

▪ Proyecto Fin de Grado ▪

Ingeniería de Computadores

Estudio de viabilidad de captura de datos  
medioambientales con sensores de bajo coste de forma  
distribuida

---

Autor

Iñaki Garrido Erausquin

Director

Julián Alberto Lafuente Rojo

Noviembre 2017



Iñaki Garrido Erausquin, 2017.

Esta obra está bajo una Licencia Creative Commons Atribución-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visite el siguiente enlace:

<http://creativecommons.org/licenses/by-sa/4.0/>.

El reconocimiento se realizará adjuntando el nombre y apellidos del autor.

## *Agradecimientos*

*A mi tutor y director del proyecto, Julián Alberto Lafuente, por ofrecerme la oportunidad de realizar este proyecto, por su dedicación en él y por haberme guiado en los momentos oportunos;*

*A Luis Gardezabal por aportar su conocimiento y haberse mostrado disponible en todo momento para ayudar, y al grupo Adian por haber suministrado el material necesario para este proyecto;*

*A Erik Fustes, porque su trabajo previo ha servido como inspiración en este, y por haber mostrado interés y disposición para ayudar en todo momento;*

*A todos mis compañeros con los que he compartido mi estancia en la facultad, en especial a mis amigos Christian y Ainhoa y a mi hermano Jon, por haberme ayudado tanto y haber hecho el trayecto más agradable;*

*A la Facultad de Informática y todo el personal que trabaja en ella por haber hecho posible esta experiencia académica y por el buen trato recibido en todo momento;*

*A mis compañeros de Magna SIS, Asier, Ainhoa, Igor, Iñaki, Jon y Christian por todo el esfuerzo hecho juntos para que Magna SIS pueda continuar existiendo y haberme hecho sentir parte de un grupo excepcional, así como a todas las generaciones previas y por venir, por su dedicación y entusiasmo ante el trabajo;*

*A mi familia y mis amigos por estar a mi lado y hacer mi vida mucho más feliz e interesante;*

*Y a mis padres, por su apoyo incondicional, su esfuerzo diario por hacer mi vida más fácil y porque sin ellos nada de esto habría sido posible;*

*Muchas gracias a todos*

# Resumen

---

En este proyecto de fin de grado se recoge el trabajo realizado con la intención de evaluar la usabilidad de un sensor de materia particulada de bajo coste (Shinyei PPD42NS) como indicador de partículas PM10 y PM2,5. Además de obtener conclusiones acerca del sensor, el trabajo tiene como fin desarrollar una metodología apropiada para este tipo de estudios, que ofrezca una base sólida para futuros proyectos de medida de la calidad del aire basados en ciencia ciudadana (crowdsourcing). Esta metodología se ha basado en el diseño y construcción de un prototipo que ha permitido toma datos con varios sensores a la vez, que luego se han analizado en base a criterios estadísticos.

## Laburpena

---

Gradu amailerako proiektu honetan aireko partikulak neurtzeko sensore merke (Shinyei PPD42NS) baten erabilgarritasuna neurtu da, konkretuki PM10 eta PM2,5 partikula moten kontzentrazioa neurtzeko ahalmena. Sensureari buruzko konklusioak ateraz aparte, lan honek ikerketa mota hauentzako metodología egoki bat eskaintzea du helburu, herritarren zientzian (crowdsourcing) oinarrituriko airearen kalitatea neurtzeko proiektuentzat oinarri sendo bat eskaintzeko. Metodologia hau, haibat sensorekin batera datuak biltzea baimendu duen prototipo baten diseinu eta eraikuntzan oinarritu da, eta datuak estatistika irizpideen bitartez analizatu dira.

# Abstract

---

In this end of grade project, the work done with the intention of evaluating the usability of a low cost particulate matter sensor (Shinyei PPD42NS) as an indicator of PM10 and PM2,5 particles is shown. In addition to obtaining some conclusions about the sensor, this work has the goal of developing an appropriate methodology for this type of studies that will offer a solid base for future projects about getting measurements of the air quality based on crowdsourcing. This methodology has been based on the design and construction of a prototype that has allowed to collect data with some sensors at the same time, data that then has been analysed based on statistic criteria.

# Índice

---

1. Introducción, objetivos y alcance.....	10
1.1.    Introducción .....	11
1.2.    Objetivo General.....	12
1.3.    Objetivos Concretos.....	12
1.4.    Exclusiones .....	13
1.5.    Metodología .....	13
1.6.    Esquema de Descomposición del Trabajo (EDT) .....	14
2. Antecedentes.....	15
2.1.    Trabajos sobre el sensor .....	16
2.2.    Estudios previos empleando el sensor.....	16
2.3.    Caracterización de las partículas PM .....	17
3. Diseño del experimento.....	19
3.1.    Objetivos concretos del experimento .....	20
3.2.    Enumeración y descripción del material y las herramientas empleadas .....	20
3.2.1 Detector de partículas (Shinyei PPD42NS) .....	20
3.2.2 Arduino Uno R3 + Base Shield v2.1 .....	22
3.2.3 GNU Octave .....	23
3.3.    Método y diseño del experimento.....	23
3.3.1 Primeros pasos.....	23
3.3.1.1 Estaciones meteorológicas del Gobierno Vasco.....	23
3.3.1.2 Ampliación de la memoria del Arduino.....	25
3.3.1.3 Preparación del software .....	26
3.3.1.4 Almacenamiento de los datos para su tratamiento .....	28
3.3.1.5 Alimentación y estimaciones de consumo.....	30
3.3.2 Variables del estudio.....	31
3.3.3 Implementación del experimento.....	32
3.3.3.1 Montaje .....	32
3.3.3.2 Toma de datos.....	33
3.3.3.3 Análisis de los datos.....	34
4. Resultados del experimento .....	35
4.1.    Observación inicial de los datos obtenidos.....	36

4.2.	Efectos de la luz y la humedad .....	39
4.3.	Tratamiento de los datos .....	39
4.4.	Prueba en condición de contaminación elevada .....	42
4.5.	Comprobaciones de consumo .....	43
5.	Implementación a gran escala .....	45
5.1	Modelos de implementación a gran escala.....	46
5.2	Posible uso del sensor en un experimento a gran escala.....	47
6.	Conclusión y trabajo futuro .....	49



## Lista de Figuras

---

<b>Figura 1:</b> Sensor Shinyei ppd42 .....	12
<b>Figura 2:</b> Esquema de Descomposición del Trabajo .....	14
<b>Figura 3:</b> Niveles de calidad del aire en función del PM .....	17
<b>Figura 4:</b> Caracterización de las partículas PM10 .....	18
<b>Figura 5:</b> Caracterización de las partículas PM2,5 .....	18
<b>Figura 6:</b> Imágenes del sensor Shinyei PPD42NS .....	21
<b>Figura 7:</b> Características del sensor Shinyei PPD42NS .....	21
<b>Figura 8:</b> Imagen del Arduino Uno .....	22
<b>Figura 9:</b> Imágenes del módulos Base Shield v2.1 .....	23
<b>Figura 10:</b> Captura de mapa con las estaciones medioambientales .....	24
<b>Figura 11:</b> Estación medioambiental de la Av. Tolosa .....	24
<b>Figura 12:</b> Inclusión de la memoria al Arduino .....	26
<b>Figura 13:</b> Consumo del sensor en función del número de sensores empleados .....	27
<b>Figura 14:</b> Parte principal del código para tomar lecturas .....	28
<b>Figura 15:</b> Relación entre el porcentaje de tiempo en baja y la concentración .....	29
<b>Figura 16:</b> Plantilla para almacenar los datos .....	30
<b>Figura 17:</b> Consumos con diferentes puertos y sensores .....	30
<b>Figura 18:</b> Estimación de duración de las baterías .....	31
<b>Figura 19:</b> Montajes para los sensores .....	33
<b>Figura 20:</b> Sensor con agujero de la cápsula con y sin tapar .....	33
<b>Figura 21:</b> Horas de medición .....	34
<b>Figura 22:</b> Coeficientes de determinación iniciales .....	36
<b>Figura 23:</b> Muestra con valores próximos .....	37
<b>Figura 24:</b> Muestra con valores muy dispares .....	37
<b>Figura 25:</b> Secuencia de medidas de cada uno de los sensores .....	38
<b>Figura 26:</b> Correlación entre sensores .....	38
<b>Figura 27:</b> Coeficientes de determinación con los sensores aislados y expuestos .....	39
<b>Figura 28:</b> Diagrama de dispersión (sensor 1) .....	40
<b>Figura 29:</b> Filtrado de datos .....	41
<b>Figura 30:</b> Valores del sensor 1 tras filtrado .....	41
<b>Figura 31:</b> Coeficientes de determinación tras filtrado .....	42
<b>Figura 32:</b> Medición en condiciones de alta contaminación .....	43
<b>Figura 33:</b> Diseño de caja para dispositivo de uso individual .....	47

# 1. Introducción, objetivos y alcance

Este primer capítulo ofrece al lector un marco de referencia para comprender la motivación del proyecto, así como aspectos relacionados con el mismo con los que probablemente no esté familiarizado. Además se establecen los objetivos y la metodología seguida durante el mismo.

## 1.1. Introducción

---

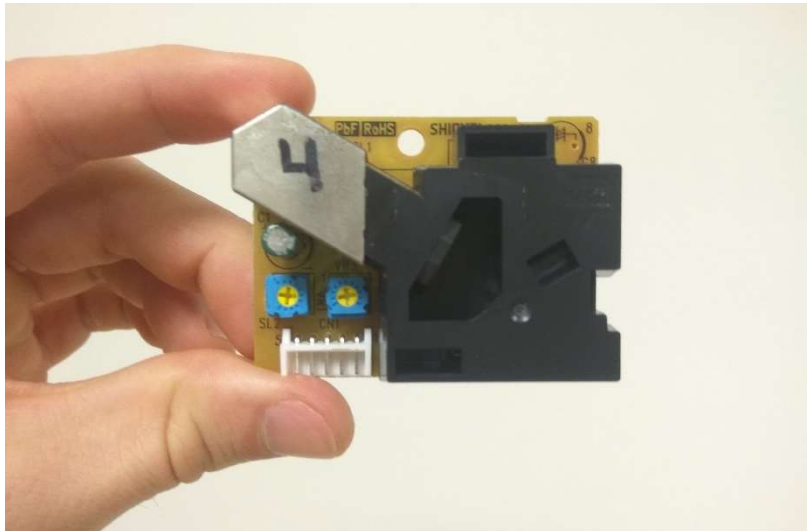
No es algo nuevo el hecho de que la contaminación ambiental y el calentamiento global son temas que preocupan a parte importante de la sociedad, que es consciente del problema y ve la necesidad de realizar cambios en sus conductas. Así queda reflejado también tanto en gobiernos y ayuntamientos que están tomando medidas para evitar que la situación empeore, como en los numerosos estudios al respecto de asociaciones independientes con el fin de concienciar aún más a la gente.

Más concretamente, el empeoramiento de la calidad del aire que respiramos y los efectos que este tiene ha sido uno de los temas más debatidos en los últimos tiempos. Según la OMS (Organización Mundial de la Salud), en 2014 el 92% de la población vivía en zonas donde no se respetaban las directrices de la propia organización en materia de calidad de aire, y se estima que la contaminación atmosférica en zonas rurales y ciudades provoca 3 millones de defunciones prematuras al año.

Es precisamente la contaminación del aire de las ciudades y zonas pobladas en general la razón de ser de este estudio. Además de numerosos gases perjudiciales para la salud, también son importantes los datos relativos a la concentración de algunas partículas en suspensión para determinar la calidad del aire de una zona. Concretamente las partículas PM<sub>2,5</sub> y PM<sub>10</sub> (partículas inhalables con tamaños entre 0 y 2,5 micrómetros, y 0 y 10 micrómetros respectivamente) que tienen origen en incendios, industria, combustión de los vehículos y el uso de aerosoles, entre otros, además de en algunas causas naturales (como por ejemplo vientos cargados de polvo y polen), aumentan considerablemente el riesgo de desarrollar cardiopatías y neumopatías, así como cáncer de pulmón.

Por esta razón la mayoría de las estaciones meteorológicas encargadas de evaluar la calidad del aire miden la concentración de estas partículas. Si bien los datos que ofrecen son datos fiables, tan solo muestran los datos de la localización concreta en la que se encuentran las estaciones, y utilizan instrumentos de medición caros. Por este motivo son numerosas las empresas que han comenzado a fabricar sensores económicos y de tamaño reducido que utilizan distintas tecnologías para detectar estas partículas, que aunque no aseguran la precisión y fiabilidad de instrumentos más caros, permiten tomar medidas en infinidad de sitios y/o de forma móvil, además de habilitar la opción de hacer estudios de forma colaborativa (crowdsourcing), todo ello con costes moderados. Un ejemplo de agrupación interesada en el uso de estos sensores es Kalapie, asociación de ciclistas de San Sebastián que promueve el uso de la bici como forma de movilidad sostenible, y que ya ha realizado pruebas con sensores de bajo coste.

En este informe queda plasmado el estudio realizado sobre unos de estos sensores (Shinyei ppd42 (Fig. 1)) con el fin de conocer su comportamiento y posibles entornos y situaciones de uso, concretando los objetivos iniciales, los experimentos llevados a cabo y las conclusiones extraídas de los mismos, así como un posible modelo de uso a gran escala.



*Figura 1: Sensor Shinyei ppd42*

## 1.2. Objetivo General

---

Diseñar e implementar un experimento para evaluar el comportamiento de los sensores de partículas Shinyei PPD42 y conocer su precisión, fiabilidad y consumo a la hora de determinar la concentración de PM, en función de los cuales se determinará la utilidad de los mismos de cara al uso como dispositivo personal para medir la calidad del aire.

Dependiendo de los resultados obtenidos, se propondrá un modelo de uso a gran escala para evaluar la calidad del aire de forma colaborativa. Esto permitiría obtener información sobre la concentración de partículas en una superficie (ciudad, pueblo, región...) con una resolución mucho mayor de la que aportan las comúnmente usadas estaciones meteorológicas, que si bien ofrecen datos fiables, tienen la limitación de ser estáticas.

## 1.3. Objetivos Concretos

---

- Buscar, recolectar y contrastar **información relativa a los sensores Shinyei PPD42**, tanto ofrecida por el propio fabricante como de trabajos independientes, así como información de **estudios previos realizados con este sensor**, similares al planteado en este trabajo. Asimismo caracterizar la materia particulada en suspensión, y en concreto la respirable (PM).
- Diseñar y realizar el montaje de un prototipo que permita la obtención de datos con los sensores, y extraer conclusiones de un experimento adaptado al material disponible para posteriormente evaluar la usabilidad de los mismos, haciendo análisis de datos con la ayuda de la herramienta de software libre matemático OCTAVE.

- Diseñar un modelo de uso a gran escala que permita obtener datos de forma colaborativa, concretando el diseño de una posible “caja de medición”, posibilidades/limitaciones en función del número de participantes, infraestructura necesaria y costes del mismo.
- Proponer posibles futuros experimentos basados en el conocimiento adquirido sobre los sensores, concretando las diferencias fundamentales con respecto al realizado, así como sus requisitos de material.

## 1.4. Exclusiones

---

El proyecto se limita al uso del sensor Shinyei PPD42, y quedan fuera de este proyecto pruebas comparativas con otros sensores, así como todas las pruebas que requiriesen de algún material no mencionado.

Asimismo, tampoco se investigará la forma en la que las estaciones meteorológicas calculan la concentración de partículas ni si existen métodos mejores para realizar los cálculos. Se tomarán como fiables sus datos por ser los datos oficiales del Gobierno Vasco.

Por último, no se llevará a la práctica el modelo de uso a gran escala, la conexión entre sensores (o sensor-servidor), ni la forma en la que el usuario final tendría acceso a los datos.

## 1.5. Metodología

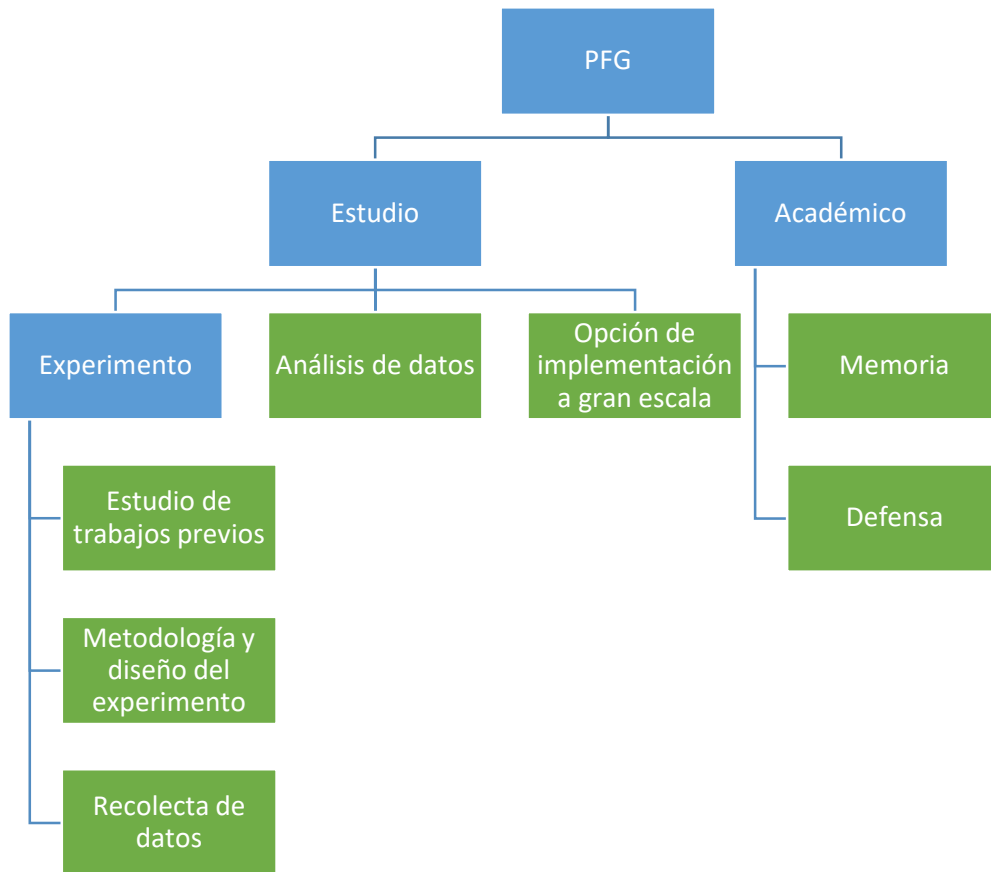
---

La metodología empleada se ha basado en la captura de datos simultánea con 5 sensores en las cercanías de alguna de las estaciones meteorológicas del Gobierno Vasco, para posteriormente contrastarlos entre ellos mismos y con los datos de referencia (los del GV). Se han seguido los siguientes pasos:

- Diseño y construcción de un prototipo que permita recolectar datos con 5 sensores a la vez.
- Recolecta de datos teniendo en cuenta la variabilidad de las circunstancias y tomando nota de ello.
- Análisis de los datos recolectados tomando como referencia los datos ofrecidos por el Gobierno Vasco.
- Extracción de conclusiones de cara a un posible uso de los sensores en un experimento a gran escala.

## 1.6. Esquema de Descomposición del Trabajo (EDT)

---



*Figura 2: Esquema de Descomposición del Trabajo*

## 2. Antecedentes

Este capítulo ofrece un contexto para este estudio, cuya idea inicial ha nacido del director del mismo con la intención de ampliar y concretar parte del conocimiento plasmado en trabajos previos. En él queda reflejada la primera parte de este trabajo, cuyo objetivo ha sido recolectar información de utilidad, tanto **sobre el sensor** en sí (características, modo de uso...) como de estudios realizados **con el sensor** (metodologías empleadas, conclusiones extraídas...).

## 2.1. Trabajos sobre el sensor

---

Uno de los antecedentes más inmediatos a este trabajo tanto por su temática como por su cercanía espacial y en el tiempo, se encuentran en el TFG [1] realizado por Erik Fustes, actualmente investigador del departamento de ATC de la Facultad de Informática de San Sebastián (UPV/EHU). El trabajo plasmado en este documento podría considerarse la continuación de parte de ese, tal y como se expresa en los posibles trabajos futuros del mismo.

En [1] se desarrolla un prototipo que utiliza diversos sensores de contaminación (entre ellos el Shinyei ppd42), todos ellos con un coste relativamente bajo como característica común, para lo cual se realiza una vasta comparativa entre los sensores actuales más relevantes, así como un breve estudio de la normativa medio-ambiental del momento.

A continuación se enumeran todos los modelos que aparecen en dicho trabajo, indicando qué agentes contaminantes detectan: Sensly (amoníaco, benceno, alcoholes, óxidos de nitrógeno, hidrocarburos, monóxido de carbono, formaldehído y tolueno, e incorpora barómetro, termómetro y sensor de humedad), AirPi (dióxido de nitrógeno y monóxido de carbono), PidgeOnPatrol (dióxido de nitrógeno, ozono y compuestos orgánicos volátiles), SensorDrone (monóxido de carbono), Birdi (monóxido de carbono), CubeSensors (luz, ruido y temperatura), FooBot (partículas PM<sub>2,5</sub> y copuestos orgánico volátiles), AirQualityEgg (dióxido de nitrógeno y monóxido de carbono), AirBeam (partículas PM<sub>2,5</sub>), Atmotube (compuestos orgánicos volátiles y monóxido de carbono).

Por otro lado, el distribuidor de los sensores (Seeedstudio) ofrece un manual de usuario [2] en el que se especifican el modo de uso y las principales características del sensor. En otro proyecto [3] se ha realizado en un exhaustivo trabajo de ingeniería inversa en el cual se dejan plasmadas las características técnicas más relevante del sensor.

## 2.2. Estudios previos empleando el sensor

---

Han servido como referencia y motivación varios estudios previos realizados con el sensor objeto del de este proyecto (Shinyei ppd42). En un trabajo [4] se realiza un sistema de monitorización de la calidad del aire empleando el mismo sensor y un Arduino, y se comparan los resultados con un detector de partículas de mayor enjundia (DC1100 Pro Air Quality). Si bien no se detalla la metodología empleada, muestra como resultado una correlación muy alta entre las lecturas de ambos sensores en la toma de datos realizada en una cocina antes, durante y tras cocinar. En este tipo de comparaciones [5] con sensores significativamente más caros y en interior, se obtienen datos bastante satisfactorios incluso con resoluciones temporales relativamente altas (12,5 minutos).

También se han hecho estudios [6] para comparar las correlaciones obtenidas ente los datos del sensor y la concentración de partículas PM<sub>2,5</sub> por un lado, y PM<sub>10</sub> por otro, obteniendo como resultado datos ambiguos, en los que se ve cómo los sensores covarían más con un tipo de partículas (PM<sub>2,5</sub>) en determinados periodos, y con el otro tipo (PM<sub>10</sub>) en otros.



Por otro lado, los dos trabajos más extensos y detallados que se han encontrado y de los cuales se ha obtenido información útil han sido “**Laboratory Evaluation of the Shinyei PPD42NS Low-Cost Particulate Matter Sensor**” [7] publicado en la revista científica PLOS ONE y “**Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California**” [8] realizado de forma conjunta por la Universidad de California y la Universidad de Washington (USA). Tanto en el que se toman los datos al aire libre como en el que se realiza en el laboratorio, se logran coeficientes de determinación notablemente elevados al comparar los datos con la contaminación medida con instrumentos más caros (aproximadamente de 0,6 y 0,8 respectivamente).

### 2.3. Caracterización de las partículas PM

---

Debido a los problemas de salud que pueden generar las partículas en suspensión, las instituciones reguladoras se preocupan de los datos sobre la materia particulada respirable, conocida de forma abreviada como PM. La categorización más habitual de estas partículas se hace en dos grupos, PM10 y PM2,5, haciendo referencia a partículas menores de 10µm y 2,5µm (aproximadamente) respectivamente. El Gobierno Vasco ofrece en su página web los rangos de concentración de partículas PM en función de los cuales se determina la calidad del aire (Fig. 3).

Estado calidad del aire	PM10	PM2,5
Muy buena	0-25	0-16
Buena	25-50	16-33
Mejorable	50-65	33-39
Mala	65-85	39-50
Muy mala	85-10000	50-10000

Figura 3: Niveles de calidad del aire en función del PM

Debido a la existencia de numerosos métodos para medir la concentración de partículas en el aire, cada uno de ellos con sus ventajas y limitaciones, las regulaciones sobre la misma se realizan en función de la tecnología empleada, habiendo varios métodos estandarizados y aceptados por las entidades reguladoras.

En ocasiones, no todos los contaminantes pueden ser medidos directamente o de forma completa. En estos casos en los que los científicos medioambientales se basan en datos incompletos o medidos de forma indirecta (como es el caso del sensor utilizado), a estos últimos se les suele denominar *indicadores*.

La distribución en función del tamaño de las partículas en suspensión no es homogénea, sino que más bien se distinguen tres grupos de distintos tamaños aproximados: gruesas, finas y ultra finas. De todas ellas, tan solo un subgrupo son respirables: PM10 (partículas inhalables), y dentro de este grupo, el subgrupo PM2,5 (partículas inhalables cuya presencia en los pulmones

es más difícil de eliminar). También es habitual mostrar los datos del PM<sub>2,5-10</sub>, extraídos de la resta PM<sub>10</sub>-PM<sub>2,5</sub>. A continuación se muestran dos gráficas a modo de ilustración (Fig. 3 y 4):

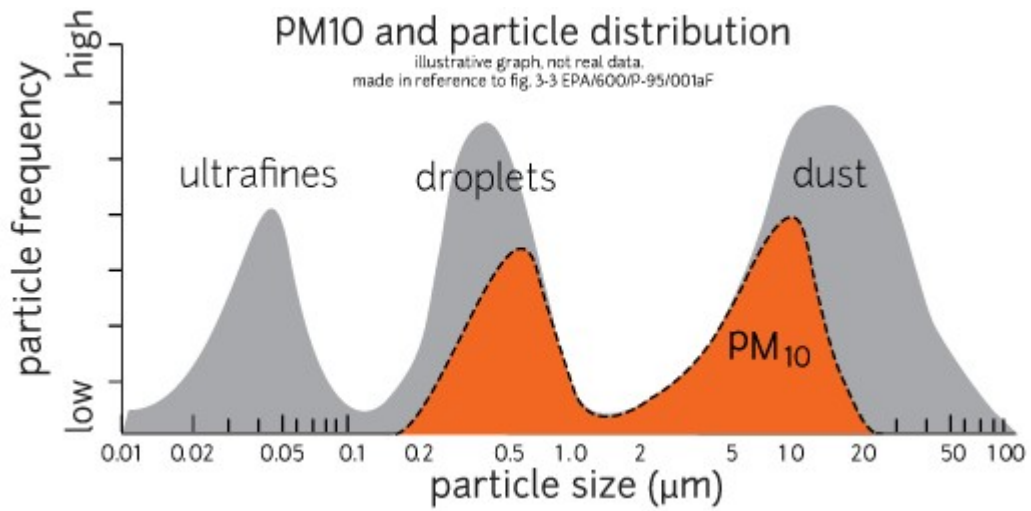


Figura 4: Caracterización de las partículas PM<sub>10</sub>

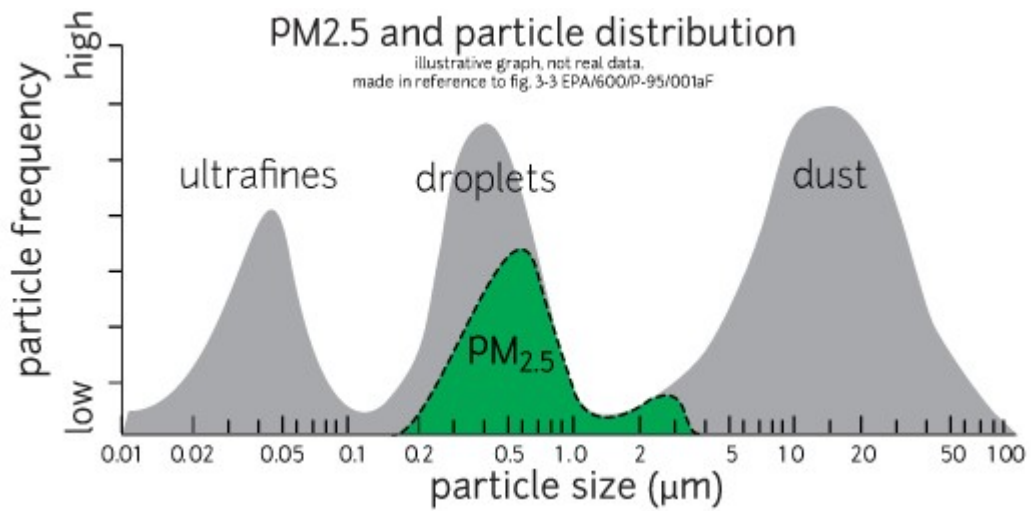


Figura 5: Caracterización de las partículas PM<sub>2,5</sub><sup>1</sup>

<sup>1</sup> Con licencia (CC BY-SA), URL: <https://publiclab.org/wiki/PM-monitoring-regulations>

### 3. Diseño del experimento

A continuación se detalla cómo se ha llevado a cabo el estudio de los sensores, dividiendo el trabajo en distintas fases: concreción de los objetivos y dudas a resolver, selección del material necesario, captura y tratamiento de datos con un método concreto y análisis de los mismos. Se han tomado como referencia las etapas de otros estudios o de métodos estándares para realizar el estudio, seleccionando las fases y adaptándolas a las características y limitaciones del proyecto.

### 3.1. Objetivos concretos del experimento

---

Conocidos el modo de empleo de los sensores y los requisitos del proyecto, así como el objetivo global, que es determinar la posibilidad de capturar datos relativos a la contaminación de forma distribuida, a continuación se detallan los objetivos fijados para la investigación:

- Determinar la efectividad de los sensores a la hora de utilizarlos como indicadores de PM10 y/o PM2,5.
- Determinar qué porcentaje de mejora es posible lograr mediante el tratamiento de los datos (eliminación de valores atípicos, aproximaciones por rangos...).
- Determinar en función de las variables del experimento qué condiciones son las óptimas para el uso de los sensores.
- Calcular el porcentaje de mejora que se logra mediante la toma de datos de forma colaborativa.

### 3.2. Enumeración y descripción del material y las herramientas empleadas

---

A continuación se enumera todo el material utilizado, describiendo posteriormente los elementos más importantes teniendo en cuenta el contexto del proyecto:

- Detector de partículas ‘Shinyei PPD42NS’
- Arduino Uno R3
- Accesorio ‘Base Shield v2.1’ para el Arduino
- Memoria extra EEPROM de 32KB (24LC256) para el Arduino
- Software GNU Octave
- Dos baterías con conectores USB de 16Ah y 10Ah respectivamente, así como un transformador de móvil con conector USB
- Software online de creación y edición de hojas de cálculo de Google (Google Sheets)
- Una caja en la que poder integrar todo el material de medición
- Un multímetro y un ordenador portátil personal

#### 3.2.1 Detector de partículas (Shinyei PPD42NS)

El funcionamiento del detector es bastante sencillo. Una resistencia colocada en la parte inferior (el sensor debe colocarse de manera que la placa quede perpendicular al plano del suelo, siendo el lateral más cercano a la resistencia el que quede de cara a este) calienta el aire de su alrededor, generando una corriente ascendente. El aire (y junto con él las partículas) entra en la cápsula a través del conducto de entrada (intake), y sale por el escape (exhaust) (Fig. 5).

Dentro, un led de luz infrarroja y un fotodiodo junto con una lente crean una zona de detección por la cual pasan las partículas. La luz que estas dispersan es recibida por el fotodiodo, generando así pulsos eléctricos. El número de pulsos por unidad de tiempo es proporcional a la

concentración de partículas en el aire, pudiendo medir de esta forma la contaminación ambiental. En este caso, las características de estos elementos permiten detectar partículas cuyo tamaño sea superior a  $1\mu\text{m}$ .

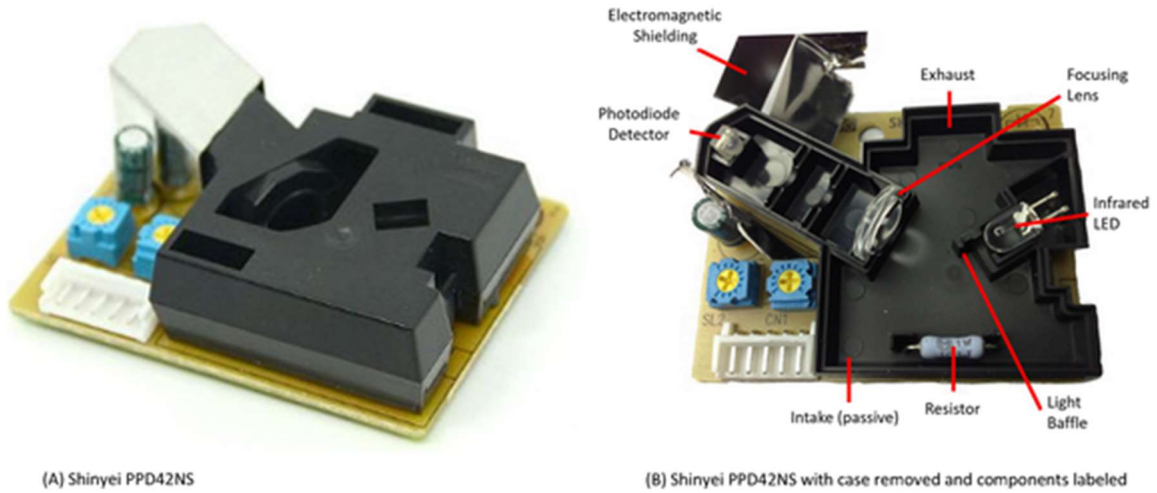


Figura 6: Imágenes del sensor Shinyei PPD42NS <sup>2</sup>

Model	PPD42NS
Detectable particle size	approx. $1\mu\text{m}$ (minimum.)
Detectable range of concentration	0~28,000 pcs/liter (0~8,000pcs/0.01 CF=283ml)
Supply Voltage	DC5V +/- 10% (CN1:Pin1=GND, Pin3=+5V) Ripple Voltage within 30mV
Operating Temperature Range	0~45°C
Operating Humidity Range	95%rh or less (without dew condensation)
Power consumption	90mA
Storage temperature	-30~60°C
Time for stabilization	1 minute after power turned on
Dimensions	59(W) × 45(H) × 22(D) [mm]
Weight	24g(approx.)
Output Method	Negative Logic, Digital output, Hi : over 4.0V(Rev.2) Lo : under 0.7V (As Input impedance : 200kΩ) OP-Amp output, Pull-up resistor : 10kΩ

Figura 7: Características del sensor Shinyei PPD42NS <sup>3</sup>

En cuanto al límite de tamaño superior de las partículas que el sensor puede detectar, ni el fabricante ofrece la información, ni existe un consenso al respecto en los estudios realizados. Al no disponer de ningún tipo de filtro y conociendo la naturaleza del sensor, cabe esperar (y así

<sup>2</sup> URL: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137789#pone-0137789-g001>

<sup>3</sup> URL: <http://wiki.timelab.org/images/f/f9/PPD42NS.pdf>

se puede comprobar en algunos trabajos) que no exista un límite estricto, sino que más bien a partir de cierto tamaño la capacidad de detección de las partículas vaya disminuyendo, cada vez de forma más severa a medida que el tamaño aumenta. Además, debido a la tecnología empleada, este sensor no es capaz de diferenciar las partículas en suspensión (todas) y las partículas respirables (PM10 y PM2,5), si bien es cierto que la mayoría de partículas en el rango 0-10µm son respirables.

### 3.2.2 Arduino Uno R3 + Base Shield v2.1

El Arduino es una tarjeta de microcontrolador (Fig. 7) basada en el microcontrolador de Atmel ATMEGA328P. Dispone de 14 pines digitales de entrada/salida, 6 entradas analógicas y un oscilador de cuarzo que trabaja a 16 MHz. Es posible alimentarla vía una conexión USB (válida también para el intercambio de datos), o mediante un alimentador externo, ya sea una batería o un transformador de corriente.

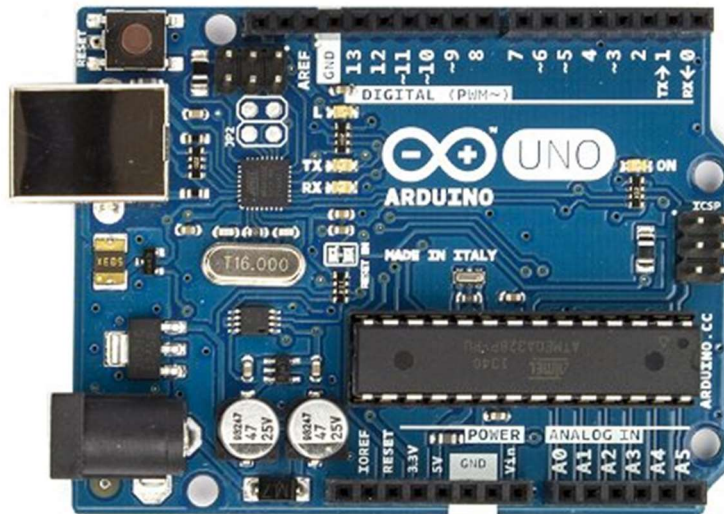


Figura 8: Imagen del Arduino Uno<sup>4</sup>

Por su parte, el adaptador Base Shield v2.1 de Seeed Studio (Fig. 8) compatible con el Arduino añade un total de 16 conexiones: cuatro conexiones analógicas, siete digitales, cuatro I2C y un UART (Siglas en inglés de Transmisor-Receptor Asíncrono Universal). Además, tiene un interruptor que permite seleccionar el voltaje de los conectores (5V o 3.3V).

---

<sup>4</sup> URL: <https://store.arduino.cc/arduino-uno-rev3>

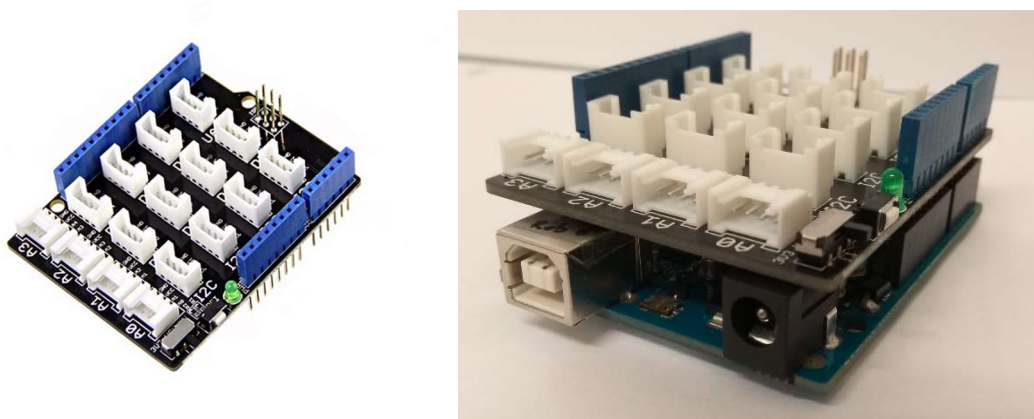


Figura 9: Imágenes del módulos Base Shield v2.1<sup>5</sup>

### 3.2.3 GNU Octave

Se trata de un lenguaje de programación de alto nivel principalmente orientado a la computación numérica, y cuya sintaxis es ampliamente compatible con la de MATLAB. Dispone de un entorno de programación propio (software libre) con su correspondiente línea de comandos que permite llevar a cabo diferentes experimentos numéricos, tales como resolver problemas de álgebra lineal, encontrar la solución de ecuaciones no lineales, manipular polinomios o generar gráficos en 2 y 3 dimensiones, entre otros. Además, es fácilmente extensible mediante la creación de funciones escritas en el propio lenguaje Octave, o utilizando módulos cargados dinámicamente escritos en otros lenguajes como C, C++ y Fortran.

## 3.3. Método y diseño del experimento

---

Previo al experimento plasmado a continuación, se ha llevado a cabo una extensa búsqueda y recolecta de información con el fin orientar la investigación y trabajar sobre una base, así como aprovechar información y partes de trabajos de investigación previos. También ha servido para familiarizarse con el uso de las distintas herramientas utilizadas y fijar los requisitos para realizar el estudio.

### 3.3.1 Primeros pasos

#### 3.3.1.1 Estaciones meteorológicas del Gobierno Vasco

Una particularidad destacable del estudio ha sido la oportunidad de contrastar los datos de los sensores con los datos ofrecidos por el Departamento de Medio Ambiente del Gobierno

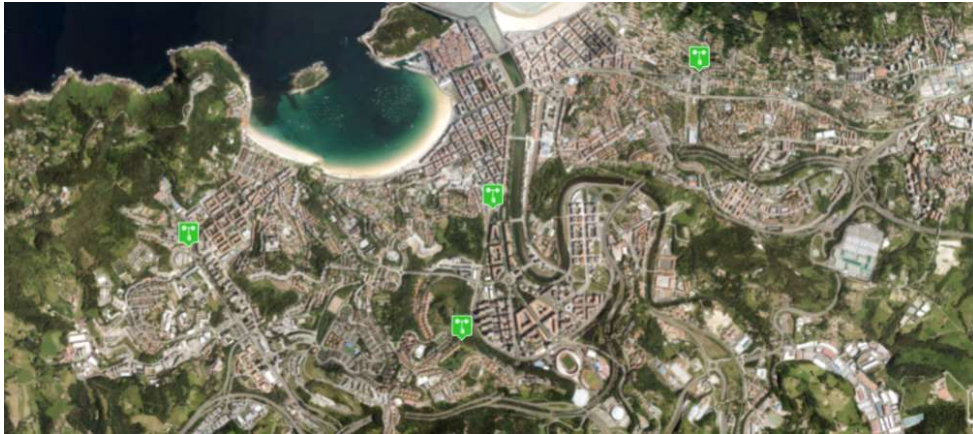
---

<sup>5</sup> URL: <https://www.seeedstudio.com/Base-Shield-V2-p-1378.html>



Vasco en su página web. A continuación se muestran las principales características del servicio que se ofrece relativo a la calidad del aire en Euskadi:

- El Gobierno Vasco dispone de un total de 83 estaciones medioambientales esparcidas por todo el territorio vasco, habiendo una mayor densidad de ellas en Guipúzcoa y Vizcaya. Es posible acceder a la información mediante un mapa disponible en la web, en el cual se muestra la localización de cada una de ellas.



*Figura 10: Captura de mapa con las estaciones medioambientales <sup>6</sup>*



*Figura 11: Estación medioambiental de la Av. Tolosa*

- Aunque no todas las estaciones miden lo mismo, la mayoría de ellas ofrecen datos de casi todos los contaminantes para los cuales se establecen límites de cara a la protección de la salud (establecidos en el Real Decreto RD 102/2011): SO<sub>2</sub> (dióxido de azufre), NO<sub>2</sub> (dióxido de nitrógeno), PM<sub>10</sub>, PM<sub>2,5</sub>, CO (monóxido de carbono), O<sub>3</sub> (ozono). En el caso

---

<sup>6</sup> URL:

<http://www.ingurumena.ejgv.euskadi.eus/r49-aa17a/es/aa17aCalidadAireWar/estacion/mapa?locale=es>



del C<sub>6</sub>H<sub>6</sub> (benceno), los metales pesados [Pb (plomo), As (arsénico), Cd (cadmio), Ni (níquel)] y el B(a)P (Benzo(a)pireno), es menor el número de estaciones debido a que su evaluación se hace de manera global para todo el territorio vasco.

- Además de los anteriormente mencionados, también se ofrecen datos de contaminantes para los cuales no se han establecido límites: Óxidos de nitrógeno [NO<sub>x</sub> (óxidos de nitrógeno) y NO (monóxido de nitrógeno)], compuestos orgánicos volátiles (COVs) (tolueno y diferentes xilenos), 16 metales pesados más (vanadio, cromo, hierro, mercurio, cobalto, selenio...) e hidrocarburos policíclicos aromáticos (HAPs) (un total de 6 diferentes).
- También se ofrecen datos medioambientales no relativos a partículas contaminantes como la dirección y velocidad del viento, humedad, temperatura y niveles de pluviosidad.
- La resolución temporal con la que se ofrecen los datos es por horas o diaria, y según la experiencia obtenida al usar el servicio durante este proyecto, lo más habitual es tener disponibles los datos transcurridas 2-3 horas desde el momento de su medición. Además, se ofrecen los datos históricos de los 6 años previos aproximadamente.

Por un lado, la posibilidad de tener acceso a datos obtenidos con instrumentos de medición muy precisos, de forma online y procedentes de una institución gubernamental (con la confianza que se le presupone) ha sido un gran punto a favor. Por otro, el no tener otra fuente de datos fiable disponible ha hecho tener que adaptar el estudio a la única fuente disponible.

Esto ha generado un par de restricciones que se han tenido que afrontar. Ha sido necesario obtener lecturas por horas (resolución con la que se muestran los datos en la web), y hacerlo en las cercanías de alguna de las estaciones meteorológicas esparcidas por la ciudad (5 en total, una al lado de la UPV/EHU). Además, ha sido necesario hacer la conversión de partículas/volumen a peso/volumen (concretamente de *partículas/0,01 pies cúbicos* a  $\mu\text{g}/\text{m}^3$ ), ya que las unidades de los datos obtenidos con el sensor y las de las estaciones no coinciden.

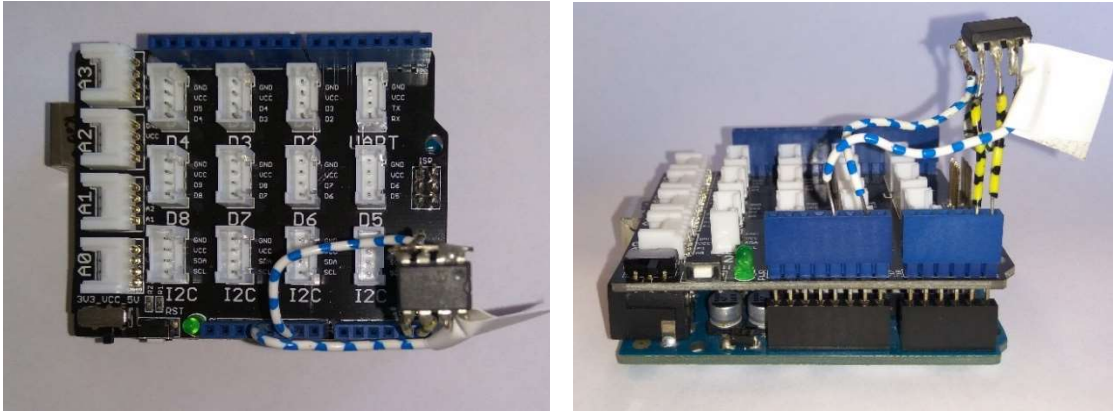
Para realizar esta conversión, es necesario conocer el peso medio de las partículas contaminantes, dato que no se ha logrado obtener de ninguna fuente. Es por ello que se ha tenido que realizar la hipótesis de que **las partículas pesan 7 picogramos** tomando como referencia los datos de las primeras lecturas realizadas. En caso de ser falsa, los datos “en bruto” se mostrarán en una escala distinta a la real, error que se solucionará al realizar una normalización de los datos.

### 3.3.1.2 Ampliación de la memoria del Arduino

El Arduino utilizado en este proyecto dispone de 32KB de memoria flash para almacenar los programas, 2KB de memoria SRAM (static random access memory) para almacenar variables en tiempo de ejecución, y tan solo 1KB de memoria EEPROM de cara a que el usuario almacene ahí datos que después quiera recuperar.

Teniendo en cuenta que el tiempo de muestreo fijado en el programa es de un minuto, y que la información de los sensores se almacena en variables de tipo Byte, tomando muestras con los cinco sensores a la vez (es decir, se almacenan 5 bytes por minuto) la memoria original tan solo permite almacenar muestras durante tres horas y media aproximadamente, lo cual es escaso de cara al proyecto.

Por este motivo se ha añadido una memoria externa EEPROM de 32KB (24LC256) conectándola al bus I2C del Arduino, además de incorporar dos funciones en el código que permiten hacer lecturas y escrituras en la nueva memoria. Teniendo 32KB de memoria disponible, el tiempo máximo de toma de muestras asciende hasta más de 100 horas.



*Figura 12: Inclusión de la memoria al Arduino*

### 3.3.1.3 Preparación del software

El software original proporcionado por el distribuidor está diseñado para ser utilizado en un Arduino y obtener los datos de un solo sensor. Debido a que se ha utilizado un único Arduino, se ha modificado el código para poder obtener lecturas de varios sensores a la vez (entre 1 y 5), repartiendo el tiempo de obtención de los datos entre todos los sensores (el Arduino no permite una programación concurrente). Lo ideal en caso de no depender de las estaciones meteorológicas del Gobierno Vasco (se ha de desplazar en varias ocasiones todo el material a un lugar cercano) habría sido utilizar un Arduino por sensor.

Partiendo de que el inicio del estudio se basa en la comprobación de las diferencias entre los sensores, se ha presupuesto que la exposición de los sensores a las partículas contaminantes es la misma a pesar de que las lecturas no las hagan de forma simultánea, sino secuencial (se captura la señal de un sensor durante 1 segundo y se pasa al siguiente). Dicho de otra forma, se presupone que el periodo de lectura fijado para los sensores (1s, de manera que por cada 5s se registra la señal de un sensor 1s) no es lo suficientemente grande como para producirse cambios (entiéndase, significativos) en la concentración ambiental de partículas, pudiendo tomar como válidos los 12 minutos de exposición (por cada hora) de cada sensor para obtener la concentración media de 1h. La recogida de datos se puede modelar de la siguiente forma:

$n$  = número de datos recolectados en 1 hora utilizando un sensor

$N$  = número de sensores utilizados

$a_i$  = Medida número  $i$  obtenida por el sensor "a" durante 1h de lecturas continuas

$a_j$  = Medida número  $j$  obtenida por el sensor "a" durante  $1/N$  horas de exposición, divididas en periodos de tiempo de 1s distribuidos de forma homogénea en 1h

$$\frac{\sum_{i=1}^n a_i}{n} = \frac{\sum_{j=1}^{n/N} a_j}{n/N}$$

Eso sí, que las diferencias entre los sensores no varíen en términos relativos al adaptar el código para poder tomar lecturas con varios no quita que sí vaya a haber diferencias (o que pueda haberlas) entre los datos que se logran con un mismo sensor utilizando el código original o el nuevo código. Se ha comprobado que la corriente suministrada a un sensor es distinta en un caso y otro debido a que la bajada de tensión que se produce en los puertos del Arduino es más acusada cuantos más sensores se conectan.

Nº de sensores conectados	1	2	3	4	5
Consumo del sensor 5	80,2mA	78,4mA	76,7mA	75,0mA	73,5mA

Figura 13: Consumo del sensor en función del número de sensores empleados

En todo caso, esto no debería suponer un impedimento para llevar a cabo el estudio, más allá de que el ajuste de los datos inicial sea distinto tomando valores con cinco sensores en vez de uno.

Por otro lado, el software original se limitaba a imprimir por pantalla 3 datos: el tiempo del sensor en baja, el porcentaje de tiempo en baja y la concentración de partículas (partículas por pie cúbico), todos ellos almacenados en variables de tipo float (4 bytes). Para llevar a cabo el proyecto ha sido necesario almacenar la información pertinente para su posterior análisis, por lo que se han llevado a cabo ligeras modificaciones. Ahora la variable "ratio" que indica el porcentaje de tiempo que el sensor ha tenido una partícula en su zona de detección se almacena en la memoria EEPROM añadida (más información en el apartado 3.3.1.3), y en vez de ser de tipo float es de tipo byte con el objetivo de ahorrar espacio en la memoria (el rango de valores permitía dicho cambio). Los cálculos necesarios para extraer la concentración a partir del ratio, dato que se busca finalmente, se han eliminado y se realizan a posteriori en la hoja de cálculo sobre la que se vuelcan los datos, ahorrando así tanto memoria como tiempo de ejecución.

```

void loop() {
  for (i=0; i<snum; i++) lowpulseoccupancy[i] += pulseIn(pin[i], LOW); //pulseIn analiza la señal durante 1s

  if ((millis()-starttime) > sampletime_ms) //Verdadero cuando el tiempo de muestreo es superado
  {
    for (i=0; i<snum; i++){
      ratio = lowpulseoccupancy[i]/((sampletime_ms/snum)); // Integer percentage 0=>100
      //concentration = (byte) 1.1*pow(ratio,3)-3.8*pow(ratio,2)+520*ratio+0.62; // using spec sheet curve
      //Serial.println(concentration);
      if (eeAddress <= 32000){
        writeEEPROM(disk1, eeAddress, ratio); //Escritura en memoria
        eeAddress += sizeof(byte);
      }
      lowpulseoccupancy[i] = 0;
    }
    starttime = millis();
  }
}

```

Figura 14: Parte principal del código para tomar lecturas

Por último se ha realizado el programa necesario para extraer los datos de forma ordenada, ya que a la hora de almacenar los datos, estos se guardan en posiciones de memoria contiguas, quedando entrelazados los datos de todos los sensores.

### 3.3.1.4 Almacenamiento de los datos para su tratamiento

Tras cada lectura de datos, éstos se han volcado en una hoja de cálculo de Google (Google Sheets) debido a las opciones de tratamiento de datos que ofrece y a la posibilidad de editar desde el lugar donde se ha almacenado el sistema de información (se ha utilizado Google Drive). Además de para almacenar los datos, la plantilla creada sirve para realizar la conversión de ratio (o porcentaje del tiempo en baja, dato que se almacena en la toma de datos) a concentración de partículas ( $\mu\text{g}/\text{m}^3$ ).

Esto se realiza en una única operación que incluye 2 conversiones: la primera convierte el ratio en concentración de partículas (*partículas/0,01 pie cúbicos*) utilizando la función ofrecida por el suministrador de los sensores (Fig. 14). La segunda realiza la conversión de *partículas/0,01 pie cúbicos* a  $\mu\text{g}/\text{m}^3$ , para lo cual se ha establecido un peso medio de las partículas de 7 picogramos. Como ya se ha comentado previamente, la primera conversión la realiza el software original en el propio Arduino, pero se ha modificado para ahorrar espacio en memoria.

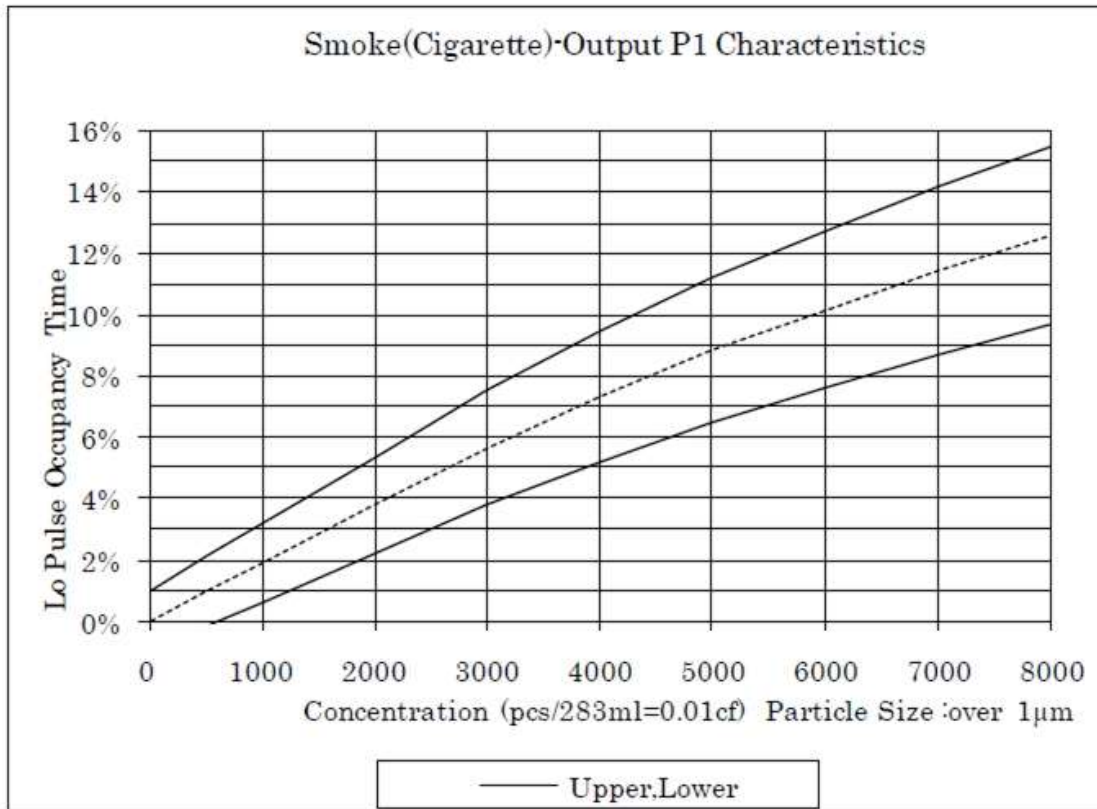


Figura 15: Relación entre el porcentaje de tiempo en baja y la concentración <sup>7</sup>

Por otro lado, la plantilla agrupa los datos que se almacenan cada minuto en horas, dato que finalmente se desea lograr, y se apuntan aquí los meta datos relativos a cada sesión de lectura (Tiempo efectivo, lugar, tiempo de muestreo, número de sensores empleados, software empleado, meteorología, tipo de alimentación, número de muestras obtenidas, hora de inicio y final y tensión utilizada).

<sup>7</sup> URL: <http://wiki.timelab.org/images/f/f9/PPD42NS.pdf>

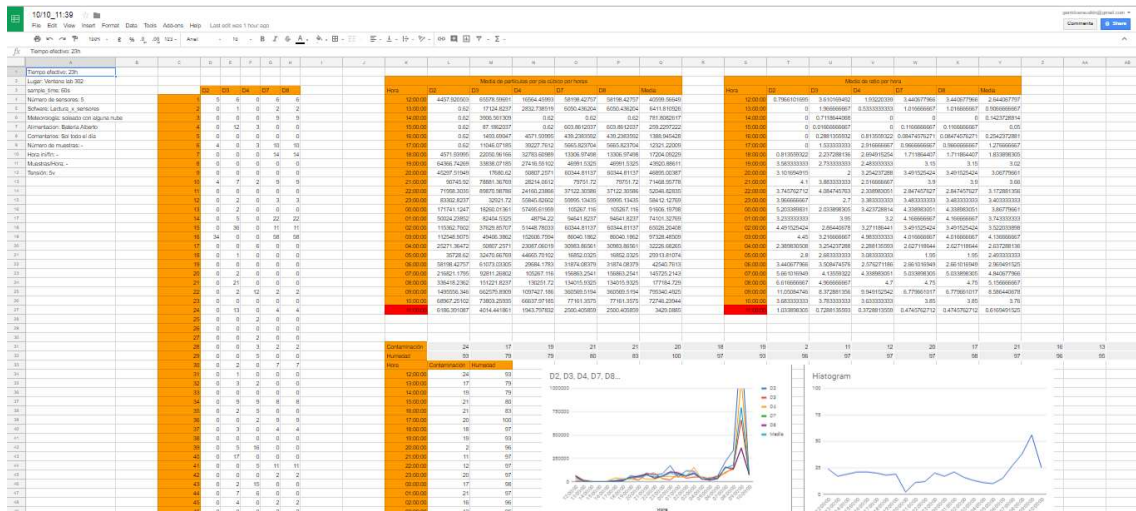


Figura 16: Plantilla para almacenar los datos

### 3.3.1.5 Alimentación y estimaciones de consumo

Con la intención de comprobar que las condiciones de los sensores a la hora de tomar datos son las mismas para todos ellos, se ha medido la corriente que suministra el Arduino a un mismo sensor conectándolo en los diferentes puertos, así como la corriente suministrada a los distintos sensores por un mismo puerto.

Puerto	D2	D3	D4	D7	D8
Sensor 8	73,3mA	73,3mA	73,2mA	73,4mA	73,4mA

Sensor	1	2	3	4	5
Puerto D8	73,4mA	73,4mA	73,2mA	73,5mA	73,4mA

Figura 17: Consumos con diferentes puertos y sensores

Las diferencias obtenidas son notables, y pueden ser debidas a que las mediciones obviamente se han realizado de forma secuencial (no simultánea), así como por el distinto calibrado de fábrica de cada uno de los sensores, que puede suponer pequeñas variaciones en la resistencia total de cada sensor.

Por otro lado, se ha estimado el consumo total del Arduino con los 5 sensores conectados, de cara a saber cuánto tiempo podrá alimentar una batería el circuito. Así como para conocer el consumo de los sensores se han realizado mediciones con un multímetro (cortando uno de los cables disponibles y haciendo un puente), el consumo del Arduino se ha obtenido de mediciones de terceros (48 mA aproximadamente).

		Consumo total (Arduino + sensores)				
		Con 1 sensor	Con 2 sensores	Con 3 sensores	Con 4 sensores	Con 5 sensores
		128 mA	205 mA	278 mA	348 mA	415 mA
Duración estimada de una batería en función de su capacidad	1000 mAh	7,8h	4,9h	3,6h	2,9h	2,4h
	10000 mAh	78h	49h	36h	29h	24h
	16000mAh	125h	78h	58h	46h	39h

*Figura 18: Estimación de duración de las baterías*

Además, debido a que el consumo en este tipo de sensores es una característica importante a tener en cuenta, se van a tomar algunas lecturas alimentando los sensores con 3,3V para determinar su posible uso con un consumo inferior.

### 3.3.2 Variables del estudio

Tras conocer el funcionamiento de los sensores y las recomendaciones de uso extraídas de experimentos de terceros, se enumeran a continuación todas las variables que pueden tener relevancia durante el estudio:

- **Humedad:** Teniendo en cuenta que los sensores funcionan mediante la captación de la luz que reflejan las partículas al pasar frente al diodo, es posible que partículas pequeñas de agua tengan el mismo efecto que las partículas contaminantes en caso de situarse en la zona de detección del sensor.
- **Temperatura:** Por un lado, variaciones en la temperatura ambiental podrían suponer variaciones en la fotorresistencia que capta los reflejos de luz producidos en las partículas, siendo el sensor más o menos sensible. Por otro, es necesario que la temperatura dentro de la cápsula del sensor sea mayor que la temperatura externa para generar una corriente de aire ascendente. Temperaturas externas elevadas podrían provocar una disminución en la diferencia de temperatura entre ambas partes y afectar o eliminar la corriente de aire necesaria para el funcionamiento del sensor.
- **Viento:** Tener la entrada y salida de aire del sensor expuestas a vientos externos podría afectar a la circulación de las partículas en la cápsula de sensor.
- **Tensión de alimentación:** Cambios de tensión afectan a la resistencia, al diodo y al fotodiodo, elementos fundamentales del sensor a la hora de tomar lecturas.
- **Exposición a la luz:** En función de lo bien o mal que esté aislado el foto diodo de la luz externa, la incidencia de luz directa sobre el sensor podría provocar variaciones en las lecturas.
- **Otros elementos:** En general, el no disponer de un filtro de aire que evite la entrada de partículas mayores de 10 micrómetros a la zona de detección puede suponer que cualquier elemento que entre en la cápsula del sensor (ej. un insecto) altere por completo las lecturas.

De cara al estudio de la incidencia de cada una de las variables, es posible aislarlas de dos formas distintas: generando entornos controlados en un laboratorio o estadísticamente (es decir, recabando muchos datos y ordenándolos por rangos de valores de las distintas variables). No es viable estudiar la incidencia de todas ellas en este trabajo de ninguna de las formas, ya sea por falta de recursos materiales y/o falta de tiempo.

Por ejemplo, aunque la temperatura pueda ser un aspecto a tener en cuenta, no existe la posibilidad de tomar las medidas en un entorno con la temperatura y la contaminación PM controlada, y en caso de decidir simplemente depender de la temperatura ambiental para estudiar su incidencia (es decir, hacerlo de forma estadística), las combinaciones con otras variables harían necesaria una cantidad de datos ingente.

Dicho lo cual, el trabajo se ha limitado a estudiar la incidencia de **la humedad y la exposición a la luz**.

### 3.3.3 Implementación del experimento

Una vez satisfechas las necesidades de conocimiento y material para llevar a cabo el experimento, en este apartado se detallan los pasos seguidos finalmente para realizar las lecturas de los datos y analizarlos. Esta etapa se ha dividido en tres fases claramente distinguibles (montaje, toma de datos y análisis) tanto por la propia naturaleza de las tareas realizadas en cada una de ellas como por su distribución en el tiempo.

#### 3.3.3.1 Montaje

Se han realizado dos montajes distintos para poder realizar las lecturas, ya que el primero de ellos ha sufrido un deterioro excesivo durante la realización del proyecto. En ambos se ha incluido todo el material necesario para la toma de datos (Arduino con la base y la memoria añadidas, los 5 sensores y la batería), se han colocado los sensores en posición adecuada (vertical, con la entrada de aire en la parte inferior), y se han realizado los orificios oportunos para permitir la entrada y la salida del aire.

El primero de ellos está realizado con una caja de cartón, motivo por el cual se ha deteriorado durante los desplazamientos y la exposición a la humedad. El segundo en cambio, más elaborado, está realizado con piezas recortadas de un tablero de madera aglomerado, y en su interior se diferencian el espacio de los sensores y el espacio del Arduino y la batería. Con este último, es posible realizar lecturas en condiciones más aisladas para los sensores (de viento y luz) y durante días de meteorología adversa, ya que impide la entrada de la lluvia.





Figura 19: Montajes para los sensores

### 3.3.3.2 Toma de datos

El objetivo principal a la hora de tomar lecturas con los sensores ha sido lograr una cantidad de datos lo suficientemente grande como para estudiar el efecto de las variables previamente mencionadas (luz y humedad) y conocer el comportamiento general del sensor.

#### 3.3.3.2.1 Control de las variables

La incidencia de la luz ha sido una variable que se han podido modificar a voluntad, y se han diferenciado dos niveles de exposición: sensor expuesto (dentro del primer montaje) y sensor aislado (dentro del segundo montaje y con el agujero de la cápsula tapado (Fig. 19)). Además, durante las lecturas con el sensor expuesto, se podrán estudiar las diferencias entre las horas de luz y la noche.

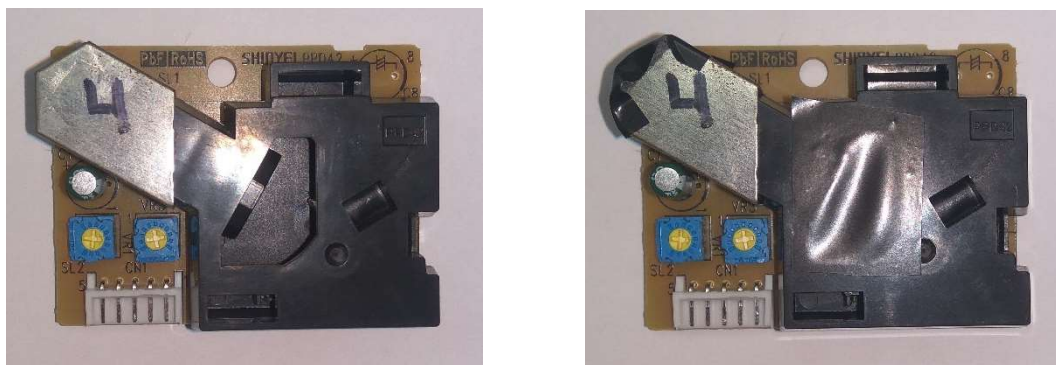


Figura 20: Sensor con agujero de la cápsula con y sin tapar

La obtención de datos con distintos niveles de humedad en cambio ha dependido de los cambios de humedad ambientales habidos durante todas las sesiones de toma de datos, y no se ha estudiado su efecto hasta tener una muestra lo suficientemente grande. Tras 216h de datos en total, un porcentaje muy alto de lecturas ha resultado ser en condiciones de humedad alta

(80%-100%), tal y como cabía esperar teniendo en cuenta el lugar y época del año en el que se ha llevado a cabo el estudio, aunque también se han obtenido lecturas con humedades ambientales más bajas.

Exposición a luz		Humedad		
Expuesto	Aislado	0%- 60%	60% - 80%	80% - 100%
99h	117h	16h	59h	141h

*Figura 21: Horas de medición*

### 3.3.3.2 Lugares de lectura

A excepción de en dos sesiones, el resto de la veces los datos se han obtenido desde la Facultad de Informática, concretamente desde la ventana del laboratorio 302 y la azotea, esta última debido a la protección que ofrecen a la lluvia las placas solares ahí instaladas.

Las otras dos ocasiones se han tomado las lecturas desde el tejado de una de las cabinas meteorológicas del gobierno vasco (la situada en Puyo) a modo de comprobación previa, para detectar (en caso de que las hubiera) variaciones significativas entre las lecturas tomadas al lado de una de estas estaciones y las tomadas a cierta distancia (concretamente la distancia entre la facultad de informática y la estación meteorológica situada en la Avenida de Tolosa, junto a la biblioteca de la universidad, 200m aproximadamente).

Nótese que la mayoría de las lecturas no han sido tomadas al lado de una estación meteorológica debido al riesgo que supone dejar el material expuesto y accesible a cualquier viandante, y/o la imposibilidad de realizar las lecturas de forma presencial (demasiadas horas).

### 3.3.3.3 Análisis de los datos

El análisis de los datos ha girado en torno al cálculo del coeficiente de determinación para determinar el grado de relación entre las fluctuaciones de unos datos y otros. Además, se ha visualizado la nube de puntos de la regresión de cada sensor en diferentes escalas para poder determinar su comportamiento y poder realizar el ajuste de datos más adecuado.

## 4. Resultados del experimento

## 4.1. Observación inicial de los datos obtenidos

---

Tomando todos los datos obtenidos<sup>8</sup> en su conjunto (216h) y sin ningún tipo de tratamiento, los resultados al calcular el coeficiente de determinación de cada uno de los sensores son los siguientes:

	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5
PM10	0,30	0,20	0,31	0,23	0,29
PM2,5	0,11	0,07	0,14	0,09	0,13

*Figura 22: Coeficientes de determinación iniciales*

Se puede observar que todos los sensores obtienen resultados parecidos, lo cual por sí solo es una buena señal. Queda además patente que los sensores presentan mayor covarianza con la contaminación PM10, y en general una mayor correlación con uno de los conjuntos de partículas supone mayor correlación con el otro también. Algo lógico teniendo en cuenta que los valores de PM10 y PM2,5 presentan un alto grado de covarianza.

Observando las gráficas se puede comprobar cómo hay momentos en los que los valores son bastante próximos (Fig. 22), y otros en los que los sensores parecen verse afectados por alguna variable, llamando especial atención los distintos picos de valores elevados, así como momentos en los que los sensores no parecen detectar casi nada (Fig. 23).

Por otro lado, los diagramas de puntos revelan de nuevo la presencia de valores espurios, habiendo puntos fuera de la zona más poblada. Asimismo, observándolos en una escala menor se aprecian numerosos valores del sensor próximos a cero no explicados por la contaminación. En general, las nubes de puntos son bastante dispersas y no muestran ninguna forma muy evidente, pero sí dejan entrever que la relación existente es lineal.

---

<sup>8</sup> Hojas de cálculo con los datos:

<https://drive.google.com/drive/folders/1BYcIVGJSjTMFwQaJqKz9Sr8hfEs7noTt?usp=sharing>

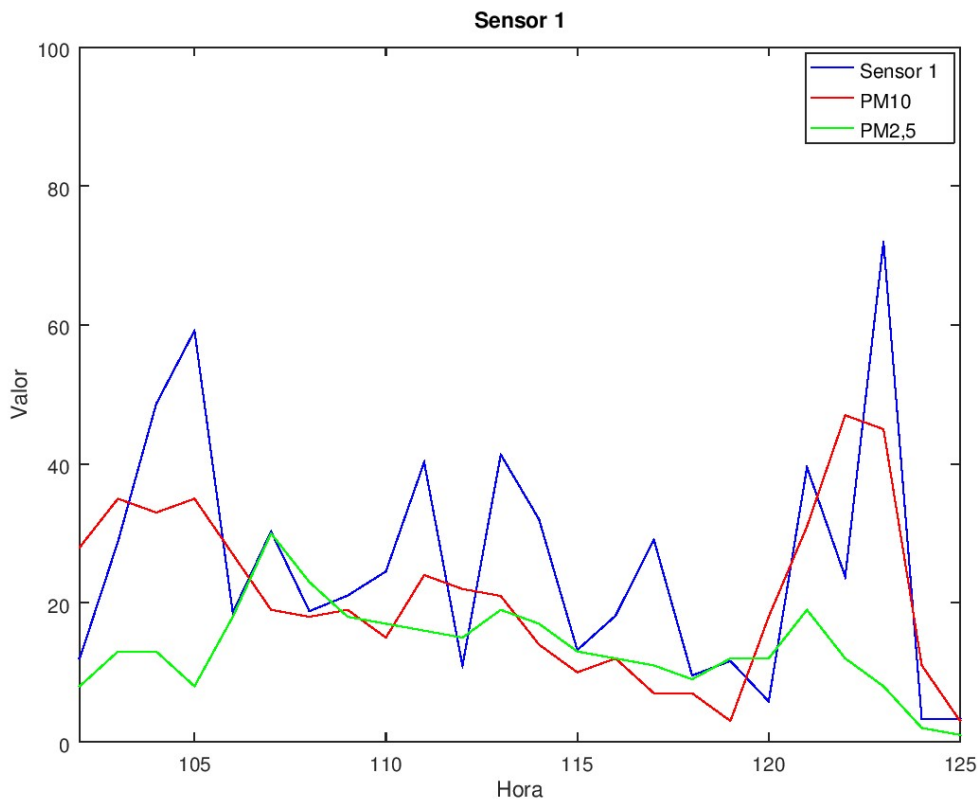


Figura 23: Muestra con valores próximos

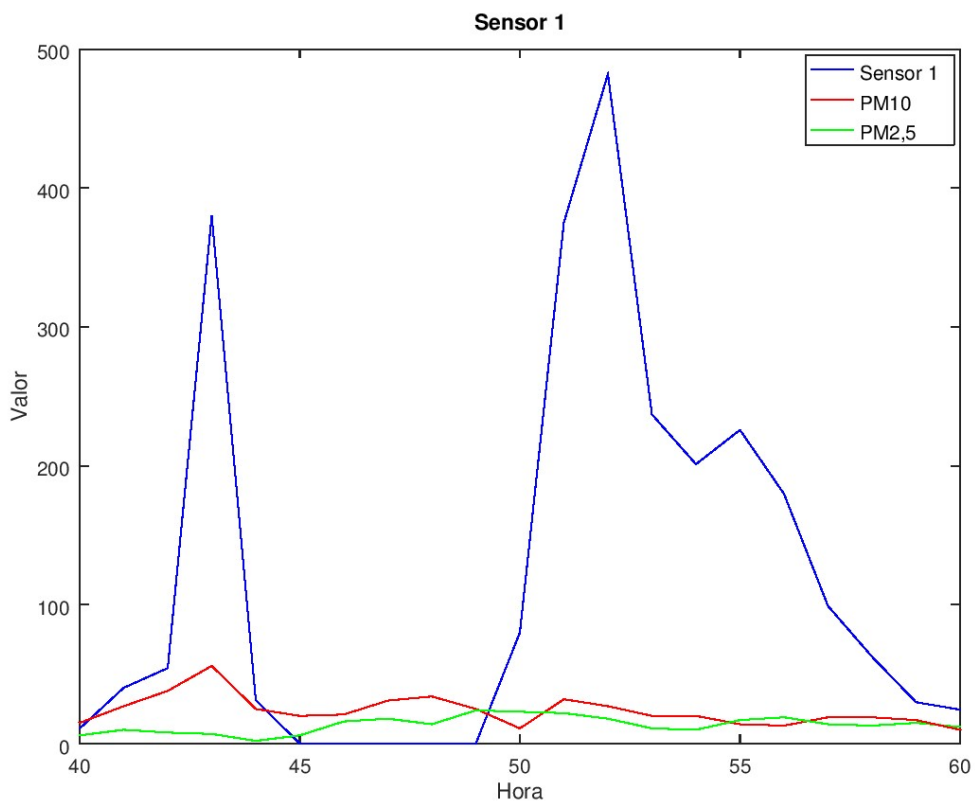


Figura 24: Muestra con valores muy dispares

En cuanto a la comparación de los sensores entre ellos mismos, todos obtienen valores similares la mayor parte del tiempo, y como consecuencia su correlación es muy alta, de manera que no ha sido necesario descartar ninguno para el análisis.

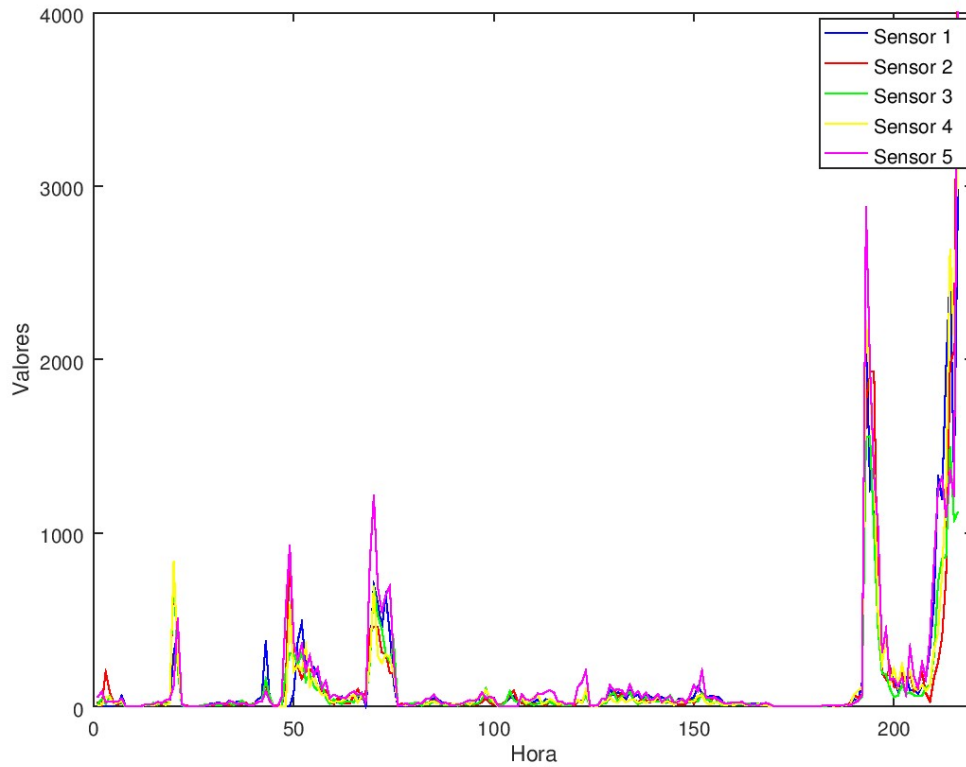


Figura 25: Secuencia de medidas de cada uno de los sensores

Sensor	1	2	3	4	5
1	1	0,87	0,91	0,94	0,91
2	-	1	0,83	0,96	0,91
3	-	-	1	0,91	0,86
4	-	-	-	1	0,91
5	-	-	-	-	1

Figura 26: Correlación entre sensores

## 4.2. Efectos de la luz y la humedad

---

Separando por un lado los datos de las horas en las que las lecturas se han tomado con los sensores expuestos a la luz ambiental y por otro los datos tomados de forma aislada, se han obtenido los siguientes coeficientes de determinación:

		Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5
Expuestos	PM10	0,40	0,18	0,29	0,19	0,40
	PM2,5	0,19	0,20	0,23	0,20	0,36
Aislados	PM10	0,38	0,28	0,38	0,33	0,32
	PM2,5	0,17	0,10	0,16	0,13	0,13

*Figura 27: Coeficientes de determinación con los sensores aislados y expuestos*

En el caso del PM2,5 la variación de todos es en la misma dirección, pero curiosamente los coeficiente disminuyen, cuando en caso de que la luz tuviera alguna incidencia lo lógico sería obtener resultados mejores al aislar los sensores. En cualquier caso, tal y como ha quedado plasmado en el apartado 4.1., el sensor no es muy buen indicador de la concentración de PM2,5, por lo que la validez de esta comparativa es escasa o nula.

En el caso del PM10, cuyo estudio tiene más sentido, estando los sensores aislados de la luz el coeficiente de determinación aumenta para algunos sensores y disminuye para otros, sin ser en ningún caso una variación llamativa. Es probable que estas diferencias disminuyesen con una mayor cantidad de datos, ya que el cálculo del coeficiente de determinación es sensible al número de muestras. Además, por si quedaran dudas, observando los datos de los días con los sensores expuestos no se detectan variaciones entre las horas de luz y la noche que indiquen la posibilidad de que la luz incida en los sensores de forma significativa.

Para comprobar los efectos de la humedad se han calculado directamente los coeficientes de determinación entre los datos de cada sensor y la humedad, y el resultado han sido valores muy próximos a 0 (del orden de  $10^{-3}$ ). Incluso aislando los datos del rango de humedad más alto (90%-100%) el resultado es el mismo a pesar de la advertencia del distribuidor de que en situaciones de más del 95% de humedad las medida pierden precisión.

## 4.3. Tratamiento de los datos

---

En el tratamiento que se muestra a continuación, los datos de los sensores tan solo se han comparado con los valores de PM10, debido a haber mostrado un mayor grado de covarianza con este tipo de partículas. Como se ha comprobado, los datos de los sensores presentan una correlación muy alta entre ellos, por lo que no es posible filtrar los datos de uno utilizando los datos del resto. Observando el diagrama de dispersión del sensor (de cualquiera de ellos) (Fig.

27), llaman la atención dos tendencias que describen el comportamiento del mismo durante este estudio.

La primera de ellas es que cuando la contaminación se mantiene en valores inferiores a  $30 \mu\text{g}/\text{m}^3$ , aunque la nube de puntos es muy dispersa y no se observa ninguna tendencia clara en ella, los valores del sensor tienden a agruparse dentro del rango [0-500] (el rango total es [0-3500] aproximadamente). Dicho lo cual, también se observan dentro de ese rango (del sensor) momentos en los cuales la contaminación es significativamente mayor, y sin embargo los sensores no muestran evidencias de ello.

La segunda y más importante, es que el sensor nunca ofrece valores altos cuando la contaminación es baja, es decir, no da falsos positivos. Para que el sensor ofrezca un determinado valor, es requisito indispensable que la contaminación sea superior a otro cierto valor, o dicho de otra forma, el valor de la contaminación determina un valor límite superior que el sensor no es capaz superar. De esta manera, valores altos del sensor aseguran un contaminación alta (entiéndase, dentro del rango de valores en los que ha oscilado durante el estudio).

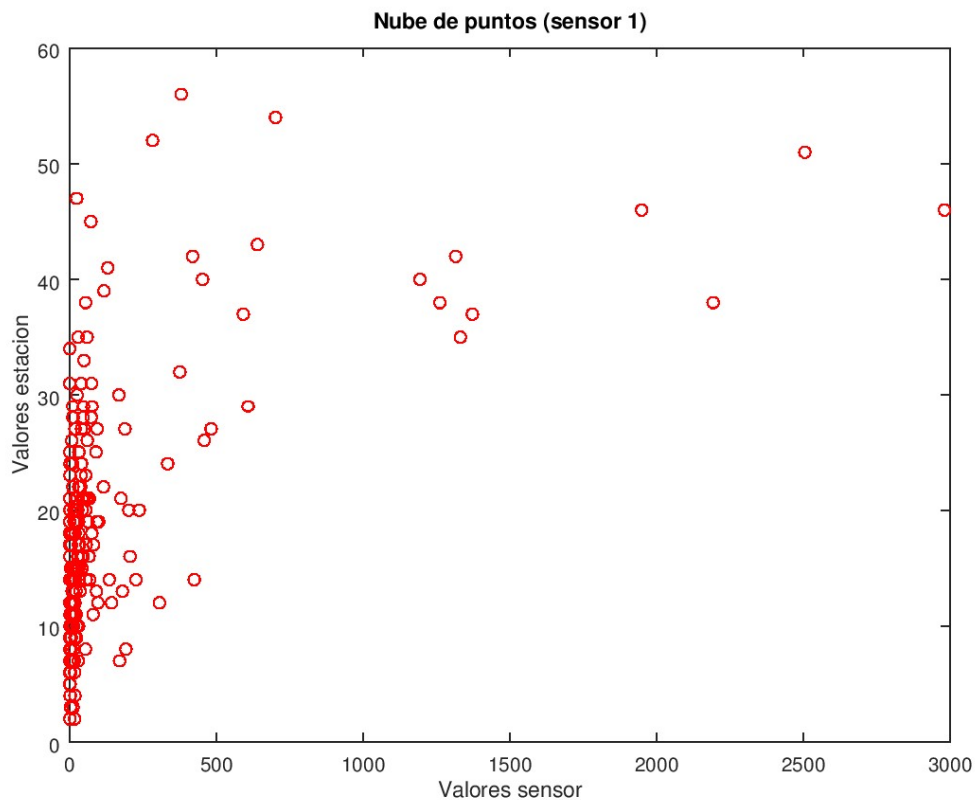


Figura 28: Diagrama de dispersión (sensor 1)



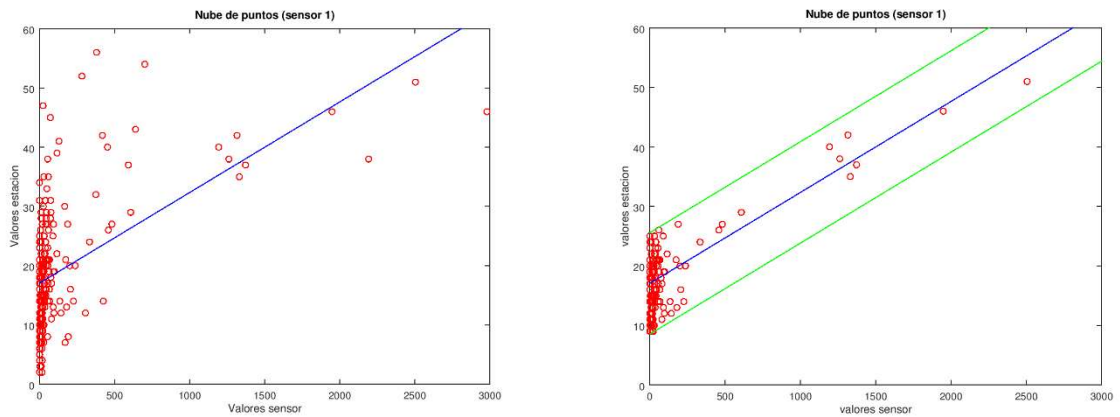


Figura 29: Filtrado de datos

Teniendo en cuenta lo anterior, a continuación se muestra un posible filtrado de los datos. Cogiendo como referencia la recta de regresión de cada sensor y calculando dos rectas paralelas a cierta distancia (una por encima y otra por debajo), se obtiene una franja que sirve para delimitar qué valores se toman como válidos y cuáles no, eliminando valores que se consideran demasiado alejados de la nube de puntos. Además, previo este filtrado, se han eliminado todos los valores de los sensores inferiores a uno por ser numerosos y claramente erráticos. Por último se ha realizado el ajuste lineal de los datos. A continuación se muestran los resultados obtenidos, indicando el porcentaje de datos conservado tras el filtrado.

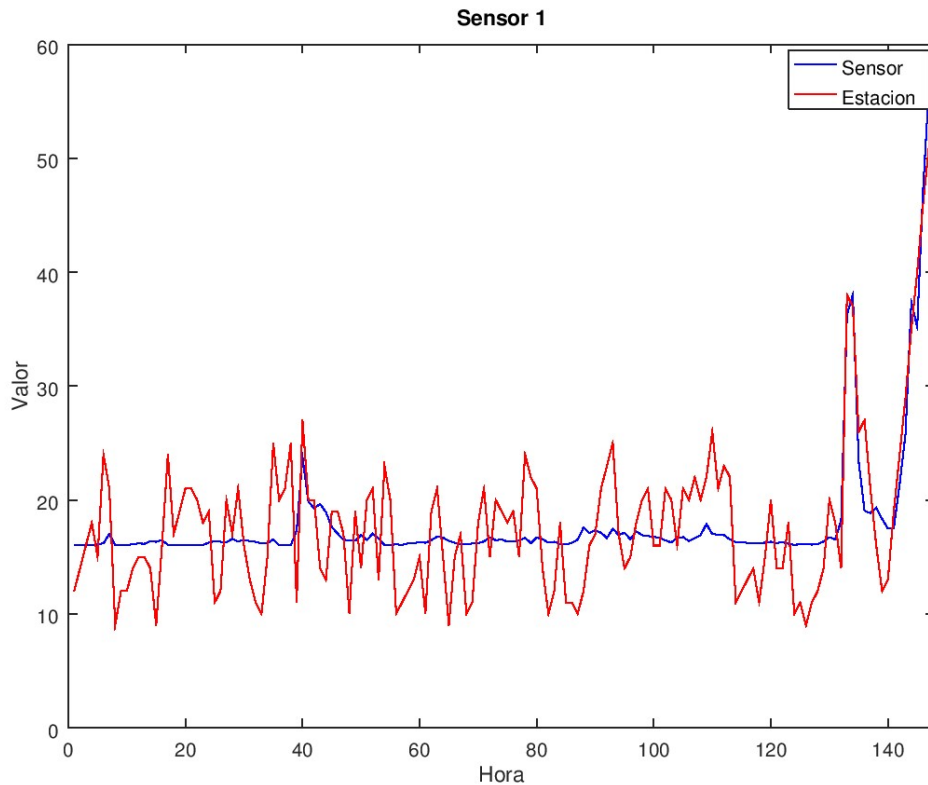


Figura 30: Valores del sensor 1 tras filtrado

	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5
Porcentaje de datos conservados	60%	64%	65%	65%	64%
Coeficiente de determinación	0,65	0,46	0,63	0,52	0,51

*Figura 31: Coeficientes de determinación tras filtrado*

Como vemos, eliminando 1/3 de los datos se consiguen coeficiente de determinación significativamente mayores. De lograrse estos coeficientes sin necesidad de filtrados, el sensor podría considerarse un buen indicador de partículas PM10. Aunque definitivamente no serviría para obtener valores precisos, especialmente con niveles de contaminación bajos, si serviría para detectar situaciones de contaminación elevada. En el caso concreto de la gráfica previa, los valores del sensor y la estación coinciden en el rango de calidad del aire en casi todo momento, si bien es cierto que los datos tan solo oscilan entre dos (*muy buena* [0-25] o *buena* [25-50]), a excepción de un único valor superior a 50 (rango *mejorable*) que el sensor también detecta.

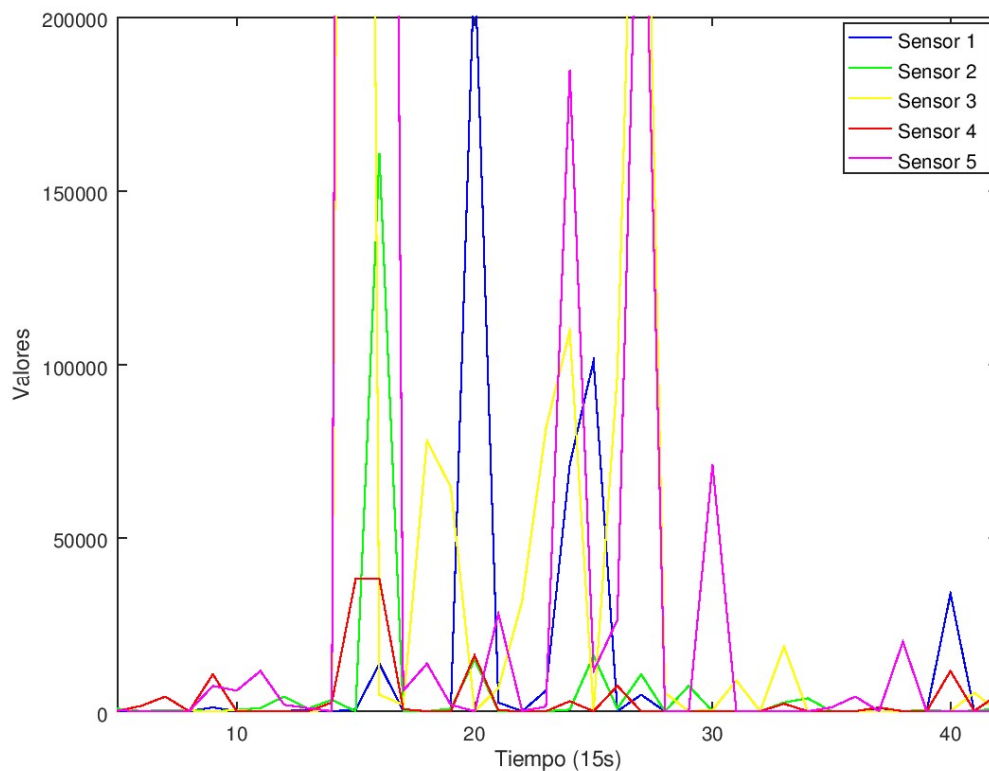
Por último, sabiendo que los valores altos del sensor siempre corresponden a valores de contaminación elevados, se ha hecho la prueba de utilizar siempre el valor más alto de los disponibles (de entre los 5 sensores el mayor) para comprobar el resultado, pero este ha sido prácticamente idéntico al obtenido por el sensor 1 (0,65 de coeficiente), una evidencia más de que todos los sensores ofrecen valores próximos.

#### **4.4. Prueba en condición de contaminación elevada**

---

Una de las limitaciones que ha tenido el proyecto ha sido no disponer de una zona de mayor contaminación con una estación meteorológica cerca con la que contrastar los datos, de manera que el rango de valores de PM10 y PM2,5 obtenidos no ha sido muy amplio. Este hecho es importante ya que cuanto menores son las fluctuaciones del dato a medir, más difícil es determinar la relación existente entre los datos de referencia y los datos obtenidos por el sensor en cuestión, y por este motivo datos estadísticos como el coeficiente de determinación son sensibles al número de datos utilizados para su cálculo (cuanto mayor es el número de datos más fiable es la muestra y mayor es la probabilidad de que el rango de los datos aumente).

Teniendo esto en cuenta, se han tomado lecturas en condiciones de alta concentración de partículas (colocando los sensores junto al escape de gases de un vehículo diésel) para comprobar la respuesta de los sensores.



*Figura 32: Medición en condiciones de alta contaminación*

Las lecturas han sido con intervalos de tiempo de 15s, y el vehículo ha estado con el motor en marcha entre los instantes 14 y 27. Como se puede comprobar todos los sensores ofrecen valores significativamente más altos en ese periodo (más incluso que en una situación de contaminación ambiental elevada tomando lecturas con este periodo). También se aprecia que los sensores no se comportan muy bien tomando lecturas durante periodos cortos, ya que fluctúan continuamente.

Por este motivo, y quizá por el hecho de hacer las lecturas por turnos (explicado en el apartado 3.3.1.3), en esta ocasión los valores entre los sensores pueden ser muy distintos en un momento dado, habiendo algunos que ofrecen valores máximos y otros más bien bajos. También por este motivo no son comparables los valores en periodos tan cortos con valores medidos en periodos largos, ya que estos se suavizan.

## 4.5. Comprobaciones de consumo

---

Las estimaciones de consumo realizadas han resultado no ser muy precisas, ya que una batería con una capacidad de 16.000mAh tan solo ha sido capaz de mantener el Arduino encendido durante 25h aproximadamente, y no 39h como se había estimado. En este caso lo más probable es que la explicación se encuentre en el consumo del propio Arduino (sin contar los sensores), cuyo dato sobre el consumo (adquirido de un trabajo previo) está medido ejecutando un software distinto, y sin ningún periférico conectado. Además, el consumo de los sensores sí se

ha medido en las mismas condiciones en las que se han tomado las lecturas, por lo que no debería haber diferencias.

En las pruebas realizadas con los sensores alimentados con 3,3v se ha obtenido una muestra de 92h. En estas circunstancias los valores obtenidos han sido por lo general muy inferiores a los obtenidos con 5v, algo que cabía esperar, y que no supone un problema ya que los datos posteriormente se normalizan. Sin embargo, y a pesar de que la contaminación ha variado bastante en este periodo (algo positivo de cara al análisis) los resultados obtenidos han sido pésimos para todos los sensores (con coeficiente de determinación en torno a 0,01) a excepción de uno, que ha obtenido un coeficiente de determinación de 0,38 debido a haber mostrado valores más altos en algunos momentos de mayor contaminación. En cualquier caso, queda patente que su funcionamiento bastante peor con este voltaje, y se recomendaría usarlo siempre con 5v.

## 5. Implementación a gran escala

Uno de los objetivos iniciales del proyecto es el estudio de la viabilidad de la implementación de un experimento crowdsourcing a gran escala que permita explotar la capacidad del sensor utilizado.

## 5.1 Modelos de implementación a gran escala

---

Para llevar a cabo un experimento de crowdsourcing a gran escala, sería necesario diseñar y construir un dispositivo completamente funcional, que fuese capaz de almacenar las medidas tomadas por el sensor durante un periodo de tiempo mínimo (tiempo que habría que establecer). En una primera etapa podría emplearse un Arduino para este propósito, pero lo ideal sería diseñar una placa específica de tamaño, precio y consumo lo más reducidos posible. Además existiría la posibilidad de incorporar en un mismo dispositivo varios sensores para medir distintos parámetros medioambientales, tal y como ha hecho Erik Fustes en su proyecto. En este sentido, habría empresas como la propia Wimbitex (con la que ha colaborado Erik) que podrían estar interesadas en participar en su construcción.

Por otro lado, en función de la resolución espaciotemporal deseada, habría que tener en cuenta la masa crítica de participantes necesaria para lograr el objetivo, y para calcularla sería necesario un estudio adicional que determinase cuánto tiempo (de media) estaría en funcionamiento cada dispositivo, las horas de máxima y mínima colaboración, y las posibles zonas que quedarían descuidadas por escasa circulación de los participantes. Además, un mecanismo basado en reputaciones daría más relevancia a los sensores más fiables. Otro aspecto importante sería el coste total de implementación, y las posibles fuentes de financiación.

En general, podríamos considerar diferentes modelos de implementación del crowdsourcing, que se pueden clasificar en función del esquema de comunicación y la disponibilidad de los datos.

Si consideramos el esquema de comunicación tenemos dos posibilidades:

- *Peer to peer*: dos o más dispositivos se comunican de forma oportunista cuando los participantes del estudio crowdsourcing se encuentren a menos de cierta distancia. Este enfoque permite prescindir de la infraestructura necesaria para acceder a Internet.
- Centralizado: cada usuario sube los datos a una plataforma online desde la que luego poder consultarlos, ya sea de forma manual o, por ejemplo, conectando el dispositivo a un smartphone que lo haga de forma automática.

Por otra parte, atendiendo al criterio de disponibilidad se pueden diferenciar dos modelos:

- Tiempo real: Los datos se van procesando a medida que están disponibles para un recalibrado continuo.
- Diferido: Los datos no se comparten hasta tener un cierto volumen que permita un análisis estadístico de mayores garantías.

Obsérvese que habría que buscar un equilibrio entre disponibilidad y precisión.

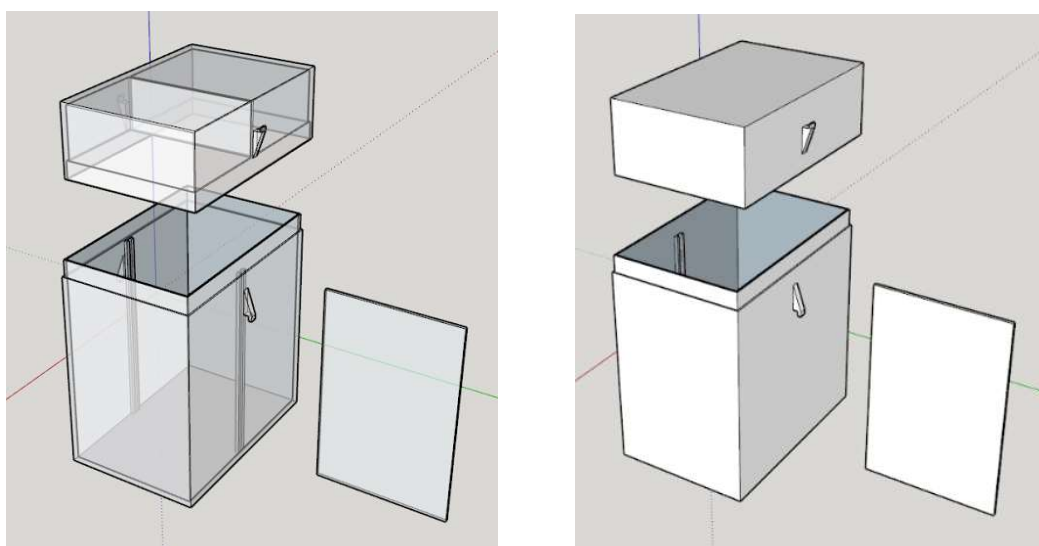
## 5.2 Posible uso del sensor en un experimento a gran escala

Si bien los sensores utilizados en este estudio ofrecen una cierta correlación con la contaminación PM10, no aportan demasiada fiabilidad en su uso como indicadores de estas partículas. En los resultados obtenidos en el capítulo anterior se observa que, si bien no ofrecen falsos positivos, son incapaces de evitar falsos negativos, lo que los descarta en ámbitos de aplicación que requieran un cierto nivel de fiabilidad.

Por otra parte, ha quedado claro en la utilización de un conjunto de estos sensores como indicadores de PM10 que todos ofrecen valores similares en periodos de lectura largos, a pesar de que en periodos cortos se muestran más inestables. Además de los problemas de fiabilidad, este hecho tendría consecuencias si se quisieran utilizar los sensores en experimentos de crowdsourcing a gran escala que involucren por ejemplo a viandantes o ciclistas.

Vistos los resultados obtenidos, es difícil imaginar el uso de estos sensores aplicando el modelo de tiempo real, por su inestabilidad en periodos de lectura cortos. La alternativa de un modelo en diferido en el que se puedan tener en cuenta muchos más datos es también dudosa y sería necesario encontrar previamente la forma de que los datos ofrecidos fueran más fiables, o utilizar otro sensor de similares características y precio con mayor fiabilidad.

De todas formas, tal y como se expone en el siguiente capítulo, durante la investigación se han conocido ciertos aspectos que podrían mejorar los resultados obtenidos, lo cual hace no poder considerar definitivo este análisis de cara a descartar el sensor, y deja la puerta abierta para que un estudio más exhaustivo o un tratamiento estadístico más elaborado descubra la forma de lograr medidas más fiables. En este sentido, se ha avanzado en el diseño de un posible dispositivo en el cual poder integrar un Arduino, una batería y un sensor:



*Figura 33: Diseño de caja para dispositivo de uso individual*

La caja tiene el tamaño justo para albergar los elementos necesarios para tomar las lecturas, y está compuesta por tres elementos: caja, tapa y separador. Este último divide la caja en dos secciones, una en la que iría el sensor y por la que pasaría el aire, y otra en la que irían el resto de elementos. El hecho de que sea extraíble permite un mejor acceso de cara a manipular tanto los elementos interiores como el propio separador, pudiendo colocar en él nuevas sujeciones.



## 6. Conclusión y trabajo futuro

Si bien el comportamiento de los sensores no ha sido del todo positivo, el estudio en sí ha resultado satisfactorio. Se ha comprobado que los sensores funcionan mucho mejor como indicadores de PM10, y que lo más lógico es que los trabajos futuros de investigación estén enfocados en esa dirección. La conclusión obtenida más relevante ha sido que en situaciones de baja contaminación, sin bien los valores oscilan considerablemente, nunca llegan a ser altos. Por algún motivo que no se ha llegado a conocer, son numerosas las ocasiones en las que aun exponiendo los sensores a situaciones de contaminación ambiental más elevadas de lo habitual (entiéndase, en el lugar del estudio), estos no responden en consecuencia. Eso sí, han quedado descartadas la incidencia de la luz y la humedad como posibles causas.

Existe la posibilidad de que el lugar desde el que se han tomado la mayoría de lecturas (ventana en cara suroeste) haya provocado este hecho, y que el edificio haya actuado de paravientos, impidiendo la exposición de los sensores a las partículas provenientes del tráfico de la avenida principal. En este sentido, queda descartada la opción de que los montajes no permitan la correcta circulación de aire por su interior, ya que en ambos casos se han tomado lecturas colocando parte de los sensores en las caras exteriores (una vez comprobado que la exposición a la luz no les afecta), y han ofrecido valores iguales a los del interior.

Otra posible causa es que el sensor no detecte partículas superiores a cierto tamaño 'x' (valor que estaría comprendido entre 2,5 y 10 micrómetros) y que las ocasiones en las que la contaminación PM10 está compuesta mayoritariamente por partículas con diámetros comprendidos entre 'x' y 10 micrómetros el sensor no sea capaz de detectarlas. Quizá el problema resida en que la absorción de aire no sea lo suficientemente potente, y podría solucionarse añadiendo un pequeño ventilador en el agujero de entrada de aire del sensor.

Sin duda, de cara a realizar estudios más exhaustivos, sería necesario recabar más datos, especialmente en ambientes de contaminación más cambiante. En caso de ser posible, lo ideal sería trabajar en colaboración con químicos expertos en la materia, y analizar el comportamiento del sensor en entornos controlados.

Por último, de cara al uso en grandes cantidades de estos sensores para recabar datos de forma colaborativa, si bien se ha concluido con el diseño de una caja de medidas, han quedado descartadas por el momento todas las opciones de implementación, y antes es necesario estudiar más en profundidad el sensor y/o buscar la forma de que sea más útil, no tanto de cara a obtener una precisión muy alta, sino más bien a su uso como indicador de situaciones de contaminación elevadas, en las cuales el riesgo de padecer una enfermedad cardiovascular aumenta.

## Bibliografía

---

- [1] Erik Fustes. *Estudio del uso de Sensores de Bajo Coste para el Análisis de Parámetros Medioambientales*, 2017.
- [2] Jiankai.li, de Seeedstudio. *Grove -Dust Sensor User Manual*, 2015. URL: [http://www.mouser.com/ds/2/744/Seeed\\_101020012-838657.pdf](http://www.mouser.com/ds/2/744/Seeed_101020012-838657.pdf)
- [3] Tracy Allen. *De-construction of the Shinyei PPD42NS dust sensor*, 2013. URL: [http://takingspace.org/wp-content/uploads/ShinyeiPPD42NS\\_Deconstruction\\_TracyAllen.pdf](http://takingspace.org/wp-content/uploads/ShinyeiPPD42NS_Deconstruction_TracyAllen.pdf)
- [4] Chris Nafis. *Air Quality Monitoring*, 2012. URL: <http://www.howmuchsnow.com/arduino/airquality/grovedust/>
- [5] Darell Tan. *Testing the Shinyei PPD42NS*, 2013. URL: <http://irq5.io/2013/07/24/testing-the-shinyei-ppd42ns/>
- [6] Ron. *The Shinyei experiment, Real-time Air Quality readings from Beijing*, 2014. URL: <http://aqicn.org/sensor/shinyei/>
- [7] Elena Austin, Igor Novosselov, Edmun Seto, Michael G.Yost. *Laboratory Evaluation of the Shinyei PPD42NS Low-Cost Particulate Matter Sensor*, 2015. URL: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137789>
- [8] D. M. Holstius, A. Pillarisetti, K. R. Smith (Universidad de California) y E. Seto (Universidad de Washington). *Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California*, 2014. URL: [https://i.publiclab.org/system/images/photos/000/003/464/original/Field\\_Calibrations\\_of\\_a\\_low-cost\\_aerosol\\_sensor\\_at\\_a\\_regulatory\\_monitoring\\_site\\_in\\_California-2014.pdf](https://i.publiclab.org/system/images/photos/000/003/464/original/Field_Calibrations_of_a_low-cost_aerosol_sensor_at_a_regulatory_monitoring_site_in_California-2014.pdf)