



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Πρωτότυπο έξυπνου θερμοστάτη
ενισχυτικής μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Χρήστος Π. Λαμπράκος

Επιβλέπων: Δημήτριος Σούντρης
Αναπληρωτής καθηγητής ΕΜΠ

Αθήνα, Οκτώβριος 2017



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Πρωτότυπο έξυπνου θερμοστάτη
ενισχυτικής μάθησης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Χρήστος Π. Λαμπράκος

Επιβλέπων: Δημήτριος Ι. Σούντρης
Αναπληρωτής καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 30η Οκτωβρίου 2017.

.....
Δημήτριος Ι. Σούντρης
Αναπληρωτής καθηγητής
ΕΜΠ

.....
Κιαμάλ Πεκμεστζί
Καθηγητής
ΕΜΠ

.....
Κωνσταντίνος Σιώζιος
Επίκουρος καθηγητής
ΑΠΘ

Αθήνα, Οκτώβριος 2017

.....

Χρήστος Π. Λαμπράκος

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών ΕΜΠ

Copyright © Χρήστος Π. Λαμπράκος, 2017

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περιεχόμενα

Σελίδα τίτλου.....	1
Περιεχόμενα.....	3
Λίστα σχημάτων.....	5
Λίστα πινάκων.....	7
Λίστα παραθέσεων αλγορίθμων.....	8
Περίληψη.....	9
Ευχαριστίες.....	11
I. Εισαγωγή.....	13
II. Συστήματα HVAC, έλεγχος και έξυπνοι θερμοστάτες.....	16
II.1. Βασικές έννοιες HVAC.....	16
II.2. Σύγχρονη έρευνα στον έλεγχο HVAC.....	17
II.2.1. Το μοντέλο θερμικής άνεσης του Fanger.....	17
II.2.2. Στοιχεία από τη βιβλιογραφία ελέγχου HVAC.....	18
II.3. Ο ρόλος του θερμοστάτη.....	19
II.4. Έξυπνοι θερμοστάτες.....	20
III. Θεωρητικό υπόβαθρο.....	23
III.1. Ενισχυτική μάθηση.....	23
III.1.1. Τυποποίηση του προβλήματος.....	24
III.2. Νευρωνικά δίκτυα.....	25
III.2.1. Το perceptron του Rosenblatt.....	27
III.2.2. Πολυεπίπεδα perceptrons (MLP's).....	28
III.2.3. Επιβλεπόμενη μάθηση και MLP's.....	29
III.3. Ο αλγόριθμος NFQ.....	30
IV. Περιγραφή του μηχανισμού αποφάσεων.....	32
IV.1. Καταστάσεις, δράσεις και αναπαράσταση εμπειρίας.....	32
IV.2. Συνάρτηση κόστους και το πρόβλημα της κανονικοποίησης.....	33
IV.3. Ζώνη λειτουργίας, τερματικές δράσεις και ο ελεγκτής έκτακτης ανάγκης.....	35
IV.4. Το νευρωνικό δίκτυο, η επιλογή δράσης και η συνάρτηση Q.....	36
IV.5. Το δίλημμα εξερεύνησης/εκμετάλλευσης και ο μηχανισμός εξερεύνησης.....	37
IV.6. Ανυπάκουα σημεία λειτουργίας.....	40
IV.7. Τελικά σχόλια και σύνοψη.....	40

V. Προσομοίωση και αξιολόγηση.....	42
V.1. Πλαίσιο εργασίας προσομοίωσης.....	42
V.2. Καιρός και κτίριο.....	43
V.3. Ζώνη λειτουργίας, ωράριο και άνθρωποι.....	45
V.4. Το πρόβλημα του ενδιάμεσου χρονικού βήματος.....	45
V.5. Αξιολόγηση επίδοσης.....	46
VI. Πειραματικά αποτελέσματα.....	48
VI.1. Χειμερινή περίοδος ($tr = 0$).....	48
VI.2. Χειμερινή περίοδος ($tr = 1$).....	53
VI.3. Χειμερινή περίοδος ($tr = 0.5$).....	56
VI.4. Καλοκαιρινή περίοδος ($tr = 0$).....	59
VI.5. Καλοκαιρινή περίοδος ($tr = 1$).....	63
VI.6. Καλοκαιρινή περίοδος ($tr = 0.5$).....	68
VI.7. Τελικά σχόλια και παρατηρήσεις.....	69
VI.7.1. Μετρήσεις σε συστήματα μικροεπεξεργαστών.....	70
VI.7.2. Παραδοχές και πιθανές επεκτάσεις.....	70

Λίστα σχημάτων

Σχήμα 1.1: Ενεργειακή κατανάλωση εμπορικών κτιρίων ΗΠΑ.....	14
Σχήμα 2.1: Διάγραμμα βασικής μονάδας συστήματος κλιματισμού.....	17
Σχήμα 2.2: Επέκταση βασικής μονάδας με σπειρώματα.....	20
Σχήμα 2.3: Επέκταση βασικής μονάδας με αναπλαστήρες.....	20
Σχήμα 2.4: Ο έξυπνος θερμοστάτης Nest.....	21
Σχήμα 3.1: Αρχή αλληλεπίδρασης πράκτορα-περιβάλλοντος.....	24
Σχήμα 3.2: Απεικόνιση βιολογικού νευρώνα.....	26
Σχήμα 3.3: Το perceptron του Rosenblatt.....	26
Σχήμα 3.4: Συνάρτηση XOR.....	27
Σχήμα 3.5: Λύση συνάρτησης XOR με MLP.....	28
Σχήμα 4.1: Συνοπτική περιγραφή μηχανισμού αποφάσεων.....	32
Σχήμα 4.2: Αρχιτεκτονική νευρωνικού δικτύου μηχανισμού αποφάσεων.....	37
Σχήμα 4.3: Αναλυτικό διάγραμμα ροής μηχανισμού αποφάσεων.....	41
Σχήμα 5.1: Πλαίσιο εργασίας προσομοίωσης.....	42
Σχήμα 5.2: Χρονοσειρές εξωτερικής θερμοκρασίας.....	43
Σχήμα 5.3: Χρονοσειρές εξωτερικής υγρασίας.....	43
Σχήμα 5.4: Χρονοσειρές ηλιακής ακτινοβολίας.....	44
Σχήμα 5.5: Απεικόνιση κτιρίου προσομοίωσης.....	44
Σχήμα 5.6: Ενδεικτικό διάγραμμα αριθμού ενοίκων ζώνης.....	45
Σχήμα 6.1: PMV τριών διαδοχικών ημερών (χειμώνας, $tr = 0$).....	48
Σχήμα 6.2: Κατανάλωση τριών διαδοχικών ημερών (χειμώνας, $tr = 0$).....	49
Σχήμα 6.3: Μεταβολή ημερησίων στατιστικών PMV (χειμώνας, $tr = 0$).....	50
Σχήμα 6.4: Μεταβολή ημερησίων στατιστικών κατανάλωσης (χειμώνας, $tr = 0$).....	50
Σχήμα 6.5: Εξοικονόμηση ενέργειας-άνεσης (χειμώνας, $tr = 0$).....	51
Σχήμα 6.6: Χρονική εξέλιξη λ και ε (χειμώνας, $tr = 0$).....	52
Σχήμα 6.7: Ιστογράμματα δράσεων-κόστους (χειμώνας, $tr = 0$).....	52
Σχήμα 6.8: PMV τριών διαδοχικών ημερών (χειμώνας, $tr = 1$).....	53
Σχήμα 6.9: Κατανάλωση τριών διαδοχικών ημερών (χειμώνας, $tr = 1$).....	54
Σχήμα 6.10: Μεταβολή ημερησίων στατιστικών PMV (χειμώνας, $tr = 1$).....	55
Σχήμα 6.11: Μεταβολή ημερησίων στατιστικών κατανάλωσης (χειμώνας, $tr = 1$).....	55
Σχήμα 6.12: Εξοικονόμηση ενέργειας-άνεσης (χειμώνας, $tr = 1$).....	56

Σχήμα 6.13: tr και ανυπάκουα σημεία λειτουργίας.....	56
Σχήμα 6.14: Χρονική εξέλιξη λ και ε (χειμώνας, $tr = 1$).....	58
Σχήμα 6.15: Ιστογράμματα δράσεων-κόστους (χειμώνας, $tr = 1$).....	58
Σχήμα 6.16: tr και μέσο ημερήσιο PMV (χειμώνας).....	59
Σχήμα 6.17: tr και μέση ημερήσια κατανάλωση (χειμώνας).....	59
Σχήμα 6.18: PMV τριών διαδοχικών ημερών (καλοκαίρι, $tr = 0$).....	60
Σχήμα 6.19: Κατανάλωση τριών διαδοχικών ημερών (καλοκαίρι, $tr = 0$).....	60
Σχήμα 6.20: Μεταβολή ημερησίων στατιστικών PMV (καλοκαίρι, $tr = 0$).....	61
Σχήμα 6.21: Μεταβολή ημερησίων στατιστικών κατανάλωσης (καλοκαίρι, $tr = 0$).....	61
Σχήμα 6.22: Εξοικονόμηση ενέργειας-άνεσης (καλοκαίρι, $tr = 0$).....	62
Σχήμα 6.23: Χρονική εξέλιξη λ και ε (καλοκαίρι, $tr = 0$).....	63
Σχήμα 6.24: Ιστογράμματα δράσεων-κόστους (καλοκαίρι, $tr = 0$).....	63
Σχήμα 6.25: PMV τριών διαδοχικών ημερών (καλοκαίρι, $tr = 1$).....	64
Σχήμα 6.26: Κατανάλωση τριών διαδοχικών ημερών (καλοκαίρι, $tr = 1$).....	64
Σχήμα 6.27: Μεταβολή ημερησίων στατιστικών PMV (καλοκαίρι, $tr = 1$).....	65
Σχήμα 6.28: Μεταβολή ημερησίων στατιστικών κατανάλωσης (καλοκαίρι, $tr = 1$).....	65
Σχήμα 6.29: Εξοικονόμηση ενέργειας-άνεσης (καλοκαίρι, $tr = 1$).....	66
Σχήμα 6.30: Χρονική εξέλιξη λ και ε (καλοκαίρι, $tr = 1$).....	67
Σχήμα 6.31: Ιστογράμματα δράσεων-κόστους (καλοκαίρι, $tr = 1$).....	67
Σχήμα 6.32: tr και μέσο ημερήσιο PMV (καλοκαίρι).....	68
Σχήμα 6.33: tr και μέση ημερήσια κατανάλωση (καλοκαίρι).....	68

Λίστα πινάκων

Πίνακας 1: Συγκεντρωτικά αποτελέσματα προσομοιώσεων..... 66

Λίστα παραθέσεων αλγορίθμων

Αλγόριθμος 1: Neural fitted Q-iteration (NFQ).....	30
Αλγόριθμος 2: Αρχή λειτουργίας ελεγκτή έκτακτης ανάγκης.....	36

Περίληψη

Οι περισσότεροι σημερινοί θερμοστάτες ρυθμίζονται χειροκίνητα από τον ίδιο τον χρήστη. Αυτό καθιστά το κόστος της λειτουργίας τους ασύμφορο, αφού κανείς δεν λαμβάνει υπ' όψιν τις διάφορες παραμέτρους που επηρεάζουν την ενεργειακή κατανάλωση, παρά επιδιώκει μονάχα την μεγιστοποίηση της θερμικής του άνεσης—στο οποίο μέγιστο σπανίως οδηγείται με την πρώτη προσπάθεια, καταλήγοντας έτσι σε ενεργοβόρες αυξομειώσεις της θερμοκρασίας.

Τα τελευταία χρόνια, ένα νέο είδος θερμοστάτη έχει εμφανιστεί. Αποκαλούμενοι έξυπνοι θερμοστάτες, οι συσκευές αυτές επιστρατεύουν τεχνικές όπως η μηχανική μάθηση προκειμένου αφ' ενός να σκιαγραφούν τις προτιμήσεις και τις συνήθειες του εκάστοτε χρήστη ή ομάδας χρηστών, αφ' ετέρου να παράγουν με τρόπο αυτόνομο και βέλτιστο (ως προς την ενεργειακή κατανάλωση και τη θερμική άνεση) τα σημεία λειτουργίας τους.

Το κίνημα των έξυπνων θερμοστατών συνοδεύεται από μια ευρύτερη προσπάθεια της επιστημονικής κοινότητας για τον αποτελεσματικό έλεγχο συστημάτων θέρμανσης, εξαερισμού, και κλιματισμού (HVAC). Πρόκειται για ένα δύσκολο ως προς την εύρεση γενικής λύσης πρόβλημα, αφού η ενεργειακή κατανάλωση και η θερμική άνεση όχι μόνο είναι αντικρουόμενα μεγέθη, αλλά εξαρτώνται επίσης από την αρχιτεκτονική και το είδος χρήσης του κτιρίου, το είδος του προς έλεγχο συστήματος, και στοχαστικές μεταβλητές όπως το πλήθος ενοίκων και ο καιρός.

Σκοπός αυτής της διπλωματικής εργασίας είναι η περιγραφή και αξιολόγηση μέσω προσομοίωσης ενός πρωτοτύπου έξυπνου θερμοστάτη καθολικής εφαρμογής και πολυμορφικής συμπεριφοράς. Γίνεται χρήση ενός αλγορίθμου ενισχυτικής μάθησης που είναι ελαφρύς υπολογιστικά, και δεν απαιτεί την ύπαρξη μοντέλου. Ο χρήστης καθορίζει την ισορροπία κατανάλωσης και άνεσης, και ο θερμοστάτης παράγει μια συμπεριφορά που αναβαθμίζεται διαρκώς και βελτιστοποιεί τα δύο μεγέθη ανάλογα. Υποστηρίζεται υλοποίηση σε ενσωματωμένες πλατφόρμες μικροελεγκτών ελάχιστου κόστους.

Λέξεις κλειδιά: Έξυπνος θερμοστάτης, Μηχανική μάθηση, Ενισχυτική μάθηση, Ευφυείς πράκτορες, Συστήματα γνώσης, Ενεργειακή αποδοτικότητα, Ενσωματωμένα συστήματα

Abstract

Most contemporary thermostats are manually regulated from the user. This results to non-beneficial function cost, since nobody takes the many different parameters regulating energy consumption into consideration, instead focusing in maximizing their thermal comfort—at which maximum hardly ever do they get with the first try, thus resolving in costly temperature fluctuations.

In recent years, a new kind of thermostats has arisen. Under the name of smart thermostats, these devices employ techniques such as machine learning in order to deduce each consumer's preferences and habits, and produce their setpoints of function in an autonomous and optimal (as regards energy consumption and thermal comfort) way.

The smart thermostats movement is complemented by a broader effort of the scientific community to effectively control heating, ventilation and air-conditioning (HVAC) systems. This is a hard problem in respect of finding a general solution, since energy consumption and thermal comfort are not only conflicting quantities, but they also depend on each building's architecture and usage purpose, the HVAC system's type, and stochastic variables like occupancy and weather.

The purpose of this thesis is to describe and evaluate, via simulation, the prototype of a universally applicable smart thermostat demonstrating polymorphic behavior. A computationally lightweight, model-free reinforcement learning algorithm is used. The user can set the tradeoff between consumption and comfort. The thermostat then produces a constantly self-refining behavior, accordingly optimizing the aforementioned quantities. Realization on embedded microcontroller platforms of minimal cost is supported.

Keywords: Smart thermostat, Machine learning, Reinforcement learning, Intelligent agents, Knowledge-based systems, Energy efficiency, Embedded systems

Ευχαριστίες

Υπό την ουσιωδέστερη έννοια, το εξώφυλλο αυτής της εργασίας λέει ψέμματα, καθώς δημιουργός της δεν είμαι μόνο εγώ. Με μεγάλη μου χαρά, λοιπόν, αφιερώνω αυτή την ενόπτηα στην αποκατάσταση της αλήθειας, ευγνωμονώντας όλους εκείνους που συνέβαλλαν στην τελμάτωση αυτού του όμορφου ταξιδιού.

Θα ξεκινίσω ευχαριστώντας τον άνθρωπο που, από την πρώτη μας συνάντηση, με εμπιστεύτηκε με ένα θέμα που στα μάτια μας τότε ήταν ένα ανοιχτό ερευνητικά πρόβλημα. Και το έκανε ενώ στο μοναδικό μάθημα που είχα την τιμή να με διδάξει, την Επιστήμη Υπολογιστών, είχα σημειώσει τον διόλου γοπτευτικό βαθμό πέντε (5). Δεν αναφέρομαι σε άλλον από τον επιβλέπων καθηγητή μου, κύριο Δημήτριο Σούντην. Οφείλω βέβαια να τονίσω πως ο κύριος Σούντης θα μείνει στη μνήμη μου για λόγους πολύ σημαντικότερους από την ανάθεση ενός θέματος διπλωματικής, λόγους τους οποίους μάλλον γνωρίζει και ο ίδιος, και συνεπώς δεν είναι ανάγκη να αναφέρω εδώ. Θα πω απλά πως στην περίπτωση της γνωριμίας μου μαζί του, η τύχη μου στάθηκε υπέρμετρη.

Συνεχίζω ευχαριστώντας από τα βάθη της καρδιάς μου τον υποψήφιο διδάκτωρα Χαράλαμπο Μάραντο, συνοδοιπόρο μου στην εργασία από την αρχή έως το τέλος. Ξέρω πως ο ίδιος ακόμα και σήμερα μάλλον υποτιμά τη συμβολή του, μα η αλήθεια είναι πως χωρίς τον Χάρη η μοίρα αυτής της εργασίας θα ήταν πολύ πιο επώδυνη, και πολύ πιο αμφισβητήσιμη. Δεν θεωρώ υπερβολή να πω πως δεν θα τα είχα καταφέρει δίχως τη βοήθειά του.

Ανάλογη ευγνωμοσύνη θέλω να εκφράσω απέναντι στους υπέροχους φίλους μου, τον Μήτσο, τον Γιαννάκη, τη Μαρία, την Ντίνα, τον Θεωρή, τον Αλέξανδρο, τον Βασίλη, τον Αργύρη, τον Γιώργο, τον Γιάννη, τον Νίκο, τον Τάσο, τον Βιτάλη, και όποιον άλλο με υπέμεινε, τόσο όταν κατενθουσιαζόμουν με ένα μάτσο σχήματα που τύχαινε να έχω δει τη Χ μέρα στο MATLAB, όσο και όταν προσπαθούσα να τους πείσω πως αυτή η διπλωματική ήταν η πιο παράλογη απόφαση που είχα πάρει ποτέ. Ευχαριστώ που κάνετε τον ορισμό μου για τη λέξη “φίλος” τόσο αυστηρό, μάγκες.

Υπάρχουν, επίσης, δύο γυναίκες στις οποίες θα χρωστάω για πάντα κάτι περισσότερο από τα πάντα, η μητέρα μου Αναστασία και η αδελφή μου Μαριάννα. Μαζί τους, όλη μου η οικογένεια (και εσύ, θεία Παναγιώτα!) μου παρείχε μέχρι τέλους την πιο ζεστή και συναρπαστικά αβίαστη στήριξη. Να είστε πάντα καλά, λοιπόν, και να λογαριάζετε σ’ εμένα.

Οι τελευταίοι τρεις μήνες εκπόνησης αυτής της εργασίας συνέπεσαν με τους τελευταίους τρεις μήνες ζωής του πατέρα μου, Παναγιώτη Θ. Λαμπράκου. Ό,τι ακολουθεί είναι αφιερωμένο στη μνήμη του. Ήσουν ο γενναιότερος άνθρωπος που θα γνωρίσω ποτέ, μπαμπά. Σ' ευχαριστώ που με έμαθες ποιος είμαι.

Σας ευχαριστώ όλους.

Αθήνα, 29 Οκτωβρίου 2017

ΧΠΛ

I. Εισαγωγή

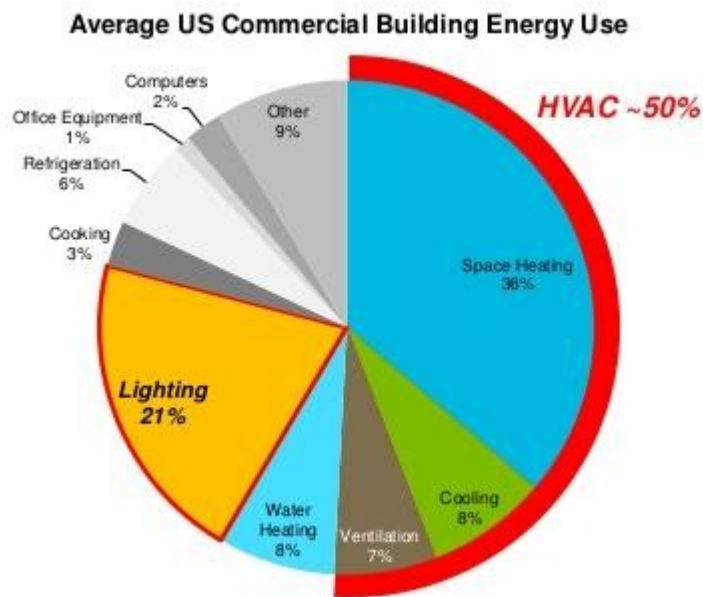
“Το τίμημα της αποτυχίας είναι υπερβολικά υψηλό”.

Έτσι ξεκινάει το έγγραφο [1] που υπεγράφη στις Βρυξέλλες τη 10η Νοεμβρίου 2010, απευθυνόμενο από την Κομισιόν στους υπόλοιπους ευρωπαϊκούς φορείς. Το έγγραφο αφορούσε την πρωτοβουλία ενεργειακής αποδοτικότητας που έκτοτε υιοθετήθηκε από την Ευρωπαϊκή Ένωση, και έχει ως στόχο την βελτίωση της ενεργειακής αποδοτικότητας κατά 20% ως το 2020.

Αυτή ήταν μόνο η αρχή. Το Νοέμβριο του 2016 η Κομισιόν ανανέωσε την πρωτοβουλία, προσθέτοντας ένα νέο στόχο, αυτή τη φορά για βελτίωση 30% ως το 2030 [2]. Παρόμοιοι στόχοι κοιτάζουν τόσο μακροπρόθεσμα όσο και έως το 2050 [3]. Αποκορύφωμα όλης αυτής της δράσης ήταν η ιστορική Συμφωνία του Παρισιού, που υπεγράφη τον Δεκέμβριο του 2015 από 195 διαφορετικές χώρες του κόσμου, προκειμένου να καταπολεμηθεί σε παγκόσμια κλίμακα η κλιματική αλλαγή [4].

Στην Ευρωπαϊκή Ένωση, τα κτίρια ευθύνονται για το 40% της ενεργειακής κατανάλωσης, και για το 36% των εκπομπών διοξειδίου του άνθρακα [5]. Η συντριπτική πλειοψηφία της κατανάλωσης προέρχεται από ανάγκες θέρμανσης και ψύξης, όπως αυτές προκύπτουν από τη λειτουργία συστημάτων θέρμανσης, εξαερισμού και κλιματισμού (HVAC). Επιπλέον, το 84% της ενέργειας που χρησιμοποιείται για λόγους θέρμανσης/ψύξης προέρχεται από ορυκτά, ενώ μόλις το 16% προκύπτει από ανανεώσιμες πηγές ενέργειας. Παραπλήσια ποσοστά ισχύουν και στις Ηνωμένες Πολιτείες της Αμερικής (βλ. Σχ. 1.1). Γίνεται έτσι προφανής η αξία λύσεων αποδοτικού ελέγχου τέτοιων συστημάτων.

Αυτή η διπλωματική εργασία παρουσιάζει το πρωτότυπο μιας τέτοιας λύσης. Σχεδιάσαμε έναν έξυπνο θερμοστάτη ενισχυτικής μάθησης που επιδεικνύει ενεργειακή εξοικονόμηση έως και 60% σε σύγκριση με την τυπική, χειροκίνητη χρήση. Ο θερμοστάτης αυτός υπακούει στην επιθυμητή ισορροπία μεταξύ κατανάλωσης και θερμικής άνεσης, και βελτιώνει τη συμπεριφορά του με το πέρασμα του χρόνου. Δεν βασίζεται σε κάποια συγκεκριμένη πληροφορία όσον αφορά το κτίριο ή το κλίμα. Η γνώση που αποκομίζει είναι οι ίδιες οι πράξεις του, και τα αποτελέσματά τους. Η τροχιά της συμπεριφοράς του οριοθετείται από τα λάθη που έκανε στο παρελθόν—και κάνει ό,τι καλύτερο προκειμένου να τα αποφύγει στο μέλλον. Αντί για ένα υπερβολικά υψηλό



Σχήμα 1.1: Διάγραμμα ενεργειακής κατανάλωσης σε εμπορικά κτίρια των ΗΠΑ. Σχεδόν το 50% δεσμεύεται σε συστήματα θέρμανσης, εξαερισμού και κλιματισμού.

τίμημα, στην προκειμένη περίπτωση η αποτυχία επιφυλάσσει εξαιρετικά πολύτιμη γνώση.

Το υπόλοιπο της εργασίας είναι οργανωμένο ως εξής: η ενότητα II παρέχει πληροφορίες για τα διάφορα είδη και τη λειτουργία των HVAC συστημάτων, τις στρατηγικές ελέγχου τους, το κίνημα των έξυπνων θερμοστατών και άλλες λύσεις που προτείνονται στη βιβλιογραφία.

Στην ενότητα III θεμελιώνεται το απαραίτητο θεωρητικό υπόβαθρο πάνω στο οποίο στηρίχτηκε ο θερμοστάτης της εργασίας. Παρουσιάζεται το οικοσύστημα ενός προβλήματος ενισχυτικής μάθησης (πράκτορες, περιβάλλον, ζεύγη κατάστασης-δράσης, συνάρτηση τιμής), και αναλυτικότερα ο στηριγμένος επ' αυτής βασικός αλγόριθμος ελέγχου που χρησιμοποιούμε (Neural Fitted Q-Iteration). Βαρύτητα δίνεται επίσης στα πολυεπίπεδα νευρωνικά δίκτυα (Multilayer Perceptrons) ως μέσα για την προσέγγιση συναρτήσεων. Στην ίδια ενότητα αναλύεται το μοντέλο ποσοτικοποίησης της θερμικής άνεσης (δείκτης PMV).

Η ενότητα IV αποτελεί την καρδιά της εργασίας, αφού περιγράφει λεπτομερώς τη διαδικασία με την οποία ο θερμοστάτης μας παράγει τα σημεία λειτουργίας του. Δίνεται έμφαση σε όλα τα στάδια που μεσολαβούν μέχρι την τελική απόφαση, στις δομές δεδομένων που αναπαριστούν τη γνώση, και στους τρόπους με τους οποίους αντιμετωπίζονται διάφορα παρελκόμενα της

συγκεκριμένης μεθόδου επίλυσης (δίλημμα εξερεύνησης/εκμετάλλευσης, συντήρηση και ανανέωση γνώσης).

Η οργάνωση και υλοποίηση των πειραμάτων/προσομοιώσεων που διεξήχθησαν για την αξιολόγηση του θερμοστάτη παρατίθενται στην ενότητα V.

Τέλος, στην ενότητα VI παρουσιάζονται και σχολιάζονται αναλυτικά τα αποτελέσματα των πειραμάτων. Αξιολογείται τόσο ο ίδιος ο έλεγχος, όσο και η αποδοτικότητα της μάθησης.

II. Συστήματα HVAC, έλεγχος και έξυπνοι θερμοστάτες

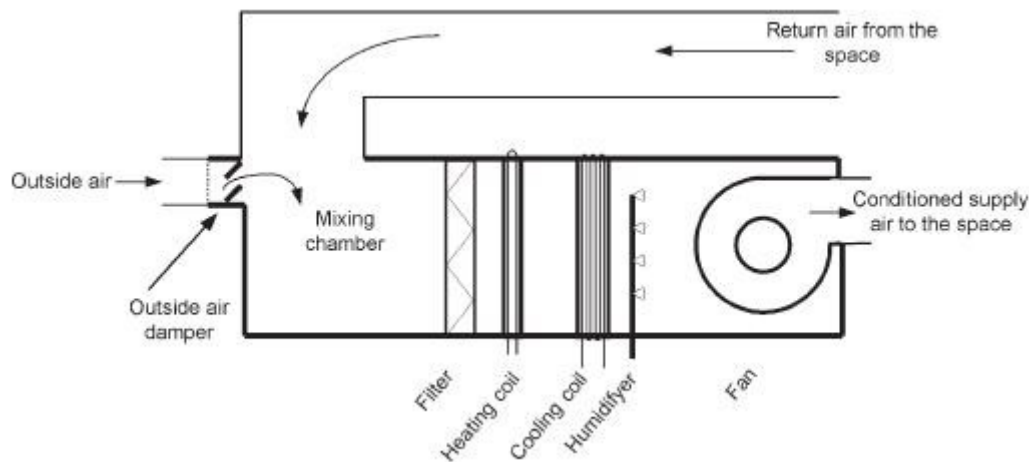
II.1. Βασικές έννοιες HVAC

Ο πρώτος ιστορικός εκπρόσωπος της οικογένειας των συστημάτων HVAC θα μπορούσε να θεωρηθεί μια αναμμένη φωτιά που ζέστανε το εσωτερικό μιας σπηλιάς. Σήμερα ο όρος περικλείει ένα πραγματικά αχανές πεδίο, από μία απλή σόμπα μέχρι το σύστημα κλιματισμού ενός υποβρυχίου ή ενός διαστημοπλοίου.

Έξι βασικές διεργασίες χαρακτηρίζουν τον πλήρη κλιματισμό ενός χώρου [6]:

- *Θέρμανση*: αφορά την προσθήκη θερμικής ενέργειας στον κλιματιζόμενο χώρο, για λόγους αύξησης ή διατήρησης της θερμοκρασίας του.
- *Ψύξη*: η αντίστροφη διεργασία της θέρμανσης.
- *Υγρανση*: αφορά την προσθήκη υδρατμών στον αέρα ενός κλιματιζόμενου χώρου προκειμένου να αυξηθεί ή να διατηρηθεί η υγρασία του.
- *Αφύγρανση*: ομοίως, η αντίστροφη διεργασία της υγρανσης.
- *Καθαρισμός*: η διεργασία που είναι υπεύθυνη για την απομάκρυνση σκόνης και βιολογικών ρύπων, όπως έντομα και γύρη, για τη διατήρηση ή τη βελτίωση της ποιότητας του αέρα ενός χώρου.
- *Εξαερισμός*: παραπλήσια διαδικασία του καθαρισμού, που αφορά όμως αέριους ρύπους και τη γενικότερη σύνθεση της ατμόσφαιρας. Ουσιαστικά είναι η ανταλλαγή εσωτερικού-εξωτερικού αέρα. Μπορεί να επιτευχθεί είτε φυσικά, όπως για παράδειγμα με δύο ανοιχτά παράθυρα που δημιουργούν ρεύμα, είτε μηχανικά (πχ ανεμιστήρες).

Η διατέλεση των παραπάνω φαίνεται συνοπτικά στο Σχ. 2.1, όπου απεικονίζεται μία βασική μονάδα κλιματισμού.



Σχήμα 2.1: Διάγραμμα βασικής μονάδας κλιματισμού [6]. Ο εξωτερικός και ο εσωτερικός αέρας αναμειγνύονται και φιλτράρονται. Κατόπιν, ανάλογα με τις ανάγκες του χώρου, το μείγμα είτε θερμαίνεται είτε ψύχεται. Ομοίως ακολουθεί είτε ύγρανση είτε αφύγρανση, πριν τη διανομή του κλιματισμένου πια αέρα, μέσω του ανεμιστήρα, στο υπόλοιπο σύστημα.

II.2. Σύγχρονη έρευνα στον έλεγχο HVAC

Σημαντική έμφαση έχει δοθεί από την επιστημονική κοινότητα σε τεχνικές ελέγχου HVAC συστημάτων. Τα δύο βασικά μεγέθη που χρησιμοποιούνται για την αξιολόγησή τους είναι η κατανάλωση ενέργειας και η θερμική άνεση των ανθρώπων. Όλη η σχετική βιβλιογραφία προϋποθέτει πρόσβαση σε αυτά.

Η ενεργειακή κατανάλωση είναι το ευκολότερο να μετρηθεί από τα δύο μεγέθη. Για παράδειγμα, στην περίπτωση μιας συνηθισμένης κατοικίας, ένα βατόμετρο μπορεί να παρεμβληθεί μεταξύ της παροχής ρεύματος και ενός κλιματιστικού και να καταγράφει τακτικές μετρήσεις των Watt (ηλεκτρική ισχύς). Στο υπόλοιπο μήκος αυτής της εργασίας υποθέτουμε την ύπαρξη ενός ανάλογου μηχανισμού.

Όμως τι συμβαίνει με την άνεση; Στην υποενότητα που ακολουθεί παρουσιάζεται ένα δημοφιλές μοντέλο που την εκτιμά με ακρίβεια. Το ίδιο μοντέλο χρησιμοποιήθηκε για τις ανάγκες της εργασίας.

III.2.1. Το μοντέλο θερμικής άνεσης του Fanger

Με τον όρο “θερμική άνεση” εννοείται η ικανοποίηση του ανθρώπου σε σχέση με το θερμικό του περιβάλλον. Φυσικά πρόκειται για προϊόν υποκειμενικής αξιολόγησης. Παρ’ όλα αυτά, στον κλάδο των μηχανικών συστημάτων

κλιματισμούν προέβαλλε από πολύ νωρίς επιτακτική η ανάγκη ποσοτικοποίησης του μεγέθους.

Το μοντέλο PMV [7] (Predicted Mean Vote) αναπτύχθηκε από τον P. O. Fanger, ο οποίος βασίστηκε σε εξισώσεις θερμικής ισορροπίας και σε εμπειρικές μελέτες που διεξήγαγε. Η κλίμακα PMV έχει τιμές στο διάστημα [-3, 3] και αντιπροσωπεύει τη θερμική άνεση που θα ένιωθε ένα μεγάλο πλήθος ανθρώπων υπό συγκεκριμένες τιμές των ακόλουθων παραμέτρων:

- Θερμοκρασία αέρα
- Μέση ακτινοβολώμενη θερμοκρασία
- Σχετική υγρασία
- Ταχύτητα αέρα
- Μεταβολικός ρυθμός
- Μόνωση ρουχισμού

Η μέση ακτινοβολώμενη θερμοκρασία αντιστοιχεί στην εκπομπή θερμότητας από τις επιφάνειες ενός κτιρίου (τοίχοι, έπιπλα). Η σχετική υγρασία εκφράζεται σε ποσοστό και αντιπροσωπεύει την παρουσία υδρατμών στην ατμόσφαιρα για μια συγκεκριμένη θερμοκρασία. Ο μεταβολικός ρυθμός σχετίζεται με το είδος δραστηριότητας των ανθρώπων ενός χώρου. Αυξάνεται με την ένταση της δραστηριότητας. Η μόνωση λόγω ρουχισμού αντανακλά την εκάστοτε εποχή: όσο μεγαλύτερη είναι, τόσο πιο “βαριά” είναι ντυμένοι οι άνθρωποι (μικρές τιμές το καλοκαίρι, μεγάλες το χειμώνα).

Το PMV αναγνωρίζεται ως έγκυρο μοντέλο αξιολόγησης της θερμικής άνεσης από οργανισμούς όπως η Αμερικανική Κοινότητα Μηχανικών Θέρμανσης, Ψύξης και Κλιματισμού (ASHRAE) [8], και βάσει αυτού οριοθετούνται οι συνθήκες θερμικής άνεσης σε εσωτερικούς χώρους. Ως αποδεκτές θεωρούνται οι τιμές στο διάστημα [-1, 1], ενώ ως προτεινόμενες αυτές στο [-0.5, 0.5].

Το πρόσημο του PMV σηματοδοτεί το αν οι άνθρωποι νιώθουν πιο κρύα (αρνητικό πρόσημο) ή πιο ζεστά (θετικό πρόσημο) από το ιδανικό, στο οποίο αντιστοιχεί η τιμή 0.

III.2.2. Στοιχεία από τη βιβλιογραφία ελέγχου HVAC

Οι τεχνικές που μελετά η σύγχρονη βιβλιογραφία όσον αφορά τον έλεγχο συστημάτων κλιματισμού παρουσιάζουν μεγάλη ποικιλία στην κλίμακα του εφαρμοζόμενου ελέγχου, υπό την έννοια του πόσο σχετίζονται με το φυσικό σύστημα οι παραγόμενες μεταβλητές εισόδου.

Για παράδειγμα, στο [9] γίνεται η εικασία πως το σημείο λειτουργίας του θερμοστάτη έχει ήδη δοθεί, και ο έλεγχος έγκειται στη γρήγορη και ακριβή μετάβαση της θερμοκρασίας σε αυτή την τιμή.

Στο [10] το σύστημα είναι αρκετά παραπλήσιο αυτού του Σχ. 2.2, και οι μεταβλητές ελέγχου είναι η πίεση και η θερμοκρασία του αέρα στον αεραγωγό. Ο έλεγχος σε αυτή την περίπτωση λοιπόν περιλαμβάνει και τμήματα του συστήματος (ανεμιστήρες που ελέγχουν την πίεση) που δεν σχετίζονται με τον θερμοστάτη όπως τον ορίσαμε παραπάνω.

Στο [11] οι μεταβλητές ελέγχου παίρνουν τιμές από ένα διάστημα πραγματικών αριθμών, οι οποίες αντιστοιχίζονται σε ηλεκτρικά σήματα από μια συνάρτηση-μαύρο κουτί.

Ιδιαίτερο ενδιαφέρον—και τη μεγαλύτερη σχέση με την παρούσα εργασία—παρουσιάζει η επιλογή των [12] και [13] να εξάγουν απευθείας τα σημεία λειτουργίας για τη θερμοκρασία της ζώνης που ελέγχεται από το θερμοστάτη. Τα αποτελέσματα της εφαρμογής του εκάστοτε σημείου συνοψίζονται από την προκύπτουσα ενεργειακή κατανάλωση και την αντίστοιχη θερμική άνεση, οι οποίες και συνθέτουν το κόστος προς ελαχιστοποίηση. Με αυτό τον τρόπο, ο έλεγχος γίνεται πιο αφαιρετικός—και συνεπώς πιο καθολικός.

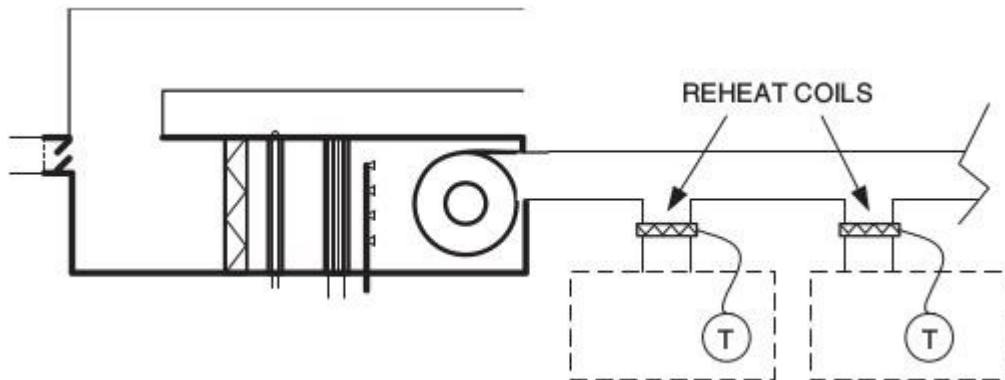
II.3. Ο ρόλος του θερμοστάτη

Στα Σχ. 2.2-3 απεικονίζονται επεκτάσεις της βασικής μονάδας κλιματισμού που εισήχθη παραπάνω σε δύο διαφορετικές ζώνες. Με τον όρο “ζώνη” εννοούμε το συγκεκριμένο τμήμα ενός κτιρίου, του οποίου η θερμοκρασία ελέγχεται από έναν θερμοστάτη. Ο θερμοστάτης αυτός μπορεί είτε να αποτελεί τμήμα του κεντρικού συστήματος κλιματισμού, είτε να έχει τη δική του μονάδα HVAC.

Στον ευρύτερο κύκλο, λοιπόν, του κλιματισμού ενός κτιρίου, ο θερμοστάτης αναλαμβάνει δύο από τις έξι προαναφερθείσες βασικές διεργασίες: τη θέρμανση και την ψύξη της εκάστοτε ζώνης.

Για παράδειγμα, σε μία διάταξη αντίστοιχη αυτής του Σχ. 2.2, ο θερμοστάτης θα μπορούσε να ελέγχει μέσω ηλεκτρικών σημάτων το σπείρωμα θέρμανσης στο πέρας του αεραγωγού. Βάσει του σημείου λειτουργίας που του έχει δοθεί, ο θερμοστάτης ζεσταίνει το σπείρωμα διαρκώς (για λόγους απλότητας θεωρούμε πως η θερμοκρασία της ζώνης είναι μικρότερη του σημείου λειτουργίας), παρακολουθώντας παράλληλα την άνοδο της θερμοκρασίας της ζώνης. Όταν αυτή γίνει ίση με το σημείο λειτουργίας, ο θερμοστάτης παύει να ζεσταίνει το σπείρωμα.

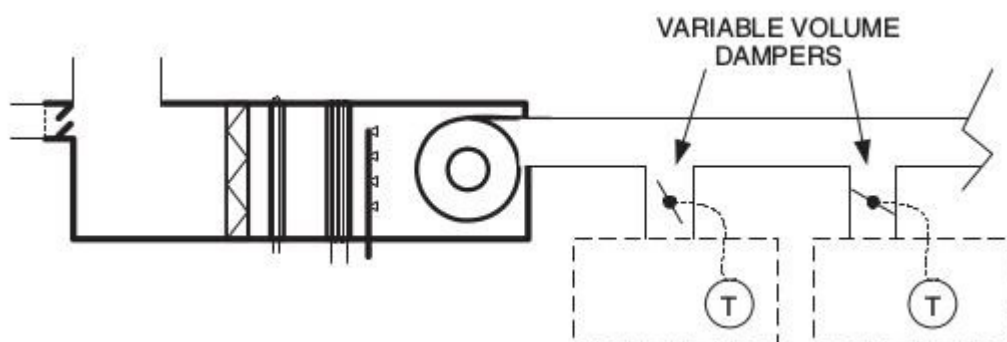
Συμπεραίνουμε, λοιπόν, πως ο έλεγχος ενός συστήματος HVAC **εσωκλείει** τον έλεγχο, υπό την έννοια των σημείων λειτουργίας, του θερμοστάτη.



Σχήμα 2.2: Επέκταση της βασικής μονάδας κλιματισμού [6] σε περισσότερες από μία ζώνες, κάθε μία εκ των οποίων ελέγχεται από ξεχωριστό θερμοστάτη. Στην προκειμένη περίπτωση, μεταξύ κάθε ζώνης και του αεραγωγού παρεμβάλλονται σπειρώματα θέρμανσης.

II.4. Έξυπνοι θερμοστάτες

Μέχρι πριν λίγα χρόνια, οι θερμοστάτες λειτουργούσαν απλοϊκά. Η τεχνολογία δεν ήταν σε θέση να τους εφοδιάσει με κάτι παραπάνω από έναν αισθητήρα θερμοκρασίας, και όσον αφορά το σημείο λειτουργίας, αυτό δινόταν απευθείας από τον χρήστη. Η εξέλιξη των ενσωματωμένων συστημάτων επέτρεψε την εμφάνιση συσκευών με περισσότερα εξαρτήματα και πιο σύνθετες λειτουργίες. Οι θερμοστάτες απέκτησαν πρόσβαση σε μετρήσεις από αισθητήρες υγρασίας, ηλιακής ακτινοβολίας και άλλων μεγεθών. Απέκτησαν επίσης μικροεπεξεργαστές και μονάδες μνήμης που επέτρεπαν τον προγραμματισμό των συσκευών αυτών.



Σχήμα 2.3: Επέκταση ανάλογη της 2.2 [6], εφοδιασμένη με αποσβεστήρες μεταβλητού όγκου αντί σπειρωμάτων.

Μέσω του Internet of Things (IoT) κατέστη δυνατή η επικοινωνία με τα κινητά τηλέφωνα των χρηστών, ο απομακρυσμένος έλεγχος (ο χρήστης μπορεί για παράδειγμα να ενεργοποιήσει τον θερμοστάτη μέσω του διαδικτύου καθώς επιστρέφει από τη δουλειά και έτσι να βρει το σπίτι ζεστό), η ανταλλαγή δεδομένων χρήσιμων για την εξαγωγή μοντέλων κινητικότητας στον εκάστοτε χώρο, θερμικής άνεσης, κ.ά.

Έτσι εμφανίστηκαν στην αγορά οι έξυπνοι θερμοστάτες, που υπόσχονται εξοικονόμηση σε λογαριασμούς ρεύματος δίχως επιβάρυνση ως προς τη θερμική άνεση των ενοίκων. Ο πιο πετυχημένος τέτοιος θερμοστάτης αυτή τη στιγμή στην αγορά είναι ο Nest (Σχ. 2.4). Στη σελίδα του προϊόντος [14] διαφημίζεται εξοικονόμηση της τάξης του 10-12%.

Ο Nest αποτέλεσε την αρχική έμπνευση για τη μέθοδο επίλυσης που ακολουθήσαμε σε αυτή την εργασία. Φημολογείται πως χρησιμοποιεί ενισχυτική μάθηση, όμως ο ακριβής αλγόριθμος δεν έχει δημοσιευτεί στη βιβλιογραφία.



Σχήμα 2.4: Ο έξυπνος θερμοστάτης Nest [14], με δυνατότητες εξοικονόμησης ενέργειας 10-12%. Η τιμή του κατά τη συγγραφή αυτής της εργασίας ανέρχεται στα 249 δολάρια.

Φυσικά, δεν ισχυριζόμαστε πως η υλοποίηση του έξυπνου θερμοστάτη μας ως έχει θα μπορούσε να αντικαταστήσει πλήρως προϊόντα όπως ο Nest. Για παράδειγμα, τόσο ο Nest όσο και αντίστοιχα προϊόντα μπορούν να διακόπτουν τη λειτουργία του υπό έλεγχο HVAC, σε περίπτωση που το κρίνουν

απαραίτητο. Μπορούν επίσης να προβλέπουν πότε ο εκάστοτε χώρος θα είναι κατειλημμένος ή όχι, και έτσι να διαμορφώνουν το δικό τους, εξατομικευμένο ωράριο λειτουργίας. Αντιθέτως, ο δικός μας θερμοστάτης λειτουργεί τις ώρες που ορίζει ο ιδιοκτήτης, και η απενεργοποίηση του συστήματος δεν είναι στις διαθέσιμες επιλογές. Εξ ου και ο χαρακτηρισμός του ως πρωτότυπο. Σε επόμενη ενότητα βέβαια προτείνονται πιθανοί τρόποι επέκτασης της λειτουργικότητάς του.

III. Θεωρητικό υπόβαθρο

III.1. Ενισχυτική μάθηση

Η ενισχυτική μάθηση είναι ένας συναρπαστικός κλάδος, περισσότερο γνωστός με την αγγλική του ονομασία, reinforcement learning (RL).

Η κεντρική ιδέα πίσω από την ενισχυτική μάθηση ακραγγίζει αυτό που αποκαλούμε φυσική μάθηση, τον τρόπο δηλαδή με τον οποίο μαθαίνει ο ίδιος ο άνθρωπος. Φανταστείτε ένα μικρό παιδί το οποίο κάνει τις πρώτες του απόπειρες να οδηγήσει ένα ποδήλατο. Αρχικά δεν έχει ιδέα για έννοιες όπως η ισορροπία ή ο συγχρονισμός των άκρων. Πιθανότατα δεν γνωρίζει καν πώς πρέπει να ανέβει στη σέλα. Έχει δει φίλους, αδέρφια ή γονείς να καβαλάνε τα δικά τους ποδήλατα, ναι, όμως αυτή τη στιγμή η σέλα δεν είναι κατειλημμένη, και τα πόδια του παιδιού αντί για τα πεντάλ στηρίζονται στο έδαφος.

Όμως το παιδί είναι αποφασισμένο να τα καταφέρει. Αν το κάνει, θα μπορεί και αυτό να παίζει με τους φίλους του, και στο κάτω-κάτω, παιχνίδι ή όχι, η ικανοποίηση της περιέργειάς του για το πώς είναι να πηγαίνεις τόσο γρήγορα είναι αρκετή για να του δώσει ένα πολύ ισχυρό κίνητρο.

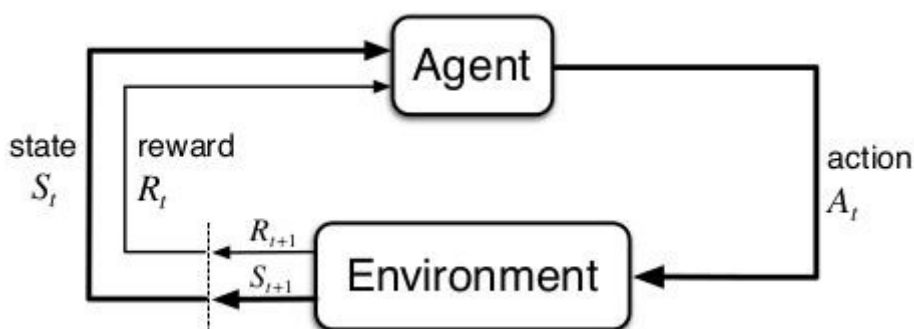
Οι πρώτες απόπειρες διαρκούν ελάχιστα. Ίσως μάλιστα κάποιες από αυτές καταλήξουν σε μικροτραυματισμούς. Σιγά-σιγά, το παιδί πέρα από φιλοδοξίες αναπτύσσει τις απαραίτητες επιφυλάξεις. Μαζί με αυτές, έστω και ασυνείδητα, έρχονται ανάλογοι συσχετισμοί μεταξύ του τι κάνει, και πότε το κάνει. Αισίως καταφέρνει να κάνει τις πρώτες πεταλιές χωρίς να χάσει την ισορροπία του, και αυτό ήταν! Πλέον είναι έτοιμο να κατακτήσει τον κόσμο (ή, τέλος πάντων, όσο από τον κόσμο αντιστοιχεί σε ένα παιδί και το ποδήλατό του).

Στην ενισχυτική μάθηση, οντότητες που ονομάζουμε πράκτορες λαμβάνουν ξανά και ξανά αποφάσεις μέσω των οποίων αλληλεπιδρούν με το περιβάλλον τους (βλ. Σχ. 3.1). Αυτό, με τη σειρά του, ανάλογα με το ποια ήταν η απόφαση επιστρέφει κάθε φορά ένα αντίστοιχο κόστος. Όπως το παιδί (τις περισσότερες φορές) δεν γνωρίζει τους νόμους της κινηματικής που διέπουν την πολυπόθητη βόλτα με το ποδήλατο, έτσι και ο πράκτορας δεν γνωρίζει και δεν μπορεί να επηρεάσει τον τρόπο με τον οποίο το κόστος εξάγεται από το περιβάλλον. Η μόνη του ελπίδα είναι, αφιερώνοντας πολύ μεγάλη προσοχή στις συνθήκες υπό τις οποίες έλαβε την εκάστοτε απόφαση, να συσσωρεύσει αρκετή

εμπειρία ώστε να καταφέρει, από κάποιο σημείο και έπειτα, να μαντεύει ποια απόφαση θα του επιστρέψει κάθε φορά το μικρότερο δυνατό κόστος.

Η ενισχυτική μάθηση έχει επιφέρει εντυπωσιακά αποτελέσματα σε διάφορες εφαρμογές της. Κάποιες από αυτές είναι πράκτορες που έχουν μάθει να παίζουν τάβλι [15], να οδηγούν πραγματικά ρομποτικά αυτοκίνητα [16] και να παίζουν το ίδιο καλά (αν όχι καλύτερα) με τον άνθρωπο μια ευρεία γκάμα βιντεοπαιχνιδιών [17]. Επίσης, σημαντική έρευνα διεξάγεται στην αξιοποίηση μεθόδων της ενισχυτικής μάθησης σε τομείς όπως η ψυχολογία και η νευροεπιστήμη [18].

Ο θερμοστάτης αυτής της εργασίας είναι ως επί το πλείστον ένας πράκτορας ενισχυτικής μάθησης.



Σχήμα 3.1: Αρχή αλληλεπίδρασης πράκτορα-περιβάλλοντος στο πλαίσιο της ενισχυτικής μάθησης [18]. Εδώ αντί για κόστος ο πράκτορας λαμβάνει μια επιβράβευση R την οποία επιδιώκει να μεγιστοποιήσει.

III.1.1. Τυποποίηση του προβλήματος

Οι βασικές δομικές μονάδες ενός προβλήματος ενισχυτικής μάθησης είναι:

- το σύνολο καταστάσεων S
- το σύνολο διαθέσιμων δράσεων A
- η συνάρτηση κόστους $c(s, a, s'): S \times A \rightarrow R$

Κατά τη λειτουργία του ένας πράκτορας ενισχυτικής μάθησης βρίσκεται ξανά και ξανά ενώπιον της επιλογής μιας δράσης $a \in A$ όντας στην κατάσταση $s \in S$. Η δράση αυτή θα προκαλέσει μέσω του περιβάλλοντος μια μετάβαση σε μια νέα κατάσταση $s' \in S$, και με τη σειρά της αυτή η μετάβαση θα αντιστοιχιστεί σε ένα κόστος c .

Ανά πάσα στιγμή, λοιπόν, μπορούμε να ορίσουμε το εξής:

$$R = \sum_{i=0}^{N-1} \gamma^i \cdot c_i(s_i, a_i, s_{i+1})$$

Το R αντιπροσωπεύει το συνολικό κόστος που απομένει να λάβει ο πράκτορας στις εναπομείνουσες N αποφάσεις του. Το $\gamma \in [0,1)$ εξασφαλίζει πως το άθροισμα συγκλίνει ακόμα και σε περιπτώσεις απείρου χρονικού ορίζοντα.

Όπως προαναφέρθηκε, ο πράκτορας δεν γνωρίζει τη συνάρτηση κόστους, και άρα δεν γνωρίζει το R . Ορίζεται η αναμενόμενη τιμή του R σε κάθε ζεύγος κατάστασης-δράσης:

$$Q(s, a) = E[R|(s, a)]$$

Η συνάρτηση Q είναι αυτή που προσπαθεί να μάθει ο πράκτορας προκειμένου να επιτύχει το σκοπό του.

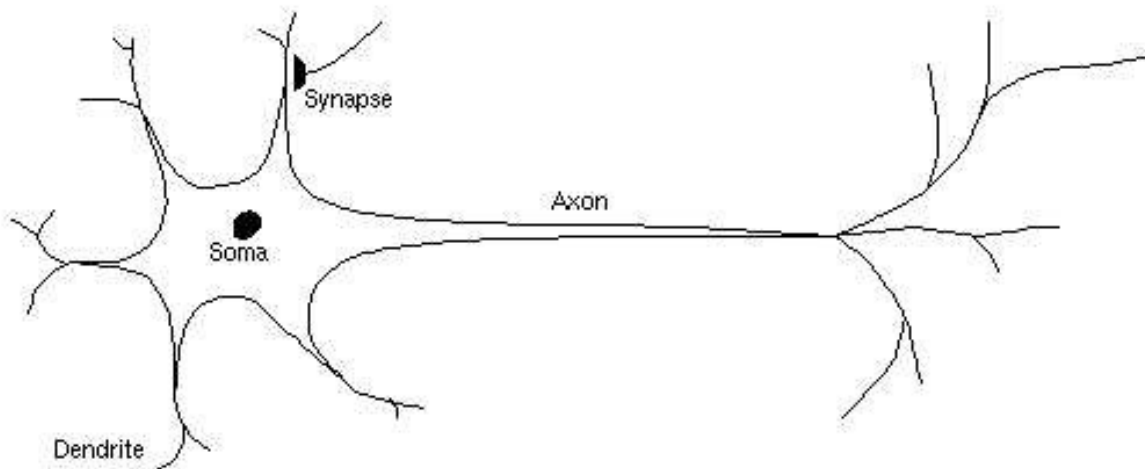
Σε περιπτώσεις όπου ο χώρος $S \times A$ είναι μικρός (πεπερασμένες, διακριτές τιμές), η συνάρτηση μπορεί να υπολογιστεί ακριβώς σε μορφή πίνακα. Σε πρακτικές εφαρμογές όπως αυτή της προκειμένης διπλωματικής εργασίας, όπου γίνεται χρήση συνεχών μεταβλητών, καθίσταται απαραίτητη η ύπαρξη κάποιας μορφής προσέγγισης συναρτήσεων.

Μία ευρέως χρησιμοποιούμενη μέθοδος είναι η αναπαράσταση της προσεγγισθείσας συνάρτησης σε παραμετρική μορφή, στα βάρη ενός πολυεπίπεδου νευρωνικού δικτύου.

III.2. Νευρωνικά δίκτυα

Η ακριβής ονομασία του όρου είναι τεχνητά νευρωνικά δίκτυα, αφού οι δημιουργοί τους εμπνεύστηκαν από τον τρόπο λειτουργίας βιολογικών νευρώνων.

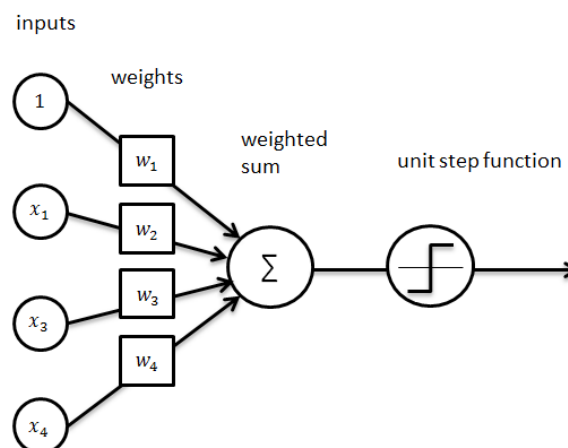
Η βασική μορφή ενός βιολογικού νευρώνα απεικονίζεται στο Σχ. 3.2. Οι απολήξεις μέσω των οποίων ο νευρώνας λαμβάνει πληροφορίες από τους υπόλοιπους ονομάζονται δενδρίτες. Οι συνάψεις παίζουν τον ρόλο της ενίσχυσης, ή της υποβάθμισης του σήματος. Το σύνολο της πληροφορίας επεξεργάζεται στο σώμα του νευρώνα. Αν το σήμα είναι αρκετά ισχυρό, δηλαδή αν προκαλεί μεταβολή του δυναμικού της μεμβράνης μεγαλύτερη από ένα συγκεκριμένο κατώφλι, ο νευρώνας εκπολώνεται. Με άλλα λόγια, παράγει ένα δυναμικό δράσης. Το δυναμικό αυτό είναι η “πληροφορία” που αναφέραμε παραπάνω, και μεταδίδεται μέσω του άξονα του νευρώνα, και των δενδριτών εξόδου, στους γύρω νευρώνες.



Σχήμα 3.2: Ένας βιολογικός νευρώνας. Σήματα από άλλους νευρώνες λαμβάνονται μέσω των δενδριτών και ενισχύονται ή υποβιβάζονται μέσω των συνάψεων. Η πληροφορία κατόπιν επεξεργάζεται στο σώμα του νευρώνα. Αν το σήμα είναι αρκετά ισχυρό, ο νευρώνας εκπολώνεται και δυναμικά δράσης μεταδίδονται μέσω του άξονα στους υπόλοιπους νευρώνες, μέσω των δενδριτών εξόδου. Εκεί, η παραπάνω διαδικασία επαναλαμβάνεται.

Τα νευρωνικά δίκτυα δεν είναι νέα ανακάλυψη. Ο πρώτος τεχνητός νευρώνας κατασκευάστηκε το 1943 από τους McCulloch και Pitts [19], όμως μεσολάβησε μεγάλο χρονικό διάστημα μέχρι η επιστημονική κοινότητα να μελετήσει σοβαρά την ιδέα.

Αυτή η εργασία χρησιμοποιεί ένα πολυεπίπεδο νευρωνικό δίκτυο, και συγκεκριμένα ένα πολυεπίπεδο perceptron.



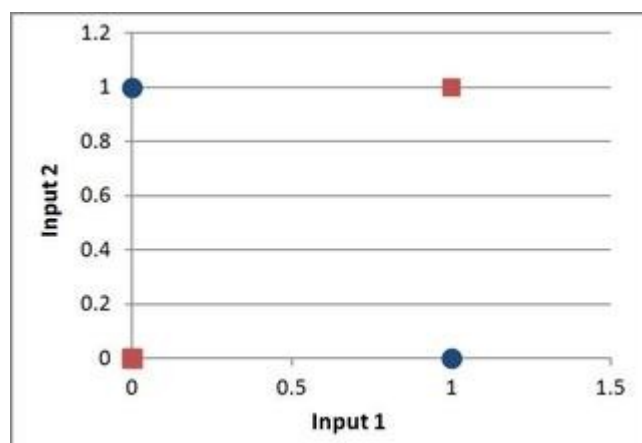
Σχήμα 3.3: Το perceptron του Rosenblatt, το οποίο μιμείται τη λειτουργία ενός βιολογικού νευρώνα. Αν το σταθμισμένο άθροισμα των εισόδων είναι θετικό, το perceptron εκπέμπει 1. Διαφορετικά, εκπέμπει 0.

III.2.1. Το perceptron του Rosenblatt

Το perceptron (που στα ελληνικά φέρει το διόλου γοπευτικό όνομα “αντίληπτρο”) εφευρέθηκε το 1957 από τον Frank Rosenblatt [20] στο Αεροναυπηγικό Εργαστήριο του Πανεπιστημίου του Κορνέλ, για λογαριασμό του Αμερικανικού Πολεμικού Ναυτικού.

Μετά από μια αρχική περίοδο ενθουσιασμού, κατά την οποία ο Rosenblatt διαφήμιζε το perceptron σαν την βασική δομική μονάδα μιας γενιάς υπολογιστών τεχνητής νοημοσύνης που θα μπορούσε να “βλέπει, να νιώθει, να αναπαράγεται και να εξελίσσει τη συνείδησή της”, οι δυνατότητες του perceptron δέχτηκαν βαριά αμφισβήτηση και η έρευνα που θα στηριζόταν πάνω του δεν ήρθε ποτέ.

Χονδρικά, η χρησιμότητα των perceptrons έγκειται στην ικανότητά τους για ταξινόμηση γραμμικά διαχωρίσιμων προτύπων. Τα βάρη εισόδου που απεικονίζονται στο Σχ. 3.3 ορίζουν για παράδειγμα μία τετραδιάστατη επιφάνεια (τόσες διαστάσεις όσα και τα βάρη), η οποία μπορεί να διαχωρίζει ομάδες δεδομένων όπως, ας πούμε, έργα τέχνης μπρεσιονισμού και μεταμοντερνισμού, ή τραγούδια ροκ και χιπ-χοπ μουσικής. Ένα σύνολο από χαρακτηριστικά δίνονται ως είσοδος στο perceptron (στην περίπτωση του Σχ. 3.3 τέσσερα) και αν τα βάρη έχουν τις σωστές τιμές, και τα χαρακτηριστικά έχουν επιλεγεί έτσι ώστε να χαρακτηρίζουν πλήρως τα μέλη των δύο πληθυσμών, τότε το perceptron κάνει σωστή ταξινόμηση.



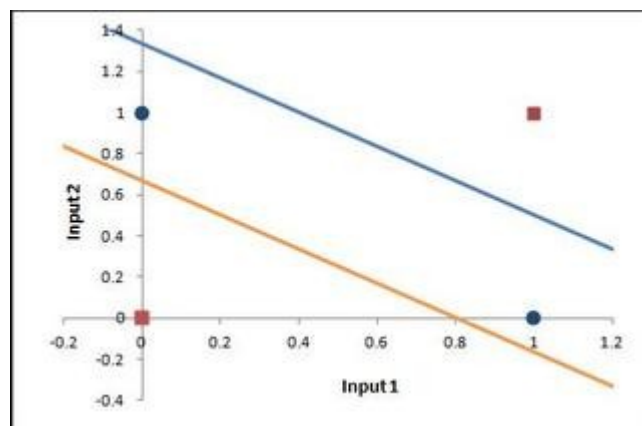
Σχήμα 3.4: Η αξίλλειος πέτρα του perceptron: η συνάρτηση XOR, η οποία δεν είναι γραμμικά διαχωρίσιμη.

Όπως είπαμε, όμως, τα perceptrons λειτουργούν μόνο όταν η ταξινόμηση γίνεται σε γραμμικά διαχωρίσιμες κλάσεις δεδομένων. Ένα απλό διαισθητικά

παράδειγμα στον διδιάστατο χώρο είναι η συνάρτηση αποκλειστικού OR (XOR). Τα χαρακτηριστικά που το perceptron θα χρησιμοποιούσε για την ταξινόμηση θα ήταν προφανώς οι δύο συντεταγμένες του εκάστοτε σημείου στο χώρο. Τα βάρη εισόδου του perceptron θα αναπαριστούσαν μια ευθεία γραμμή, η οποία θα έπρεπε να μπορεί να χαραχτεί με τέτοιο τρόπο, ώστε στο ένα ημιεπίπεδο που όριζε να βρίσκονται πρότυπα μόνο της μίας κλάσης, και στο άλλο πρότυπα μόνο της άλλης. Είναι εμφανές πως στην περίπτωση της XOR αυτό δεν μπορεί να γίνει.

III.2.2. Το πολυεπίπεδο perceptron (MLP)

Η απάντηση στον περιορισμό των γραμμικά διαχωρίσιμων προβλημάτων ήρθε από τα πολυεπίπεδα perceptrons, για συντομία MLP (βλ. Σχ. 3.5).



Σχήμα 3.5: Η λύση του προβλήματος της XOR όπως προκύπτει από τη χρήση ενός νευρωνικού δικτύου perceptron δύο επιπέδων. Οι νευρώνες του πρώτου επιπέδου ορίζουν τις δύο γραμμές που φαίνονται στο σχήμα, ενώ ο νευρώνας του δεύτερου επιτελεί τη λογική πράξη OR (η οποία είναι γραμμικά διαχωρίσιμη).

Κατά κάποιον τρόπο, τα MLP απεκατέστησαν το όνομα των perceptrons. Δίχως να μπορούμε σε λεπτομέρειες, θα αναφέρουμε το Θεώρημα Καθολικής Προσέγγισης [21], σύμφωνα με το οποίο ένα MLP με ένα κρυφό επίπεδο νευρώνων (δηλαδή ένα επίπεδο που παρεμβάλλεται μεταξύ των νευρώνων εισόδου και εξόδου) μπορεί να προσεγγίσει, υπό χαλαρές προϋποθέσεις, οποιαδήποτε συνεχή συνάρτηση.

Φυσικά, το παραπάνω θεώρημα δεν αποτελεί πανάκεια. Η πλειάδα των προβλημάτων με πρακτική εφαρμογή στον φυσικό κόσμο είναι πολύ πιο σύνθετη από μια απλή ταξινόμηση στον διδιάστατο χώρο. Συχνά σαν είσοδος

δίνονται πρότυπα με δεκάδες ή εκατοντάδες χαρακτηριστικά. Τα βάρη του δικτύου, προκειμένου να ταξινομήσουν σωστά, δεν καθορίζονται εποπτικά, αλλά αποτελούν προϊόντα επαναληπτικών διαδικασιών εκπαίδευσης. Το Θεώρημα Καθολικής Προσέγγισης δεν εγγυάται πως αυτές οι διαδικασίες εκπαίδευσης θα συγκλίνουν πάντα σε πεπερασμένο χρόνο, παρά μόνο πως *κάποια στιγμή* (για παράδειγμα, στην αρχή της επόμενης χιλιετίας) θα συνέκλιναν.

III.2.3. Επιβλεπόμενη μάθηση και MLP's

Στην προηγούμενη ενότητα αναφέρθηκαν διαδικασίες εκπαίδευσης μέσα από τις οποίες καθορίζονται τα βάρη ενός MLP για τη διατέλεση μιας συγκεκριμένης λειτουργίας. Οι συνηθέστερες των λειτουργιών είναι η ταξινόμηση και η παλινδρόμηση.

Στην περίπτωση μας, το δίκτυο χρησιμοποιείται για παλινδρόμηση. Με αυτό τον όρο ορίζεται η ικανότητα πρόβλεψης της τιμής μιας άγνωστης συνάρτησης έχοντας ως είσοδο χαρακτηριστικά που την περιγράφουν. Η διαδικασία με την οποία τα βάρη ενός νευρωνικού δικτύου ρυθμίζονται προκειμένου να επιτύχουν επιτυχή παλινδρόμηση (ή ταξινόμηση) ονομάζεται, στην περίπτωση που έχουμε στη διάθεσή μας ιστορικό παλαιότερων ζευγών προτύπων και τιμών της συνάρτησης, επιβλεπόμενη μάθηση.

Η επιβλεπόμενη μάθηση συνοψίζεται ως εξής:

- Τα βάρη του δικτύου αρχικοποιούνται σε τυχαίες τιμές.
- Για κάθε ένα από τα ζεύγη προτύπων-τιμών, τα χαρακτηριστικά δίνονται ως είσοδος στο δίκτυο. Το σφάλμα μεταξύ της εξόδου και της σωστής τιμής (διαθέσιμης από τα δεδομένα), και οι μερικές παράγωγοι του σφάλματος ως προς το κάθε βάρος του δικτύου, χρησιμοποιούνται για την ανάλογη ανανέωση των βαρών.
- Η διαδικασία επαναλαμβάνεται επαναληπτικά (κάθε μία από τις επαναλήψεις ονομάζεται εποχή) μέχρι την εκπλήρωση κάποιου κριτηρίου τερματισμού. Για παράδειγμα, μέχρι το μέσο τετραγωνικό σφάλμα να φτάσει σε ένα επιθυμητό κατώτατο όριο, ή μέχρι η επίδοση του δικτύου να μη βελτιώνεται για έναν προκαθορισμένο αριθμό διαδοχικών εποχών.

Ο κλάδος της μηχανικής μάθησης είναι πολύ πιο ενδιαφέρων και ανεπτυγμένος από την εντύπωση που μπορεί να σχηματίσει ο αναγνώστης από τις παραπάνω πληροφορίες, οι οποίες είναι περισσότερο ενδεικτικές για το

υπόβαθρο των τεχνικών γνώσεων του συγγραφέα (και τη διάθεσή του να εμβαθύνει σε έννοιες που στην παρούσα εργασία αντιμετωπίζονται ως μαύρο κουτί), παρά για οτιδήποτε άλλο. Παραπέμπουμε στο εξαιρετικό σύγγραμμα του [22] για μια αξιοπρεπή και λεπτομερέστατη ξενάγηση στον κλάδο.

III.3. Ο αλγόριθμος NFQ

Ο πυρήνας του μηχανισμού αποφάσεων που περιγράφεται σε αυτή την εργασία είναι ο αλγόριθμος NFQ. Η πλήρης ονομασία του είναι Neural Fitted Q-iteration, και η δημιουργία του χρεώνεται στον Martin Riedmiller [23].

Ο αλγόριθμος περιγράφεται ακολούθως:

```

NFQ_main() {
input: a set of transition samples  $D$ ; output: Q-value function  $Q_N$ 
  k=0
  init_MLP()  $\rightarrow Q_0$ ;
  Do {
    generate_pattern_set  $P = \{(input^l, target^l), l = 1, \dots, \#D\}$  where:
       $input^l = s^l, u^l$ ,
       $target^l = c(s^l, u^l, s'^l) + \gamma \min_b Q_k(s'^l, b)$ 
    Rprop_training( $P$ )  $\rightarrow Q_{k+1}$ 
    k:= k+1
  } WHILE ( $k < N$ )

```

Αλγόριθμος 1: Ο αλγόριθμος Neural Fitted Q-iteration [23].

Όπως προαναφέρθηκε, η συνάρτηση Q αναπαριστά, για κάθε ζεύγος κατάστασης-δράσης, την αναμενόμενη τιμή του συσσωρευμένου κόστους. Ο πράκτοράς μας επιδιώκει να ελαχιστοποιήσει αυτό το κόστος. Η παρούσα εκτίμησή του για τη συνάρτηση Q βρίσκεται αποθηκευμένη σε παραμετρική μορφή στα βάρη ενός MLP.

Το MLP αρχικοποιείται τυχαία. Ο πράκτορας αλληλεπιδρά με το περιβάλλον, αποθηκεύοντας διαδοχικά τις εμπειρίες του στη μορφή τριπλέτων (s, a, c) . Κάθε τέτοια τριπλέτα αντιπροσωπεύει μια μετάβαση από την κατάσταση s , μέσω της δράσης a , στην κατάσταση s' . Το αποτέλεσμα αυτής της μετάβασης συνοψίζεται από την τιμή του κόστους c . Όταν έχουν συσσωρευτεί αρκετές μεταβάσεις, η γνώση του πράκτορα για την Q ανανεώνεται offline μέσω της διαδικασίας που φαίνεται στον Αλγ. 1. Κατόπιν, ο πράκτορας συνεχίζει να συλλέγει εμπειρίες, πλέον δρώντας βασιζόμενος στη νέα του

εκτίμηση για την Q. Οι νέες αυτές εμπειρίες προστίθενται στις ήδη υπάρχουσες, και την επόμενη φορά που το δίκτυο θα εκπαιδευτεί, θα χρησιμοποιηθούν όλες μαζί.

Τα πλεονεκτήματα του NFQ συνοψίζονται ως εξής:

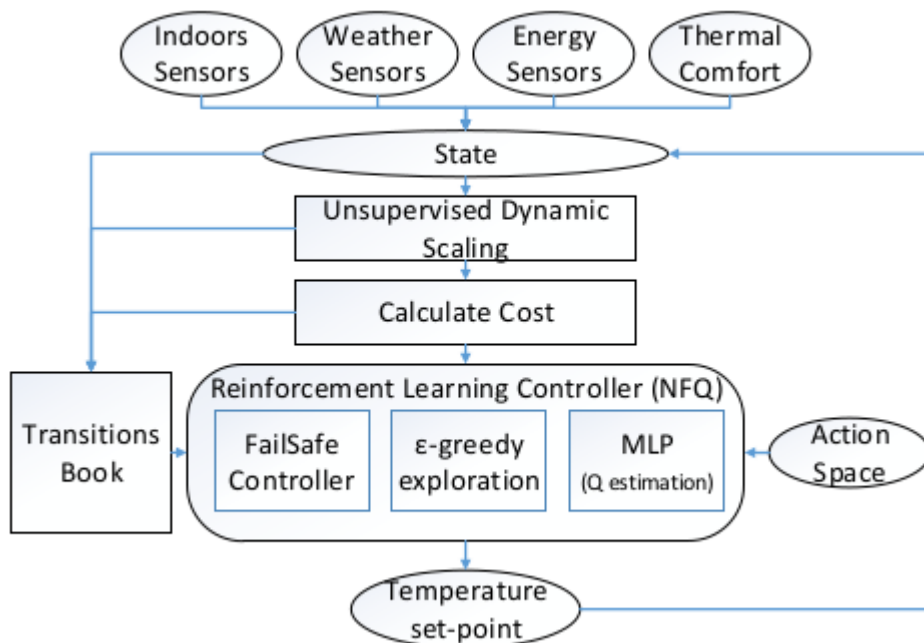
- η μέθοδος δεν βασίζεται σε κάποιο μοντέλο. Η μόνη πληροφορία που απαιτείται για τον έλεγχο είναι οι μεταβάσεις που παράγονται από τον ίδιο τον έλεγχο.
- εκπαιδύοντας το δίκτυο offline σε όλες τις ως τώρα μεταβάσεις, εξασφαλίζεται αφ' ενός η διατήρηση της γνώσης, καθίσταται εφικτή αφ' ετέρου η χρήση αλγορίθμων εκπαίδευσης batch learning, οι οποίες συγκλίνουν πιο γρήγορα. Ο αλγόριθμος εκπαίδευσης που χρησιμοποιεί ο NFQ χρεώνεται επίσης στον Riedmiller, και όπως φαίνεται και από τον Αλγ. 1 είναι ο RPROP [24].
- λόγω του παραπάνω, η μέθοδος είναι αποδοτική ως προς τη χρήση δεδομένων, και παράγει πολιτικές υψηλής ποιότητας με σχετικά λίγες μεταβάσεις.

Φυσικά, δεν έχουμε ακόμα αναφερθεί στον τρόπο με τον οποίο ο πράκτορας συλλέγει τις μεταβάσεις του στην περίπτωσή μας, και στο πότε οι μεταβάσεις αυτές θεωρούνται αρκετές προκειμένου το νευρωνικό δίκτυο να επανεκπαιδευτεί. Η επόμενη ενότητα αποσαφηνίζει αυτά τα ερωτήματα, αφού περιγράφει τη μέθοδο ελέγχου μας στο σύνολό της.

IV. Περιγραφή του μηχανισμού αποφάσεων

Η λειτουργία του μηχανισμού που παρουσιάζεται σε αυτή την εργασία έγκειται στη διαδοχική παραγωγή σημείων λειτουργίας για το θερμοστάτη ενός συστήματος κλιματισμού, βελτιστοποιημένων ως προς την επίτευξη της επιθυμητής ισορροπίας μεταξύ ενεργειακής κατανάλωσης και θερμικής άνεσης, η οποία και δίδεται από τον χρήστη.

Σε αυτή την ενότητα αναλύεται ο κύκλος ενεργειών που λαμβάνουν χώρα κάθε φορά, μέχρι την παραγωγή του εν λόγω σημείου λειτουργίας, μια συνοπτική περιγραφή του οποίου μπορεί να βρεθεί στο Σχ. 4.1, και μια αναλυτικότερη στο Σχ. 4.3.



Σχήμα 4.1: Συνοπτική περιγραφή του μηχανισμού παραγωγής για τα σημεία λειτουργίας του θερμοστάτη.

IV.1. Καταστάσεις, δράσεις και αναπαράσταση εμπειρίας

Κάθε κύκλος εκκινεί με τη λήψη του σήματος παρούσας κατάστασης. Το σήμα αυτό απαρθίζεται από μετρήσεις για τις εξής μεταβλητές:

- Εξωτερική θερμοκρασία
- Ηλιακή ακτινοβολία
- Επίπεδο εσωτερικής υγρασίας
- Εσωτερική θερμοκρασία

Προκειμένου να είναι εφικτή η προσέγγιση της συνάρτησης Q , το σήμα κατάστασης ενός ελεγκτή ενισχυτικής μάθησης πρέπει να πληροί την μαρκοβιανή ιδιότητα: να περιέχει δηλαδή όλη την πληροφορία που είναι απαραίτητη για την περιγραφή της συνάρτησης κόστους—και κατά συνέπεια της συνάρτησης Q . Στην περίπτωσή μας, αυτό σημαίνει πως το σήμα κατάστασης πρέπει να είναι σε θέση να περιγράψει την ενεργειακή κατανάλωση και τη θερμική άνεση που προέκυψαν από την εφαρμογή του εκάστοτε σημείου λειτουργίας. Τα αποτελέσματα της ενότητας VI δείχνουν πως το συγκεκριμένο σήμα κατάστασης είναι επιτυχές υπό αυτή την έννοια.

Ο όρος “δράση”, τώρα, αναφέρεται στο ίδιο το σημείο λειτουργίας που δίνεται κάθε φορά στον θερμοστάτη. Ο χώρος που ορίζεται από το σύνολο των καταστάσεων και το σύνολο των δράσεων πρέπει να τηρείται όσο το δυνατόν μικρότερος, προκειμένου η μάθηση να είναι όσο το δυνατόν ταχύτερη. Οι δράσεις, λοιπόν, στην περίπτωσή μας είναι:

- Διατήρηση της εσωτερικής θερμοκρασίας
- Αύξηση της εσωτερικής θερμοκρασίας κατά 1°C
- Μείωση της εσωτερικής θερμοκρασίας κατά 1°C

Η γνώση του ευφυούς μας πράκτορα είναι το σύνολο των μεταβάσεων που είδαμε ωρίτερα να χρησιμοποιεί ο αλγόριθμος NFQ. Αποτελείται από όλες τις καταστάσεις που έχουν συναντηθεί, όλες τις δράσεις που έχουν ληφθεί, και όλα τα κόστη που έχουν προκύψει σαν αποτέλεσμα. Θα αναφερόμαστε σε αυτό το ιστορικό ως TB, από την αγγλική φράση Transitions Book. Το TB αποτελεί τη θεμελιώδη δομή δεδομένων από την οποία, σύμφωνα με τον Αλγ. 1, εξάγονται κάθε φορά τα πρότυπα και οι στόχοι εκπαίδευσης του MLP που αναπαριστά τη συνάρτηση Q .

IV.2. Συνάρτηση κόστους και το πρόβλημα της κανονικοποίησης

Η συνάρτηση κόστους υπολογίζεται ως εξής:

$$c(s, a, s') = \begin{cases} tr \cdot E_{std} + (1 - tr) \cdot |PMV|_{std} & \text{(non-terminal)} \\ 10 & \text{else} \end{cases}$$

Η μεταβλητή tr παίρνει τιμές στο διάστημα $[0, 1]$ και εκφράζει το πόσο επιθυμεί ο χρήστης ο έλεγχος να εστιάσει στη βελτιστοποίηση της κατανάλωσης (και κατά συνέπεια να αγνοήσει τη θερμική άνεση). Στις ακραίες περιπτώσεις

$tr=0$ και $tr=1$, ο έλεγχος που προκύπτει είναι καθαρά εστιασμένος στην άνεση και την κατανάλωση αντίστοιχα.

Οι τιμές E_{std} και $|PMV|_{std}$ είναι η ενεργειακή κατανάλωση σε Watts και η θερμική άνεση όπως ορίζεται από την κλίμακα του Fanger που αναλύθηκε σε ποηγούμενη ενότητα, κανονικοποιημένες στο διάστημα $[-1, 1]$. Όσον αφορά την άνεση, χρησιμοποιούμε την απόλυτο τιμή της διότι θέλουμε ο έλεγχος να ελαχιστοποιεί την απόστασή της από το ιδανικό, που αναπαρίσταται από την τιμή 0. Κατά συνέπεια, δράσεις που έχουν ως αποτέλεσμα κάτι εξίσου πιο κρύο ή πιο ζεστό από το ιδανικό θεωρούνται ισότιμες.

Ένα πρόβλημα κατά την εκτέλεση on-line ελέγχου που δεν βασίζεται σε κάποιο μοντέλο είναι η απουσία οποιασδήποτε πληροφορίας σχετικά με τα δεδομένα που συναντιώνται. Η ενεργειακή κατανάλωση και η θερμική άνεση ανήκουν σε τελείως διαφορετικές κλίμακες, όμως πρέπει να αντιμετωπίζονται ισότιμα από τη συνάρτηση κόστους, προκειμένου η τιμή του tr να έχει ουσιώδη σημασία.

Κατά συνέπεια, μια κανονικοποίηση των τιμών τους είναι αναγκαία. Όμως πώς μπορεί να επιτευχθεί μια τέτοια, ακριβής κανονικοποίηση, όταν δεν έχουμε ιδέα για το γενικό εύρος των τιμών; Μια ικανοποιητική λύση, την οποία και χρησιμοποιούμε σε αυτή την εργασία, είναι το *unsupervised dynamic scaling* [25]. Η μέθοδος αυτή κανονικοποιεί ένα μέγεθος βάσει της τρέχουσας μέσης τιμής του, και της αντίστοιχης τυπικής απόκλισης, ως εξής:

$$X_{std} = \frac{X - \mu^k}{\delta^k}$$

όπου μ^k η τρέχουσα μέση τιμή και $\delta^k = \sqrt{\frac{s^k}{k-1}}$ η τρέχουσα τυπική απόκλιση.

Οι τύποι για την ανανέωσή τους είναι:

$$\mu^k = \mu^{k-1} + \frac{X - \mu^{k-1}}{k} \quad \text{και} \quad s^k = s^{k-1} + (X - \mu^{k-1})(X - \mu^k)$$

Σε κάθε νέα μετάβαση, λοιπόν, οι ανωτέρω εκτιμήσεις των στατιστικών χαρακτηριστικών για την κατανάλωση και τη θερμική άνεση ανανεώνονται, και βάσει αυτών κανονικοποιούνται τα μεγέθη. Τονίζουμε επίσης πως κάθε ανανέωση συνεπάγεται και αντίστοιχα “νέα” συνάρτηση κόστους, οι έως τώρα τιμές της οποίας είναι αποθηκευμένες στην ΤΒ. Φροντίζουμε λοιπόν να ανανεώνουμε όλη αυτή τη γνώση σε κάθε χρονικό βήμα.

IV.3. Ζώνη λειτουργίας, τερματικές δράσεις και ο ελεγκτής έκτακτης ανάγκης

Στα πρωταρχικά στάδια λήψης αποφάσεων, κανείς δεν εγγυάται πως ο ελεγκτής μας θα πράξει σωστά, αφού δεν έχει προλάβει να συσσωρεύσει αρκετή γνώση. Είναι λοιπόν απαραίτητος ο ορισμός μιας ζώνης ασφαλούς λειτουργίας, έξω από την οποία ο έλεγχος απαγορεύεται αυστηρά να οδηγήσει το σύστημα. Δανειστήκαμε αυτή την ιδέα από το [16], όπου μια αντίστοιχη ζώνη είχε οριστεί προκειμένου να εξασφαλιστεί πως ο έλεγχος δεν θα οδηγούσε το αμάξι έξω από τα όρια του δρόμου.

Έχοντας λοιπόν υπ' όψιν τα όρια αποδεκτών τιμών για το PMV στη βιβλιογραφία, ορίσαμε αυτή τη ζώνη λειτουργίας ως $|PMV| \leq 0.75$. Μια εύλογη ερώτηση είναι γιατί δεν χρησιμοποιήσαμε απ' ευθείας την περιοχή $[-0.5, 0.5]$ η οποία και θεωρείται η ιδανική. Η απάντηση είναι πως αποφασίσαμε να θυσιάσουμε λίγη από την άνεση των χρηστών, σε περίπτωση που επιθυμιά τους είναι κυρίως η βελτιστοποίηση της ενεργειακής κατανάλωσης, η οποία και είναι χαμηλότερη σε περιοχές όπου το PMV είναι υψηλότερο.

Μια δράση, λοιπόν, χαρακτηρίζεται τερματική (και αντιστοιχίζεται στην τερματική τιμή κόστους 10, όπως φαίνεται στον ορισμό της συνάρτησης κόστους της προηγούμενης υποενότητας) στις ακόλουθες δύο περιπτώσεις:

- η δράση οδήγησε το PMV έξω από τη ζώνη λειτουργίας
- πριν τη λήψη της δράσης, το PMV ήταν ήδη εκτός της ζώνης λειτουργίας, και ακολούθως η δράση αύξησε, αντί να μειώσει, την απόλυτη τιμή του.

Ο δεύτερος όρος εξασφαλίζει πως ο έλεγχος θα οδηγή την άνεση εντός της ζώνης λειτουργίας σε ελάχιστο χρόνο. Τονίζουμε δε πως η τήρηση της ζώνης λειτουργίας ισχύει ανεξάρτητα από την τιμή του tr .

Δύο πράγματα συμβαίνουν όταν ο πράκτορας λάβει τερματικό κόστος. Αφ' ενός, ο έλεγχος για το επόμενο χρονικό βήμα μεταφέρεται στον ελεγκτή έκτακτης ανάγκης, του οποίου καθήκον είναι να επανορθώσει για το λάθος του πράκτορα. Αφ' ετέρου, λαμβάνει χώρα μία ακόμα εκπαίδευση του νευρωνικού δικτύου σύμφωνα με τον αλγόριθμο NFQ.

Η δράση του ελεγκτή έκτακτης ανάγκης συνοψίζεται από τον ακόλουθο αλγόριθμο:

```
if  $PMV > 0$  then  
    decrease indoor temperature by  $1^{\circ}C$   
else  
    increase indoor temperature by  $1^{\circ}C$   
end if
```

Αλγόριθμος 2: Αρχή λειτουργίας του ελεγκτή έκτακτης ανάγκης.

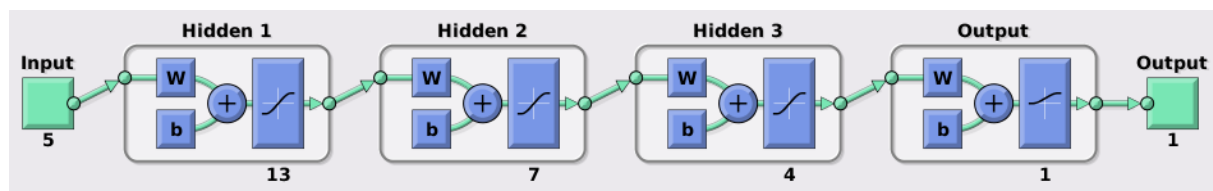
IV.4. Το νευρωνικό δίκτυο, η επιλογή δράσης και η συνάρτηση Q

Έχει ήδη αναφερθεί πως η συνάρτηση Q, η οποία υπενθυμίζουμε πως αντιστοιχίζει ζεύγη καταστάσεων-δράσεων σε εκτιμήσεις του συνολικού κόστους που συνεπάγονται, αναπαρίσταται από τα συναπτικά βάρη ενός MLP.

Υπενθυμίζουμε ακόμα τον συντελεστή γ στον ορισμό του συνολικού κόστους, ο οποίος εξασφαλίζει πως αυτό θα είναι πεπερασμένο. Μια δεύτερη ερμηνεία του συντελεστή είναι το εύρος του χρονικού ορίζοντα βάσει του οποίου δρα κάθε φορά ο πράκτορας. Μπορεί, για παράδειγμα, το άμεσο κόστος μιας δράσης να είναι μεγαλύτερο από αυτό μιας άλλης, όμως στην πρώτη περίπτωση ο έλεγχος να οδηγηθεί σε μια τροχιά της οποίας τα αθροισμένα κόστη είναι λιγότερα από τα αντίστοιχα που θα προέκυπταν με την επιλογή της δεύτερης δράσης. Αν το γ ισούται με μηδέν, ο πράκτορας δρα με βάση αποκλειστικά τα άμεσα κόστη.

Σε αυτή την εργασία, και λόγω του ότι ορίσαμε τις διαθέσιμες δράσεις του ελεγκτή ως βηματικές αυξομειώσεις της τρέχουσας θερμοκρασίας, επιλέξαμε τιμή $\gamma=0.98$. Για κάθε κατάσταση που συναντά, ο πράκτορας εξετάζει την έξοδο του νευρωνικού δικτύου (η οποία ταυτίζεται με την εκτίμηση της εκάστοτε τιμής της συνάρτησης Q) δοκιμάζοντας επαναληπτικά ως είσοδο τα τρία διαθέσιμα ζεύγη κατάστασης-δράσης. Τελικά επιλέγει τη δράση που αντιστοιχεί στη μικρότερη έξοδο του νευρωνικού, δηλαδή σε αυτή που θεωρεί πως θα του αποφέρει το ελάχιστο συνολικό κόστος στο μέλλον.

Η αρχιτεκτονική του νευρωνικού μας περιγράφεται στο Σχ. 4.2:



Σχήμα 4.2: Αρχιτεκτονική νευρωνικού δικτύου για την αναπαράσταση της συνάρτησης Q . Η είσοδος έχει 5 μεταβλητές, εκ των οποίων οι 4 πρώτες αφορούν την κατάσταση και η τελευταία την υπό εξέταση δράση. Το δίκτυο έχει 13 κόμβους στην είσοδο, έναν κόμβο στην έξοδο, και επίσης δύο κρυφά επίπεδα με 7 και 4 νευρώνες αντίστοιχα. Όλοι οι νευρώνες πλιν της εξόδου χρησιμοποιούν τη σιγμοειδή υπερβολικής εφαπτομένης ως συνάρτηση ενεργοποίησης, ενώ ο νευρώνας εξόδου χρησιμοποιεί τη λογιστική σιγμοειδή. Η ανωτέρω αρχιτεκτονική είναι προϊόν εξαντλητικού trial & error.

IV.5. Το δίλημμα εξερεύνησης/εκμετάλλευσης και ο μηχανισμός εξερεύνησης

Κάθε φορά, ο πράκτοράς μας καλείται να αποφασίσει αν θα εκμεταλλευτεί τη γνώση που έχει συγκεντρώσει ως τώρα (και άρα θα επιλέξει τη δράση που αντιστοιχεί στη μικρότερη έξοδο του MLP), ή αν θα αναζητήσει μια καλύτερη λύση. Η γνώση του πράκτορα πηγάζει από μεταβάσεις, πράγματα που ο ίδιος έχει δοκιμάσει στο παρελθόν. Δεν μπορεί να γνωρίζει τη βελτιστότητα μιας συγκεκριμένης δράσης σε μια συγκεκριμένη κατάσταση, αν ποτέ δεν επιλέξει αυτή τη δράση. Αυτό είναι το δίλημμα εξερεύνησης/εκμετάλλευσης, και αποτελεί θεμελιώδες πρόβλημα της ενισχυτικής μάθησης.

Σε αυτή την εργασία αντιμετωπίζουμε το δίλημμα ορίζοντας ϵ -άπληστη επιλογή δράσεων. Κάθε φορά ο πράκτορας εκμεταλλεύεται τη γνώση του με πιθανότητα $(1-\epsilon)$. Στις υπόλοιπες περιπτώσεις δρα τυχαία.

...όμως αυτό δεν είναι αρκετό. Πώς θα καθορίσουμε την τιμή του ϵ ; Μπορεί σε συγκεκριμένα τμήματα του χώρου καταστάσεων-δράσεων ο πράκτορας να έχει κατορθώσει να προσεγγίσει εξαιρετικά καλά τη συνάρτηση Q , και συνεπώς να μην έχει ανάγκη περεταίρω εξερεύνηση. Αποφασίσαμε λοιπόν να υιοθετήσουμε τον μηχανισμό εξερεύνησης που περιγράφεται στο [26].

Πιο αναλυτικά, ο μηχανισμός αυτός ονομάζεται Self-Regulating Action Exploration (SRE), και ρυθμίζει το ϵ βάσει ποσοστών επιτυχίας που υπολογίζονται ανά τακτά χρονικά διαστήματα ως εξής:

$$\epsilon^{\text{new}} = f(1 - \lambda) \left\{ \kappa(1 - \lambda) + (1 - \kappa)\epsilon^{\text{old}} \right\}$$

$$f(x) = \begin{cases} 1 & \text{when } x > 0 \\ 0 & \text{when } x \leq 0 \end{cases}$$

Το κ είναι ο ρυθμός μάθησης του ϵ , ενώ το λ είναι το ποσοστό επιτυχίας που προαναφέραμε. Η ρύθμιση χωρίζεται σε δύο στάδια, αυτό του “ανοιχτού παραθύρου”, εντός του οποίου μετράται το λ , και “κλειστού παραθύρου”, που αποτελεί ουσιαστικά την αναπροσαρμογή του ϵ βάσει του παραπάνω τύπου.

Όταν το παράθυρο είναι ανοιχτό (και άρα το λ δεν έχει καθοριστεί πλήρως), το ϵ ρυθμίζεται από την ακόλουθη σχέση:

$$\epsilon^{new} = \epsilon^{old} - \theta$$

$$\theta = \frac{\epsilon_0}{N_\delta N_0}$$

ϵ_0 είναι η αρχικοποίηση του ϵ

N_0 είναι μια αρχική εκτίμηση του αριθμού των φορών που θα χρειαστεί να ανανεωθεί το ϵ μέχρι τη σύγκλιση του πράκτορα σε μια πολιτική που γνωρίζει καλά.

N_δ είναι η μερίδα εκείνη του τελικού αριθμού επαναλήψεων μέχρι τη σύγκλιση, στην οποία ο πράκτορας δεν εξερευνά καθόλου.

Ο αλγόριθμος δεν γνωρίζει εξ αρχής πόσες επαναλήψεις θα χρειαστούν μέχρι ο πράκτορας να εκπαιδευτεί καλά. Συνεπώς, σε κάθε κλείσιμο του παραθύρου, όπου έχουμε στα χέρια μας τις νέες τιμές για τα λ και ϵ , ο αριθμός αυτός των επαναλήψεων ανανεώνεται βάσει της σχέσης:

$$N = N_e + N_r$$

$$N_r = \frac{\epsilon^{new}}{N_\delta \theta} - \left(\frac{\epsilon^{new}}{N_\delta \theta} \bmod N_w \right) + N_w$$

N_e είναι ο αριθμός των επαναλήψεων που έχουν προηγηθεί ήδη.

N_w είναι το μήκος του παραθύρου στο οποίο μετράμε το λ .

Φυσικά, δεν έχουμε θίξει ακόμα τον ορισμό του λ , ο οποίος δεν είναι άλλος από τον αριθμό “θετικών εκβάσεων” προς το μήκος του παραθύρου. Στη δημοσίευση του [26] δεν γίνεται λόγος για τη φύση αυτών των θετικών εκβάσεων.

Πήραμε την πρωτοβουλία να συσχετίσουμε τις εκβάσεις αυτές με την απόλυτη τιμή του σφάλματος χρονικής διαφοράς (Temporal Difference—TD error):

$$TD = c(s, a, s') + \gamma \cdot \min_a Q(s', a) - Q(s, a)$$

Το σφάλμα αυτό εκφράζει την ποιότητα της πρόβλεψης του νευρωνικού δικτύου για την τιμή της συνάρτησης Q όσον αφορά το ζεύγος κατάστασης-δράσης (s, a) . Υπενθυμίζουμε πως οι πρώτοι δύο όροι του ορισμού συνθέτουν τους στόχους μάθησης όπως ορίζονται από τον αλγόριθμο NFQ.

Η πρωτοβουλία αυτή εμπνεύστηκε από τη δουλειά του [27]. Εκεί επιχειρηματολογείται πως το απόλυτο σφάλμα χρονικής διαφοράς αποτελεί έναν ικανοποιητικό δείκτη της ελεγχιμότητας ενός ζεύγους κατάστασης-δράσης. Με άλλα λόγια, όσο μικρότερο είναι, τόσο καλύτερα γνωρίζει ο πράκτορας το συγκεκριμένο ζεύγος.

Μια έκβαση λοιπόν στην περίπτωσή μας θεωρείται θετική αν το απόλυτο σφάλμα χρονικής διαφοράς όσον αφορά την τελευταία δράση του πράκτορά μας είναι μικρότερο του 15%. Το όριο αυτό είναι αρκετά μικρό ώστε να σηματοδοτεί μια αξιοπρεπή πρόβλεψη, και επίσης αρκετά ελαστικό προκειμένου να ενθαρρύνει τα αρχικά στάδια της μάθησης, στα οποία ο πράκτορας δεν έχει συγκεντρώσει αρκετή πληροφορία. Ενδεικτικά αναφέρουμε πως στο πέρας μίας προσομοίωσης λειτουργίας του θερμοστάτη για 3 μήνες, το μεγαλύτερο μέσο τετραγωνικό απόλυτο σφάλμα χρονικής διαφοράς που συγκεντρώνει ο πράκτοράς μας είναι 3.2%.

Επίσης για τις υπόλοιπες παραμέτρους του μηχανισμού εξερεύνησης έχουμε επιλέξει τις ακόλουθες τιμές:

$$\begin{aligned} \epsilon_0 &= 1 \\ N_0 &= 2 \cdot (\text{schedule length}) \\ N_\delta &= 0.25 \\ N_w &= (\text{schedule length}) \end{aligned}$$

IV.6. Ανυπάκουα σημεία λειτουργίας

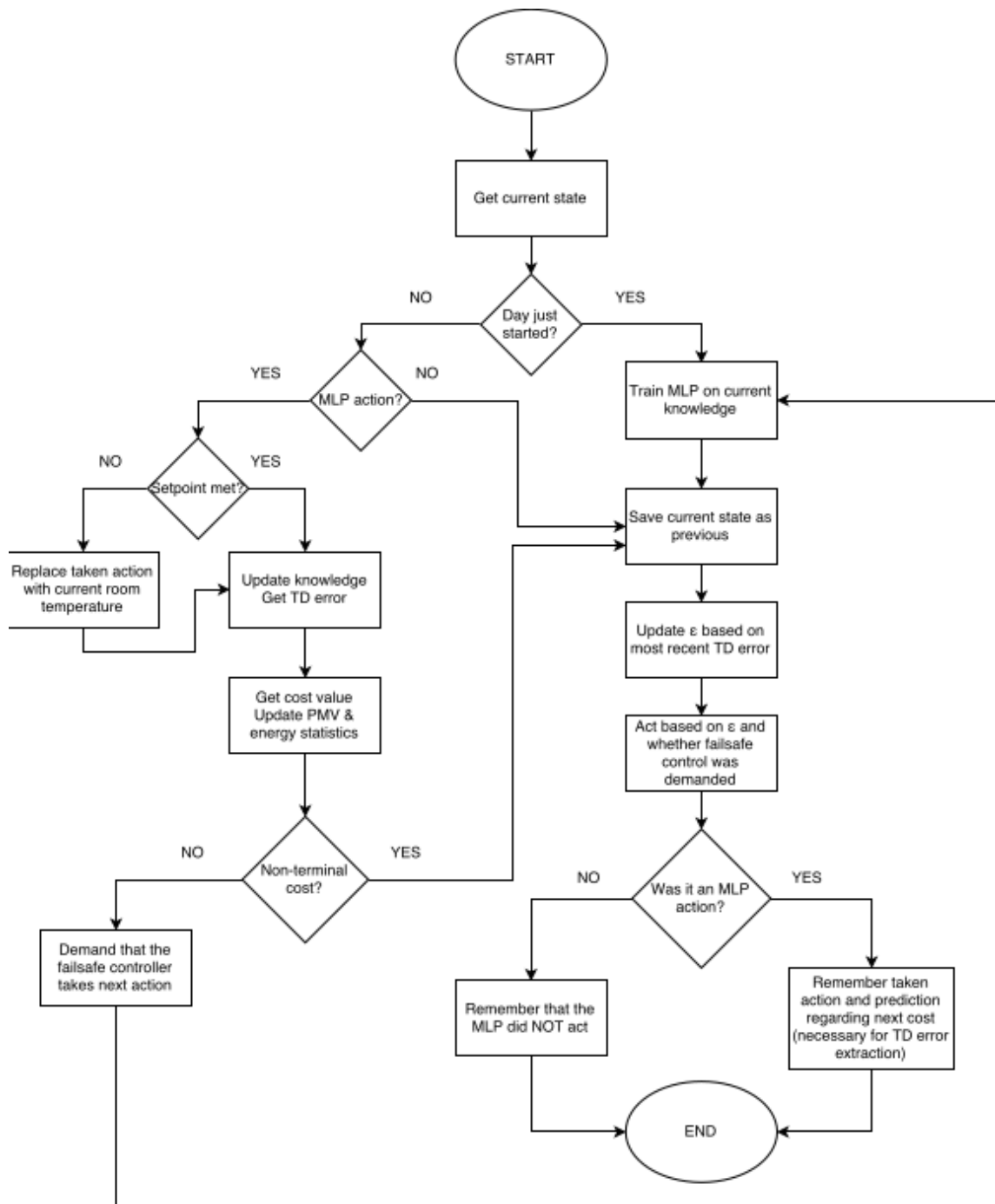
Κατά τα αρχικά στάδια ανάπτυξης του μηχανισμού αποφάσεων που παρουσιάζεται σε αυτή την εργασία, παρατηρήσαμε πως κάποια σημεία λειτουργίας δεν τηρούνταν από το σύστημα κλιματισμού. Με άλλα λόγια, παρατηρήθηκε σημαντικό σφάλμα (πολλές φορές μεγαλύτερο του ενός βαθμού Κελσίου) ανάμεσα στο εκάστοτε σημείο λειτουργίας, και στη μετρηθείσα θερμοκρασία του δωματίου κατά την εξαγωγή του σήματος κατάστασης στο επόμενο χρονικό βήμα. Το σφάλμα αυτό δυσχέραινε τη μάθηση σε πολύ μεγάλο βαθμό, αφού πρακτικά έλεγε ψέματα στον πράκτορα, συσχετίζοντας τα λάθος αποτελέσματα στις λάθος δράσεις.

Φροντίσαμε λοιπόν, αργότερα, τις φορές που παρατηρούνταν τέτοιο σφάλμα, η ληφθείσα δράση του πράκτορα να αντικαθίσταται με τη μετρηθείσα θερμοκρασία δωματίου μετά την—αποτυχημένη—εφαρμογή του σημείου λειτουργίας.

Αποφασίσαμε, επίσης, να κάνουμε μια διερεύνηση των λόγων για τους οποίους μπορεί να εμφανίζονται τέτοιες ανυπάκουες μεταβάσεις. Τα αποτελέσματα συμπεριλαμβάνονται στην ενότητα VI (Σχ. 6.13).

IV.7. Τελικά σχόλια και σύνοψη

Ο κύκλος παραγωγής των σημείων λειτουργίας του θερμοστάτη συνοψίζεται στο διάγραμμα ροής του Σχ. 4.3. Ο κύκλος αυτός επαναλαμβάνεται κάθε μέρα ανά δέκα λεπτά, έως το πέρας του ωραρίου. Θα μπορούσε κανείς να επιχειρηματολογήσει ενάντια στο μικρό μέγεθος του συγκεκριμένου χρονικού βήματος. Αυτή η επιλογή όμως (έναντι της πιο κλασικής των 20 λεπτών, για παράδειγμα) επιτρέπει στον πράκτορα να συσσωρεύσει περισσότερη εμπειρία σε λιγότερο χρόνο, και συνεπώς να μάθει πιο γρήγορα. Επίσης, ένα αρκούντως μικρό χρονικό βήμα εξασφαλίζει πως, σε περίπτωση λάθους του πράκτορα (δηλαδή λήψη τερματικής δράσης), οι συνέπειες του λάθους δεν θα επιβαρύνουν για πολύ τους χρήστες.



Σχήμα 4.3: Διάγραμμα ροής για τον κύκλο παραγωγής κάθε επόμενου σημείου λειτουργίας. Πρώτα λαμβάνεται η παρούσα κατάσταση του συστήματος, κατόπιν αξιολογείται η τελευταία δράση του πράκτορα, στη συνέχεια ανανεώνεται το ϵ βάσει του μηχανισμού εξερεύνησης και τέλος λαμβάνεται η επόμενη δράση. Το νευρωνικό δίκτυο επανεκπαιδεύεται στην αρχή της κάθε μέρας, και σε περίπτωση λήψης τερματικού κόστους.

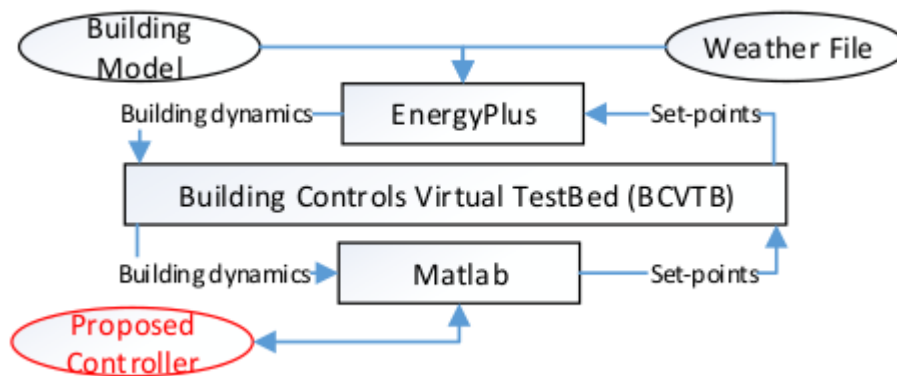
V. Προσομοίωση και αξιολόγηση

V.1. Πλαίσιο εργασίας προσομοίωσης

Το πλαίσιο εργασίας που ακολουθήσαμε για την προσομοίωση της λειτουργίας του μηχανισμού αποφάσεων είναι ευρέως δεδομένο στη βιβλιογραφία. Λεπτομερής περιγραφή του μπορεί να βρεθεί στα [28] και [29], ενώ δουλειές που το ακολουθούν είναι για παράδειγμα οι [12] και [13]. Χρησιμοποιούνται τρία περιβάλλοντα:

- Η δυναμική του κτιρίου, η επίδραση του καιρού, η ενεργειακή κατανάλωση αλλά και η θερμική άνεση προσομοιώνονται στο περιβάλλον EnergyPlus.
- Η επεξεργασία των μετρήσεων και γενικότερα ο κώδικας του ελεγκτή μας, ο οποίος παράγει σε κάθε χρονικό βήμα το επόμενο σημείο λειτουργίας, υλοποιήθηκε στη γλώσσα MATLAB.
- Η ανταλλαγή δεδομένων μεταξύ του EnergyPlus και του MATLAB υλοποιείται μέσω του Building Controls Virtual Test Bed (BCVTB).

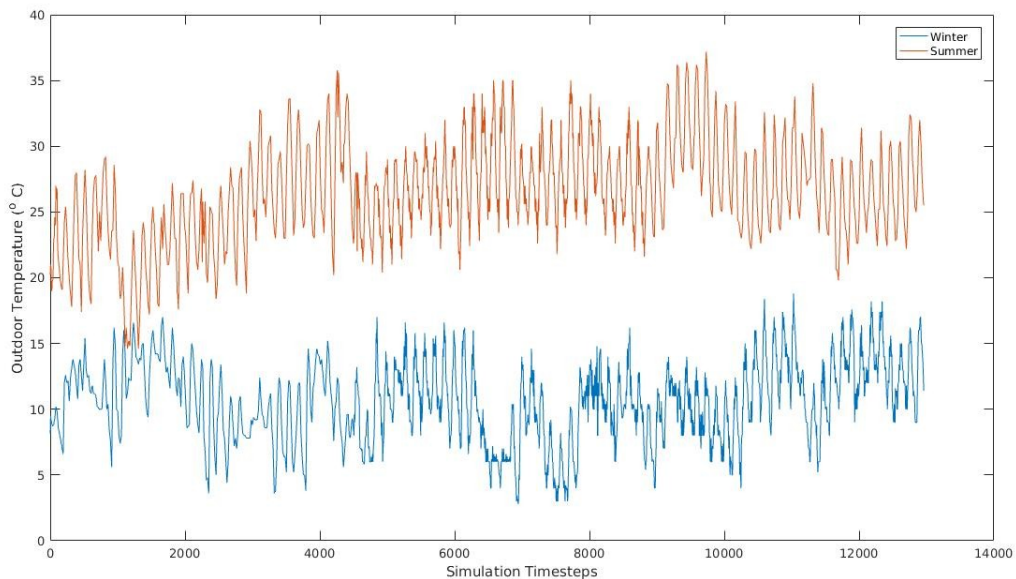
Τα παραπάνω απεικονίζονται στο Σχ. 5.1.



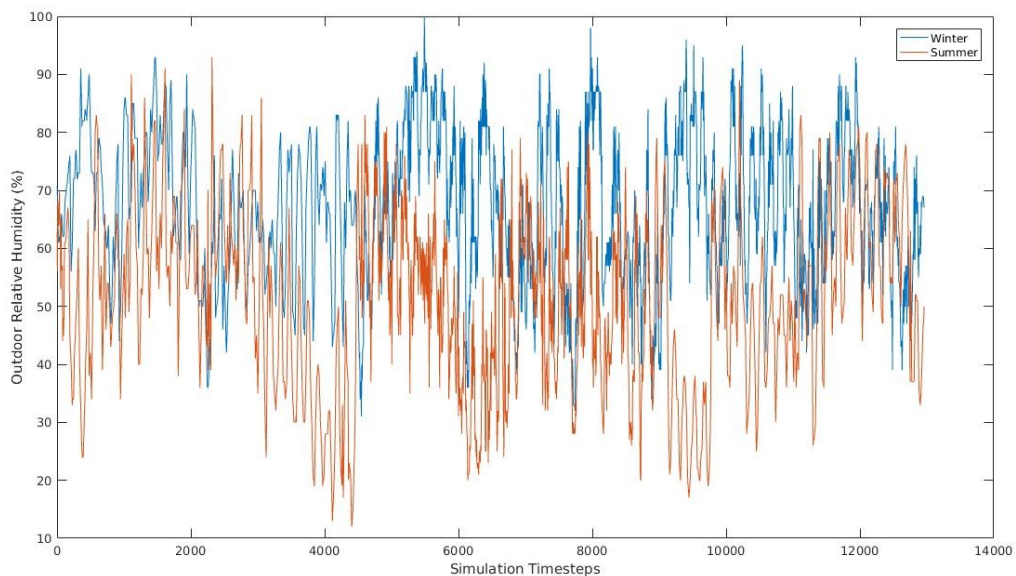
Σχήμα 5.1: Πλαίσιο εργασίας για την προσομοίωση του μηχανισμού αποφάσεων της εργασίας. Το μοντέλο του κτιρίου και το αρχείο καιρού (εξωτερική θερμοκρασία, ηλιακή ακτινοβολία κτλ.) δίνονται ως είσοδοι στο EnergyPlus. Συναρτήσει αυτών προκύπτει η κατάσταση του συστήματος, η οποία μεταβιβάζεται στο MATLAB μέσω του BCVTB. Εκεί εκτελείται ο κύκλος επεξεργασίας που περιγράφηκε στην ενότητα IV, και παράγεται το επόμενο σημείο λειτουργίας του θερμοστάτη, που πάλι μέσω του BCVTB δίνεται στο EnergyPlus, η προσομοίωση της δυναμικής του κτιρίου προχωράει στο επόμενο χρονικό βήμα, κ.ο.κ.

V.2. Καιρός και κτίριο

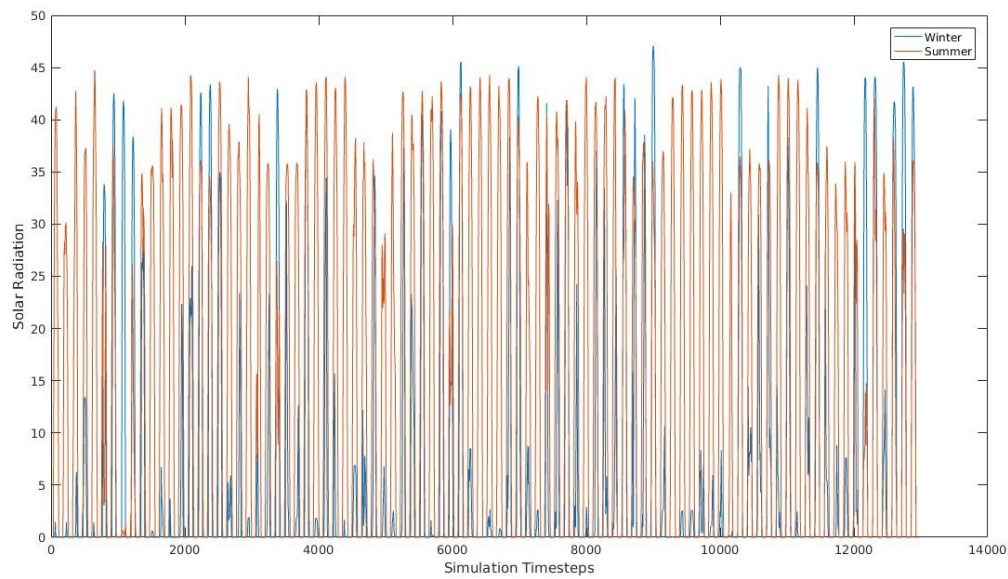
Όσον αφορά τον καιρό, χρησιμοποιήθηκαν δημοσίως προσβάσιμα δεδομένα από μετρήσεις που έλαβαν χώρα στην Αθήνα το 2010 [30]. Ενδεικτικές χρονοσειρές της εξωτερικής θερμοκρασίας, υγρασίας και ακτινοβολίας φαίνονται στα Σχ. 5.2-4.



Σχήμα 5.2: Εξωτερική θερμοκρασία για τις δύο περιόδους προσομοίωσης.

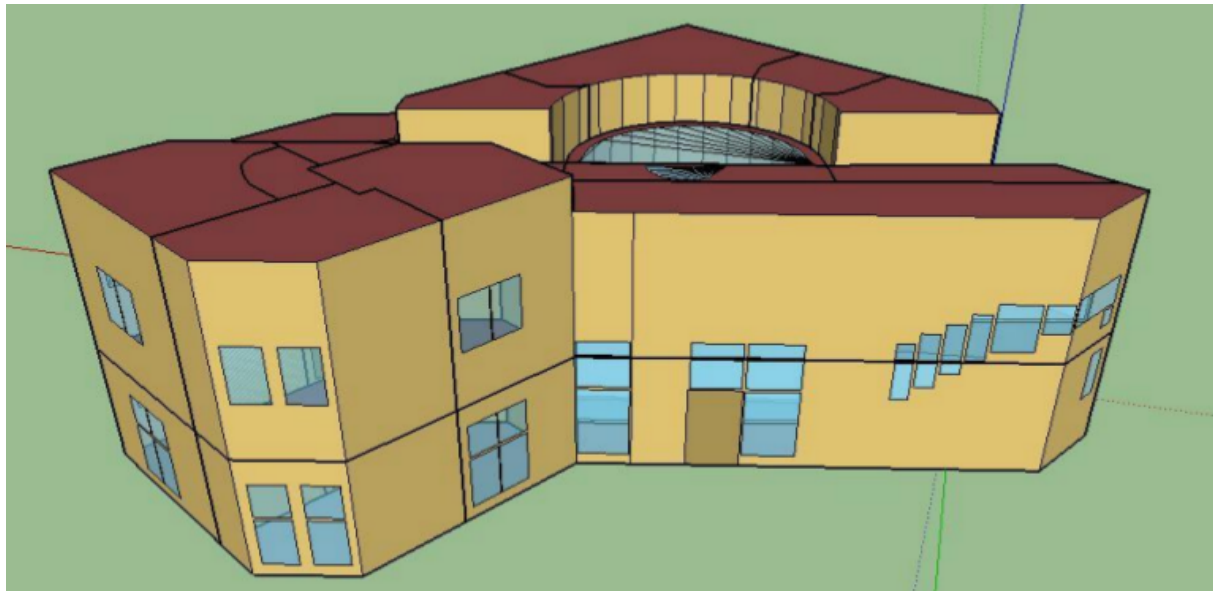


Σχήμα 5.3: Εξωτερική υγρασία για τις δύο περιόδους προσομοίωσης.



Σχήμα 5.4: Ηλιακή ακτινοβολία για τις δύο περιόδους προσομοίωσης.

Το μοντέλο του κτιρίου αφορά πραγματικό κτίριο στα Χανιά της Κρήτης με δέκα θερμικές ζώνες, και η μοντελοποίηση ήταν μέρος ευρύτερου πρότζεκτ χρηματοδοτημένου από την Ευρωπαϊκή Κομισιόν. Άποψή του απεικονίζεται στο Σχ. 5.5.

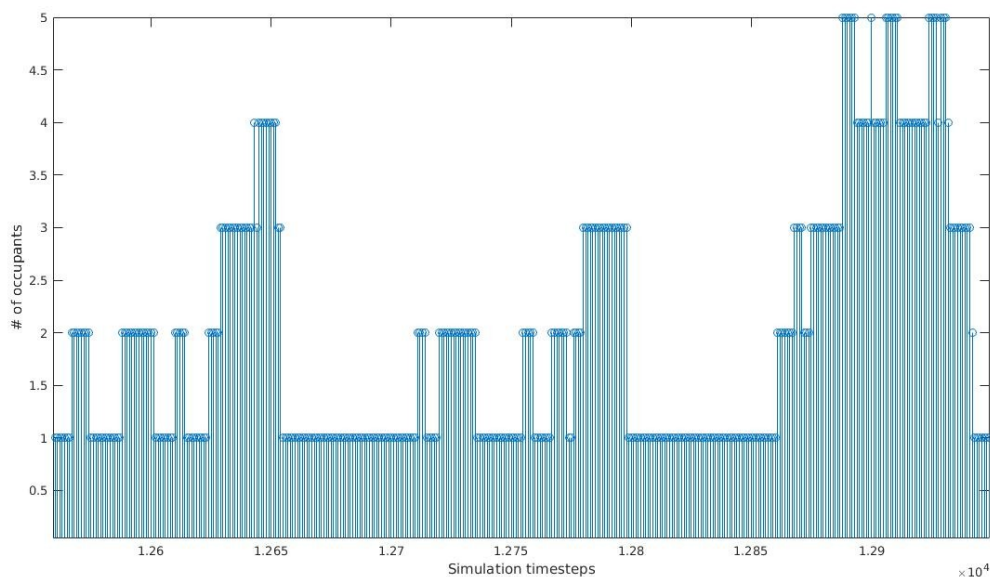


Σχήμα 5.5: Απεικόνιση του κτιρίου στο οποίο έλαβαν χώρα οι προσομοιώσεις.

V.3. Ζώνη λειτουργίας, ωράριο και άνθρωποι

Στις προσομοιώσεις του κτιρίου, τροφοδοτήσαμε με τα σημεία λειτουργίας του μηχανισμού αποφάσεων τον θερμοστάτη της δεύτερης θερμικής ζώνης. Ο θερμοστάτης λειτουργεί από τις 6 το πρωί έως και τις 9 το βράδυ, εφαρμόζοντας όπως αναφέρθηκε στην προηγούμενη ενότητα τα σημεία λειτουργίας του ανά 10 λεπτά.

Όσον αφορά τους ενοίκους της ζώνης ανά πάσα χρονική στιγμή, ο αριθμός τους κυμαίνεται από 1 έως 5 και μεταβάλλεται τυχαία. Υποθέτουμε μόνο πως κατά τις ώρες λειτουργίας υπάρχει πάντα τουλάχιστον ένα άτομο στη ζώνη. Ενδεικτικά αποτελέσματα για τρεις μέρες προσομοίωσης δείχνονται στο Σχ. 5.6.



Σχήμα 5.6: Μεταβολή αριθμού ενοίκων θερμικής ζώνης για τρεις διαδοχικές μέρες προσομοίωσης. Τα ενδιάμεσα διαστήματα όπου ο αριθμός παραμένει σταθερός και ίσος με 1 αντιστοιχούν σε ώρες μη-λειτουργίας του θερμοστάτη, και συνεπώς αγνοούνται.

V.4. Το πρόβλημα του ενδιάμεσου χρονικού βήματος

Όπως αναφέρθηκε στην αρχή του κεφαλαίου, η επικοινωνία του ελεγκτή μας με το κτίριο—δηλαδή η ανταλλαγή δεδομένων—επιτυγχάνεται μέσω του BCVTB. Μια ιδιαιτερότητα του BCVTB είναι ότι, προκειμένου να εξασφαλιστεί πως η προσομοίωση δεν θα υποπέσει σε αλγεβρικούς βρόχους, εισάγει καθυστέρηση ενός χρονικού βήματος ανάμεσα στην εφαρμογή ενός σημείου λειτουργίας και στα αποτελέσματά του. Με άλλα λόγια, οι αποφάσεις που παίρνει στο MATLAB ο μηχανισμός αποφάσεων εφαρμόζονται δύο χρονικά βήματα

αργότερα. Αυτό δημιουργεί σοβαρό πρόβλημα για την ποιότητα του ελέγχου, αφού πραγματοποιείται online, και συνεπώς χρειάζεται τα αντίστοιχα δεδομένα στο αμέσως επόμενο χρονικό βήμα.

Η λύση που ακολουθήσαμε είναι η εξής: ξεκινώντας από την αρχή του ωραρίου κάθε μέρας, ο κύκλος παραγωγής σημείων λειτουργίας εκτελείται **ανά δύο** χρονικά βήματα. Στο χρονικό βήμα που μεσολαβεί, ως σημείο λειτουργίας δίνεται το ίδιο με αυτό που αποφασίστηκε ακριβώς πριν. Την επόμενη φορά που ο ελεγκτής θα λάβει το σήμα κατάστασης, αυτό θα περιέχει πράγματι τα σωστά αποτελέσματα.

V.5. Αξιολόγηση επίδοσης

Η αξιολόγηση του πράκτορα αυτής της εργασίας μπορεί να διαχωριστεί σε δύο τμήματα: ένα που αφορά τη μάθηση, και ένα δεύτερο για την επίδοση του θερμοστάτη.

Όσον αφορά τη μάθηση, εποπτεύουμε για κάθε ένα από τα πειράματα που παρουσιάζουμε γραφικές απεικονίσεις του λόγου θετικών μεταβάσεων λ και της πιθανότητας εξερεύνησης ϵ , όπως αυτή διαμορφώνεται από το πρώτο. Δίνονται επίσης ιστογράμματα για τις δράσεις που λήφθηκαν και τα άμεσα κόστη που προέκυψαν. Τέλος, προσφέρονται μετρήσεις του μέσου τετραγωνικού απόλυτου σφάλματος χρονικής διαφοράς, καθώς και του αριθμού των φορών που το νευρωνικό δίκτυο χρειάστηκε να επανεκπαιδευτεί. Όλα αυτά δείχνουν:

- την ικανότητα του πράκτορα πράγματι να *μαθαίνει* (και άρα να διατηρεί υψηλό λόγο θετικών μεταβάσεων, με την εξαίρεση περιόδων όπου συναντά πρωτόγνωρες συνθήκες. Επίσης, να επιδεικνύει χαμηλό μέσο τετραγωνικό απόλυτο σφάλμα χρονικής διαφοράς)
- τον πολυμορφισμό στη συμπεριφορά του πράκτορα ανάλογα με την τιμή του tradeoff (δηλαδή τη λήψη διαφορετικών δράσεων για διαφορετικές τιμές)
- την αποδοτικότητα ως προς τα δεδομένα (η οποία συσχετίζεται άμεσα με τον αριθμό των φορών επανεκπαίδευσης του δικτύου)

Για την επίδοση του θερμοστάτη, τώρα, συγκρίνουμε την ενεργειακή κατανάλωση και τη θερμική άνεση όπως εκφράζεται από την κλίμακα PMV με μια ευρεία γκάμα ελέγχων σταθερού σημείου λειτουργίας (RBC). Οι περισσότεροι χειροκίνητοι θερμοστάτες τείνουν να λειτουργούν σε ένα σταθερό σημείο λειτουργίας το χειμώνα, και ένα αντίστοιχο το καλοκαίρι [31]. Παρόμοια, ένας έξυπνος θερμοστάτης θα παρήγαγε σημεία λειτουργίας σε ένα λογικό όσον

αφορά τη θερμική άνεση εύρος (στην περίπτωσή μας, από 20 έως 27 βαθμούς Κελσίου). Διαβαθμίσεις σε αυτά τα σημεία φυσικά και υπάρχουν, λόγω άλλων παραγόντων που επηρεάζουν κατανάλωση και άνεση (όπως για παράδειγμα ανοιχτά παράθυρα, μαγείρεμα κτλ.). Και πάλι, όμως, οι διαβαθμίσεις αυτές δεν θα δημιουργούσαν κάτι διαφορετικό από μια τροχιά μεταξύ των σημείων λειτουργίας που επιλέξαμε. Συμπεριλαμβάνοντας, λοιπόν, όσο περισσότερα γίνεται, προκύπτει μια ουσιώδης αξιολόγηση έναντι στον συνήθη χειροκίνητο ή έξυπνο έλεγχο.

VI. Πειραμαματικά αποτελέσματα

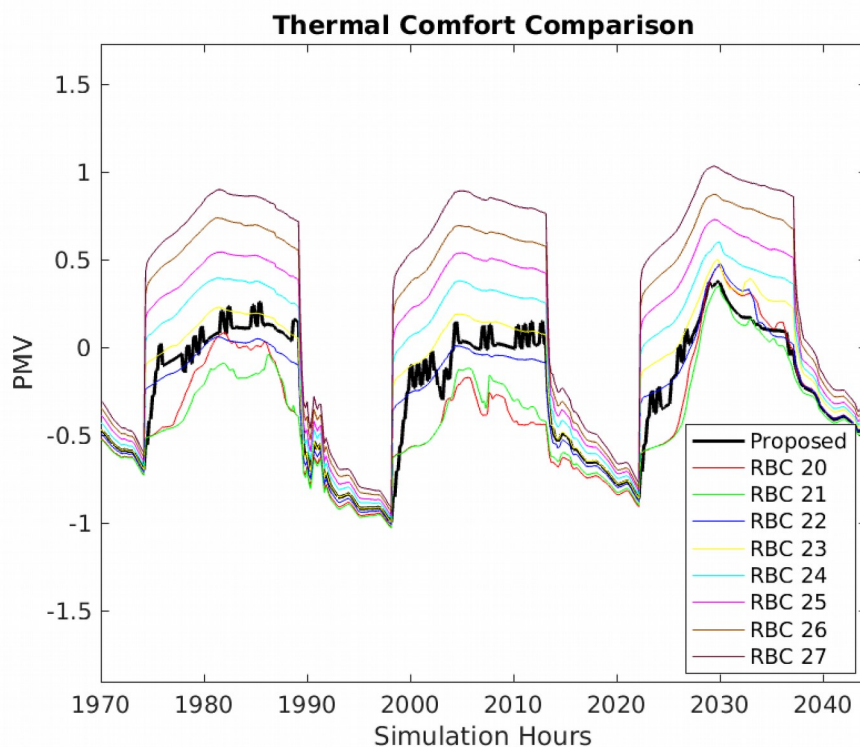
Τα αποτελέσματα που ακολουθούν αφορούν δύο περιόδους, διάρκειας 90 ημερών η κάθε μία.

Η πρώτη περίοδος είναι χειμωνιάτικη, από 01 Ιανουαρίου έως 31 Μαρτίου. Η δεύτερη περίοδος είναι καλοκαιρινή, από 01 Ιουνίου έως 31 Αυγούστου.

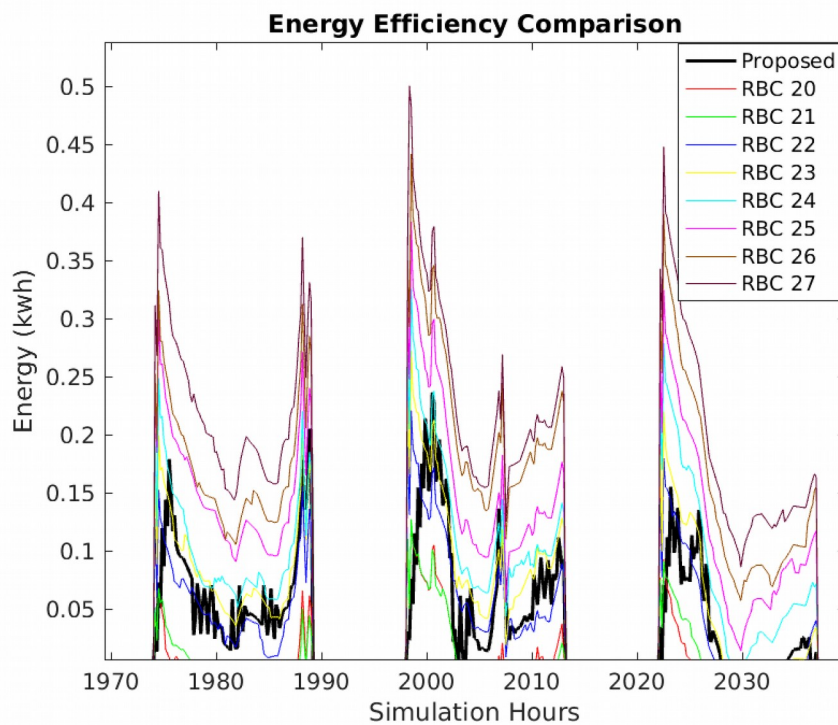
Για κάθε περίοδο, τροφοδοτήσαμε το θερμοστάτη μας από τις θερμικές ζώνες του κτιρίου με τα σημεία λειτουργίας του μηχανισμού αποφάσεων που περιγράφηκε στην ενότητα IV. Δώσαμε τρεις διαφορετικές τιμές στο tr , την τιμή δηλαδή που καθορίζει την έμφαση στην ελαχιστοποίηση της ενεργειακής κατανάλωσης: 0, 0.5 και 1.

Ακολουθούν τα αποτελέσματα που προέκυψαν.

VI.1. Χειμερινή περίοδος ($tr = 0$)



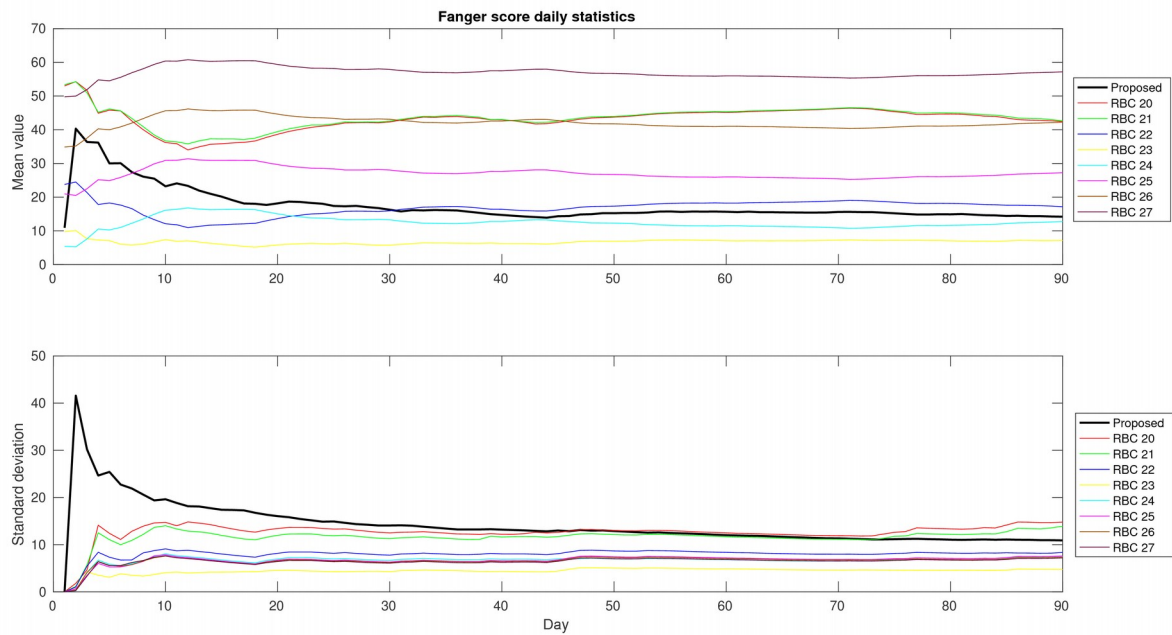
Σχήμα 6.1: Σύγκριση θερμικής άνεσης για τρεις διαδοχικές χειμερινές ημέρες (χειμώνας, $tr = 0$).



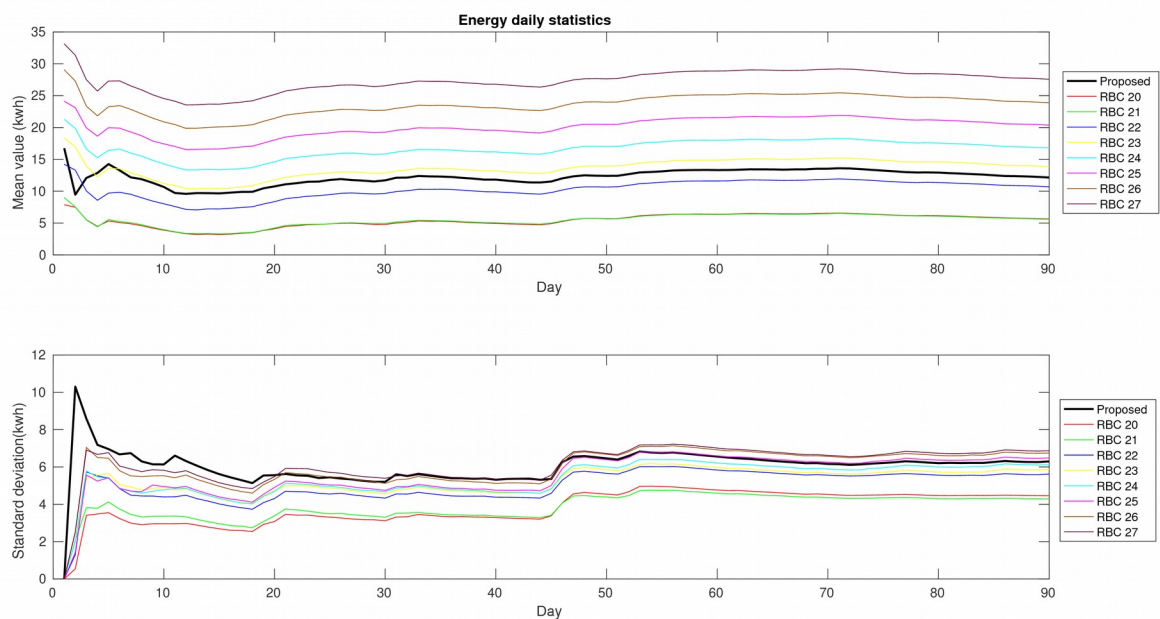
Σχήμα 6.2: Σύγκριση ενεργειακής κατανάλωσης για τρεις διαδοχικές χειμερινές ημέρες (χειμώνας, $t_r = 0$).

Ήδη από τα παραπάνω δύο σχήματα, δύο σημαντικές παρατηρήσεις μπορούν να γίνουν:

- ο θερμοστάτης φαίνεται να ακολουθεί μια πολιτική που κρατάει τη θερμική άνεση εξαιρετικά κοντά στο ιδανικό. Αυτό φαίνεται περισσότερο τις δύο πρώτες μέρες, οι οποίες είναι πιο κρύες από την τρίτη, που ενδεχομένως αντιστοιχεί σε πρωτόγνωρες για τον πράκτορα καιρικές συνθήκες.
- ο πράκτορας έχει μάθει, κατά την πρώτη ώρα λειτουργίας του, να ζεσταίνει διαρκώς το δωμάτιο μέχρι μια θερμοκρασία, γύρω από την οποία κατόπιν ταλαντώνεται. Αυτό αφ' ενός τον καθυστερεί σε σχέση με τα RBC όσον αφορά την επίτευξη της βέλτιστης θερμικής άνεσης, αφ' ετέρου τον κάνει πολύ πιο αποδοτικό ενεργειακά ακόμα και τώρα, που η κατανάλωση απουσιάζει πλήρως, λόγω της τιμής του t_r , από τη συνάρτηση κόστους (ο αναγνώστης μπορεί να παρατηρήσει τα πελώρια spikes στην αρχή της λειτουργίας των RBC κατά την αρχή της μέρας).

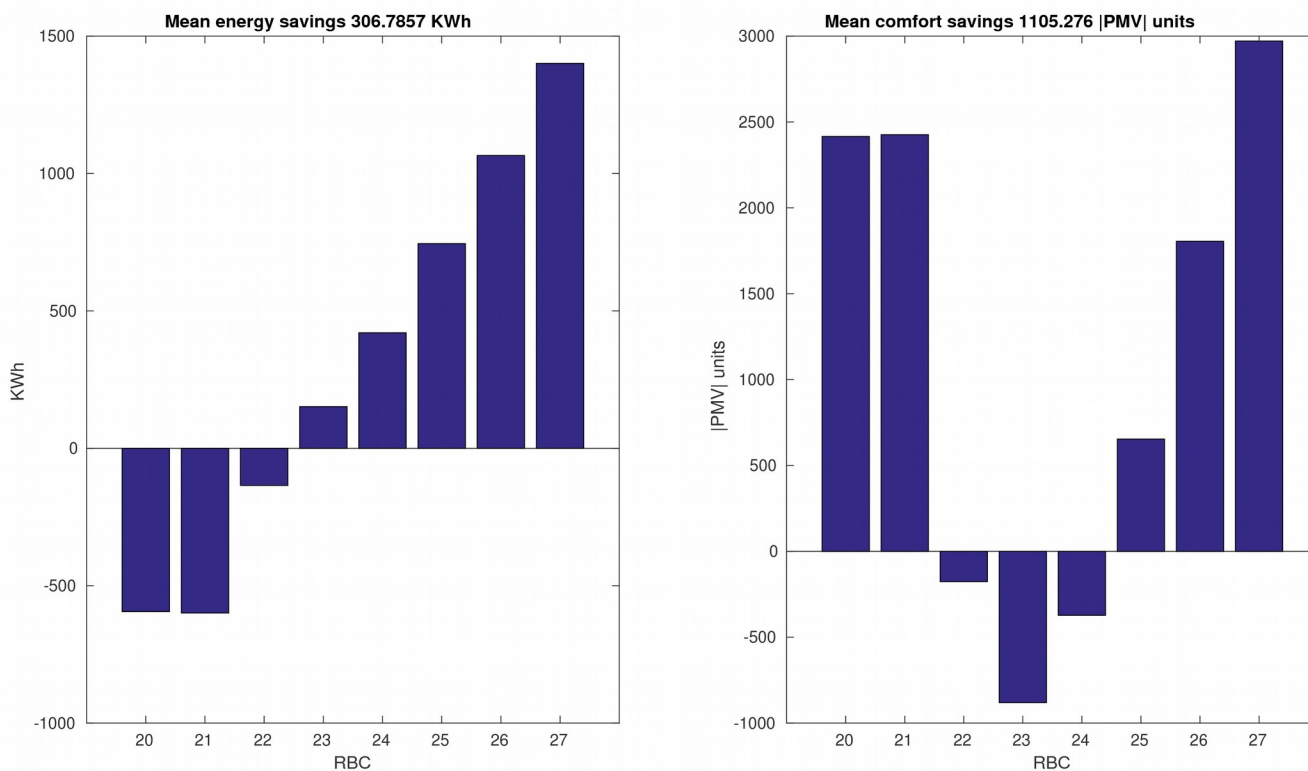


Σχήμα 6.3: Μεταβολή μέσης ημερήσιας θερμικής άνεσης, και της αντίστοιχης τυπικής απόκλισης, με την πάροδο του χρόνου (χειμώνας, $tr = 0$).



Σχήμα 6.4: Μεταβολή μέσης ημερήσιας ενεργειακής κατανάλωσης, και της αντίστοιχης τυπικής απόκλισης, με την πάροδο του χρόνου (χειμώνας, $tr = 0$).

Τα παραπάνω ποσοστά είναι ημερήσια, και ασφαλώς δεν κολακεύουν τον πράκτορά μας. Έχοντας κατά νου όμως το Σχ. 6.1, καταλαβαίνουμε πως η ημερήσια επίδοση του θερμοστάτη ως προς τη θερμική άνεση φαίνεται χειρότερη λόγω της πολιτικής σταδιακής αύξησης της θερμοκρασίας κατά την πρώτη ώρα. Είναι εμφανές πως πέρα από αυτή την ώρα, ο πράκτορας ανταγωνίζεται με άνεση τα RBC 23 και 24. Τονίζουμε επίσης την ενεργειακή του ανωτερότητα όσον αφορά τα τελευταία, η οποία φαίνεται καθαρά στο Σχ. 6.4. Τέλος, παρατηρούμε την φθίνουσα τάση τόσο της μέσης θερμικής άνεσης όσο και της αντίστοιχης τυπικής απόκλισης, η οποία υποδηλώνει διαρκή μάθηση και βελτίωση συμπεριφοράς.

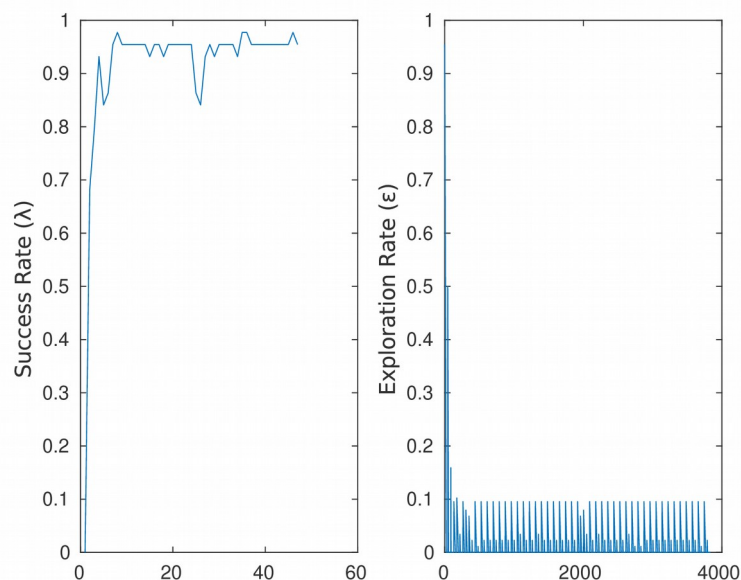


Σχήμα 6.5: Γραφήματα εξοικονόμησης ενέργειας-άνεσης (χειμώνας, $tr = 0$).

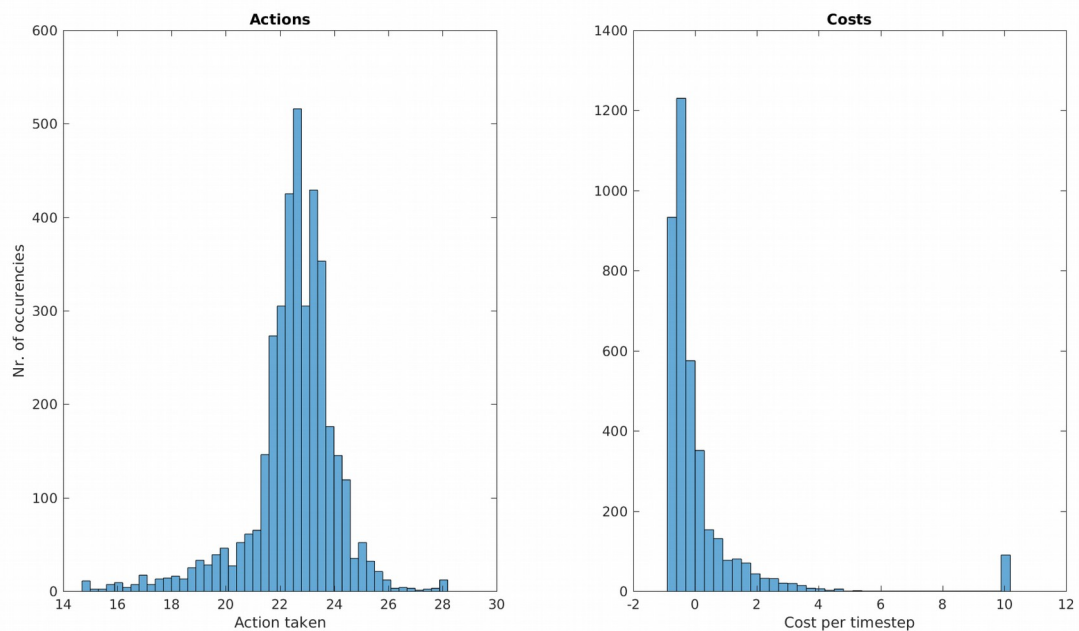
Το ανωτέρω γράφημα δικαιώνει πλήρως τον πράκτορά μας, αφού τονίζει την—ενδογενή—ενεργειακή του εξοικονόμηση, αλλά δείχνει επίσης πως το μόνο RBC το οποίο φαίνεται καλύτερό του είναι αυτό των 22 βαθμών. Στο Σχ. 6.3 όμως είναι προφανές πως η ανωτερότητα αυτή (όσον αφορά την άνεση) είναι πλασματική, και οφείλεται στην μεγάλη διαφορά που επιδεικνύει το

συγκεκριμένο RBC τον πρώτο μήνα, από το πέρας του οποίου και έπειτα ο πράκτορας συμπεριφέρεται καλύτερα.

Εξετάζουμε στη συνέχεια κάποια γραφήματα σχετικά με τη μάθηση του ελεγκτή.



Σχήμα 6.6: Λόγος θετικών μεταβάσεων και πιθανότητα εξερεύνησης (χειμώνας, $tr = 0$).

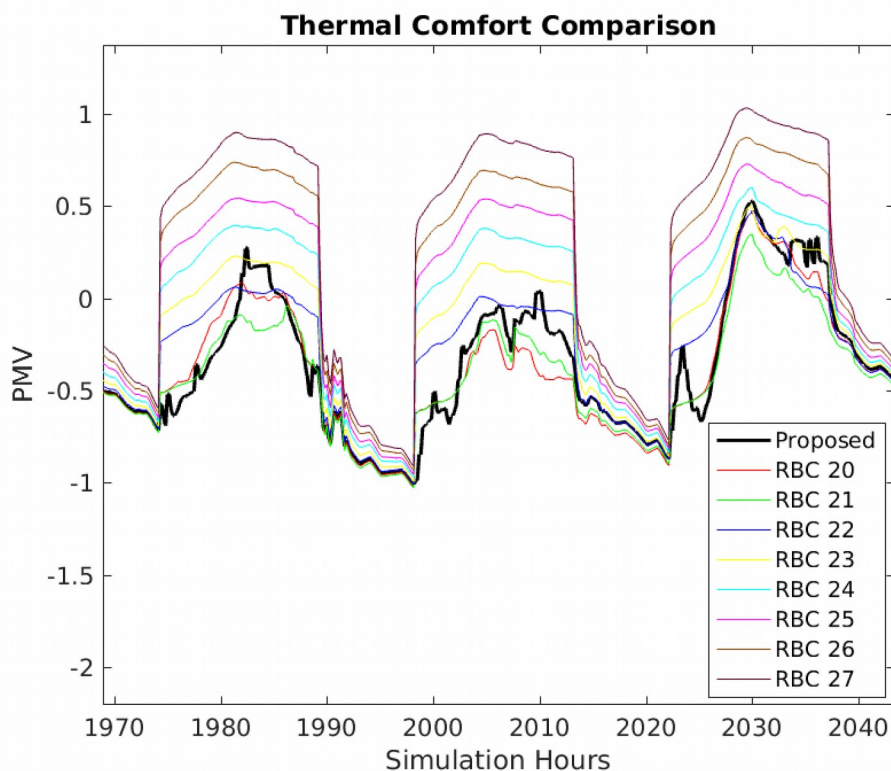


Σχήμα 6.7: Ιστογράμματα δράσεων-κόστους (χειμώνας, $tr = 0$).

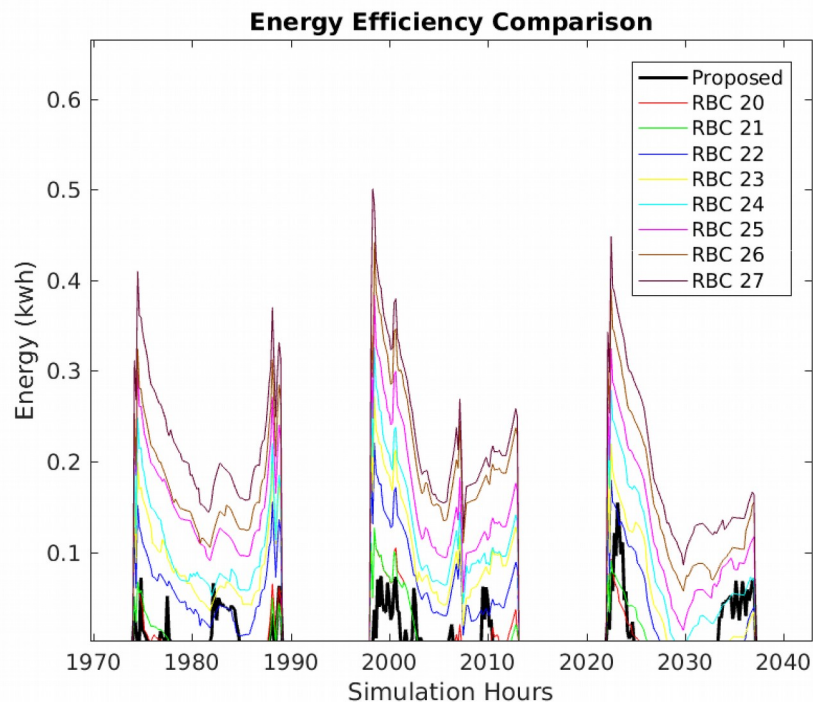
Ο πράκτορας φαίνεται να μαθαίνει πολύ καλά, διατηρώντας από νωρίς έναν υψηλό λόγο θετικών μεταβάσεων και συγκεντρώνοντας επί το πλείστον αρνητικά άμεσα κόστη. Όσον αφορά τις δράσεις του, και κατά πόσο διαφοροποιούνται καθώς μεταβάλλεται το tradeoff, θα το διαπιστώσουμε στη συνέχεια. Τονίζουμε πάντως πως η συντριπτική πλειοψηφία των δράσεων του συσσωρεύεται στην περιοχή των 22 έως 24 βαθμών, που παραπάνω είδαμε πως είναι τα κυρίαρχα RBC ως προς τη θερμική άνεση.

VI.2. Χειμερινή περίοδος ($tr = 1$)

Εστιάζουμε τώρα στην έταιρη ακραία περίπτωση, αυτής της αποκλειστικής έμφασης στην ελαχιστοποίηση της ενεργειακής κατανάλωσης. Αντίστοιχα σχήματα των παραπάνω ακολουθούν. Τονίζεται πως, για ουσιώδη σύγκριση, όσον αφορά την πρώτη δυάδα σχημάτων οι ίδιες τρεις μέρες απεικονίζονται και εδώ. Η ίδια τακτική ακολουθείται και στις καλοκαιρινές προσομοιώσεις παρακάτω.



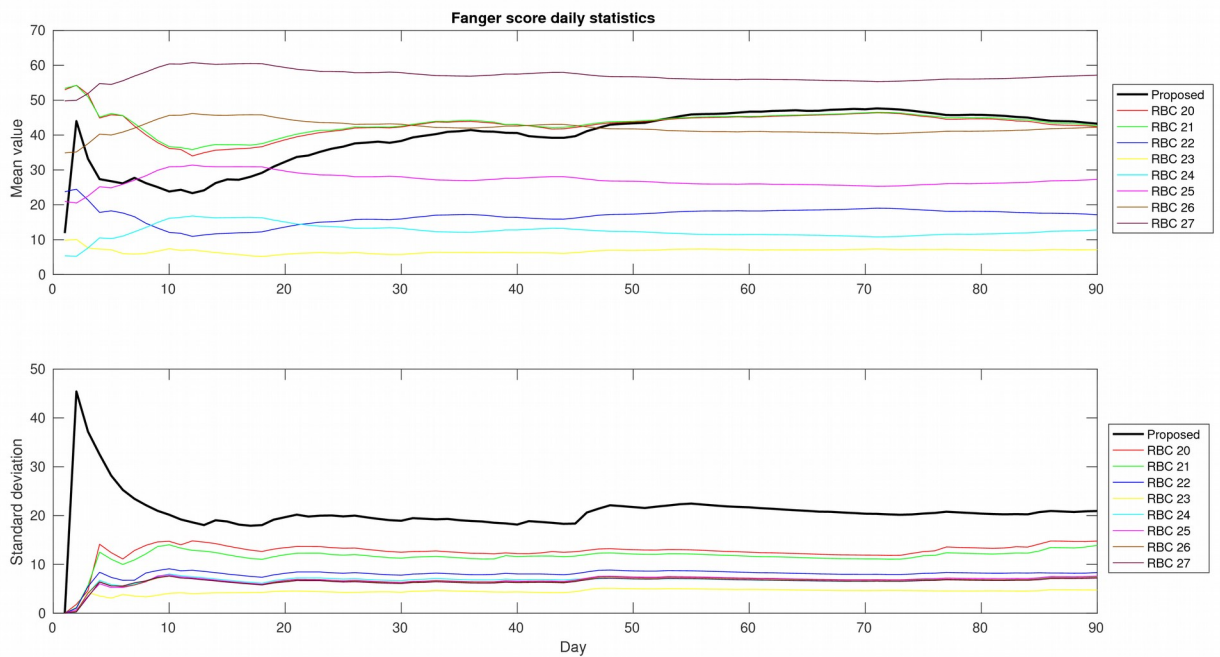
Σχήμα 6.8: Σύγκριση θερμικής άνεσης για τρεις διαδοχικές χειμερινές ημέρες (χειμώνας, $tr = 1$).



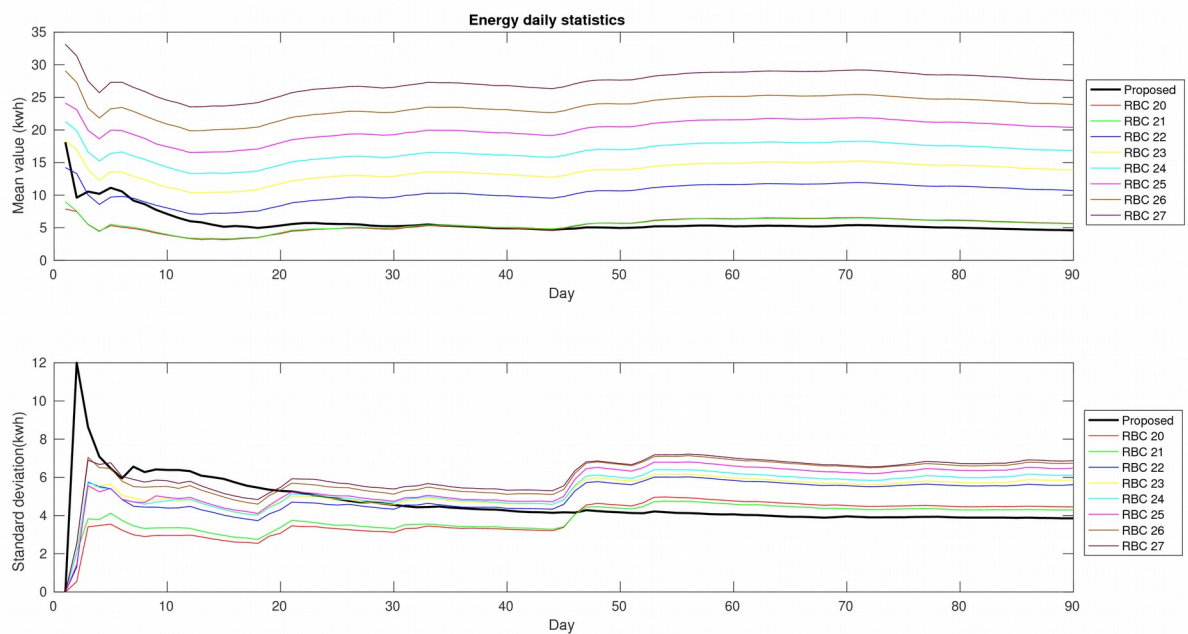
Σχήμα 6.9: Σύγκριση ενεργειακής κατανάλωσης για τρεις διαδοχικές χειμερινές ημέρες (χειμώνας, $tr = 1$).

Είναι ήδη προφανές πως η μορφή του ελέγχου έχει αλλάξει σημαντικά. Ο θερμοστάτης πλέον ξοδεύει ελάχιστο έως καθόλου χρόνο προκειμένου να ζεστάνει το δωμάτιο. Για παράδειγμα, τη δεύτερη μέρα, αυξάνει τη θερμοκρασία ίσα-ίσα προκειμένου το PMV να παραμείνει στη ζώνη λειτουργίας. Την περισσότερη ώρα καταναλώνεται μηδενική ενέργεια. Αναφέρουμε μια εξαιρετικά πιθανή εξήγηση στο τέλος της παρούσας ενότητας.

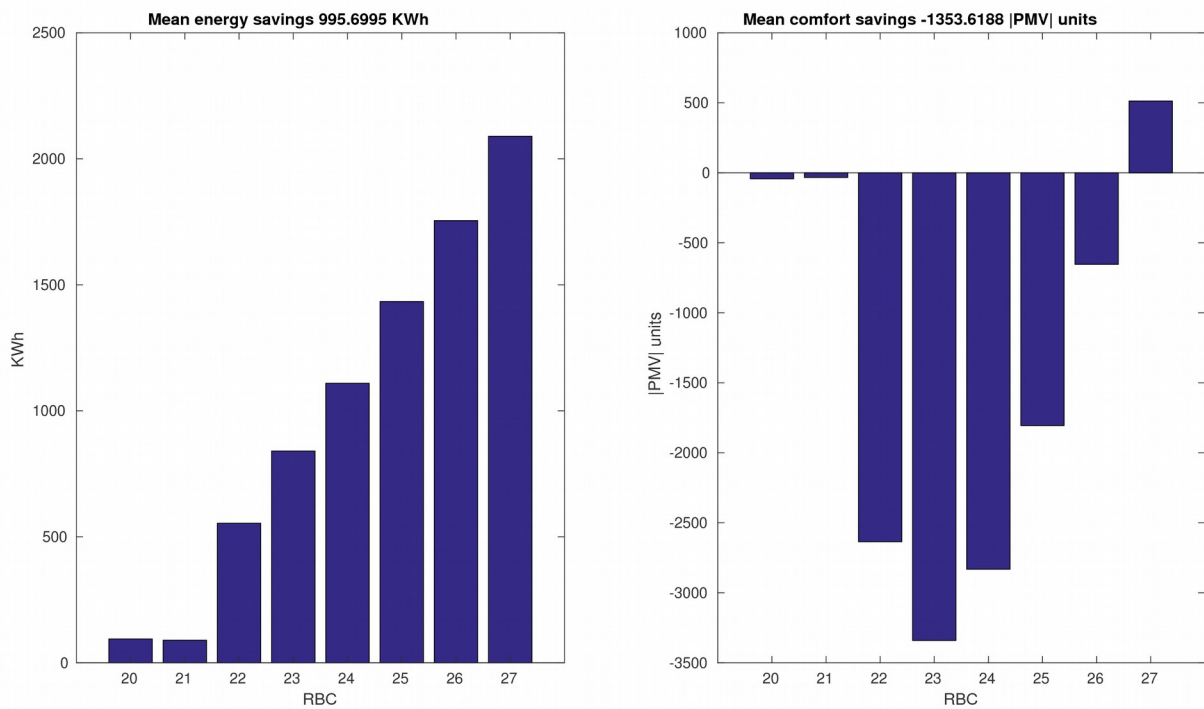
Στα Σχ. 6.10-11 τα επιχειρήματα υπέρ της αποτελεσματικότητας του πράκτορά μας πληθαίνουν σημαντικά. Η θερμική άνεση είναι, όπως και αναμέναμε, κακή, αφού ο θερμοστάτης εστιάζει σε ψυχρές θερμοκρασίες (πιο κοντινές με την εξωτερική) προκειμένου να οικονομήσει ενεργειακά. Η κατανάλωση, από την άλλη, δεν παύει να βελτιώνεται, τόσο από την άποψη της μέσης τιμής όσο και ως προς την τυπική της απόκλιση. Παίρνει περίπου ενάμισο μήνα στον πράκτορά μας να υπερνικήσει όλα τα RBC. Το Σχ. 6.12 επιβεβαιώνει πανηγυρικά τα παραπάνω, με τελική μέση εξοικονόμηση σχεδόν 1000 kWh. Σύμφωνα με στατιστικές μελέτες της Ευρωπαϊκής Ένωσης [32], για ένα ελληνικό νοικοκυριό στο δεύτερο μισό του 2015 αυτό θα μεταφραζόταν σε 180 ευρώ λιγότερα στο λογαριασμό τη ΔΕΗ.



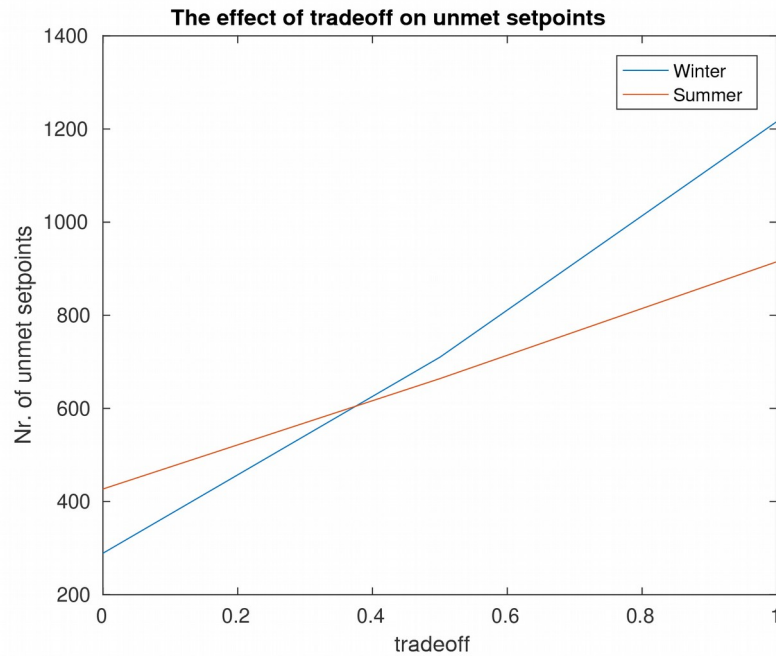
Σχήμα 6.10: Μεταβολή μέσης ημερήσιας θερμικής άνεσης, και της αντίστοιχης τυπικής απόκλισης, με την πάροδο του χρόνου (χειμώνας, $tr = 1$).



Σχήμα 6.11: Μεταβολή μέσης ημερήσιας ενεργειακής κατανάλωσης, και της αντίστοιχης τυπικής απόκλισης, με την πάροδο του χρόνου (χειμώνας, $tr = 1$).



Σχήμα 6.12: Γραφήματα εξοικονόμησης ενέργειας-άνεσης (χειμώνας, $tr = 1$).



Σχήμα 6.13: Η σχέση ανάμεσα στο tradeoff και τον αριθμό των ανυπάκουων σημείων λειτουργίας, όπως αυτά περιγράφησαν στο τέλος του κεφαλαίου IV.

Πριν προχωρήσουμε στα αντίστοιχα σχετικά με τη μάθηση σχήματα για αυτή την προσομοίωση, θα προσπαθήσουμε να λύσουμε το μυστήριο της μηδενικής κατανάλωσης ενέργειας που θίξαμε λίγο παραπάνω.

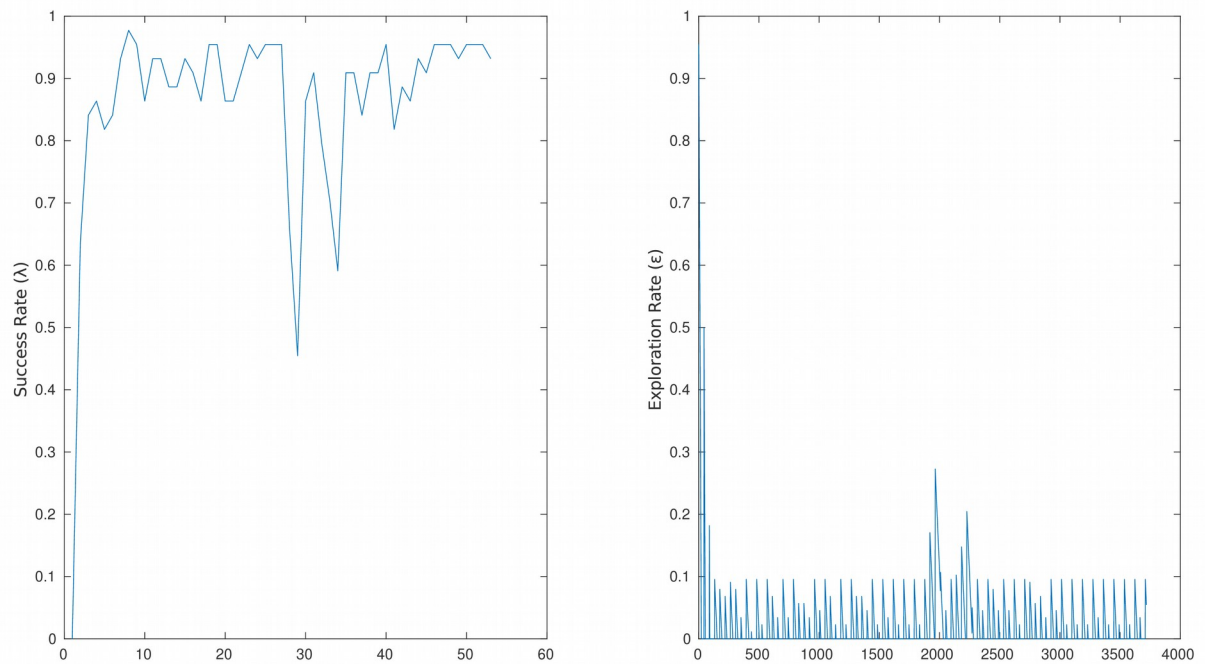
Φροντίσαμε, κατά τις προσομοιώσεις, να μετράμε τον αριθμό των σημείων λειτουργίας τα οποία αποδείχθηκαν “ανυπάκουα”. Με άλλα λόγια αυτό σημαίνει τον αριθμό των φορών που η διαφορά ανάμεσα σε ένα εφαρμοσθέν σημείο λειτουργίας και την προκύπτουσα θερμοκρασία δωματίου που μετρήθηκε στο επόμενο χρονικό βήμα ήταν μεγαλύτερη από μισό βαθμό Κελσίου.

Κατόπιν ως θεωρήσουμε πως ένα σημείο λειτουργίας αποδεικνύεται ανυπάκουο, όταν η αντίστοιχη θερμοκρασία είναι έξω από το εύρος θέρμανσης/ψύξης του θερμοστάτη. Για παράδειγμα, το χειμώνα, προφανώς το σύστημα βρίσκεται σε λειτουργία θέρμανσης. Συνεπώς, τι μπορεί να σημαίνει για το σύστημα ένα αίτημα της μορφής “ζέστανε το δωμάτιο στους 19/20/21 βαθμούς”, ενώ είναι πιθανό αυτή να ήταν η θερμοκρασία του δωματίου *απουσία* θέρμανσης;

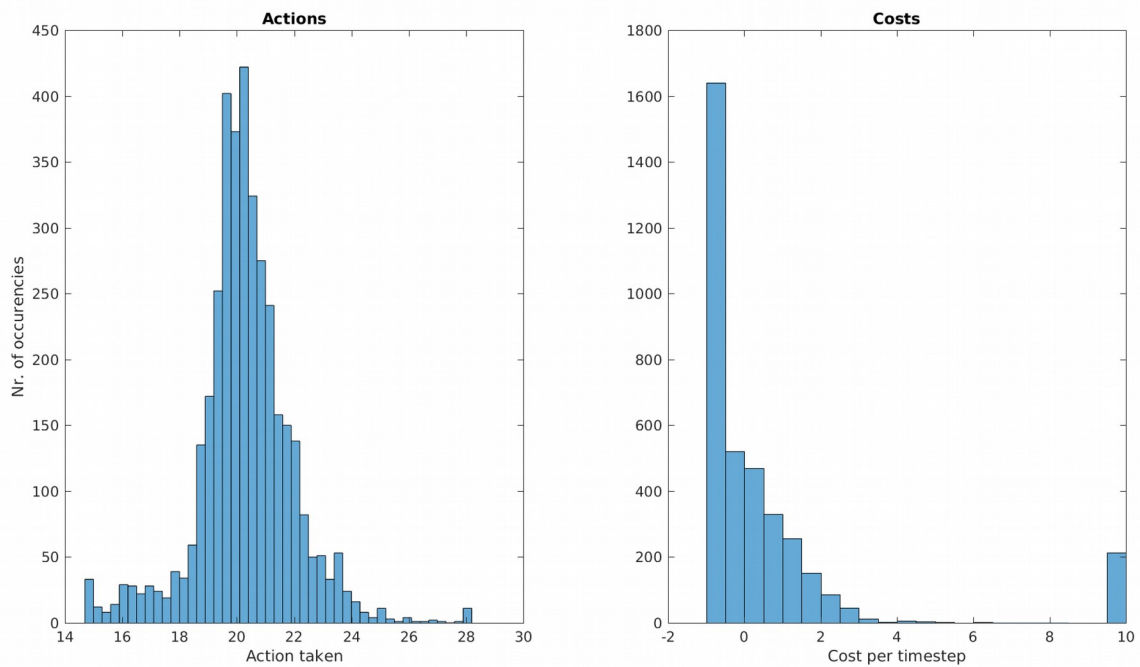
Τίποτα. Ο θερμοστάτης δεν θα μπει καν στον κόπο. Τουλάχιστον, προτείνουμε πως στην περίπτωση του συγκεκριμένου κτιρίου, αυτό κάνει. Και συνεπώς προτείνουμε πως τα περισσότερα, αν όχι όλα, ανυπάκουα σημεία λειτουργίας συνοδεύονται από μηδενική κατανάλωση ενέργειας. Και τονίζουμε πως περίοδοι μηδενικής κατανάλωσης φαίνονται στο Σχ. 6.9 και για τα RBC 20 και 21. Πράγμα που ενισχύει τη θεωρία μας για την ύπαρξη θερμοκρασιών που υπό συγκεκριμένες συνθήκες απαιτούν μηδενική ενέργεια. Πράγμα που μας φέρνει, τελικά, στο Σχ. 6.13 και το τελικό μας επιχείρημα: *ο πράκτορας εξαναγκάζει το σύστημα σε όλο και περισσότερα ανυπάκουα σημεία λειτουργίας καθώς αυξάνεται το tradeoff, διότι έτσι σπαταλά όλο και λιγότερη ενέργεια.*

Μπορεί, βέβαια, να κάνουμε και λάθος. Γι αυτό δεν θα επεκταθούμε περαιτέρω. Παραμένει, παρ’ όλα αυτά, μια παρατήρηση που κατά την ταπεινή γνώμη του συγγραφέα αξίζει να διερευνηθεί περισσότερο.

Τα Σχ. 6.14-15 συνεχίζουν να στηρίζουν τις ικανότητες μάθησης του πράκτορά μας. Παρατηρούμε πως βουτιές του λόγου θετικών μεταβάσεων (και άρα της ποιότητας πρόβλεψης του νευρωνικού δικτύου όσον αφορά τη συνάρτηση Q) συνοδεύονται, όπως απαιτεί ο μηχανισμός εξερεύνησης, με ραγδαίες αυξήσεις της πιθανότητας εξερεύνησης. Επίσης, το ιστόγραμμα των ληφθέντων δράσεων είναι πια πολύ πιο διαφορετικό του Σχ. 7, επιδεικνύοντας την εμμονή του πράκτορά μας πλέον σε χαμηλές θερμοκρασίες. Τα άμεσα κόστη παραμένουν αρνητικά στην πλειοψηφία τους.



Σχήμα 6.14: Λόγος θετικών μεταβάσεων και πιθανότητα εξερεύνησης (χειμώνας, $tr = 1$).

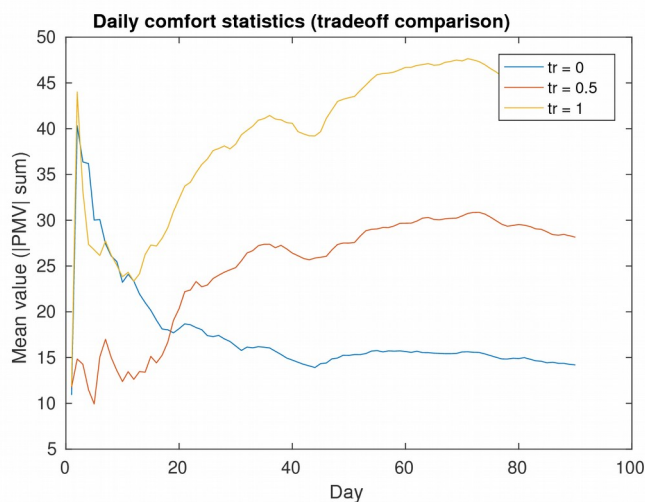


Σχήμα 6.15: Ιστογράμματα δράσεων-κόστους (χειμώνας, $tr = 1$).

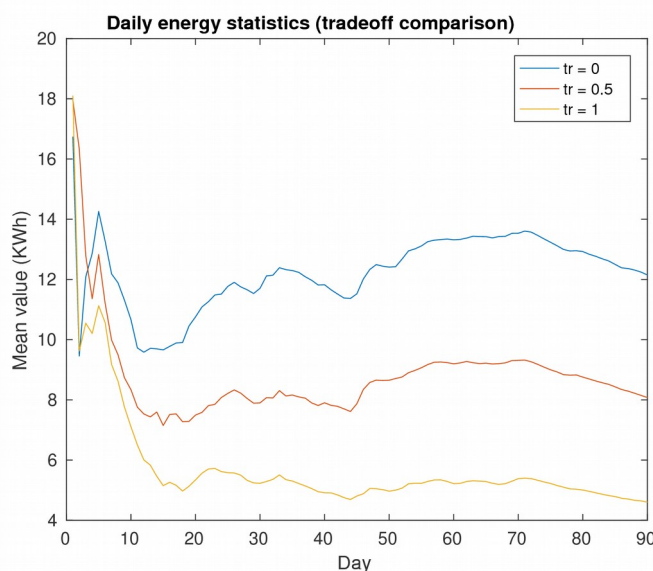
VI.3. Χειμερινή περίοδος ($tr = 0.5$)

Για λόγους συντομίας, δεν θα παραθέσουμε το ίδιο πλήθος σχημάτων για αυτή την περίπτωση. Πιστεύουμε όμως πως τα ακόλουθα δύο είναι ούτως ή άλλως αρκετά.

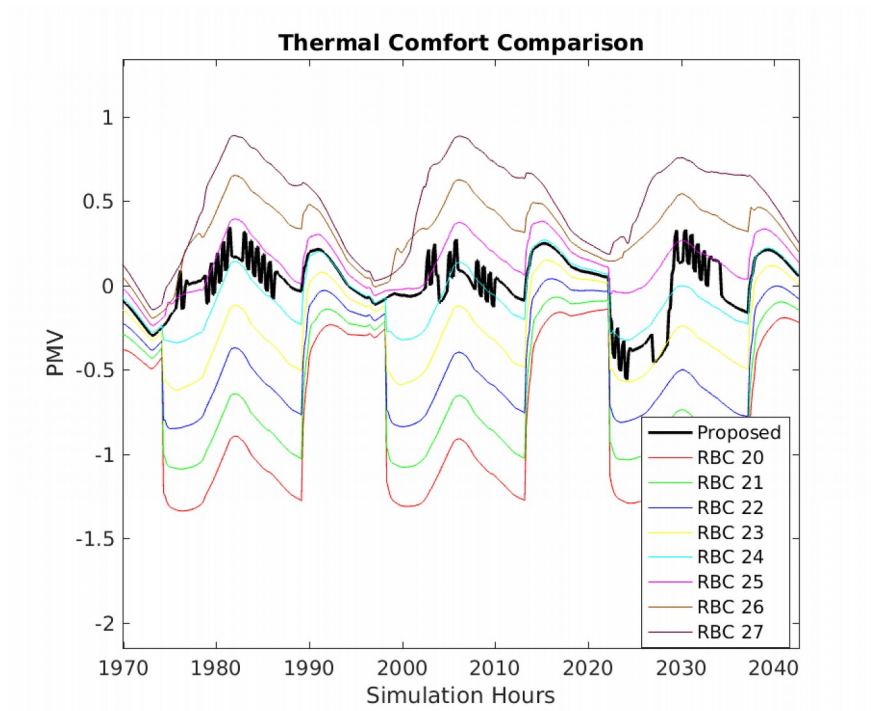
Συγκεκριμένα δείχνουμε στα Σχ. 6.16-17 τον τρόπο με τον οποίο μεταβάλλονται η μέση ημερήσια άνεση και κατανάλωση για τις δύο ακραίες περιπτώσεις που εξετάσαμε παραπάνω, και για μία ενδιάμεση. Βλέπουμε πως πράγματι, η μέση περίπτωση αποφέρει μέσα αποτελέσματα.



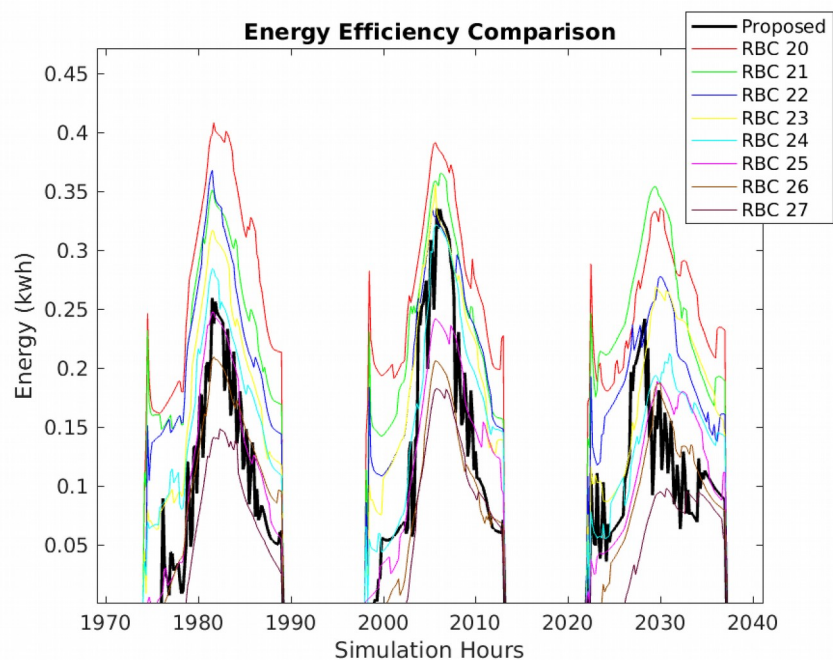
Σχήμα 6.16: Μεταβολή μέσης ημερήσιας θερμικής άνεσης για διάφορες τιμές του tradeoff (χειμώνας).



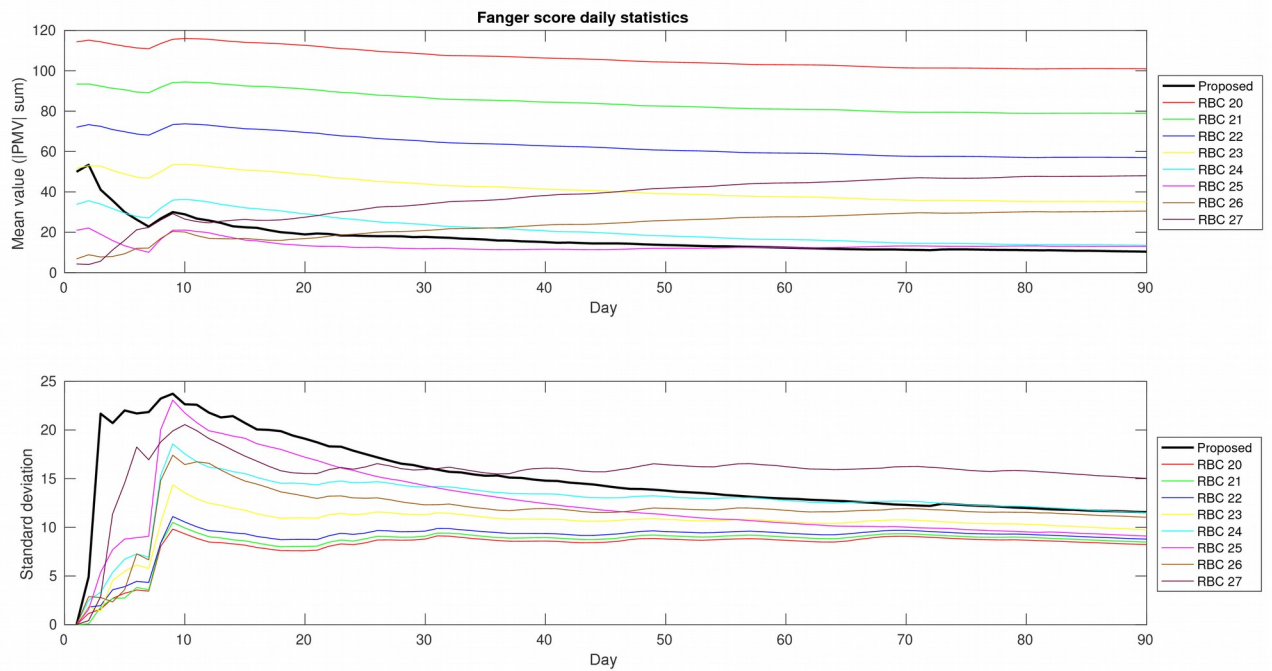
Σχήμα 6.17: Μεταβολή μέσης ημερήσιας ενεργειακής κατανάλωσης για διάφορες τιμές του tradeoff (χειμώνας).

VI.4. Καλοκαιρινή περίοδος ($tr = 0$)

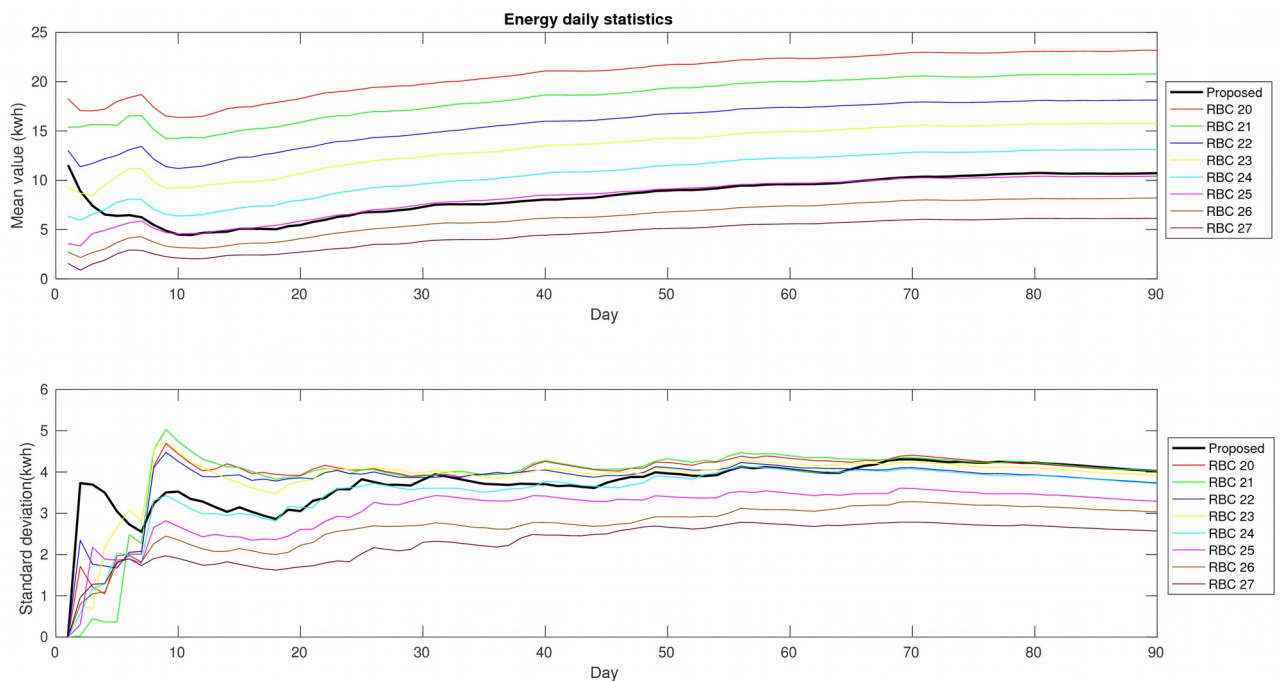
Σχήμα 6.18: Σύγκριση θερμικής άνεσης για τρεις διαδοχικές θερινές ημέρες (καλοκαίρι, $tr = 0$).



Σχήμα 6.19: Σύγκριση ενεργειακής κατανάλωσης για τρεις διαδοχικές θερινές ημέρες (καλοκαίρι, $tr = 0$).



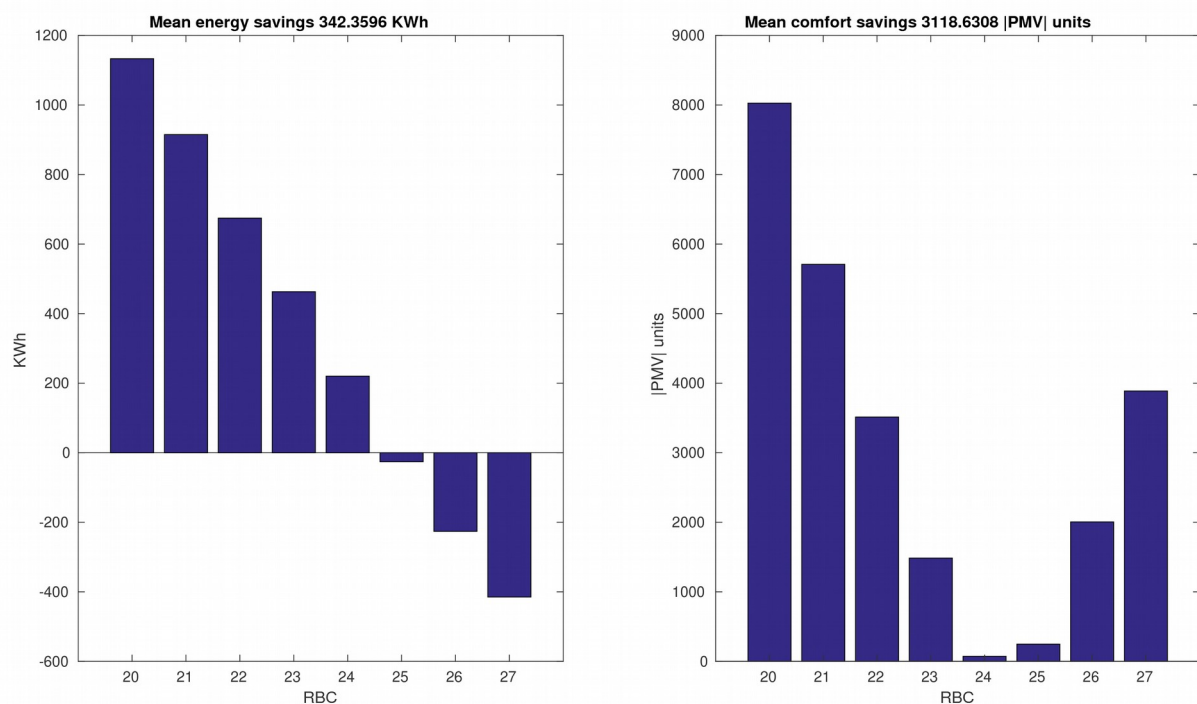
Σχήμα 6.20: Μεταβολή μέσης ημερήσιας θερμικής άνεσης, και της αντίστοιχης τυπικής απόκλισης, με την πάροδο του χρόνου (καλοκαίρι, $tr = 0$).



Σχήμα 6.21: Μεταβολή μέσης ημερήσιας ενεργειακής κατανάλωσης, και της αντίστοιχης τυπικής απόκλισης, με την πάροδο του χρόνου (καλοκαίρι, $tr = 0$).

Ομοίως με την περίπτωση του χειμώνα, τα ανωτέρω αποτελέσματα μας οδηγούν στις ακόλουθες παρατηρήσεις:

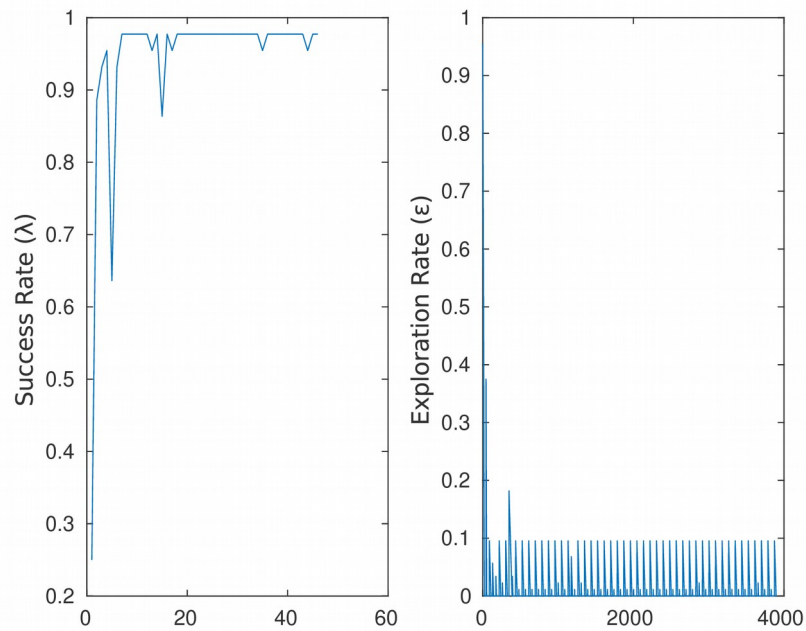
- το καλοκαίρι δεν χρειάζεται περίοδος warmup, γιατί η θερμοκρασία ξεκινά ήδη από ικανοποιητικά επίπεδα όσον αφορά τη θερμική άνεση. Έτσι είναι εύκολο για τον θερμοστάτη, ύστερα από δύομιση περίπου μήνες λειτουργίας, να επιδείξει βέλτιστη συμπεριφορά έναντι όλων των RBC.
- η μέση ενεργειακή κατανάλωση παραμένει ικανοποιητική.



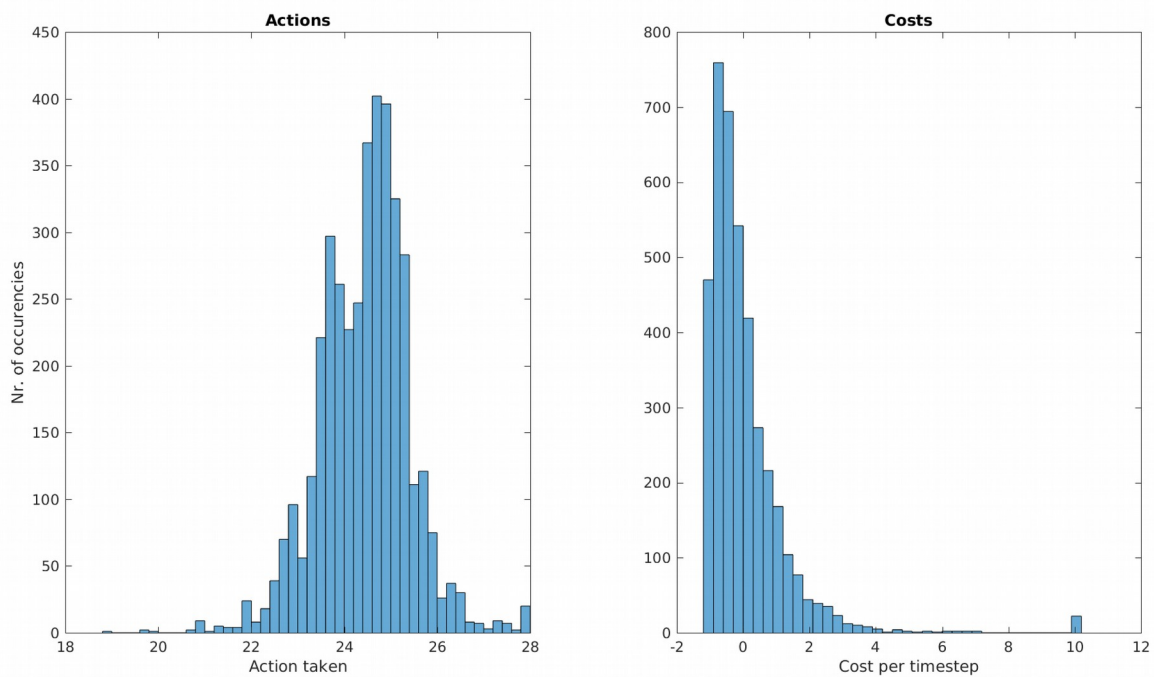
Σχήμα 6.22: Γραφήματα εξοικονόμησης ενέργειας-άνεσης (καλοκαίρι, $tr = 0$).

Τα Σχ. 6.23-24 αφορούν κατ' αναλογία με πριν τη μάθηση του πράκτορα. Ο λόγος θετικών μεταβάσεων παραμένει υψηλός, ρυθμίζοντας αντίστοιχα σε μικρές τιμές την πιθανότητα εξερεύνησης. Η πλειοψηφία των δράσεων που ανατέθηκαν στο θερμοστάτη κυμαίνεται αυτή τη φορά στο διάστημα [24, 25], που στα προηγούμενα σχήματα φαίνονται να αποτελούν τις βέλτιστες (εκ των RBC) θερμοκρασίες για τη θερμική άνεση. Τα άμεσα κόστη παραμένουν ως επί το πλείστον αρνητικά, δείχνοντας πως και εδώ ο μηχανισμός αποφάσεων λειτουργεί άρτια.

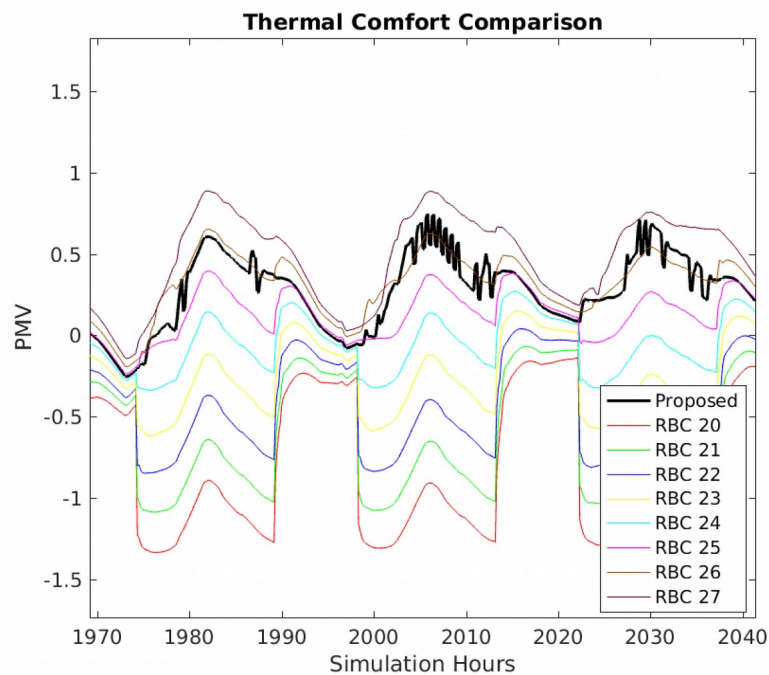
Ιδιαίτερη έμφαση δίνουμε στον εξαιρετικά χαμηλό αριθμό φορών που προέκυψε τερματικό κόστος, όπως φαίνεται στο Σχ. 6.24.



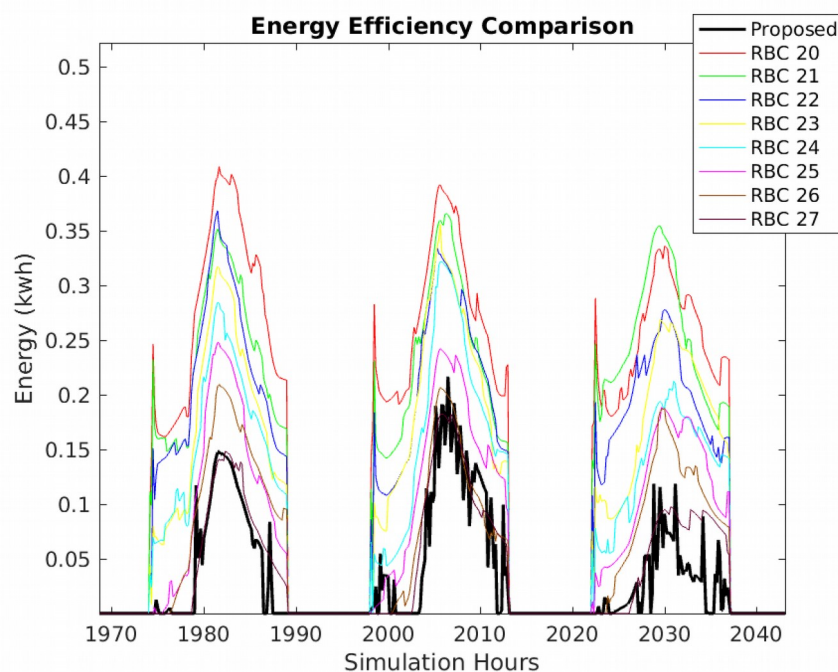
Σχήμα 6.23: Λόγος θετικών μεταβάσεων και πιθανότητα εξερεύνησης (καλοκαίρι, $tr = 0$).



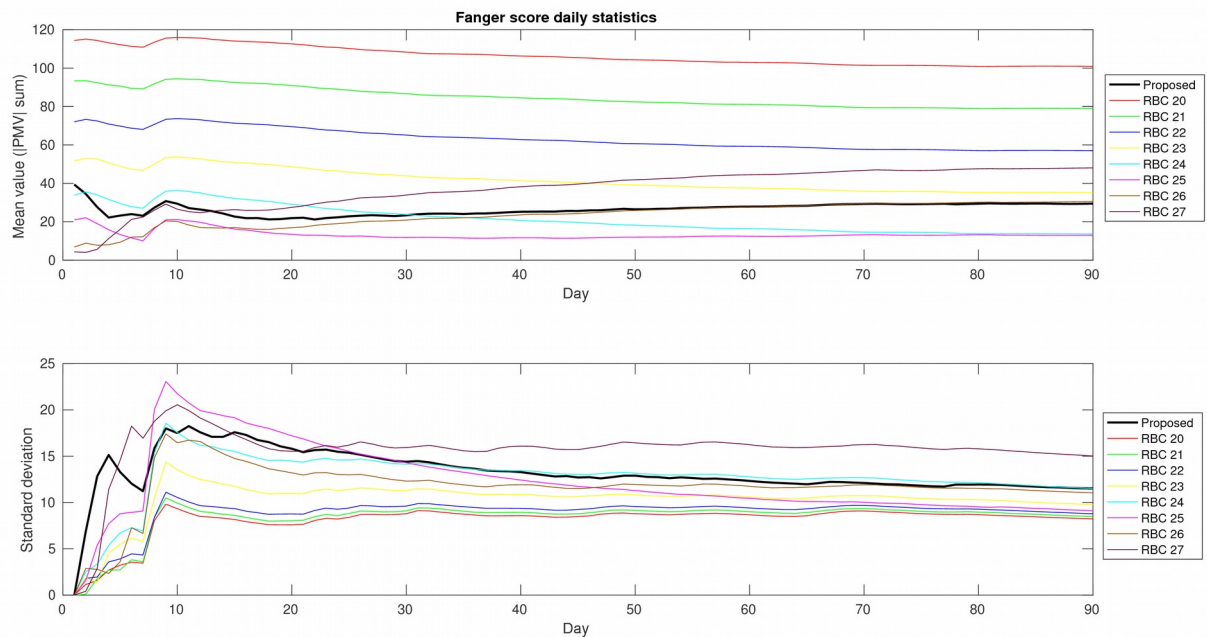
Σχήμα 6.24: Ιστογράμματα δράσεων-κόστους (καλοκαίρι, $tr = 0$).

VI.5. Καλοκαιρινή περίοδος ($tr = 1$)

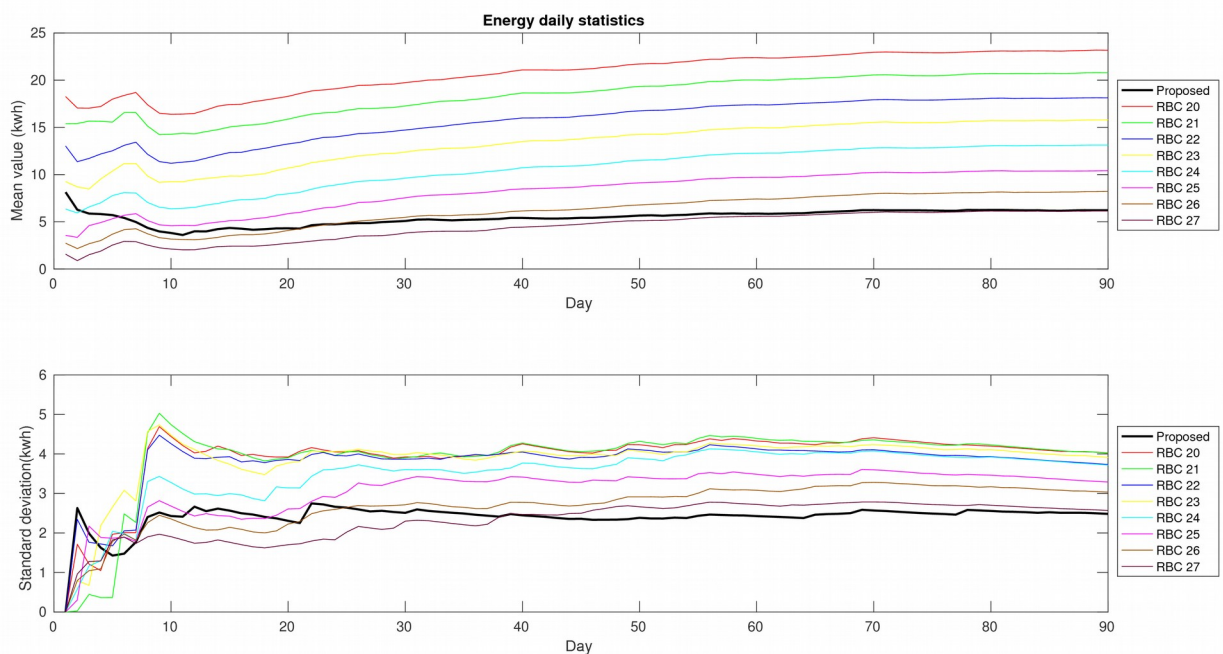
Σχήμα 6.25: Σύγκριση θερμικής άνεσης για τρεις διαδοχικές θερινές ημέρες (καλοκαίρι, $tr = 1$).



Σχήμα 6.26: Σύγκριση ενεργειακής κατανάλωσης για τρεις διαδοχικές θερινές ημέρες (καλοκαίρι, $tr = 1$).



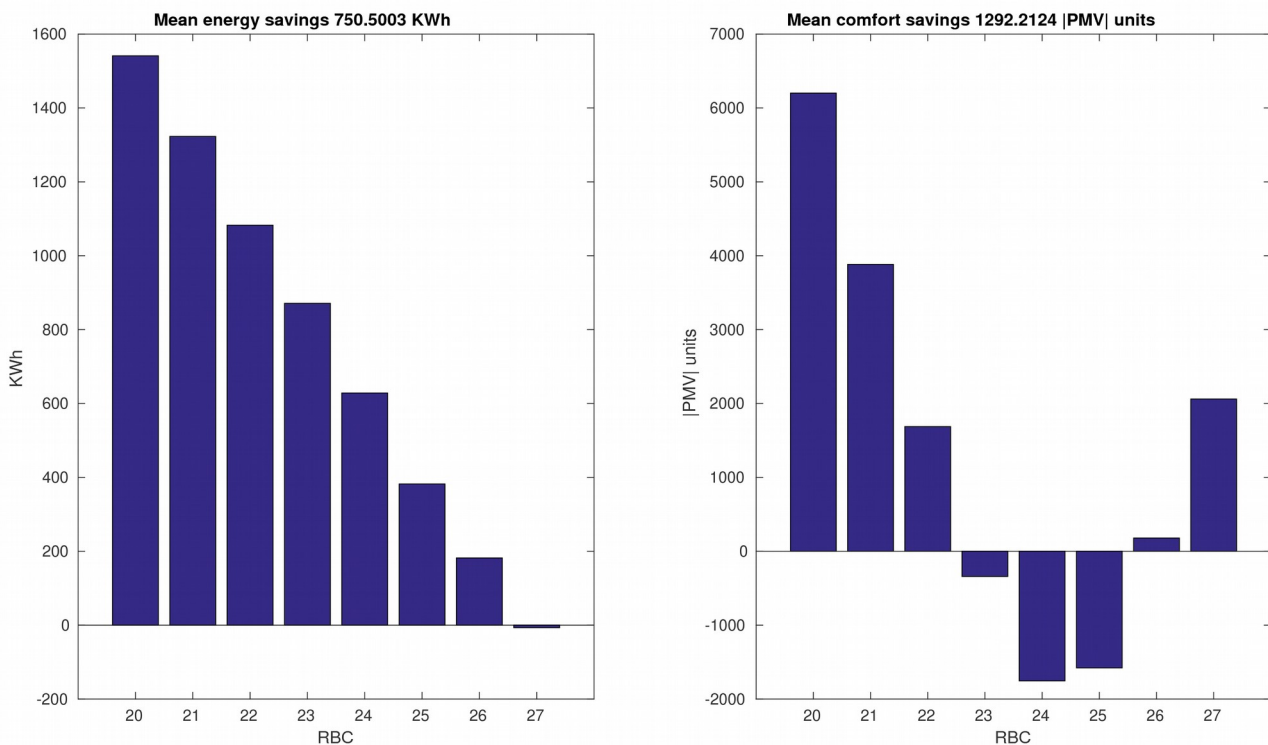
Σχήμα 6.27: Μεταβολή μέσης ημερήσιας θερμικής άνεσης, και της αντίστοιχης τυπικής απόκλισης, με την πάροδο του χρόνου (καλοκαίρι, $tr = 1$).



Σχήμα 6.28: Μεταβολή μέσης ημερήσιας ενεργειακής κατανάλωσης, και της αντίστοιχης τυπικής απόκλισης, με την πάροδο του χρόνου (καλοκαίρι, $tr = 1$).

Και εδώ, λοιπόν, ο θερμοστάτης καταφέρνει μέσω του μηχανισμού αποφάσεων να φτάσει στο πέρας της προσομοίωσης με ελάχιστη ενεργειακή κατανάλωση.

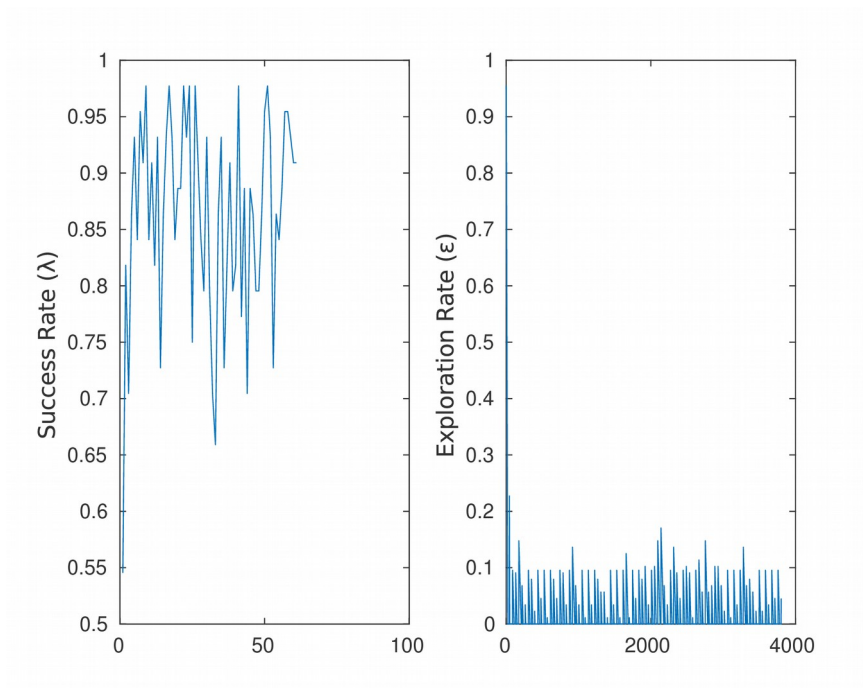
Ιδιαίτερη μνεία αξίζει να δοθεί στον τρόπο με τον οποίο, στο Σχ. 6.25, προσεγγίζει την καμπύλη των 27 βαθμών Κελσίου δίχως να την φτάνει ποτέ, αφού φαίνεται να παραβιάζει τη ζώνη λειτουργίας του 0.75 PMV. Λαμβάνοντας αυτό υπ' όψιν, καταλαβαίνουμε και τον λόγο για τον οποίο το RBC 27 φαίνεται για αρκετό καιρό, στο Σχ. 6.28, να έχει την καλύτερη ενεργειακή απόδοση.



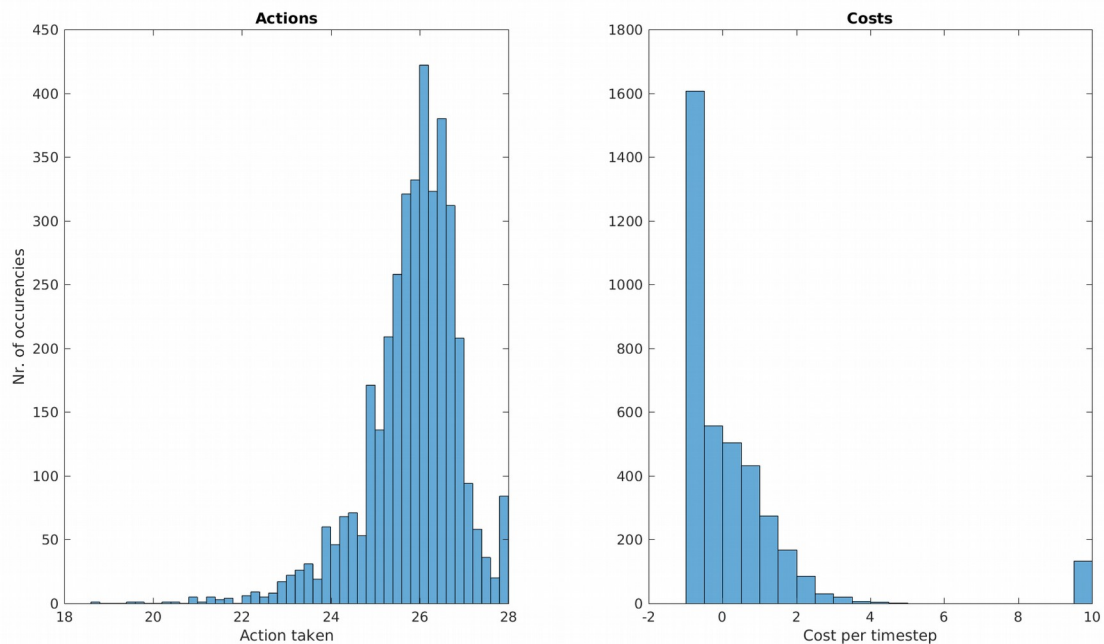
Σχήμα 6.29: Γραφήματα εξοικονόμησης ενέργειας-άνεσης (καλοκαίρι, $tr = 1$).

Learning Performance		
Tradeoff	Retrain sessions (winter/summer)	MSATD (winter/summer)
0 (Optimize Comfort)	180 / 112	0.0135 / 0.0068
0.5 (Optimize both)	207 / 157	0.0198 / 0.0159
1 (Optimize Energy)	303 / 223	0.0284 / 0.0322
Control Performance		
Tradeoff	Mean energy savings (winter/summer)	Mean comfort savings (winter/summer)
0 (Optimize Comfort)	-7.8% / 10.8%	11.9% / 41.8%
0.5 (Optimize both)	28.4% / 32.4%	-3.9% / 27.4%
1 (Optimize Energy)	59.2% / 48.3%	-23.3% / 5.8%

Πίνακας 1: Συγκεντρωτικά αποτελέσματα προσομοιώσεων



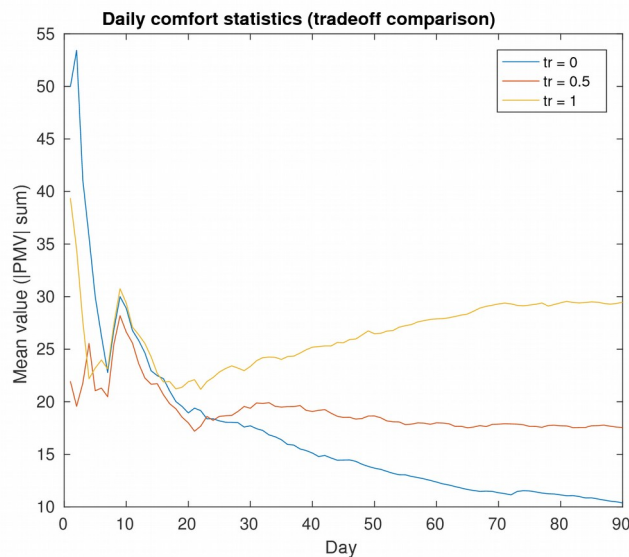
Σχήμα 6.30: Λόγος θετικών μεταβάσεων και πιθανότητα εξερεύνησης (καλοκαίρι, $tr = 1$).



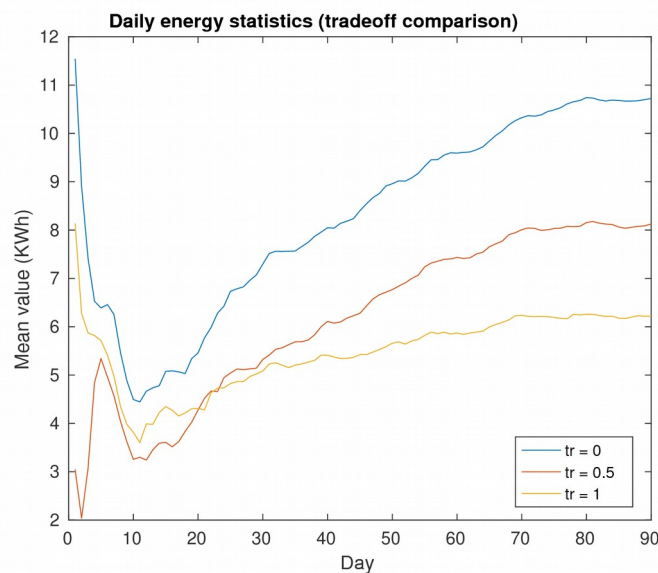
Σχήμα 6.31: Ιστογράμματα δράσεων-κόστους (καλοκαίρι, $tr = 1$).

VI.6. Καλοκαιρινή περίοδος ($tr = 0.5$)

Δίχως επιπλέον σχόλια, στα ακόλουθα σχήματα δείχνουμε πως η υπακοή του μηχανισμού αποφάσεων στο tradeoff συμβαίνει ανάλογα και στις καλοκαιρινές προσομοιώσεις.



Σχήμα 6.32: Μεταβολή μέσης ημερήσιας θερμικής άνεσης για διάφορες τιμές του tradeoff (καλοκαίρι).



Σχήμα 6.33: Μεταβολή μέσης ημερήσιας ενεργειακής κατανάλωσης για διάφορες τιμές του tradeoff (καλοκαίρι).

VI.7. Τελικά σχόλια και παρατηρήσεις

Στις προηγούμενες ενότητες αυτού του κεφαλαίου αποδείχθηκε σχολαστικά η ικανότητα του πράκτορά μας, ανεξαρτήτως καιρικών συνθηκών ή τιμής του tradeoff, να επιδεικνύει βέλτιστη και διαρκώς βελτιούμενη συμπεριφορά. Ο Πίνακας 1 της σελίδας 63 παραθέτει συγκεντρωτικά τα αποτελέσματα όσον αφορά την επίδοση τόσο της μάθησης, όσο και του ίδιου του ελέγχου—τα τελευταία σε ποσοστιαία μορφή. Επ’ αυτών κάνουμε τις ακόλουθες παρατηρήσεις:

- μέσα σε τρεις μήνες, οι περισσότερες φορές που χρειάστηκε να επανεκπαιδευτεί το νευρωνικό δίκτυο του πράκτορα (συνυπολογίζοντας τις επανεκπαιδεύσεις στην αρχή κάθε επόμενης ημέρας) ήταν 303. Αυτό μεταφράζεται σε λιγότερες από τέσσερις—**3.37**—επανεκπαιδεύσεις ανά μέρα, κατά μέσο όρο. Πέραν του ότι κατά τη διάρκειά τους ο έλεγχος θα έχει αναληφθεί από τον ελεγκτή έκτακτης ανάγκης, και άρα ούτως ή άλλως η παραγωγή του επόμενου σημείου λειτουργίας δεν θα καθυστερεί, το αποτέλεσμα αυτό μας ενθαρρύνει να διερευνήσουμε την πιθανότητα υλοποίησης του μηχανισμού αποφάσεων σε σύστημα ενσωματωμένου μικροεπεξεργαστή. Η εκπαίδευση νευρωνικών δικτύων είναι υπολογιστικά απαιτητική και γι αυτό αποφεύγεται συνήθως η εκτέλεσή της σε τέτοιες πλατφόρμες, όμως θεωρούμε τον ελεγκτή μας ικανοποιητικά ελαφρύ προκειμένου να αποτελέσει εξαίρεση σε αυτόν τον κανόνα. Στην επόμενη ενότητα αναφέρουμε τους χρόνους εκπαίδευσης που μετρήσαμε με βάση τον κώδικα και τα δεδομένα μας σε δύο δημοφιλείς πλατφόρμες.
- το μέγιστο μέσο τετραγωνικό απόλυτο σφάλμα χρονικών διαφορών (Mean Absolute Temporal Difference Error—MSATD) που προέκυψε ήταν **3.22%**. Αυτό τεκμηριώνει τους υψηλά σταθερούς λόγους θετικών μεταβάσεων στα σχήματα των προηγούμενων ενοτήτων, και επιβεβαιώνει τον ισχυρισμό πως ο NFQ μαθαίνει γρήγορα, καλά, με λίγα δεδομένα.
- τόσο το χειμώνα, όσο και το καλοκαίρι, παρατηρείται σταθερά ανάλογη σχέση μεταξύ των φορών επανεκπαίδευσης του MLP και του MSATD. Λαμβάνοντας υπ’ όψιν πως το νευρωνικό επανεκπαιδεύεται—πέραν των φορών στην αρχή της κάθε μέρας—όποτε ο πράκτορας σφάλνει, και άρα όποτε ο πράκτορας δεν έχει προβλέψει ορθά το αναμενόμενο κόστος από ένα ζεύγος κατάστασης-δράσης, άρα όποτε το TD-σφάλμα του είναι μεγάλο, το αποτέλεσμα αυτό είναι απόλυτα λογικό, αναμενόμενο και απαραίτητο.

- οι ποσοστιαίες βελτιώσεις κατανάλωσης και άνεσης δείχνουν αφ' ενός την υπακοή στο tradeoff, αφ' ετέρου, στην περίπτωση του χειμώνα, το χαρακτήρα “μαύρου προβάτου” που έχει η θερμική άνεση λόγω της αρχικής περιόδου του warmup. Και πάλι, όμως, αυτό ανταμείβεται με εξοικονόμηση ενέργειας έως και **59.2%**.

VI.7.1 Μετρήσεις σε συστήματα μικροεπεξεργαστών–υλοποισιμότητα

Προκειμένου να αποδείξουμε την υλοποισιμότητα του πράκτορα αυτής της εργασίας σε πλατφόρμες ενσωματωμένων μικροεπεξεργαστών, πήραμε μετρήσεις για τους χρόνους *πρόβλεψης* και *εκπαίδευσης* του νευρωνικού δικτύου πάνω στα δεδομένα ενός TB (υπενθυμίζουμε, Transitions Book, δηλαδή η βάση στην οποία ο πράκτορας αποθηκεύει όλες τις μεταβάσεις του συστήματος) με 4000 καταχωρήσεις. Το μέγεθος αυτό αντιστοιχεί σε περίπου 1.5 μήνα λειτουργίας και 153 KB μνήμης.

Ο κώδικας υλοποιήθηκε στο scripting περιβάλλον Anaconda, με τη βοήθεια της βιβλιοθήκης scikit-learn. Δοκιμάστηκε, κατόπιν, επαναληπτικά σε δύο πλατφόρμες, το BeagleBoard xm (ARM37x CortexA8 1-core up to 1GHz) και το Raspberry Pi Zero (ARMv6 BCM2835 1-core up to 1GHz). Οι μέσοι χρόνοι για εκπαίδευση και πρόβλεψη στις δυο πλατφόρμες ήταν **48.94 sec** και **64.32 sec** (εκπαίδευση) και **0.0013 sec** και **0.0025 sec** (πρόβλεψη) αντίστοιχα.

VI.7.2 Παραδοχές και πιθανές επεκτάσεις

Θα ήταν παράλογο να μην τονιστούν οι παραδοχές υπό τις οποίες εκπονήθηκε αυτή η εργασία, παρ' όλο που κάποιες έχουν ήδη αναφερθεί σε προηγούμενα σημεία. Ο συγγραφέας τρέφει την καλή πίστη πως αντί να αφαιρούν αξία από το παρόν έργο, οι παραδοχές αυτές επιφυλάσσουν εξαιρετικά πρόσφορο έδαφος για μελλοντική έρευνα, επεκτάσεις και βελτιώσεις.

- η συνάρτηση κόστους έχει πρόσβαση σε μια μέτρηση του PMV. Υπό τις συνθήκες των προσομοιώσεών μας, η μέτρηση αυτή ερχόταν από το ακριβές περιβάλλον του EnergyPlus. Σε μια πρακτική εφαρμογή, δεν γνωρίζουμε κατά πόσο κάτι τέτοιο θα ήταν εφικτό—παρ' όλο που πρόσφατη έρευνα έχει διεξαχθεί πάνω στην εκτίμηση της θερμικής άνεσης σε πραγματικό χρόνο [33].
- σε χώρους με εντονότερη κινητικότητα από αυτή των πέντε ατόμων με την οποία τροφοδοτήσαμε την ζώνη των προσομοιώσεών μας, η εισαγωγή μας μεταβλητής σχετικής με τους ενοίκους της ζώνης στο σήμα

κατάστασης του ελεγκτή θα ήταν απαραίτητη (είτε παρούσα μέτρηση, είτε κάποιου είδους πρόβλεψη). Στην περίπτωση μας η επίδραση των ατόμων στην κατανάλωση και την άνεση που προκύπτουν από την εφαρμογή ενός σημείου λειτουργίας αποδείχθηκαν αμελητέες, όμως για μεγαλύτερους αριθμούς αυτό δεν θα ισχύει. Επιπλέον, μια τέτοια μεταβλητή θα επέτρεπε την εισαγωγή μίας ακόμα δράσης, αυτής της απενεργοποίησης του θερμοστάτη, στις περιπτώσεις εκείνες που ο πράκτορας προβλέπει, για παράδειγμα, ότι ο χώρος θα πάψει να είναι κατειλημμένος. Αυτό θα ήταν και το κλειδί προκειμένου να οδηγηθούμε κατόπιν σε έναν θερμοστάτη που καθορίζει αυτόνομα το ωράριό του.

- παρόμοιες επεκτάσεις του σήματος κατάστασης θα μπορούσαν να αφορούν παραμέτρους όπως για παράδειγμα το αν τα παράθυρα του χώρου είναι ανοιχτά ή κλειστά, ή πόσο ανοιχτά είναι, και γενικότερα άλλες μετρήσιμες μεταβλητές ενός χώρου που επηρεάζουν με ρεαλιστικό τρόπο το κόστος. Βέβαια, τέτοιες επεκτάσεις δημιουργούν ταυτόχρονα την ανάγκη για κάποια μέθοδο προσέγγισης συναρτήσεων πιο εξελιγμένη από το MLP, με δυνατότητες ταχύτερης μάθησης.
- δεν πρέπει να αγνοούμε το γεγονός ότι όλα τα ανωτέρω αποτελέσματα προέρχονται από την προσομοίωση του ελέγχου μίας από τις δέκα ζώνες ενός κτιρίου. Πήραμε την πρωτοβουλία να θεωρήσουμε πως αυτό δεν προκαλεί βλάβη της γενικότητας, αφού η βιβλιογραφία ορίζει τις θερμικές ζώνες των κτιρίων ως ανεξάρτητες μεταξύ τους. Εξακολουθεί όμως να διατηρεί την έντασή της η περιέργειά μας για μια προσομοίωση πλήρους κλίμακας.

ΤΕΛΟΣ ΕΡΓΑΣΙΑΣ

Βιβλιογραφία

- [1] Ανακοίνωση της Κομισιόν προς το Ευρωπαϊκό Κοινοβούλιο, το Συμβούλιο, την Ευρωπαϊκή Οικονομική και Κοινωνική Επιτροπή και την Επιτροπή Περιφερειών. *Ενέργεια 2020: Μια στρατηγική για ανταγωνιστική, αειφόρο και ασφαλή ενέργεια.*
- [2] Ανακοίνωση της Κομισιόν προς το Ευρωπαϊκό Κοινοβούλιο, το Συμβούλιο, την Ευρωπαϊκή Οικονομική και Κοινωνική Επιτροπή και την Επιτροπή Περιφερειών. *Πλαίσιο πολιτικής για το κλίμα και την ενέργεια κατά την περίοδο από το 2020 έως το 2030.*
- [3] Ανακοίνωση της Κομισιόν προς το Ευρωπαϊκό Κοινοβούλιο, το Συμβούλιο, την Ευρωπαϊκή Οικονομική και Κοινωνική Επιτροπή και την Επιτροπή Περιφερειών. *Ενεργειακός χάρτης πορείας για το 2050.*
- [4] Άρθρο της wikipedia για τη Συμφωνία του Παρισιού. https://en.wikipedia.org/wiki/Paris_Agreement
- [5] Σελίδα της Κομισιόν για το ρόλο των τριών στην ενεργειακή αποδοτικότητα. <https://ec.europa.eu/energy/en/topics/energy-efficiency/buildings>
- [6] McDowall, R. (2006). *Fundamentals of HVAC systems*. Academic Press.'
- [7] Fanger, P. O. (1970). Thermal comfort. Analysis and applications in environmental engineering. *Thermal comfort. Analysis and applications in environmental engineering.*
- [8] American society of heating refrigerating and air conditioning engineers. (1992). *ASHRAE STANDARD: An American Standard: Thermal Environmental Conditions for Human Occupancy*. American Society of Heating refrigerationg and air conditioning engineers.
- [9] Riedmiller, M. (1998). High quality thermostat control by reinforcement learning-A case study. In *Proceedings of the Conald Workshop* (pp. 1-2).

- [10] Kusiak, A., Tang, F., & Xu, G. (2011). Multi-objective optimization of HVAC system with an evolutionary computation algorithm. *Energy*, 36(5), 2440-2449.
- [11] Ruelens, F., Iacovella, S., Claessens, B. J., & Belmans, R. (2015). Learning agent for a heat-pump thermostat with a set-back strategy using model-free reinforcement learning. *Energies*, 8(8), 8300-8318.
- [12] Danassis, P., Siozios, K., Korkas, C., Soudris, D., & Kosmatopoulos, E. (2017). A low-complexity control mechanism targeting smart thermostats. *Energy and Buildings*, 139, 340-350.
- [13] Marantos, C., Siozios, K., & Soudris, D. (2017). A Flexible Decision-Making Mechanism Targeting Smart Thermostats. *IEEE Embedded Systems Letters*.
- [14] Κεντρική σελίδα παρουσίασης του έξυπνου θερμοστάτη της Nest. <https://nest.com/thermostats/nest-learning-thermostat/overview/>
- [15] Tesauro, G. (1995). Td-gammon: A self-teaching backgammon program. In *Applications of Neural Networks* (pp. 267-285). Springer US.
- [16] Riedmiller, M., Montemerlo, M., & Dahlkamp, H. (2007, October). Learning to drive a real car in 20 minutes. In *Frontiers in the Convergence of Bioscience and Information Technologies, 2007. FBIT 2007* (pp. 645-650). IEEE.
- [17] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.
- [18] Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1, No. 1). Cambridge: MIT press.

- [19] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- [20] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- [21] Άρθρο της wikipedia για το Θεώρημα Καθολικής Προσέγγισης.
https://en.wikipedia.org/wiki/Universal_approximation_theorem
- [22] Haykin, S. S., Haykin, S. S., Haykin, S. S., & Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Upper Saddle River, NJ, USA:: Pearson.'
- [23] Riedmiller, M. (2005, October). Neural fitted Q iteration-first experiences with a data efficient neural reinforcement learning method. In *ECML* (Vol. 3720, pp. 317-328).
- [24] Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *Neural Networks, 1993., IEEE International Conference on* (pp. 586-591). IEEE.
- [25] Bollegala, D. (2017). Dynamic feature scaling for online learning of binary classifiers. *Knowledge-Based Systems*, 129, 97-105.
- [26] Teng, T. H., Tan, A. H., & Tan, Y. S. (2012). Self-regulating action exploration in reinforcement learning. *Procedia Computer Science*, 13, 18-30.
- [27] Gehring, C., & Precup, D. (2013, May). Smart exploration in reinforcement learning using absolute temporal difference errors. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems* (pp. 1037-1044). International Foundation for Autonomous Agents and Multiagent Systems.

[28] Sagerschnig, C., Gyalistras, D., Seerig, A., Prívará, S., Cigler, J., & Vana, Z. (2011, September). Co-simulation for building controller development: The case study of a modern office building. In *Proc. CISBAT* (pp. 14-16).

[29] Kontes, G. D., Giannakis, G. I., Kosmatopoulos, E. B., & Rovas, D. V. (2012, October). Adaptive-fine tuning of building energy management systems using co-simulation. In *Control Applications (CCA), 2012 IEEE International Conference on* (pp. 1664-1669). IEEE.

[30] Αρχεία καιρού για το EnergyPlus. https://energyplus.net/weather-location/europe_wmo_region_6/GRC//GRC_Athens.167160_IWEC

[31] Urban, B. & Roth, K. (2014). A data-driven framework for comparing residential thermostat energy performance. Technical report. Fraunhofer Center for Sustainable Energy Systems, Cambridge, MA.

[32] Στατιστικά στοιχεία τιμών ενέργειας για νοικοκυριά το δεύτερο μισό του 2015. [http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Electricity_prices_for_household_consumers,_second_half_2015_\(%C2%B9\) \(EUR per kWh\) YB16.png](http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Electricity_prices_for_household_consumers,_second_half_2015_(%C2%B9)_%28EUR_per_kWh%29_YB16.png)

[33] Ruz, M. L., Fragoso, S., Rodríguez, D., & Vázquez, F. (2015, June). Real-time estimation of thermal comfort indices in an office building with a solar powered HVAC system. In *Control and Automation (MED), 2015 23th Mediterranean Conference on* (pp. 803-808). IEEE.