δάκου – Ψεκασμών – Μετεωρολογικών Δεδομένων .....	60
5.4.1.	Κατασκευή του πίνακα .....	60
5.4.2.	Εκτίμηση της επιρροής του ποσοστού θηλυκών, των μετεωρολογικών μεταβλητών και των ψεκασμών στη μεταβολή του πληθυσμού .....	61
5.4.2.1.	Εκτίμηση της συσχέτισης μεταξύ πληθυσμού και ποσοστού θηλυκών .....	61
5.4.2.2.	Εκτίμηση της επιρροής της διενέργειας ψεκασμών και των καιρικών φαινομένων στη μεταβολή του δακοπληθυσμού, με τον υπολογισμό του μέσου όρου .....	63
5.5.	Πρόγνωση της μεταβολής του πληθυσμού με την τεχνική της Κατηγοριοποίησης ..	65
5.5.1.	Κατηγοριοποίηση σε 3 κλάσεις. Δέντρο Απόφασης .....	66
5.5.2.	Κατηγοριοποίηση σε 2 κλάσεις. Δέντρο Απόφασης .....	68
5.5.3.	Σύγκριση της κατηγοριοποίησης σε δύο και σε τρεις κλάσεις .....	70
5.5.4.	Κατηγοριοποίηση σε 3 κλάσεις, ξεχωριστά για κάθε νομό. Δέντρο Απόφασης .....	70

5.5.5. Κατηγοριοποίηση σε 3 κλάσεις, Support Vector Machine (Μηχανή Διανυσμάτων Υποστήριξης) .....	73
5.5.6. Σύγκριση όλων των μοντέλων και των πινάκων .....	73
<b>6 Επίλογος .....</b>	<b>75</b>
<b>Βιβλιογραφία .....</b>	<b>77</b>

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Το Πρόγραμμα Συλλογικής Καταπολέμησης του Δάκου του Υπουργείου Αγροτικής Ανάπτυξης και Τροφίμων

#### 1.1.1 Η ελιά και ο Δάκος της ελιάς

Η **ελιά** (*Olea europaea*) είναι καρποφόρο δέντρο της οικογένειας των ελαιοειδών (Oleaceae), του οποίου η καλλιέργεια είναι γνωστή στη Μεσόγειο και την Ανατολική Μεσόγειο ειδικότερα, από την προϊστορική εποχή. Από τον καρπό της, παράγεται το ελαιόλαδο, το μοναδικό έλαιο που εξάγεται με μηχανική επεξεργασία, βασικό συστατικό της



Μεσογειακής Διατροφής και γνωστό για τις πολύτιμες οργανοληπτικές του ιδιότητες. Ο ίδιος ο ελαιόκαρπος καταναλώνεται και ως επιτραπέζιος.

Η ελιά καλλιεργείται στις περισσότερες περιοχές της Ελλάδας, κυρίως όμως στις θερμότερες και ξηρότερες περιοχές (Κ. Ποντίκης, 1992). Όσο

αφορά στο βιολογικό της κύκλο, **η άνθηση και η καρπόδεση** λαμβάνουν χώρα από το 3<sup>ο</sup> δεκαήμερο του Απριλίου έως τέλος Ιουνίου, ενώ η **ανάπτυξη του καρπού** από αρχές Ιουλίου μέχρι τα μέσα Νοεμβρίου, οπότε αρχίζει και η περίοδος συγκομιδής (Κ. Ποντίκης, 1992). Η Ελλάδα είναι η τρίτη ελαιοπαραγωγός χώρα στον κόσμο, ενώ το ελαιόλαδο και οι επιτραπέζιες ελιές είναι από τα κυριότερα εξαγωγικά της προϊόντα.

Η ελιά προσβάλλεται από μία σειρά από παράσιτα. Το σοβαρότερο από αυτά θεωρείται ο **Δάκος της Ελιάς**. Ο Δάκος (*Bactocera oleae*) είναι ένα είδος μύγας (τάξη Diptera). Ο βιολογικός του κύκλος, όπως και των περισσότερων εντόμων, αποτελείται από 4

στάδια: **ωό (αυγό)** – **προνύμφη (larva) (κάμπια)** – **νύμφη (pupa) (χρυσάλιδα)** – **ακμαίο (τέλειο)**.



**Εικόνα 1:** Στην παραπάνω εικόνα φαίνονται, με αντίστοιχη σειρά, τα τέσσερα στάδια ανάπτυξης του εντόμου. Παρατηρούμε το αυγό και την προνύμφη που βρίσκονται στο εσωτερικό του ελαιοκάρπου, με την προνύμφη να έχει φάει μέρος του μεσοκαρπίου.

Το θηλυκό ακμαίο εναποθέτει, με τον ωοθέτη του, στον καρπό της ελιάς ένα **ωό** (κατά κανόνα ένα ωό ανά καρπό). Στη θέση εναπόθεσης, στην επιφάνεια του καρπού, είναι εμφανές ένα **νύγμα**. Αν το αυγό δεν εκκολαφθεί, τότε μιλάμε για **άγονο νύγμα**. Αν το αυγό εκκολαφθεί, μετά από ένα αριθμό ημερών βγαίνει η **προνύμφη**, η οποία ορύσσει στοές τρώγοντας το μεσοκάρπιο του καρπού. Όταν η προνύμφη ολοκληρώσει την ανάπτυξή της, μετακινείται προς την επιφάνεια του καρπού, όπου διευρύνει τη στοά και μετατρέπεται σε **νύμφη**. Μετά από λίγες μέρες σπάει την εξωτερική μεμβράνη του καρπού (εφυμενίδα) και εξέρχεται ως **ακμαίο**. Η οπή την οποία ανοίγει για να εξέλθει είναι η χαρακτηριστική **οπή εξόδου**. Συχνά, στο νύγμα του δάκου αναπτύσσεται ο μύκητας *Macrophoma dalmatica*, ο οποίος προξενεί **δευτερογενείς προσβολές** στον καρπό και προκαλεί τα χαρακτηριστικά συμπτώματα **ξεροβούλα** (στις άγουρες ελιές) και **σαπιοβούλα** (στις ώριμες ελιές).



**Εικόνα 2:** Ξεροβούλα σε άγουρο καρπό ελιάς από **δευτερογενή προσβολή** του μύκητα *Macrophoma dalmatica*.

Η δραστηριότητα των ακμαίων και η διάρκεια του κάθε σταδίου ανάπτυξης εξαρτάται από την θερμοκρασία περιβάλλοντος και την εξωτερική σχετική υγρασία. Συνοπτικά, οι άριστες θερμοκρασίες για την εναπόθεση αυγών είναι μεταξύ 20° και 30°C και η άριστη σχετική υγρασία του ατμοσφαιρικού αέρα μεταξύ 70% και 100% (Εγκύκλιος του Υπουργείου Γεωργίας, 2002).

Ο Δάκος προξενεί σημαντικές οικονομικές **ζημιές** στην ελαιοπαραγωγή (βρώσιμες και ελαιοποιήσιμες ελιές). Οι ζημιές είναι **ποσοτικές** (πρόωρη πτώση του καρπού πριν τη συλλογή, μείωση σάρκας ελαιοκάρπου) και **ποιοτικές** (ποιοτική υποβάθμιση του ελαιολάδου λόγω αύξησης οξύτητας, αλλοίωση οργανοληπτικών ιδιοτήτων του ελαιολάδου, μείωση της εμπορικής αξίας – μέχρι και ακαταλληλότητα – των βρώσιμων ελιών). (Εγκύκλιος του Υπουργείου Γεωργίας, 2002)

Η αντιμετώπισή του στις συμβατικές καλλιέργειες γίνεται με χημικό τρόπο (εντομοκτόνα), είτε με **καθολικούς ψεκασμούς** (ψεκάζεται ολόκληρο το δέντρο) είτε με **δολωματικούς ψεκασμούς** (ψεκάζεται μέρος του δέντρου με εντομοκτόνο μαζί με κάποια ελκυστική ουσία).

### **1.1.2 Η αντιμετώπιση του Δάκου από το Ελληνικό Κράτος**

Λόγω της σοβαρότητας του δάκου ως κινδύνου για τις ελαιοκαλλιέργειες, ξεκίνησε ήδη από το 1953 το Πρόγραμμα Συλλογικής Καταπολέμησης του Δάκου (πρόγραμμα Δακοκτονίας) από το τότε Υπουργείο Γεωργίας και τα κατά τόπους Ταμεία Προστασίας Ελαιοπαραγωγής (ΤΠΕ) των ελαιοκομικών περιοχών τα οποία ιδρύθηκαν. Το 1996, με τη δημιουργία των Νομαρχιακών Αυτοδιοικήσεων, οι αρμοδιότητες των ΤΠΕ πέρασαν στις Νομαρχιακές Αυτοδιοικήσεις, ενώ από το 2011, με την εφαρμογή του νόμου Καλλικράτης, στις αντίστοιχες Περιφερειακές Ενότητες των Περιφερειών. Το Υπουργείο (τότε Γεωργίας και τώρα Αγροτικής Ανάπτυξης και Τροφίμων) αναλαμβάνει την Υποστήριξη και την Προμήθεια των Υλικών και οι κατά τόπους Διευθύνσεις Αγροτικής Οικονομίας και Κτηνιατρικής των Περιφερειακών Ενοτήτων αναλαμβάνουν την υλοποίηση.

Για να ενταχθεί μία περιοχή στο ετήσιο πρόγραμμα δακοκτονίας, πρέπει αυτό να ζητηθεί και να εγκριθεί και πρέπει το ποσοστό καρποφορίας να είναι 25% και 20% μιας πλήρους σοδειάς, για τις ελαιοποιήσιμες και βρώσιμες ελιές αντίστοιχα.

Προσλαμβάνεται και εποχιακό προσωπικό (γεωπόνοι, παγιδοθέτες, ψεκαστές, ενίοτε και εργολάβοι) από 15 Μαΐου έως 30 Νοεμβρίου κάθε έτους, η οποία είναι και η περίοδος διάρκειας του προγράμματος κάθε έτους.

(Τα στοιχεία του κειμένου αυτής της ενότητας προέρχονται από τη Σειρά Δαιφανειών με τίτλο: «*Πρόγραμμα Συλλογικής Καταπολέμησης του Δάκου*» που είναι αναρτημένη στον ιστότοπο του ΥΠΑΑΤ)

### 1.1.3 Μέτρηση του πληθυσμού των δάκων

Ο έλεγχος των δακοπληθυσμών γίνεται με γυάλινες παγίδες τύπου McPhail.



Εικόνα 3: Παγίδες McPhail.

Οι παγίδες γεμίζονται με 400-500 ml νερού με διαλυμένη 2% φωσφορική ή θειική αμμωνία το καλοκαίρι ή πρωτεΐνη και βόρακα 4% και 1,5% αντίστοιχα το φθινόπωρο, ως ελκυστική ουσία, για να συλλαμβάνονται τα ακμαία του εντόμου. (Εγκύκλιος του Υπουργείου Γεωργίας, 2002)

Η ανάρτηση των παγίδων γίνεται στην αρχή κάθε περιόδου. Τοποθετείται περίπου **μία παγίδα ανά 1000 δέντρα**. Οι παγίδες τοποθετούνται στο εσωτερικό της κόμης του δέντρου. (Εγκύκλιος του Υπουργείου Γεωργίας, 2002)

Οι μετρήσεις γίνονται ανά πενήνήμερο, στις ημερομηνίες 1-2, 6-7, 11-12, 16-17, 21-22, 26-27 του κάθε μήνα (δηλαδή 6 μετρήσεις το μήνα). Σε κάθε μέτρηση καταγράφεται η ημερομηνία και ο αριθμός των «δακοσυλλήψεων» (των ατόμων που παγιδεύτηκαν) και ειδικότερα ο αριθμός αρσενικών και θηλυκών ατόμων. Τα θηλυκά ξεχωρίζουν από τα αρσενικά λόγω της ύπαρξης του ωοθήτη (είναι το όργανο που μοιάζει με κεντρί, με το οποίο εναποθέτει τα αυγά). Αν υπάρχουν τα κατάλληλα όργανα, καταγράφεται και η θερμοκρασία και η υγρασία.



Εικόνα 4: Αρσενικό (αριστερά) και θηλυκό (δεξιά) ακμαίο. Είναι ευδιάκριτος ο ωοθήτης στο πίσω μέρος (κοιλία) του θηλυκού.

Στο τέλος κάθε μέτρησης, η κάθε παγίδα αδειάζει και ξαναγεμίζει με καινούργιο διάλυμα.

#### 1.1.4 Δειγματοληψία ελαιοκάρπου

Σε τακτά χρονικά διαστήματα (συνήθως μία φορά στο τέλος κάθε μήνα για τις περιπτώσεις που εξετάζονται στην παρούσα Διατριβή), λαμβάνουν χώρα δειγματοληψίες ελαιοκάρπων για τον προσδιορισμό της υπάρχουσας δακοπροσβολής. Από κάθε κέντρο δειγματοληψίας (10000 δέντρων περίπου), επιλέγονται κατά τύχη **40 δέντρα** διάσπαρτα και κατανεμημένα σε όλη την έκτασή του. Από **κάθε δέντρο** λαμβάνονται **24 καρποί** από 5 θέσεις: 8 από την κορυφή και 16 από τις ποδιές (4 καρποί ανά σημείο του ορίζοντα). Έτσι συνολικά συγκεντρώνονται  $40 \times 24 = 960$  **καρποί για κάθε κέντρο δειγματοληψίας**.

Κατά την εξέταση των καρπών του δείγματος, οι καρποί διαχωρίζονται σε υγιείς και προσβεβλημένους και οι προσβεβλημένοι σε έχοντες **άγωνα νύγματα, αυγά, προνύμφες 1<sup>ου</sup>, 2<sup>ου</sup> και 3<sup>ου</sup> σταδίου, νύμφες, οπές εξόδου και παρασιτισμό (δευτερογενή προσβολή από το μύκητα)** (βλ. παράγραφο 1.1.1.) κι έτσι υπολογίζεται το % ποσοστό της συνολικής προσβολής και των επί μέρους σταδίων της. Το άθροισμα ζώντα αυγά + ζώσες προνύμφες + ζώσες νύμφες + έξοδοι αποτελεί τη **γόνιμη προσβολή**.

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από: Εγκύκλιος του Υπουργείου Γεωργίας, 2002)

#### 1.1.5 Διενέργεια ψεκασμών

Οι ψεκασμοί διακρίνονται σε **θεραπευτικούς** και **προληπτικούς (δολωματικούς)**.

Σκοπός των **θεραπευτικών** είναι η θανάτωση τόσο των ακμαίων εντόμων, όσο και των διαφόρων σταδίων προνύμφης που βρίσκονται μέσα στον ελαιοκάρπο. Εφαρμόζονται ι) **όταν δεν ελέγχεται ο πληθυσμός** με δολωματικούς ψεκασμούς και ιι) όταν το ποσοστό δακοπροσβολής είναι μεγαλύτερο από 5% στις ελιές που προορίζονται για ελαιόλαδο ή 2% σε αυτές που προορίζονται για βρώσιμες.

Συνήθως οι ψεκασμοί που διενεργούνται είναι **δολωματικοί**. Σε αυτή την περίπτωση ψεκάζεται μια μικρή περιοχή στο εσωτερικό του δέντρου, είτε σε κάθε δέντρο είτε ανά δύο ή τρία. Το ψεκαστικό υγρό περιέχει, εκτός από το εντομοκτόνο, **ελκυστική ουσία**. Σκοπός είναι η προσέλκυση και θανάτωση των ακμαίων πριν αρχίσει η εναπόθεση αυγών στον ελαιοκάρπο. Οι δολωματικοί ψεκασμοί δεν γίνονται όταν η θερμοκρασία υπερβαίνει τους 28°C (γιαυτό και αρχίζουν νωρίς τι πρωί), όταν η σχετική υγρασία είναι

μικρότερη από 25% ή όταν πρόκειται να βρέξει. **Οι ισχυρές βροχοπτώσεις που οδηγούν σε απόπλυση του δολώματος επιβάλλουν επανάληψη του ψεκασμού.**

Ο προσδιορισμός του χρόνου του **πρώτου ψεκασμού** εξαρτάται από τις εξής παραμέτρους:

- Ύψος δακοπληθυσμού (5-20 δάκοι ανά παγίδα και πενήμερο)
- Αναλογία θηλυκών/αρσενικών (τουλάχιστον 1)
- Ποσοστό γόνιμων θηλυκών (τουλάχιστον 5%) (σημ.: όπως φαίνεται παρακάτω, δεν δίνεται πουθενά τέτοιο στοιχείο στα δεδομένα που διαθέτουμε)
- Κατάσταση ελαιοκάρπου (μέσο βάρος καρπού >0,20gr και έναρξη πήξης του πυρήνα)
- Κλιματολογικές συνθήκες (θερμοκρασία – υγρασία)

Για τον προσδιορισμό του χρόνου εφαρμογής των **επόμενων ψεκασμών** λαμβάνονται υπόψη το **ύψος του δακοπληθυσμού**, η **αναλογία θηλυκών/αρσενικά**, το **ποσοστό γόνιμων θηλυκών**, οι **καιρικές συνθήκες**, καθώς και τα **αποτελέσματα των δειγματοληψιών**, η **ανοδική ή καθοδική τάση των πληθυσμών** και η **τοπική εμπειρία**.

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από: Εγκύκλιος του Υπουργείου Γεωργίας, 2002)

## **1.2 Ολοκληρωμένη Διαχείριση Καλλιεργειών**

Οι σύγχρονες ανάγκες στην πρωτογενή παραγωγή και στην παραγωγή γενικότερα, απαιτούν την αύξηση της απόδοσης αλλά και την παραγωγή προϊόντων υψηλής ποιότητας από τη μία πλευρά, με τη μικρότερη όμως οικονομική και περιβαλλοντική επιβάρυνση από την άλλη. Η Ολοκληρωμένη Διαχείριση Καλλιεργειών έχει σκοπό την ορθολογική χρήση των εισροών (πχ. λιπάσματα, νερό, φυτοφάρμακα), στα πλαίσια ενός γενικότερου καλλιεργητικού πλάνου, ώστε αυτές να περιοριστούν κατά το δυνατόν.

Η φυτοπροστασία είναι ο κλάδος της Γεωπονίας που ασχολείται με τη μελέτη και αντιμετώπιση των παραγόντων που δρουν περιοριστικά στην ανάπτυξη των φυτών (ασθένειες, εχθροί, ζιζάνια). Δεδομένων όλων των παραπάνω, η φυτοπροστασία στρέφεται, στα πλαίσια ενός συνδυασμού καλλιεργητικών, βιολογικών και χημικών μέσων, στον περιορισμό χρήσης των φυτοφαρμάκων και στην ορθολογική χρήση τους.



Αυτό απαιτεί πολύ καλή γνώση όλων των χαρακτηριστικών που επηρεάζουν την πορεία της καλλιέργειας. **Η εφαρμογή της Πληροφορικής επιστήμης για την ανάλυση όλων αυτών των δεδομένων είναι καθοριστική και αναγκαία.**

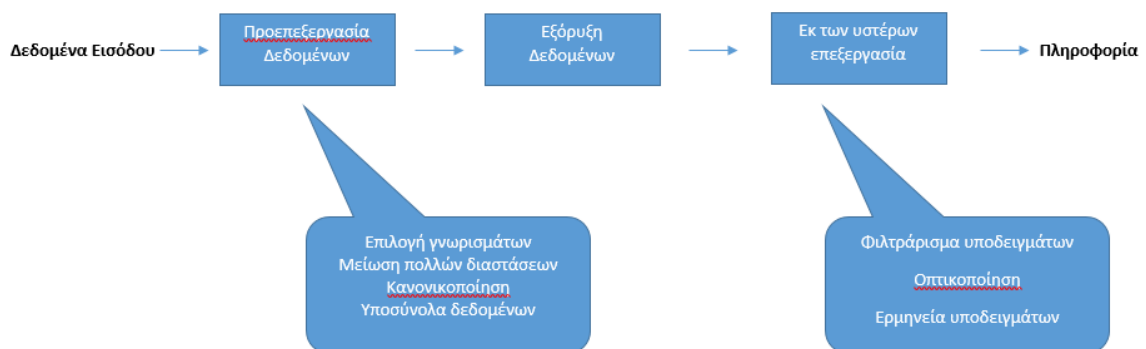
# Κεφάλαιο 2

## Εξόρυξη Δεδομένων

### 2.1 Τι είναι η Εξόρυξη Δεδομένων

Εξόρυξη Δεδομένων είναι η διαδικασία της αυτόματης ανακάλυψης χρήσιμων πληροφοριών μέσα από μεγάλες δεξαμενές δεδομένων (Tan et al, 2006). Χρησιμοποιώντας τις τεχνικές της Εξόρυξης Δεδομένων μπορούμε να εξάγουμε πρότυπα που εξηγούν τις σχέσεις που διέπουν τα δεδομένα σε μεγάλους όγκους δεδομένων ή ακόμα και να προγνώσουμε την τιμή κάποιου χαρακτηριστικού σε νέα δεδομένα, δεδομένης της τιμής των υπολοίπων.

Αποτελεί μέρος της Ανακάλυψης Γνώσης από τις Βάσεις Δεδομένων, η οποία είναι η συνολική διεργασία της μετατροπής ακατέργαστων δεδομένων σε σημαντικές πληροφορίες.



**Σχήμα 1:** Η διεργασία της ανακάλυψης γνώσης από τις Βάσεις Δεδομένων (Tan et al, 2006)

Όπως φαίνεται και στο Σχήμα 1, τα κύρια στάδια της διαδικασίας ανακάλυψης γνώσης από τις Βάσεις Δεδομένων είναι η **Προεπεξεργασία (pre-processing)**, η **Εξόρυξη Δεδομένων (data mining)** και η **εκ των υστέρων επεξεργασία (post-processing)**.

Με την **προεπεξεργασία** διαμορφώνουμε τα δεδομένα εισόδου έτσι ώστε να είναι κατάλληλα για ανάλυση με μεθόδους εξόρυξης δεδομένων. Κατά την ενσωμάτωση των αποτελεσμάτων της εξόρυξης δεδομένων σε συστήματα λήψης αποφάσεων, τα δεδομένα υφίστανται μία **εκ των υστέρων επεξεργασία** ώστε μόνο τα έγκυρα και χρήσιμα αποτελέσματα να ενσωματωθούν στο εκάστοτε σύστημα λήψης αποφάσεων.

Οι εργασίες της **εξόρυξης δεδομένων**, χωρίζονται γενικά σε δύο βασικές κατηγορίες:

- **Προγνωστικές εργασίες (predictive tasks):** Στόχος τους είναι η πρόβλεψη της τιμής ενός συγκεκριμένου χαρακτηριστικού, βασιζόμενες στις τιμές των υπόλοιπων χαρακτηριστικών. Το χαρακτηριστικό που θέλουμε να προβλέψουμε λέγεται **στόχος (target)** ή **εξαρτημένη μεταβλητή (dependent variable)**. Μοντέλα που χρησιμοποιούνται για την πρόγνωση είναι η **κατηγοριοποίηση (classification)** για διακριτές μεταβλητές στόχους και η **παλινδρόμηση (regression)** για συνεχείς μεταβλητές στόχους.
- **Περιγραφικές εργασίες (descriptive tasks):** Στόχος τους είναι να εξάγουν πρότυπα που εξηγούν τις βασικές σχέσεις που διέπουν τα δεδομένα. Μοντέλα που χρησιμοποιούνται για περιγραφή δεδομένων είναι η **ανάλυση συσχέτισης (association analysis)**, η **ανάλυση συστάδων (cluster analysis)** και η **ανάλυση ανωμαλιών (anomaly detection)**.

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

## 2.2 Δεδομένα

### 2.2.1 Τύποι Δεδομένων

Ένα **σύνολο δεδομένων (dataset)** μπορεί να θεωρηθεί ως μια συλλογή **αντικειμένων δεδομένων (data objects)**. Ένα αντικείμενο είναι πχ. μία εγγραφή μιας βάσης δεδομένων ή αλλιώς μια γραμμή ενός πίνακα δεδομένων. Τα αντικείμενα περιγράφονται από ένα πλήθος **χαρακτηριστικών (attributes)**. Ένα χαρακτηριστικό είναι μια ιδιότητα ή γνώρισμα ενός αντικειμένου.

Τα **δεδομένα** μπορεί να είναι είτε **πειραματικά** είτε **δεδομένα παρατηρήσεων**.

Τα χαρακτηριστικά μπορεί να είναι:

- **Διακριτά (discrete)**: έχουν πεπερασμένο ή αριθμήσιμα άπειρο σύνολο τιμών. Ειδική κατηγορία των διακριτών είναι τα **δυναδικά (binary)**
- **Συνεχή (continuous)**: Το σύνολο τιμών τους είναι μη αριθμήσιμα άπειρο

**Ασύμμετρα χαρακτηριστικά (asymmetric attributes)** είναι αυτά για τα οποία επικεντρωνόμαστε στο αν η τιμή τους είναι μηδενική ή μη μηδενική.

Χαρακτηριστικά συνόλων δεδομένων είναι:

- Η **διάσταση**: Είναι το πλήθος των χαρακτηριστικών που περιέχουν τα αντικείμενα του συνόλου
- Η **σποραδικότητα**: Σποραδικά είναι για παράδειγμα τα σύνολα με ασύμμετρα χαρακτηριστικά, όπου τα περισσότερα χαρακτηριστικά έχουν μηδενική τιμή και ένας μικρός αριθμός χαρακτηριστικών έχει μη μηδενικές τιμές.
- Η **ανάλυση**: Διαφορετικός βαθμός ανάλυσης των δεδομένων μπορεί να κάνει ορατά διαφορετικά πρότυπα

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

### 2.2.2 Ποιότητα Δεδομένων

Για ποικίλους λόγους, τα δεδομένα που αποκτούμε δεν είναι ποτέ σε τέλεια μορφή για επεξεργασία. Διάφορα προβλήματα που παρουσιάζουν είναι, μεταξύ άλλων, τα εξής:

- **Θόρυβος:** Πρόκειται για διαστρέβλωση τιμών ή για προσθήκη μη αληθινών αντικειμένων. Όσο περισσότερος είναι ο θόρυβος, τόσο δυσκολότερος, έως και αδύνατος, ο εντοπισμός των προτύπων που διέπουν τα δεδομένα.
- **Ορθότητα:** Είναι η εγγύτητα των μετρήσεων με την πραγματική τιμή της ποσότητας που μετριέται (Tan et al, 2006). Η ορθότητα εξαρτάται από την **ακρίβεια** και τη **μεροληψία**
- **Ακραίες τιμές, αγνοούμενες τιμές, μη συνεπείς τιμές, διπλότυπα δεδομένα**

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

### 2.2.3 Προεπεξεργασία Δεδομένων

Τα βήματα που μπορεί να ακολουθηθούν κατά την προεπεξεργασία των δεδομένων είναι τα εξής:

- **Συνάθροιση (aggregation):** Είναι η σύνοψη δύο ή περισσότερων χαρακτηριστικών σε ένα.
- **Δειγματοληψία (sampling):** Όταν η επεξεργασία ολόκληρου του συνόλου των δεδομένων είναι ακριβή ή χρονοβόρα
- **Μείωση των πολλών διαστάσεων (dimensionality reduction):** Οδηγεί σε καλύτερη κατανόηση του μοντέλου και αντιμετωπίζει την κατάρα των πολλών διαστάσεων
- **Επιλογή υποσυνόλου γνωρισμάτων (feature subset selection)**
- **Δημιουργία γνωρισμάτων (feature creation)**
- **Διακριτοποίηση και δυαδικοποίηση δεδομένων (data discretization and binarization):** Μετατροπή ενός συνεχούς χαρακτηριστικού σε κατηγορικό, ή μετατροπή ενός συνεχούς ή διακριτού χαρακτηριστικού σε δυαδικό
- **Μετασχηματισμοί μεταβλητών (variable transformation):** Για παράδειγμα αλλαγή μεγέθους μέτρησης (πχ. από km σε m). Ένας μετασχηματισμός είναι και η **κανονικοποίηση**, χάρη στην οποία μπορούμε να κάνουμε συγκρίσιμα δύο ή

περισσότερα σύνολα τιμών. Για παράδειγμα μπορούμε να μετασχηματίσουμε ένα σύνολο τιμών με τον εξής τύπο:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

## 2.2.4 Μέτρα Ομοιότητας και Ανομοιότητας

Η ομοιότητα και η ανομοιότητα είναι αριθμητικά μέτρα για το πόσο όμοια ή ανόμοια είναι δύο αντικείμενα.

Παραδείγματα μέτρησης της ομοιότητας είναι:

- **Η Ευκλείδεια απόσταση:**  $\text{dist} = \sqrt{\sum_{k=1}^n (x(k) - y(k))^2}$
- **Η απόσταση Minkowski:**  $\text{dist} = (\sum_{k=1}^n (x(k) - y(k))^r)^{\frac{1}{r}}$
- **Η ομοιότητα συνημιτόνου:** Χρησιμοποιείται για σύγκριση κειμένων. Αν  $d_1, d_2$  δύο κείμενα, που παριστάνονται ως δύο διανύσματα, τότε  $\cos(d_1, d_2) = \frac{d_1 \cdot d_2}{\|d_1\| \|d_2\|}$  όπου ο αριθμητής παριστάνει το εσωτερικό τους γινόμενο και  $\|d\|$  είναι το μήκος του διανύσματος  $d$

Η **συσχέτιση (correlation)** μετράει τη γραμμική σχέση μεταξύ αντικειμένων.

Ο **συντελεστής συσχέτισης** του Pearson μεταξύ δύο αντικειμένων  $p$  και  $q$ , κυμαίνεται μεταξύ -1 και 1 και ορίζεται ως:

$$\text{corr}(p, q) = \frac{\text{cov}(p, q)}{\text{std}(p) \cdot \text{std}(q)} = \frac{s_{pq}}{s_p \cdot s_q}$$

Όπου:

- $\text{cov}(p, q)$  είναι η συνδιακύμανση (covariance) των  $p$  και  $q$   
 $\text{cov}(p, q) = s_{pq} = \frac{1}{n-1} \sum_{k=1}^n (p_k - \bar{p})(q_k - \bar{q})$
- $\text{std}(p), \text{std}(q)$  είναι οι τυπικές αποκλίσεις (standard deviations).
- $\text{std}(p) = s_p = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (p_k - \bar{p})^2}$
- $\text{std}(q) = s_q = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (q_k - \bar{q})^2}$

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

## 2.3 Κατηγοριοποίηση (classification)

### 2.3.1 Τι είναι και πώς λειτουργούν τα μοντέλα κατηγοριοποίησης

Ένα **μοντέλο κατηγοριοποίησης (classification model)**, είναι μία συνάρτηση η οποία παίρνει ως **δεδομένα εισόδου** μια συλλογή από εγγραφές, καθεμία από τις οποίες χαρακτηρίζεται από ένα σύνολο χαρακτηριστικών και ένα ειδικό χαρακτηριστικό που ορίζεται ως ετικέτα της κατηγορίας και είναι πάντα διακριτό. Η συνάρτηση αυτή απεικονίζει ένα σύνολο χαρακτηριστικών σε μια από τις προκαθορισμένες κατηγορίες (κλάσεις).

Τα μοντέλα κατηγοριοποίησης μπορούν να χρησιμοποιηθούν τόσο για **περιγραφή** (πχ για την επεξήγηση του πώς τα χαρακτηριστικά κατατάσσουν ένα αντικείμενο σε μια κατηγορία) όσο και για **πρόγνωση** (ανάθεση ετικέτας σε μη κατηγοριοποιημένες εγγραφές).

Κάθε κατηγοριοποιητής χρησιμοποιεί έναν **αλγόριθμο μάθησης (learning algorithm)**. Βασικός στόχος του αλγορίθμου μάθησης είναι να δημιουργήσει μοντέλα που να προβλέπουν με ακρίβεια τις ετικέτες των κατηγοριών για εγγραφές, οι οποίες προηγουμένως ήταν άγνωστες.

Τα δεδομένα χωρίζονται σε **σύνολο εκπαίδευσης (training set)** και **σύνολο ελέγχου (test set)**. Το πρώτο χρησιμοποιείται για την κατασκευή του μοντέλου κατηγοριοποίησης. Κατόπιν το μοντέλο εφαρμόζεται στο δεύτερο, αποδίδονται στις εγγραφές του ετικέτες κατηγοριών και συγκρίνονται με τις ήδη υπάρχουσες, για να εκτιμηθεί η απόδοση. Οι μετρήσεις των εγγραφών που έχουν προβλεφθεί σωστά και λανθασμένα, τοποθετούνται σε ένα πίνακα που ονομάζεται **μήτρα σύγχυσης (confusion matrix)**.

		Προβλεφθείσα κατηγορία	
		Κατηγορία = 1	Κατηγορία = 0
Πραγματική κατηγορία	Κατηγορία = 1	$f_{11}$	$f_{10}$
	Κατηγορία = 0	$f_{01}$	$f_{00}$

**Πίνακας 1:** Μήτρα σύγχυσης ενός δυαδικού προβλήματος. Τα  $f_{11}$ ,  $f_{10}$ ,  $f_{01}$ ,  $f_{00}$ , είναι πλήθη εγγραφών (Tan et al 2006).

Από τη μήτρα σύγχυσης μπορούν να εκτιμηθούν:

- Η **ακρίβεια (accuracy)** =  $\frac{\text{αριθμός σωστών προβλέψεων}}{\text{πλήθος εγγραφών}}$
- Ο **ρυθμός σφάλματος (error rate)** =  $\frac{\text{αριθμός λανθασμένων προβλέψεων}}{\text{πλήθος εγγραφών}}$

και η σχέση μεταξύ τους είναι  **$accuracy = 1 - error\ rate$**

Για παράδειγμα, στη μήτρα σύγχυσης της **Εικόνας 2**, είναι:

- **$accuracy = \frac{f_{11}+f_{00}}{f_{11}+f_{10}+f_{01}+f_{00}}$**
- **$error\ rate = \frac{f_{10}+f_{01}}{f_{11}+f_{10}+f_{01}+f_{00}}$**

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

### 2.3.2 Υπερπροσαρμογή μοντέλου

Τα σφάλματα που προέρχονται από ένα μοντέλο κατηγοριοποίησης χωρίζονται σε **σφάλματα εκπαίδευσης (training errors)** και **σφάλματα γενίκευσης (generalization errors)**. Σφάλμα εκπαίδευσης είναι το πλήθος εγγραφών που κατηγοριοποιούνται λανθασμένα στο σύνολο εκπαίδευσης (misclassification errors).



Σφάλμα γενίκευσης είναι το αναμενόμενο σφάλμα κατηγοριοποίησης σε εγγραφές που πρόκειται να κατηγοριοποιηθούν. Για να είναι ένα μοντέλο κατηγοριοποίησης καλό, θα πρέπει να είναι μικρά και τα δύο είδη σφαλμάτων. Αν ένα μοντέλο προσαρμόζεται πολύ καλά στα δεδομένα εκπαίδευσης και έχει δυσανάλογα μεγάλο σφάλμα γενίκευσης, μεγαλύτερο και από ένα μοντέλο με μεγαλύτερο σφάλμα εκπαίδευσης, τότε μιλάμε για **υπερπροσαρμογή μοντέλου**.

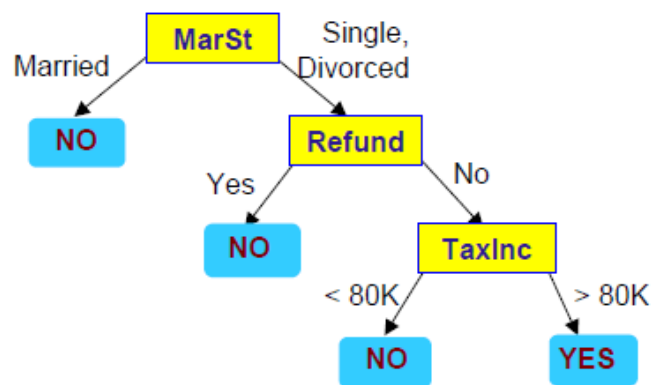
Η υπερπροσαρμογή μοντέλου μπορεί να οφείλεται σε παρουσία θορύβου, σε έλλειψη αντιπροσωπευτικών δειγμάτων ή σε άλλους παράγοντες. Οδηγεί σε δένδρα περισσότερο πολύπλοκα από όσο χρειάζεται.

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

### 2.3.3 Δένδρα απόφασης (decision trees)

Τα δένδρα απόφασης είναι μια απλή αλλά ευρέως χρησιμοποιούμενη τεχνική κατηγοριοποίησης. Ένα δείγμα δένδρου απόφασης φαίνεται στην παρακάτω εικόνα:

<i>Tid</i>	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



**Εικόνα1:** Δένδρο απόφασης με δεδομένα εισόδου τον πίνακα που παρουσιάζεται αριστερά (Tan et al, 2006)

Ο **αλγόριθμος του Hunt** είναι ένας αλγόριθμος στο οποίο βασίζονται πολλοί αλγόριθμοι δένδρων απόφασης (πχ ID3, C4.5, CART). Σε αυτόν, το δέντρο απόφασης μεγαλώνει αναδρομικά και οι εγγραφές εκπαίδευσης διαιρούνται σε όλο και πιο αμιγή σύνολα. Η γενική μορφή του αλγορίθμου είναι η εξής:

1. Έστω  $D_t$  το σύνολο των εγγραφών του συνόλου εκπαίδευσης που φτάνουν στον κόμβο  $t$
2.
  - a. Εάν το  $D_t$  περιέχει εγγραφές που ανήκουν στην ίδια κλάση  $y_t$ , τότε ο  $t$  γίνεται κόμβος φύλλο με ετικέτα  $y_t$
  - b. Εάν το  $D_t$  περιέχει εγγραφές που ανήκουν σε διαφορετικές κλάσεις, χρησιμοποιούμε ένα έλεγχο γνωρίσματος για να διασπάσουμε τα δεδομένα σε μικρότερα υποσύνολα
3. Το βήμα 2 ακολουθείται αναδρομικά για κάθε θυγατρικό κόμβο

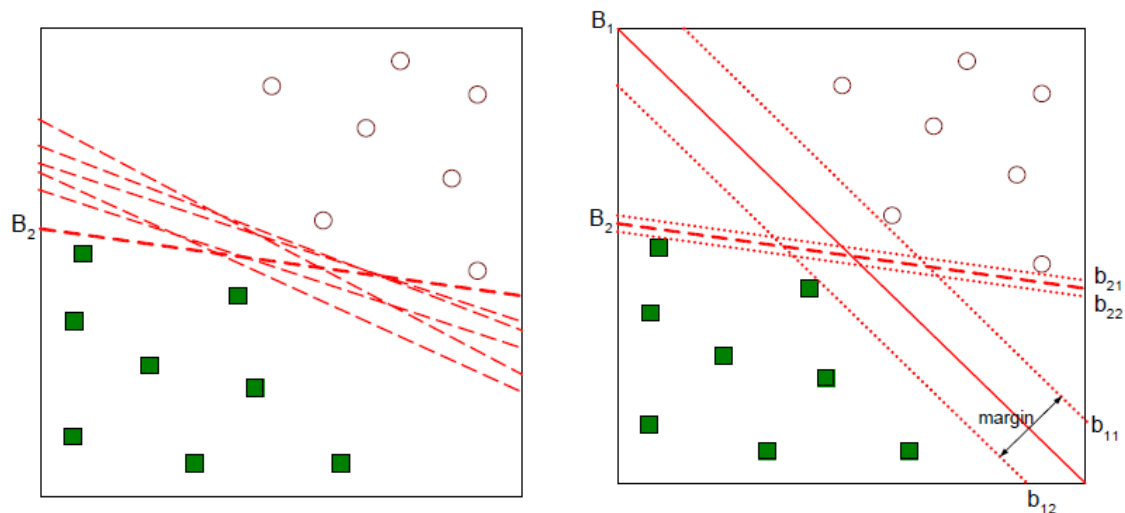
Τα θέματα που προκύπτουν κατά την εφαρμογή του αλγορίθμου είναι ο **καθορισμός του τρόπου διαχωρισμού** και ο καθορισμός του **πότε σταματά η ανάπτυξη** του δένδρου.

Θέματα που αφορούν τον τρόπο διαχωρισμού είναι τα σημεία διάσπασης των συνεχών χαρακτηριστικών και το αν η διάσπαση θα είναι δυαδική ή πολλαπλή. Επίσης, επειδή η διάσπαση γίνεται έτσι ώστε να προκύπτουν κόμβοι με τη μέγιστη δυνατή καθαρότητα, χρησιμοποιούνται διάφοροι τρόποι εκτίμησης της καθαρότητας, όπως το **GINI**, η **εντροπία** και το λάθος κατηγοριοποίησης (**classification error**). Τέλος, χρειάζεται να ξέρουμε πότε πρέπει να σταματήσει η ανάπτυξη του δένδρου ώστε να αντιμετωπίσουμε τα προβλήματα υπερπροσαρμογής.

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

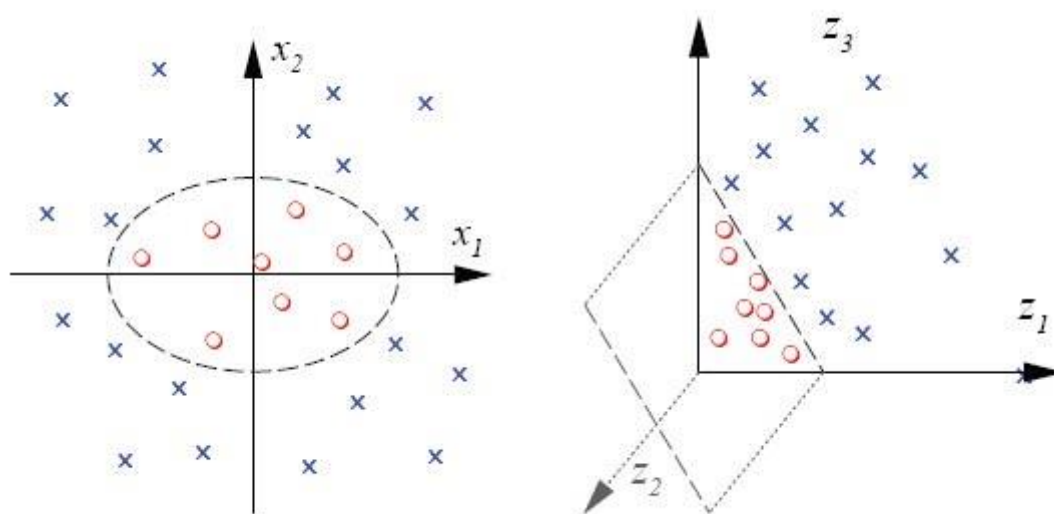
### 2.3.4 Μηχανές διανυσμάτων υποστήριξης (support vector machines, SVM)

Οι μηχανές διανυσμάτων υποστήριξης είναι ένα μοντέλο το οποίο χρησιμοποιήθηκε, μαζί με τα δένδρα απόφασης, στην παρούσα διατριβή. Σκοπός του μοντέλου είναι η κατασκευή ενός υπερεπιπέδου το οποίο διαχωρίζει πλήρως τα δεδομένα δύο κατηγοριών (όριο απόφασης).



**Εικόνα 2:** Παράδειγμα διαχωρισμού ενός συνόλου δεδομένων (Tan et al, 2006)

Όπως φαίνεται στα σχήματα της **Εικόνας 2**, υπάρχουν άπειρα υπερεπίπεδα που διαχωρίζουν πλήρως τις δύο κατηγορίες. Για την επιλογή του καλύτερου χρησιμοποιείται η έννοια του **υπερεπιπέδου μέγιστου περιθωρίου (maximal margin hyperplane)**. Κάθε υπερεπίπεδο  $B_i$  το οποίο διαχωρίζει πλήρως τις δύο κατηγορίες, σχετίζεται με ένα ζεύγος υπερεπιπέδων  $b_{i1}$  και  $b_{i2}$  τα οποία ακουμπούν στα μέλη των δύο συνόλων (εικόνα 4, δεξιά). Η απόσταση μεταξύ τους λέγεται **περιθώριο** και το βέλτιστο υπερεπίπεδο είναι αυτό που βρίσκεται στο μέσο του μέγιστου περιθωρίου. Στην περίπτωση που το όριο απόφασης δεν είναι γραμμικό, τα δεδομένα μετασχηματίζονται σε ένα νέο χώρο  $\Phi(x)$ , έτσι ώστε να μπορεί να χρησιμοποιηθεί ένα γραμμικό όριο απόφασης.



**Εικόνα 3:** Κατηγοριοποίηση δεδομένων με ένα μη γραμμικό όριο απόφασης. Αριστερά φαίνεται το όριο απόφασης στον αρχικό χώρο και δεξιά στο μετασχηματισμένο

Για την περίπτωση των μη γραμμικών συναρτήσεων οι SVMs χρησιμοποιούν συναρτήσεις γνωστές ως **συναρτήσεις πυρήνα (kernel functions)**.

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

### 2.3.5 Άλλοι κατηγοριοποιητές

Άλλα μοντέλα κατηγοριοποίησης που χρησιμοποιούνται είναι:

- ο **κατηγοριοποιητής κανόνων (rule-based classifier)**: Χρησιμοποιεί ένα σύνολο κανόνων (rule set) της μορφής “if...then...else”
- ο **κατηγοριοποιητής πλησιέστερου γείτονα**: Επιλέγει τα δείγματα του συνόλου εκπαίδευσης που έχουν μεγαλύτερη ομοιότητα με το δείγμα ελέγχου και αποδίδει σε αυτό την ετικέτα που έχει η πλειοψηφία τους
- ο κατηγοριοποιητής **Bayes**: Χρησιμοποιεί το θεώρημα του Bayes, το οποίο έχει να κάνει με τις δεσμευμένες πιθανότητες
- τα **τεχνητά νευρωνικά δίκτυα (artificial neural networks)**: Αποτελούνται από ένα δίκτυο κόμβων, που προσομοιάζουν στους νευρώνες του ανθρώπινου εγκεφάλου, οι οποίοι συνδέονται μεταξύ τους με ένα σύνολο προσανατολισμένων βελών.

Επίσης χρησιμοποιούνται, **μέθοδοι συνδυασμού κατηγοριοποιητών** (classifier combination) όπως η **εμφωλίαση (bagging)**, η **ενίσχυση (boosting)** και τα **τυχαία δάση (random forests)**. Ένας αλγόριθμος ενίσχυσης που χρησιμοποιείται είναι ο **αλγόριθμος AdaBoost**.

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

## 1.4 Άλλες τεχνικές Εξόρυξης Δεδομένων

Οι κυριότερες τεχνικές εξόρυξης δεδομένων εκτός από την κατηγοριοποίηση είναι:

- η **Ανάλυση Συστάδων (Cluster Analysis)**: Χωρίζει τα δεδομένα σε κατηγορίες (συστάδες), με βάση τη μεταξύ τους ομοιότητα ή ανομοιότητα
- η **Ανάλυση Συσχέτισης (Association Analysis)**: Αναλύει ένα είδος δεδομένων γνωστό ως **δεδομένα καλαθιού αγοράς (market basket transactions)** και βρίσκει **κανόνες συσχέτισης (association rules)** μεταξύ τους.

(Μέρος του κειμένου αυτής της ενότητας προέρχεται από Tan et al, 2006)

# Κεφάλαιο 3

## Η Ανάλυση Δεδομένων στη Φυτοπροστασία

### 3.1 Εφαρμογή τεχνικών Εξόρυξης Δεδομένων στη φυτοπροστασία

Οι αλγόριθμοι μηχανικής μάθησης στη γεωργία χρησιμοποιούνται τα τελευταία χρόνια ευρέως στα GIS, στην εδαφολογία, στην υδρολογία, στη γεωργία ακριβείας, στην πρόγνωση της απόδοσης της καλλιέργειας και στην διασφάλιση ποιότητας της παραγωγής (Hill et al, 2014). Η εφαρμογή μεθόδων εξόρυξης δεδομένων στη φυτοπροστασία φαίνεται να είναι ένα ανερχόμενο πεδίο έρευνας.

Την τελευταία εικοσαετία έχουν αρκετές φορές χρησιμοποιηθεί στη φυτοπροστασία μοντέλα πρόγνωσης χρησιμοποιώντας μεθόδους όπως την παλινδρόμηση, κατηγοριοποιητές Bayes, μηχανές διανυσμάτων υποστήριξης και νευρωνικά δίκτυα. Τα μοντέλα αυτά χρησιμοποιώντας ως δεδομένα εισόδου χαρακτηριστικά όπως η θερμοκρασία, η ατμοσφαιρική υγρασία, η εδαφική υγρασία, η ηλιακή ακτινοβολία κ.α., έχουν χρησιμοποιηθεί για την πρόγνωση της υγρασίας στα φύλλα, του μεγέθους της προσβολής εχθρών και ασθενειών διαφόρων καλλιεργειών, κυρίως του σιταριού και της παρουσίας δεσοξινιβαλενόλης, μίας τοξίνης που παράγεται από μύκητες που προσβάλλουν τα σιτηρά και είναι επιβλαβής για την ανθρώπινη υγεία (Yun Hwan Kim et al, 2014).

Επίσης την τελευταία εικοσιπενταετία η έρευνα έχει στραφεί στις μεθόδους μηχανικής μάθησης προκειμένου να ανιχνευτεί σε πολύ αρχικά στάδια η ανάπτυξη ασθενειών, εχθρών και ζιζανίων με την ανάλυση δεδομένων που προέρχονται από οπτικούς αισθητήρες υψηλής ανάλυσης. Οι μέθοδοι που χρησιμοποιήθηκαν είναι η παλινδρόμηση, οι μηχανές διανυσμάτων υποστήριξης, τα νευρωνικά δίκτυα αλλά και η ανάλυση συστάδων.

Τα αποτελέσματα δείχνουν ότι η τεχνολογία αυτή βρίσκεται ακόμα σε αρχικά στάδια αλλά είναι πολλά υποσχόμενη (Behmann et al, 2015).

Στη Νέα Ζηλανδία διερευνήθηκε η αποτελεσματικότητα πέντε κατηγοριοποιητών (Δέντρο Απόφασης, Bayes, Τυχαία Δάση (Random Forest), AdaBoost και Μηχανές Υποστήριξης Διανυσμάτων), για την πρόγνωση του αν ο πληθυσμός του φυλλοδέτη, ενός εντόμου που προσβάλλει το ακτινίδιο, βρίσκεται πάνω ή κάτω από το όριο της απόφασης για ψεκασμό. Η απόφαση για ψεκασμό δεν έγινε δυνατόν να προβλεφθεί σε ικανοποιητικό βαθμό ενώ, αντίθετα, η απόφαση για μη ψεκασμό προβλέφθηκε επιτυχώς, με μεγαλύτερη ακρίβεια να δίνουν ο AdaBoost και ο Bayes (Hill et al, 2014).

Στο Πακιστάν, μια από τις μεγαλύτερες βαμβακοπαραγωγούς χώρες στον κόσμο, ψηφιοποιήθηκαν δεδομένα που συλλέγονταν συστηματικά από τις κρατικές υπηρεσίες τα οποία κρατούνταν σε φυσικό αρχείο έντυπων καρτελών. Τα δεδομένα αφορούσαν, μεταξύ άλλων παρακολούθηση της ανάπτυξης εχθρών και ασθενειών, ημερολόγιο ψεκασμών και μετεωρολογικές μετρήσεις. Με μία πρωτότυπη μέθοδο ανάλυσης συστάδων διερευνήθηκε το πού οφείλεται η αρνητική συσχέτιση, η οποία έχει παρατηρηθεί, μεταξύ της χρήσης φυτοφαρμάκων και της απόδοσης της καλλιέργειας. Η ανάλυση αποκάλυψε συγκεκριμένα σφάλματα στη χρήση των φυτοφαρμάκων. (Abdullah et al, 2003) (Abdullah et al, 2006).

### **3.2 Η Ανάλυση δεδομένων στην αντιμετώπιση του Δάκου της ελιάς**

Στο Ισραήλ, κατασκευάστηκε μοντέλο το οποίο κατάφερε να προβλέψει με μεγάλη ακρίβεια τις διακυμάνσεις του πληθυσμού των δάκων για διάρκεια 11 χρόνων, γνωρίζοντας τον αρχικό αριθμό αυγών, προνυμφών, νυμφών, ακμαίων και γόνιμων ακμαίων και την επίδραση της θερμοκρασίας στη διάρκεια καθενός από τα στάδια αυτά. Η θερμοκρασία της κόμης του κάθε ελαιοδέντρου υπολογίστηκε από δεδομένα δορυφόρου που μετρά με οπτικούς αισθητήρες την αντανάκλαση του φωτός (Blum et al, 2013)

Στη Χαλκιδική, μελετήθηκε η επίδραση του υψομέτρου σε χωροχρονικά πρότυπα του Δάκου της ελιάς. Σε μια περιοχή με ποικίλο υψόμετρο τοποθετήθηκαν παγίδες McPhail,

καταγράφηκαν οι γεωγραφικές συντεταγμένες τους και μετρήθηκε ο πληθυσμός του Δάκου στη διάρκεια μίας περιόδου δακοκτονίας (βλ. παρ. 1.1.2). Τα δεδομένα αναλύθηκαν με τη βοήθεια εφαρμογής στατιστικής ανάλυσης γεωγραφικών δεδομένων και εντοπίστηκαν συγκεκριμένα hotspots σε περιοχές διαφορετικού υψομέτρου ανάλογα με την εποχή, που δείχνουν μετακινήσεις πληθυσμών σε διαφορετικό υψόμετρο ανάλογα με την εποχή. (Kounatidis et al, 2008).

Τέλος, βρίσκεται σε εξέλιξη η κατασκευή ενός συστήματος λήψης αποφάσεων το οποίο, με τη βοήθεια παγίδων με οπτικούς αισθητήρες στα ελαιόδεντρα, θα καταγράφει τον πληθυσμό του δάκου και θα αποφασίζει αν πρέπει να γίνει ψεκασμός – σε συγκεκριμένο δέντρο κάθε φορά – και αν ναι, θα ειδοποιεί σχετικά το χρήστη μέσω γραφικής διεπαφής. (Pontikakos et al, 2009 (Pontikakos et al, 2012) (Ποντικάκος κ.α. 2015). Οι ψεκασμοί έτσι είναι εντοπισμένοι και όχι καθολικοί. Άλλη έρευνα δείχνει ότι αντίστοιχοι αισθητήρες θα μπορούσαν να είναι και ακουστικοί (Potamitis et al 2015)



# Κεφάλαιο 4

## Περιγραφή και Συλλογή Δεδομένων

### 4.1 Πίνακες από τη Βάση Δεδομένων DACUS του

#### ΥΠΑΑΤ

##### 4.1.1 Χορήγηση πινάκων από το Υπουργείο

Αρχικός σκοπός της παρούσας Διατριβής, ήταν η επεξεργασία αποκλειστικά των στοιχείων της Βάσης Δεδομένων DACUS του Υπουργείου Αγροτικής Ανάπτυξης και Τροφίμων της Ελλάδας (ΥΠΑΑΤ), όπως αυτά έχουν καταχωρηθεί από τις Διευθύνσεις Αγροτικής Οικονομίας και Κτηνιατρικής (ΔΑΟΚ) των κατά τόπους Περιφερειακών Ενοτήτων (ΠΕ) Σάμου και Χαλκιδικής, για τα τελευταία 5 – 10 χρόνια. Επιλέχθηκαν οι νομοί Σάμου και Χαλκιδικής και ζητήθηκαν, με γραπτή αίτηση στο ΥΠΑΑΤ, στοιχεία που αφορούν στο Πρόγραμμα Δακοκτονίας, στην αναλυτικότερη δυνατή μορφή και για όσα έτη έχουν καταγραφεί. Όπως διαπιστώθηκε με την παραλαβή των δεδομένων, η Βάση άρχισε να ενημερώνεται το 2011 με στοιχεία από ελάχιστες περιοχές. Το 2012 και το 2013 δεν ενημερώθηκε καθόλου, ενώ μόλις την περίοδο 2014 προστέθηκαν σε αυτή στοιχεία από μεγάλο αριθμό περιοχών. Όσο αφορά τους νομούς ενδιαφέροντός μας, λάβαμε δεδομένα του 2011 για το νομό Σάμου (Σάμος & Ικαρία) ενώ για το νομό Χαλκιδικής μόνο όσο αφορά στις δειγματοληψίες για τη δακοπροσβολή, του 2014 για τους νομούς Σάμου και Χαλκιδικής και μερικές δοκιμαστικές εγγραφές του 2008.

Τα δεδομένα δόθηκαν σε ένα αρχείο .xls το οποίο περιείχε 9 πίνακες σε ισάριθμα φύλλα εργασίας. Οι πίνακες είναι οι εξής:

##### 4.1.1.1 Πίνακας ΔΗΜΟΙ

Στον πίνακα καταγράφονται οι Δήμοι με τη μορφή του νόμου «Καποδίστρια», τα Δημοτικά Διαμερίσματα και οι οικισμοί. Τα πεδία του πίνακα είναι τα εξής:

*DIR\_ID*: είναι το ID του Νομού. Προφανώς είναι εξωτερικό κλειδί που δείχνει πίνακα όπου καταγράφονται οι Νομοί, ο οποίος δεν έχει δοθεί από το Υπουργείο.

*DHMOS\_CODE*: ένας 8-ψήφιος κωδικός όπου τα δύο πρώτα ψηφία αντιστοιχούν στο Νομό, τα επόμενα δύο στο Δήμο (Καποδιστριακό), τα επόμενα δύο στο Δημοτικό Διαμέρισμα και τα τελευταία δύο στον οικισμό, με μοναδική τιμή για κάθε εγγραφή,

*DHMOS\_DESCR*: είναι το όνομα του Δήμου ή του Δημοτικού Διαμερίσματος ή του οικισμού,

*DHMOS\_ID*: είναι πενταψήφιο ID με αύξουσα σειρά καταγραφής της εγγραφής, με μοναδική τιμή για κάθε εγγραφή.

#### **4.1.1.2 Πίνακας ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ**

Καταγράφονται μετρήσεις του δακοπληθυσμού. Όπως φαίνεται από τα πεδία του πίνακα, οι προδιαγραφές της Βάσης Δεδομένων DACUS είναι να καταγράφονται οι μετρήσεις ανά παγίδα, ενώ προβλέπεται ακόμα και καταγραφή θερμοκρασίας και υγρασίας. Στην πράξη καταγραφές της θερμοκρασίας και της υγρασίας υπάρχουν μόνο για το νομό Σάμου το 2011. Επίσης οι εγγραφές για τη Σάμο το 2011 και για τη Χαλκιδική το 2014 είναι συνολικά **ανά Δημοτικό Διαμέρισμα**. Εγγραφές ανά παγίδα έχουμε μόνο για τη Σάμο το 2014, χωρίς να καταγράφονται οι γεωγραφικές συντεταγμένες των παγίδων. Το σύνολο των εγγραφών του πίνακα είναι 4399. Τα κυριότερα πεδία του πίνακα είναι τα εξής:

*CHANGE\_DATE*: είναι η ημερομηνία αλλαγής (δηλ. μέτρησης και αλλαγής του υγρού της παγίδας),

*MALE, FEMALE*: Αριθμός αρσενικών και θηλυκών ατόμων αντίστοιχα

*MO\_DACUS*: Μέσος όρος δάκων ανά παγίδα

*FEMALE\_POSOSTO*: Η αναλογία θηλυκά/αρσενικά

*GONIMOTHTA*: Το % ποσοστό γόνιμων θηλυκών. **Δεν υπάρχουν καταγραφές εκτός από τις δοκιμαστικές εγγραφές του 2008.**

*DHMOS\_ID*: Το ID του Δήμου, που δείχνει στον αντίστοιχο πίνακα.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O		
1	CHANGE DATE	TEMPERATURE	HUMIDITY	MALE	FEMALE	MO DACUS	FEMALE POSOSTO	GONIMOTHTA	DHMOS_AA	PAGIDA_AA	DAC_METRHSH	DACUS_ID	TOMEARX_ID	PK_ID	PK_POSOSTO	DHS
2	07/03/2008	30	75	20	20	2		25	2	1			101	46	49	45
3	12/03/2008	30	75	30	25	3		30	3	1			102	46	49	45
4	17/03/2008	30	75	20	20	2		15	4	1			103	46	49	45
5	22/03/2008	30	75	5	5	1		5	5	1			104	46	49	45
6	20/08/2008	1	1	7	5	1	0.42	1	2	1			114	45	49	1
7	20/08/2008	1	1	7	6	1	0.46	1	2	2			115	45	49	1
8	27/03/2008	30	75	15	10	2		7	6	1			105	46	49	45
9	02/04/2008	30	75	10	5	1		10	7	1			106	46	49	45
10	15/08/2008	1	1	5	5	1	0.5	1	1	2			108	45	49	1
11	15/08/2008	1	1	1	3	1	0.75	1	1	3			109	45	49	1
12	15/08/2008	1	1	1	1	1	1	1	1	1			53	45	49	1
13	20/08/2008	1	1	2	3	1	0.6	1	2	3			126	45	49	1
14	02/03/2008	30	75	10	10	1	0.5	10	1	2			128	46	49	45
15	07/03/2008	30	75	20	20	2	0.5	25	2	2			129	46	49	45
16	12/03/2008	30	75	30	25	3	0.45	30	3	2			130	46	49	45
17	17/03/2008	30	75	20	20	2	0.5	15	4	2			131	46	49	45
18	15/03/2009	25	65	12	8	0.5		45	1	1			97	86	81	52
19	02/03/2008	30	75	10	10	1	0.5	10	1	1			62	46	49	45
20	10/08/2011	30	56	38	21	4.21	0.36		1	14			26476	20820	20904	10
21	20/07/2011	27	30.5	8	9	3.4	0.53		83	5			26970	20820	20904	
22	04/10/2011	27	30.5	10	12	4.4	0.55		84	5			26971	20820	20904	
23	12/10/2011	27	30.5	9	15	4.8	0.63		85	5			26972	20820	20930	
24	17/10/2011	26	30.9	2	2	0.8	0.5		86	5			26973	20820	20904	
25	26/10/2011	26	30.9	7	13	4	0.65		87	5			26974	20820	20904	
26	21/07/2011	30	47.5	12	23	6.9	0.66		88	5			26975	20820	20930	
27	28/07/2011	29	47.8	19	21	8	0.53		89	5			27036	20820	20904	
28	03/08/2011	30	50.9	8	5	2.5	0.38		90	5			27037	20820	20930	
29	10/08/2011	30	51	6	7	2.5	0.54		91	5			27038	20820	20904	
30	17/08/2011	30	51	9	11	4	0.61		92	5			27039	20820	20930	

**Εικόνα 4:** Το αρχείο .xls που παραχωρήθηκε από το ΥΠΑΑΤ. Φαίνεται ο Πίνακας ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ, ενώ κάτω φαίνονται τα φύλλα εργασίας με τους τίτλους των Πινάκων.

#### 4.1.1.3 Πίνακας ΔΕΙΓΜΑΤΟΛΗΨΙΑ

Στον πίνακα καταγράφονται τα αποτελέσματα από τις δειγματοληψίες για την εκτίμηση της δακοπροσβολής. Οι εγγραφές είναι ανά Δημοτικό Διαμέρισμα. Από τα πεδία του πίνακα φαίνεται ότι οι προδιαγραφές της Βάσης Δεδομένων DACUS είναι για να καταγράφονται όλοι οι υπάρχοντες τύποι των προσβολών. Σε αρκετές εγγραφές υπάρχουν κενά πεδία και δεν είναι σαφές αν αυτό υπονοεί τον αριθμό 0 ή αν απλά δεν έχει γίνει καταγραφή. Γενικά, στις περιπτώσεις που σε ένα συγκεκριμένο Δημοτικό Διαμέρισμα για ένα συγκεκριμένο έτος οι τιμές ενός πεδίου είναι κενές, μπορούμε να υποθέσουμε ότι δεν έχει γίνει καταγραφή. Στην αντίθετη περίπτωση μπορούμε να υποθέσουμε ότι υπονοείται η τιμή 0.

Το σύνολο των εγγραφών του πίνακα είναι 174. Τα κυριότερα πεδία του πίνακα είναι τα εξής:

*DEIGMA\_DATE:* Είναι η ημερομηνία που έγινε η δειγματοληψία

*PR\_CARP\_AGNOS, PR\_CARP\_GONIMOS, PRONYFH\_A, PRONYFH\_B, PRONYFH\_G, NYFH, EXITS, PARASITISMOS, AYGA:* Ο αριθμός των καρπών οι οποίοι βρέθηκαν να έχουν άγωνα νύγματα, γόνιμες προσβολές, προνύμφες α, β και γ σταδίου, νύμφες, σπές εξόδου, δευτερογενείς προσβολές και αυγά αντίστοιχα (βλ. παράγραφο 1.1.1.)

*DHMOS\_ID:* Το ID του Δήμου, που δείχνει στον αντίστοιχο πίνακα.

#### 4.1.1.4 Πίνακας ΨΕΚΑΣΜΟΙ

Καταγράφονται στοιχεία για τους ψεκασμούς που διενεργήθηκαν. Οι εγγραφές είναι ανά Δημοτικό Διαμέρισμα. Ο πίνακας περιέχει 224 εγγραφές. Τα κυριότερα πεδία του πίνακα είναι τα εξής:

*SPRAY\_STDATE, SPRAY\_ENDATE*: Οι ημερομηνίες έναρξης και λήξης του ψεκασμού αντίστοιχα.

*FAR\_ID*: Το ID φαρμάκου που χρησιμοποιήθηκε. Προφανώς είναι εξωτερικό κλειδί που δείχνει πίνακα όπου καταγράφονται τα εντομοκτόνα, ο οποίος δεν έχει δοθεί από το Υπουργείο.

*SPRAYED\_TREES, LIQUID\_USED, FARMAK\_USED, PULL\_USED, LAND\_PROTECT*: Ο αριθμός δέντρων που ψεκάστηκαν, η ποσότητα ψεκαστικού υγρού που χρησιμοποιήθηκε (ψεκαστικό υγρό = νερό + εντομοκτόνο + ελκυστική ουσία), η ποσότητα εντομοκτόνου που χρησιμοποιήθηκε, η ποσότητα ελκυστικής ουσίας που χρησιμοποιήθηκε και η έκταση της περιοχής που ψεκάστηκε αντίστοιχα.

*DHMOS\_ID*: Το ID του Δήμου, που δείχνει στον αντίστοιχο πίνακα.

#### 4.1.1.5 Άλλοι Βοηθητικοί Πίνακες

Πρόκειται για τους πίνακες ΕΠΟΠΤΕΙΕΣ, ΤΟΜΕΙΣ, ΠΕΡΙΟΔΟΙ, ΤΟΜΕΑΡΧΕΣ. Στο πρόγραμμα δακοκτονίας κάθε νομός χωρίζεται σε Εποπτείες και κάθε Εποπτεία σε Τομείς. Οι πίνακες φαίνεται να συνδέονται μεταξύ τους ως εξής:

- Ο πίνακας ΕΠΟΠΤΕΙΕΣ διαθέτει, μεταξύ άλλων, το πεδίο *DIR\_ID*, το οποίο δείχνει το Νομό και το οποίο αναφέρεται παραπάνω στην παράγραφο 4.1.1.1 για τον πίνακα ΔΗΜΟΙ
- Ο πίνακας ΤΟΜΕΙΣ διαθέτει, μεταξύ άλλων, το πεδίο *EP\_ID* οποίο δείχνει στο ομώνυμο πρωτεύον κλειδί του πίνακα ΕΠΟΠΤΕΙΕΣ.
- Ο πίνακας ΠΕΡΙΟΔΟΙ καταγράφει πληροφορίες για τις Περιόδους του Προγράμματος Δακοκτονίας (πχ. 1-7-2014 έως 30-11-2014), ανά Τομέα. Διαθέτει, μεταξύ άλλων, το πεδίο *TOM\_ID* οποίο δείχνει στο ομώνυμο πρωτεύον κλειδί του πίνακα ΤΟΜΕΙΣ.
- Ο πίνακας ΤΟΜΕΑΡΧΕΣ καταγράφει πληροφορίες για τους Τομεάρχες που απασχολήθηκαν μία ορισμένη Περίοδο (κάθε Τομεάρχης είναι υπεύθυνος για τον Τομέα του). Διαθέτει, μεταξύ άλλων, το πεδίο *PERIOD\_ID* οποίο δείχνει στο ομώνυμο πρωτεύον κλειδί του πίνακα ΠΕΡΙΟΔΟΙ.

- Το πεδίο *TOMEARX\_ID* είναι το πρωτεύον κλειδί του πίνακα *TOMEAPXES*, το οποίο είναι και εξωτερικό κλειδί στους πίνακες *ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ*, *ΔΕΙΓΜΑΤΟΛΗΨΙΑ* και *ΨΕΚΑΣΜΟΙ*. Μέσω αυτού του κλειδιού, κάθε εγγραφή αυτών των πινάκων σχετίζεται «αλυσιδωτά» με τους βοηθητικούς πίνακες που αναφέρονται σε αυτή την παράγραφο.

#### **4.1.1.6 Πίνακας ΚΑΡΠΟΦΟΡΙΑ**

Στον πίνακα καταγράφονται, ανά Δημοτικό Διαμέρισμα, οι εκτιμήσεις του ποσοστού καρποφορίας που γίνονται κάθε χρόνο την Άνοιξη, οι οποίες παίζουν καθοριστικό ρόλο στο αν η περιοχή θα ενταχθεί στο ετήσιο πρόγραμμα δακοκτονίας (βλ. παρ. 1.1.2.). Διαθέτει και αυτός τα πεδία *DHMOS\_ID* και *TOMEARX\_ID*.

#### **4.1.2 Δημιουργία Βάσης Δεδομένων**

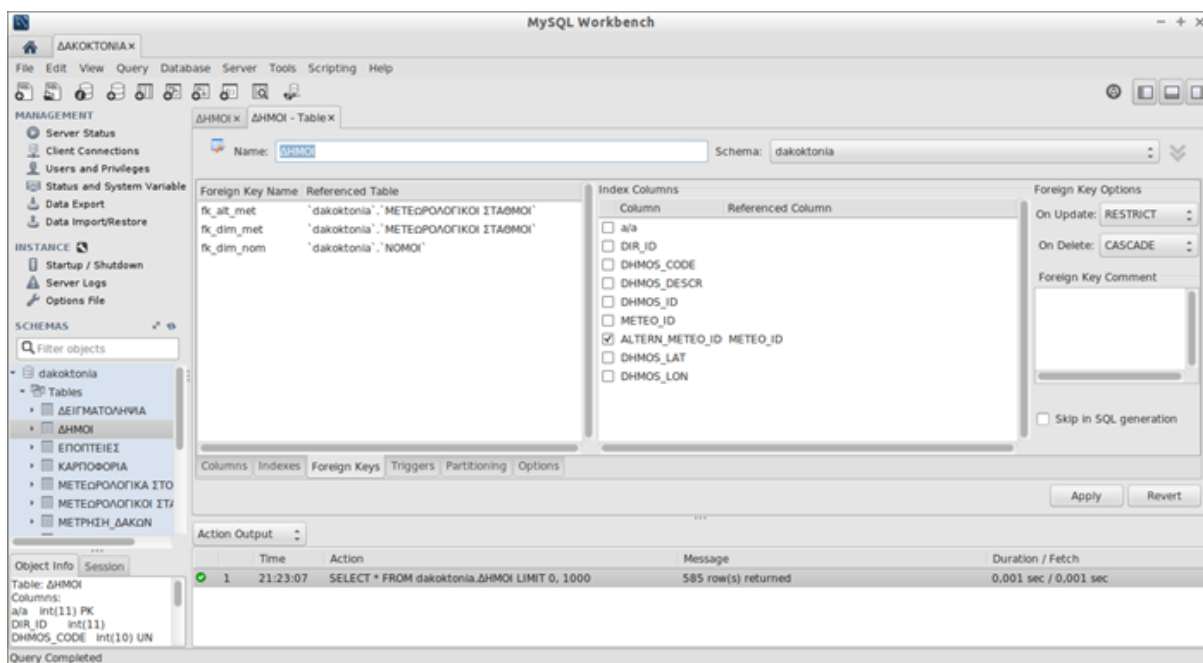
Για τη διευκόλυνση της επεξεργασίας των στοιχείων που χορηγήθηκαν από το ΥΠΑΑΤ δημιουργήθηκε Βάση Δεδομένων σε MySQL, με τη βοήθεια του εργαλείου **MySQL Workbench**.

##### **4.1.2.1 Πίνακες**

Για λόγους συμβατότητας με τους πίνακες που χορηγήθηκαν από το Υπουργείο και ευκολίας στην εισαγωγή των δεδομένων, δεν έγινε κάποια κανονικοποίηση της Βάσης Δεδομένων. Δηλαδή τα πεδία των Πινάκων της Βάσης είναι αυτούσια τα πεδία των πινάκων που χορηγήθηκαν από το Υπουργείο. Δημιουργήθηκε επιπλέον ο Πίνακας «*NOMOI*» με τα εξής τρία πεδία: *a/a*, *NOMOI\_ID* και *NOMOI\_DESCR*, με το πρώτο να είναι το πρωτεύον κλειδί και στο τρίτο να καταγράφεται το όνομα του νομού. Το δεύτερο ορίστηκε να έχει μοναδική τιμή για κάθε εγγραφή, ενώ σχετίζεται (με σχέση ένα προς πολλά) με το πεδίο *DIR\_ID* του Πίνακα *ΔΗΜΟΙ*.

### 4.1.2.2 Σχέσεις μεταξύ Πινάκων

Στους πίνακες ορίστηκαν εξωτερικά κλειδιά με τρόπο ώστε να δημιουργηθούν σχέσεις όπως αυτές περιγράφονται αναλυτικά στις περιγραφές των πινάκων στην παράγραφο 4.1.1.



Εικόνα 5: Η πλατφόρμα **MySQL Workbench**. Αριστερά φαίνονται τα ονόματα μερικών πινάκων, ενώ στην κύρια οθόνη τα εξωτερικά κλειδιά του πίνακα ΔΗΜΟΙ, όπως διαμορφώθηκαν μετά την προσθήκη και των μετεωρολογικών δεδομένων (βλ. παράγραφο 4.3)

### 4.1.2.3 Καταχώρηση Δεδομένων

Για την καταχώρηση των δεδομένων χρησιμοποιήθηκε η εφαρμογή **Base** της σουίτας **LibreOffice**. Δημιουργήθηκε ένα αρχείο .odb το οποίο αποτελεί μία σύνδεση με τη MySQL Database που έχει δημιουργηθεί. Η καταχώρηση των δεδομένων έγινε με επικύρωση, δεδομένου ότι, όπως αναφέρθηκε, οι αντίστοιχοι πίνακες που δόθηκαν από το Υπουργείο έχουν ακριβώς τα ίδια πεδία.



**Εικόνα 6:** Η εφαρμογή **LibreOffice Base**. Διακρίνονται τα ονόματα των Πινάκων όπως διαμορφώθηκαν μετά την προσθήκη και των μετεωρολογικών δεδομένων (βλ. παράγραφο 4.3)

#### 4.1.2.4 Δημιουργία Ερωτημάτων

Τα ερωτήματα δημιουργήθηκαν ως **Views** στο **MySQL Workbench** και είναι τα εξής:

ΕΡΩΤ\_ΜΕΤΡ\_ΔΑΚΩΝ, ΕΡΩΤ\_ΔΕΙΓΜΑΤΟΛΗΨΙΑ και ΕΡΩΤ\_ΨΕΚΑΣΜΟΙ.

Βασίζονται στους αντίστοιχους πίνακες και περιέχουν όλα τα σημαντικά πεδία τους, μαζί με τα πεδία των πινάκων ΔΗΜΟΙ και ΝΟΜΟΙ (συμπεριλαμβανομένου και του ονόματος του Δήμου (DHMOS\_DESCR) και του Νομού (NOMOS\_DESCR). Επιπλέον:

Στο ερώτημα ΕΡΩΤ\_ΜΕΤΡ\_ΔΑΚΩΝ:

- το πεδίο FEMALE\_POSOSTO, δεν είναι το αντίστοιχο πεδίο του πίνακα ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ αλλά υπολογίζεται εκ νέου, αυτή τη φορά ως  $FEMALE / (MALE + FEMALE)$ , για καλύτερο υπολογισμό στην περίπτωση που το MALE είναι 0.
- προστέθηκε το πεδίο MIN, στο οποίο υπολογίζεται η τιμή  $MIN(MALE, FEMALE)$ , η οποία απεικονίζει τον αριθμό των ζευγαριών (1:1) αρσενικών-θηλυκών που σχηματίζονται
- προστέθηκε το πεδίο SYNOLO\_DAKON όπου υπολογίζεται το άθροισμα  $MALE + FEMALE$

Στο ερώτημα ΕΡΩΤ\_ΔΕΙΓΜΑΤΟΛΗΨΙΑ:

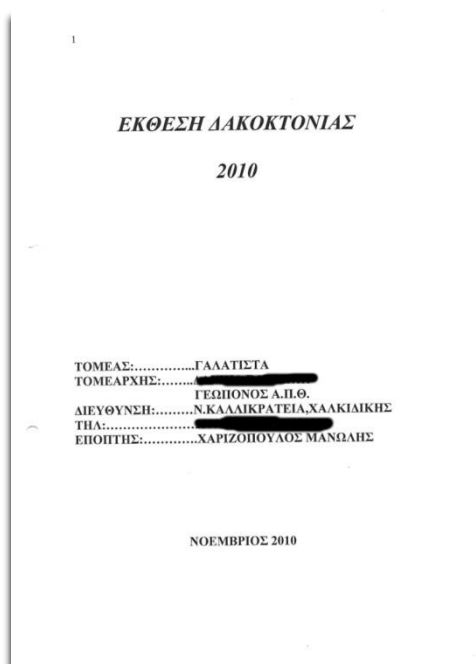
- το πεδίο PR\_CARP\_GONIMOS δεν είναι το αντίστοιχο πεδίο του πίνακα ΔΕΙΓΜΑΤΟΛΗΨΙΑ αλλά υπολογίζεται εκ νέου ως: PRONYFH\_A + PRONYFH\_B + PRONYFH\_G + NYFH + AYGA + EXITS (βλ. παράγραφο 1.1.4.)
- προστέθηκε το πεδίο PROSVOLLES όπου υπολογίζεται το άθροισμα όλων των τύπων προσβολών.

## 4.2 Αρχείο της Διεύθυνσης Αγροτικής Οικονομίας και Κτηνιατρικής Π.Ε. Χαλκιδικής

### 4.2.1 Πρόσβαση στο αρχείο

Μετά από τηλεφωνική επικοινωνία με τους γεωπόνους της Διεύθυνσης Αγροτικής Οικονομίας και Κτηνιατρικής της Περιφερειακής Ενότητας Χαλκιδικής, μας παραχωρήθηκε ευγενικά το αρχείο με τις εκθέσεις των Τομεαρχών για τα έτη 2009 έως 2013. Πρόκειται για φυλλάδια μεγέθους Α4 των 30 – 50 σελίδων στα οποία οι Τομεάρχες κατέγραφαν αναλυτικά την πορεία του προγράμματος στον Τομέα τους για τη συγκεκριμένη περίοδο.

Οι εκθέσεις εστάλησαν (μετά από γραπτή αίτηση) με courier σε δύο δόσεις. Με τη βοήθεια ενός σκάνερ, μετατρέπονταν σε αρχεία pdf και επιστρέφονταν πάλι με courier.



**Εικόνα 7:** Το εξώφυλλο της Έκθεσης Δακοκτονίας για τον Τομέα Γαλάτιστας το 2010



#### 4.2.2 Περιεχόμενο Εκθέσεων Δακοκτονίας

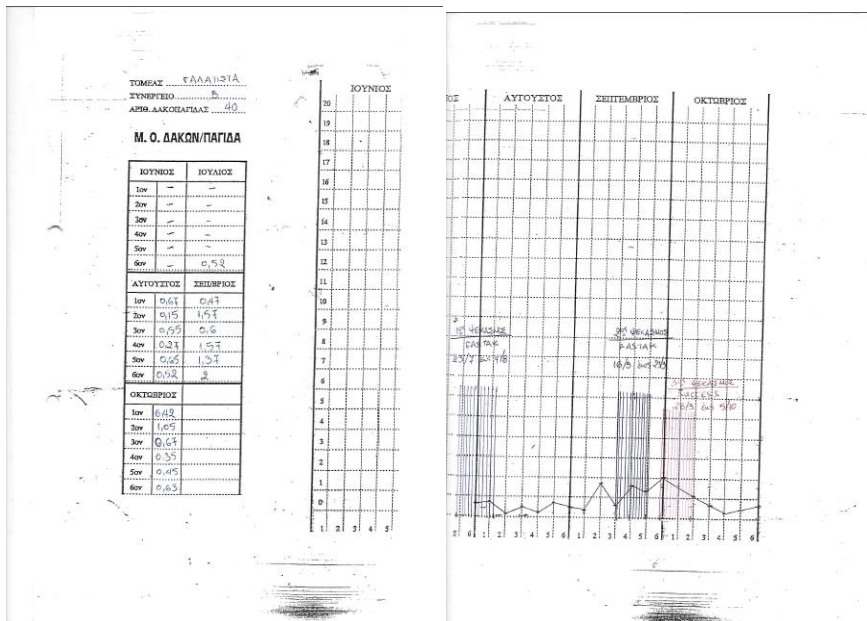
Στις Εκθέσεις περιλαμβάνονται κυρίως τα αποτελέσματα των μετρήσεων δακοπληθυσμού και των δειγματοληψιών για τη μέτρηση της δακοπροσβολής στον ελαιόκαρπο, καθώς και πληροφορίες για τους ψεκασμούς που διενεργήθηκαν. Σε κάθε Τομέα δρουν δύο ή περισσότερα Συνεργεία και τα στοιχεία παρατίθενται ανά Συνεργείο. Οι εκθέσεις περιλαμβάνουν κι άλλα στοιχεία όπως γενικές πληροφορίες για τον τομέα, γενικά κλιματολογικά στοιχεία, πληροφορίες για τη δράση των συνεργείων κλπ.

##### 4.2.2.1 Μετρήσεις δακοπληθυσμού

Ο βασικότερος τρόπος παράθεσης των στοιχείων μέτρησης των δακοπληθυσμών είναι η «Κατάσταση», η οποία είναι μία καρτέλα μεγέθους περίπου Α3, στην οποία σημειώνεται χειρόγραφα το όνομα και η θέση κάθε παγίδας, ο αριθμός των συλληφθέντων δάκων (αρσενικά + θηλυκά) για κάθε πενήτημερο και στο τέλος της στήλης, το σύνολο των συλληφθέντων δάκων για το κάθε πενήτημερο. Επίσης, στις περισσότερες εκθέσεις, υπάρχει και άλλη μία καρτέλα ίδιου μεγέθους, όπου καταγράφεται ο Μέσος Όρος Δάκων ανά Παγίδα, καθώς κι ένα γράφημα που απεικονίζει τις αλλαγές στον πληθυσμό των δάκων (Μ.Ο. Δάκων / Παγίδα) στη διάρκεια του χρόνου και τους ψεκασμούς που έλαβαν χώρα.

The figure consists of two forms. The left form is titled 'ΚΑΤΑΣΤΑΣΗ' and is for the 'Α' Συνεργείο του Τομέα Γαλάτιστας το 2010'. It has columns for 'ΙΟΥΝΙΟΣ' and 'ΙΟΥΛΙΟΣ' (days 1-6) and rows for trap types like 'Μαύρα', 'Κανίτες', and 'Φαλιό'. The right form is also titled 'ΚΑΤΑΣΤΑΣΗ' and is for the 'Α' Συνεργείο του Τομέα Γαλάτιστας το 2010'. It has columns for 'ΑΥΓΟΥΣΤΟΣ', 'ΣΕΠΤΕΜΒΡΙΟΣ', and 'ΟΚΤΩΒΡΙΟΣ' (days 1-6) and rows for trap types like 'Μαύρα', 'Κανίτες', and 'Φαλιό'. Both forms have a 'ΣΥΝΟΛΟ' row at the bottom.

**Εικόνα 8:** Η «Κατάσταση» για το Α' Συνεργείο του Τομέα Γαλάτιστας το 2010, όπως αυτή ψηφιοποιήθηκε σε δύο μέρη



**Εικόνα 9:** Η καρτέλα με τους Μ.Ο. Δάκων / Παγίδα και το γράφημα για το Β' Συνεργείο του Τομέα Γαλάτιστας το 2010, όπως αυτή ψηφιοποιήθηκε σε δύο μέρη

Στις εκθέσεις του έτους 2009 συμπεριλαμβάνονταν και μία σειρά Καταστάσεων με το όνομα «ΕΝΤΥΠΟ 1», σε μηχανογραφημένη μορφή όπου για κάθε πενήθερο σημειώνεται ο Μέσος Όρος Αρσενικών και Θηλυκών, ξεχωριστά, ανά Παγίδα, για κάθε συνεργείο. Αυτού του είδους τα έγγραφα δεν συμπεριλαμβάνονταν στις εκθέσεις των άλλων ετών και ως εκ τούτου δεν ήταν δυνατόν να έχουμε τους πληθυσμούς αρσενικών και θηλυκών ξεχωριστά και κατά συνέπεια ό,τι άλλο προκύπτει από αυτούς (πχ αναλογία θηλυκών/αρσενικά κλπ).

ΕΝΤΥΠΟ 1 ΚΑΤΑΣΤΑΣΗ  
ΣΥΝΑΡΤΗΣΗ ΑΡΣΕΝΙΚΩΝ (ΑΡΣΕ-ΘΗΛ) ΑΝΑ ΤΟΜΕΑ ΚΑΙ ΔΗΗΘΙΚΟ ΔΙΑΜΕΡΙΣΜΑ

Έτος: 2009  
Μήνας: ΟΚΤΩΒΡΙΟΣ

ΠΟΚΙΝΗΣ ΒΛΑΣΙ: ΧΑΛΚΙΔΙΚΗΣ

ΝΟΜΟΣ: Χαλκιδικής  
ΕΠΙΧΕΙΡΗΣΙΑ: Ν. ΜΑΡΜΑΡΑ  
ΤΟΜΕΑΣ: Ν. ΜΑΡΜΑΡΑ  
ΤΟΜΕΑΡΧΗΣ: Κ. ΒΑΣΙΛΕΙΑΔΟΥ

ΔΗΗΘΙΚΟ ΔΙΑΜΕΡΙΣΜΑ	ΑΡΙΘΜΟΣ ΕΠΙΧΕΙΡΗΣΕΩΝ	ΑΡΙΘΜΟΣ ΠΑΓΙΔΩΝ	Μ.Ο. ΔΑΚΩΝ ΑΝΑ ΠΑΓΙΔΑ ΚΑΙ ΠΕΝΘΗΜΕΡΟ ΑΝΑ ΔΙΑΜΕΡΙΣΜΑ	
			Αρσενικών	Θηλυκών
Ν. ΜΑΡΜΑΡΑ	05000	01	19,47	17,02
Ν. ΜΑΡΜΑΡΑ	05000	01	14,02	14,35
ΣΥΝΟΛΟ			17,15	15,85
		Μ.Ο. ΣΥΝΟΛΟΥ	32,83	

Το Μ.Ο. ΤΟΥ ΣΥΝΟΛΟΥ ΕΙΝΑΙ ΤΟ ΑΘΡΩΣΜΑ Μ.Ο. ΑΡΣΕΝΙΚΩΝ + Μ.Ο. ΘΗΛΥΚΩΝ Ο ΤΟΜΕΑΡΧΗΣ (ΥΠΟΓΡΑΦΗ)

© 2009 ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ - ΔΙΕΥΘΥΝΣΗ ΓΕΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ - ΜΑΡΜΑΡΑ 1

**Εικόνα 10:** Το ΕΝΤΥΠΟ 1 για το 1ο Πενθήμερο του Οκτωβρίου στην Έκθεση του τομέα Ν. Μαρμαρά 2009

Σε κάποιες εκθέσεις επίσης, τα αποτελέσματα των μετρήσεων του δακοπληθυσμού παρουσιάζονταν και σε μηχανογραφημένους πίνακες ή σε μηχανογραφημένα διαγράμματα. Σε αυτές τις περιπτώσεις, είτε καταγράφονταν ο Μέσος Όρος Δάκων ανά Παγίδα είτε και ειδικότερα ο Μέσος Όρος Αρσενικών και Θηλυκών ανά παγίδα, για κάθε πενθήμερο μέτρησης.

#### 4.2.2.2 Στοιχεία δειγματοληψιών για τη μέτρηση της δακοπροσβολής

Τα αποτελέσματα των δειγματοληψιών καταγράφονται σε μικρούς πίνακες με τη μορφή **ποσοστού προσβεβλημένων καρπών**. Τις περισσότερες φορές καταγράφονται ως εκατοστιαίο ποσοστό, ενώ κάποιες άλλες ως κλάσμα του 1.

**ΓΕΝΙΚΗ ΠΡΟΣΒΟΛΗ ΚΑΙ ΠΟΣΟΣΤΟ ΔΑΚΟΠΡΟΣΒΟΛΗΣ ΕΠΙ ΤΟΙΣ % ΤΩΝ ΕΠΗΜΕΡΟΥΣ ΣΤΑΔΙΩΝ ΩΣ ΠΡΟΣ ΤΗ ΓΕΝΙΚΗ ΠΡΟΣΒΟΛΗ ΣΕ ΔΕΙΓΜΑΤΑ ΕΛΑΙΟΚΑΡΠΟΥ**

ΤΟΜΕΑΣ: ΠΟΡΤΟ ΚΟΥΦΟ  
ΣΥΝΕΡΓΕΙΟ: ██████████  
ΤΟΜΕΑΡΧΗΣ: ██████████

α/α	ΜΗΝΑΣ	ΠΡΟΝΥΜΦ Η 1 <sup>η</sup> ΣΤΑΔΙΟΥ (1)	ΠΡΟΝΥΜΦ Η 2 <sup>η</sup> ΣΤΑΔΙΟΥ (2)	ΝΥΜΦΗ (3)	ΕΣΘΛΟΣ (4)	ΓΟΝΙΜΗ ΠΡΟΣΒΟΛΗ (1+2+3+4) (5)	ΞΕΡΟΒΟΥΛΑ Δ (6)	ΑΓΟΝΑ ΝΥΓΜΑΤΑ (7)	ΣΥΝΟΛΙΚΗ ΠΡΟΣΒΟΛΗ (5+6+7) (8)
1	ΙΟΥΛΙΟΣ						7,25	1,63	8,88
2	ΑΥΓΟΥΣΤΟΣ	1,58			0,22	1,8	13,58	1,13	16,51
3	ΣΕΠΤΕΜΒΡΙΟΣ	2,35			1,02	3,37	16,08	1,27	20,72
4	ΟΚΤΩΒΡΙΟΣ	1,42	1,3	2,2	2,08	7	17,23	1,25	25,48

**Εικόνα 11:** Πίνακας καταγραφής της δακοπροσβολής του συνεργείου Πόρτο Κουφό του τομέα Δήμου Τορώνης για το 2011. Τα αποτελέσματα καταγράφονται ως εκατοστιαία ποσοστά.

Σημειώνεται ότι σε κάποιες εκθέσεις δεν βρέθηκαν πίνακες καταγραφής της δακοπροσβολής, καθώς και σε κάποιες άλλες τα παρατιθέμενα στοιχεία ήταν ασαφή (δεν μπορούσε να διευκρινιστεί αν καταγράφουν εκατοστιαίο ποσοστό ή κλάσμα του 1). Επίσης, το **2009 δεν καταγράφονται** καθόλου οι **δευτερογενείς προσβολές**, ενώ αντίθετα είναι **το μοναδικό έτος** στο οποίο καταγράφονται οι καρποί στους οποίους βρέθηκαν **αυγά**.

#### 4.2.2.3 Στοιχεία για τους διενεργηθέντες ψεκασμούς

Οι ημερομηνίες έναρξης και λήξης των ψεκασμών φαίνονται τόσο στην «Κατάσταση» όσο και στην καρτέλα με το γράφημα (βλ. παράγραφο 4.2.2.1). Στις περισσότερες εκθέσεις δακοκτονίας καταγράφονται και πληροφορίες για την ποσότητα του ψεκαστικού υγρού και του φαρμάκου και της ελκυστικής ουσίας που χρησιμοποιήθηκαν και, σε κάποιες εκθέσεις, τον αριθμό δέντρων που ψεκάστηκαν και

την έκταση της περιοχής που ψεκάστηκε. Οι πληροφορίες αυτές καταγράφονται είτε περιφραστικά σε μορφή κειμένου είτε (σπανιότερα) σε μορφή πίνακα.

### 4.2.3 Καταγραφή στοιχείων και καταχώρησή τους στη Βάση Δεδομένων

Κατασκευάστηκαν τρία αρχεία .xls με ισάριθμους πίνακες που διέθεταν τα ίδια πεδία με τους Πίνακες ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ, ΔΕΙΓΜΑΤΟΛΗΨΙΑ και ΨΕΚΑΣΜΟΙ (τα αρχεία ονομάστηκαν new\_entries\_ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ, new\_entries\_ΔΕΙΓΜΑΤΟΛΗΨΙΑ και new\_entries\_ΨΕΚΑΣΜΟΣ αντίστοιχα). Η καταχώρηση των στοιχείων στους Πίνακες αυτούς έγιναν με πληκτρολόγηση. Σε κάθε εγγραφή στους παραπάνω Πίνακες, συμπληρώνονταν τα πεδία DHMOS\_ID και TOMEARX\_ID.

Για να υπάρχει η μεγαλύτερη δυνατή συμβατότητα με τους πίνακες που είχαν παραχωρηθεί από το ΥΠΑΑΤ και με τη Βάση Δεδομένων που δημιουργήσαμε, προστέθηκαν νέες εγγραφές στους Βοηθητικούς Πίνακες ΕΠΟΠΤΕΙΕΣ, ΤΟΜΕΙΣ, ΠΕΡΙΟΔΟΙ και ΤΟΜΕΑΡΧΕΣ (βλ. παράγραφο 4.1.1.4). Οι καταχωρήσεις των ID (πχ. TOMEARX\_ID) δεν ακολούθησαν τη λογική της Βάσης DACUS του Υπουργείου (πενταψήφιος κωδικός που αυξάνεται κατά ένα), αλλά του τριψήφιου ή τετραψήφιου κωδικού όπου στο πρώτο μέρος φαίνεται το έτος και στο δεύτερο η σειρά καταγραφής (πχ. **TOM\_ID = 901** είναι η πρώτη εγγραφή Τομέα για το έτος 2009, **TOM\_ID = 1102** είναι η δεύτερη εγγραφή Τομέα για το έτος 2011).

#### 4.2.3.1 Μετρήσεις δακοπληθυσμού

Για λόγους που επεξηγούνται αναλυτικά σε επόμενο κεφάλαιο (βλ. παρ. 5.1.1.4), η καταχώρηση των στοιχείων έγινε **ανά συνεργείο**. Ο αύξων αριθμός του συνεργείου (για ίδιο DACUS\_ID), καταχωρήθηκε στη στήλη DHMOS\_AA.

Στις περιπτώσεις που καταγράφεται μόνο ο Μέσος Όρος Δάκων ανά Παγίδα (βλ. παράγραφο 4.2.2.1), συμπληρώθηκε μόνο το πεδίο MO\_DACUS. Στις περιπτώσεις που καταγράφονται και οι αριθμοί ατόμων αρσενικών και θηλυκών (πεδία MALE και FEMALE), το πεδίο FEMALE\_POSOSTO συμπληρώνονταν αυτόματα με συνάρτηση (πχ. =H743/(G743+H743)). Τέλος, για τις περιπτώσεις που καταγράφονταν οι μέσοι όροι αρσενικών και θηλυκών ανά παγίδα, προστέθηκαν δύο νέες στήλες με ονόματα “male”

και “female”, όπου καταχωρούνταν οι μέσοι όροι, ενώ τα πεδία “MALE” και “FEMALE” συμπληρώνονταν αυτόματα με συνάρτηση η οποία πολλαπλασίαζε το μέσο όρο ανά παγίδα επί τον εκάστοτε αριθμό παγίδων. Στο τέλος, ο πίνακας αντιγράφηκε σε καινούργιο Φύλλο Εργασίας, δημιουργήθηκαν καινούργιες στήλες “MALE”, “FEMALE” και “FEMALE\_POSOSTO” στις οποίες έγινε **Ειδική Επικόλληση** των τιμών των παλιών ομώνυμων στηλών, οι οποίες στη συνέχεια Διαγράφηκαν μαζί με τις στήλες “male” και “female”. Έτσι ο τελικός πίνακας είχε τα ίδια πεδία και με την ίδια σειρά με τον Πίνακα ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ της Βάσης Δεδομένων.

#### 4.2.3.2 Στοιχεία δειγματοληψιών για τη μέτρηση της δακοπροσβολής

Η καταχώρηση των στοιχείων έγινε **ανά συνεργείο**, όπως ακριβώς είναι καταγεγραμμένα και στις Εκθέσεις Δακοκτονίας. Ο αύξων αριθμός του συνεργείου (για ίδιο DACUS\_ID), καταχωρήθηκε στη στήλη DEIGMA\_AA.

Όπως αναφέρθηκε στην παράγραφο 4.2.2.2, τα αποτελέσματα των δειγματοληψιών στις Εκθέσεις Δακοκτονίας καταγράφονται με τη μορφή **ποσοστού προσβεβλημένων καρπών**. Αντίθετα, στα δεδομένα που δόθηκαν από το ΥΠΑΑΤ οι αριθμοί απεικονίζουν τον **αριθμό προσβεβλημένων καρπών**. Για να υπάρχει ομοιομορφία στις εγγραφές, προστέθηκαν, στον πίνακα του αρχείου new\_entries\_ΔΕΙΓΜΑΤΟΛΗΨΙΑ, νέες στήλες στις οποίες πληκτρολογούνταν τα *εκατοστιαία ποσοστά προσβεβλημένων καρπών*, ενώ η καταχώρηση του *αριθμού προσβεβλημένων καρπών* υπολογίζονταν αυτόματα, δεχόμενοι ότι το δείγμα είναι πάντα 960 καρποί (βλ. παράγραφο 1.1.4).

Στους πίνακες καταγραφής της δακοπροσβολής, ενώ καταγράφεται ο μήνας, δεν καταγράφεται ποτέ η ημερομηνία δειγματοληψίας, ενώ δεν καταγράφεται τέτοια πληροφορία σε καμία Έκθεση Δακοκτονίας. Στον αντίστοιχο Πίνακα του Υπουργείου παρατηρήθηκε ότι οι περισσότερες δειγματοληψίες γίνονταν προς το τέλος του μήνα, (στο νομό Χαλκιδικής όλες). Για το λόγο αυτό, δόθηκε **συμβατική ημερομηνία δειγματοληψίας**, σε όλες τις εγγραφές **η 30<sup>η</sup> του εκάστοτε μήνα**.

Ο τελικός πίνακας διαμορφώθηκε με τρόπο παρόμοιο με αυτόν που περιγράφεται στην προηγούμενη παράγραφο (4.2.3.1).

Σημειώνεται ότι στις περιπτώσεις που δεν βρέθηκαν πίνακες καταγραφής της δακοπροσβολής, ή που τα παρατιθέμενα στοιχεία ήταν ασαφή (βλ. παράγραφο 4.2.2.2),

προστέθηκαν εγγραφές στις οποίες όλα τα πεδία καταγραφής των αποτελεσμάτων ήταν κενά.

#### 4.2.3.3 Στοιχεία για τους διενεργηθέντες ψεκασμούς

Αν και τα στοιχεία που παρατίθενται στις εκθέσεις είναι ανά συνεργείο, η καταχώρηση έγινε **ανά Δημοτικό Διαμέρισμα**, όπως είναι καταχωρημένα στις εγγραφές που ήδη υπήρχαν στη Βάση Δεδομένων.

Η καταχώρηση έγινε ως εξής:

- Στο πεδίο *FAR\_ID* καταχωρήθηκε **το όνομα** του φαρμάκου που χρησιμοποιήθηκε, και όχι ο κωδικός όπως ήταν στα δεδομένα που παραχωρήθηκαν από το ΥΠΑΑΤ (βλ. παράγραφο 4.1.1.3).
- Στο πεδίο *SPRAY\_STDATE*, καταχωρήθηκε η πρωιμότερη ημερομηνία έναρξης ψεκασμού μεταξύ των συνεργείων του ίδιου τομέα. Πχ, σε ένα τομέα με τρία συνεργεία, αν στο Συνεργείο Α ο ψεκασμός άρχισε στις 7-8-2010, στο Συνεργείο Β στις 6-8-2010 και στο Συνεργείο Γ στις 8-8-2010, καταχωρείται η ημερομηνία 6-8-2010).
- Αντίστοιχα στο πεδίο *SPRAY\_ENDATE*, καταχωρήθηκε η οψιμότερη ημερομηνία λήξης ψεκασμού μεταξύ των συνεργείων του ίδιου τομέα.
- Στα πεδία *SPRAYED\_TREES*, *LIQUID\_USED*, *FARMAK\_USED*, *PULL\_USED* και *LAND\_PROTECT*, οι καταχωρήσεις των στοιχείων –όπου αυτά υπήρχαν- γίνονταν κατ' εκτίμηση:
  - Στις περιπτώσεις εκθέσεων όπου τα στοιχεία καταγράφονταν σε μορφή πίνακα, καταχωρήθηκε το άθροισμα των αντίστοιχων τιμών για τα συνεργεία (πχ σε ένα τομέα με δύο συνεργεία, αν στο Συνεργείο Α χρησιμοποιήθηκαν 2 lt φαρμάκου και στο Συνεργείο Β χρησιμοποιήθηκαν 3 lt φαρμάκου, τότε στο Πεδίο *FARMAK\_USED* καταχωρείται η τιμή 5(=2+3)).
  - Επίσης, στις περιπτώσεις όπου τα στοιχεία καταγράφονται περιφραστικά σε μορφή κειμένου, συχνά οι τιμές δίνονταν ανά ημέρα, οπότε γίνονταν αναγωγή πολλαπλασιάζοντας επί το σύνολο των ημερών (πχ. Σε Τομέα με τρία Συνεργεία ο ψεκασμός έγινε μεταξύ 7-8-2010 και

14-8-2010 (8 ημέρες) και χρησιμοποιήθηκαν 800 lt ψεκαστικού υγρού ανά ημέρα για κάθε συνεργείο. Σε αυτή την περίπτωση καταχωρείται η τιμή 19200(= 8 x 800x 3)

## 4.3 Μετεωρολογικά Δεδομένα

Δεδομένου ότι, όπως αναφέρθηκε και παραπάνω, η δραστηριότητα των ακμαίων και η διάρκεια του κάθε σταδίου ανάπτυξης εξαρτάται από την θερμοκρασία περιβάλλοντος και την εξωτερική σχετική υγρασία, αποφασίστηκε η συλλογή πληροφοριών για τις μετεωρολογικές συνθήκες που επικράτησαν, ανά ημέρα, στις περιοχές που μελετάμε, κατά τις περιόδους που λάμβανε χώρα το Πρόγραμμα Δακοκτονίας.

### 4.3.1 Στοιχεία από το Εθνικό Αστεροσκοπείο Αθηνών

Στο διαδίκτυο, στον ιστότοπο του Εθνικού Αστεροσκοπείου Αθηνών, υπάρχει ελεύθερη πρόσβαση στη Βάση Δεδομένων **meteoSearch** και συγκεκριμένα στην ιστοσελίδα <http://penteli.meteo.gr/meteosearch/>. Εκεί μπορεί ο καθένας να αναζητήσει και να βρει τα δεδομένα μετρήσεων από ένα μεγάλο δίκτυο μετεωρολογικών σταθμών. Επιλέγοντας μετεωρολογικό σταθμό, μήνα και έτος, αποκτά πρόσβαση σε πίνακα που για κάθε μέρα του μήνα παραθέτει πληροφορίες που αφορούν στη θερμοκρασία, τη βροχόπτωση και τον άνεμο.



Εικόνα 12: Η αρχική σελίδα του meteoSearch

MONTHLY CLIMATOLOGICAL SUMMARY for AUG. 2009

NAME: straton1 CITY: STATE:  
ELEV: 0 m LAT: LONG:

TEMPERATURE (°C), RAIN (mm), WIND SPEED (km/hr)

DAY	MEAN TEMP	HIGH	TIME	LOW	TIME	HEAT DEG DAYS	COOL DEG DAYS	RAIN	AVG WIND SPEED	HIGH	TIME	DOM DIR
1												
2												
3												
4	28.2	29.7	19:50	25.5	00:00	0.0	3.7	0.0	3.1	12.9	15:10	E
5	26.8	29.4	18:10	23.9	4:30	0.0	8.4	0.0	5.6	24.1	12:20	ESE
6	27.2	30.3	14:10	23.2	5:30	0.0	8.8	0.0	7.2	41.8	9:20	WNW
7	27.5	29.8	17:40	24.5	7:20	0.0	9.2	0.0	7.7	29.0	11:20	N
8	26.3	28.3	17:20	23.3	3:20	0.0	7.9	0.0	6.4	29.0	21:20	ESE
9	25.8	28.7	19:20	21.7	6:40	0.0	7.5	0.0	5.8	27.4	10:00	ESE
10	24.0	26.2	1:50	20.5	11:10	0.0	5.7	0.0	4.5	32.2	2:10	NE
11	23.9	26.7	14:30	20.0	00:00	0.0	5.6	4.0	7.7	22.5	8:00	ESE
12	22.4	25.4	15:20	18.3	5:20	0.0	4.1	0.0	5.0	30.6	12:00	ESE
13	24.6	28.5	18:40	19.8	7:10	0.0	6.2	0.0	4.7	24.1	3:00	WNW
14	24.6	27.4	16:00	20.9	7:00	0.0	6.3	0.0	4.2	19.3	10:10	ESE
15	25.4	29.3	13:20	22.3	9:50	0.0	7.1	0.4	4.7	25.7	13:30	WNW
16	27.7	31.5	19:10	22.3	6:10	0.0	9.3	0.0	6.6	30.6	14:30	NNE
17	26.6	29.6	12:10	23.2	6:50	0.0	8.2	0.0	5.3	25.7	12:40	NNE
18	25.8	30.5	17:20	21.7	7:00	0.0	7.4	0.0	4.0	17.7	12:50	W
19	27.2	33.2	12:40	22.1	6:20	0.0	8.8	0.0	4.3	24.1	11:00	E
20	27.9	32.2	12:20	23.4	4:10	0.0	9.6	0.0	6.4	30.6	11:40	ESE
21	25.3	28.1	18:40	21.7	7:00	0.0	7.0	0.0	6.3	29.0	0:40	ESE
22	24.4	27.7	19:20	20.5	7:10	0.0	6.1	0.0	4.7	29.0	10:30	W
23	24.7	28.2	10:40	20.6	7:00	0.0	6.4	0.0	5.0	24.1	11:20	ESE
24	24.6	27.4	19:30	20.9	5:20	0.0	6.3	0.0	5.6	32.2	10:40	ESE
25	24.5	27.9	14:00	20.9	8:50	0.0	6.2	0.4	7.1	29.0	4:20	N
26	25.1	29.1	13:20	20.6	3:30	0.0	6.8	0.0	5.3	25.7	14:00	WNW
27	25.1	29.2	11:20	20.6	6:50	0.0	6.7	0.0	5.1	25.7	12:40	WNW
28	23.9	27.6	18:40	20.1	7:00	0.0	5.6	0.0	4.5	12.9	10:50	WNW
29	24.1	27.2	16:40	20.3	6:30	0.0	5.7	0.0	4.7	12.9	10:10	WNW
30	24.6	27.6	12:20	21.1	6:50	0.0	6.2	0.2	4.3	16.1	00:00	W
31	24.9	27.3	14:10	22.4	7:20	0.0	6.6	0.2	4.7	22.5	4:10	ESE
25.5	33.2	19	18.3	12	0.0	193.5	5.2	5.4	41.8	6	ESE	

Max >= 32.0: 2  
Max <= 0.0: 0  
Min <= 0.0: 0  
Min <= -15.0: 0  
Max Rain: 3.99 ON 11/08/09  
Days of Rain: 5 (> .2 mm) 1 (> 2 mm) 0 (> 20 mm)  
Heat Base: 18.3 Cool Base: 18.3 Method: Integration

**Εικόνα 13:** Πίνακας με τα μετεωρολογικά στοιχεία του Μετεωρολογικού Σταθμού Στρατωνίου (Χαλκιδικής) για τον Αύγουστο του 2009

Αποφασίστηκε να γίνει συλλογή πληροφοριών από τους μετεωρολογικούς σταθμούς που βρίσκονται στα όρια των νομών που εξετάζουμε και συγκεκριμένα τους σταθμούς: **Κασσανδρείας, Πολυγύρου, Στρατωνίου και Μαρμαρά (Ν. Χαλκιδικής)** και **Καρλοβασιού και Ραχών Ικαρίας (Ν. Σάμου)**, για τους μήνες **Ιούνιο έως Νοέμβριο** των ετών **2009 – 2015**. Σε τηλεφωνική επικοινωνία με το αρμόδιο τμήμα του Εθνικού Αστεροσκοπείου διαπιστώθηκε ότι υπάρχουν και αντίστοιχα στοιχεία που αφορούν στην εξωτερική υγρασία. Τα στοιχεία αυτά ζητήθηκαν με γραπτή αίτηση και μας εστάλησαν σε μορφή πινάκων παρόμοια με αυτή των πινάκων που βρήκαμε στον ιστότοπο του meteosearch. Μαζί με αυτά εστάλη κι ένας πίνακας όπου καταγράφεται για καθένα από τους ζητούμενους σταθμούς ο **κωδικός** του, οι **συντεταγμένες** του και το **υψόμετρο** στο οποίο βρίσκεται.

Σημειώνεται πάντως ότι ο σταθμός της Κασσανδρείας λειτουργεί από το 2011, του Στρατωνίου και της Ικαρίας από τον Αύγουστο του 2009 και του Μαρμαρά μόλις από το 2014.



### 4.3.2 Καταχώρηση στοιχείων στη Βάση Δεδομένων

Τα δεδομένα αυτά συγκεντρώθηκαν σε ένα αρχείο .xls. Σε αυτό κατασκευάστηκαν πίνακες που περιείχαν τα πεδία τόσο από τους πίνακες που υπήρχαν στο διαδίκτυο όσο και από αυτούς που μας εστάλησαν. Για λόγους καλύτερης οργάνωσης, κατασκευάστηκαν ξεχωριστοί πίνακες ανά μετεωρολογικό σταθμό, σε ξεχωριστά φύλλα εργασίας.

Στη συνέχεια, στη Βάση Δεδομένων δημιουργήθηκε ο Πίνακας **ΜΕΤΕΩΡΟΛΟΓΙΚΑ ΣΤΟΙΧΕΙΑ**, με τα ίδια ακριβώς πεδία, με την ίδια σειρά, τα οποία είναι τα εξής:

*a/a*

*METEO\_ID*: ο κωδικός του σταθμού. Ταυτίζεται με τον κωδικό που χρησιμοποιεί το Αστεροσκοπείο

*DATE*

*MEAN\_TEMP, HIGH\_TEMP, LOW\_TEMP*: Μέση, υψηλότερη και χαμηλότερη θερμοκρασία ημέρας

*TIME\_HIGH\_TEMP, TIME\_LOW\_TEMP*: Ώρα που σημειώθηκε η υψηλότερη και η χαμηλότερη θερμοκρασία

*HEAT\_DEG\_DAYS, COOL\_DEG\_DAYS*: Βαθμοημέρες θέρμανσης και βαθμοημέρες ψύξης

*RAIN*: Χιλιοστά βροχής

*AVG\_WIND\_SPEED, HIGH\_WIND\_SPEED, TIME\_HIGH\_WIND\_SPEED*: Μέση και υψηλότερη ταχύτητα ανέμου και ώρα όπου σημειώθηκε η υψηλότερη

*DOM\_DIR*: Επικρατούσα κατεύθυνση ανέμου

*OUT\_HUM, OUT\_HUM\_MAX, OUT\_HUM\_MIN*: Μέση, υψηλότερη και χαμηλότερη εξωτερική σχετική υγρασία.

Έπειτα τα δεδομένα του αρχείου .xls καταχωρήθηκαν στον Πίνακα.

Το Πεδίο *METEO\_ID* του παραπάνω πίνακα ορίστηκε ως εξωτερικό κλειδί το οποίο δείχνει στο ομώνυμο πεδίο ενός επίσης καινούργιου Πίνακα που δημιουργήθηκε με το όνομα **ΜΕΤΕΩΡΟΛΟΓΙΚΟΙ ΣΤΑΘΜΟΙ**, όπου καταγράφονται στοιχεία για τους μετεωρολογικούς σταθμούς. Τα υπόλοιπα πεδία του Πίνακα είναι τα εξής:

*a/a*

*METEO\_CODE*: Κωδικός με τέσσερα αλφαριθμητικά. Τα πρώτα δύο είναι χαρακτηριστικά του νομού στον οποίο βρίσκεται ο μετεωρολογικός σταθμός και τα

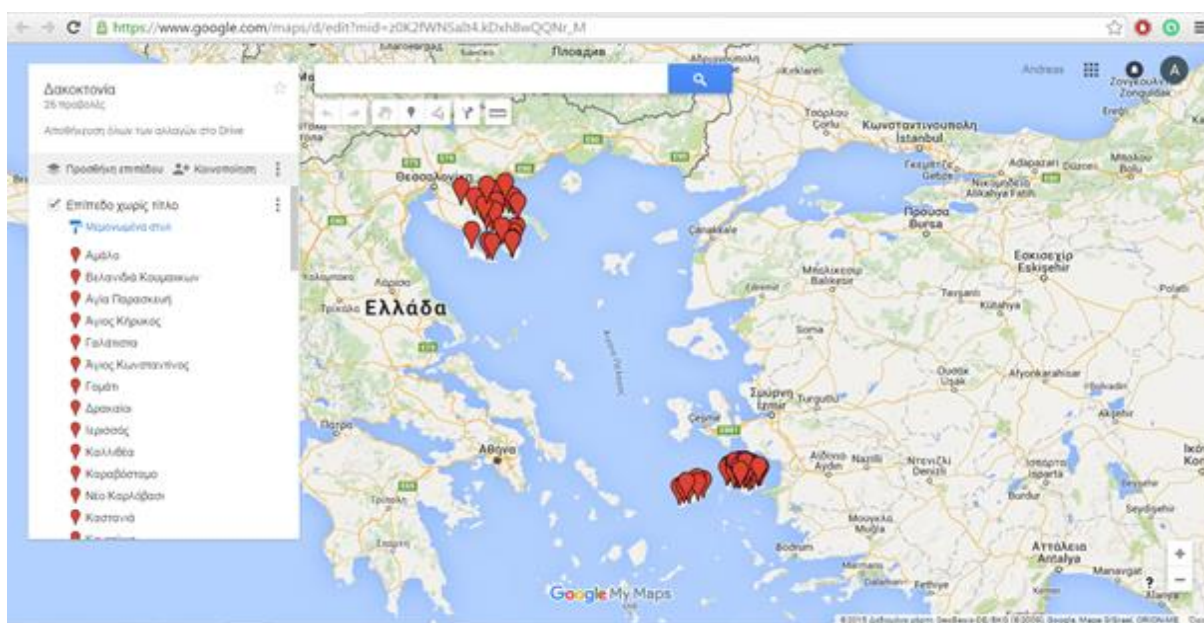
επόμενα δύο ο αύξων αριθμός του σταθμού μέσα στο νομό του. πχ sa01 είναι ο πρώτος σταθμός που καταγράφηκε του νομού Σάμου, ενώ ch04, ο τέταρτος σταθμός που καταγράφηκε του νομού Χαλκιδικής.

*METEO\_DESCR*: Όνομα περιοχής που βρίσκεται ο σταθμός

*METEO\_LAT*, *METEO\_LON*, *YPSOMETRO*: Γεωγραφικό Πλάτος, Γεωγραφικό Μήκος, Υψόμετρο

### 4.3.3 Συσχέτιση του Πίνακα ΜΕΤΕΩΡΟΛΟΓΙΚΑ ΣΤΟΙΧΕΙΑ με άλλους Πίνακες της Βάσης Δεδομένων

Καταρχήν, με τη βοήθεια του Google Maps, επισημάνθηκαν στο χάρτη όλα τα Δημοτικά Διαμερίσματα στα οποία έχουν γίνει μετρήσεις, δειγματοληψίες και ψεκασμοί, καθώς και όλοι οι μετεωρολογικοί σταθμοί που καταγράφονται στη Βάση Δεδομένων, και δημιουργήθηκε ένα αρχείο Google My Maps, το οποίο φαίνεται στην παρακάτω εικόνα:

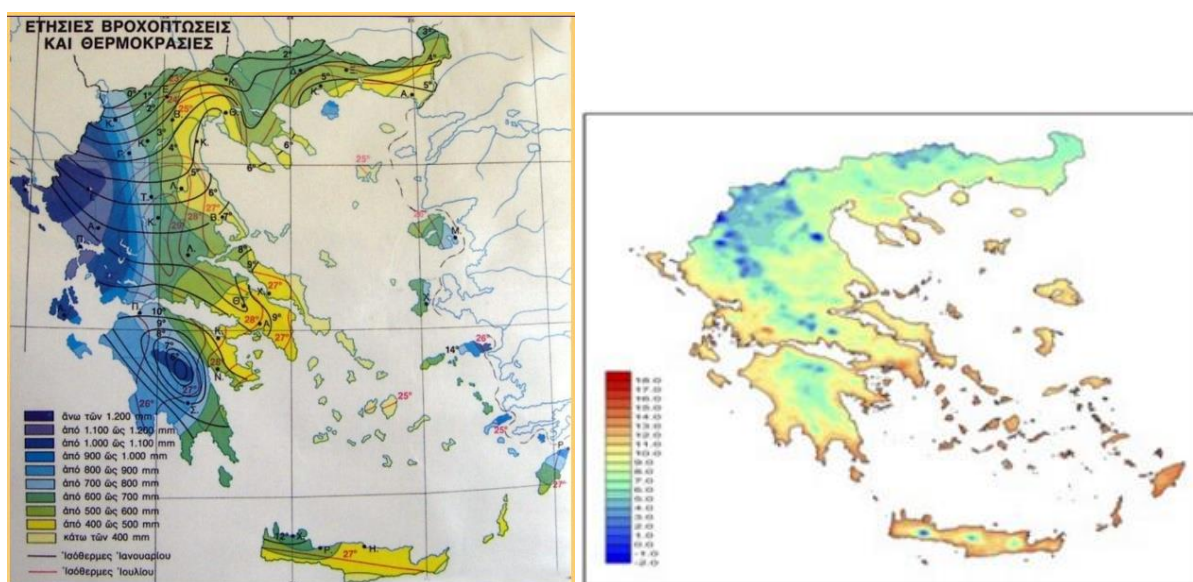


**Εικόνα 14:** Ο χάρτης Google My Maps με επισημασμένα όλα τα Δημοτικά Διαμερίσματα που έχουν λάβει μέρος σε Προγράμματα Δακοκτονίας και τους Μετεωρολογικούς Σταθμούς

Στον Πίνακα ΔΗΜΟΙ, προστέθηκαν τα πεδία *METEO\_ID* και *ALTERN\_METEO\_ID*. Το πρώτο καταγράφει τον εγγύτερο μετεωρολογικό σταθμό στο Δήμο. Επειδή, όπως αναφέρεται στο τέλος της παραγράφου 4.3.1, κάποιοι σταθμοί άρχισαν να λειτουργούν αργότερα, καθώς και επειδή πολλές φορές, λόγω βλαβών, δεν γίνονταν καταγραφή

δεδομένων για μερικές μέρες, προστέθηκε και η δεύτερη στήλη που καταγράφει έναν εναλλακτικό μετεωρολογικό σταθμό. Ταυτόχρονα, προστέθηκαν και τα πεδία DHMOS\_LAN και DHMOS\_LOT στα οποία καταγράφονται οι γεωγραφικές συντεταγμένες του Δήμου. Τα τέσσερα νέα πεδία δεν συμπληρώθηκαν σε όλες τις εγγραφές, παρά μόνο σε αυτές των Δημοτικών Διαμερισμάτων που έχουν λάβει μέρος τουλάχιστον σε ένα από τα Προγράμματα Δακοκτονίας των ετών 2009 - 2015.

Ο ορισμός του εγγύτερου μετεωρολογικού σταθμού και του εναλλακτικού του, δεν έγινε με μοναδικό κριτήριο τη μικρότερη απόσταση. Για το σκοπό αυτό αναζητήθηκαν στο διαδίκτυο και βρέθηκαν χάρτες στους οποίους μπορεί κάποιος να διακρίνει κλιματολογικές ζώνες. Οι χάρτες αυτοί φαίνονται στην **Εικόνα 15**



**Εικόνα 15:** Χάρτες της Ελλάδας με κλιματολογικές ζώνες. Αριστερά απεικονίζονται ζώνες ετήσιων βροχοπτώσεων και θερμοκρασιών. Δεξιά μια γεωγραφική κατανομή της γραμμικής τάσης του ολικού ετήσιου ύψους βροχής (σε mm/year) για την περίοδο 1998-2011.

Κατόπιν, ο ορισμός του εγγύτερου μετεωρολογικού σταθμού και του εναλλακτικού του έγινε κατ' εκτίμηση, για κάθε Δημοτικό Διαμέρισμα χωριστά, με γνώμονα την προσεκτική οπτική παρατήρηση του χάρτη Google My Maps παράλληλα με τους δύο παραπάνω χάρτες, καθώς και υπολογίζοντας κατ' εκτίμηση το υψόμετρο του κάθε Δημοτικού Διαμερίσματος.

#### 4.3.4 Δημιουργία νέου ερωτήματος και συμπλήρωση πεδίων στα ήδη υπάρχοντα

Δημιουργήθηκε το Ερώτημα ΕΡΩΤ\_ΜΕΤΕΩΡ\_ΣΤΟΙΧΕΙΑ, το οποίο περιέχει όλα τα πεδία των Πινάκων ΜΕΤΕΩΡΟΛΟΓΙΚΑ\_ΣΤΟΙΧΕΙΑ και ΜΕΤΕΩΡΟΛΟΓΙΚΟΙ\_ΣΤΑΘΜΟΙ.

Επίσης, στα ερωτήματα ΕΡΩΤ\_ΜΕΤΡ\_ΔΑΚΩΝ και ΕΡΩΤ\_ΔΕΙΓΜΑΤΟΛΗΨΙΑ, προστέθηκαν τα πεδία ΜΕΤΕΟ\_ID και ALTERN\_ΜΕΤΕΟ\_ID από τον Πίνακα ΔΗΜΟΙ καθώς και τα πεδία ΜΕΤΕΟ\_DESCR και ALTERN\_ΜΕΤΕΟ\_DESCR τα οποία αντιστοιχούν στο πεδίο ΜΕΤΕΟ\_DESCR του Πίνακα ΜΕΤΕΩΡΟΛΟΓΙΚΟΙ\_ΣΤΑΘΜΟΙ και απεικονίζουν το όνομα του εγγύτερου και του εναλλακτικού μετεωρολογικού σταθμού. Για να επιτευχθεί αυτό, δόθηκαν στον Πίνακα ΜΕΤΕΩΡΟΛΟΓΙΚΑ\_ΣΤΟΙΧΕΙΑ, κατά τη δημιουργία του εκάστοτε ερωτήματος σε γλώσσα SQL, τα ονόματα “master.meteo” και “altern.meteo”. Το πεδίο ΜΕΤΕΟ\_ID του Πίνακα ΔΗΜΟΙ, δείχνει στο Πρωτεύον κλειδί ΜΕΤΕΟ\_ID του “master.meteo”, ενώ το πεδίο ALTERN\_ΜΕΤΕΟ\_ID του Πίνακα ΔΗΜΟΙ, δείχνει στο Πρωτεύον Κλειδί ΜΕΤΕΟ\_ID του “altern.meteo”.

# Κεφάλαιο 5

## Προεπεξεργασία, Ανάλυση Δεδομένων και Συμπεράσματα

Η προεπεξεργασία των δεδομένων έγινε κυρίως με τη χρήση τη στατιστικής γλώσσας προγραμματισμού R. Η R είναι ένα περιβάλλον εντολών που χρησιμοποιείται για ανάλυση δεδομένων, υπολογισμούς και δημιουργία γραφημάτων.

Πρόκειται για λογισμικό ανοικτού κώδικα (open source) και μπορεί να θεωρηθεί σα μια εφαρμογή της γλώσσας προγραμματισμού S. Χρησιμοποιείται είτε με τη χρήση απευθείας εντολών, είτε με τη δημιουργία προγραμμάτων.

Ο προγραμματισμός σε R έγινε με τη χρήση του περιβάλλοντος εργασίας RStudio.

### 5.1. Επεξεργασία πινάκων

#### 5.1.1. Δημιουργία και επεξεργασία των data frames

Τα ερωτήματα ΕΡΩΤ\_ΜΕΤΡ\_ΔΑΚΩΝ, ΕΡΩΤ\_ΔΕΙΓΜΑΤΟΛΗΨΙΑ, ΕΡΩΤ\_ΨΕΚΑΣΜΟΙ και ΕΡΩΤ\_ΜΕΤΕΩΡ\_ΣΤΟΙΧΕΙΑ (βλ. παρ. 4.1.2.4 και 4.3.4) επικολλώνται σε φύλλα εργασίας .xls, τα οποία αποθηκεύονται ως αρχεία .csv. Από αυτά τα αρχεία (*n\_metrisi\_dakon.csv*, *n\_deigmato.csv*, *n\_psekasmoi.csv*, *meteorologika\_stoixeia.csv*) δημιουργούνται τα πρώτα data frames για επεξεργασία στην R.

Με την εντολή `read.csv` δημιουργούνται τα αρχικά data frames **metr\_dakon**, **deigmato**, **psekasmoi** και **meteo**, από τους πίνακες των αντίστοιχων αρχείων csv. Π.χ.:

```
metr_dakon <- read.csv ("n_metrisi_dakon.csv", encoding='UTF-8')
```

#### 5.1.1.1. Μετατροπές και προσθήκες σχετικές με τις ημερομηνίες

Όλες οι στήλες που περιέχουν **ημερομηνίες** διαμορφώνονται σε μορφή αναγνωρίσιμη από την R ως ημερομηνία, για να μπορούν να συγκρίνονται μεταξύ τους. Π.χ.:

```
metr_dakon$CHANGE_DATE<-as.Date(metr_dakon$CHANGE_DATE , format="%m/%d/%y")
```

Έπειτα, προστίθεται η στήλη YEAR, όπου καταγράφεται το έτος στο οποίο αφορά η εκάστοτε εγγραφή και το οποίο εξάγεται από την ημερομηνία (πχ. `metr_dakon$YEAR<-format(metr_dakon$CHANGE_DATE , "%Y")` ).

Σε κάποιες περιπτώσεις καταγράφονται σε ξεχωριστές στήλες, και ο μήνας και το δεκαπενθήμερο (MONTH, HALF). Για τον υπολογισμό του δεκαπενθημέρου, δημιουργήθηκε η συνάρτηση *halfmon*, η οποία υπολογίζει αν βρισκόμαστε στο πρώτο ή στο δεύτερο δεκαπενθήμερο του εκάστοτε μήνα.

Επίσης, επειδή οι μετρήσεις του πληθυσμού του δάκου γίνονται ανά **πενθήμερο** (βλ. παρ. 1.1.3), αποφασίστηκε να χωριστεί η κάθε περίοδος σε πενήμερα, όπου το πρώτο ξεκινάει από τον Ιούλιο. Έτσι πχ η 16-8-2014, βρίσκεται στο 10<sup>ο</sup> πενήμερο του 2014. Θεωρώντας ότι οι τυπικές ημερομηνίες μέτρησης είναι οι 1,6,11,16,21 και 26 κάθε μήνα, ορίσαμε ως πενήμερο την περίοδο από 2 ημέρες πριν από αυτές τις ημερομηνίες έως 2 ημέρες μετά. Πχ, το πρώτο πενήμερο του 2014 είναι από 30-6-2014 έως 3-7-2014, ενώ το δεύτερο από 4-7-2014 έως 8-7-2014. Η 29<sup>η</sup> του εκάστοτε μήνα κατατάσσεται στο πενήμερο της 26<sup>ης</sup>. Σε κάθε πίνακα προστίθεται η στήλη PENTH, όπου καταγράφεται το πενήμερο στο οποίο αφορά η κάθε εγγραφή, το οποίο υπολογίζεται με ειδική συνάρτηση.

#### 5.1.1.2. Προσθήκες στηλών με τους μέσους όρους δάκων ανά παγίδα

Επειδή, όπως αναλύεται στην μεθεπόμενη παράγραφο, αποφασίστηκε ότι η επεξεργασία των δεδομένων πρέπει να γίνει ανά Δημοτικό Διαμέρισμα και επειδή σε κάθε Δημοτικό Διαμέρισμα έχει τοποθετηθεί διαφορετικός αριθμός παγίδων, στις περιπτώσεις που μελετήθηκαν οι μετρήσεις της δακοπροσβολής χρησιμοποιήθηκε ο Μέσος Όρος Δάκων ανά Παγίδα.

Για το σύνολο των δάκων χρησιμοποιήθηκε η στήλη MO\_DACUS. Για τα θηλυκά, τα αρσενικά και τα ζευγάρια (στήλη MIN), προστέθηκαν οι στήλες MO\_FEMALE, MO\_MALE και MO\_MIN αντίστοιχα. Οι παραπάνω στήλες υπολογίστηκαν με την καταρχήν

κατασκευή μιας επιπλέον στήλης με όνομα AR\_PAGIDON στην οποία έχει υπολογιστεί ο εκάστοτε αριθμός παγίδων ως εξής:

```
metr_dakon$AR_PAGIDON<-metr_dakon$SYNOLO_DAKON/metr_dakon$MO_DACUS
```

Με τη βοήθεια του Αριθμού Παγίδων κατασκευάστηκαν οι στήλες με τους μέσους όρους θηλυκών και αρσενικών και ζευγαριών ανά παγίδα. Πχ.:

```
metr_dakon$MO_FEMALE<- metr_dakon$FEMALE/metr_dakon$AR_PAGIDON
```

Υπενθυμίζουμε ότι στα δεδομένα του αρχείου της ΔΑΟΚ Χαλκιδικής, σε κάποιες περιπτώσεις καταγράφονταν οι μέσοι όροι θηλυκών και αρσενικών ανά παγίδα, ενώ σε ακόμα περισσότερες περιπτώσεις καταγράφονταν ο αριθμός των παγίδων (βλ. παρ. 4.2.2.1). Παρόλα αυτά οι πληροφορίες αυτές δεν είχαν καταγραφεί στη Βάση Δεδομένων επειδή δεν υπήρχε τέτοιο πεδίο στον πίνακα ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ.

### 5.1.1.3. Συναθροίσεις (aggregations) δεδομένων

Στους πίνακες (data frames) metr\_dakon και deigματο, τα δεδομένα συναθροίστηκαν ανά Δήμο και Ημερομηνία ή Πενθήμερο (ανάλογα με την περίπτωση). **Οι λόγοι που η συνάθροιση έγινε ανά Δήμο είναι οι εξής:**

- Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο, οι εγγραφές στον πίνακα ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ είναι ανά παγίδα μόνο για το νομό Σάμου το 2014.
- Στις Εκθέσεις Δακοκτονίας της ΔΑΟΚ Χαλκιδικής, στις «Καταστάσεις» (βλ. παρ. 4.2.2.1), οι πληθυσμοί των δάκων (θηλυκά + αρσενικά) είναι καταγεγραμμένοι ανά παγίδα. Για κάθε παγίδα καταγράφεται ο αύξων αριθμός, η περιοχή και σε κάποιες περιπτώσεις, το όνομα του κτήματος. Η καταγραφή των δεδομένων ανά παγίδα, θα προϋπέθετε ότι η κάθε παγίδα θα έπρεπε να εντοπιστεί σε όλες τις εκθέσεις, όλων των ετών για το συγκεκριμένο Δήμο που ανήκει και να κωδικοποιηθεί. Αυτό θα ήταν δυνατό στις εκθέσεις στις οποίες καταγράφονται και τα ονόματα των κτημάτων, αλλά η διαδικασία ταυτοποίησης και καταχώρησης θα ήταν χρονοβόρα, εφόσον δεν γίνεται αυτόματα και θα υπερέβαινε τα χρονικά περιθώρια περάτωσης της παρούσας διατριβής. Κάτι τέτοιο θα ήταν ιδιαίτερα χρήσιμο αν ήταν δυνατός και ο χωρικός εντοπισμός του κάθε κτήματος σε ένα χάρτη, (αν πχ δίνονταν οι γεωγραφικές συντεταγμένες κάθε κτήματος), ώστε να μελετηθούν οι μεταβολές της πυκνότητας του πληθυσμού στην πορεία του χρόνου. Δεν διατίθεται όμως τέτοια πληροφορία. Γι αυτούς τους λόγους η καταγραφή των στοιχείων έγινε ανά συνεργείο.

- Στις Εκθέσεις Δακοκτονίας της ΔΑΟΚ Χαλκιδικής, ενώ η καταγραφή των μετρήσεων δακοπληθυσμού και των δειγματοληψιών δακοπροσβολής έγινε ανά συνεργείο, συνήθως δεν μπορεί να εξακριβωθεί αν η χωρική έκταση του εκάστοτε συνεργείου για μία συγκεκριμένη χρονιά είναι ίδια για όλες τις χρονιές. Σε κάθε έκθεση δακοκτονίας, τα συνεργεία ονοματίζονται συνήθως ως «Συνεργείο Α», «Συνεργείο Β» κλπ, αλλά δεν κατέστη δυνατόν να εξακριβωθεί με ασφάλεια αν πχ το «Συνεργείο Α» της Μεγάλης Παναγιάς το 2009 είναι το ίδιο με το «Συνεργείο Α» της Μεγάλης Παναγιάς το 2010. Έτσι, στη μετέπειτα επεξεργασία των δεδομένων δεν επιχειρήθηκε συνάθροιση των δεδομένων ανά συνεργείο.

Η συνάθροιση (aggregation) των δεδομένων ανά Δήμο και Ημερομηνία ή Πενθήμερο γίνονταν με τη χρήση της εντολής aggregate, σε επιλεγμένα πεδία και υπολογίζονταν συνήθως ο μέσος όρος τους. Π.χ., στην παρακάτω εντολή γίνεται ταυτόχρονα συνάθροιση, στο μέσο όρο, των πεδίων MO\_FEMALE, MO\_DACUS, FEMALE\_POSOSTO και MO\_MIN ανά Δήμο και Πενθήμερο. Συμπεριλαμβάνεται πάντα και το Έτος γιατί πρόκειται για πενθήμερο συγκεκριμένου έτους.

Έτσι η οποιαδήποτε επεξεργασία έγινε και αφορούσε στους πίνακες ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ και ΔΕΙΓΜΑΤΟΛΗΨΙΑ, γίνονταν με τη χρήση των data frames **aggr\_metr** και **aggr\_deig** αντίστοιχα, τα οποία προκύπταν από τις συναθροίσεις των data frames metr\_dakon και deigmato αντίστοιχα.

## **5.2. Κατασκευή διαγραμμάτων (plots) που απεικονίζουν τις μεταβολές του δακοπληθυσμού στο χρόνο**

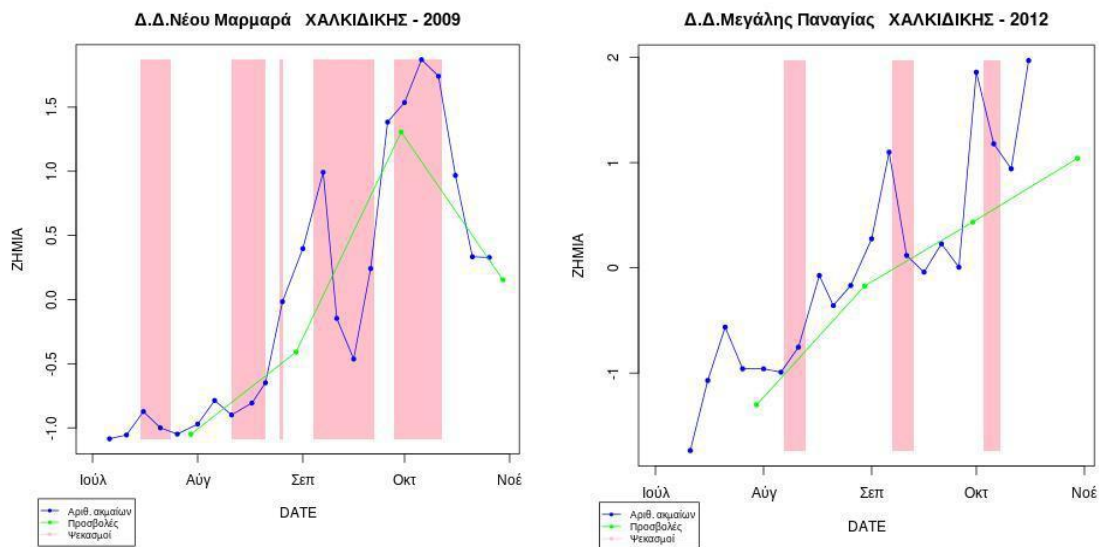
Προκειμένου να απεικονιστούν οι αυξομειώσεις του δακοπληθυσμού και η αλληλεπίδρασή τους με άλλους παράγοντες, κατασκευάστηκαν σειρές από διαγράμματα. Κατασκευάζονταν ένα διάγραμμα για κάθε Δήμο ανά Έτος, με σκοπό να μπορεί να γίνει αντιπαραβολή πολλών διαγραμμάτων το ένα μετά το άλλο και να εντοπιστούν οπτικά τυχόν ομοιότητες.



Η κατασκευή γίνονταν με τη δημιουργία μιας συνάρτησης με παραμέτρους το πρωτεύον κλειδί του Δήμου (DHMOS\_ID) και τη χρονολογία. Π.χ.:

```
dak_pros_psek<-function(dimos, year){
```

### 5.2.1. Πληθυσμός δάκου σε συνδυασμό με τη δακοπροσβολή και τις ημερομηνίες των ψεκασμών



**Εικόνα 16:** Διαγράμματα που απεικονίζουν τις μεταβολές του δακοπληθυσμού και της δακοπροσβολής σε συνδυασμό με τις ημερομηνίες ψεκασμών.

Κατασκευάστηκαν διαγράμματα σαν αυτά που φαίνονται στην εικόνα 16. Όπως δείχνει και το υπόμνημα, η μπλε γραμμή απεικονίζει τον πληθυσμό του δάκου (θηλυκά + αρσενικά) και η πράσινη γραμμή το σύνολο των προσβολών. Τα ροζ παραλληλόγραμμα απεικονίζουν τα χρονικά διαστήματα κατά τα οποία έλαβαν χώρα ψεκασμοί.

Τα δεδομένα συναθροίστηκαν ανά Δήμο και Ημερομηνία

Προκειμένου να απεικονιστούν στο ίδιο διάγραμμα, ο πληθυσμός του δάκου και το ύψος της προσβολής **κανονικοποιήθηκαν** με βάση τον τύπο:

$$x_{norm} = \frac{x_i - \bar{x}}{s}$$

όπου s είναι η τυπική απόκλιση.

Στα data frames με τις επιλεγμένες εγγραφές (για Δήμο και έτος που εξετάζεται στο διάγραμμα), οι τιμές των στηλών με τον πληθυσμό του δάκου και με το ύψος των προσβολών, αντικαθίσταντο με τις κανονικοποιημένες τιμές. Π.χ.:

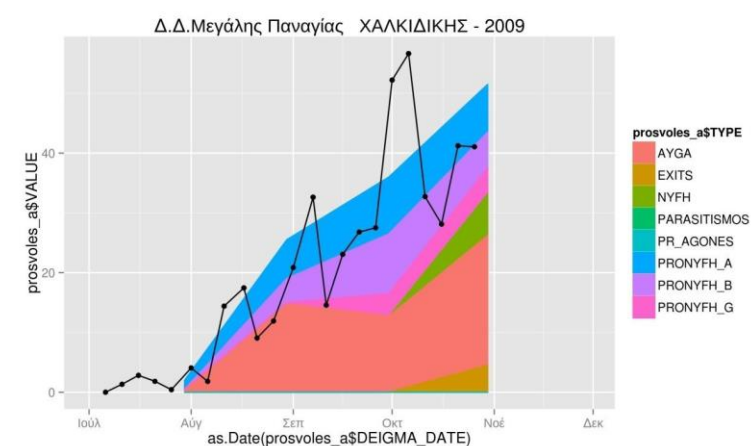
```
aggr_met_a$DAKOI <- norm_mean(aggr_met_a$DAKOI)
```

### 5.2.1.1. Παρατηρήσεις επί των διαγραμμάτων

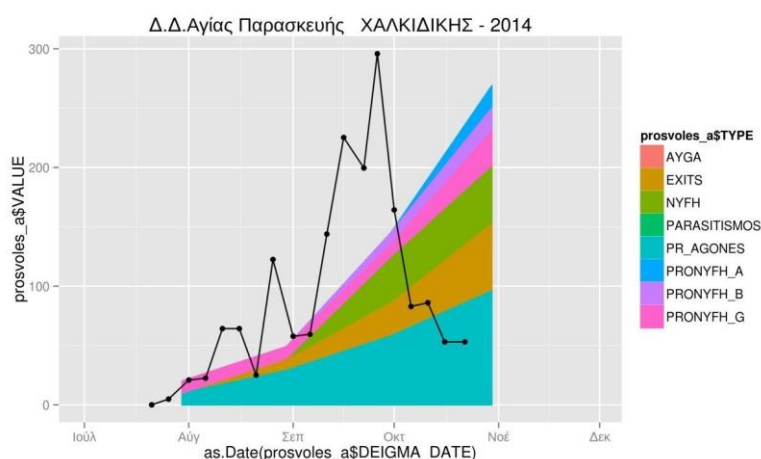
Στα περισσότερα διαγράμματα παρατηρήθηκε ένας, σε γενικές γραμμές, παραλληλισμός της αυξομείωσης του πληθυσμού των δάκων και της προσβολής.

Επίσης παρατηρήθηκε ότι η εφαρμογή ενός ψεκασμού δεν συνεπάγεται πάντα και μία πτώση του δακοπληθυσμού. Αυτό είναι ως ένα βαθμό αναμενόμενο, δεδομένου ότι ποικίλοι άλλοι παράγοντες μπορούν να επηρεάσουν την αυξομειωτική πορεία του πληθυσμού, όπως οι καιρικές συνθήκες ή ο αριθμός των προνυμφών, νυμφών και αυγών μέσα στους καρπούς. Όπως αναφέρθηκε (βλ. παρ. 1.1.5), οι ψεκασμοί είναι συνήθως δολωματικοί και αντιμετωπίζουν τα ακμαία και όχι τα αυγά ή τις προνύμφες που βρίσκονται μέσα στον καρπό.

### 5.2.2. Πληθυσμός δάκου σε συνδυασμό με τους τύπους της δακοπροσβολής αναλυτικά



**Εικόνα 17:** Διαγράμματα που απεικονίζουν τις μεταβολές του δακοπληθυσμού και όλων των τύπων προσβολής



Κατασκευάστηκαν διαγράμματα σαν αυτά που φαίνονται στην εικόνα 17. Η μαύρη γραμμή απεικονίζει τον πληθυσμό του δάκου (θηλυκά + αρσενικά). Με τα διάφορα χρώματα απεικονίζεται το μέγεθος των διαφόρων τύπων προσβολής, όπως φαίνεται στο υπόμνημα. Η σειρά των διαφόρων τύπων προσβολής στο διάγραμμα είναι διαφορετική από αυτή στο υπόμνημα. Η τοποθέτηση είναι τέτοια ώστε κάτω κάτω να είναι οι άγονες προσβολές και ο παρασιτισμός και οι υπόλοιποι τύποι από πάνω είναι αυτοί που αποτελούν τη γόνιμη προσβολή. Έτσι μπορεί κάποιος να έχει εικόνα και για το μέγεθος της γόνιμης προσβολής συνολικά.

	DHMOS	DEIGMA_DATE	VALUE	TYPE	YEAR
376	14260	2014-10-30	0.000000	PARASITISMOS	2014
377	14265	2014-10-30	0.000000	PARASITISMOS	2014
378	14287	2014-10-30	0.000000	PARASITISMOS	2014
379	14298	2014-10-30	0.000000	PARASITISMOS	2014
380	14309	2014-10-30	0.000000	PARASITISMOS	2014
381	14326	2014-10-30	0.000000	PARASITISMOS	2014
382	14336	2014-10-30	0.000000	PARASITISMOS	2014
383	16848	2014-10-30	97.000000	PARASITISMOS	2014
384	16852	2014-10-30	24.000000	PARASITISMOS	2014
385	16867	2014-10-30	43.000000	PARASITISMOS	2014
386	14112	2009-07-30	0.000000	PR_AGONES	2009
387	14231	2009-07-30	106.000000	PR_AGONES	2009
388	14254	2009-07-30	0.000000	PR_AGONES	2009
389	14256	2009-07-30	0.000000	PR_AGONES	2009
390	14260	2009-07-30	0.000000	PR_AGONES	2009
391	14265	2009-07-30	60.000000	PR_AGONES	2009
392	14268	2009-07-30	24.000000	PR_AGONES	2009
393	14287	2009-07-30	43.500000	PR_AGONES	2009
394	14298	2009-07-30	0.000000	PR_AGONES	2009
395	14309	2009-07-30	0.000000	PR_AGONES	2009
396	14315	2009-07-30	0.000000	PR_AGONES	2009
397	14326	2009-07-30	0.000000	PR_AGONES	2009
398	14331	2009-07-30	0.000000	PR_AGONES	2009

Showing 376 to 398 of 3,080 entries

**Εικόνα 18:** Το data frame “prosvoles”, από το οποίο εξήχθησαν τα area plots. Πίνακας από το περιβάλλον του RStudio.

Όπως φαίνεται στην εικόνα, τα ονόματα των τύπων προσβολής, δεν είναι πεδία αλλά τιμές του πεδίου TYPE. Οι τιμές του πεδίου VALUE, απεικονίζουν το μέγεθος της προσβολής για κάθε εγγραφή. Για την κατασκευή του συγκεκριμένου πίνακα κατασκευάστηκαν καταρχήν ξεχωριστά dataframes με συναθροισμένα δεδομένα (ανά Δήμο και Ημερομηνία) για καθένα τύπο προσβολής. Η στήλη με το μέγεθος της προσβολής μετονομαζόταν σε VALUE. Στη συνέχεια προσθέτονταν μία ακόμα στήλη με το όνομα TYPE και τιμές, σε όλες τις εγγραφές, το όνομα του τύπου προσβολής. Στο τέλος όλα τα dataframes που δημιουργήθηκαν ενώθηκαν και δημιουργήθηκε ένα ενιαίο.

Αφού δημιουργούνται το area plot, και για να «χωρέσει» μέσα σε αυτό η γραμμή που απεικονίζει τον πληθυσμό του δάκου, γινόταν μια προσαρμογή των τιμών του δακοπληθυσμού ως εξής:

- Κανονικοποίηση των τιμών με βάση τον τύπο:

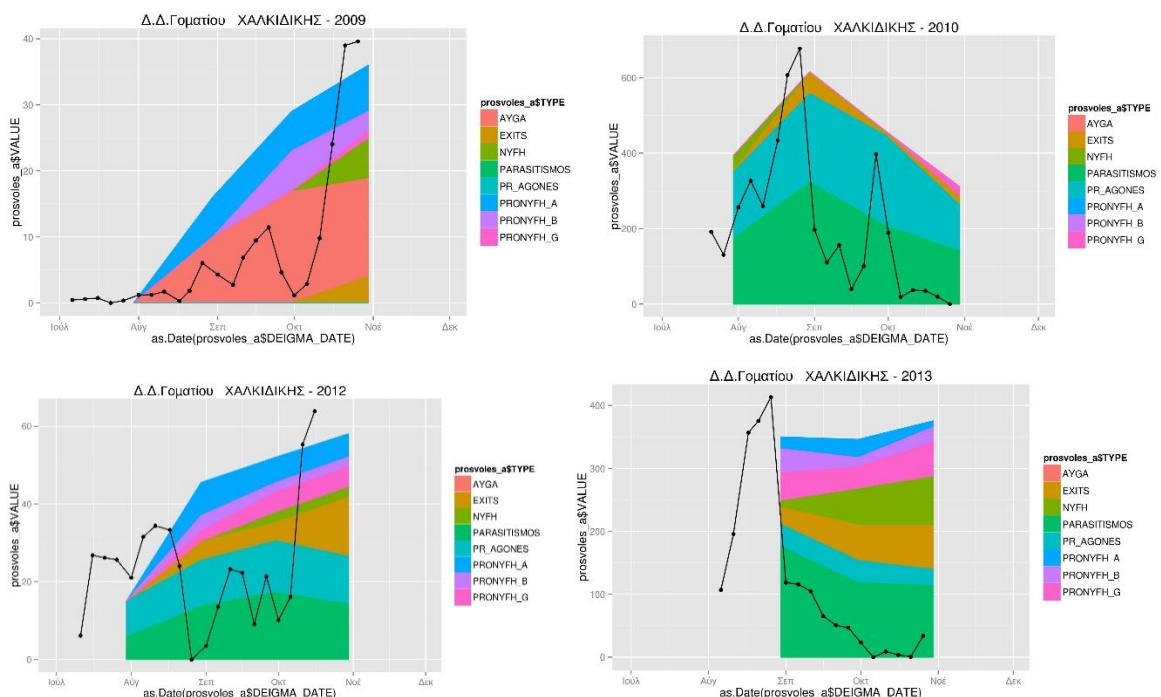
$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- Οι κανονικοποιημένες τιμές πολλαπλασιάζονται επί το εύρος του άξονα y, το οποίο είχε προηγουμένως υπολογιστεί:

```
aggr_met_a$DAKOI <- (y_range)*normm(aggr_met_a$DAKOI)
```

### 5.2.2.1. Παρατηρήσεις επί των διαγραμμάτων

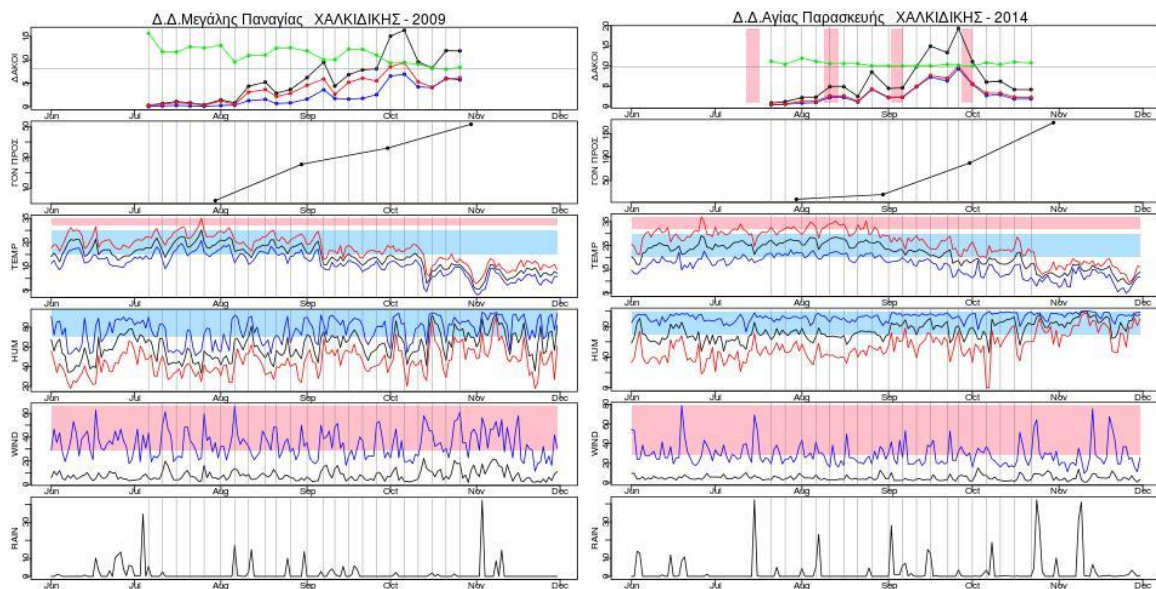
Ένα γενικό συμπέρασμα που μπορεί να προκύψει είναι ότι η πορεία αυξομείωσης του μεγέθους των γόνιμων προσβολών δεν ακολουθεί πάντα αυτή του συνόλου των προσβολών. Επίσης, ποικιλία παρατηρείται και στην αναλογία γόνιμων / (άγονων προσβολών + παρασιτισμού), καθώς σε κάποιες περιπτώσεις είναι περίπου στο 1:1 (πχ. Γομάτι Χαλκιδικής 2013), σε άλλες οι άγονες προσβολές είναι μηδαμινές (πχ. Γομάτι Χαλκιδικής 2009), σε άλλες οι γόνιμες προσβολές είναι σημαντικά λιγότερες (πχ. Γομάτι Χαλκιδικής 2010), ενώ σε άλλες η αναλογία μεταβάλλεται στη διάρκεια του χρόνου (πχ. Γομάτι Χαλκιδικής 2012).



**Εικόνα 19:** Τα διαγράμματα με τους τύπους προσβολών για το Γομάτι Χαλκιδικής τα έτη 2009, 2010, 2012, 2013, όπου διακρίνονται διαφορές στην αναλογία γόνιμων προσβολών προς άγονες + παρασιτισμό.

Τέλος, όσο αφορά **στα ποσοστά** προνυμφών α' σταδίου, προνυμφών β' σταδίου, προνυμφών γ' σταδίου και νυμφών, **επί του συνόλου των γόνιμων προσβολών**, παρατηρείται μία, σε γενικές γραμμές, μείωση των προνυμφών α' σταδίου στην πορεία του χρόνου, μία αύξηση και μετά μείωση των προνυμφών β' σταδίου και μία αύξηση των προνυμφών γ' σταδίου και των νυμφών (βλ. για παράδειγμα την Εικόνα 19).

### 5.2.3. Πολλαπλά διαγράμματα με πληθυσμό δάκου (σύνολο, θηλυκά, αρσενικά, ποσοστό θηλυκών) μαζί με ψεκασμούς, γόνιμες προσβολές και μετεωρολογικά στοιχεία



**Εικόνα 20:** Διαγράμματα που απεικονίζουν τις μεταβολές του δακοπληθυσμού (σύνολο, θηλυκά, αρσενικά, ποσοστό θηλυκών) μαζί με ψεκασμούς, των γόνιμων προσβολών και μετεωρολογικά στοιχεία

Κατασκευάστηκαν διαγράμματα σαν αυτά που φαίνονται στην εικόνα 20. Όπως φαίνεται στην εικόνα, το κάθε διάγραμμα αποτελείται από έξι μικρά διαγράμματα:

- Στο πρώτο απεικονίζεται ο πληθυσμός του δάκου. Με μαύρο είναι το σύνολο (θηλυκά + αρσενικά), με κόκκινο τα θηλυκά, με μπλε τα αρσενικά. Με πράσινο το ποσοστό θηλυκών (πεδίο FEMALE\_POSOSTO). Έχει πολλαπλασιαστεί επί το εύρος του άξονα των y. Με ροζ παραλληλόγραμμα απεικονίζονται τα χρονικά διαστήματα κατά τα οποία έλαβαν χώρα ψεκασμοί.

- Στο δεύτερο απεικονίζονται οι γόνιμες προσβολές. Είναι σε ξεχωριστό διάγραμμα από το δακοπληθυσμό επειδή δεν έγινε κανονικοποίηση, έτσι ώστε να φαίνεται στο άξονα των y το μέγεθος του δακοπληθυσμού και της δακοπροσβολής
- Στο τρίτο απεικονίζεται η θερμοκρασία. Με μαύρο η μέση ημερήσια, με μπλε η ελάχιστη και με κόκκινο η μέγιστη. Το γαλάζιο παραλληλόγραμμο απεικονίζει το εύρος θερμοκρασιών μεταξύ 20 και 30 °C, όπου είναι η άριστη θερμοκρασία για την εναπόθεση αυγών και το ροζ τις θερμοκρασίες πάνω από 32 °C, όπου σταματά η ανάπτυξη του δάκου (βλ. παρ. 1.1.1)
- Στο τέταρτο απεικονίζεται η υγρασία. Με μαύρο η μέση ημερήσια, με μπλε η μέγιστη και με κόκκινο η ελάχιστη. Το γαλάζιο παραλληλόγραμμο απεικονίζει το εύρος μεταξύ 70% και 100%, όπου είναι η άριστη υγρασία για την εναπόθεση αυγών
- Στο πέμπτο απεικονίζεται η ταχύτητα ανέμου. Το ροζ παραλληλόγραμμο απεικονίζει τις τιμές που είναι μεγαλύτερες από 28.8 km/h, όπου οι ψεκασμοί καθίστανται αναποτελεσματικοί επειδή παρεμποδίζονται οι πτήσεις των εντόμων, οπότε αυτά δεν τρέφονται από το ψεκαστικό υγρό και επιζούν (Pontikakos et al. 2010)
- Στο έκτο απεικονίζεται το ημερήσιο ύψος βροχής

Σχεδιάστηκαν αχνές κάθετες γραμμές στις ημερομηνίες μέτρησης του δακοπληθυσμού, σε όλα τα μικρά διαγράμματα, για να είναι ευκολότερη η αντιπαραβολή μεταξύ τους. Επίσης, στο διάγραμμα που απεικονίζει τον πληθυσμό του δάκου, σχεδιάστηκε μια αχνή οριζόντια γραμμή στο μέσο του εύρους του άξονα y, η οποία παριστάνει το 0,5 ποσοστό θηλυκών (FEMALE\_POSOSTO)

Για τα μετεωρολογικά δεδομένα, για κάθε Δήμο και Έτος, επιλέγονταν οι εγγραφές από το μετεωρολογικό σταθμό που αντιστοιχεί στον εξεταζόμενο Δήμο:

```
meteo_a <- meteo[meteo$METEO_ID==aggr_met_a$METEO_ID[1] &meteo$YEAR == year ,]
```

Σε περίπτωση που δεν υπήρχαν εγγραφές για το συγκεκριμένο μετεωρολογικό σταθμό στο συγκεκριμένο έτος, επιλέγονταν οι εγγραφές του εναλλακτικού μετεωρολογικού σταθμού. Σαν δείκτης, χρησιμοποιούνταν οι εγγραφές της στήλης OUT\_HUM, με την παραδοχή ότι η μέση ημερήσια εξωτερική υγρασία δεν μπορεί να είναι 0.

### 5.2.3.1. Παρατηρήσεις επί των διαγραμμάτων

Παρατηρήθηκε, στις περισσότερες περιπτώσεις, μια ελαφρά υπεροχή του πληθυσμού των θηλυκών, γύρω στο 0,6.

Δεν διαπιστώθηκε οπτικά κάποια συσχέτιση μεταξύ των αυξομειώσεων του πληθυσμού και των μετεωρολογικών φαινομένων.

## 5.3. Εκτίμηση της συσχέτισης μεταξύ πληθυσμού δάκων και προσβολών στους καρπούς

Εκτιμήθηκε η συσχέτιση μεταξύ των στηλών του συνόλου δάκων (MO\_DACUS), των θηλυκών (MO\_FEMALE), του ποσοστού θηλυκών (FEMALE\_POSOSTO) και των ζευγών (MO\_MIN) με τη στήλη του συνόλου των προσβολών (PROSVOLLES).

Επειδή σε κάθε περιοχή και έτος, η μεταβολή του πληθυσμού των δάκων μπορεί να επιδρά διαφορετικά στη μεταβολή του μεγέθους της προσβολής λόγω διαφορετικής επίδρασης ποικίλων παραγόντων (πχ. Διαφορετικό μικροκλίμα. Έτσι μπορεί πχ στη Γαλάτιστα το 2011 να είχαμε χαμηλούς πληθυσμούς δάκων και υψηλές προσβολές, ενώ αλλού, στο ίδιο έτος, το ανάποδο), δημιουργήθηκε μία συνάρτηση η οποία για κάθε περιοχή (δήμο) και έτος:

- επιλέγει τα δεδομένα δημιουργώντας ένα νέο πίνακα. Π.χ.:

```
(aggr_month_a<-aggr_month[aggr_month$DHMOS==dimos&aggr_month$YEAR == year,]),
```

- τα κανονικοποιεί. Π.χ.:

```
(aggr_month_a$MO_DACUS <- norm_mean(aggr_month_a$MO_DACUS))
```

(για την κανονικοποίηση με τη συνάρτηση `norm_mean` βλ. παρ. 5.2.1.)

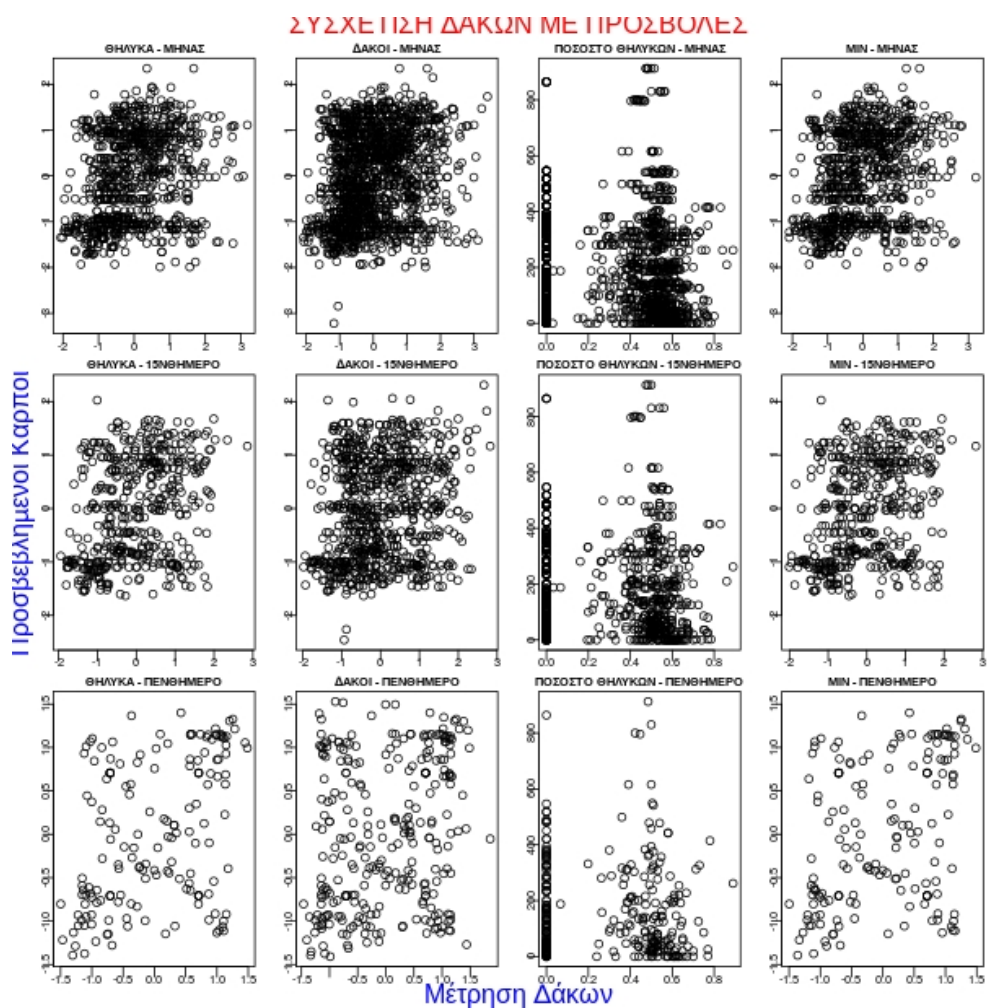
- και τα προσαρτά σε καινούργιο πίνακα που περιέχει κανονικοποιημένα δεδομένα. Π.χ.:

```
(month_dakoi <- append(month_dakoi , aggr_month_a$MO_DACUS))
```

Στη συνέχεια υπολογιζόταν ο **συντελεστής συσχέτισης** μεταξύ του πληθυσμού των δάκων (σύνολο, θηλυκά ή ζευγάρια) ή σχεδιαζόταν το σχετικό **διάγραμμα συνδιασποράς**. Π.χ.:

### 5.3.1. Εκτίμηση της συσχέτισης χωρίς τη χρήση των μετεωρολογικών στοιχείων

Αρχικά κατασκευάστηκε το συγκεντρωτικό διάγραμμα που φαίνεται στην Εικόνα 21.



**Εικόνα 21:** Συγκεντρωτικό διάγραμμα συσχέτισης του πληθυσμού δάκων με τις προσβολές το επόμενο πενήνημερο

- στην 1η γραμμή παρουσιάζονται οι εγγραφές μέτρησης του πληθυσμού που βρίσκονται στον ίδιο μήνα με την ημερομηνία δειγματοληψίας και μέτρησης της προσβολής



- στη δεύτερη γραμμή αντίστοιχα οι εγγραφές μέτρησης που βρίσκονται στο ίδιο 15νθήμερο με τη δειγματοληψία
- στην τρίτη γραμμή αντίστοιχα οι μετρήσεις πληθυσμού που έγιναν **ένα πενθήμερο πιο πριν** από τη δειγματοληψία
- στην 1η στήλη παρατίθενται οι μετρήσεις του μέσου όρου των θηλυκών ανά παγίδα
- στη 2η στήλη οι μετρήσεις του μέσου όρου δάκων (θηλυκά + αρσενικά) ανά παγίδα
- στην 3η στήλη οι μετρήσεις του ποσοστού θηλυκών προς συνολικό πληθυσμό.
- στην 4η στήλη έχει υπολογιστεί το  $\min(\text{αρσενικά}, \text{θηλυκά})$ , όπου απεικονίζει τον αριθμό των ζευγαριών που μπορούν να σχηματιστούν

Η συγχώνευση του πίνακα των δειγματοληψιών με αυτόν των μετρήσεων του δακοπληθυσμού **ένα πενθήμερο νωρίτερα**, γινόταν με τη χρήση της παρακάτω εντολής, αμέσως μετά τη δημιουργία του dataframe *aggr\_deig* (βλ. παρ. 5.1.1.4):

```
aggr_deig$PENTH <- aggr_deig$PENTH- 1
```

Υπολογίστηκε επίσης ο **συντελεστής συσχέτισης**, για τα διαγράμματα της τελευταίας γραμμής (εκτός από αυτό με το ποσοστό):

Για τα **θηλυκά** είναι: 0.3837354

Για το **σύνολο των δάκων** είναι: 0.2252035

Για τα **ζευγάρια** είναι: 0.3717653

Επειδή οι συντελεστές είναι μικροί, ξανατρέξαμε τον αλγόριθμο μειώνοντας κάθε φορά τη στήλη PENTH (αναφέρεται στο πενθήμερο) κατά ένα, στον πίνακα με τις δειγματοληψίες. Π.χ.:

```
(aggr_deig$PENTH <- aggr_deig$PENTH - 2)
```

```
(aggr_deig$PENTH <- aggr_deig$PENTH - 3)
```

Έτσι συσχετίσαμε τον αριθμό των προσβολών με τον πληθυσμό των δάκων όπως ήταν πριν από 2 ή 3 ή 5 πενθήμερα.

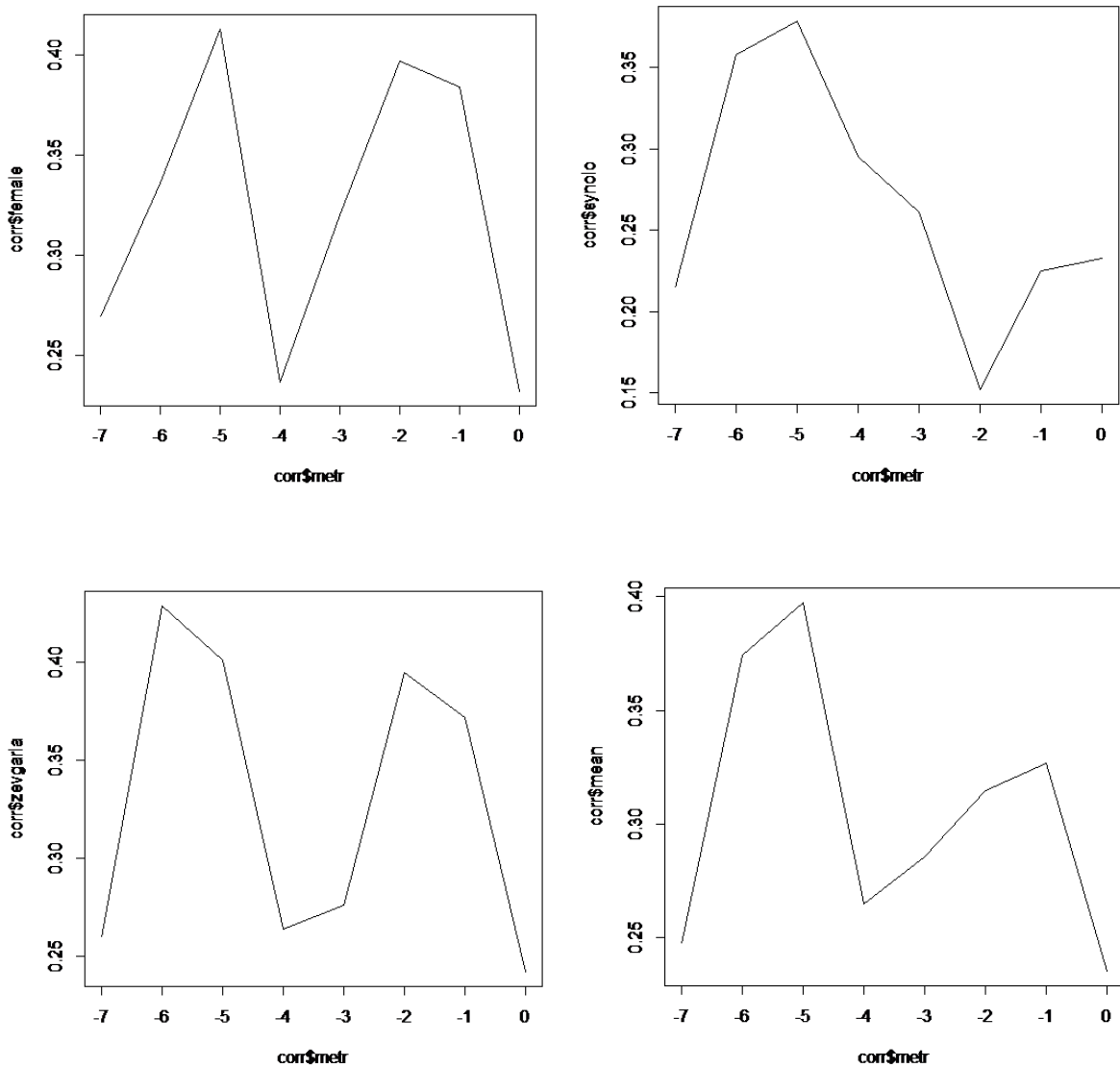
Τα αποτελέσματα φαίνονται στον Πίνακα 2.

metr	female	synolo	zevgaria	mean
0	0,2320808	0,2326720	0,2421453	0,2356327
-1	0,3837354	0,2252035	0,3717653	0,3269014
-2	0,3967355	0,1522658	0,3948190	0,3146068
-3	0,3200265	0,2611465	0,2757734	0,2856488
-4	0,2365216	0,2950666	0,2639493	0,2651792
-5	0,4126511	0,3784516	0,4008926	0,3973318
-6	0,3358884	0,3579700	0,4285669	0,3741418
-7	0,2687729	0,2146220	0,2594848	0,2476266
mean	0,3233015	0,2646748	0,3296746	

**Πίνακας 2:** Συντελεστές συσχέτισης της δακοπροσβολής με τον δακοπληθυσμό (θηλυκά, συνολικός, ζευγάρια) σε ποικίλα πενήθημερα μέτρησης

- Στην 1<sup>η</sup> στήλη είναι το πόσα πενήθημερα πριν τη δειγματοληψία έγινε η μέτρηση δακοπληθυσμού με την οποία συγκρίνονται
- Στις επόμενες στήλες είναι οι συντελεστές συσχέτισης των προσβολών με τον πληθυσμό των θηλυκών, του συνόλου των δάκων και των ζευγαριών αντίστοιχα
- Στην τελευταία στήλη και στην τελευταία γραμμή είναι οι μέσοι όροι των συντελεστών συσχέτισης των στηλών και των γραμμών αντίστοιχα.

Βάσει του παραπάνω πίνακα κατασκευάστηκαν και τα διαγράμματα που φαίνονται στην Εικόνα 22.



**Εικόνα 22:** Διαγράμματα που απεικονίζουν τη μεταβολή του συντελεστή συσχέτισης πληθυσμού / προσβολών σε συνάρτηση με το πόσα πενήμερα νωρίτερα από τη δειγματοληψία ελαιοκάρπου έγινε η μέτρηση του δακοπληθυσμού

### 5.3.1.1. Παρατηρήσεις επί των αποτελεσμάτων

Από το πολλαπλό διάγραμμα συνδιασφοράς δακοπληθυσμού / προσβολών παρατηρείται ότι, όσο αφορά το ποσοστό θηλυκών, οι μεγαλύτερες τιμές προσβολών βρίσκονται κοντά στην τιμή 0.5 (αναλογία 1:1), αφού προφανώς αξιοποιείται το σύνολο των ατόμων για τη δημιουργία νέας γενιάς.

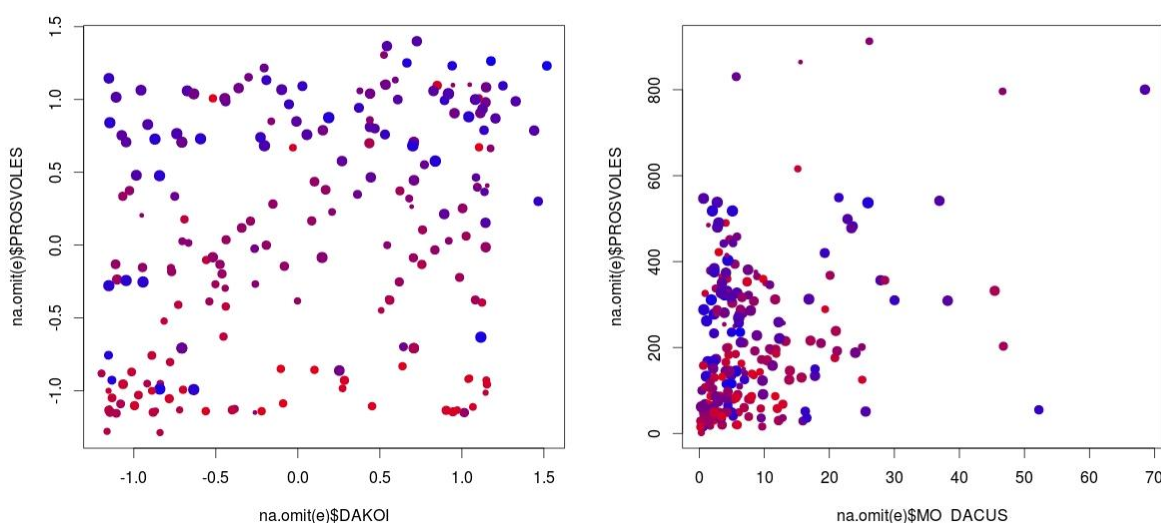
Υπάρχει μία θετική συσχέτιση μεταξύ πληθυσμού δάκων και προσβολών στους καρπούς, όχι όμως πολύ μεγάλη. Είναι γύρω στο 20% - 40%.

Η συσχέτιση αυτή είναι σε γενικές γραμμές μεγαλύτερη για τα θηλυκά και για τα ζευγάρια και μικρότερη για το σύνολο των δάκων.

Μεγαλύτερες συσχετίσεις μεταξύ των προσβολών και του πληθυσμού των δάκων παρατηρούνται **5 ή 6 πενήμερα πριν τη δειγματοληψία, δηλαδή περίπου 1 μήνα πριν.**

### 5.3.2. Εκτίμηση της συσχέτισης συνεκτιμώντας και τα μετεωρολογικά στοιχεία

Κατασκευάστηκε μια σειρά από διαγράμματα συνδιασποράς παρόμοια με αυτά της 3<sup>ης</sup> σειράς του συγκεντρωτικού διαγράμματος της προηγούμενης παραγράφου, αλλά για μετρήσεις δακοπληθυσμού που έγιναν **5 πενήμερα πριν τη δειγματοληψία ελαιοκάρπου** λόγω των μεγαλύτερων συσχετίσεων που είχαν παρατηρηθεί (βλ. παρ. 5.3.1.1). Άλλη μία σειρά διαγραμμάτων κατασκευάστηκε με δεδομένα που δεν ήταν κανονικοποιημένα.



**Εικόνα 23:** Διαγράμματα συνδιασποράς δακοπληθυσμού/προσβολής ελαιοκάρπου όπου λαμβάνονται υπόψη και η θερμοκρασία και η υγρασία. Στο δεξί διάγραμμα ο δακοπληθυσμός και οι προσβολές δεν είναι κανονικοποιημένες

Η απεικόνιση της θερμοκρασίας και της υγρασίας στα διαγράμματα γίνεται ως εξής:

- Θερμοκρασία: Μπλε=10 °C έως Κόκκινο=30 °C
- Υγρασία: Ανάλογη με το μέγεθος του εκάστοτε σημείου

Τα μετεωρολογικά δεδομένα που εκτιμήθηκαν, αφορούσαν στο μέσο όρο όλων των ημερών μεταξύ μέτρησης δακοπληθυσμού και δειγματοληψίας ελαιοκάρπου. Για την απόδοση της θερμοκρασίας με το χρώμα των κουκκίδων δημιουργήθηκε η στήλη TMPRGB η οποία κάνει αναγωγή αυτού του εύρους 10-30 °C στο εύρος 0-255, για να χρησιμοποιηθεί στο RGB. Πχ:

```
eniaios_dakoi$TMPRGB <- (eniaios_dakoi$TEMP-10)*255/20
```

Επίσης, για την απόδοση της υγρασίας με το μέγεθος των κουκκίδων δημιουργήθηκε η στήλη HUNOR η οποία κάνει αναγωγή της μέσης ημερήσιας **υγρασίας** στο διάστημα 0-1. Πχ:

```
norm <- function(x){  
  (x-min(na.omit(x)))/(max(na.omit(x))-min(na.omit(x)))  
}  
eniaios$HUNOR<-norm(eniaios$HUM)
```

Οι παραπάνω στήλες χρησιμοποιούνταν κατάλληλα, όπως φαίνεται στο παρακάτω παράδειγμα, ώστε να κατασκευάζεται το χρώμα και το μέγεθος των κουκκίδων του διαγράμματος ανάλογα με τις παραπάνω στήλες

```
plot(na.omit(eniaios)$MO_DACUS , na.omit(eniaios)$PROSVOLLES, pch=19,  
col=rgb(na.omit(eniaios)$TMPRGB, 0, 255-na.omit(eniaios)$TMPRGB,max=255),  
cex=2*eniaios$HUNOR)
```

#### 5.3.2.1. Παρατηρήσεις επί των αποτελεσμάτων

Από την παρατήρηση των διαγραμμάτων συνδιασποράς που κατασκευάστηκαν, γίνεται σχετικά εύκολα αντιληπτή η συσχέτιση μεταξύ του μεγέθους των προσβολών και της θερμοκρασίας αφού υπάρχει μια συγκέντρωση των κόκκινων κουκκίδων (υψηλές θερμοκρασίες) στις χαμηλές τιμές του άξονα των προσβολών και των μπλε κουκκίδων στις ψηλές τιμές του άξονα των προσβολών.

Αντίθετα, δεν είναι ευδιάκριτη κάποια συσχέτιση μεταξύ του μεγέθους των προσβολών και της υγρασίας.

Επίσης, δεν διαφαίνεται κάποια συσχέτιση μεταξύ πληθυσμού δάκων και θερμοκρασίας ή υγρασίας.

## 5.4. Δημιουργία Ενιαίου Πίνακα Μετρήσεων Δάκου – Ψεκασμών – Μετεωρολογικών Δεδομένων

### 5.4.1. Κατασκευή του πίνακα

Προκειμένου να κατασκευαστεί ένα μοντέλο κατηγοριοποίησης που να προβλέπει την άνοδο ή την πτώση του πληθυσμού με βάση τα μετεωρολογικά δεδομένα και τους ψεκασμούς, κατασκευάστηκε ένας ενιαίος πίνακας, σε κάθε εγγραφή του οποίου εμφανίζονται οι μετρήσεις του πληθυσμού των δάκων και οι αντίστοιχες του επόμενου πενθημέρου. Τα μετεωρολογικά δεδομένα εμφανίζονται ως μέσος όρος των ημερών μεταξύ των δύο πενθημέρων, αλλά δημιουργήθηκαν και πεδία με τις μετεωρολογικές μετρήσεις κατά τις ημερομηνίες κύριας και επόμενης μέτρησης δακοπληθυσμού. Ειδικές στήλες σημαίες δείχνουν αν το πενθήμερο της κάθε εγγραφής βρίσκεται σε περίοδο ψεκασμού. Οι στήλες αυτές είναι οι εξής:

**AFT\_STRT:** παίρνει τιμή '1' αν η ημ/νία κύριας μέτρησης είναι μετά την έναρξη ψεκασμού

**BFR\_END:** παίρνει τιμή '1' αν η ημ/νία επόμενης μέτρησης είναι πριν τη λήξη ψεκασμού

**PSEK\_BETW:** παίρνει τιμή '1' αν η ολόκληρη η διάρκεια ψεκασμού λαμβάνει χώρα στο μεσοδιάστημα μεταξύ των δύο μετρήσεων

#### 5.4.2. Εκτίμηση της επιρροής του ποσοστού θηλυκών, των μετεωρολογικών μεταβλητών και των ψεκασμών στη μεταβολή του πληθυσμού

Προκειμένου να εκτιμηθεί η επιρροή των διαφόρων παραγόντων που απεικονίζονται στον πίνακα στη μεταβολή του πληθυσμού, προστέθηκαν τα εξής πεδία:

**DD:** Μεταβολή πληθυσμού ανάμεσα στις δύο μετρήσεις ( $eniaios\$DD < -eniaios\$n\_MO\_DACUS - eniaios\$MO\_DACUS$ )

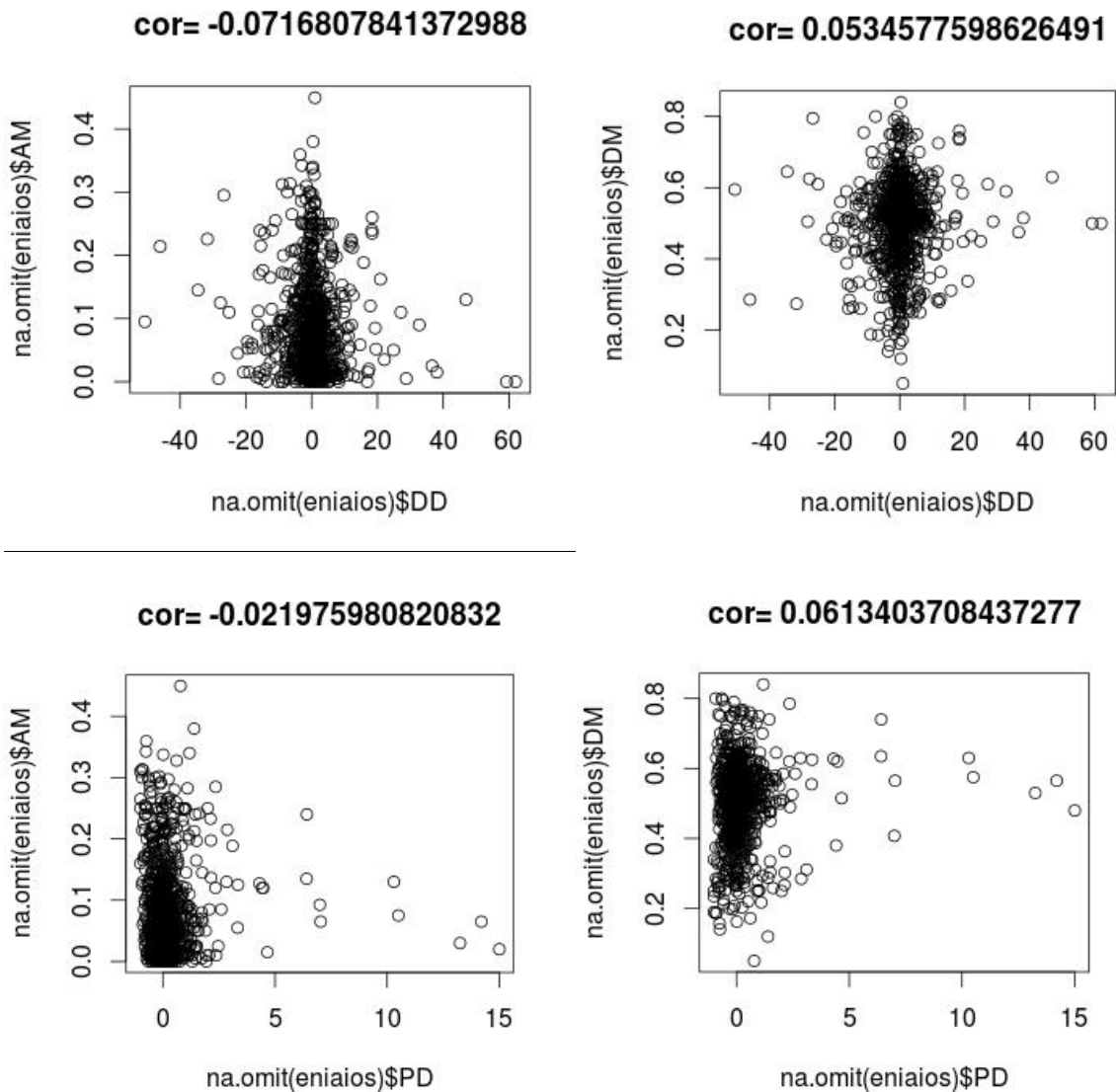
**DM:** Μέσος όρος ποσοστού θηλυκών στις δύο μετρήσεις ( $eniaios\$DD < -eniaios\$n\_MO\_DACUS - eniaios\$MO\_DACUS$ )

**AM:** Απόσταση μέσου όρου θηλυκών από το 0,5 ( $eniaios\$AM < -abs(eniaios\$DM - 0.5)$ )

**PD:** Ποσοστό μεταβολής του πληθυσμού ανάμεσα στις δύο μετρήσεις ( $eniaios\$PD < -eniaios\$DD / eniaios\$MO\_DACUS$ )

##### 5.4.2.1. Εκτίμηση της συσχέτισης μεταξύ πληθυσμού και ποσοστού θηλυκών

Κατά την παρατήρηση των διαγραμμάτων της παρ. 5.2.3., φάνηκε ότι πιθανόν η αναλογία θηλυκών / αρσενικών επηρεάζει την άνοδο ή την πτώση του πληθυσμού και συγκεκριμένα ότι όταν αυτή είναι κοντά στο 0,5 σημειώνεται σε κάποιες περιπτώσεις άνοδος του πληθυσμού, ενώ όταν απομακρύνεται από το 0,5 και κυρίως όταν υπερτερούν τα αρσενικά, σημειώνεται σε κάποιες περιπτώσεις μείωση του πληθυσμού. Για να μελετηθεί αν πράγματι ισχύει κάτι τέτοιο, δημιουργήθηκαν τα διαγράμματα συνδιασποράς που φαίνονται στην Εικόνα 24.



**Εικόνα 24:** Διαγράμματα συνδιασποράς μαζί με τους αντίστοιχους συντελεστές συσχέτισης. Οι συνδυασμοί είναι: μεταβολή δακοπληθυσμού (DD) με ποσοστό θηλυκών (DM) και απόσταση ποσοστού από το 0,5 (AM) και ποσοστό μεταβολής δακοπληθυσμού (PD) με ποσοστό θηλυκών (DM) και απόσταση ποσοστού από το 0,5 (AM).

Τα διαγράμματα κατασκευάστηκαν με εντολές όπως η παρακάτω:

```
plot(na.omit(eniaios)$PD, na.omit(eniaios)$DM, main=paste('cor=', cor(na.omit(eniaios)$PD,
na.omit(eniaios)$DM)))
```

Στα διαγράμματα δεν διαφαίνεται καμία συσχέτιση της αναλογίας θηλυκών / αρσενικών με τη μεταβολή του δακοπληθυσμού σε δύο συνεχόμενα πενήθημερα μέτρησης, με εξαίρεση το γεγονός ότι τα περιστατικά μεγάλης εκτόξευσης του πληθυσμού συνέβησαν με αναλογία θηλυκών / αρσενικά γύρω στο 0,5. Αυτό είναι



φανερότερο στα δύο κάτω διαγράμματα που απεικονίζουν το ποσοστό μεταβολής του δακοπληθυσμού.

#### 5.4.2.2. Εκτίμηση της επιρροής της διενέργειας ψεκασμών και των καιρικών φαινομένων στη μεταβολή του δακοπληθυσμού, με τον υπολογισμό του μέσου όρου

Δεδομένου ότι ο μέσος όρος της στήλης PD (ποσοστό μεταβολής του δακοπληθυσμού) υπολογίστηκε και βρέθηκε να είναι **0,24**, υπολογίστηκε ο μέσος όρος της στήλης υπό διάφορες συνθήκες. Για τα μετεωρολογικά δεδομένα χρησιμοποιήθηκαν τα πεδία του πίνακα με τους μέσους όρους των μετεωρολογικών στοιχείων (βλ. παρ. 5.4.1) Συγκεκριμένα:

- Για τις μετρήσεις που έγιναν σε περιόδους ψεκασμών: **0,15**
- Για τις μετρήσεις που ΔΕΝ έγιναν σε περιόδους ψεκασμών: **0,30**

Για τις περιόδους ψεκασμών προστέθηκε στον ενιαίο πίνακα η στήλη **PSEK**:

```
eniaios$PSEK <- eniaios$AFT_STRT | eniaios$BFR_END | eniaios$PSE_BETW
```

Ως μετρήσεις που έγιναν σε περιόδους ψεκασμών, εννοούνται αυτές των εγγραφών όπου PSEK = TRUE

- Για ημέρες με βροχή: **0,16**
- Για ημέρες χωρίς βροχή: **0,36**

Ημέρες με βροχή είναι αυτές όπου **RAIN>0**

- Για μέση ταχύτητα ανέμου>=8.5: **0,15**
- Για μέση ταχύτητα ανέμου<8.5: **0,25**

Η τιμή 8,5 ορίστηκε επειδή ο μέσος όρος της μέσης ταχύτητας ανέμου για μέγιστη ταχύτητα ανέμου>= 28,8 km/hείναι8,2km/h

```
mean(na.omit(meteo$AVG_WIND_SPEED)[na.omit(meteo$HIGH_WIND_SPEED)>=28.8])
```

Ο αριθμός των εγγραφών όπου **eniaios\$m\_WIND<8.5** είναι 1444, ενώ ο αριθμός των εγγραφών όπου **eniaios\$m\_WIND>=8.5** είναι 127, πολύ μικρότερος δηλαδή.

- Για μέση ημερήσια υγρασία>= 70%: **-0,03\***
- Για μέση ημερήσια υγρασία< 70%: **0,40\***
- Για ελάχιστη ημερήσια υγρασία>= 70%: **-0,29**
- Για μέγιστη ημερήσια υγρασία< 70%: **0,60**

Ο αριθμός των εγγραφών όπου  $eniaios\$MN\_HUM \geq 70$  είναι 37, ενώ ο αριθμός των εγγραφών όπου  $eniaios\$MX\_HUM < 70$  είναι 266.

- Για μέση ημερήσια θερμοκρασία μεταξύ 20 °C και 30 °C: **0,31**
- Για μέση ημερήσια θερμοκρασία εκτός ζώνης 20 °C - 30 °C: **0,12\***
- Για ελάχιστη ημερήσια θερμοκρασία  $\leq 20$  °C: **0,23**
- Για ελάχιστη ημερήσια θερμοκρασία  $> 20$  °C: **0,28**
- Για μέγιστη ημερήσια θερμοκρασία  $\geq 30$  °C: **0,41\***
- Για μέγιστη ημερήσια θερμοκρασία  $< 30$  °C : **0,19\***

Οι υπολογισμοί έγιναν με εντολές όπως η παρακάτω:

```
mean( na.omit(eniaios)$PD [na.omit(eniaios)$m_TEMP<=20 &na.omit(eniaios)$m_TEMP>=30] )
```

Παρατηρούμε ότι ο μέσος όρος μεταβολής του πληθυσμού σε περιόδους ψεκασμών είναι χαμηλότερος από ότι σε περιόδους που δεν υπάρχει ψεκασμός. Παρόλα αυτά παραμένει αυξητικός (θετικός).

Το ίδιο συμβαίνει και με τις ημέρες με βροχή σε σχέση με αυτές χωρίς βροχή.

Στις μέρες με δυνατό άνεμο, ο μέσος όρος μεταβολής του πληθυσμού είναι χαμηλότερος από ότι σε αυτές με μικρότερη ταχύτητα ανέμου. Όμως πρέπει να ληφθεί υπόψη ότι, αφενός στην πρώτη περίπτωση, έχουμε πολύ μικρότερο αριθμό εγγραφών και αφετέρου, στη δεύτερη περίπτωση, ο μέσος όρος μεταβολής του πληθυσμού δεν είναι αυξημένος σε σχέση με το γενικό μέσο όρο.

Για τη υγρασία και την θερμοκρασία εξετάστηκαν οι ζώνες 70% - 100% και 20°C - 30°C αντίστοιχα, ως άριστες συνθήκες για την εναπόθεση αυγών (βλ. παρ. 1.1.1).

Όσο αφορά στην υγρασία, όταν η μέση ημερήσια υγρασία είναι μέσα στο άριστο εύρος, διαφαίνονται, αντίθετα, τάσεις μείωσης του δακοπληθυσμού ενώ όταν είναι μικρότερη από αυτό το διάστημα διαφαίνονται τάσεις αύξησης.

Τέλος διαφαίνεται μία μείωση των αυξητικών τάσεων του πληθυσμού όταν η θερμοκρασία βρίσκεται εκτός του άριστου εύρους. Επίσης, παρατηρούμε ότι ο μέσος

όρος μεταβολής του πληθυσμού όταν η μέγιστη ημερήσια θερμοκρασία υπερβαίνει τους 30°C είναι υψηλότερος, ενώ όταν η μέγιστη ημερήσια θερμοκρασία είναι κάτω από τους 30°C είναι χαμηλότερος.

## 5.5. Πρόγνωση της μεταβολής του πληθυσμού με την τεχνική της Κατηγοριοποίησης

Η κατηγοριοποίηση έγινε με τη χρήση της εφαρμογής Rattle. Το Rattle είναι μία εφαρμογή της R και συνιστά ένα Γραφικό Περιβάλλον Διεπαφής Χρήστη (Graphic User Interface, GUI) για Εξόρυξη Δεδομένων.

Οι πίνακες οι οποίοι αποτέλεσαν πηγές δεδομένων, προέκυψαν από το πίνακα που περιγράφεται στο κεφάλαιο 5.4. Σε αυτόν προστέθηκε το πεδίο **TARGET**, το οποίο χωρίζει τα δεδομένα του πεδίου PD (ποσοστό μεταβολής δακοπληθυσμού) στις κλάσεις στις οποίες θέλουμε να ανατεθούν οι εγγραφές.

Έπειτα επιλέχθηκαν τα πεδία με τους μέσους όρους των μετεωρολογικών δεδομένων (m\_TEMP, H\_TEMP, L\_TEMP, m\_HUM, MX\_HUM, MN\_HUM, WIND, H\_WIND, RAIN), τα πεδία με τα στοιχεία για τους ψεκασμούς και το πεδίο DM (ποσοστό θηλυκών). Επίσης, σε κάποιες περιπτώσεις προστέθηκε το πεδίο **PSEK** το οποίο περιέχει το λογικό άθροισμα των πεδίων AFT\_STRT, BFR\_END και PSEK\_BETW (στοιχεία για τους ψεκασμούς) και δείχνει αν βρισκόμαστε γενικά σε περίοδο ψεκασμών.

Τέλος, αφαιρέθηκαν οι εγγραφές που περιείχαν κενά πεδία. Οι πίνακες που προέκυψαν περιείχαν 867 εγγραφές.

Σε όλες τις περιπτώσεις, τα δεδομένα χωρίστηκαν σε σύνολο εκπαίδευσης (**training set**) 70% και σύνολο ελέγχου (**testing set**) 30%.

### 5.5.1. Κατηγοριοποίηση σε 3 κλάσεις. Δέντρο Απόφασης

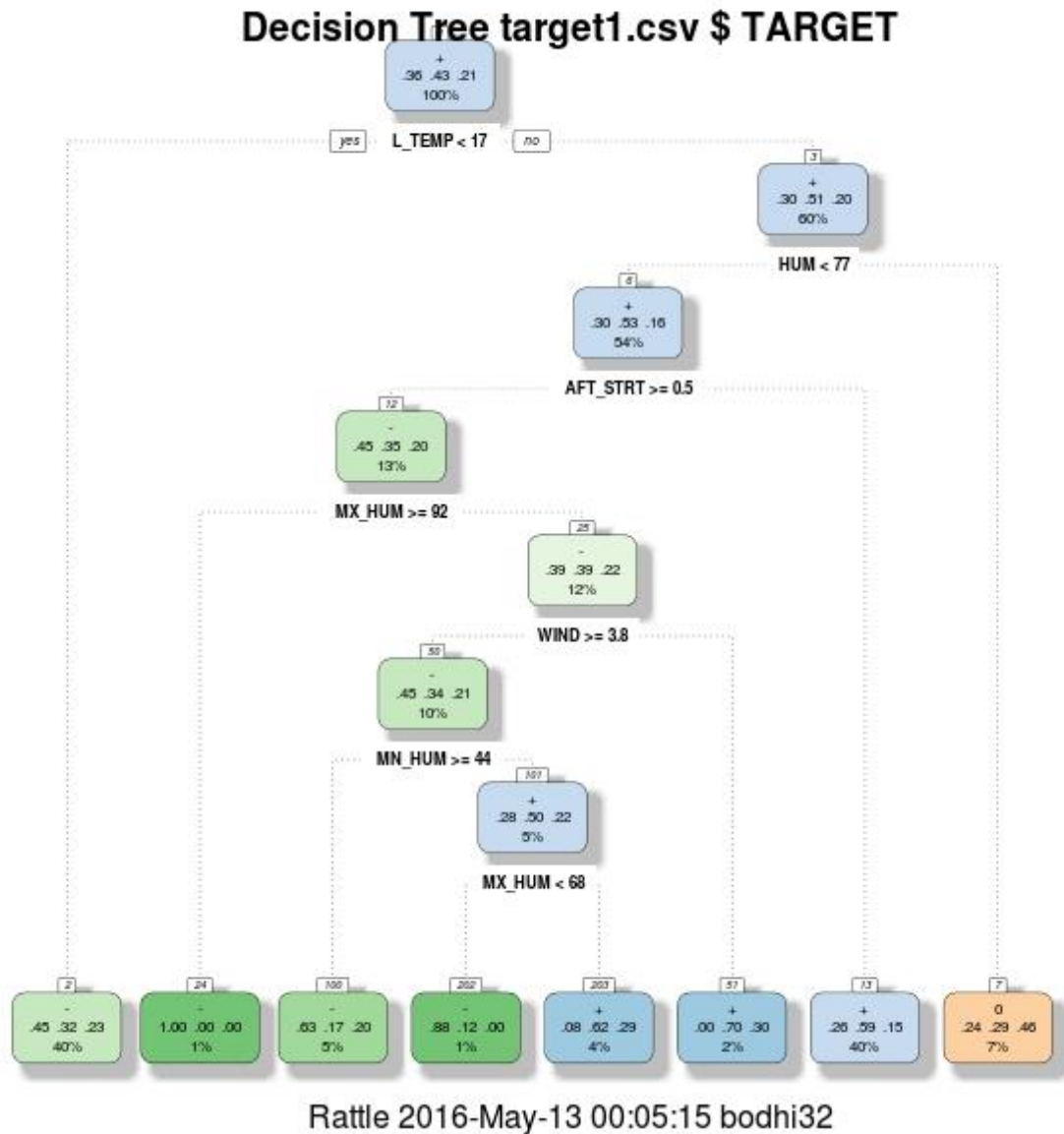
Σε αυτή την περίπτωση το πεδίο TARGET παίρνει τρεις τιμές: +, - και 0, ανάλογα με το αν έχουμε σημαντική αύξηση, σημαντική μείωση ή η μεταβολή του πληθυσμού θεωρείται αμελητέα.

Ως όριο για το αν η μεταβολή θεωρείται αμελητέα ορίστηκε αυθαίρετα η τιμή **0,1** οπότε στη στήλη TARGET δόθηκαν οι τιμές:

- **0** για  $-0.1 < PD < 0.1$
- **+** για  $PD \geq 0.1$
- **-** για  $PD \leq 0.1$

Εφαρμόστηκε το δέντρο απόφασης (decision tree) σε πίνακα (αρχείο "**target1.csv**") που περιείχε όλα τα πεδία των μετεωρολογικών δεδομένων, το ποσοστό θηλυκών και τα πεδία AFT\_STRT, BFR\_END και PSEK\_BETW (σύνολο 13 πεδία). Δοκιμάστηκε η κατασκευή δέντρων με διαφορετικές τιμές πολυπλοκότητας (**complexity**) και τελικά ορίστηκε η τιμή **0,01**

Επίσης εφαρμόστηκε σε πίνακα (αρχείο "**target3.csv**") που αντί για τα πεδία AFT\_STRT, BFR\_END και PSEK\_BETW είχε το πεδίο PSEK (βλ. αρχή της παρ. 5.5) (σύνολο 11 πεδία). Δοκιμάστηκε η κατασκευή δέντρων με διαφορετικές τιμές πολυπλοκότητας (**complexity**) και τελικά ορίστηκε η τιμή **0,004**. Μεγαλύτερη ακρίβεια είχε το δέντρο που κατασκευάστηκε από τον πρώτο πίνακα.



**Εικόνα 25:** Δέντρο Απόφασης για τον πίνακα target1.csv

Στο δέντρο απόφασης για τον πίνακα target1.csv δεν χρησιμοποιήθηκαν όλες οι μεταβλητές, παρά μόνο οι **AFT\_STRT\_HUM**, **L\_TEMP**, **MN\_HUM**, **MX\_HUM** και **WIND**

Ο ρυθμός σφάλματος (**overall error**) είναι **49%**, οπότε η ακρίβεια (**accuracy**) είναι  $100 - 49 = 51\%$

Σημειώνεται ότι, δέντρο απόφασης που κατασκευάστηκε από πίνακα στον οποίο εξαρχής αγνοήθηκαν (ignore) τα πεδία που δεν χρησιμοποιήθηκαν παραπάνω, έδωσε αποτελέσματα με μικρότερη ακρίβεια (49%).

Το δέντρο απόφασης που κατασκευάστηκε από τον πίνακα **target3.csv**, είχε ακρίβεια 48%.

Επίσης κατασκευάστηκε άλλος ένας πίνακας με όνομα **02\_target.csv**, παρόμοιος με τον target1.csv και τον target3.csv, όπου όμως ως όριο για το αν η μεταβολή θεωρείται αμελητέα ορίστηκε η τιμή **0,2** οπότε η στήλη TARGET πήρε τις τιμές:

- **0** για  $-0.2 < PD < 0.2$
- **+** για  $PD \geq 0.2$
- **-** για  $PD \leq 0.2$

Ο πίνακας περιείχε (συμπεριλαμβανομένων των υπολοίπων) μαζί τα πεδία AFT\_STRT, BFR\_END και PSEK\_BETW (όπως ο target1.csv και PSEK (όπως ο target3.csv) και η επιλογή του ποια πεδία θα χρησιμοποιηθούν έγινε από το Rattle, θέτοντας τα πεδία που δεν θέλουμε ανενεργά (ignore).

Τα αποτελέσματα όμως έδωσαν μεγαλύτερο ολικό σφάλμα. Πχ για πίνακα και συνθήκες παρόμοιες με τον target1.csv, το ολικό σφάλμα ήταν 57% και για πίνακα και συνθήκες παρόμοιες με τον target3.csv, το ολικό σφάλμα ήταν 53%.

### 5.5.2. Κατηγοριοποίηση σε 2 κλάσεις. Δέντρο Απόφασης

Σε αυτή την περίπτωση το πεδίο TARGET παίρνει τρεις τιμές: + 0 και -, ανάλογα με το αν έχουμε σημαντική αύξηση, σημαντική μείωση του πληθυσμού. Η στήλη TARGET πήρε τις τιμές:

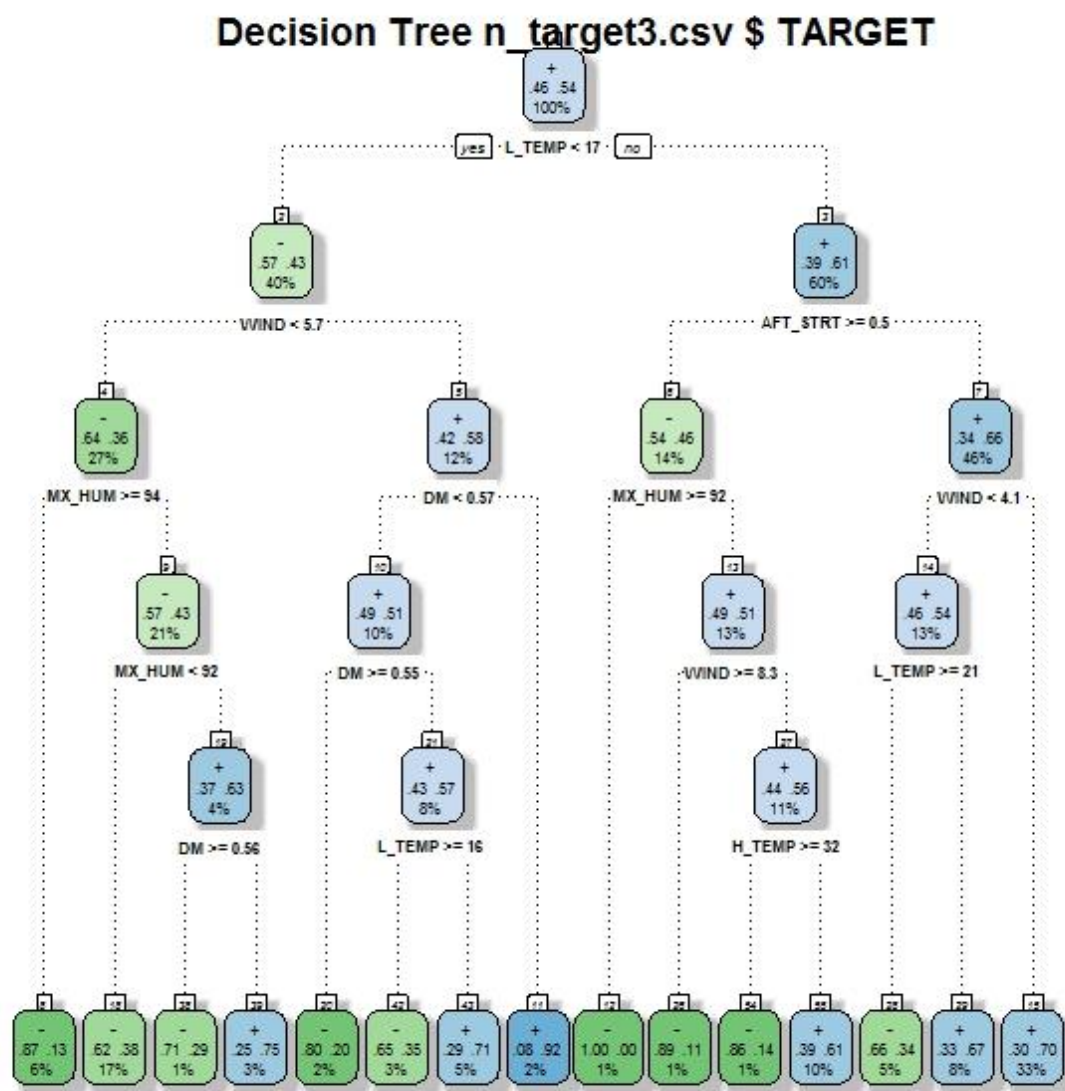
- **+** για  $PD > 0$
- **-** για  $PD \leq 0$

Τα δεδομένα χωρίστηκαν κι πάλι σε σύνολο εκπαίδευσης (**training set**) 70% και σύνολο ελέγχου (**testing set**) 30%.

Εφαρμόστηκε το δέντρο απόφασης (decision tree) σε πίνακα (αρχείο "**n\_target3.csv**") που περιείχε ίδια πεδία με τον πίνακα **02\_target.csv**.

Δοκιμάστηκε η κατασκευή δέντρων με διαφορετικές τιμές πολυπλοκότητας (**complexity**) και τελικά ορίστηκε complexity η τιμή **0,004**.

Δοκιμάστηκαν δύο εκδοχές, μία με τα πεδία AFT\_STRT, BFR\_END και PSEK\_BETW και μία με το πεδίο PSEK και **μεγαλύτερη ακρίβεια είχαμε με τη δεύτερη εκδοχή**. Στο δέντρο που κατασκευάστηκε δεν χρησιμοποιήθηκαν οι μεταβλητές TEMP και RAIN. Έτσι εφαρμόστηκε ξανά ο αλγόριθμος σε πίνακα που αγνοήθηκαν εξαρχής τα δύο αυτά πεδία και επιπλέον το πεδίο H\_TEMP.



Rattle 2016-Μαϊ-13 00:16:34 kalog-netbook

**Εικόνα 26:** Δέντρο Απόφασης για τον πίνακα n\_target3.csv με complexity 0.01

Για την εκδοχή με τη μεγαλύτερη ακρίβεια, ο ρυθμός σφάλματος (**overall error**) είναι **39%**, οπότε η ακρίβεια (**accuracy**) είναι  $100 - 39 = 61\%$

### **5.5.3. Σύγκριση της κατηγοριοποίησης σε δύο και σε τρεις κλάσεις**

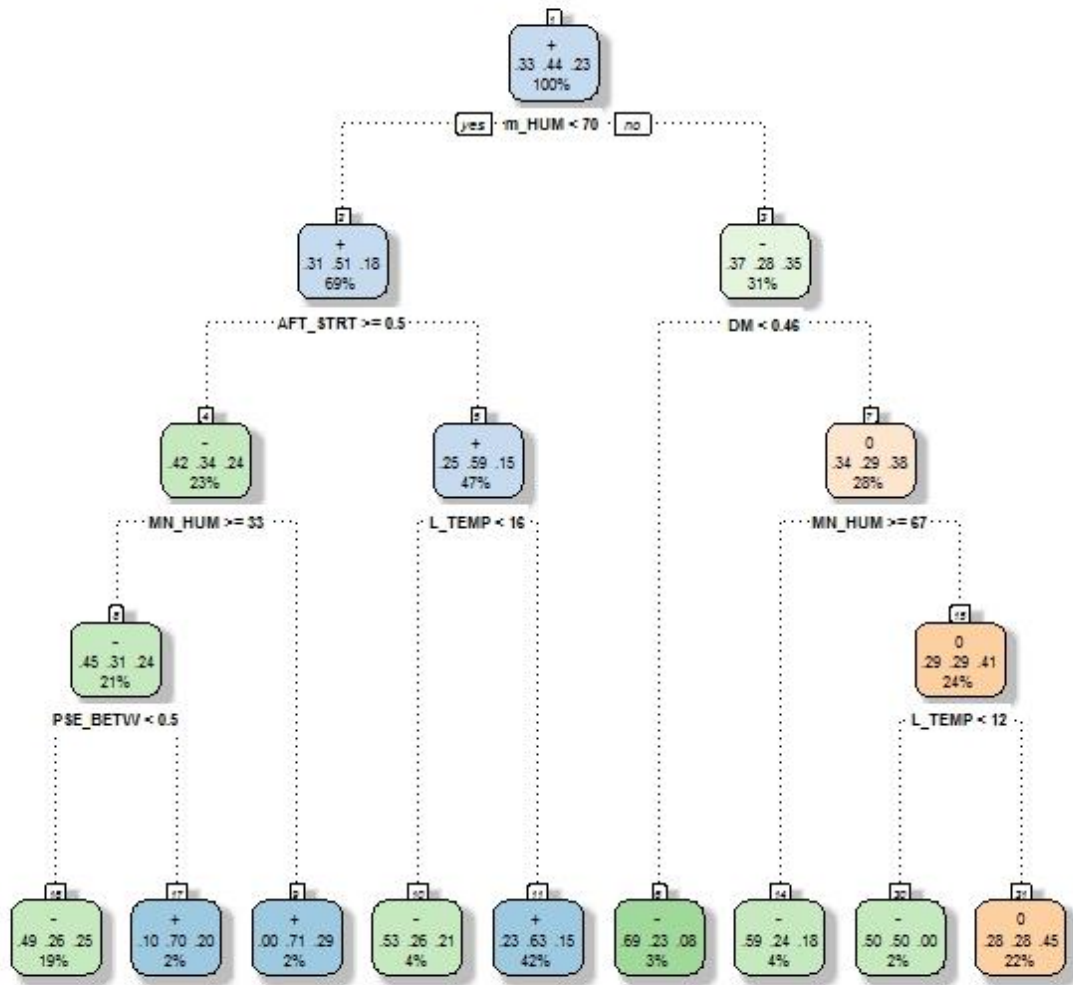
Ενώ η κατηγοριοποίηση σε δύο κλάσεις έδωσε φαινομενικά μεγαλύτερη ακρίβεια (61%) από αυτήν σε τρεις κλάσεις (51%), στην πρώτη περίπτωση, η πιθανότητα να προβλέψει το σύστημα τη σωστή κλάση στην τύχη είναι  $1/3$  δηλαδή 33%, επομένως έχουμε μία **βελτίωση της ευστοχίας κατά 18%**. Στη δεύτερη περίπτωση, η πιθανότητα αυτή είναι 50%, οπότε έχουμε βελτίωση της ευστοχίας κατά 11%. Επομένως μπορεί να θεωρηθεί ότι το μοντέλο με τις τρεις κλάσεις κάνει καλύτερη πρόβλεψη. Επιπλέον η κατηγοριοποίηση σε τρεις κλάσεις (αύξηση, σταθερό, μείωση) είναι καλύτερος τρόπος μοντελοποίησης του προβλήματος από ότι σε δύο (αύξηση, μείωση), οπότε συνεχίζουμε παρακάτω χρησιμοποιώντας αυτό το μοντέλο.

### **5.5.4. Κατηγοριοποίηση σε 3 κλάσεις, ξεχωριστά για κάθε νομό. Δέντρο Απόφασης**

Κατασκευάστηκαν οι πίνακες `t_chalkidiki.csv` και `t_samos.csv` οι οποίοι είναι παρόμοιοι με τον πίνακα `n_target3.csv` και περιέχουν μόνο τις εγγραφές των νομών Χαλκιδικής και Σάμου αντίστοιχα. Ο πρώτος περιέχει 616 εγγραφές και ο δεύτερος 278 εγγραφές. Χρησιμοποιήθηκαν τα ίδια πεδία και οι ίδιες παράμετροι με τον πίνακα `target1.csv`. Τα αποτελέσματα φαίνονται παρακάτω:



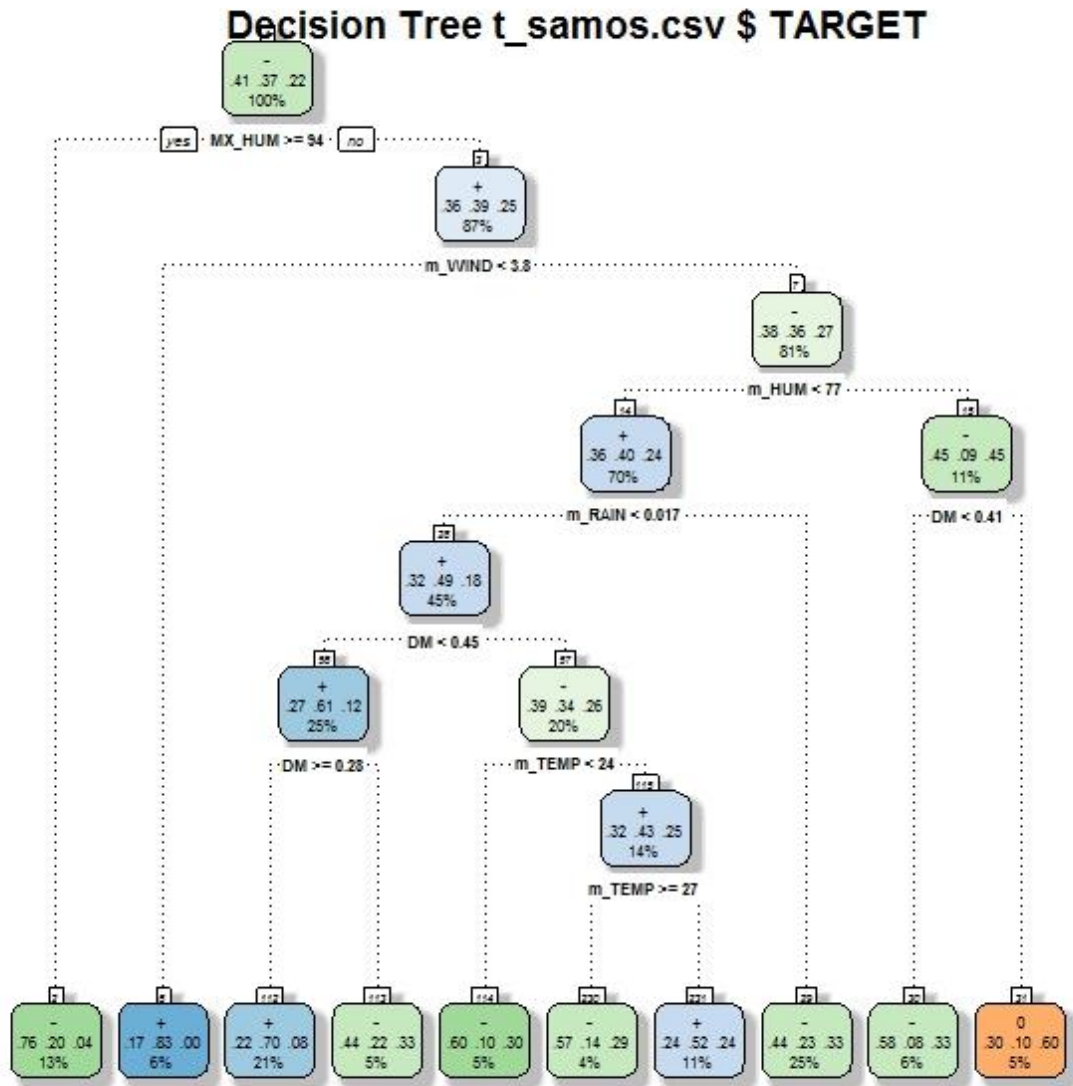
## Decision Tree t\_chalkidiki.csv \$ TARGET



Rattle 2016-Μαΐ-13 00:21:39 kalog-netbook

Εικόνα 27: Δέντρο Απόφασης για τον πίνακα t\_chalkidiki.csv

Ο ρυθμός σφάλματος (**overall error**) για τον πίνακα t\_chalkidiki.csv είναι **48%**, οπότε η ακρίβεια (**accuracy**) είναι  $100 - 48 = 52\%$



**Εικόνα 28:** Δέντρο Απόφασης για τον πίνακα t\_samos.csv

Ο ρυθμός σφάλματος (**overall error**) για τον πίνακα t\_samos.csv είναι **52%**, οπότε η ακρίβεια (**accuracy**) είναι  $100 - 52 = 48\%$

### 5.5.5. Κατηγοριοποίηση σε 3 κλάσεις. Support Vector Machine (Μηχανή Διανυσμάτων Υποστήριξης)

Τα δεδομένα του πίνακα *target1.csv* κατηγοριοποιήθηκαν επίσης με το μοντέλο Support Vector Machine.

Δοκιμάστηκε το τρέξιμο του αλγορίθμου με διαφορετικές kernel functions και τα καλύτερα αποτελέσματα τα έδωσε το Radial Basis (rbfdot).

Ο ρυθμός σφάλματος (**overall error**) είναι **50%**, οπότε η ακρίβεια (**accuracy**) είναι  $100 - 50 = 50\%$

### 5.5.6. Σύγκριση όλων των μοντέλων και των πινάκων

Μία σύνοψη για τον πίνακα *target1.csv* και τους *chalkidiki.csv* και *samos.csv* φαίνεται στον Πίνακα 3

ΠΙΝΑΚΑΣ	ΕΝΙΑΙΟΣ			ΧΑΛΚΙΔΙΚΗ	ΣΑΜΟΣ
ΑΡΙΘΜΟΣ ΚΛΑΣΕΩΝ	3 (όριο: 0,1)	3 (όριο: 0,2)	2	3 (όριο: 0,1)	3 (όριο: 0,1)
ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΗΣ	TREE	SVM	TREE	TREE	TREE
OVERALL ERROR (%) (Testing data)	49	50	53	39	48
ACCURACY (%) (Testing data)	51	50	47	61	52

**Πίνακας 3:** Σύγκριση αποτελεσματικότητας ανάμεσα στα δύο μοντέλα και ανάμεσα στον ενιαίο και στους εξιδεικευμένους ανά νομό πίνακες

#### Παρατηρήσεις:

- Μεταξύ των δύο ορίων για το αν η μεταβολή θεωρείται αμελητέα, καλύτερη ακρίβεια δίνει το 0,1
- Το μοντέλο με δύο κλάσεις φαίνεται ότι δίνει μεγαλύτερη ακρίβεια. Όμως, όπως αναλύεται στην παράγραφο 5.5.3, η μοντελοποίηση με τρεις κλάσεις είναι προτιμότερη.

- Οι δύο κατηγοριοποιητές που δοκιμάστηκαν δίνουν περίπου την ίδια ακρίβεια, **η οποία πάντως δεν μπορεί να θεωρηθεί μεγάλη ακρίβεια**
- Με εξειδίκευση σε συγκεκριμένους νομούς δεν πετυχαίνουμε καλύτερη ακρίβεια. Αντίθετα μάλιστα, στη Σάμο η ακρίβεια ήταν μικρότερη

# Κεφάλαιο 6

## Επίλογος

Η δημιουργία της Βάσης Δεδομένων DACUS του Υπουργείου Αγροτικής Ανάπτυξης και Τροφίμων της Ελλάδας έγινε για την ηλεκτρονική καταγραφή όλων των στοιχείων που αφορούν στις μετρήσεις δακοπληθυσμού, δακοπροσβολών και τα στοιχεία για τους πραγματοποιηθέντες ψεκασμούς, όπως αυτά προκύπτουν κατά την εφαρμογή των ετήσιων Προγραμμάτων Δακοκτονίας από τις κατά τόπους Διευθύνσεις Αγροτικής Οικονομίας και Κτηνιατρικής των Περιφερειακών Αυτοδιοικήσεων.

Τα στοιχεία αυτά μπορούν να καταστούν πολύτιμος αρωγός σε μια προσπάθεια ορθολογικότερης εφαρμογής της χημικής καταπολέμησης που και οικονομικό όφελος έχει να προσφέρει και στην προστασία του περιβάλλοντος θα συμβάλλει και μπορεί να βελτιώσει την ποιότητα του προϊόντος (λόγω της μείωσης των υπολειμμάτων φυτοφαρμάκων), ενός προϊόντος (επιτραπέζια ελιά, ελαιόλαδο) που βρίσκεται στην κορυφή των εξαγωγίμων προϊόντων της Ελλάδας. **Θα ήταν χρήσιμη η ανάπτυξη ενός συστήματος για την ετήσια αξιολόγηση του Προγράμματος Δακοκτονίας, το οποίο θα οδηγούσε στην ορθότερη εφαρμογή των χημικών επεμβάσεων. Θα ήταν επίσης χρήσιμη η ανάπτυξη ενός μοντέλου Ανακάλυψης Γνώσης το οποίο θα προέβλεπε την αύξηση του πληθυσμού του δάκου, ώστε να γίνει χημική επέμβαση νωρίτερα και οι ψεκασμοί να γίνουν λιγότεροι και περισσότερο αποτελεσματικοί.** Το μοντέλο αυτό θα μπορούσε να ενσωματωθεί στο σύστημα λήψης αποφάσεων που ήδη βρίσκεται σε εξέλιξη (Pontikakos et al, 2009, 2012), (Ποντικάκος κ.α. 2015), (Potamitis et al 2015). **Όπως φαίνεται στην παρούσα εργασία, η ανάπτυξη τέτοιων μοντέλων είναι εφικτή αλλά έχει ακόμα πολύ δρόμο να διανύσει.**

Ένα σημαντικό πρόβλημα είναι η μη καταγραφή και καταχώρηση δεδομένων σε σημαντικά πεδία της Βάσης Δεδομένων DACUS. Τέτοια είναι ο πληθυσμός **γόνιμων θηλυκών**, που δεν καταγράφονται πουθενά, των **αυγών** και των **δευτερογενών προσβολών**, που δεν καταγράφονται στις περισσότερες εγγραφές και των προνυμφών που κάποιες φορές δεν καταγράφεται ο αριθμός τους σε όλα τα στάδια. Όλα αυτά τα

στοιχεία είναι πολύτιμα για μία ακριβή πρόβλεψη του μελλοντικού πληθυσμού του Δάκου(Blum et al, 2013).

Επίσης, ενώ υπάρχουν τέτοια πεδία στον πίνακα ΜΕΤΡΗΣΗ\_ΔΑΚΩΝ, δεν καταγράφονται οι μετρήσεις της **θερμοκρασίας** και της **υγρασίας**. Η γνώση των μετεωρολογικών και άλλων συνθηκών που επικρατούν δίπλα στην παγίδα μπορεί να δώσει πιο ακριβείς προβλέψεις για τη μελλοντική εξέλιξη του δακοπληθυσμού (Blum et al, 2013). Σημειώνουμε ότι τα μετεωρολογικά στοιχεία που χρησιμοποιήθηκαν στην παρούσα εργασία ήταν από μετεωρολογικούς σταθμούς που συνήθως βρίσκονται σε μεγάλη απόσταση από τις περιοχές ενδιαφέροντος και ,πολλές φορές, σε περιοχές με διαφορετικό μικροκλίμα. Οι αυξομειώσεις του πληθυσμού των δάκων έχει διαπιστωθεί ότι επηρεάζονται από το υψόμετρο (Kounatidis et al, 2008) ή ακόμα και από τη θέση της παγίδαςMcPhail μέσα στον αγρό (Dimouetal, 2003).

Στην παρούσα εργασία, η επεξεργασία έγινε σε δεδομένα συναθροισμένα ανά Δήμο, για λόγους που αναφέρονται στην παράγραφο 5.1.1.4. Η επεξεργασία δεδομένων που θα αφορούσαν συγκεκριμένες παγίδες και όχι ολόκληρο Δήμο θα βελτίωνε σημαντικά των ακρίβεια των αποτελεσμάτων και θα μπορούσε ίσως να οδηγήσει στην κατασκευή ενός χρήσιμου μοντέλου. Για να γίνει αυτό δυνατό, θα πρέπει να υπάρχει συστηματική καταγραφή των τοποθεσιών που βρίσκονται οι παγίδες και καταγραφή των συντεταγμένων, ώστε να μπορούμε να έχουμε συγκρίσιμες μετρήσεις.

Επειδή οι μετρήσεις γίνονται από τους παγιδοθέτες, τα δεδομένα μπορεί να υπάρχουν μεγάλα **σφάλματα μέτρησης**. Φαίνεται όμως να είναι εφικτή η αυτοματοποίηση της μέτρησης του πληθυσμού των δάκων με τη χρήση οπτικών ή ακουστικών αισθητήρων (Pontikakos et al, 2009, 2012), (Ποντικάκος κ.α. 2015), (Potamitis et al 2015) και η μέτρηση των τοπικών κλιματικών συνθηκών με τη χρήση δεδομένων από δορυφόρους (Blumetal, 2013).

Περιθώρια βελτίωσης μπορούν να υπάρξουν και στην εφαρμογή των μοντέλων Κατηγοριοποίησης. Δεν έχουν συμπεριληφθεί στα δεδομένα εισόδου τα δεδομένα των δειγματοληψιών για την εκτίμηση της δακοπροσβολής. Αυτό βέβαια προϋποθέτει την εκτίμηση του χρόνου που μεσολαβεί μεταξύ του κάθε σταδίου ανάπτυξης του δάκου και της εμφάνισης ακμαίων. Επίσης θα πρέπει να εκτιμηθεί η αποτελεσματικότητα και άλλων κατηγοριοποιητών.

# Βιβλιογραφία

Ποντικάκος Κ., Περδίκης Δ., Τσιλιγκιρίδης Θ. (2015), Γραφική διεπαφή χρήστη ενός ολοκληρωμένου συστήματος υπηρεσιών θέσης για την παρακολούθηση και την αντιμετώπιση επιβλαβών δίπτερων εντομών σε δεντρώδεις καρποφόρες καλλιέργειες (1<sup>ο</sup> Συνέδριο Γεωγραφικών Πληροφοριακών Συστημάτων και Χωρικής Ανάλυσης στη Γεωργία και στο Περιβάλλον, Αθήνα, Μάιος 2015, Πρακτικά Περιλήψεων)

Ποντίκης, Κ., (1992) Ελαιοκομία. *Εκδόσεις Σταμούλη*.

Υπουργείο Γεωργίας (Ελλάδα), Γενική Δ/νση Φυτικής Παραγωγής, Δ/νση Προστασίας Φυτικής Παραγωγής (2002), Εγκύκλιος προς Νομαρχ. Αυτοδιοικήσεις - Δ/νσεις Γεωργίας Ελαιοκομικών Νομών (Αρ. Πρωτ. 94631/15-4-2002).

Υπουργείο Αγροτικής Ανάπτυξης και Τροφίμων (άγνωστη χρονολογία) *Πρόγραμμα Συλλογικής Καταπολέμησης του Δάκου*. Σειρά διαφανειών αναρτημένη στον διαδικτυακό ιστότοπο του ΥΠΑΑΤ (<http://www.minagric.gr/index.php/el/for-farmer-2/crop-production/fytoprostasiamenu/dakoktonia>)

Abdullah A., Brobst S., Pervaiz I., Umer M., Nisar A. (2003), Learning Dynamics of Pesticide Abuse through Data Mining (Copyright Australian Computer Society, Inc, The Australasian Workshop on Data Mining and Web Intelligence (AWDM&WI2004), Dunedin, New Zealand)

Abdullah A., Amir H. (2006), Data Mining a New Pilot Agriculture Extension Data Warehouse (Journal of Research and Practice in Information Technology, Vol. 38, No. 3, August 2006)

Behmann J., Mahlein A-K, Rumpf T, RömerC.,Plümer L. (2015) A review of advanced machine learning methods for the detection of biotic stress in precision cropprotection (Precision Agric (2015) 16:239–260)

Blum M., Lensky I. M., Rempoulakis P., Nestel D. (2013) Modeling insect population fluctuations with satellite land surface temperature (Agricultural and Forest Meteorology 176 (2013) 90– 93)

Dimou I., Koutsikopoulos C., Economopoulos A., Lykakis J (2003) The Distribution of Olive Fruit Fly Captures with McPhail Traps within an Olive Orchard (Phytoparasitica 31(2):124-131)

Hill M.G., Connolly P.G., Reutemann P., Fletcher D. (2014) The use of data mining to assist crop protection decisions on kiwifruit in New Zealand (M.G. Hill et al. / Computers and Electronics in Agriculture 108 (2014) 250–257)

Kounatidis I., Papadopoulos N.T., Mavragani-Tsipidou P., Cohen Y., Tertivanidis K., Nomikou M. & Nestel D. (2008) Effect of elevation on spatio-temporal patterns of olive fly (*Bactrocera oleae*) populations in northern Greece (J. Appl. Entomol. 132 (2008) 722–733)

Pontikakos C. M., Tsiligriridis T.A., Drougka M.E. (2009) Location-Aware System For Olive Fruit Fly Spray Control (Computers and Electronics in Agriculture 70 (2010) 355–368)

Pontikakos C. M., Tsiligriridis T.A., Yalouris C. P., Kontodimas D. C. (2012), Pest management control of olive fruit fly (*Bactrocera oleae*) based on a location-aware agro-environmental system (Computers and Electronics in Agriculture 87 (2012) 39–50)

Potamitis I., Eliopoulos P., Kontodimas D. C. (2015) The electronic McPhail Trap and a potential revision of the decision protocol (7th MEETING OF THE IOBC/WPRS WG INTEGRATED PROTECTION OF OLIVE CROPS)

Tan P-N, Steinbach M., Kumar V. (2006) Εισαγωγή Στην Εξόρυξη Δεδομένων (Μετάφραση: Βερούκιος Β.Σ., Σουραβλάς Σ., Εκδόσεις Τζιόλα 2015)

Yun Hwan Kim, Seong Joon Yoo, Yeong Hyeon Gu, Jin Hee Lim, Dongil Han, Sung Wook Baik (2014) Crop Pests Prediction Method using Regression and Machine Learning Technology: Survey (IERI Procedia 6 ( 2014 ) 52 – 56)