

ΟΔΗΓΟΣ ΧΡΗΣΗΣ ΑΙΣΘΗΤΗΡΩΝ PurpleAir

Περιεχόμενα

- Εισαγωγή
- Περιγραφή αισθητήρα
- Εγκατάσταση αισθητήρα
- Αντιμετώπιση προβλημάτων αισθητήρα
- Λήψη δεδομένων
- Διαχείριση δεδομένων
- Περιγραφική στατιστική

ΕΙΣΑΓΩΓΗ

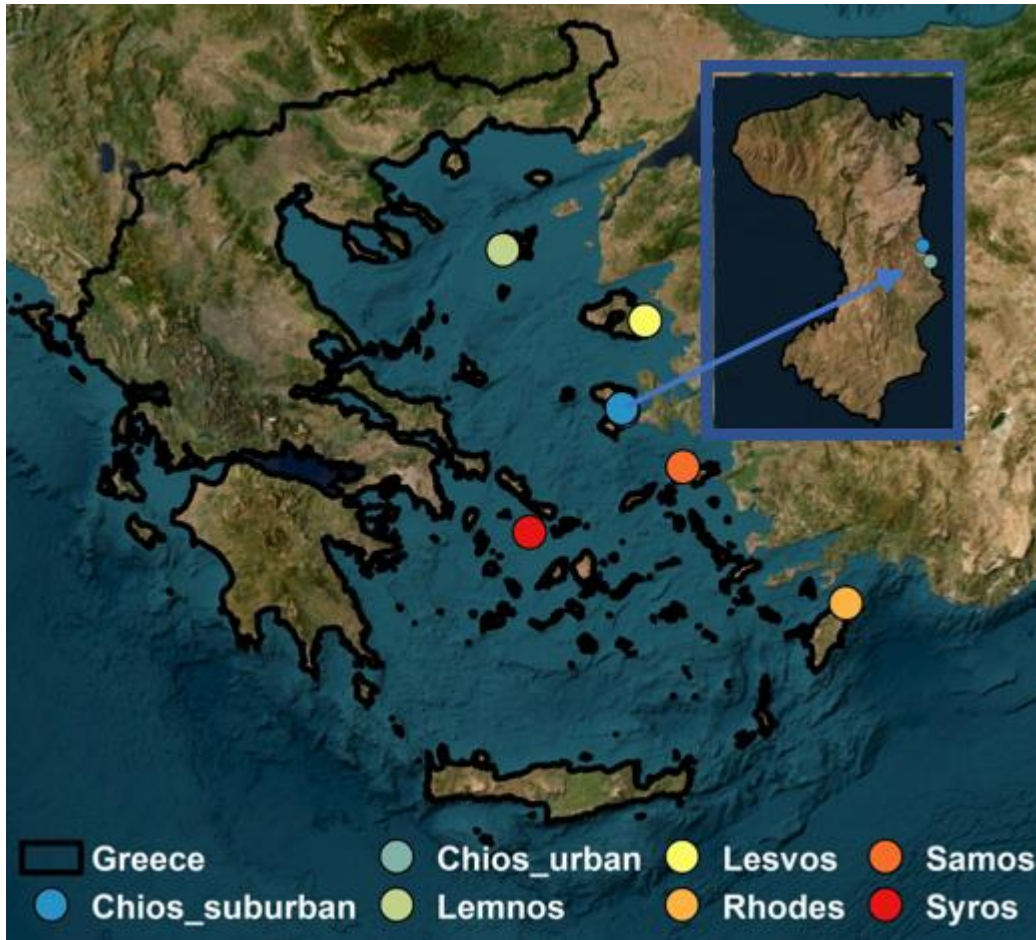
Παραδοσιακά, η ατμοσφαιρική ρύπανση για ρύπους όπως CO, NO, NO₂ και αιωρούμενα σωματίδια (Particulate Matter - PM) παρακολουθείται σε σταθερές τοποθεσίες με ακριβή και υψηλού κόστους όργανα μέτρησης (Rai and Kumar, 2017). Την τελευταία δεκαετία, οι μέθοδοι παρακολούθησης της ποιότητας του αέρα, το ποιος μπορεί να πραγματοποιήσει τις μετρήσεις και η χωρική και χρονική κλίμακα παρακολούθησης έχουν αλλάξει ριζικά. Αυτές οι αλλαγές οφείλονται στην ανάπτυξη των αισθητήρων χαμηλού κόστους (Low-Cost Sensors - LCSs), στους οποίους συμπεριλαμβάνονται και αισθητήρες μέτρησης PM (Giordano et al, 2021). Λόγω των εγγενών χαρακτηριστικών τους (χαμηλό κόστος, μικρό μέγεθος, φορητότητα), οι LCSs είναι εξαιρετικά εύρηστοι και μπορούν να χρησιμοποιηθούν για τον εντοπισμό σημείων υψηλής ρύπανσης, τον εντοπισμό των πηγών ρύπανσης, τη συμπλήρωση των δικτύων των σταθερών σταθμών παρακολούθησης, τη μέτρηση της ατομικής έκθεσης σε ρύπους, την εκπαίδευση και την ευαισθητοποίηση των πολιτών για την ποιότητα του αέρα (EPA n.d.).

Ωστόσο, υπάρχουν ορισμένοι περιορισμοί (μειονεκτήματα) που θα πρέπει να ξεπεραστούν πριν οι αισθητήρες χαμηλού κόστους μπορέσουν να αντικαταστήσουν τις παραδοσιακές μεθόδους παρακολούθησης της ρύπανσης. Οι περιορισμοί περιλαμβάνουν τη χαμηλότερη ακρίβεια, τη μικρότερη αξιοπιστία και το μικρότερο εύρος μέτρησης σε σχέση με τις κλασσικές μεθόδους. Τα μειονεκτήματα αυτά μπορεί να περιορισθούν με τον έλεγχο της επαναληψιμότητας και της ακρίβειας των αισθητήρων πριν ή/και κατά την χρήση τους και την διόρθωση της απόκρισής τους ώστε να βελτιώνεται η ακρίβεια. Στην ενότητα “6. Διαχείριση Δεδομένων” παρουσιάζεται η διαδικασία που ακολουθήθηκε προς την κατεύθυνση αυτή για τους αισθητήρες PurpleAir που χρησιμοποιήθηκαν στη μελέτη.

Οι αισθητήρες PurpleAir χρησιμοποιούνται εκτεταμένα παγκοσμίως από πολίτες και κοινότητες για τη συλλογή δεδομένων για τα αιωρούμενα σωματίδια. Τα δεδομένα αποστέλλονται μέσω ενός δικτύου WiFi στο υπολογιστικό νέφος (cloud), παρουσιάζονται σε χάρτη και μπορεί είτε να παρακολουθούνται από εκεί είτε να ληφθούν σε ένα αρχείο για περαιτέρω ανάλυση. Η επισκόπηση των δεδομένων σε πραγματικό χρόνο σε έναν χάρτη είναι πολύ χρήσιμη γιατί δίνει άμεση εικόνα της ποιότητας του αέρα για άτομα που ασκούνται σε εξωτερικούς χώρους, για ευάλωτα άτομα κ.λπ. Σε πολλές περιπτώσεις αυτή η πληροφορία είναι αρκετή. Ωστόσο, η λήψη των δεδομένων στον υπολογιστή του χρήστη, η διαχείριση και η ανάλυσή τους προσφέρει περαιτέρω πληροφορίες για τα επίπεδα συγκεντρώσεων για μεγάλα χρονικά διαστήματα, βοηθά στην ανίχνευση και διόρθωση πιθανών ανωμαλιών των δεδομένων, στην ανίχνευση συγκεκριμένων μοτίβων και στην παρουσίαση επεξεργασμένων αποτελεσμάτων στο ευρύτερο κοινό.

Το 2021 ένα δίκτυο 7 αισθητήρων PurpleAir εγκαταστάθηκε στις μεγαλύτερες πόλεις 6 νησιών του Αιγαίου για την παρακολούθηση των επιπέδων $PM_{2.5}$ (εικόνα 1). Ένας επιπλέον αισθητήρας εγκαταστάθηκε στο Θησείο της Αθήνας, στις εγκαταστάσεις του Αστεροσκοπείου Αθηνών. Για πρώτη φορά αξιολογήθηκε η ποιότητα του αέρα στην περιοχή σε τόσο μεγάλη χωρική και χρονική κλίμακα. Η έρευνα διεξήχθη στα πλαίσια του έργου ENIRISST+ που συγχρηματοδοτείται από την Ελλάδα και την Ευρωπαϊκή Ένωση.

Σχετικά με την χωροθέτηση των αισθητήρων, ελήφθησαν υπόψη οι προβλέψεις της οδηγίας 2008/50/ΕΚ “για την ποιότητα του ατμοσφαιρικού αέρα και καθαρότερο αέρα για την Ευρώπη” που προβλέπει ότι τα σημεία δειγματοληψίας που προορίζονται για την προστασία της ανθρώπινης υγείας τοποθετούνται έτσι ώστε να παρέχουν στοιχεία για τις περιοχές υψηλότερων συγκεντρώσεων και για τις περιοχές αντιπροσωπευτικών συγκεντρώσεων μέσα σε οικισμούς. Παράλληλα, ελήφθη υπόψη η ανασκόπηση της επιστημονικής βιβλιογραφίας, βάσει της οποίας οι επιστημονικές μελέτες που αφορούν το Ανατολικό Αιγαίο είναι περιορισμένες χωρικά και χρονικά. Συναξιολογήθηκε το γεγονός ότι Εθνικό Δίκτυο Παρακολούθησης της Ατμοσφαιρικής Ρύπανσης (ΕΔΠΑΡ) του ΥΠΕΝ δεν περιλαμβάνει σταθμούς στα νησιά του Ανατολικού Αιγαίου. Συνυπολογίστηκε ότι με την πανελλαδική υποδομή ΠΑΝΑΚΕΙΑ, παρακολουθείται η ατμοσφαιρική ρύπανση ($PM_{2.5}$) με ένα δίκτυο αισθητήρων σε όλη τη χώρα εκτός των νησιών Λήμνος, Χίος, Σάμος και Σύρος. Η χωροθέτηση αισθητήρων σε αυτά τα νησιά μέσω της υποδομής ENIRISST+ θα λειτουργήσει συμπληρωματικά και συνεργιστικά με την υποδομή ΠΑΝΑΚΕΙΑ. Τέλος, μέσω επιτόπιας έρευνας που διενεργήθηκε σε όλα τα νησιά (συνεντεύξεις με κατοίκους) καταγράφηκε η υποκειμενική αίσθηση τους για την ποιότητα του αέρα σε κάθε πόλη και οι σημαντικότερες πηγές ατμοσφαιρικής ρύπανσης (μεταφορές, οικιακή θέρμανση, βιομηχανικές/βιοτεχνικές πηγές συμπεριλαμβανόμενων και των σταθμών παραγωγής ηλεκτρικής ενέργειας, επίδραση του τουρισμού). Από τα ανωτέρω, προκρίθηκε η εγκατάσταση ενός αισθητήρα σε κάθε μία από τις μεγαλύτερες πόλεις 6 νησιών του Ανατολικού Αιγαίου (Μύρινα (Λήμνος), Μυτιλήνη (Λέσβος), Χίο (Χίος), Καρλόβασι (Σάμος), Ρόδο (Ρόδος) και Ερμούπολη (Σύρος)), σε σημείο εντός του αστικού ιστού, που να αντιστοιχεί σε αντιπροσωπευτικές τιμές ρύπανσης και αντιπροσωπευτική, μέση έκθεση του πληθυσμού (για παράδειγμα τηρήθηκε απόσταση των μετρητών από βιομηχανικές πηγές ή κόμβους κυκλοφορίας ή/και οι μετρητες χωροθετήθηκαν ανάντη διαφόρων σημειακών πηγών ρύπανσης σε σχέση με τους επικρατούντες ανέμους). Εάν από την μελέτη αυτή προκύψουν αυξημένες τιμές ρύπανσης σε κάποιο νησί ή για κάποια χρονικά διαστήματα, η περίπτωση αυτή μπορεί να μελετηθεί περαιτέρω με ένα πυκνότερο δίκτυο.



Εικόνα 1. Το δίκτυο των σταθμών παρακολούθησης της ποιότητας του αέρα στα νησιά του Αιγαίου

Σκοπός του οδηγού

Ο οδηγός αυτός στοχεύει στην παροχή πληροφοριών στο κοινό (πολίτες και κοινότητες πολιτών, μη κυβερνητικές οργανώσεις, τοπικές αρχές, εκπαιδευτικοί οργανισμοί) σχετικά με την χρήση των αισθητήρων PurpleAir. Αρχικά περιγράφεται η αρχή λειτουργίας των αισθητήρων PurpleAir. Στη συνέχεια παρουσιάζεται μια διαδικασία βήμα προς βήμα για την εγκατάσταση των αισθητήρων και παρουσιάζονται τα συνήθη προβλήματα και οι λύσεις τους. Στη συνέχεια, προτείνεται μια απλή διαδικασία λήψης, διαχείρισης, ανάλυσης και παρουσίασης των δεδομένων ποιότητας αέρα των PurpleAir.

Οι απαιτήσεις για τη χρήση αυτού του οδηγού είναι:

- Πρόσβαση σε αισθητήρα ποιότητας αέρα PurpleAir ή ανάλογο,
- Βασικές γνώσεις περιγραφικής στατιστικής και κάποιου λογισμικού υπολογιστικών φύλλων (πχ του Excel).

Σημείωση 1:

→ Σε αυτό το έγγραφο, δίνονται οδηγίες για τη χρήση αισθητήρων ενός συγκεκριμένου κατασκευαστή (PurpleAir) και γίνονται αναφορές στην χρήση συγκεκριμένων λογισμικών. Οι συγγραφείς δεν συνιστούν τη συγκεκριμένη ή άλλη μάρκα αισθητήρων ούτε συγκεκριμένα λογισμικά.

→ Η χρήση των LCSs προορίζεται να είναι εκπαιδευτική και μη κανονιστική. Αυτό σημαίνει ότι οι LCSs δεν μπορεί να χρησιμοποιηθούν για αδειοδοτήσεις, έλεγχο συμμόρφωσης με όρια ή εκτίμηση επιπτώσεων στην υγεία.

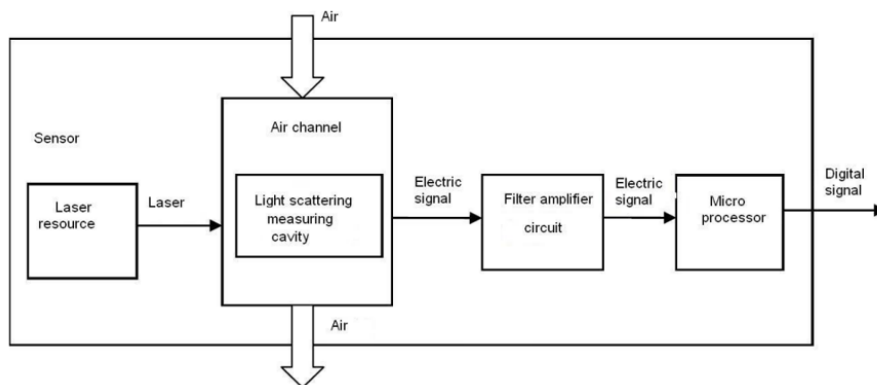
Σημείωση 2:

Η συζήτηση σε αυτόν τον οδηγό περιορίζεται στα $PM_{2.5}$ (και όχι στα PM_{10}) λόγω της υψηλότερης επίπτωσης του λεπτόκοκκου κλάσματος των PM στην ανθρώπινη υγεία. Εκτός αυτού, οι μετρήσεις των $PM_{2.5}$ από τους αισθητήρες PurpleAir είναι πιο ακριβείς σε σύγκριση με τις μετρήσεις των PM_{10} . Για παράδειγμα αναφέρεται στην βιβλιογραφία (Kosmopoulos et al, 2020) ότι όταν τα PM_{10} περιλαμβάνουν ένα σημαντικό χονδρόκοκκο κλάσμα (όπως συμβαίνει στα επεισόδια σκόνης από την Σαχάρα), οι αναφερόμενες συγκεντρώσεις PM_{10} από τους αισθητήρες χαμηλού κόστους φαίνεται ότι δεν είναι ακριβείς.

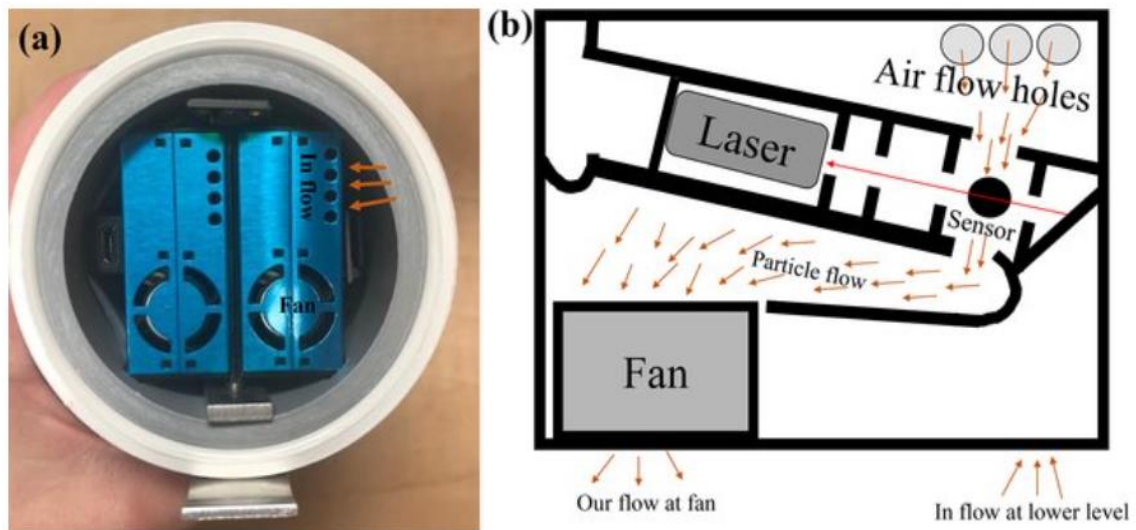
ΠΕΡΙΓΡΑΦΗ ΑΙΣΘΗΤΗΡΑ

Η λειτουργία των αισθητήρων PurpleAir βασίζεται στην σκέδαση του φωτός (Εικόνες 2 και 3). Κάθε μονάδα περιέχει δύο αισθητήρες Plantower PMS5003 που ονομάζονται κανάλια A και B. Ένας ενσωματωμένος ανεμιστήρας τραβάει αέρα και σωματίδια στον θάλαμο μέτρησης σε κάθε αισθητήρα. Ο αέρας με τα σωματίδια περνούν μέσα από μια δέσμη λέιζερ μήκους κύματος 680 ± 10 nm. Το φως της δέσμης σκεδάζεται από τα σωματίδια και προσπίπτει σε έναν ανιχνευτή. Το σήμα του ανιχνευτή χρησιμοποιείται για τον υπολογισμό των συγκεντρώσεων PM_1 , $PM_{2.5}$ και PM_{10} σε μικρογραμμάρια ανά κυβικό μέτρο ($\mu\text{g}/\text{m}^3$) με έναν αποκλειστικό αλγόριθμο.

Όταν ο αισθητήρας συνδέεται στο διαδίκτυο, στέλνει τα δεδομένα στο αποθετήριο δεδομένων της PurpleAir και παρουσιάζονται σε έναν χάρτη. Οι χρήστες μπορούν να ορίσουν αν τα δεδομένα του αισθητήρα τους θα είναι διαθέσιμα στο κοινό ή θα είναι ιδιωτικά. Κάθε μονάδα PurpleAir περιλαμβάνει και αισθητήρες σχετικής υγρασίας (Relative Humidity – RH) και θερμοκρασίας T.



Εικόνα 2. Διάγραμμα περιγραφής της αρχής λειτουργίας του αισθητήρα Plantower PMS5003 sensor (Πηγή: 2016 product data manual of PLANTOWER)



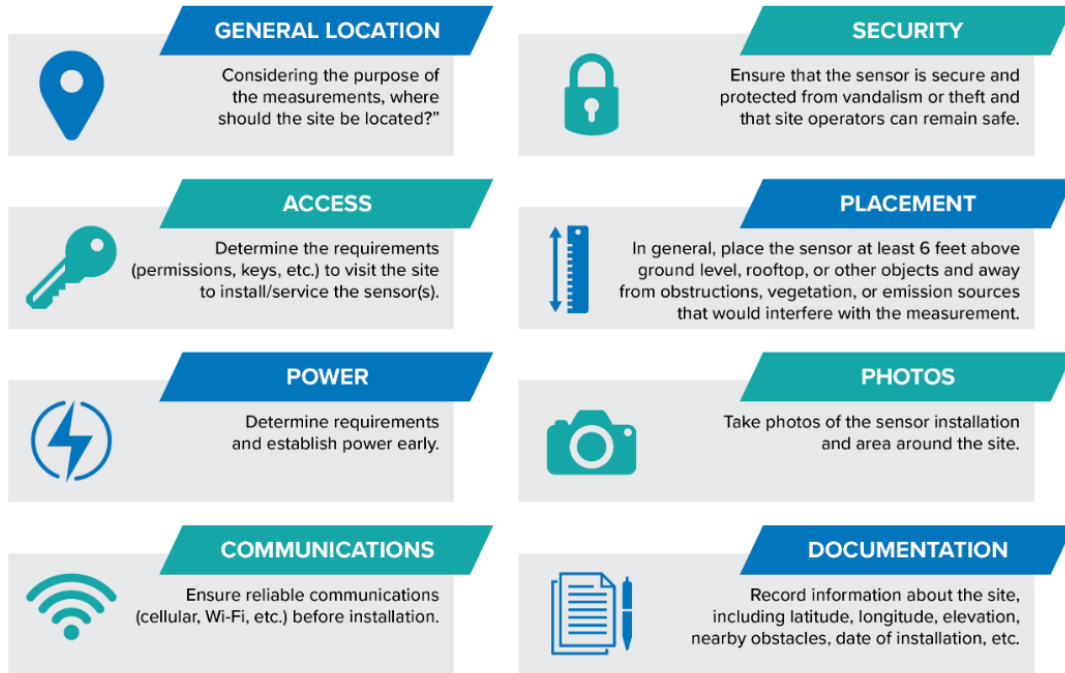
Εικόνα 3. (a) Κάτω μέρος του αισθητήρα (b) Εσωτερικό του αισθητήρα PMS5003. Ένας ανεμιστήρας αντλεί τα σωματίδια μέσω της εισροής (στρογγυλές οπές) στο χαμηλότερο επίπεδο του αισθητήρα. Τα σωματίδια κινούνται προς το πάνω μέρος του αισθητήρα, περνούν μέσα από τη διαδρομή της δέσμης λέιζερ προκαλώντας σκέδαση της δέσμης. Τέλος εξέρχονται από τον ανεμιστήρα.

ΕΓΚΑΤΑΣΤΑΣΗ ΑΙΣΘΗΤΗΡΑ

Το σημείο εγκατάστασης του αισθητήρα εξαρτάται από τον σκοπό της έρευνας. Εάν για παράδειγμα ο σκοπός είναι να βρεθούν οι συνήθεις απαντώμενες (αντιπροσωπευτικές) συγκεντρώσεις σε μια αστική περιοχή, τότε θα πρέπει να αποφευχθεί η εγκατάσταση σε κάποιον πολυσύχναστο δρόμο ή δίπλα σε μια πηγή όπως μια καμινάδα ή μια ψησταριά ή όπου εκτελούνται δομικά έργα. Για μελέτες έκθεσης πληθυσμού, τα συστήματα αισθητήρων θα πρέπει να αναπτύσσονται σε τοποθεσίες με διαφορετικά χαρακτηριστικά πληθυσμού, π.χ. πολυσύχναστους δρόμους, εσωτερικούς και εξωτερικούς χώρους, αγροτικές τοποθεσίες και τοποθεσίες υποβάθρου. Η διάρκεια της συλλογής δεδομένων πρέπει να επιλέγεται ώστε να καλύπτει ημερήσιες, εβδομαδιαίες και εποχιακές κατανομές των ρύπων (Yatkin et al, 2022).

Εκτός των ανωτέρω, υπάρχουν και πρακτικά θέματα που θα πρέπει να αντιμετωπισθούν όπως η ύπαρξη δικτύου και ρεύματος, η προσβασιμότητα και η ασφάλεια του σημείου, θέματα τεκμηρίωσης και άλλα που συνοψίζονται στην εικόνα 4.

Logistical Considerations for Finding a Place to Locate Air Sensors



Learn more at: <https://www.epa.gov/air-sensor-toolbox>

Εικόνα 4. Υλικοτεχνικές εκτιμήσεις που θα πρέπει να γίνουν προκειμένου να βρεθεί ο κατάλληλος χώρος τοποθέτησης του αισθητήρα (Πηγή: EPA, 2022)

Πριν ρυθμίσετε τον αισθητήρα σας, θα χρειαστείτε:

- το αναγνωριστικό συσκευής του αισθητήρα (device ID),
- το email που σχετίζεται με τον αισθητήρα και δόθηκε κατά την αγορά αυτού (περισσότερα στον σύνδεσμο <https://community.purpleair.com/t/associated-email-vs-owners-email/643>),
- ένα τοπικό δίκτυο WiFi και
- Μια συσκευή με δυνατότητα σύνδεσης σε δίκτυο WiFi, όπως τηλέφωνο ή υπολογιστής.

Η διαδικασία εγκατάστασης περιγράφεται λεπτομερώς στα παρακάτω βήματα.

1ο βήμα: Συνδέστε το καλώδιο στο τροφοδοτικό ρεύματος. Ελέγξτε εάν χρειάζεστε ευρωπαϊκό ανάπτερο, επειδή το παρεχόμενο καλώδιο έχει αμερικανική έξοδο βύσματος. Εάν όλα είναι εντάξει, μια λυχνία θα ανάψει στο κάτω μέρος του αισθητήρα (εικόνα 5).



Εικόνα 5. Το κάτω μέρος του αισθητήρα PurpleAir. Η λυχνία που ανάβει όταν ο αισθητήρας βρίσκεται σε λειτουργία επισημαίνεται με κόκκινο κύκλο.

2ο βήμα: Μετά από λίγα λεπτά θα δημιουργηθεί ένα hotspot του αισθητήρα με το όνομα PurpleAir-**** (οι αστερίσκοι θα είναι ένας κωδικός 2-4 χαρακτήρων που καθορίζεται από τους τελευταίους χαρακτήρες του αναγνωριστικού συσκευής (device ID) του αισθητήρα). Ανοίξτε τις ρυθμίσεις δικτύου στη συσκευή σας (τηλέφωνο, φορητός υπολογιστής κ.λπ.) και συνδεθείτε στο hotspot του αισθητήρα.

3ο βήμα: Συνήθως εμφανίζεται ένα αναδυόμενο παράθυρο. Εάν όχι, ανοίξτε ένα πρόγραμμα περιήγησης ιστού (browser) και πληκτρολογήστε "http://192.168.4.1/config " στη γραμμή διευθύνσεων. Εάν εξακολουθείτε να μην μπορείτε να μεταβείτε στην παραπάνω σελίδα, δοκιμάστε να θέσετε προσωρινά σε παύση ή να απενεργοποιήσετε τα δεδομένα στη συσκευή σας και να εισαγάγετε ξανά την παραπάνω διεύθυνση URL.

4ο βήμα: Επιλέξτε την καρτέλα "Ρυθμίσεις Wi-Fi". Συνήθως εμφανίζονται αρκετά δίκτυα Wi-Fi, από τα οποία θα πρέπει να επιλέξετε αυτό στο οποίο θέλετε να συνδεθεί ο αισθητήρας. Γράψτε τον κωδικό πρόσβασης (password) του δικτύου WI-Fi στον συγκεκριμένο χώρο στο κάτω μέρος και κάντε κλικ στο " Save " (Αποθήκευση).

5ο βήμα: Ο αισθητήρας σας θα αρχίσει να προσπαθεί να συνδεθεί στο επιλεγμένο δίκτυο Wi-Fi. Αυτή η διαδικασία μπορεί να διαρκέσει μερικά λεπτά. Μόλις η σύνδεση είναι επιτυχής, το μήνυμα στο επάνω μέρος της σελίδας θα λέει, "Looking Good" και το σήμα θα είναι πράσινο. Από τώρα και στο εξής δεν θα βλέπετε πλέον το δίκτυο PurpleAir-**** στις ρυθμίσεις δικτύου σας.

ΑΝΤΙΜΕΤΩΠΙΣΗ ΠΡΟΒΛΗΜΑΤΩΝ ΑΙΣΘΗΤΗΡΑ

Σε αυτή την ενότητα παρέχονται χρήσιμες πληροφορίες με τη μορφή ερωτήσεων και απαντήσεων για να βοηθήσουν τους χρήστες στην αντιμετώπιση προβλημάτων που σχετίζονται με τη ρύθμιση, συντήρηση και χρήση των αισθητήρων PurpleAir.

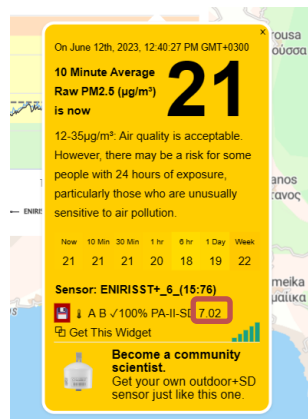
Ερώτηση 1. Δεν μπορώ να συνδέσω τον αισθητήρα στο δίκτυο Wi-Fi.

Πρώτον, θα πρέπει να ελέγξετε ότι το δίκτυό σας Wifi είναι 2.4GHz (συνήθως οικιακά και προσωπικά δίκτυα). Δεύτερον, πρέπει να δώσετε τον κωδικό πρόσβασης του δικτύου σας στις "Ρυθμίσεις Wi-Fi" που αναφέρονται στην ενότητα ΕΓΚΑΤΑΣΤΑΣΗ ΑΙΣΘΗΤΗΡΑ.

Εάν τίποτα από τα παραπάνω δεν επιλύει το πρόβλημα, πιθανώς έχετε χάσει την πρόσφατη ενημέρωση λογισμικού. Επομένως, πρέπει να ελέγξετε την έκδοση του αισθητήρα σας (εικόνα 6, η έκδοση αυτού του αισθητήρα είναι 7.02). Κλικάροντας το σημείο ενός άλλου αισθητήρα στον χάρτη PurpleAir ([Real-Time Air Quality Map | PurpleAir](#)) μπορείτε να δείτε την έκδοση που εμφανίζεται στο πλαίσιο πληροφοριών του αισθητήρα. Σε περίπτωση που η έκδοση του αισθητήρα σας δεν είναι η ίδια θα πρέπει να την ενημερώσετε, ακολουθώντας τα παρακάτω βήματα:

Συνδέστε τον αισθητήρα σας στο δίκτυο Wi-Fi του κινητού σας τηλεφώνου (πρέπει να χρησιμοποιήσετε τον κωδικό πρόσβασης του δικτύου του κινητού σας τηλεφώνου).

Κρατήστε τη σύνδεση για περίπου 1.30 ώρα ώστε να ολοκληρωθεί η σχετική ενημέρωση. Βεβαιωθείτε ότι η έκδοση που εμφανίζεται στο πλαίσιο πληροφοριών του αισθητήρα σας είναι η ίδια με τους υπόλοιπους αισθητήρες που εμφανίζονται στο [Real-Time Air Quality Map | PurpleAir](#).



Εικόνα 6. Το πλαίσιο πληροφοριών που εμφανίζεται κάνοντας κλικ στον κύκλο ενός αισθητήρα στον χάρτη PurpleAir ([Real-Time Air Quality Map | PurpleAir](#)). Η έκδοση του αισθητήρα επισημαίνεται με κόκκινο τετράγωνο.

Σε περίπτωση που θέλετε να συνδέσετε τον αισθητήρα σας σε [captive portal networks](#), όπως καφετέριες, πανεπιστήμια και άλλες τοποθεσίες που χρησιμοποιούν πύλες υποδοχής, θα χρειαστείτε ειδική εξουσιοδότηση από τον διαχειριστή του δικτύου για να επιτρέψει στην εταιρία PurpleAir να επικοινωνήσει με το συγκεκριμένο δίκτυο (διαβάστε περισσότερα στο [Sensor Won't Connect to your WiFi Network - Sensors / Troubleshooting - PurpleAir Community](#)).

Ερώτηση 2. Δεν μπορώ να δω το δίκτυο PurpleAir-xxx.

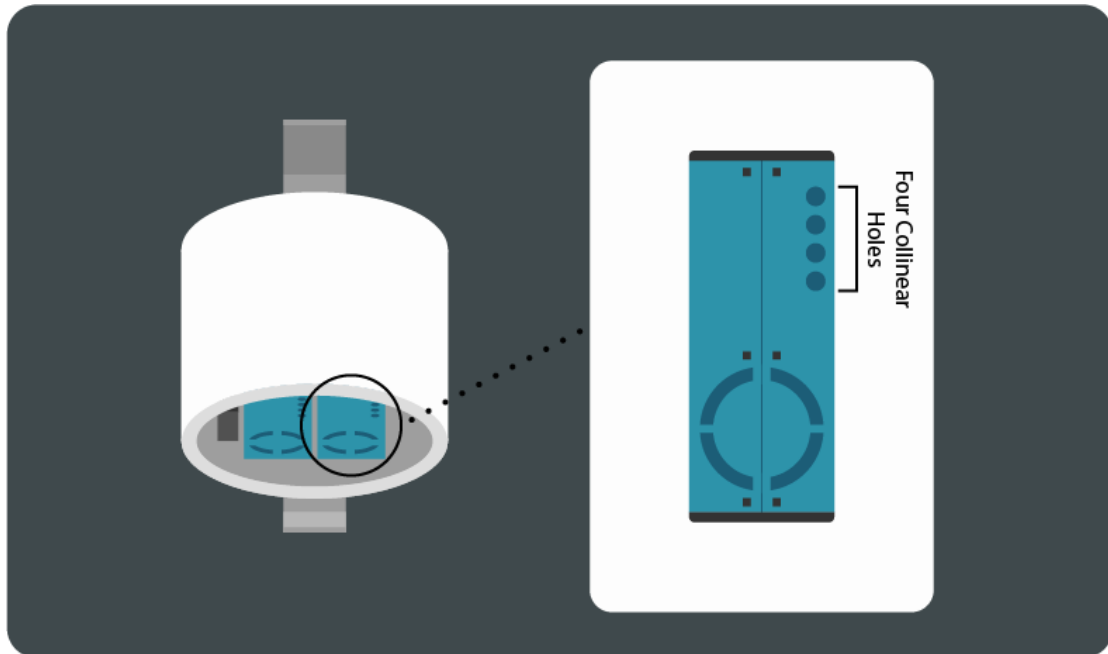
Αφού συνδέσετε τον αισθητήρα σας, ελέγξτε ότι η κόκκινη λυχνία στο κάτω μέρος του αισθητήρα σας είναι αναμμένη και μπορείτε να ακούσετε τη μηχανική δραστηριότητα του ανεμιστήρα. Εάν τα ανωτέρω είναι αρνητικά, ελέγξτε ότι το καλώδιο τροφοδοσίας λειτουργεί σωστά συνδέοντάς το σε άλλη συσκευή. Σε κάθε άλλη περίπτωση, παρακαλούμε επικοινωνήστε με την PurpleAir (contact@purpleair.com).

Ερώτηση 3. Υπάρχει μεγάλη διαφορά στις τιμές που μετράνε τα δύο κανάλια ή ένα από τα κανάλια μετρά μια επαναλαμβανόμενη τιμή ή μια τιμή που είναι κοντά στο μηδέν.

Είναι αναμενόμενο τα σωματίδια που μετρώνται από το κανάλι A να είναι λίγο διαφορετικά από αυτά που μετρώνται από το κανάλι B. Ωστόσο, εάν αυτή η διαφορά είναι μεγαλύτερη από 5 µg/m³ για περισσότερο από 5 ώρες ή ένα από τα κανάλια μετράει τις ίδιες τιμές (π.χ. 1.1,

1.4, 1.2 κ.λπ.) για μεγάλο χρονικό διάστημα, πρέπει να ελέγξετε εάν ένας από τους μετρητές λέιζερ στον αισθητήρα σας έχει γεμίσει με γύρη, έντομα ή σκόνη. Σε αυτή την περίπτωση είναι απαραίτητο να καθαρίσετε τους μετρητές λέιζερ με πεπιεσμένο αέρα. Για αισθητήρες PA-II και PA-II-SD, πρέπει να εστιάσετε στις τέσσερις οπές (Four Collinear Holes) σε κάθε μετρητή λέιζερ (βλ. εικόνα 7). Εάν το πρόβλημα παραμένει, πρέπει να αλλάξετε τον μετρητή λέιζερ. Μπορείτε να αγοράσετε το σωστό από το ηλεκτρονικό κατάστημα PurpleAir ([online store](#)). Για τη διαδικασία αντικατάστασης, τα παρακάτω βίντεο παρέχουν τις απαραίτητες οδηγίες (πηγή: [Sensor Maintenance - Sensors / Troubleshooting - PurpleAir Community](#)):

- PA-I: [PA-I-Indoor Laser Counter Replacement Video.mp4 - Google Drive](#)
- PA-II or PA-II-SD: [PA-II-SD Laser Counter Replacement.mp4 - Google Drive](#)
- PurpleAir Flex: [PA-II-FLEX Laser Counter Replacement.mp4 - Google Drive](#)
- PurpleAir Zen: [PurpleAir Zen Laser Counter Replacement.mp4 - Google Drive](#)



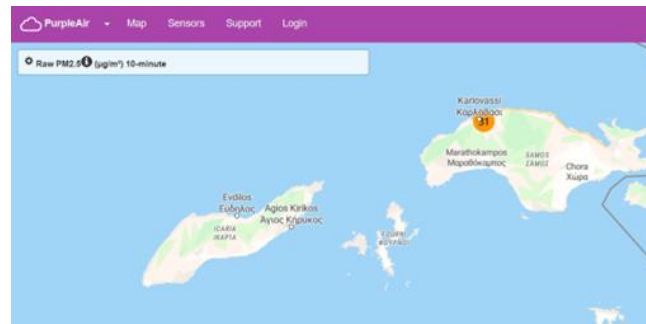
Εικόνα 7. Οι τέσσερις οπές (Four Collinear Holes) που πρέπει να καθαρίζονται με πεπιεσμένο αέρα σε περίπτωση που οι μετρητές λέιζερ έχουν γεμίσει με γύρη, υπολείμματα εντόμων ή σκόνη (πηγή: [Sensor Maintenance - Sensors / Troubleshooting - PurpleAir Community](#))

ΛΗΨΗ ΔΕΔΟΜΕΝΩΝ

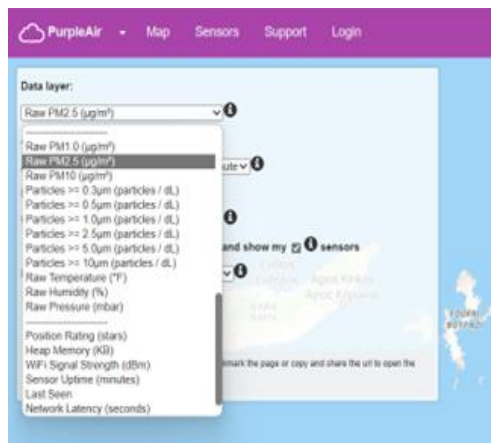
Τα δεδομένα που παρουσιάζονται στον χάρτη PurpleAir καθώς και τα ιστορικά δεδομένα μπορούν να μεταφορτωθούν για περαιτέρω χρήση. Υπάρχουν τρεις τρόποι λήψης των δεδομένων που καταγράφονται από κάθε αισθητήρα.

1) Λήψη μέσω του χάρτη

Αφού ζουμάρετε στον αισθητήρα για τον οποίο θέλετε να κατεβάσετε δεδομένα (εικόνα 8α), μπορείτε να κάνετε κλικ στη γραμμή (πάνω αριστερά, εικόνα 8β) για να επιλέξετε στο «Data layer» την παράμετρο που θα εμφανίζεται στον χάρτη (π.χ. Raw PM_{2.5}, Raw PM₁₀, Humidity κ.λπ.) καθώς και τη μέση περίοδο (average period, εικόνα 8γ).



(α)



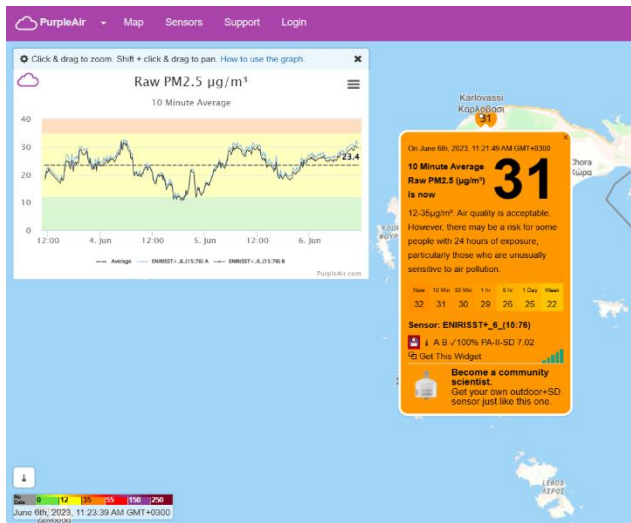
(β)



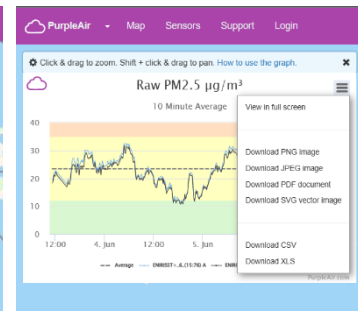
(γ)

Εικόνα 8. Διαδικασία λήψης δεδομένων από τον χάρτη [Real-Time Air Quality Map | PurpleAir](https://purpleair.com/).

Κάνοντας κλικ στον κύκλο του αισθητήρα, εμφανίζεται το πλαίσιο πληροφοριών του αισθητήρα (το πορτοκαλί πλαίσιο στην εικόνα 9α) χρωματισμένο ανάλογα με τη γραμμή υπομνήματος που φαίνεται κάτω αριστερά στο χάρτη (στο παράδειγμα είναι πορτοκαλί επειδή η μέση συγκέντρωση PM_{2.5} 10-λέπτου είναι 31 $\mu\text{g}/\text{m}^3$ και ανήκει στην τρίτη κατηγορία του υπομνήματος). Το γράφημα (εικόνα 9β) δείχνει τα δεδομένα για την επιλεγμένη περίοδο. Η διακεκομμένη γραμμή παρουσιάζει τη μέση τιμή για αυτήν την περίοδο. Μετακινώντας το ποντίκι πάνω από το γράφημα ή πατώντας σε ένα συγκεκριμένο σημείο στο γράφημα θα εμφανιστεί η τιμή, η ημερομηνία και η ώρα. Οι διαθέσιμες επιλογές για να κατεβάσετε μια εικόνα του γραφήματος είναι ως PNG, JPG, PDF ή SVG. Τα συγκεκριμένα δεδομένα μπορούν επίσης να μεταφορτωθούν είτε σε μορφή CSV είτε σε μορφή XLS κάνοντας κλικ στις τρεις οριζόντιες γραμμές (μενού hamburger) (βλ. εικόνα 9β). Ανάλογα με την περίοδο υπολογισμού του μέσου όρου που έχετε επιλέξει, υπάρχει ένα όριο στις τιμές που περιλαμβάνονται στο ληφθέν αρχείο csv ή xls. Όταν επιλέγετε "Reporting or modified within: All-time", η περίοδος λήψης των τιμών είναι όπως παρουσιάζεται στον πίνακα 1, ενώ οι τιμές για τους δύο αισθητήρες παρουσιάζονται ξεχωριστά. Οι συγκεντρώσεις στο αρχείο λήψης παρουσιάζονται σε τοπική ώρα (για την Ελλάδα δηλαδή UTC +2 για τη χειμερινή ώρα και UTC + 3 για τη θερινή ώρα). Πιο συγκεκριμένα, γράφονται τέσσερις στήλες. 1η στήλη: τοπική ώρα, 2η στήλη η μέση τιμή, 3η στήλη: συγκέντρωση για το κανάλι A και 4η στήλη: συγκέντρωση για το κανάλι B.



(α)



(β)

Εικόνα 9. Το πλαίσιο πληροφοριών καθώς και το γράφημα που εμφανίζονται επιλέγοντας τον κύκλο που αντιπροσωπεύει τον συγκεκριμένο αισθητήρα.

Πίνακας 1. Η χωρητικότητα των αρχείων που λαμβάνονται σύμφωνα με την επιλεγμένη μέση περίοδο συγκεντρώσεων (Averaging period).

Χρονική Περίοδος	Αριθμός τιμών στο αρχείο λήψης
Real-time	1140 τιμές
10-minute	περίπου 432 τιμές
30-minute	περίπου 336 τιμές
1-hour	περίπου 336 τιμές
6-hour	περίπου 360 τιμές
1-day	365 τιμές (the last year)
1-week	περίπου 79 τιμές
1-Month	68 τιμές (περίπου από την εγκατάσταση του αισθητήρα)
1-Year	Συμπεριλαμβάνονται όλα τα έτη που ο αισθητήρας λειτουργεί

Για τις υπόλοιπες επιλογές είναι απαραίτητη η ύπαρξη API Keys. Για να μπορέσει ο χρήστης να αλληλεπιδράσει με το PurpleAir API θα πρέπει να ζητήσει τα προσωπικά του API Keys στέλνοντας αίτημα στο μείλ contact@purpleair.com. Υπάρχουν δύο τύποι API Keys:

READ keys χρησιμοποιούνται για την ανάγνωση καταχωρήσεων.

WRITE keys χρησιμοποιούνται για την προσθήκη, ενημέρωση ή διαγραφή καταχωρήσεων.

2) Λήψη χρησιμοποιώντας το PurpleAir API

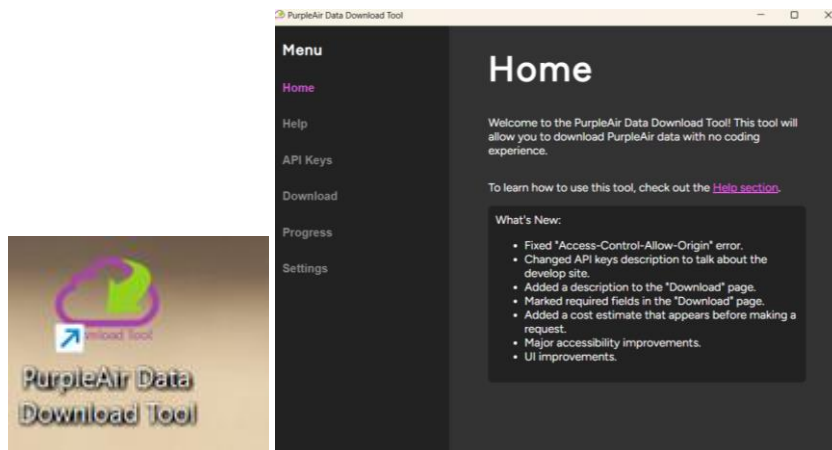
Σε περίπτωση που θέλετε να πραγματοποιήσετε κλήσεις API με το PurpleAir API, παρακαλούμε διαβάστε τις οδηγίες που δίνονται στον παρακάτω σύνδεσμο: [Making API Calls with the PurpleAir API - Data / API - PurpleAir Community](#)

3) Εργαλείο λήψης δεδομένων PurpleAir (PurpleAir Data Download Tool)

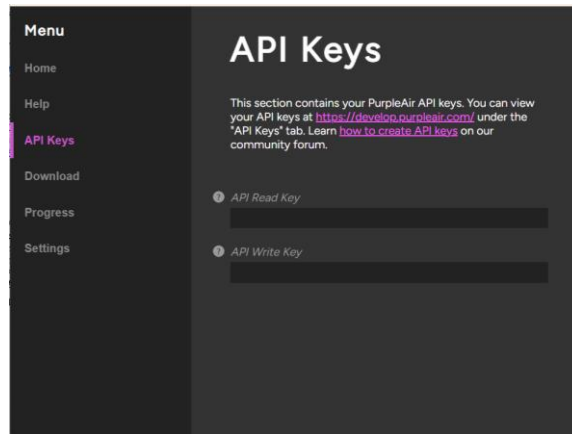
Πιθανώς ο ευκολότερος τρόπος για να κατεβάσετε δεδομένα, για μεγαλύτερο χρονικό διάστημα από αυτό που παρέχεται από το γράφημα του χάρτη, είναι το PurpleAir Data Download Tool το οποίο είναι ένα λογισμικό ανοιχτού κώδικα. Ανάλογα με το λειτουργικό σύστημα του υπολογιστή σας μπορείτε να ακολουθήσετε τους παρακάτω συνδέσμους για να κατεβάσετε και να εγκαταστήσετε το πρόγραμμα ([PurpleAir Data Download Tool - General / Announcements - PurpleAir Community](#)).

- MacOS
- Windows (32 bit)
- Windows (64 bit)

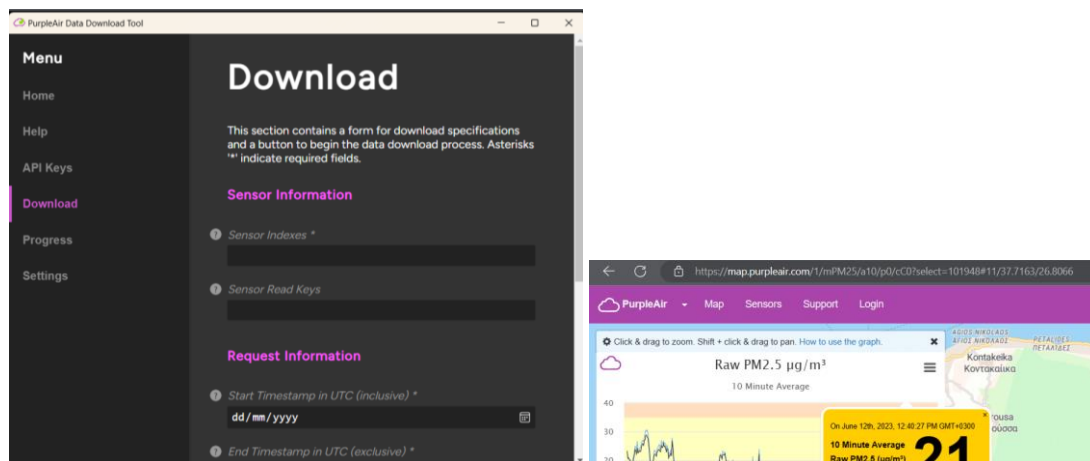
Στη συνέχεια, πρέπει να ανοίξετε το φάκελο όπου έγινε λήψη του αρχείου και να εκτελέσετε το εκτελέσιμο αρχείο κάνοντας κλικ δύο φορές. Εάν εμφανιστεί ένα αναδυόμενο μήνυμα από το Windows Defender που αναφέρει ότι το λογισμικό είναι κακόβουλο, κάντε κλικ στο "περισσότερες πληροφορίες" και, στη συνέχεια, "εκτελέστε ούτως ή άλλως". Αυτό το μήνυμα εμφανίζεται επειδή το λογισμικό δεν είναι γνωστό στη βάση δεδομένων του Windows Defender. Ένα εικονίδιο βρίσκεται τώρα στην επιφάνεια εργασίας σας. Κάνοντας δεξί κλικ ανοίγει το PurpleAir Data Download Tool. Τα παρακάτω βήματα θα σας καθοδηγήσουν στη χρήση του.



1ο βήμα: Στη γραμμή στα αριστερά, κάντε κλικ στην καρτέλα API Keys. Σε αυτή την ενότητα πρέπει να γράψετε τα προσωπικά σας API Read and Write Keys (αυτά που λάβατε από το email επικοινωνίας με την εταιρεία PurpleAir).

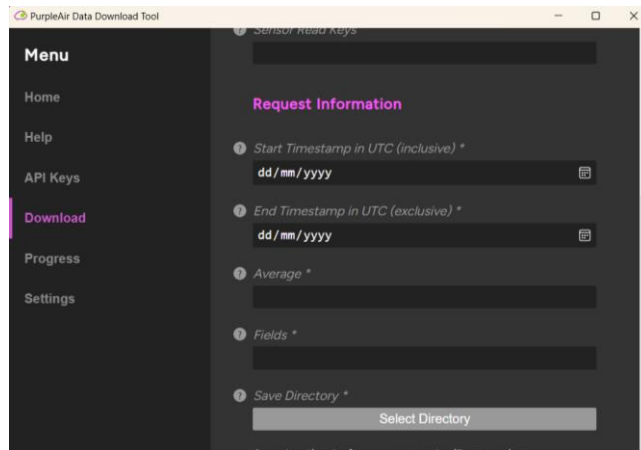


2ο βήμα: Κάντε κλικ στην καρτέλα *Download* (Λήψη) και συμπληρώστε τον κωδικό του αισθητήρα που θέλετε να κατεβάσετε δεδομένα. Μπορείτε να βρείτε τον κωδικό του αισθητήρα στο url (μετά τη λέξη "select=") κάνοντας κλικ στον αισθητήρα στο χάρτη. Για πολλαπλούς κωδικούς διαχωρίστε τους αριθμούς με κόμματα (π.χ. 101948,101361, προσέξτε να μην χρησιμοποιήσετε κενό μεταξύ των αριθμών!). Για δημόσιους αισθητήρες δεν χρειάζεται να συμπληρώσετε τα "Sensor read Keys".

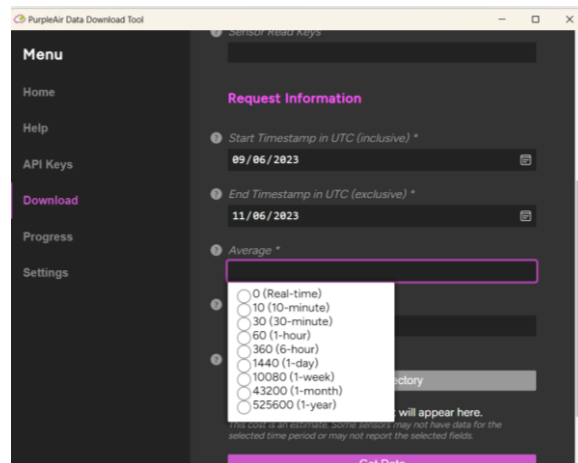


3ο βήμα – Request information (Ζητήστε πληροφορίες):

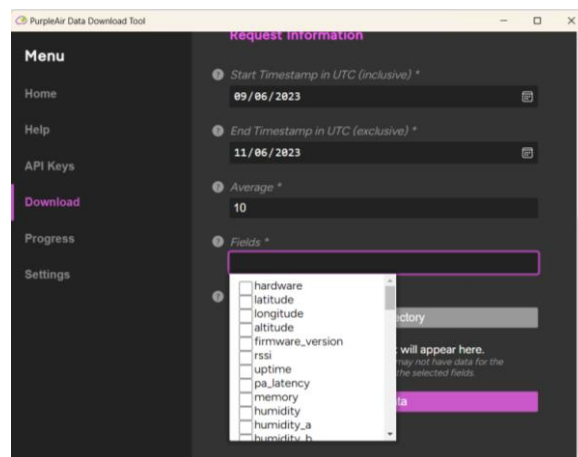
- Συμπληρώστε το χρονικό διάστημα έναρξης (start timestamp) και λήξης (end timestamp) του αιτήματός σας. Προσοχή!!! Η περίοδος λήψης δεδομένων προσέξτε πρέπει να δηλωθεί σε UTC!!! Μπορείτε επίσης να κάνετε κλικ στο εικονίδιο ημερολογίου στα δεξιά.



- Ο μέσος όρος (Average) αντιπροσωπεύει την περίοδο που θέλετε να υπολογιστεί η μέση συγκέντρωση. Κάνοντας κλικ στην καρτέλα εμφανίζονται πολλαπλές επιλογές.



- Τα πεδία (Fields) αντιπροσωπεύουν την παράμετρο για την οποία θέλετε να κάνετε λήψη δεδομένων. Μπορείτε να επιλέξετε περισσότερα από ένα πεδία με ένα αίτημα (κάνοντας κλικ στην καρτέλα). Ομοίως, πολλά πεδία πρέπει να διαχωρίζονται με κόμματα και χωρίς κενό μεταξύ τους. Η επεξήγηση κάθε πεδίου μπορεί να βρεθεί εδώ [API - PurpleAir](#) στην καρτέλα "Sensors - Get Sensors Data" (μετακινηθείτε προς τα κάτω για να βρείτε το πεδίο που σας ενδιαφέρει).



- Τέλος, θα πρέπει να επιλεγεί ο φάκελος όπου θα αποθηκευτεί το ληφθέν αρχείο (Save directory).
- Κάντε κλικ στην καρτέλα "Get Data" και θα ξεκινήσει η διαδικασία λήψης.

Χρήσιμη σημείωση 1: Οι τιμές που κατεβαίνουν μέσω του γραφήματος στον χάρτη είναι σε τοπική ώρα, ωστόσο στο Tool πρέπει να συμπληρώσετε την προτιμώμενη ώρα σε UTC.

Χρήσιμη σημείωση 2: Η εταιρεία PurpleAir έχει εφαρμόσει ένα σύστημα βασισμένο σε πόντους για τη χρήση μέσω του API. Η προβολή και η λήψη δεδομένων στον χάρτη PurpleAir παραμένει δωρεάν. Ωστόσο, εάν θέλετε να κατεβάσετε δεδομένα μέσω του εργαλείου *PurpleAir Data Download Tool*, πρέπει να επικοινωνήσετε με την PurpleAir (contact@purpleair.com) για να ενημερώσετε το λογαριασμό σας προσθέτοντας πόντους για να ξεκινήσετε (θα πρέπει να δώσετε τα κλειδιά API που αντιπροσωπεύουν τον "οργανισμό" σας). Τα API Keys σας θα λειτουργούν μόνο όταν προστεθούν πόντοι στον "οργανισμό" σας. Μπορείτε να προσθέσετε αυτούς τους πόντους στο [PurpleAir Develop](https://community.purpleair.com/t/new-api-online-dashboard/3981). Μπορείτε να βρείτε περισσότερες πληροφορίες σχετικά με την προσθήκη πόντων και τη χρήση του πίνακα ελέγχου εδώ: <https://community.purpleair.com/t/new-api-online-dashboard/3981>.

ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ

Μέγεθος αρχείου και απαιτήσεις αποθήκευσης δεδομένων

Είτε επιλέξετε να κατεβάσετε τα δεδομένα σε μορφή .csv είτε σε μορφή .xls, το μέγεθος του αρχείου είναι πολύ μικρό (μέγιστο 220 KB/αισθητήρα). Σε περίπτωση που επιλέξετε να κατεβάσετε το γράφημα ως εικόνα (PNG, JPG, PDF ή SVG) το μέγεθος του αρχείου είναι μικρότερο από 110 KB ανά αισθητήρα.

Επιθεώρηση ενός νέου αρχείου δεδομένων

Κατά τη λήψη δεδομένων από την πλατφόρμα PurpleAir, πρέπει να διερευνηθούν τα ακόλουθα:

1. Ποιος είναι ο τύπος των αρχείων και πώς γίνεται η διαχείριση τους;
2. Πόσες παρατηρήσεις (γραμμές) και πόσες μεταβλητές (στήλες) υπάρχουν σε κάθε αρχείο;
3. Έχουν μεταφερθεί σωστά οι τιμές, τα ονόματα και οι ετικέτες των μεταβλητών;
4. Υπάρχουν χρονικοί προβληματισμοί στις χρονοσειρές των δεδομένων;
5. Βρίσκονται όλες οι τιμές δεδομένων (πχ οι συγκεντρώσεις) εντός εύλογου εύρους;
6. Υπάρχουν ελλείπουσες τιμές (missing values); Σε τι ποσοστό και σε ποια μοτίβα;

Διαχείριση αρχείων

Τα δεδομένα μπορούν να ληφθούν (να κατεβαστούν) από τον χάρτη PurpleAir υπό μορφή csv ή xls. Ένα αρχείο xls είναι ένας ορθογώνιος τύπος αρχείου δεδομένων. Σε αυτά τα αρχεία, κάθε σειρά αντιπροσωπεύει μια συγκεκριμένη παρατήρηση ή περίπτωση (observation or case) και κάθε στήλη αντιπροσωπεύει μια μεταβλητή (variable). Αυτός ο τύπος αρχείων μπορεί κατόπιν να εισαχθεί σε ένα πιο σύνθετο στατιστικό πρόγραμμα.

Είναι πολύ σημαντικό να αποθηκευτεί κάπου στον υπολογιστή το αρχικό αρχείο δεδομένων, χωρίς καμία επεξεργασία. Επιπλέον, μετά από κάθε σημαντική επεξεργασία, καλό είναι να αποθηκεύεται μια νέα έκδοση του αρχείου. Αυτό γιατί είναι πιθανόν να αναιρέσουμε κάποια απόφαση επεξεργασίας και να χρειαζόμαστε την προηγούμενη έκδοση ή γιατί μπορεί κάποια έκδοση του αρχείου να καταστραφεί.

Αριθμός παρατηρήσεων και μεταβλητών

Θα πρέπει να γνωρίζετε πόσες παρατηρήσεις (ή περιπτώσεις) αναμένεται να υπάρχουν στο αρχείο δεδομένων που έχετε κατεβάσει. Για παράδειγμα, στην περίπτωση ωριαίων μέσων όρων, για μετρήσεις που λαμβάνονται για 15 ημέρες, αναμένεται ότι το αρχείο θα περιέχει $24 \times 15 = 360$ σειρές (ή περιπτώσεις) συν τις επικεφαλίδες. Επιπλέον πρέπει να επιβεβαιώνετε ότι ο σωστός αριθμός μεταβλητών (στηλών) περιλαμβάνονται στο αρχείο. Τέλος είναι καλή πρακτική να ελέγχετε εάν ο αριθμός σειρών και μεταβλητών είναι σωστός μετά από κάθε γύρο επεξεργασίας των δεδομένων.

Σε περίπτωση που το αρχείο excel πρόκειται να εισαχθεί σε ένα στατιστικό πρόγραμμα (όπως το SPS), θα πρέπει να αναζητηθούν οι κανόνες ονοματολογίας των μεταβλητών σε κάθε πρόγραμμα ώστε τα ονόματα τους να είναι συμβατά με όλα τα προγράμματα που θα χρησιμοποιηθούν.

Θέματα σχετικά με τον χρόνο

Σωστή ημερομηνία: Γενικά μια ημερομηνία αποθηκεύεται ως ένας αριθμός ημερών (ή άλλης μονάδας χρόνου) σε σχέση με μια ημερομηνία αναφοράς. Πχ το Excel αποθηκεύει τις ημερομηνίες με τη μορφή διαδοχικών σειριακών αριθμών, όπου κάθε αριθμός αντιστοιχεί σε μια μέρα, σε σχέση με το 1900. Τα διάφορα προγράμματα χρησιμοποιούν διαφορετικές συμβάσεις (άλλο έτος αναφοράς, άλλη μονάδα χρόνου) με αποτέλεσμα κάποιες φορές οι ημερομηνίες να μην μεταφέρονται σωστά μεταξύ προγραμμάτων.

Διαφορές ζωνών ώρας: Πολλές φορές οι συγκεντρώσεις των ρύπων αναφέρονται σε ώρα UTC (Coordinated Universal Time - Συντονισμένη Παγκόσμια Ώρα) αντί σε τοπική ώρα. Η ώρα UTC αποτελεί το διεθνές «σημείο αναφοράς χρόνου». Όλες οι ζώνες ώρας ορίζονται σε σχέση με τη μετατόπισή τους από την ώρα UTC. Για παράδειγμα, στην Ελλάδα, η τοπική ώρα είναι Ώρα Ανατολικής Ευρώπης (UTC+02:00). Η θερινή ώρα, η οποία μετακινείται μία ώρα μπροστά (UTC+03:00) εφαρμόζεται από την τελευταία Κυριακή του Μαρτίου έως την τελευταία Κυριακή του Οκτωβρίου. Αυτό σημαίνει ότι όταν μια συγκέντρωση PM_{2.5} παρατηρείται κάπου στην Ελλάδα την 1η Απριλίου, 11:40 π.μ. ώρα UTC, η τοπική ώρα Ελλάδας είναι 14:40.

Όταν κατεβάζουμε δεδομένα από διεθνείς πλατφόρμες, μερικές φορές δίνεται η δυνατότητα να τα κατεβάσουμε σε ώρα UTC ή σε τοπική ώρα. Συχνά όμως οι πλατφόρμες αυτές δεν λαμβάνουν υπόψη την σύμβαση της θερινής ώρας για την μετατροπή από UTC σε τοπική ώρα. Έτσι είναι προτιμότερο να κατεβάζουμε τα δεδομένα μας σε ώρα UTC και να κάνουμε εμείς τη μετατροπή σε τοπική ώρα Ελλάδας.

Συγχρονισμός δεδομένων: Ο συγχρονισμός των δεδομένων είναι απαραίτητος στις συγκρίσεις των μετρήσεων LCSs με τις συγκεντρώσεις οργάνων αναφοράς που λαμβάνονται σε τοπική ώρα. Επιπλέον όταν παρουσιάζονται δεδομένα μετρήσεων, είναι χρήσιμο να γίνεται αναφορά σε τοπική ώρα.

Χρονικό διάστημα μέσου όρου: Οι αισθητήρες PurpleAir δίνουν τον μέσο όρο της συγκέντρωσης σωματιδίων κάθε 2 λεπτά. Υπάρχει η επιλογή λήψης δεδομένων από τον χάρτη σε μέσους όρους 10 λεπτών, 30 λεπτών, 1 ώρας, 6 ωρών, 1 ημέρας, 1 μήνα και 1 έτους. Η επιλογή του χρονικού διαστήματος εξαρτάται από τον σκοπό της έρευνας. Εάν ο σκοπός είναι να διερευνηθούν οι συγκεντρώσεις PM λόγω βραχύβιων γεγονότων όπως η άφιξη ενός πλοίου στο λιμάνι, ή το κάπνισμα (για τη ρύπανση εσωτερικών χώρων) τότε καλό είναι να επιλεγεί ένα μικρό διάστημα (2 λεπτά ή 10 λεπτά). Σημειώστε ότι ο μέσος όρος επί 2 λεπτά πιάνει μεν βραχύβιες κορυφές αλλά μπορεί να οδηγήσει σε υψηλά επίπεδα θορύβου μέτρησης και άρα να χειροτερέψει το όριο ανίχνευσης του αισθητήρα.

Εάν ο σκοπός είναι να συγκριθούν οι μετρήσεις των αισθητήρων με τα όρια της ΕΕ ή τις κατευθυντήριες γραμμές του ΠΟΥ, τότε αρκεί η λήψη των δεδομένων σε ωριαίους ή ημερήσιους μέσους όρους, ανάλογα με τη διάμετρο των PM (PM_{2.5} ή PM₁₀).

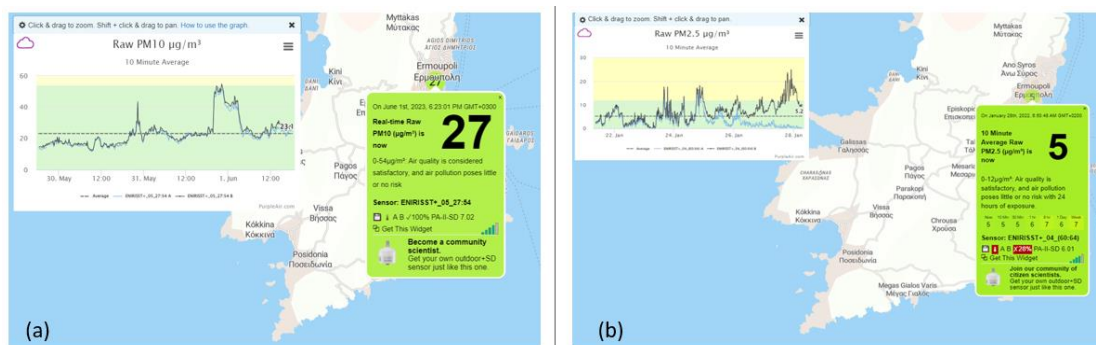
Σε ποιο χρονικό διάστημα αντιστοιχεί ο μέσος όρος; Για τους αισθητήρες PurpleAir, μια ωριαία μέση συγκέντρωση που αναφέρεται στις 14:00 είναι το αποτέλεσμα του μέσου όρου των μετρήσεων ανά διαστήματα 2 λεπτών - της επόμενης ώρας (ώρα έναρξης 14:00 – ώρα λήξης 14:58). Αυτή η πληροφορία είναι χρήσιμη όταν πρόκειται να συγχρονισθούν μετρήσεις διαφορετικών οργάνων.

Ποιότητα δεδομένων

Είναι παρόμοιες οι ενδείξεις των καναλιών A και B του αισθητήρα;

Ο αισθητήρας PA-II PurpleAir διαθέτει δύο μετρητές λέιζερ (αισθητήρες Plantower PMS 5003) που αναφέρονται ως "κανάλι A" και "κανάλι B". Υπό κανονικές συνθήκες τα δύο κανάλια θα πρέπει να δείχνουν παρόμοιες συγκεντρώσεις PM. Αν η απόδοση του αισθητήρα χειροτερέψει τότε μπορεί να παρατηρηθεί απόκλιση μεταξύ των μετρήσεων των A και B.

Στην εικόνα 10 παρουσιάζονται δεδομένα από έναν αισθητήρα στην Ερμούπολη της Σύρου, στην Ελλάδα από τον Ιούνιο 2023 (a) και τον Ιανουάριο 2022 (b). Η μπλε γραμμή στο γράφημα αντιπροσωπεύει το κανάλι A και η μαύρη το κανάλι B του αισθητήρα. Συγκρίνοντας τη συσχέτιση αυτών των δύο καναλιών δημιουργείται ένας δείκτης που ονομάζεται confidence score. Όταν η τιμή του δείκτη πλησιάζει το 100% σημαίνει ότι τα κανάλια A και B συμφωνούν και η ποιότητα της μέτρησης είναι αποδεκτή και αντίστροφα. Το confidence score στο γράφημα (a) είναι 100% που σημαίνει ότι υπάρχει καλή ομοιότητα μεταξύ των δύο καναλιών και καλή ποιότητα δεδομένων ενώ στο (b) είναι 28% υποδηλώνοντας κακή ομοιότητα και κακή ποιότητα δεδομένων.



Εικόνα 10. Τα δύο κανάλια ενός αισθητήρα PurpleAir στην Ερμούπολη της Σύρου και το confidence score (100% για την (a) περίπτωση και 28% για την περίπτωση (b))

Ένας περιορισμός αυτής της προσέγγισης είναι ότι δεν λαμβάνει υπόψη την πιθανότητα να υποβαθμιστούν και τα δύο κανάλια A και B με παρόμοιο τρόπο. Εάν υπάρχει τέτοια υποψία, συγκρίνονται οι μετρήσεις με κάποιον αισθητήρα σε γειτονικό σημείο ή με κάποιο όργανο αναφοράς.

Επιπλέον, μπορεί να πραγματοποιηθεί ποιοτικός έλεγχος των μέσων ημερήσιων συγκεντρώσεων από τα κανάλια A και B με βάση τη μεθοδολογία που προτείνουν οι Barkjohn et al (2021). Πιο συγκεκριμένα, οι ημερήσιες μέσες συγκεντρώσεις από τα δύο κανάλια συγκρίνονται με βάση την εξίσωση

$$24 \text{ h percent difference} = \frac{(A - B) \times 2}{(A + B)}$$

όπου A είναι η μέτρηση από το κανάλι A και B είναι η μέτρηση από το κανάλι B.

Σε περίπτωση που η ποσοστιαία διαφορά ήταν μεγαλύτερη από 61%, οι τιμές αυτή την ημέρα ελέγχονται περαιτέρω προκειμένου να αφαιρεθούν οι τιμές.

Δεύτερον, οι ωριαίες συγκεντρώσεις από τα δύο κανάλια ελέγχονται με βάση τη μεθοδολογία που προτείνουν οι Bi et al (2020) για την απόρριψη ακραίων τιμών. Η εξίσωση που χρησιμοποιείται είναι:

$$APB = \left| \frac{PM2.5B - PM2.5A}{PM2.5A} \right| \times 100\%$$

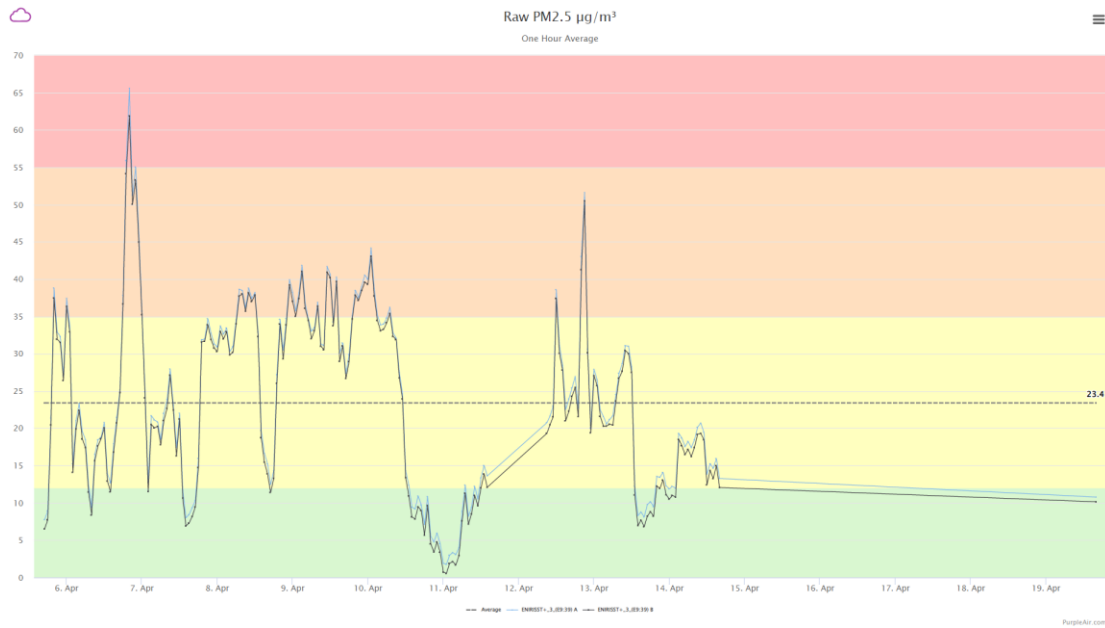
Το 5% των μετρήσεων με τις κορυφαίες τιμές APB αφαιρούνται.

Βρίσκονται τα δεδομένα στο αναμενόμενο εύρος?

Το επόμενο βήμα είναι να εξεταστεί αν οι συγκεντρώσεις του αισθητήρα φαίνονται λογικές (έχουν νόημα). Αυτό μπορεί να γίνει με επισκόπηση των διακυμάνσεων τους σε συνάρτηση με τον χρόνο είτε on line είτε αφού κατεβούν τα δεδομένα, χρησιμοποιώντας ένα λογισμικό υπολογιστικών φύλλων όπως το excel. Άλλες διαδικασίες είναι να χρησιμοποιηθεί η επιλογή «φίλτρο» του excel ή να γίνει διάγραμμα συχνοτήτων μέσω κάποιου στατιστικού πακέτου.

Ένας λόγος για μη-λογικά δεδομένα είναι η βλάβη στον ίδιο τον αισθητήρα ή στη μετάδοση δεδομένων. Για παράδειγμα, στην περίπτωση της εικόνας 11, ο αισθητήρας δεν μετέδιδε δεδομένα μεταξύ 12-13 και 14-19 Απριλίου και αυτές οι τιμές θα πρέπει να αφαιρεθούν από την περαιτέρω ανάλυση.

Σε περίπτωση που καταγράφονται εξωπραγματικά υψηλές ή χαμηλές συγκεντρώσεις για μεγάλο χρονικό διάστημα, οι μετρήσεις του αισθητήρα θα πρέπει να συγκριθούν με άλλες, από αισθητήρες που βρίσκεται κοντά και να ελεγχθεί αν εμφανίζεται και εκεί το ίδιο μοτίβο. Εάν δεν εμφανίζεται, τότε ίσως (σε περίπτωση υψηλών τιμών) μια πηγή κοντά στον αισθητήρα δίνει υψηλά επίπεδα ρύπανσης και πρέπει να διερευνηθεί περαιτέρω. Για να έχετε μια ιδέα για το ποια επίπεδα ρύπανσης θεωρούνται υψηλά ή χαμηλά, μπορείτε να συμβουλευτείτε τις συγκεντρώσεις του Ευρωπαϊκού Δείκτη Ποιότητας Αέρα (αναφέρονται ως ημερήσιοι μέσοι όροι) που αντιστοιχούν σε καλή, αρκετά καλή, μέτρια ή κακή ποιότητα αέρα (πίνακας 2).



Εικόνα 11. Διάγραμμα των συγκεντρώσεων του αισθητήρα σε συνάρτηση με τον χρόνο όπου φαίνεται αδυναμία μετάδοσης δεδομένων μεταξύ 12-13 και μετά τις 14 Απριλίου

Πίνακας 2. Ο Ευρωπαϊκός Δείκτης Ποιότητας Αέρα με 24-ώρες μέσες τιμές για τα PM₁₀ και PM_{2.5} (Source: EEA (n.d.))

Pollutant	Index level (based on pollutant concentrations in µg/m3)					
	Good	Fair	Moderate	Poor	Very poor	Extremely poor
Particles less than 2.5 µm (PM _{2.5})	0-10	10-20	20-25	25-50	50-75	75-800
Particles less than 10 µm (PM ₁₀)	0-20	20-40	40-50	50-100	100-150	150-1200

Ελλείπουσες τιμές (missing data)

Το τελευταίο βήμα πριν από την ανάλυση δεδομένων είναι η εξέταση του ποσοστού των δεδομένων που λείπουν και των μοτίβων με τα οποία συμβαίνει αυτό. Κατά την συλλογή οποιασδήποτε κατηγορίας δεδομένων, σχεδόν πάντα παρατηρούνται ελλείπουσες τιμές. Παρόλο που εμφανίζεται συχνά, το πρόβλημα αυτό είναι δύσκολο να αντιμετωπιστεί. Ο πρώτος στόχος είναι να βρει κανείς πόσα δεδομένα λείπουν. Ο δεύτερος στόχος είναι να εξεταστούν τα μοτίβα των ελλειπουσών τιμών για τις διάφορες μεταβλητές. Για παράδειγμα, οι ελλείπουσες τιμές εμφανίζονται σε συγκεκριμένες μεταβλητές ενώ σε άλλες όχι; Υπάρχουν

μεταβλητές που εμφανίζουν ελλείπουσες τιμές σε μεγάλα ποσοστά ενώ άλλες σε μικρά ποσοστά; Ποιοι είναι οι λόγοι για τους οποίους λείπουν δεδομένα;

Τα δεδομένα μπορεί να λείπουν για πολλούς λόγους που είναι χρήσιμο να καταγράφονται στο σύνολο των δεδομένων. Αυτό γίνεται με διάφορες μεθόδους, συχνά χρησιμοποιώντας συγκεκριμένους κωδικούς δεδομένων για διαφοροποίηση μεταξύ τους, πχ χρησιμοποιώντας τιμές (όπως -8 και -9) που δεν μπορούν να εμφανιστούν ως αληθινές τιμές για την εν λόγω μεταβλητή. Για τους αισθητήρες Purpleair, συνηθισμένοι λόγοι είναι η βλάβη του εξοπλισμού ή τα προβλήματα στη μετάδοση δεδομένων μέσω WiFi.

Οι ελλείπουσες τιμές δημιουργούν δύο προβλήματα: Αφενός μειώνουν τον αριθμό των παρατηρήσεων που είναι διαθέσιμες για ανάλυση, μειώνοντας έτσι τη στατιστική ισχύ (δηλ. την ικανότητά να ανιχνεύονται στατιστικά σημαντικές διαφορές στα δεδομένα) και αφετέρου μπορεί να εισάγουν συστηματικό σφάλμα (bias) στην ανάλυση. Το πρώτο σημείο βασίζεται στο γεγονός ότι, εφόσον όλα τα υπόλοιπα είναι ίδια, η στατιστική ισχύς αυξάνεται καθώς αυξάνεται ο αριθμός των παρατηρήσεων. Για το δεύτερο σημείο (bias) δίνεται παράδειγμα στις επόμενες παραγράφους.

Οι ελλείπουσες τιμές ταξινομούνται σε 3 κατηγορίες: Οι τελείως τυχαίες (missing completely at random – MCR), οι τυχαίες (missing at random – MAR) και οι μη αγνοήσιμες (nonignorable) ελλείπουσες τιμές.

MCAR σημαίνει ότι λείπει ένα μέρος των δεδομένων μιας μεταβλητής αλλά αυτό δεν σχετίζεται ούτε με τις συγκεκριμένες τιμές ούτε με τις τιμές των άλλων μεταβλητών στο συγκεκριμένο σύνολο δεδομένων. Αυτός είναι ο πιο εύκολος τύπος ελλειπουσών τιμών να αντιμετωπισθεί διότι όλες οι τιμές μπορεί να θεωρηθούν ως τυχαίο δείγμα ολόκληρου του συνόλου δεδομένων. Αυτή είναι η περίπτωση των ελλειπουσών τιμών των αισθητήρων Purpleair, καθώς οι συνήθεις λόγοι για την έλλειψη δεδομένων είναι η αστοχία του εξοπλισμού ή η αστοχία μετάδοσης δεδομένων (πρόβλημα στο δίκτυο).

Οι ελλείπουσες τιμές τύπου MAR σημαίνουν ότι το γεγονός ότι λείπουν τα δεδομένα δεν σχετίζεται με τη δική τους τιμή, αλλά σχετίζεται με τις τιμές άλλων μεταβλητών στην ανάλυση. Για παράδειγμα, ηλεκτροχημικοί αισθητήρες που μετρούν αέρια όπως NOx, O₃ ή CO παρέχουν λιγότερο ακριβή δεδομένα σε ακραίες καιρικές συνθήκες (>30 Κελσίου, <30% RH) και τέτοια δεδομένα δεν αναφέρονται από ορισμένες διαδικασίες διασφάλισης ποιότητας (π.χ. <https://www.aqmesh.com/techsupport/aqmesh-technical-support/faqs/extreme-environment-flagging/>) με αποτέλεσμα να λείπουν τιμές του τύπου MAR.

Η κατηγορία ελλειπουσών τιμών nonignorable αναφέρεται σε δεδομένα των οποίων η έλλειψη σχετίζεται με τη δική τους τιμή. Αυτή η κατηγορία είναι που μπορεί να εισάγει συστηματικό σφάλμα στην ανάλυση. Αυτό θα μπορούσε να συμβεί (για παράδειγμα) εάν ένας αισθητήρας σταματούσε να μεταδίδει δεδομένα σε περίπτωση πολύ χαμηλών ή πολύ υψηλών συγκεντρώσεων ρύπων. Οι αισθητήρες PurpleAir δεν εμφανίζουν αυτό το πρόβλημα.

Οι συνήθεις μέθοδοι στατιστικής ανάλυσης προϋποθέτουν ότι τα δεδομένα είναι πλήρη. Στην περίπτωση που ένα σύνολο δεδομένων περιέχει μεγάλο ποσοστό ελλειπουσών τιμών, θα πρέπει να αποφασιστεί ο τρόπος αντιμετώπισης του προβλήματος. Η καλύτερη λύση είναι να καταβληθεί επιπλέον προσπάθεια για τη συλλογή των δεδομένων που λείπουν. Στην περίπτωση προβλημάτων μετάδοσης δεδομένων, η έκδοση SD των αισθητήρων PurpleAir περιλαμβάνει ενσωματωμένο καταγραφικό SD (Secure Digital) για εγγραφή και αποθήκευση δεδομένων χωρίς σύνδεση WiFi. Το SD μπορεί να αποθηκεύσει δεδομένα για περίπου 2 χρόνια. Με τον τρόπο αυτό, συμπληρώνονται οι ελλείπουσες τιμές.

Μια άλλη λύση είναι να συμπεριληφθεί μια ψευδομεταβλητή (dummy variable) (0, 1) στην ανάλυσή, που υποδεικνύει τις ελλείπουσες τιμές, μαζί με μια τεκμαρτή τιμή (imputed value)

που αντικαθιστά τα δεδομένα που λείπουν. Υπάρχουν διαφορετικές μέθοδοι καταλογισμού δεδομένων (imputation) και οι πιο εξελιγμένες μπορεί να απαιτούν την βοήθεια από έναν ειδικό ή τη χρήση λογισμικού που έχει σχεδιαστεί ειδικά για την διαχείριση ελλειπουσών τιμών (π.χ. SPSS MVA). Μια απλή μέθοδος καταλογισμού είναι η αντικατάσταση των τιμών που λείπουν με μια τιμή όπως ο ωριαίος μέσος όρος μιας επιλεγμένης περιόδου (μιας εποχής ή ολόκληρου του έτους). Ωστόσο, αυτό δεν συνιστάται, καθώς σχεδόν πάντα οδηγεί σε ακραία υποεκτίμηση της διακύμανσης (variance).

Τέλος, σε περιπτώσεις υψηλού ποσοστού ελλειπουσών τιμών, η λύση μπορεί να είναι η απόρριψη των αντίστοιχων παρατηρήσεων ή μεταβλητών και η αλλαγή του σκοπού της ανάλυσης. Για παράδειγμα, εάν υπάρχουν συχνές βλάβες στη μετάδοση δεδομένων κατά τη διάρκεια του χειμώνα λόγω διακοπών ρεύματος, τότε η ανάλυση μπορεί να επικεντρωθεί στις άλλες εποχές.

Αξιολόγηση της ακρίβειας και της επαναληψιμότητας ενός αισθητήρα στο πεδίο (in situ)

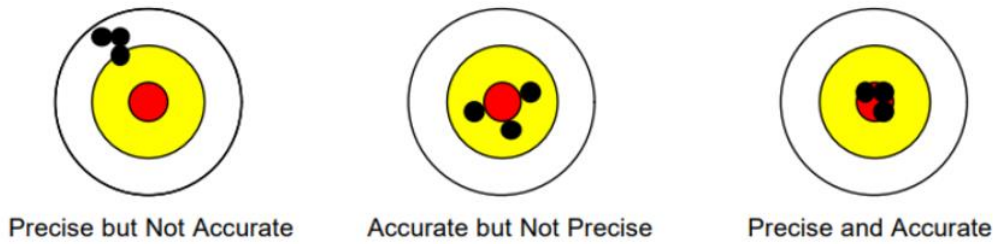
Οι αισθητήρες PurpleAir ή άλλοι ανάλογοι δεν μετρούν απευθείας τις συγκεντρώσεις μάζας των σωματιδίων (σε μικρογραμμάρια ανά κυβικό μέτρο). Χρησιμοποιούν οπτικές μεθόδους που βασίζονται στον σκεδασμό του φωτός από σωματίδια διαμέτρου (τυπικά) $> \sim 0.3 \text{ mm}$. Κατόπιν, χρησιμοποιώντας παραδοχές που μπορεί να εισάγουν σφάλματα στα αποτελέσματα, μετατρέπουν την σκέδαση του φωτός σε συγκεντρώσεις μάζας (de Souza, 2023). Η μεθοδολογία μετατροπής από κάθε κατασκευαστή αισθητήρα δεν είναι γνωστή.

Επιπλέον, είναι γνωστό ότι τα αιωρούμενα σωματίδια στην ατμόσφαιρα απορροφούν νερό καθώς αυξάνεται η σχετική υγρασία (RH) της ατμόσφαιρας (Kosmopoulos et al, 2020). Αυτό οδηγεί σε αύξηση τόσο του μεγέθους των σωματιδίων όσο και της μάζας τους. Η σκέδαση του φωτός από τα σωματίδια αυξάνεται επίσης. Όμως οι συγκεντρώσεις PM δεν πρέπει να περιλαμβάνουν αυτό το νερό που απορροφήθηκε. Τα όργανα αναφοράς είναι εξοπλισμένα με ξηραντήρες ή θερμοαντήρες για την απομάκρυνση του νερού από τα σωματίδια. Οι αισθητήρες χαμηλού κόστους, όπως οι PurpleAir, δεν περιλαμβάνουν ξηραντήρες ή θερμοαντήρες. Έτσι, οι συγκεντρώσεις PM που μετρούν οι αισθητήρες μπορεί να είναι κατά πολύ εσφαλμένες (biased) λόγω της υγροσκοπικής ανάπτυξης των σωματιδίων όταν η σχετική υγρασία περιβάλλοντος (RH) είναι υψηλή (de Souza, 2023). Ως αποτέλεσμα, οι αισθητήρες PurpleAir υπερεκτιμούν τις συγκεντρώσεις λεπτών σωματιδίων κάτω από υψηλή υγρασία σε σύγκριση με τα όργανα αναφοράς που λειτουργούν στην ίδια τοποθεσία (EPA, n.d.).

Η ακρίβεια και η επαναληψιμότητα των μετρήσεων ορίζονται ως εξής (EPA, 2001):

- Ακρίβεια: Ο βαθμός απόκλισης μιας μέτρησης από την πραγματική τιμή,
- Επαναληψιμότητα: Ο βαθμός της διακύμανσης σε επαναλαμβανόμενες μετρήσεις της ίδιας ποσότητας μιας παραμέτρου.

Για παράδειγμα, εάν πραγματοποιηθούν δέκα μετρήσεις μιας παραμέτρου την ίδια στιγμή στην ίδια τοποθεσία με την ίδια μέθοδο, η ακρίβεια υποδεικνύεται από το πόσο καλά ο μέσος όρος των δέκα μετρήσεων αντικατοπτρίζει την πραγματική συγκέντρωση και η επαναληψιμότητα υποδεικνύεται από τη διακύμανση των αποτελεσμάτων των δέκα μετρήσεων. Μια κλασική απεικόνιση ακρίβειας και επαναληψιμότητας απεικονίζεται στην εικόνα 12.



Εικόνα 12. Ακρίβεια και επαναληψιμότητα (precision and accuracy)

Η ακρίβεια και η επαναληψιμότητα των αισθητήρων του δικτύου ENIRISST+ εξετάστηκε στα πλαίσια του έργου, με χρήση ωριαίων μέσων όρων των συγκεντρώσεων $PM_{2.5}$. Αρχικά ελέγχθηκε η επαναληψιμότητα των ενδείξεων των 8 αισθητήρων. Οι 8 αισθητήρες τοποθετήθηκαν μαζί (co-location) σε σημείο στα προάστια της πόλης της Χίου για περίοδο 17 ημερών (27.10.2021 – 12.11.2021). Ελλείπει σφαλμάτων, οι συγκεντρώσεις που μετρώνται από τους αισθητήρες σε ζεύγη θα πρέπει να είναι ίσες ($y=x$). Εφαρμόστηκε ένα μοντέλο γραμμικής παλινδρόμησης για να απεικονίσει την πραγματική σχέση μεταξύ των μετρήσεων των οκτώ αισθητήρων, ανά ζευγή. Ο συντελεστής προσδιορισμού R^2 ήταν $\geq 0,9976$, οι κλίσεις της γραμμικής παλινδρόμησης κυμαίνονταν μεταξύ 0,950-1,04 και ο σταθερός όρος ήταν κοντά στο μηδέν. Αυτά τα ευρήματα υποδηλώνουν καλή επαναληψιμότητα μεταξύ των μετρήσεων των ζευγών αισθητήρων.

Στη συνέχεια, επιλέχθηκε τυχαία ένας από τους οκτώ αισθητήρες και τοποθετήθηκε δίπλα σε ένα όργανο ισοδύναμο με όργανο αναφοράς (beta attenuation monitor) στον Υπερσταθμό Παρακολούθησης Ατμοσφαιρικής Ρύπανσης Αστικού Υποβάθρου του Εθνικού Αστεροσκοπείου Αθηνών στο Θησείο της Αθήνας, για την περίοδο 23/12/2021 - 28/02/2022, με σκοπό να ελεγχθεί η ακρίβεια των αισθητήρων. Οι συγκεντρώσεις $PM_{2.5}$ που αναφέρονται από το όργανο αναφοράς θεωρούνται οι πραγματικές τιμές.

Εφαρμόστηκε γραμμική παλινδρόμηση μεταξύ των μετρήσεων του αισθητήρα και της RH (ανεξάρτητες μεταβλητές) και της συγκέντρωσης του οργάνου αναφοράς (εξαρτημένη μεταβλητή) και βρέθηκε ότι η εξίσωση που περιγράφει καλύτερα τη σχέση μεταξύ τους είναι:

$$y_{ref}=0.423*x_{CFI}-0.117*RH+11.051$$

Όπου: y_{ref} ($\mu g/m^3$) είναι η συγκέντρωση που μετράται από το όργανο αναφοράς, x_{CFI} ($\mu g/m^3$) είναι η συγκέντρωση που μετράται από τον αισθητήρα PurpleAir και RH (%) είναι η σχετική υγρασία. Ο συντελεστής προσδιορισμού R^2 του μοντέλου ήταν 0.81. Ο συντελεστής προσδιορισμού R^2 εκφράζει το ποσοστό της μεταβλητότητας των τιμών της εξαρτημένης μεταβλητής που εξηγείται, μέσω του μοντέλου, από τις ανεξάρτητες μεταβλητές. Η αλλιώς υποδεικνύει την ποιότητα προσαρμογής της γραμμής παλινδρόμησης στα δεδομένα.

Κατόπιν, εισάγοντας την συγκέντρωση του αισθητήρα x_{CFI} και την σχετική υγρασία RH (όπως μετράται από τον αισθητήρα) προσδιορίζουμε τις διορθωμένες τιμές των συγκεντρώσεων. Έτσι για παράδειγμα στην περίπτωση που ο αισθητήρας αναφέρει συγκέντρωση $30 \mu g/m^3$, και RH 50%, η διορθωμένη συγκέντρωση είναι $18 \mu g/m^3$. Στην περίπτωση που ο αισθητήρας αναφέρει συγκέντρωση $30 \mu g/m^3$, σε RH 80%, η διορθωμένη συγκέντρωση είναι $14 \mu g/m^3$.

Οι ανεπεξέργαστες μετρήσεις του αισθητήρα, οι διορθωμένες μετρήσεις, οι μετρήσεις αναφοράς και η RH (σχετική υγρασία) για 2 ημέρες αναλύσεων (23-25/12/2021)

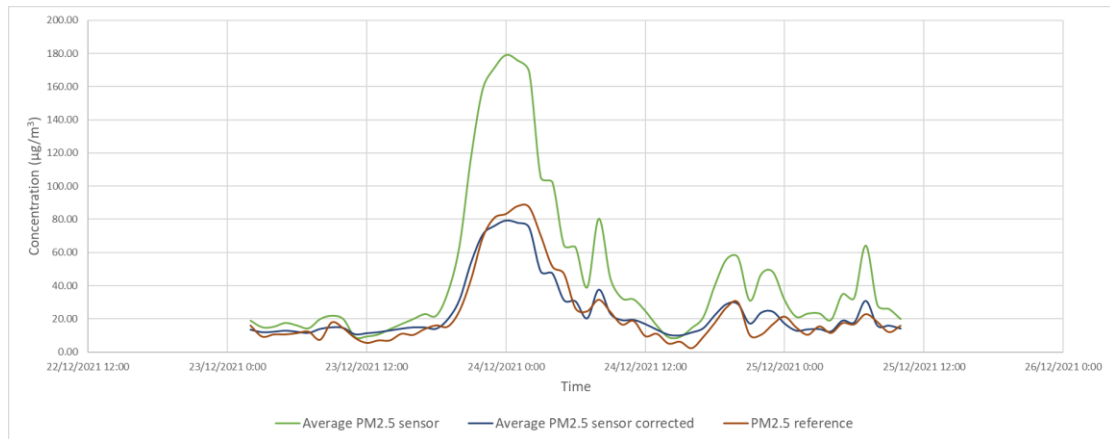
παρουσιάζονται στον πίνακα 3. Η βελτίωση στην ακρίβεια των διορθωμένων τιμών απεικονίζεται στην εικόνα 13.

Πίνακας 3. Οι ανεπεξέργαστες μετρήσεις του αισθητήρα, οι διορθωμένες μετρήσεις, οι μετρήσεις αναφοράς και η RH για περίοδο μετρήσεων 2 ημερών

DateTime (UTC+2)	PM _{2.5} sensor (μg/m ³)	PM _{2.5} sensor corrected (μg/m ³)	PM _{2.5} reference (μg/m ³)	Humidity (%)
23/12/2021 2:00	19.1	13.5	15.9	48
23/12/2021 3:00	15.1	12.0	9.2	47
23/12/2021 4:00	15.4	12.3	10.5	45
23/12/2021 5:00	17.7	12.9	10.5	48
23/12/2021 6:00	16.2	12.2	11.3	48
23/12/2021 7:00	14.5	11.6	12.2	48
23/12/2021 8:00	20.0	14.0	7.3	47
23/12/2021 9:00	22.0	14.9	17.8	46
23/12/2021 10:00	19.9	14.5	14.2	42
23/12/2021 11:00	9.1	10.7	8.4	36
23/12/2021 12:00	9.5	11.3	5.4	32
23/12/2021 13:00	10.7	12.0	6.9	31
23/12/2021 14:00	13.9	13.0	6.9	34
23/12/2021 15:00	17.0	14.1	10.9	35
23/12/2021 16:00	19.9	14.9	10.2	39

23/12/2021 17:00	23.0	14.9	13.8	51
23/12/2021 18:00	22.1	14.1	16.0	54
23/12/2021 19:00	36.2	19.8	15.3	56
23/12/2021 20:00	64.2	31.6	24.8	57
23/12/2021 21:00	117.8	54.1	43.9	58
23/12/2021 22:00	158.5	70.9	69.0	61
23/12/2021 23:00	171.3	76.2	81.0	63
24/12/2021 0:00	179.2	79.5	83.4	63
24/12/2021 1:00	175.9	78.1	88.1	63
24/12/2021 2:00	169.0	75.2	87.5	63
24/12/2021 3:00	105.5	48.7	70.3	60
24/12/2021 4:00	102.6	47.5	51.6	59
24/12/2021 5:00	64.6	31.4	47.2	60
24/12/2021 6:00	63.2	30.6	26.2	62
24/12/2021 7:00	39.2	20.3	24.5	62
24/12/2021 8:00	80.5	37.7	31.5	64
24/12/2021 9:00	44.5	23.2	24.1	57
24/12/2021 10:00	32.5	19.2	16.5	48

24/12/2021 11:00	31.9	19.4	18.7	44
24/12/2021 12:00	25.0	16.9	9.5	40
24/12/2021 13:00	16.0	13.5	10.9	37
24/12/2021 14:00	9.1	10.5	5.0	38
24/12/2021 15:00	9.2	10.0	6.1	42
24/12/2021 16:00	14.6	11.8	2.2	46
24/12/2021 17:00	21.2	14.4	9.0	48
24/12/2021 18:00	40.4	22.3	17.6	50
24/12/2021 19:00	56.1	28.9	26.8	50
24/12/2021 20:00	57.1	28.9	29.8	54
24/12/2021 21:00	31.1	17.2	10.1	60
24/12/2021 22:00	47.2	23.9	10.4	61
24/12/2021 23:00	48.4	24.3	16.7	61
25/12/2021 0:00	31.6	17.2	21.3	61
25/12/2021 1:00	21.4	12.9	14.8	62



Εικόνα 13. Οι συγκεντρώσεις του οργάνου αναφοράς (κόκκινη γραμμή), οι ανεπεξέργαστες συγκεντρώσεις του αισθητήρα (πράσινη γραμμή) και οι διορθωμένες συγκεντρώσεις του αισθητήρα (μπλε γραμμή)

ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Η περιγραφική στατιστική ασχολείται με τις μεθόδους οργάνωσης, σύνοψης και παρουσίασης των δεδομένων με τρόπο εύχρηστο και κατανοητό. Οι τεχνικές που χρησιμοποιούνται είναι τα γραφήματα που οπτικοποιούν τα δεδομένα και οι αριθμητικοί δείκτες που αντιπροσωπεύουν ιδιότητες των δεδομένων. Η περιγραφική στατιστική επιτρέπει την εξαγωγή συμπερασμάτων και τη λήψη αποφάσεων

Δείκτες κεντρικών τάσεων (μέσος όρος και διάμεσος)

Ο μέσος όρος και η διάμεσος είναι οι πρώτοι δείκτες που υπολογίζονται σε ένα νέο σύνολο δεδομένων.

Ο μέσος όρος υπολογίζεται αθροίζοντας όλες τις τιμές και διαιρώντας το άθροισμα με τον αριθμό των τιμών. Ο μέσος όρος δεν αποτελεί καλό δείκτη της κεντρικής τάσης στην περίπτωση που υπάρχουν ακραίες τιμές (outliers).

Η διάμεσος είναι η μεσαία τιμή όταν ένα σύνολο δεδομένων ταξινομείται από το ελάχιστο στο μέγιστο. Εάν υπάρχουν n τιμές, η διάμεσος ορίζεται ως η $(n+1)/2$ η τιμή. Έτσι αν $n = 7$, η διάμεσος τιμή είναι η $(7+1)/2$ η ή η τέταρτη τιμή. Εάν ο αριθμός των τιμών είναι ζυγός, η διάμεσος είναι ο μέσος όρος των δύο μεσαίων τιμών. Αυτό ορίζεται ως ο μέσος όρος της $(n/2)$ και $((n/2)+1)$ ης τιμής. Έτσι αν υπάρχουν έξι τιμές, η διάμεσος είναι ο μέσος όρος της $(6/2)$ ης και $((6/2)+1)$ ης τιμής ή η τρίτη και τέταρτη τιμή. Και οι δύο τεχνικές παρουσιάζονται παρακάτω:

Μονός αριθμός τιμών: 1, 2, 3, 4, 7, 8, 10 Διάμεσος = 4

Ζυγός αριθμός τιμών: 1, 2, 3, 4, 5, 6 Διάμεσος = $(3+4)/2 = 3.5$

Η διάμεσος είναι καλύτερος δείκτης μέτρησης της κεντρικής τάσης από τον μέσο όρο για δεδομένα που είναι ασύμμετρα ή περιέχουν ακραίες τιμές. Αυτό συμβαίνει επειδή η διάμεσος βασίζεται στην σειρά των τιμών των δεδομένων και όχι στις πραγματικές τους τιμές: το 50 % των τιμών δεδομένων βρίσκονται κάτω από τη διάμεσο και το 50 % πάνω από τη διάμεσο, χωρίς να λαμβάνονται υπόψη οι πραγματικές τιμές. Επομένως, δεν έχει σημασία αν το σύνολο δεδομένων περιέχει κάποιες εξαιρετικά μεγάλες ή μικρές τιμές, γιατί αυτές δεν επηρεάζουν τον δείκτη περισσότερο από τις λιγότερο ακραίες. Για παράδειγμα, η διάμεσος και των τριών παρακάτω κατανομών είναι 4:

Κατανομή A: 1, 1, 3, 4, 5, 6, 7

Κατανομή Β: 0.01, 3, 3, 4, 5, 5, 5

Κατανομή Γ: 1, 1, 2, 4, 5, 100, 2000

Σε μια συμμετρική κατανομή (όπως είναι η κανονική κατανομή), ο μέσος όρος και η διάμεσος τιμή συμπίπτουν. Σε μια ασύμμετρη (skewed) κατανομή οι δύο αυτοί δείκτες διαφέρουν και ο βαθμός κατά το οποίο διαφέρουν είναι ένας τρόπος αξιολόγησης της ασυμμετρίας μιας κατανομής.

Δείκτες διασποράς

Οι δείκτες διασποράς είναι εξίσου σημαντικές με τους δείκτες κεντρικής τάσης διότι καθορίζουν τον τρόπο με τον οποίο οι τιμές είναι διεσπαρμένες γύρω από τις παραμέτρους κεντρικής τάσης.

Για παράδειγμα, δύο σύνολα συγκεντρώσεων ατμοσφαιρικής ρύπανσης PM_{2,5} μπορεί να έχουν και τα δύο τον ίδιο μέσο όρο 20 μg/m³, αλλά το ένα θα μπορούσε να έχει εύρος από 0 έως 40 μg/m³ ενώ το άλλο να έχει εύρος από 0 έως 70 μg/m³ υποδηλώνοντας διαφορετικά μοτίβα ρύπανσης.

Το εύρος

Το απλούστερο μέτρο διασποράς είναι το εύρος, το οποίο είναι η διαφορά μεταξύ της υψηλότερης και της χαμηλότερης τιμής ενός συνόλου δεδομένων.

Διακύμανση, τυπική απόκλιση και συντελεστής μεταβλητότητας

Τα πιο κοινά μέτρα διασποράς για συνεχή δεδομένα είναι η διακύμανση και η τυπική απόκλιση. Και τα δύο περιγράφουν πόσο διαφέρουν οι μεμονωμένες τιμές σε ένα σύνολο δεδομένων από τη μέση τιμή. Η διακύμανση είναι ο μέσος όρος των τετραγωνικών αποκλίσεων από τον μέσο όρο και η τυπική απόκλιση είναι η τετραγωνική ρίζα της διακύμανσης.

Ο τύπος για τη διακύμανση s^2 ενός συνόλου δεδομένων με n παρατηρήσεις είναι:

$$s^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2$$

Όπου x_i είναι η τιμή i του συνόλου δεδομένων ενώ \bar{x} είναι ο μέσος όρος των τιμών του συνόλου δεδομένων.

Οι μονάδες μέτρησης της διακύμανσης είναι οι μονάδες του αρχικού μεγέθους στο τετράγωνο. Έτσι αν μετράμε τις συγκεντρώσεις των PM_{2,5} σε μg/m³, τότε η διακύμανση των τιμών θα μετράται σε μg/m³ εις το τετράγωνο. Αυτό σημαίνει ότι θα υπήρχε δυσκολία σύγκρισης του μέσου όρου και της διακύμανσης. Για να αποφευχθεί αυτό, παίρνουμε την τετραγωνική ρίζα της διακύμανσης και το νέο μέγεθος ονομάζεται τυπική απόκλιση. Ο τύπος για την τυπική απόκλιση ενός συνόλου δεδομένων είναι:

$$s = \sqrt{s^2} = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Γενικά, για δύο μεταβλητές που μετρώνται με τις ίδιες μονάδες, η ομάδα με τη μεγαλύτερη διακύμανση και τυπική απόκλιση έχει μεγαλύτερη την μεταβλητότητα. Ωστόσο, η μονάδα

μέτρησης επηρεάζει το μέγεθος της διακύμανσης: οι συγκεντρώσεις $PM_{2.5}$, εκφρασμένες σε mg/m^3 αντί για $\mu g/m^3$, θα είχαν μικρότερη διακύμανση και τυπική απόκλιση. Ο συντελεστής διακύμανσης (CV – Coefficient of Variation), ξεπερνά αυτή τη δυσκολία και καθιστά δυνατή τη σύγκριση της μεταβλητότητας μεταξύ των μεταβλητών που μετρώνται σε διαφορετικές μονάδες. Ο CV υπολογίζεται διαιρώντας την τυπική απόκλιση με τον μέσο όρο και στη συνέχεια πολλαπλασιάζοντας το αποτέλεσμα με το 100:

$$CV = \frac{s}{\bar{x}} \times 100\%$$

Οι δείκτες περιγραφικής στατιστικής για τα διορθωμένα δεδομένα του αισθητήρα PurpleAir από τον πίνακα 3, παρουσιάζονται στον πίνακα 4.

Πίνακας 4. Δείκτες περιγραφικής στατιστικής των διορθωμένων δεδομένων του αισθητήρα PurpleAir για τα $PM_{2.5}$ ($\mu g/m^3$)

Mean	25.8
Median	17.1
Minimum	10.0
Maximum	79.5
Range	70
Variance	403.7
Standard deviation	20.1
Coefficient of variation (CV)	78%

Γραφήματα

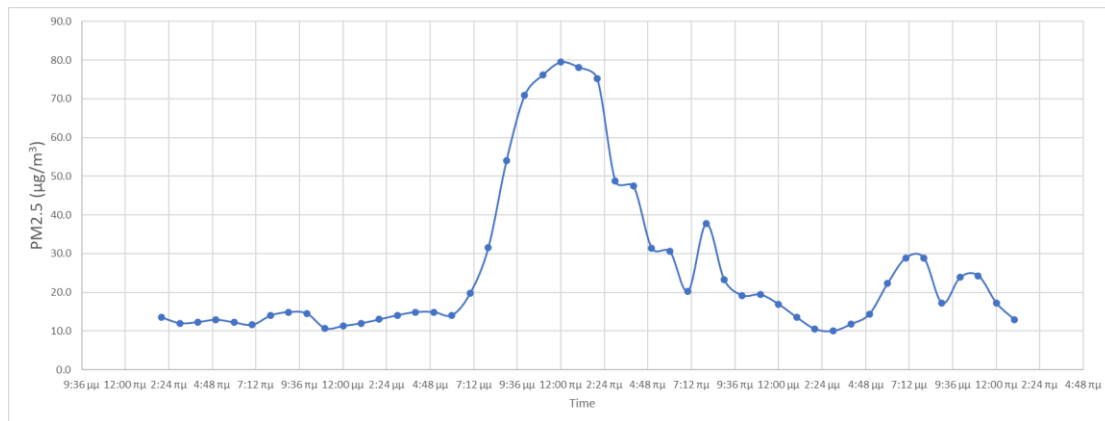
Τα γραφήματα είναι εργαλεία που χρησιμοποιούνται για να παρουσιάζονται πληροφορίες σχετικά με τα δεδομένα. Υπάρχουν πολλές μέθοδοι γραφικής αναπαράστασης δεδομένων, κάποιες είναι βασικές και περιλαμβάνονται σε λογισμικά υπολογιστικών φύλλων όπως το Microsoft Excel και άλλες είναι πολύπλοκες και αναπτύσσονται σε γλώσσες προγραμματισμού όπως η R. Η απλούστερη γραφική αναπαράσταση είναι συχνά η καλύτερη.

Τα γραφήματα γραμμής χρησιμοποιούνται για την εμφάνιση της σχέσης μεταξύ δύο μεταβλητών, συχνά μεταξύ χρόνου στον άξονα x και κάποιας άλλης μεταβλητής στον άξονα y.

Στην εικόνα 14, παρουσιάζεται ένα γράφημα γραμμής μεταξύ των διορθωμένων συγκεντρώσεων $PM_{2.5}$ από έναν αισθητήρα PurpleAir (δεδομένα από τον πίνακα 3) σε σχέση με το χρόνο για μια περίοδο 2 ημερών (παραμονή Χριστουγέννων και ημέρα των Χριστουγέννων 2021). Την πρώτη ημέρα, οι συγκεντρώσεις $PM_{2.5}$ αρχίζουν να αυξάνονται αργά το απόγευμα (από τις 6:00 μ.μ. περίπου) και φτάνουν στην μέγιστη τιμή τους ($80 \mu g/m^3$) στις 12:00 τα μεσάνυχτα. Στη συνέχεια, οι συγκεντρώσεις μειώνονται σταδιακά μέχρι τις 7:00 π.μ.

Την επόμενη μέρα, μια μικρότερη κορυφή εμφανίζεται το πρωί, στις 8:00 π.μ. ($29 \mu\text{g}/\text{m}^3$). Το απόγευμα και το βράδυ της δεύτερης ημέρας ανιχνεύονται δύο μικρές κορυφές στις 7:00 μ.μ. ($30 \mu\text{g}/\text{m}^3$) και στις 22:30 μ.μ. ($24 \mu\text{g}/\text{m}^3$).

Αυτό το διαφορετικό μοτίβο ατμοσφαιρικής ρύπανσης κατά τη διάρκεια των δύο ημερών οφείλεται πιθανώς στις ποσότητες των $\text{PM}_{2.5}$ που εκπέμπονται αλλά και στις μετεωρολογικές συνθήκες. Η θέρμανση των κατοικιών αποτελεί σημαντική πηγή $\text{PM}_{2.5}$ κατά τους χειμερινούς μήνες, ιδιαίτερα όταν χρησιμοποιείται βιομάζα (ξύλα) σε σόμπες και τζάκια. Οι κορυφές εμφανίζονται συνήθως το απόγευμα και το βράδυ, όταν οι άνθρωποι βρίσκονται στα σπίτια τους και χρησιμοποιούν την θέρμανση (κορυφή κατά την πρώτη ημέρα). Η οδική κυκλοφορία δίνει επίσης κορυφές, ιδιαίτερα κατά τη διάρκεια της ημέρας. Στην κυκλοφορία θα μπορούσε να οφείλεται η κορυφή που παρατηρείται το πρωί της δεύτερης μέρας. Οι μετεωρολογικές συνθήκες παίζουν σημαντικό ρόλο διότι οι υψηλές ταχύτητες ανέμου ευνοούν τη διασπορά της ρύπανσης και η βροχή ξεπλένει τα PM από την ατμόσφαιρα. Ίσως αυτό να συμβαίνει τη δεύτερη ημέρα των μετρήσεων. Αυτές είναι ενδεικτικές εξηγήσεις των μοτίβων ατμοσφαιρικής ρύπανσης οι οποίες πρέπει να υποστηριχθούν από περαιτέρω έρευνα.



Εικόνα 14. Γράφημα γραμμής των συγκεντρώσεων $\text{PM}_{2.5}$ σε συνάρτηση με τον χρόνο

Πηγές

2016 product data manual of PLANTOWER, Digital universal particle concentration sensor, available at https://www.aqmd.gov/docs/default-source/aq-spec/resources-page/plantower-pms5003-manual_v2-3.pdf

Ardon-Dryer, K., Dryer, Y., Williams, J. N., & Moghimi, N. (2020). Measurements of $\text{PM}_{2.5}$ with PurpleAir under atmospheric conditions. *Atmospheric Measurement Techniques*, 13(10), 5441-5458.

Barkjohn, K. K., Gantt, B., & Clements, A. L. (2021). Development and application of a United States-wide correction for $\text{PM}_{2.5}$ data collected with the PurpleAir sensor. *Atmospheric Measurement Techniques*, 14(6), 4617-4637. <https://doi.org/10.5194/amt-14-4617-2021>

Bi, J.; Wildani, A.; Chang, H.H.; Liu, Y. Incorporating low-cost sensor measurements into high-resolution $\text{PM}_{2.5}$ modeling at a large spatial scale. *Environ. Sci. Technol.* 2020, 54, 2152-2162. <https://doi.org/10.1021/acs.est.9b06046>

Boslaugh, S. (2012). *Statistics in a nutshell: A desktop quick reference*. O'Reilly Media, Inc.

deSouza, P., Barkjohn, K., Clements, A., Lee, J., Kahn, R., Crawford, B., & Kinney, P. (2023). An analysis of degradation in low-cost particulate matter sensors. *Environmental Science: Atmospheres*, 3(3), 521-536.

EEA (n.d.). European Air Quality Index, European Environment Agency, available at <https://airindex.eea.europa.eu/Map/AQI/Viewer/>

EPA (2001). The Precision and Accuracy of Environmental Measurements for the Building Assessment Survey and Evaluation Program Previously, U.S. Environmental Protection Agency, available at <https://www.epa.gov/sites/default/files/2014-08/documents/precisionandaccuracy.pdf>

EPA (2022). A Guide to Siting and Installing Air Sensors, available at <https://www.epa.gov/air-sensor-toolbox/guide-siting-and-installing-air-sensors>

EPA (n.d.). How to Evaluate Low-Cost Sensors by Collocation with Federal Reference Method Monitors, National Exposure Research Laboratory Office of Research and Development, available at https://www.epa.gov/sites/default/files/2018-01/documents/collocation_instruction_guide.pdf

Giordano, M. R., Malings, C., Pandis, S. N., Presto, A. A., McNeill, V. F., Westervelt, D. M., ... & Subramanian, R. (2021). From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors. *Journal of Aerosol Science*, 158, 105833.

Kosmopoulos, G., Salamalikis, V., Pandis, S. N., Yannopoulos, P., Bloutsos, A. A., & Kazantzidis, A. (2020). Low-cost sensors for measuring airborne particulate matter: Field evaluation and calibration at a South-Eastern European site. *Science of The Total Environment*, 748, 141396.

PurpleAir (n.d.). What are channel A and channel B? available at <https://community.purpleair.com/t/what-are-channel-a-and-channel-b/3643>

Rai A.C., Kumar P (2017). Summary of Air Quality sensors and recommendations for application, iScape. Available online: https://www.iscapeproject.eu/wp-content/uploads/2017/09/iSCAPE_D1.5_Summary-of-air-qualitysensors-and-recommendations-for-application.pdf

Yatkin, S., Gerboles, M, Borowiak, A and Signorini, M. (2022). Guidance on low-cost sensors deployment for air quality monitoring experts based on the AirSensEUR experience, Publications Office of the European Union, Luxembourg, doi:10.2760/14893, JRC130050.

Χρηματοδότηση: Η έρευνα χρηματοδοτήθηκε από την Ερευνητική Υποδομή EN.I.R.I.S.S.T.+ (MIS 5047041), που υλοποιείται στο πλαίσιο της Δράσης «Ενίσχυση των Υποδομών Έρευνας και Καινοτομίας», που χρηματοδοτείται από το Επιχειρησιακό Πρόγραμμα ΕΠΑνΕΚ (ΕΣΠΑ 2014-2020) και συγχρηματοδοτείται από την Ελλάδα και την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης).