



Universidad
Politécnica
de Cartagena



Universidad
Politécnica
de Cartagena

Campus
de Excelencia
Internacional

***Contribución al campo del IoT
mediante el desarrollo de sensores
inteligentes basados en
espectrofotometría de longitud de
onda variable.***

***Aplicación a la monitorización en
continuo de la carga contaminante
en aguas residuales urbanas***

Doctorado en Tecnologías Industriales

Autor:

Daniel Carreres Prieto

Directores:

Dr. Juan Tomás García Bermejo

Dr. Juan Suardíaz Muro

Cartagena 2021



Universidad
Politécnica
de Cartagena | Campus
de Excelencia
Internacional

**Contribución al campo del IoT mediante
el desarrollo de sensores inteligentes
basados en espectrofotometría de longitud
de onda variable.**

**Aplicación a la monitorización en
continuo de la carga contaminante en
aguas residuales urbanas**

Doctorado en Tecnologías Industriales

Autor: D. Daniel Carreres Prieto

Directores: Dr. Juan Tomás García Bermejo

Dr. Juan Suardíaz Muro

2021

**CONFORMIDAD DE SOLICITUD DE AUTORIZACIÓN DE DEPÓSITO DE
TESIS DOCTORAL POR EL/LA DIRECTOR/A DE LA TESIS**

D. Juan Tomás García Bermejo y D. Juan Suardíaz Muro, directores de la Tesis doctoral titulada “*Contribución al campo del IoT mediante el desarrollo de sensores inteligentes basados en espectrofotometría de longitud de onda variable. aplicación a la monitorización en continuo de la carga contaminante en aguas residuales urbanas*”.

INFORMA:

Que la referida Tesis Doctoral, ha sido realizada por D. Daniel Carreres Prieto, dentro del Programa de Doctorado Tecnologías Industriales, dando mi conformidad para que sea presentada ante el Comité de Dirección de la Escuela Internacional de Doctorado para ser autorizado su depósito.

Informe positivo sobre el plan de investigación y documento de actividades del doctorando/a emitido por el Director/Tutor (**RAPI**).

La rama de conocimiento en la que esta tesis ha sido desarrollada es:

- Ciencias
 Ciencias Sociales y Jurídicas
 Ingeniería y Arquitectura

En Cartagena, a 15 de marzo de 2021

LOS DIRECTORES DE LA TESIS

JUAN
TOMAS|
GARCIA|
BERMEJO

Firmado digitalmente por JUAN
TOMAS|GARCIA|BERMEJO
Nombre de reconocimiento (DN):
cn=JUAN TOMAS|GARCIA|BERMEJO,
serialNumber=██████████
givenName=JUAN TOMAS,
sn=GARCIA BERMEJO,
ou=Ciudadanos, o=ACCV, c=ES
Fecha: 2021.03.15 09:55:24 +01'00'

Fdo: Juan Tomás García Bermejo

JUAN|
SUARDIA|
Z|MURO

Firmado
digitalmente por
JUAN|SUARDIAZ|
MURO
Fecha: 2021.03.15
09:18:06 +01'00'

Fdo: Juan Suardíaz Muro

COMITÉ DE DIRECCIÓN ESCUELA INTERNACIONAL DE DOCTORADO

CONFORMIDAD DE DEPÓSITO DE TESIS DOCTORAL
POR LA COMISIÓN ACADÉMICA DEL PROGRAMA

D^a.Sonia Busquier Sáez, Secretaria de la Comisión Académica del Programa de Doctorado en Tecnologías Industriales

INFORMA:

Que la Tesis Doctoral titulada, “*Contribución al campo del IoT mediante el desarrollo de sensores inteligentes basados en espectrofotometría de longitud de onda variable. aplicación a la monitorización en continuo de la carga contaminante en aguas residuales urbanas*”, ha sido realizada, dentro del mencionado Programa de Doctorado, por D. Daniel Carreres Prieto bajo la dirección y supervisión de los Drs Juan Tomás García Bermejo y Juan Suardiáez Muro

En reunión de la Comisión Académica, visto que en la misma se acreditan los indicios de calidad correspondientes y la autorización del Director/a de la misma, se acordó dar la conformidad, con la finalidad de que sea autorizado su depósito por el Comité de Dirección de la Escuela Internacional de Doctorado.

Evaluación positiva del plan de investigación y documento de actividades por el Presidente de la Comisión Académica del programa (**RAPI**).

La Rama de conocimiento por la que esta tesis ha sido desarrollada es:

- Ciencias
- Ciencias Sociales y Jurídicas
- Ingeniería y Arquitectura

En Cartagena, a 15 de marzo de 2021

EL PRESIDENTE DE LA COMISIÓN ACADÉMICA

SONIA|
BUSQUIER|SAEZ

Firmado digitalmente por SONIA|BUSQUIER|
SAEZ
Nombre de reconocimiento (DN): cn=SONIA|
BUSQUIER|SAEZ, serialNumber=
givenName=SONIA, sn=BUSQUIER SAEZ,
ou=Ciudadanos, o=ACCV, c=ES
Fecha: 2021.03.15 10:04:47 +01'00'

Fdo: Sonia Busquier Sáez

COMITÉ DE DIRECCIÓN ESCUELA INTERNACIONAL DE DOCTORADO

Agradecimientos

Son muchas las personas que me han ayudado a lo largo de estos años de tesis, y sin lugar a dudas, sin ellas habría sido posible llevar a cabo este trabajo. En primer lugar, quiero expresar mi profundo agradecimiento a mis directores, Dr. Juan Tomás García Bermejo y Dr. Juan Suardíaz Muro y al que considero mi “tercer director”, el Dr. Fernando Cerdán Cartagena, catedrático del área de ingeniería telemática.

Quisiera agradecer a la Fundación Séneca, el apoyo recibido a para la realización de la tesis doctoral, en el marco de las becas para la Formación del Personal Investigador (FPI) con referencia Mod. B, Ref. 20320/FPI/17, así como a la empresa HIDROGEA y más concretamente a Don Andrés Martínez Gumbau, Director Gerente de HIDROGEA, a Don Francisco Antonio Reyes Bas, Director de Operaciones de HIDROGEA y a Don Fulgencio Díaz Madrid, antiguo Director Gerente, a Don Xavier Ybarra Moreno, antiguo Director Regional y a Inmaculada Serrano Sánchez, actual Directora Regional de HIDROGEA, por la confianza depositada, proporcionándome todos el apoyo necesario para su realización.

Así mismo, quiero extender mi agradecimiento a todo el personal de las estaciones de depuración que durante estos años me han ayudado durante las campañas de muestreo, y que sin su ayuda, nada de esto habría sido posible.

Resta agradecer a mis padres, Pedro y Matilde, por todo su cariño, comprensión y paciencia durante estos años, sin los cuales nada de lo que he logrado habría sido posible.

A todos vosotros, gracias.

Dedicatoria

A mis padres, Pedro y Matilde,

A mis directores, Juan García, Juan Suardíaz

Al que considero mi “tercer director”, Fernando Cerdán,

A la Fundación Séneca y a HIDROGEA por el apoyo recibido.

Resumen

Conocer la evolución de la carga contaminante de la red de saneamiento es clave para un control más fehaciente de la misma, especialmente durante episodios de lluvia intensa debido al riesgo de desbordamiento de la red, lo que puede implicar el vertido de una alta concentración de contaminantes al medio receptor.

Tradicionalmente, el estudio de la evolución de la carga contaminante es llevado a cabo mediante la toma de muestras puntuales que son analizadas en laboratorio, lo que imposibilita llevar a cabo una monitorización en tiempo real. No obstante, en las últimas décadas se han llevado a cabo diversos estudios que ponen de manifiesto la capacidad de caracterizar ciertos parámetros contaminantes a partir de mediciones indirectas, basadas en la espectroscopia de absorción molecular, como es el caso de un espectrofotómetro para un rango amplio de longitudes de onda o de un medidor de turbidez para una única longitud de onda. Actualmente ya se están comercializando equipos que por medio de un análisis espectrofotométrico basado en lámparas de xenón, estiman ciertos parámetros de contaminación como la DQO o la concentración de nutrientes sin necesidad de reactivos o pretratamientos.

A pesar del avance que este tipo de sistemas representan, el uso de lámparas halógenas o de xenón para la generación del espectro de trabajo puede representar también una limitación a la hora de desarrollar equipos de bajo coste, dimensiones y consumo que puedan operar de forma autónoma en la red.

La presente tesis doctoral, realizada en colaboración con la empresa HIDROGEA gracias a una beca para la Formación del Personal Investigador (FPI) de la Fundación Séneca, se centra en el diseño e implementación de una nueva generación de equipos inteligentes basados en espectrofotometría LED, para el análisis en tiempo real de una amplia variedad de parámetros contaminantes presentes en las aguas residuales que recorren las redes de saneamiento, mediante diversos modelos de estimación.

Esto se ha concretado en el desarrollo de dos equipos de bajo coste, uno para su uso en laboratorio, el cual se ha denominado Espectrofotómetro Desktop, y otro para su uso en la red de saneamiento, como un equipo autónomo capaz de llevar a cabo la adquisición, caracterización, almacenamiento y envío de la información de las muestras de agua en un corto periodo de tiempo a la nube. Además, se han calculado diversos modelos matemáticos para la estimación de la DQO, DBO_5 , SST, P, NT, $\text{NO}_3^- \text{N}$ a partir de la respuesta espectral de las muestras, con una alta precisión y sin necesidad de químicos o pretratamientos.

Se trata de una tesis con un enfoque multidisciplinar, desarrollada bajo la supervisión de dos grupos de investigación de la Universidad Politécnica de Cartagena (UPCT). Por una parte, la División de Innovación en Sistemas Telemáticos y Tecnología Electrónica (DINTEL) y por otra, el grupo de investigación en Ingeniería Hidráulica Marítima y Medio Ambiental (Hidr@m) de la UPCT, que desarrolla su trabajo en la investigación aplicada en sistemas de abastecimiento y saneamiento y, en especial, en el estudio del transporte de contaminantes así como la monitorización y el estudio de las medidas de remediación en redes de saneamiento.

Abstract

Knowing the evolution of the pollutant load of the sewer network is key for a more reliable control, especially during episodes of heavy rainfall due to the risk of overflowing, which could involve the discharge of a high concentration of pollutants into the receiving water bodies.

Traditionally, this study is carried out by taking point samples that are analysed in the laboratory, which makes it impossible to carry out real-time monitoring. However, in recent decades, studies have begun to emerge that show that it is possible to measure certain pollutant parameters using indirect measurements based on molecular absorption spectroscopy, such as a spectrophotometer for a wide range of wavelengths or a turbidity meter for a single wavelength. Currently, equipment is already on the market that, by means of a spectrophotometric analysis based on xenon lamps, estimate certain parameters such as COD or nutrient concentration without the need for reagents or pre-treatment.

Despite the progress that this type of system represents, the use of halogen or xenon lamps for the generation of the working spectrum represents a limitation when it comes to developing low-cost, low-dimension and low-consumption equipment that can operate autonomously in the network.

This doctoral thesis, carried out in collaboration with the company HIDROGEA thanks to a grant for the Training of Research Personnel (FPI) from the Seneca Foundation, focuses on the design and implementation of a new generation of intelligent equipment based on LED spectrophotometry, for the real-time analysis of a wide variety of pollutant parameters present in the wastewater along the sewer networks, by means of various estimation models.

This has resulted in the development of two low-cost devices, one for laboratory use, called Desktop Spectrophotometer, and the other for use in the sewerage network, as a stand-alone device capable of acquiring, characterising, storing and sending the information from the water samples in a short period of time to the cloud. In addition, several mathematical models have been calculated for the estimation of COD, BOD₅, TSS, P, NT, NO₃⁻N from the spectral response of the samples, with high accuracy and without the need for chemicals or pre-treatment.

This is a thesis with a multidisciplinary approach, developed under the supervision of two research groups of the Polytechnic University of Cartagena (UPCT). On the one hand, the Division of Innovation in Telematic Systems and Electronic Technology (DINTEL) and on the other hand, the research group in Hydraulic, Maritime and Environmental Engineering (Hidr@m) of the UPCT, which develops its work in applied research in supply and sanitation systems and, in particular, in the study of the transport of pollutants as well as the monitoring and study of remediation measures in sanitation networks.

Índice general

Capítulo 1 Estado del arte	1
1.1 Introducción	1
1.2 Normativa	4
1.3 Estimación de la carga contaminante de forma indirecta: Técnicas y estudios	5
1.3.1 Turbidímetro y Espectrofotómetro	6
1.3.2 Regresión Lineal Multivariable.....	8
1.3.2.1 Determinación de SST en la Cuenca Ecully, Francia por Bertrand-Krajewski a partir de la turbidez	10
1.3.2.2 Determinación de SST, DQO y NT en la Cuenca de la ciudad de Pau por Tomas Bersinger a partir de la turbidez y la conductividad	11
1.3.2.3 Determinación de DQO, DBO ₅ y SST a partir de turbidez por Del Río, H.....	13
1.3.2.4 Determinación de DQO, DBO, NT a partir de espectrofotometría de fluorescencia por ultravioleta	14
1.3.3 Regresión por mínimos cuadrados parciales (PLSR)	16
1.3.3.1 Determinación de DQO y DBO ₅ a partir de NIR utilizando modelos PLSR	17
1.3.4 Regresión no lineal multivariada basada en técnicas de machine learning	18
1.3.4.1 Determinación de SST a partir de UV-Vis mediante Boosting-IPW-PLS.....	22
1.3.4.2 Determinación del nitrógeno total mediante redes neuronales y SVM.....	23
1.3.4.2 Determinación de DQO, DBO ₅ y SST a partir de las características de la red de captación urbana mediante técnicas de inteligencia artificial.....	24
1.4 Equipos para la caracterización de la carga contaminante.....	26
1.4.1 Estación para agua residual S::can.....	26
1.4.2 Sistema de espectrómetro UV-Vis Roice Water™.....	27
1.4.3 Analizador de aguas residuales VWA-3000UV de la empresa Vasthi Engineers Private Limited™	28
1.4.4 Sistema ISA con sonda espectrómetro UV-Vis de GO-Systemelektronik™.....	29
1.5 Líneas de actuación de la tesis doctoral	30
Capítulo 2 Espectrofotómetro Desktop	33

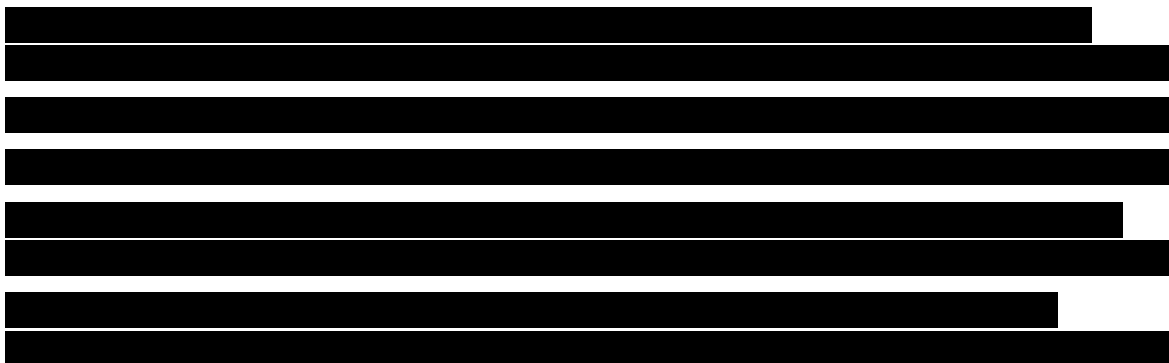


5.2.3	Extensión del rango de trabajo	125
5.2.4	Generación de ecuación genérica de calibración	130
5.3	Resultados del proceso de calibración.....	133
5.4	Conclusiones	136
Capítulo 6 Calibración de LED RGB		137
6.1	Introducción.....	137
6.2	Resultados preliminares.....	138
6.3	Extensión del rango de trabajo	148
6.4	Resultados del proceso de calibración.....	156
6.5	Conclusiones	160
Capítulo 7 Modelos de estimación de la carga contaminante.....		163
7.1	Introducción.....	163
7.2	Materiales y métodos	164
7.2.1	Campaña experimental.....	164
7.2.2	Características de los datos.....	166
7.2.2.1	Agua bruta.....	167
7.2.2.2	Agua tratada.....	169
7.3	Modelos estadísticos de estimación de la carga contaminante	171
7.3.2	Modelos de regresión lineal.....	171
7.3.2.1	Demanda Química de Oxígeno (DQO)	175
7.3.2.2	Demanda biológica de oxígeno a 5 días (DBO ₅).....	177
7.3.2.3	Sólidos suspendidos totales (SST)	178
7.3.3	Modelos basados en algoritmo genético.....	180
7.3.3.1	Demanda Química de Oxígeno (DQO)	181
7.3.3.2	Demanda biológica de oxígeno a 5 días (DBO ₅).....	184
7.3.3.3	Sólidos suspendidos totales (SST)	186
7.3.3.4	Fósforo (P).....	188
7.3.3.5	Nitrógeno total (NT).....	189
7.3.3.6	Nitrato Nitrógeno (NO ₃ -N)	192
7.4	Comparación de modelos	194
7.5	Conclusiones	197
Capítulo 8 Conclusiones, aportaciones y desarrollos futuros.....		199
8.1	Conclusiones y aportaciones.....	199
8.2	Desarrollos futuros.....	200
Anexo A.....		201
Anexo B.....		211

Bibliografía 251

Índice de Figuras

Figura 1.1. Esquema básico de un turbidímetro y espectrofotómetro.....	7
Figura 1.2. Ejemplo de diagrama de dispersión entre pares de variables donde existen variables correladas.....	9
Figura 1.3. Relación SST-Turbidez para la cuenca Ecully	11
Figura 1.4. Curvas de correlación entre turbidez y (A) DQO, (B) SST. (C).Curva de correlación entre conductividad y NT.....	12
Figura 1.5. Correlación entre la turbidez y (A) SST, (B) DQO/DBO5 para la cuenca Ensanche	14
Figura 1.6. Componentes detectados mediante EEM y PARAFAC	15
Figura 1.7. Diagramas de correlación para la estimación de (A) DQO, (B) DBO, (C) NT.....	16
Figura 1.8. Diagrama de dispersión entre NIR y (A) DQO (B).....	18
Figura 1.9. Ejemplos de hiperplanos en Máquinas de vector de soportes	19
Figura 1.10. Conjuntos de datos que no pueden ser separados de forma perfecta por un hiperplano.....	20
Figura 1.11. Ejemplo de red neuronal convolucional.....	21
Figura 1.12. Esquema de árbol evolutivo en algoritmos genéticos.....	22
Figura 1.13. Estimación por medio de la técnica Boosting-IPW-PLS de (A) DQO, (B) SST y (C) Aceites y grasas.....	23
Figura 1.14. Comparativa entre los valores de nitrógeno total medidos con respecto a los estimados	24
Figura 1.15. Comparativa de resultados obtenidos mediante máquinas de vector de soporte y programación genética.....	25
Figura 1.16. Estación para agua residual S::can™	27
Figura 1.17. Sistema de espectrómetro UV-Vis Roice Water™	28
Figura 1.18. Analizador de aguas residuales VWA-3000UV, de la empresa Vasthi Engineers Private Limited ™	29
Figura 1.19. Sistema ISA con sonda espectrómetro UV-Vis de GO-Systemelektronik™	30



[Redacted Table-Of-Contents]

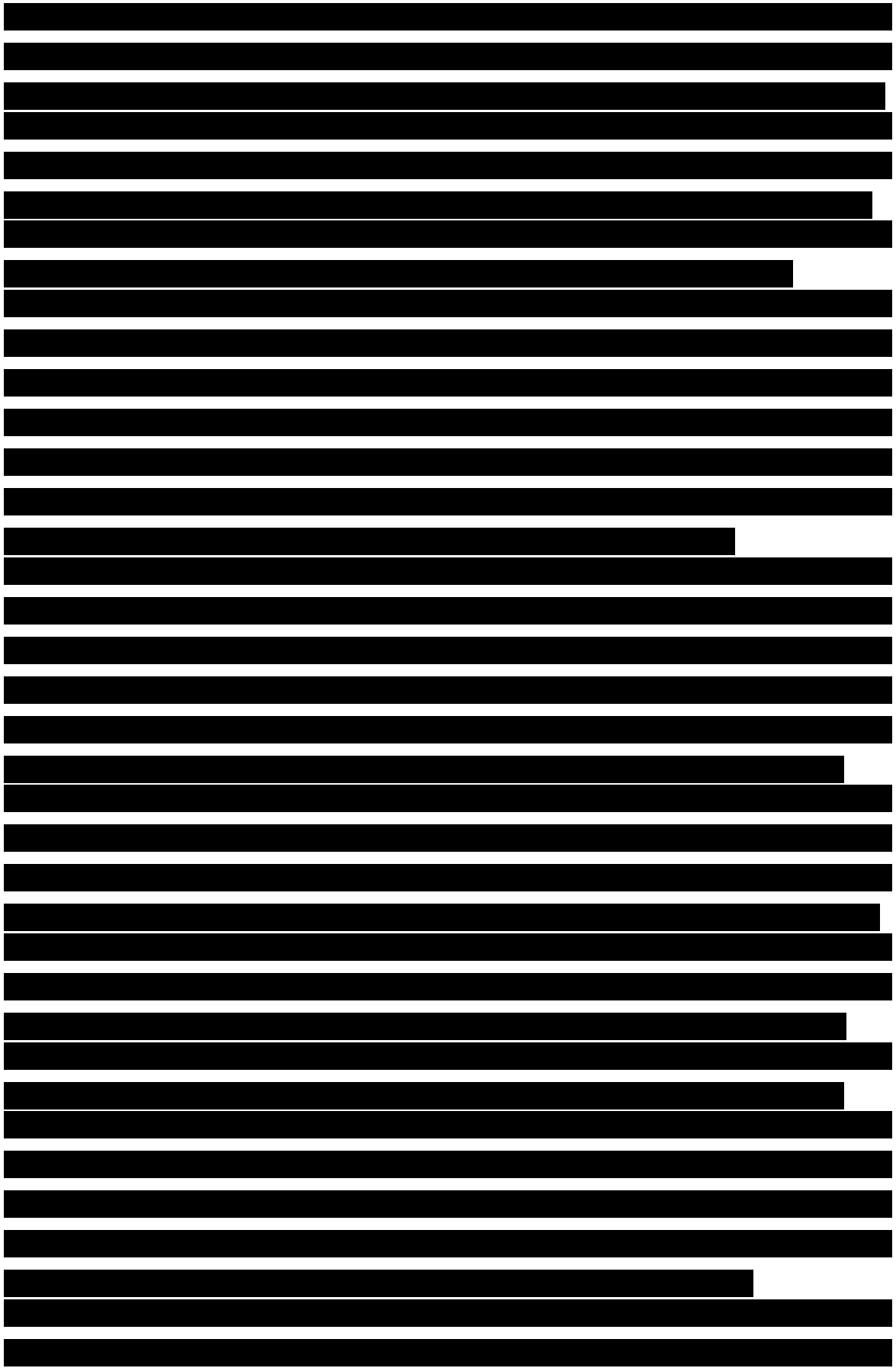


Figura 4.1. Panel de control servidor web..... 97

Figura 4.2. Tabla “users” de la base de datos con muestra de credenciales.....	99
Figura 4.3. Interfaz para la creación de nuevos usuarios	100
Figura 4.4. Respuesta del servidor tras el inicio de sesión de un usuario mediante Postman ..	101
Figura 4.5. Lista de muestras analizadas por los equipos desarrollados	102
Figura 4.6. Respuesta servicio web para la recuperación de la lista de muestras almacenadas	104
Figura 4.7. Aplicación para la subida de datos analíticos medidos en laboratorio	104
Figura 4.8. Aplicación de control de los equipos desarrollados.....	105
Figura 4.9. Vista en detalle de una muestra contenida en el servidor	106
Figura 4.10. Servicio web para la recuperación de los datos de una muestra en Postman	108
Figura 4.11. Estimación de parámetros contaminantes de una muestra de agua bruta almacenada en el servidor, acompañada de una comparativa entre modelos y los valores medidos en laboratorio	109
Figura 4.12. Lista de equipos asociados al usuario sobre los que se puede consultar la evolución de la carga contaminante en el tiempo	111
Figura 4.13. Interfaz para la edición de equipos.....	111
Figura 4.14. Evolución de la carga contaminante por equipo.....	112
Figura 4.15. Lista de alertas detectadas.....	113
Figura 4.16. Partes de la aplicación para PC para el control de los equipos	114
Figura 4.17. Pestaña “In-Sewer” de la aplicación para PC.....	116
Figura 5.1. Espectro de emisión del LED VAOL-5EUV8T4 extraído del datasheet del fabricante	118
Figura 5.2. Comparación de los niveles de transmitancia a 385nm medidos con un equipo comercial basado en lámparas incandescentes y el LED VAOL-5EUV8T4 con dos brillos diferentes	119
Figura 5.3. Vista del software desarrollado para la gestión de los equipos descritos en los Capítulos 2 y 3, durante el proceso de obtención de los valores de transmitancia para distintos niveles de brillo.....	121
Figura 5.4. Efecto del nivel de brillo en los valores de transmitancia registrados en la muestra M1 por medio de un LED de 385nm de longitud de onda pico.....	122
Figura 5.5. Comparación de valores de transmitancia entre 385 y 700nm, medidos con 59 LEDs de diferentes longitudes de onda pico (rojo) y con un equipo comercial basado en lámparas incandescentes (azul)	123
Figura 5.6. Diagrama de dispersión entre los valores de transmitancia medidos con un equipo basado en lámparas incandescentes y los obtenidos por un LED de 470nm de pico y a un nivel de brillo 40 sobre 19 muestras	124
Figura 5.7. Comparación entre los valores de transmitancia a 470nm medidos con un equipo comercial y los obtenidos con un LED de 470nm de pico tras aplicar el modelo de calibración sobre 19 muestras.....	124
Figura 5.8. Representación del rango de trabajo útil de un LED de longitud de onda pico de 470nm	125

Figura 5.9. Comparación de valores de transmitancia obtenida a 450nm con el equipo comercial y un LED de longitud de onda pico de 470nm 126

Figura 5.10. Diagrama de dispersión entre el valor de referencia de transmitancia a 450nm con el LED de 470nm de pico y un cierto nivel de brillo (T40)..... 127

Figura 5.11. Comparación de los valores de transmitancia obtenida a 500nm con el equipo comercial y el prototipo desarrollado a partir de un LED con una longitud de onda de pico de 470nm y 500nm 127

Figura 5.12. Rectas de calibración entre 380 y 700nm..... 130

Figura 5.13. Valores de transmitancia calculados en el punto de inflexión de las líneas de calibración de la Figura 5.12. Transmitancia LED 50% 131

Figura 5.14. Espectro de color de la luz visible 131

Figura 5.15. Cálculo de las líneas de tendencia de la pendiente m y de la ordenada en el origen n en función de λ 132

Figura 5.16. Comparativa de los valores de transmitancia antes (verde), después (rojo) con respecto al equipo comercial (azul) para las siguientes muestras: (A) M2, (B) M3, (C) M5, (D) M8, (E) M9, (F) M10, (G) M13 y (H) M19..... 134

Figura 5.17. Respuesta espectral de muestra de agua residual 135

Figura 6.1. Espectro de emisión del LED RGB HV-5RGB25..... 139

Figura 6.2. Esquema para la calibración de LED RGB 140

Figura 6.3. Comparativa entre los valores de transmitancia medidos con el LED verde del RGB lámpara incandescente a 525nm 141

Figura 6.4. Comparativa entre los valores de transmitancia medidos con el LED rojo y la lámpara incandescente a 625nm 141

Figura 6.5. Comparativa entre los valores de transmitancia medidos con el LED azul y la lámpara incandescente a 460nm 142

Figura 6.6. Comparativa del ajuste para el espectro de emisión del LED rojo: (A) 600nm, (B) 610nm, (C) 625nm, (D) 630nm, y (E) 645nm 143

Figura 6.7. Comparativa del ajuste para el espectro de emisión del LED verde: (A) 500nm, (B) 510nm, (C) 525nm, (D) 530nm, y (E) 550nm 144

Figura 6.8. Comparativa del ajuste para el espectro de emisión del LED azul: (A) 435nm, (B) 445nm, (C) 460nm, (D) 470nm, y (E) 485nm 145

Figura 6.9. Ajuste del LED azul, con longitud de onda máxima es de 460nm 146

Figura 6.10. Diagrama de bandas del LED azul. Elaboración propia..... 146

Figura 6.11. Espectro de emisión útil del LED RGB HV-5RGB25 147

Figura 6.12. Espectro visible comparativo 148

Figura 6.13. Valores heurísticos RGB para las longitudes de onda visibles según la representación del color..... 152

Figura 6.14. Combinación de longitudes de onda y correlación para: (A) 560, (B) 570, (C) 580, (D) 590, y (E) 600nm 153

Figura 6.15. Comparativa entre el valor de transmitancia registrado con el LED RGB y el obtenido con la lámpara incandescente, a 580nm para 20 muestras analizadas	154
Figura 6.16. Combinación de longitudes de onda y correlación para: (A) 670, (B) 505, y (C) 460nm	155
Figura 6.17. Comparativa entre los valores de transmitancia medidos con un equipo comercial basado en la lámpara incandescente (azul) y el dispositivo LED RGB desarrollado (rojo) después del proceso de calibración llevado a cabo dentro de 510-645nm, para las siguientes muestras: (A) S7, (B) S12, (C) S15, (D) S16, y (E) S20	156
Figura 6.18. Comparativa entre los valores de transmitancia medidos con un equipo comercial basado en la lámpara incandescente (azul) y el dispositivo LED RGB desarrollado (rojo) después del proceso de calibración en el rango de 510-645nm, para las siguientes muestras: (A) S21, (B) S22, (C) S23, (D) S43, (E) S44, (F) S45, (G) S46, (H) S47, y (I) S48	157
Figura 7.1. Ejemplo de caracterización de muestras de agua a la entrada de la planta de tratamiento proporcionada por el laboratorio (Fichero Excel).....	165
Figura 7.2. Fichero CSV con la información espectrofotométrica medida con el Espectrofotómetro Desktop y los parámetros contaminantes medidos en el laboratorio de la planta de tratamiento	166
Figura 7.3. Datos espectrofotométricos y parámetros de contaminantes para muestras agua bruta.....	167
Figura 7.4. Datos espectrofotométricos y parámetros de contaminantes para las muestras tratadas	170
Figura 7.5. Esquema de cálculo de los modelos de regresión lineal	172
Figura 7.6. Histograma con combinación de muestras de agua bruta y tratada. (A) DQO, (B) .	174
Figura 7.7. Comparación entre los valores de DQO medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos por el modelo de regresión lineal multivariante de la Ecuación (7.1), para el agua bruta	176
Figura 7.8. Gráfica de dispersión de las variables utilizadas en el modelo MLR para calcular la DQO con respecto a: (A) Transmitancia a 380nm, (B) Transmitancia a 521nm, (C) Transmitancia a 555nm y (D) Transmitancia a 580nm	176
Figure 7.9. Comparación entre los valores de	177
Figura 7.10. Diagrama de dispersión de las variables utilizadas en el modelo de la ecuación (7.2) para calcular la	178
Figura 7.11. Comparación entre los valores de SST medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos por el modelo de regresión lineal, para las aguas residuales (agua bruta)	179
Figura 7.12. Diagrama de dispersión de las variables utilizadas en el modelo de la Ecuación (7.3) para calcular los SST con respecto a (A) Transmitancia a 380nm, (B) Absorbancia a 656nm....	179
Figura 7.13. Curva de aprendizaje para el cálculo de DQO mediante algoritmo genético.....	182
Figura 7.14. Comparación entre los valores de DQO medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.4).....	183
Figura 7.15. Comparación entre los valores de	185

Figura 7.16. Comparación entre los valores de SST medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.6).....	187
Figura 7.17. Comparación entre los valores de P medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.7).....	189
Figura 7.18. Comparación entre los valores de NT medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.8).....	191
Figura 7.19. Comparación entre los valores de NO₃-N medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.9).....	193

Índice de tablas

Tabla 1.1. Datos utilizados para la generación de modelos de Máquinas de vectores de soporte y programación genética en estudio.	25
Tabla 5.1. Características LED VAOL-5EUV8T4 117	
Tabla 5.2. Modelos de calibración calculados para longitudes de onda comprendidas entre 380-700nm	129
Tabla 5.3. RMSD e Índice de Error	136
Tabla 6.1. Características de las aguas residuales mostradas en la Figura 6.18.....	159
Tabla 6.2. Error cuadrático medio (RMSD) e índice de error entre 510-645nm.....	160
Tabla 7.1. Variables analizadas durante la campaña de muestreo	165
Tabla 7.2. Pruebas de normalidad relativas al agua bruta	175
Tabla 7.3. Índices de impacto de las variables utilizadas para el modelo de DQO basado en algoritmo genético	182
Tabla 7.4. Ejemplo de estimación DQO de 15 muestras aleatorias	183
Tabla 7.5. Índices de impacto de las variables utilizadas para el modelo de.....	184
Tabla 7.6. Ejemplo de estimación.....	185
Tabla 7.7. Índices de impacto de las variables utilizadas para el modelo de SST basado en algoritmo genético	186
Tabla 7.8. Ejemplo de estimación SST de 15 muestras aleatorias.....	187
Tabla 7.9. Índices de impacto de las variables utilizadas para el modelo del P basado en algoritmo genético	188
Tabla 7.10. Ejemplo de estimación P de 15 muestras aleatorias	189
Tabla 7.11. Índices de impacto de las variables utilizadas para el modelo del NT basado en algoritmo genético	190
Tabla 7.12. Ejemplo de estimación NT de 15 muestras aleatorias.....	191
Tabla 7.13. Índices de impacto de las variables utilizadas para el modelo del NO3-N basado en algoritmo genético	192
Tabla 7.12. Ejemplo de estimación NO3-N de 15 muestras aleatorias.....	193
Tabla 7.13. Error cuadrático medio e índice de error (Er) de los diferentes modelos calculados	194
Tabla 7.14. Resumen de la relevancia de las longitudes de onda en los diferentes modelos de estimación calculados.....	196

Capítulo 1

Estado del arte

1.1 Introducción

Los sistemas de saneamiento urbanos están conformados por redes de alcantarillado y por estaciones depuradoras de aguas residuales (EDAR). Una gestión integrada de dichos sistemas es necesaria para poder hacer frente a diversos problemas como los asociados a los episodios de lluvia, dado que éstos, en un periodo corto de tiempo, tienden a producir la sobrecarga de la red, provocando la descarga en el medio receptor de considerables volúmenes con una alta concentración de contaminantes, con el grave daño medioambiental que esto supone.

Las aguas de escorrentía generadas durante los periodos de lluvia, producen también un gran impacto en la eficacia de los procesos de depuración en las plantas de tratamiento [1-4], dado que la entrada de un mayor caudal y/o mayor carga contaminante, pueden hacer que las plantas no logren los niveles de depuración establecidos por normativa, haciendo esa aguas no válidas para otras aplicaciones como podría ser el regadío.

Los sistemas de saneamiento urbano deben cumplir las políticas de la UE para detener la degradación de las aguas y del medio ambiente: Directiva Marco del Agua (DMA, 2000/60/CE), Directiva de Normas de Calidad Ambiental (NCA, 2008/105/CE), Directiva 91/271/CEE o Directiva de Aguas Residuales Urbanas (DAR), Directiva 2006/7/CE o Directiva de Aguas de Baño de la UE (Directiva 2006/7/CE) y Directiva Marco de Estrategia Marina (Directiva 2008/56/CE).

Disponer de información de carácter cuantitativo, de forma continua y en tiempo real sobre la evolución de los contaminantes en la red de alcantarillado (sobre todo durante episodios de lluvia, donde estos niveles son mayores), resulta clave para una mejor optimización de las plantas de tratamiento y de dichas redes, lo cual puede contribuir a un mejor cumplimiento de las políticas medioambientales de la UE y a alcanzar los objetivos del Pacto Verde Europeo [5].

El desarrollo de sensores para la monitorización continua de la carga contaminante de las aguas residuales en la red de alcantarillado es un gran reto desde el punto de vista técnico, debido a la variabilidad de parámetros a caracterizar, a los diferentes procedimientos para llevarlos a cabo, y las condiciones de ambiente extremo existentes en el interior de dichas redes [6,7].

Durante más de una década se han utilizado técnicas ópticas, como la espectroscopia Ultravioleta-Visible (UV-Vis, 100-700nm) y la espectroscopia de infrarrojo cercano (NIR), es decir entre 750 y 2500nm, para caracterizar de forma indirecta y fiable a la vez parámetros contaminantes en las aguas residuales como los sólidos en suspensión, la materia orgánica y los nitratos en las aguas residuales [8-25]. La UV-Vis se refiere a la interacción entre las muestras y la radiación en el rango de longitudes de onda de 200-780nm en una o varias longitudes, con el fin de estimar una serie de parámetros [19].

El éxito de este tipo de análisis indirecto radica en la rapidez en la caracterización, el no ser necesario alterar las muestras (eliminando el uso de químicos), y por ser respetuosa con el medio ambiente.

Los análisis que permiten relacionar la carga contaminante del agua residual con los valores obtenidos mediante las caracterizaciones ópticas se fundamentan en el uso de técnicas estadísticas como la regresión de mínimos cuadrados parciales (PLS), para generar un modelo de regresión basado en los datos espectrales para estimar los parámetros de calidad del agua [8, 26-28]. Varios estudios han demostrado una buena concordancia entre los valores estimados con este tipo de técnicas (espectrofotometría), con respecto a aquellos medidos en laboratorio, como es el caso de la estimación de la Demanda Química de Oxígeno (DQO) [8, 11, 16-18, 20, 22-25], o los sólidos en suspensión totales (SST) mediante UV-Vis y NIR [11,17,18,22-24].

Así mismo, el nitrógeno en forma de nitrato ($\text{NO}_3\text{-N}$) demostró un error entorno al 25% en las estimaciones realizadas con longitudes de onda pertenecientes al ultravioleta y al visible (UV-Vis) [24]. Otros por su parte, ha utilizado la segunda derivada del espectrograma de absorción en UV-Vis para la caracterización del $\text{NO}_3\text{-N}$ [10]. En la actualidad hay equipos de espectrofotometría UV que funcionan en continuo, para la determinación de por ejemplo, la concentración de nitratos y nitritos presentes el agua [14, 22].

En ciudades como Linz [21], Graz, Ecully y Viena [23], se han obtenido resultados prometedores sobre la estimación de ciertos parámetros a partir de mediciones indirectas, lo que posibilita en el largo plazo llevar a cabo una monitorización en continuo de la red de alcantarillado.

El uso de técnicas estadísticas, se han convertido en un elemento clave en el desarrollo de este tipo de equipos. La regresión lineal (LR) y otras más avanzadas como el machine learning, como las máquinas de vectores de soporte (SVM), el uso de algoritmos evolutivos (EVO) y las redes neuronales artificiales (ANN), han permitido la detección de patrones dentro del espectro de absorción para estimar parámetros como la demanda bioquímica de oxígeno a los cinco días (DBO_5) o la DQO [11-13, 17, 19, 25].

El despliegue de sensores basados en espectroscopia a lo largo de toda la red de saneamiento para monitorizar la carga contaminante en tiempo real, precisa de un gran

número de equipos, es decir, de su implantación masiva en la red, lo cual lleva aparejado un coste elevado de despliegue.

En la bibliografía se puede encontrar sistemas basados en espectrofotometría válidos para llevar a cabo un análisis puntual de determinados compuestos existentes en las aguas residuales que alcanzan las plantas de tratamiento [29-40].

A partir de las curvas espectrales de absorbancia medidas en un amplio rango de longitudes de onda, en combinación con la caracterización de las aguas llevadas a cabo en los laboratorios de las plantas de tratamiento, es posible generar modelos de correlación entre ambos [41].

Por otro lado, el uso de tanques para el almacenamiento y sedimentación de las aguas pluviales, no suele representar una opción económica, por lo que el uso de sistemas de monitorización, puede suponer un complemento al almacenamiento y laminación alertando y proporcionando información de la calidad del agua, lo cual puede ofrecer probablemente la solución más respetuosa con el medio ambiente [16].

A pesar de su importancia, el número de estudios en relación a la caracterización de la carga contaminante en continuo en el interior de las redes de saneamiento sigue siendo limitado debido a ciertos inconvenientes, como la variabilidad de la composición de la muestra, la concentración de gases corrosivos como el ácido sulfhídrico, entre otros, lo que dificulta el encontrar correlaciones entre la carga contaminante y la respuesta espectral de las muestras [19].

Enmarcado en este objetivo, la presente tesis doctoral se centra en el desarrollo de sistemas que permitan llevar a cabo una caracterización de las aguas en un periodo corto de tiempo a partir del análisis espectrofotométrico de las mismas. Para ello, se han implementado dos equipos que hacen uso de tecnología LED para la generación del espectro de trabajo, con una reducción significativa de tamaño, coste y dimensiones de los dispositivos, facilitando su implantación masiva en la red.

A lo largo del presente trabajo de investigación, se han desarrollados dos tipos de equipo:

- **Espectrofotómetro Desktop:** Diseñado como un equipo de sobremesa para su uso en laboratorio.
- **Espectrofotómetro In-Sewer:** Equipo para su implantación en la red de saneamiento, capaz de captar muestras del agua residual circulante y caracterizar de forma indirecta y enviar los resultados de las muestras directamente a la nube, sin necesidad de la intervención del usuario.

Así mismo, se han desarrollado modelos estadísticos, basados en regresión lineal y algoritmos genéticos, para estimar la carga contaminante de las muestras a partir de la respuesta espectral de las mismas, sin necesidad de reactivos químicos o de someter a las aguas a pretratamiento [41].

Con el fin de mostrar cómo han evolucionado los sistemas para la monitorización de la calidad de las aguas a partir de mediciones indirectas, en el presente capítulo se describen diversos estudios llevados a cabo para encontrar modelos de estimación a partir de la respuesta espectrofotométrica de muestras de agua residuales, así como los principales

equipos existentes en el mercado basados en la espectrofotometría para la monitorización de la calidad de las aguas.

1.2 Normativa

La lucha con la contaminación de las aguas ha sido un principio rector de la política medioambiental de la Unión Europea y de sus estados miembros. El objetivo 6 de la Agenda 2030 Agua Limpia y Saneamiento [42], establece que *“de aquí a 2030, mejorar la calidad del agua reduciendo la contaminación, eliminando los vertidos y minimizando la emisión de productos químicos y materiales peligrosos, como contaminantes emergentes y prioritarios reduciendo a la mitad el porcentaje de aguas residuales sin tratar y aumentando considerablemente el reciclado y la reutilización sin riesgos a nivel mundial.”*

A lo largo de los años, este compromiso se ha ido evidenciando en los diferentes marcos normativos, pudiendo destacar los siguientes en materia de protección de las aguas:

1. Real Decreto Legislativo 1/2001, de 20 de julio, texto refundido Ley de Aguas, de carácter territorial.
2. Real Decreto 927/1988, de 29 de julio, que aprueba el Reglamento de la Administración Pública del agua y de la Planificación Hidrológica amplia los títulos preliminares II y III del Real Decreto Legislativo 1/2001, Ley de Aguas.
3. Real Decreto 849/1986, de 11 de abril, que se aprueba el Reglamento del Dominio Público Hidráulico, desarrolla artículos preliminares -del I al VI de la Ley 29/1985, de 2 de agosto, de aguas.
4. Real Decreto 817/2015, de 11 de septiembre, establece criterios de seguimiento y evaluación del estado de las aguas superficiales y normas de calidad ambiental.
5. Real Decreto Ley 11/1995, de 28 de diciembre, que establece normas de aplicación al tratamiento de las aguas residuales urbanas.
6. Real Decreto 509/1996, de 15 de marzo, que desarrolla Real Decreto Ley 11/1995.
7. Real Decreto 1620/2007, de 7 de diciembre, establece el régimen jurídico de reutilización de las aguas depuradas.
8. Real Decreto 1290/2012, de 7 de septiembre, por el que se modifica el Reglamento del Dominio Público Hidráulico, aprobado por Reales Decretos 849/1986, 509/1996 y desarrollo Real Decreto Ley 11/1995, por el que se establecen las normas aplicables al tratamiento de las aguas residuales urbanas.

El Real Decreto 1290/2012, surge debido al hecho de que no existe una norma específica que regule los desbordamientos de los sistemas de saneamiento en episodios de lluvia, siendo necesario la incorporación al Reglamento de Dominio Público Hidráulico artículos que permitan limitar la contaminación producida por dichos desbordamientos.

Su artículo 246, Iniciación del procedimiento de autorización de vertidos, recoge en su sección 2.e, que la declaración de vertidos debe incluir *“características de la red de saneamiento y los sistemas de aliviaderos, y las medidas, actuaciones e instalaciones previstas para limitar la contaminación por desbordamiento en episodios de lluvias”*.

Este de concepto de “medidas” es profundizado en su sección 3.c (art 246.3.c), donde alega que los estudios técnicos deben contener, además de las características de la red y el régimen de lluvias, debe incluir las “...*características de las sustancias presentes en los desbordamientos de los sistemas de saneamiento en episodios de lluvia*”.

Así mismo, los titulares de autorización de vertidos deben informar anualmente de los desbordamientos producidos así como de los sistemas utilizados para cuantificar los mismos. Por otro lado, las autorizaciones de vertidos establecen las condiciones en que estos deben realizarse, fijando medidas y actuaciones que deban ser llevadas a cabo para limitar su desbordamiento, en especial en cuanto a la contaminación emitida al medio receptor.

Este artículo ya pone de manifiesto la importancia de un conocimiento temprano de qué sucede con la contaminación en la red de saneamiento para lograr una respuesta mejor coordinadas en la limitación de la contaminación al medio durante episodios de desbordamiento.

Orden AAA / 2056/2014 de 27 de octubre, por la que se aprueban los modelos oficiales de solicitud de autorización y de declaración de vertido, que trata de forma parcial los temas contenidos en el Real Decreto 1290/2012, vuelve a incidir en la importancia de llevar a cabo un **control de contaminantes en continuo de, pH, conductividad y turbidez, combinado con un control periódico de otros contaminantes**.

De esta forma, queda patente la importancia de una monitorización de la calidad de las aguas para lograr una mejor gestión medioambiental.

1.3 Estimación de la carga contaminante de forma indirecta: Técnicas y estudios

Técnicas como la espectrofotometría o la turbidez, son técnicas simples que permiten obtener información relevante sobre las muestras objeto de análisis sin necesidad de llevar a cabo alteraciones en las mismas. En los últimos años, numerosos estudios han sido llevados a cabo para la obtención de modelos que permitan estimar la carga contaminante de las aguas a partir de la respuesta espectral de las mismas.

En este epígrafe se describirán los principales trabajos en esta materia. Con el fin de facilitar la exposición de los resultados y los tipos de análisis llevados a cabo, los diferentes estudios recogidos en las bibliografía, se clasificarán entorno a dos de las técnicas más utilizadas para la generación de estos modelos:

- Regresión lineal multivariable [43, 44]
- Regresión por mínimos cuadrados parciales (PLSR, *Partial Least Square Regression*) [45, 46]
- Regresión no lineal multivariada basada en técnicas de machine learning, como las máquinas de vectores de soporte (SVM) [47-50], redes neuronales artificiales (ANN) [51], o el uso de algoritmos evolutivos (EVO) [52].

Antes de abordar cada una de las técnicas antes mencionadas, se hace necesario describir los aspectos claves de aquellos equipos sobre el que los estudios que se describirán a lo largo del presente capítulo se basan, es decir, los turbidímetros y espectrofotómetros.

1.3.1 Turbidímetro y Espectrofotómetro

Los espectrofotómetros y los turbidímetros son equipos que permiten obtener información relevante sobre las muestras líquidas sin necesidad de tener que someterlas a alteraciones. Aunque ambos equipos presentan diferencias en cuanto a su composición y operación, su principio de funcionamiento es idéntico: Contabilizar la cantidad de luz que logra atravesar la muestra a una determinada longitud de onda.

Esto se fundamenta en la Ley de Beer-Lambert [53], que proporciona una relación de carácter experimental entre la cantidad de luz que logra atravesar la muestra con las propiedades físico-químicas de la misma.

$$T = \frac{I}{I_0} \quad (1.1)$$

En la Ecuación (1.1) se muestra dicha relación, donde la transmitancia T es el cociente entre la cantidad de luz que logra atravesar la muestra I , y la cantidad de luz que incide sobre esta, I_0 la cual puede entenderse como la cantidad de luz emitida por la fuente (si consideramos que no hay pérdidas por difracción antes de incidir) o la cantidad de luz que logra atravesar una muestra de referencia, comúnmente agua destilada.

A partir de esta relación, se puede definir el término absorbancia, A , Ecuación (1.2), como el logaritmo de la transmitancia, y cambiado de signo para que proporcione un valor positivo.

$$A = -\log(T) \quad (1.2)$$

En la Figura 1.1 se muestra un esquema simplificado de un turbidímetro (Figura 1.1 A) y un espectrofotómetro (Figura 1.1 B). Aunque ambos presentan una configuración similar, estos presentan diferencias significativas.

El turbidímetro (Figura 1.1 A), consta de una fuente de luz infrarroja (880nm) que es dirigida hacia la muestra a través de una serie de lentes. Este tipo de equipos miden lo que se conoce como “Luz Difusa” que es la luz irradiada por la muestra de forma perpendicular a la fuente de luz, de ahí que el sensor se coloque formando un ángulo de 90° con esta, estando esto regulado por las normas USEPA 180.1 e ISO 7027.

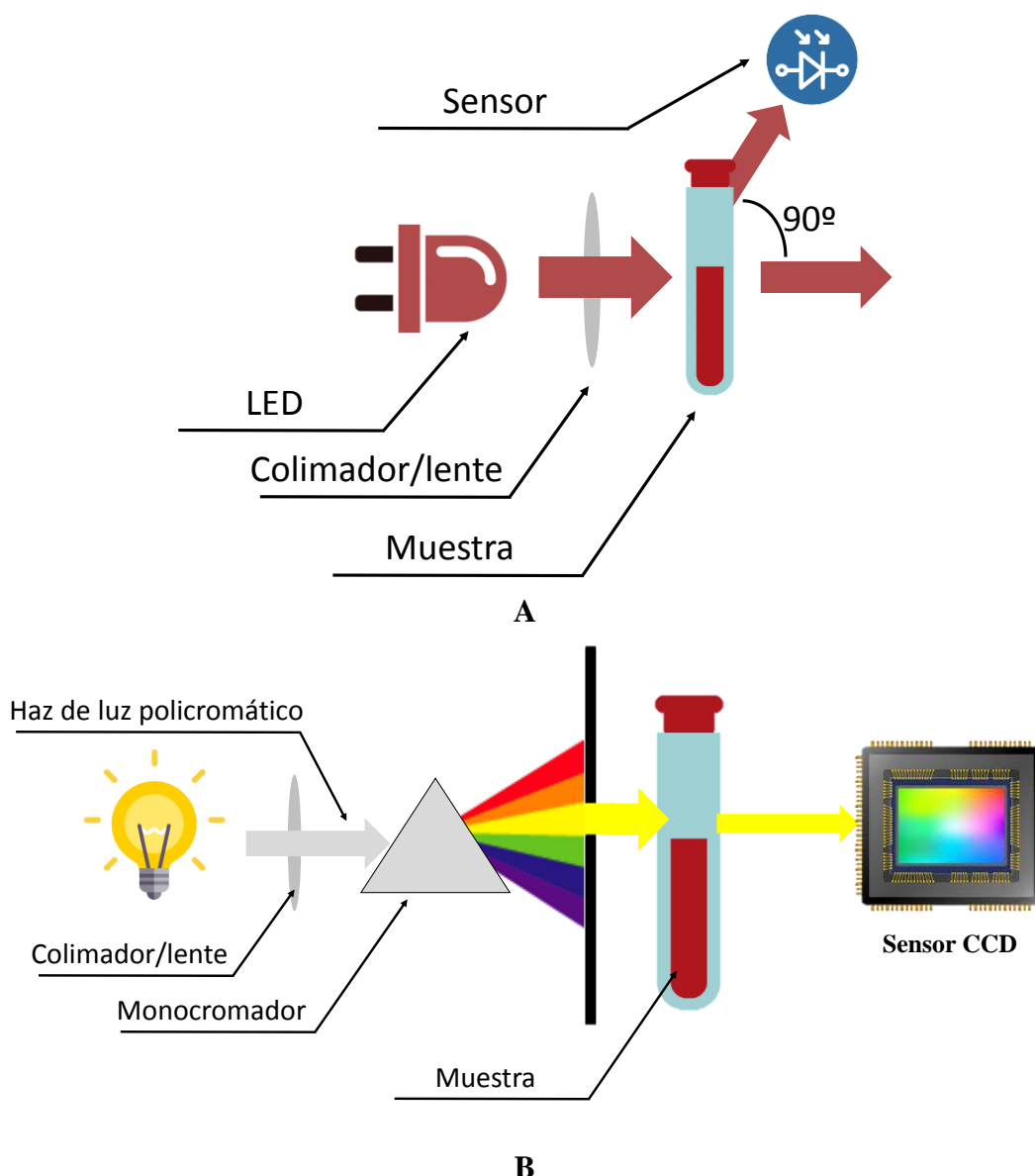


Figura 1.1. Esquema básico de un turbidímetro y espectrofotómetro.

Un espectrofotómetro (Figura 1.1 B), presenta una estructura similar al turbidímetro, con la diferencia de utilizar un rango de longitudes de onda mucho más amplio (desde el ultravioleta hasta el infrarrojo, es decir, 100-2500nm) y al hecho de que la fuente de luz, la muestra y el sensor se encuentran alineados, al contrario que un turbidímetro donde el sensor se dispone de forma perpendicular.

El uso de lámparas incandescentes es habitual en este tipo de equipos para la generación del espectro de trabajo, siendo las más habituales las lámparas de tungsteno para la generación del espectro Vis-NIR y lámparas de deuterio para UV-Vis. Dado que estas lámparas emiten una gran variedad de longitudes de onda de forma simultánea, se hace uso de monocromadores o redes de difracción para descomponer (difractar) ese haz de luz en cada una de las longitudes de onda que la constituye. Luego mediante un sistema de lentes, se logra direccionar una longitud de onda específica hacia la muestra rumbo al sensor. Aunque también pueden encontrarse configuraciones donde el haz de luz es difractado después de atravesar la muestra y justo antes de incidir en el sensor.

Esta es la principal diferencia con respecto a los turbidímetros: Un espectrofotómetro lleva a cabo un análisis a múltiples longitudes de onda, al contrario que el turbidímetro, que se limita a una única longitud de onda, 880nm.

Estos equipos constituyen la base de numerosos estudios de medición indirecta de la carga contaminante de las aguas a partir de la respuesta espectral de la misma, como los que se describirán a continuación.

1.3.2 Regresión Lineal Multivariable

Es una de las técnicas más extendidas en el campo del análisis estadístico [54] debido a su simplicidad de uso. Permite la obtención de modelos que a partir de un conjunto de variables de entrada (llamadas regresores), que son capaces de estimar una variable de salida (variable respuesta) con la que dichos regresores presentan cierta correlación. Se trata de una extensión natural de los modelos de regresión simple [55], donde se obtienen una expresión de primer orden basada en una única variable de entrada.

Los modelos de regresión lineal multivariable presentan una estructura como la mostrada en Ecuación (1.3):

$$Y_j = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + u_j \quad (1.3)$$

Donde 'Y'; es la variable respuesta la cual se desea estimar a partir de los regresores 'x'; los cuales se encuentran ponderados por unos coeficientes 'b'; y 'u'; es el residuo.

La aplicación de este tipo de técnicas viene condicionada por el cumplimiento de ciertos requisitos por parte de los datos, tanto de los regresores como de la variable respuesta. En términos generales estos requisitos se pueden resumir en dos aspectos: La independencia entre los datos y el carácter normal de la variable respuesta.

Cuando se introducen variable en un modelo, un aspecto que debe tenerse en cuenta es la existencia de regresores altamente correlados (multicolinealidad) [56-58], es decir, la existencia de variables que puedan ser explicadas a partir de las otras ya presentes en el modelo. Para evitar este tipo de problemas, se suelen recurrir a dos técnicas estadísticas que permiten detectar la existencia de multicolinealidad:

- **Diagrama de dispersión entre pares de variables**

Permite analizar de forma gráfica, qué relación guarda una variable con respecto a otras, de tal manera que aquellas que parezcan correlacionarse con una línea recta (como se muestra a modo de ejemplo en la Figura 1.2 remarcados con un cuadrado de color azul), indican que las variables comparadas pueden explicarse una a partir de la otra y por lo tanto ambas no deben introducirse conjuntamente en el modelo, siendo uno de ellos redundante. Esto también se puede verificar de forma cuantitativa mediante la aparición de coeficientes de Pearson muy altos entre la variables, lo cual también indica la existencia de multicolinealidad. Este procedimiento también permite observar la existencia de valores atípicos de una forma simple, como son los marcados con un círculo de color verde en la Figura 1.2.

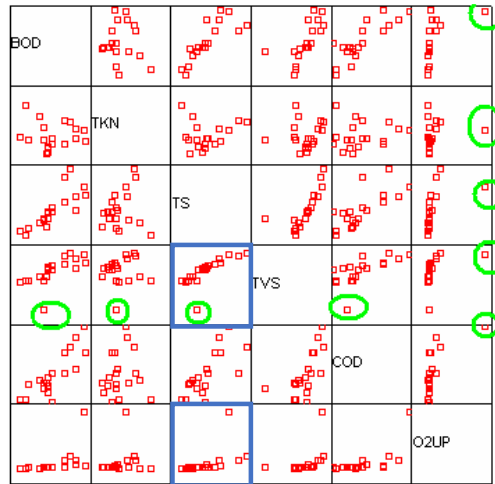


Figura 1.2. Ejemplo de diagrama de dispersión entre pares de variables donde existen variables correladas.

– **Factor de varianza inflada**

La varianza de cada uno de los coeficientes ‘b’ del modelo de regresión lineal descrito en la Ecuación (1.3), puede obtenerse por medio de la Ecuación (1.4)

$$VAR(\widehat{B}_k) = \frac{\widehat{\sigma}^2}{(n-1)(1-R_k^2)S_k^2} \quad (1.4)$$

Donde ‘ R_k^2 ’ representa el coeficiente de determinación de la variable ‘ x_k ’ con respecto al resto de variables, el cual se calcula como el cociente entre la covarianza de una pareja de variables, con respecto al producto de la varianza de dichas variables. ‘ S_k^2 ’ es la varianza de la covariable ‘ x_l ’. Y ‘ $\widehat{\sigma}^2$ ’ es la dispersión de los datos del modelo, y ‘n’ es el tamaño de la muestra.

El problema para estimar la existencia de regresores altamente correlados es que no existe ningún contraste estadístico propiamente dicho para su determinación, por ello surgieron con el paso de los años técnicas como el Factor de Varianza Inflada (FIV) [59], aunque este tipo de técnicas no es infalible [60].

Si un regresor no presentara multicolinealidad, la expresión de la Ecuación (1.4) podría escribirse como se muestra en la Ecuación (1.5):

$$VAR(\widehat{B}_k) = \frac{\widehat{\sigma}^2}{(n-1)S_k^2} \quad (1.5)$$

Donde el cociente entre las Ecuaciones (1.4) y (1.5) nos permite obtener la expresión del Factor de Varianza Inflada (Ecuación (1.6)):

$$FVI(\widehat{B}_k) = \frac{1}{1-R_k^2} \quad (1.6)$$

En la mayoría de la bibliografía establece que la existencia de multicolinealidad viene determinada para un $FVI > 10$ [61], debido a la incertidumbre que puede

estar asociada a este indicador, consideramos que un umbral de 7 resulta más adecuado en base a la propia experiencia con el uso de este tipo de indicadores.

El otro aspecto clave para la aplicación de técnicas de regresión lineal multivariable es que la variable respuesta siga una distribución normal. Para ello se puede recurrir a pruebas de normalidad basadas en contrastes estadísticos, como la de Kolmogorov-Smirnov y Shapiro-Wilk [62, 63] donde la obtención de un P-valor mayor a 0.1 (o 0.05 para un intervalo de confianza del 95%) indican que los test no paramétricos de normalidad reflejan que la variable respuesta sigue una distribución normal.

No obstante, no siempre es posible que la variable respuesta siga una distribución normal. Aunque muchos fenómenos de la naturaleza pueden ajustarse a un modelo gaussiano, esto no siempre se cumple [64]. En ocasiones se puede llevar a cabo la transformación de los datos mediante modelos exponenciales [65] o logarítmicos [66], pero no siempre es así, sobre todo si las variables respuestas presentan mucha varianza entre ellas, como ocurre cuando se aglutinan datos sobre un mismo fenómeno medido en lugares y condiciones diferentes.

Un aspecto clave de este tipo de técnicas es que si el número de datos de partida es muy bajo, es muy fácil que los modelos calculados presenten una alta precisión, es decir un R^2 cercano a la unidad. Más allá del número de muestras, otro aspecto clave es el número de regresores con respecto al número de registros (datos) disponibles. Si un dataset dispone de 50 registros pero cada uno de ellos incluye un número mayor de regresores (por ejemplo 80), al introducir muchas variables en el modelo (siempre y cuando no estén correladas las unas con las otras) se tenderá al obtener un mejor ajuste (R^2) [67], pero en la práctica ese modelo carecerá de validez.

A pesar de estas limitaciones en cuanto a la aplicación de este tipo de modelos, su simplicidad de implementación y facilidad de interpretación de los resultados, hacen que sean una de las técnicas más utilizadas.

Con el fin de mostrar cómo este tipo de técnicas pueden ser utilizadas para la generación de modelos que permitan caracterizar la carga contaminante de las aguas residuales a partir de la información espectrofotométrica de las muestras dentro de un cierto rango de longitudes de onda, a continuación se describirán alguno de los trabajos más relevantes llevados a cabo en este ámbito.

1.3.2.1 Determinación de SST en la Cuenca Ecully, Francia por Bertrand-Krajewski a partir de la turbidez

Los trabajos de Bertrand-Krajewski [68] sobre la cuenca Ecully situada en la ciudad de Lyon en Francia entre el 28 de Mayo y el 18 de Junio de 2003, permitieron obtener unos modelos de correlación lineal entre los sólidos en suspensión totales y la turbidez de las muestras. Para llevar a cabo su estudio, empleó el turbidímetro CUS 1 de la casa Endress+Hauser [69], el cual opera a 880nm de acuerdo con la norma NF EN 27027 [70], asociado a un transmisor Mycom CUM 121-100 para el envío de la información de forma conjunta. Así mismo la información relativa a la turbidez fue registrada con un datalogger (Sofrel S50), que registra los valores de turbidez a intervalos de tiempo de 2 minutos y así como otro que tomaba muestras cada 3 segundos (Ahlborn Almemo V5).

Las muestras registradas por el turbidímetro, fueron corregidas por medio de la ecuación de calibración del equipo, que el autor obtuvo en base a un procedimiento descrito por él en el año 2000 [71] a partir de cuatro valores de referencia de turbidez (20, 200, 1000 y 1800 NTU), los cuales fueron medidos 12 veces cada uno con el turbidímetro, dando lugar a la expresión de corrección de la Ecuación (1.7).

$$T_0 = 0.9858 * T_1 + 3.4745 \quad (1.7)$$

Donde T_1 es el valor de turbidez medido por el equipo antes del proceso de calibración.

Tras la corrección en los datos, aplicó un análisis de regresión lineal simple (dado que el único regresor era la turbidez) obteniendo la expresión mostrada en la Ecuación (1.8), con un nivel de confianza del 95%, denotando la robustez de este tipo de técnicas de medición indirectas a partir de la turbidez.

$$SST_{(mg/l)} = 1.71 * Turbidez_{(NTU)} + 24.6 \quad (1.8)$$

La Figura 1.3 muestra la alta correlación entre los valores de turbidez y la concentración de sólidos en suspensión.

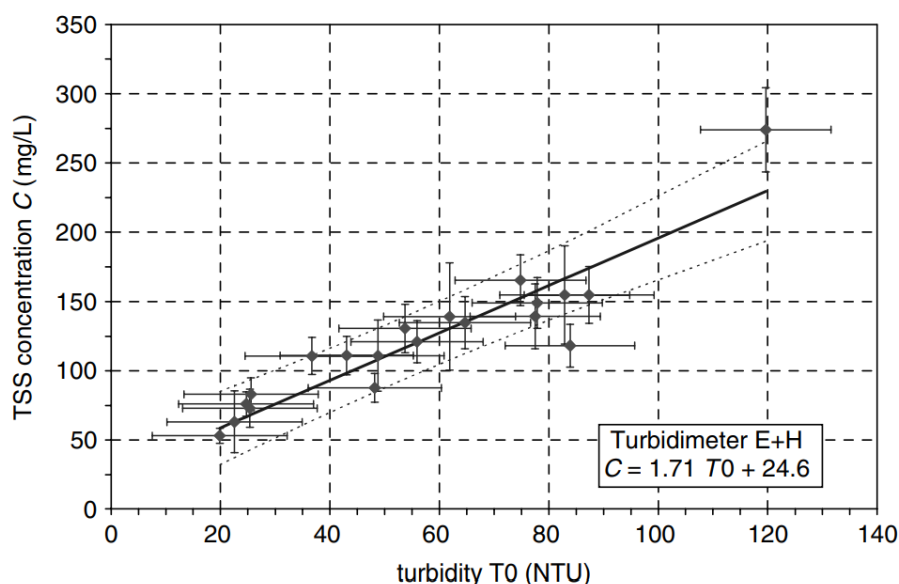


Figura 1.3. Relación SST-Turbidez para la cuenca Ecully. [68]

1.3.2.2 Determinación de SST, DQO y NT en la Cuenca de la ciudad de Pau por Tomas Bersinger a partir de la turbidez y la conductividad

Los estudios llevados a cabo por Bersinger [72] en la Cuenca de la Ciudad de Pao, situada en el suroeste de Francia, tenían como objeto la obtención de modelos que permitan estimar a partir de la turbidez de las muestras los sólidos en suspensión totales, la demanda química de oxígeno. Así mismo también hizo uso de la conductividad de las muestras para la estimación del nitrógeno total (NT).

Para ello, llevó a cabo una campaña de muestreo entre Marzo y Noviembre de 2012 con la ayuda de un toma muestras automático (Muestreador de aguas residuales 3700 de la marca ISCO), obteniendo un total de 108 muestras de agua residual de ámbito doméstico,

comercial e industrial obtenidas en una red de alcantarillado de 800km que daba servicio al 75% de la población.

De estas 108 muestras se utilizaron 70 para llevar a cabo el análisis estadístico de las mismas. Tal y como se ha comentado al inicio de este epígrafe, en la generación de modelos de regresión lineal, las variables respuestas deben seguir una distribución normal. Las muestras obtenidas en este estudio, no presentaba esta normalidad, motivo por el cual fueron sometidas a un proceso de normalización, consistente en el logaritmo neperiano de la variable respuesta (DQO, SST y NT).

De este estudio, se obtuvieron las expresiones mostradas en las Ecuaciones (1.9) – (1.11).

$$DQO_{(mg/l)} = Turbidez_{(NTU)}^{0.917} * e^{1.559} \quad (1.9)$$

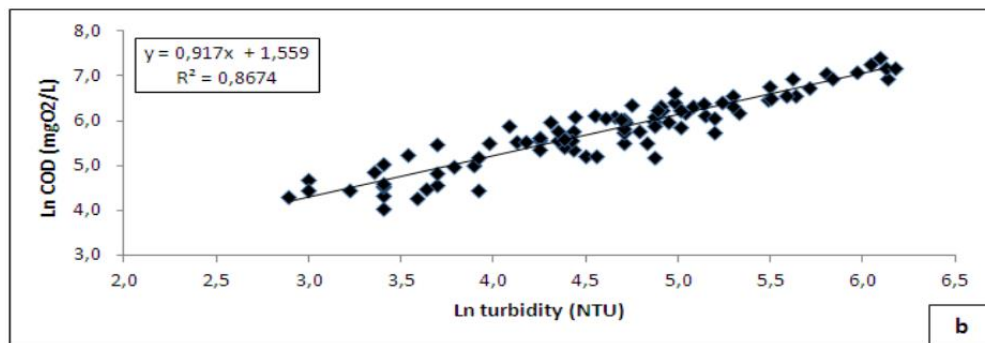
$$SST_{(mg/l)} = Turbidez_{(NTU)}^{1.0124} * e^{0.2894} \quad (1.10)$$

$$NT_{(mg/l)} = Conductividad_{(S/m)}^{1.1165} * e^{-4.0074} \quad (1.11)$$

En la Figura 1.4, se muestra los diagramas de dispersión obtenidos para cada parámetro.

Como se puede observar, las correlaciones son elevadas, especialmente en el cálculo de los SST que supera el 90% de ajuste. Este hecho es relevante, porque confirma que las longitudes de onda próximas al infrarrojo [41] (como las que hacen uso los turbidímetros) guardan una alta correlación con los sólidos en suspensión.

Así mismo, el resto de modelos ha presentado un ajuste algo inferior. El cálculo de la DQO a partir de la turbidez ha presentado una precisión de 86%, un valor alto si tenemos en cuenta que la materia orgánica tiende a presentar una mejor respuesta ante longitudes de onda próximas al ultravioleta [73, 74]. Esto redundaría en la adecuación de este tipo de equipos para llevar a cabo un análisis de la calidad de las aguas de forma simple e in-situ.



A

Figura 1.4. Curvas de correlación entre turbidez y (A) DQO, (B) SST. (C).Curva de correlación entre conductividad y NT. [72]

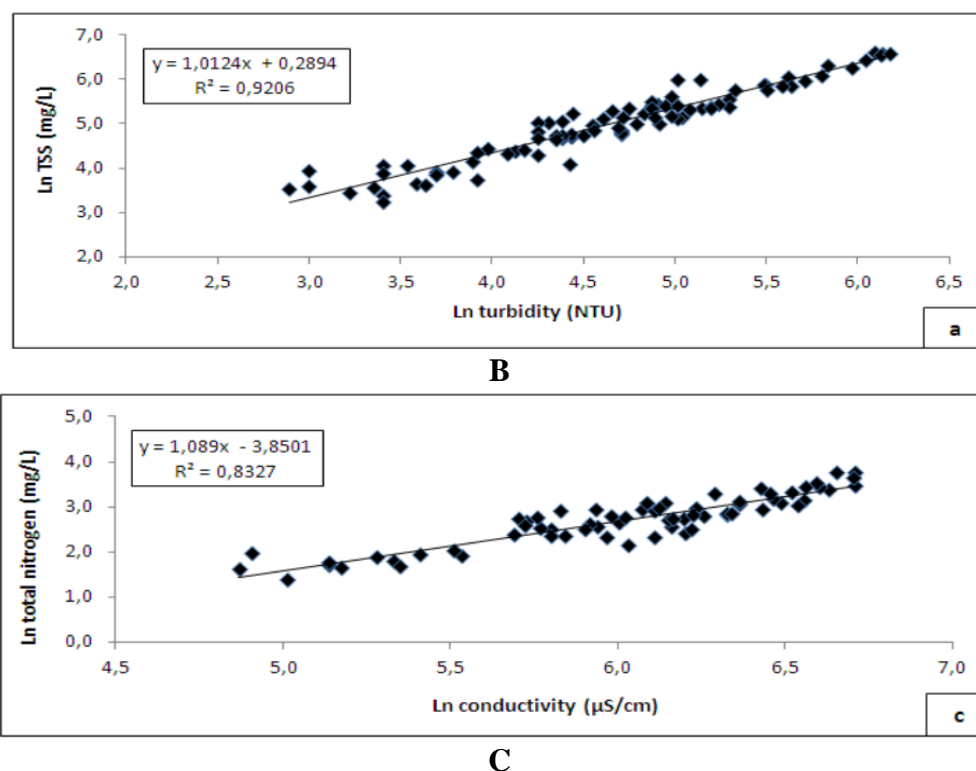


Figura 1.4 Continuación. Curvas de correlación entre turbidez y (A) DQO, (B) SST. (C).Curva de correlación entre conductividad y NT. [72]

1.3.2.3 Determinación de DQO, DBO₅ y SST a partir de turbidez por Del Río, H

Los estudios llevados por Del Río en 2011 [26, 75] en Santiago de Compostela, se obtuvieron los modelos de correlación mostrados en las Ecuaciones (1.12) – (1.14), a partir del análisis de muestras tomadas tanto en tiempo seco como de lluvia.

$$DQO_{(mg/l)} = 3.049 * Turbidez + 174.86 \quad (1.12)$$

$$DBO5_{(mg/l)} = 1.7713 * Turbidez + 112.76 \quad (1.13)$$

$$SST_{(mg/l)} = 1.1878 * Turbidez + 36.419 \quad (1.14)$$

Estos modelos son capaces de explicar a partir de la turbidez la DQO, DBO₅ y SST con una precisión cercana al 80% (o del 90% para la DBO₅)

En la Figura 1.5, se muestra la correlación entre la turbidez y los SST (Figura 1.5 A) así como con la DQO y DBO₅ (Figura 1.5 B).

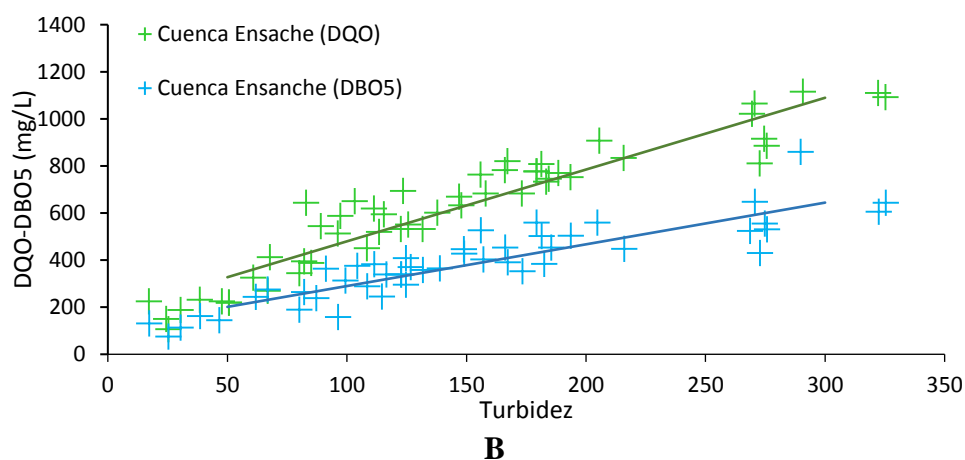
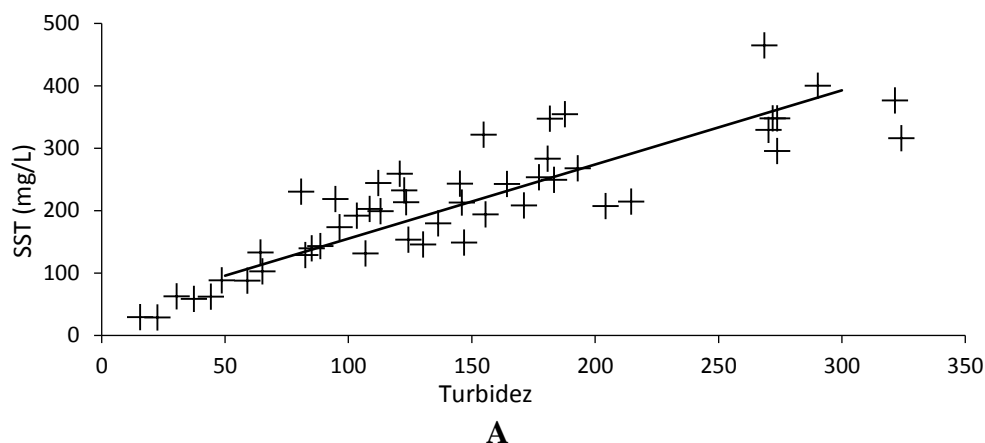


Figura 1.5. Correlación entre la turbidez y (A) SST, (B) DQO/DBO5 para la cuenca Ensanche. [26, 75]

Este estudio constata que la demanda de oxígeno por parte de los microorganismos puede ser estimada a partir de la turbidez de las muestras, más concretamente a partir del espectro infrarrojo. Tal y como se ha mencionado con anterioridad, este hecho tiene gran relevancia dado que la materia orgánica tiende a ser más sensible a la radiación ultravioleta, por lo que al ser posible su medición con este tipo de longitudes de onda, permite el desarrollo de sistemas más sencillos para la caracterización de este tipo de parámetros sin necesidad de recurrir a longitudes de onda más cortas de mayor consumo energético y coste (Ultravioleta).

1.3.2.4 Determinación de DQO, DBO, NT a partir de espectrofotometría de fluorescencia por ultravioleta

Tras enunciar las posibilidades de la medida de la turbidez para la caracterización de ciertos parámetros como la DQO, DBO o los SST, se hace necesario ilustrar cómo es posible el uso de otras partes del espectro para poder llevar a cabo la caracterización de la carga contaminante de las aguas, y sobre todo, la adecuación de los modelos de regresión lineal para estos fines.

El trabajo llevado a cabo por Jin Hur y Jinwoo Cho de la Universidad de Sejong en Korea [73] para la monitorización de la carga contaminante en el Río Gap representa un avance respecto a los sistemas de caracterización antes ilustrados, dado que su estudio no se basa en el uso de turbidímetros (880nm), sino en espectrofotometría de fluorescencia UV, más concretamente, a partir de los valores de absorbancia de las muestras a 220nm y 254nm y matrices de excitación-emisión de fluorescencia (EEM) [76] con análisis factorial paralelo (PARAFAC) [73, 77].

El análisis PARAFAC se utiliza cada vez más para descomponer las matrices de emisión-excitación de fluorescencia (EEM) en sus componentes químicos subyacentes. Cuando la muestra es irradiada con una determinada longitud de onda, esta emite otra longitud de onda con una cierta atenuación, lo que representa con un cierto nivel de absorbancia. Esto proporciona información muy útil sobre las propiedades físico-químicas de las muestras analizadas.

A partir del EEM y PARAFAC, y con la ayuda de Matlab, los investigadores determinaron la tres componente características, los cuales designaron como C1, C2 y C3, cuyos resultados se muestran en la Figura 1.6:

- C1: Sustancias orgánicas de tipo húmico microbiano.
- C2: Sustancias orgánicas de tipo húmico terrestre.
- C3: Sustancias orgánicas de tipo proteico.

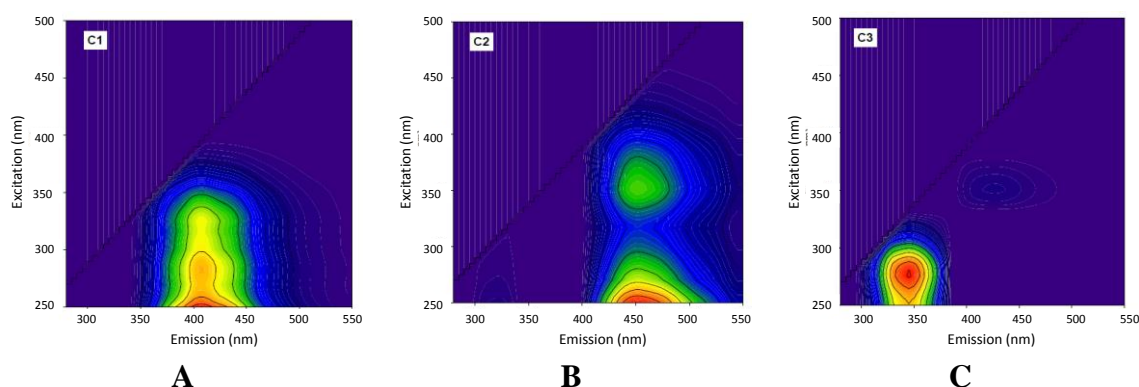


Figura 1.6. Componentes detectados mediante EEM y PARAFAC. [73]

En la Figura 1.6 se muestra la relación entre las longitudes de onda emitidas por la fuente de luz que alcanza a la muestra, con respecto a las longitudes de onda que esta muestra emite, representando los niveles de absorbancia captados de menor (violeta) a mayor (rojo).

De estos tres compuestos, los investigadores utilizaron el designado como C3 (Figura 1.6 C) al ser el que presentaba un cambio más significativo en comparación a los otros (Figura 1.6 A-B), por lo que este componente puede ser utilizado como un buen trazador para determinar el grado de contaminación orgánica.

En las Ecuaciones (1.15)-(1.17), se muestran los modelos obtenidos en el estudio, los cuales ha presentado un alto nivel de correlación.

$$DQO_{(mg/l)} = 177 * Absorbancia_{254nm} - 0.8893 \quad R^2 = 97.3\% \quad (1.15)$$

$$DBO_{(mg/l)} = 0.0082 * C_3 - 0.5679 \quad R^2 = 94.8\% \quad (1.16)$$

$$NT_{(mg/l)} = 5.26 * Absorbancia_{220nm} + 0.456 \quad R^2 = 91.1\% \quad (1.17)$$

En la Figura 1.7, se muestran los ajustes obtenidos. Como puede observarse, en todos los casos los modelos de regresión han permitido obtener un ajuste superior al 90%.

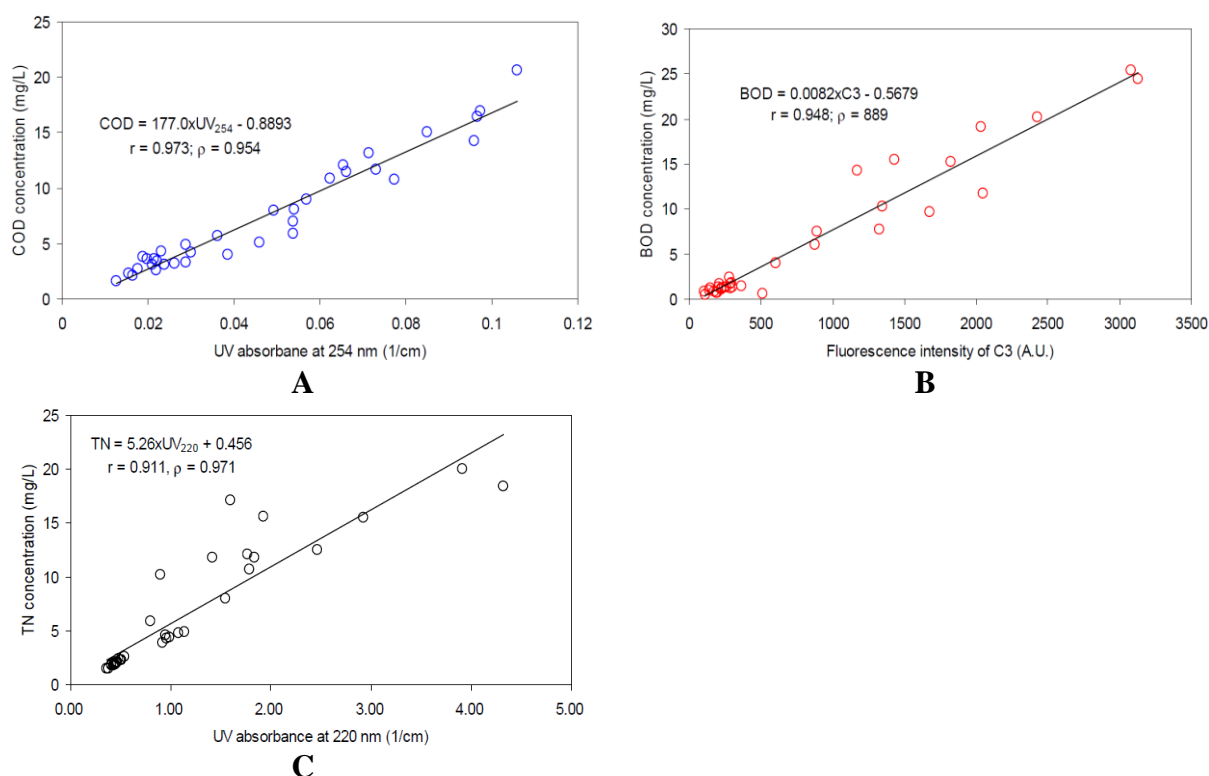


Figura 1.7. Diagramas de correlación para la estimación de (A) DQO, (B) DBO, (C) NT. [73]

1.3.3 Regresión por mínimos cuadrados parciales (PLSR)

Los modelos de regresión lineal comentados en el apartado anterior resultan adecuados para la estimación de la carga contaminante a partir de variables como la turbidez o la respuesta espectral a ciertas longitudes de onda. No obstante, su uso implica que los datos cumplan ciertas características:

- Que las variables sigan una distribución normal
- No exista multicolinealidad
- El número de datos debe ser mayor al número de regresores

En base a estas limitaciones, numerosos investigadores ha abordado otras técnicas de análisis que permiten superar estas limitaciones, como es la técnica de regresión por mínimos cuadrados parciales o PLSR por sus siglas en inglés.

Esta técnica, entre otros motivos, surge para solventar los problemas de multicolinealidad en los modelos de regresión, sobre todo cuando el número de regresores es mayor (o muy alto) en comparación con el número de datos.

Lo que hace esta técnica es crear nuevos regresores que sean combinación de otros ya existentes, de forma que se consigue generar un número inferior de parámetros (regresores) que permite solventar los problemas de multicolinealidad y poder realizar análisis válidos incluso si el número de variables supera al número de muestras. Un ejemplo de esto es el caso planteado por Svante Wold, hijo de Herman Wold que es el creador de esta técnica, el cual fue capaz de realizar mediante PLSR una regresión con 10 regresores y un único caso, frente al mínimo de 5 casos por variable que nos recomienda la estadística tradicional [78].

La regresión PLS es muy utilizada en el campo de la industria química. Una aplicación típica es modelar la relación entre mediciones espectrales pertenecientes al NIR (cerca del infrarrojo), IR (infrarrojo) y UV (ultravioleta), dado que estas suelen estar correlacionadas entre sí, y guardan relación con la propiedad físico-química que se desea analizar del producto industrial, como pueden ser los polímeros.

Este tipo de técnicas, a diferencia de los modelos de regresión, no están orientadas a descargar regresores, sino a obtener estimaciones a partir de un cierto número de variables artificiales.

A lo largo de las décadas han surgido diferentes algoritmos de regresión PLS, donde uno de los más habituales es el algoritmo iterativo no lineal de mínimos cuadrados parciales (NIPALS) [79] desarrollado por Herman Wold. Este algoritmo reduce el número de regresores mediante el uso de una técnica similar al análisis de componentes principales, con el fin de determinar qué variables guarda una mayor relación con la variable respuesta, es decir, qué variables están más correlacionadas con esta.

Además, esta técnica permite, al contrario de la regresión por mínimos cuadrados, el cálculo de múltiples variables respuesta en un único modelo, lo que dota de mayor flexibilidad a la hora de llevar a cabo estimaciones de múltiples parámetros de forma simultánea.

Con el fin de mostrar cómo este tipo de técnicas pueden ser utilizadas para la generación de modelos que permitan caracterizar la carga contaminante de las aguas residuales, a continuación se describirán alguno de los trabajos más relevantes llevados a cabo con esta técnica.

1.3.3.1 Determinación de DQO y DBO₅ a partir de NIR utilizando modelos PLSR

Los trabajos realizados por Qiong YANG, Zhenyao LIU y Jidong YANG [80], investigadores de la Escuela de Química e Ingeniería Química de Chongqing, China, muestran como la espectrofotometría NIR puede ser combinada con las técnicas PLSR para estimar la DQO y DBO₅ de las muestras, con una precisión superior al 95%, más concretamente del 95.42% y 96.52% respectivamente.

En su estudio se analizaron un total de 120 muestras mediante el espectrofotómetro U-4100 UV-Vis-NIR [81], con rango espectral entre 240-2600nm, aunque el estudio se centró en el rango de 800-1800nm con saltos de 2nm. Los datos fueron analizados utilizando el software Unscrambler™, que es una herramienta de análisis y modelación muy utilizada en entornos industriales, y que tiene soporte para aplicaciones de PLSR [82].

Como paso previo al análisis del modelo, se llevó a cabo la eliminación de valores atípicos en la respuesta espectral de las muestras mediante el procedimiento conocido como “Método de comprobación cruzada” [83].

En la Figura 1.8, se muestran los diagramas de dispersión obtenidos al correlacionar la DQO (Figura 1.8 A) y DBO₅ (Figura 1.8 B) con un regresor artificial generador de la combinación de los valores espectrales de las 500 longitudes de onda medidas (800-1800nm en saltos de 2nm).

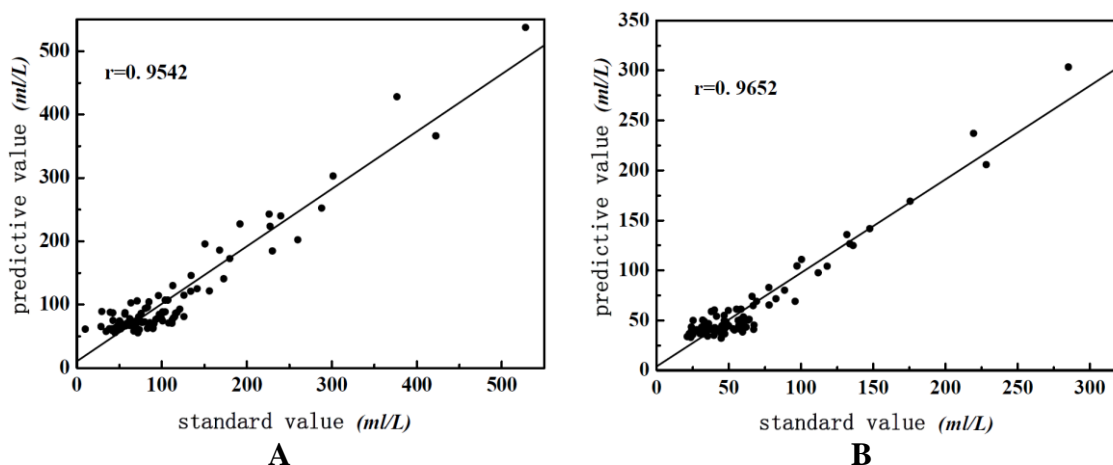


Figura 1.8. Diagrama de dispersión entre NIR y (A) DQO (B) DBO₅. [80]

Este procedimiento ha permitido condensar un gran número de regresores en una única variable que guarda una alta correlación con las variables respuestas estudiadas.

No obstante, se observa una gran concentración de datos al inicio del diagrama de dispersión, donde el efecto de datos más alejados tiende a mejorar el ajuste del modelo. Si esos datos extremos fueran eliminados, se podría evaluar con mayor claridad el grado de precisión en la estimación. Dado que los investigadores no han divulgado su modelo, este punto es difícil de evaluar.

En todo caso, las estimaciones mediante este procedimiento resultan adecuadas para la caracterización de las aguas residuales in-situ.

1.3.4 Regresión no lineal multivariada basada en técnicas de machine learning

Aunque los modelos de regresión lineal son muy utilizados por su facilidad de implementación, su uso viene condicionado a la naturaleza de los datos, dado que estos deben seguir una distribución normal. Cuando se trabajan con conjuntos de datos que han sido obtenidos de diferentes fuentes, como podría ser diferentes puntos de vertidos en el caso del análisis de las aguas, es habitual que exista una varianza significativa entre cada conjunto, lo cual impide en muchas ocasiones la linealización de los mismos.

Los modelos lineales tratan de aproximar la variable respuesta a un conjunto de rectas, sin embargo, esta forma de aproximación no siempre resulta óptima.

En las últimas décadas, con el avance de la computación, han proliferado los modelos de regresión no lineal, que permiten modelizar relaciones de naturaleza no lineal entre

variables, las cuales se han fundamentado en técnicas de machine learning [47]. Esto permite, no sólo lograr modelos más precisos, sino que permite correlacionar variables que no podían ser analizadas bajo técnicas de regresión lineal.

Entre las técnicas más comunes en este campo destacamos las siguientes:

1. Máquinas de vector de soporte [47-50]
2. Redes neuronales [51]
3. Algoritmos evolutivos [52]

Las **Máquinas de vectores de soporte** [47-50], es una técnica desarrollada durante la década de los 90, la cual estaba orientada a la clasificación binaria de conjuntos de datos, sin embargo, su uso se ha extendió a un enfoque de clasificación múltiples.

Esta se fundamenta en la definición de un hiperplano afín [47-50] de $p-1$ dimensiones, es decir un subespacio plano que no tiene por qué pasar por el origen. En una clasificación binaria, el hiperplano se reduce a un espacio en dos dimensiones ($p=2$), es decir, un subespacio de 1 dimensión ($p-1$), lo que equivale a una recta que divide al conjunto de datos en dos zonas, tal y como se muestra en la Figura 1.9 A. Mientras que en un espacio tridimensional, el hiperplano es un subespacio de dos dimensiones (Figura 1.9 B)

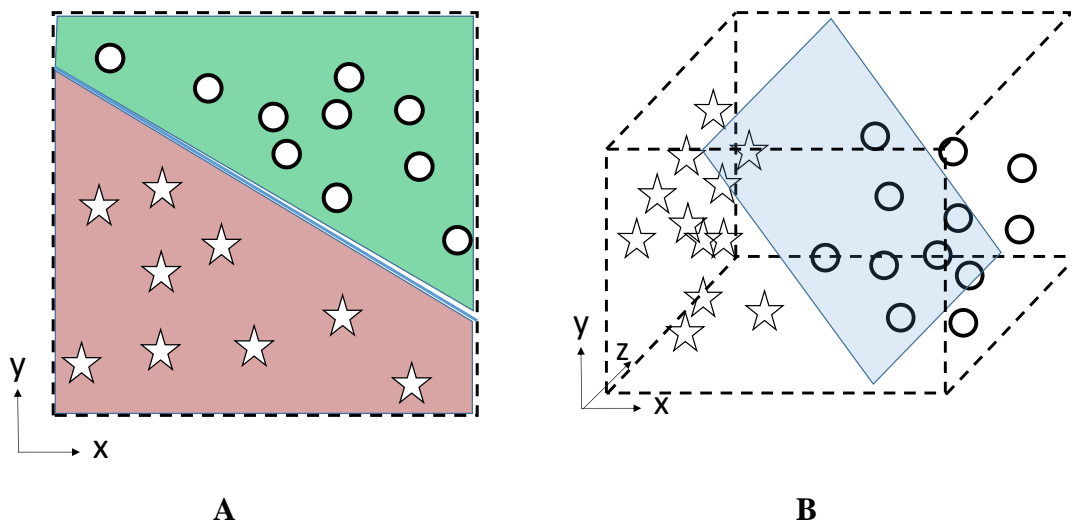


Figura 1.9. Ejemplos de hiperplanos en Máquinas de vector de soportes.

A nivel matemático, este hiperplano para un espacio de p dimensiones se puede definir como se muestra en la Ecuación (1.18):

$$\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p = 0 \quad (1.18)$$

Donde ‘ β ’ son coeficientes propios del hiperplano. Las parejas de puntos ‘ X ’ (por ejemplo X_1 y X_2), que satisfagan la igualdad de la Ecuación (1.18), serán aquellos puntos que pertenezcan al hiperplano. En caso de no cumplirse dicha igualdad, significaría que esa pareja de puntos se sitúa a un lado u otro del mismo, lo que permite la clasificación de los datos en diferentes grupos o categorías.

No obstante, en muy pocas ocasiones será posible encontrar un hiperplano que separe a los conjuntos de datos de forma perfecta como se ha mostrado en la Figura 1.9, sino que lo más normal es encontrar configuraciones como las mostradas en la Figura 1.10.

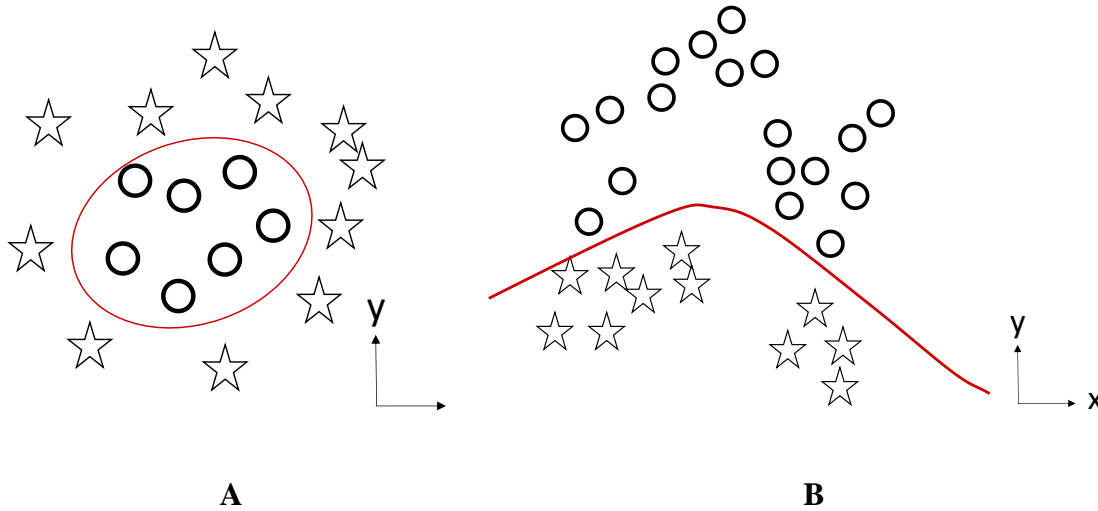


Figura 1.10. Conjuntos de datos que no pueden ser separados de forma perfecta por un hiperplano.

En ninguno de los casos mostrados en la Figura 1.10, es posible encontrar un plano que separe a los datos de forma perfecta debido a la complejidad de los mismos. Por ello, la Máquinas de vectores de soporte usa la Función de Kernel [84], que en esencia, permite la proyección de la información de un espacio de característica a otro de mayor dimensión logrando obtener una solución lineal, es decir que puede ser separada por una línea en ese nuevo espacio.

En la bibliografía existen diferentes tipos de funciones Kernel que pueden ser implementadas, donde alguna de las más relevantes son las descritas en las Ecuaciones (1.19)-(1.21):

$$\text{Lineal} \rightarrow K(X_i, X_j) = (X_i * X_j)^n \quad (1.19)$$

$$\text{Perceptron} \rightarrow K(X_i, X_j) = \|X_i - X_j\| \quad (1.20)$$

$$\text{Gaussiana} \rightarrow K(X_i, X_j) = e^{-\frac{\|X_i - X_j\|^2}{2\sigma^2}} \quad (1.21)$$

Donde X_i, X_j son dos puntos cualquiera, σ es la desviación típica de los datos y n un coeficiente.

Este tipo de técnicas se limitan a establecer una clasificación de los datos, es decir, no son válidas para obtener información de carácter cuantitativo, sólo cualitativo.

Esto resulta útil en el caso de querer determinar si una muestra de agua contiene o no unos ciertos parámetros contaminantes o si estos superan un cierto umbral. No obstante, este tipo de técnicas de clasificación se suelen utilizar de forma conjunta con otras basadas en redes neuronales [51] para poder obtener información de carácter cuantitativo.

Las **redes neuronales** están constituidas por diferentes capas, donde en cada una de ellas existen nodos, también denominados neuronas, las cuales se interconectan entre sí. Cada una de esas interconexiones está ponderadas por un coeficiente llamado “peso” (Figura 1.11). Es precisamente la determinación de esos pesos lo que se persigue cuando se entrena una red neuronal.

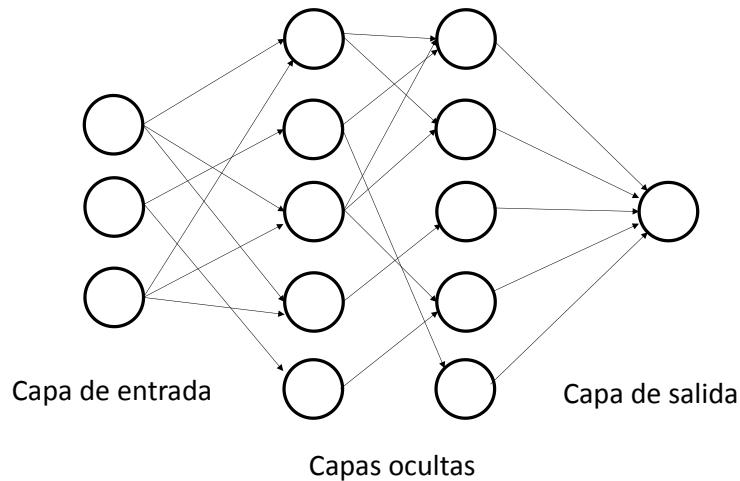


Figura 1.11. Ejemplo de red neuronal convolucional.

En cada iteración, la red toma una serie de datos de entrada y los procesa (pondera) en su paso por cada una de las neuronas hasta alcanzar una salida. Esta es comparada durante la fase de entrenamiento con el valor esperado (valor de referencia), donde el error, es decir, la diferencia entre el valor estimado y el de referencia, se traslada a las capas anteriores en un proceso conocido como *backpropagation* [85] para tratar de realizar cambios en los pesos de las distintas neuronas para poder reducir este error.

Un caso particular de redes neuronales son los llamados **algoritmos genéticos** o **algoritmos evolutivos**. De forma análoga a como un proceso de selección natural funciona, estos parten de una población inicial de funciones de estimación (que pueden ser generadas al azar o mediante diferentes técnicas como la regresión lineal) que son sometidas a proceso de mutación, con la esperanza de alcanzar en alguna generación, un modelo que permita obtener estimaciones con el nivel de precisión deseado.

El problema que presentan los algoritmos basados en redes neuronales, es que no se sabe a ciencia cierta cómo funcionan. Aunque se puede visualizar las capas, neuronas y pesos que los constituye, no es posible saber el porqué de dichos pesos y relaciones entre las neuronas.

Esto plantea una dificultad a la hora de tratar de implementar este tipo de modelos en equipos de baja capacidad de computación, dado que estos precisarían de toda la red para poder funcionar, aspecto que no siempre es viable de implementar.

Para solventar esta problemática, dentro de los algoritmos genéticos, encontramos un tipo de estudio denominado Regresión simbólica [86]. Esto permite expresar la red como una expresión matemática de mayor o menor complejidad (número de variables y operaciones). Este tipo de algoritmos se basan en árboles evolutivos como el mostrado en la Figura 1.12.

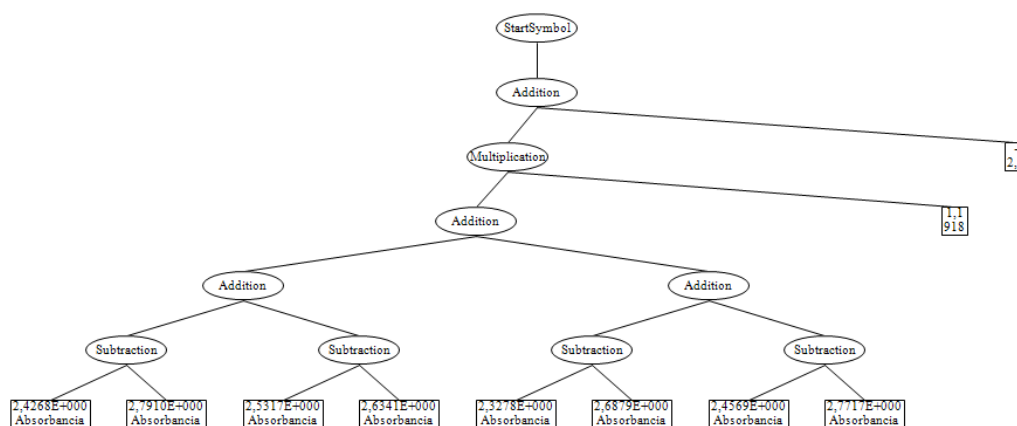


Figura 1.12. Esquema de árbol evolutivo en algoritmos genéticos.

Cada árbol se genera a partir de un nodo de adición, del que se deriva un número aleatorio de nodos, que pueden ser constantes, variables u operaciones matemáticas de cualquier tipo. Los nodos que consisten en operaciones, tendrán nuevos nodos descendentes, que pueden ser de nuevo, constantes, variables u operaciones. El proceso de ramificación del árbol continúa hasta que todas las terminaciones son constantes o variables, o se supera la longitud total y/o la profundidad del árbol.

Estos son generados aleatoriamente, se prueba con los datos de entrada para comprobar cuán cerca está la variable de respuesta de los valores de referencia. Los árboles más cercanos a este resultado serán mutados (combinados) para generar otros árboles, repitiéndose este proceso hasta un cierto número de generaciones y/o alcanzar el nivel de estimación deseado.

1.3.4.1 Determinación de SST a partir de UV-Vis mediante Boosting-IPW-PLS

Los trabajos llevados a cabo por los investigadores Xusong Qin, Furong Gao, Guohua Chen de la universidad de Hong Kong [87], tuvieron como punto de partida el desarrollo de un sistema para la monitorización de la calidad del agua a partir del uso de un espectrómetro UV-Vis y un turbidímetro para monitorizar la DQO, los SST y las grasas y aceites tiempo real.

Para ello, los investigadores implementaron un modelo basado en redes neuronales fruto de la fusión de diferentes técnicas de análisis, con el fin de lograr modelos más robustos. Estos modelos se desarrollaron mediante las técnicas Boosting-IPW-PLS [87] desarrollada por ellos sobre más de 200 muestras de aguas residuales.

El método Boosting-IPW-PLS incorpora IPW (Inverse Probability Weighting o Ponderación de Probabilidad Inversa) [88] en el esquema de boosting para suprimir las variables irrelevantes para determinar la calidad de las aguas, mediante la asignación de pequeñas ponderaciones. El IPW es una técnica estadística que permite utilizar una seudopoblación diferente a la original, lo cual resulta de utilidad cuando no es posible utilizar directamente los datos de partida, por ejemplo cuando existen datos ausentes.

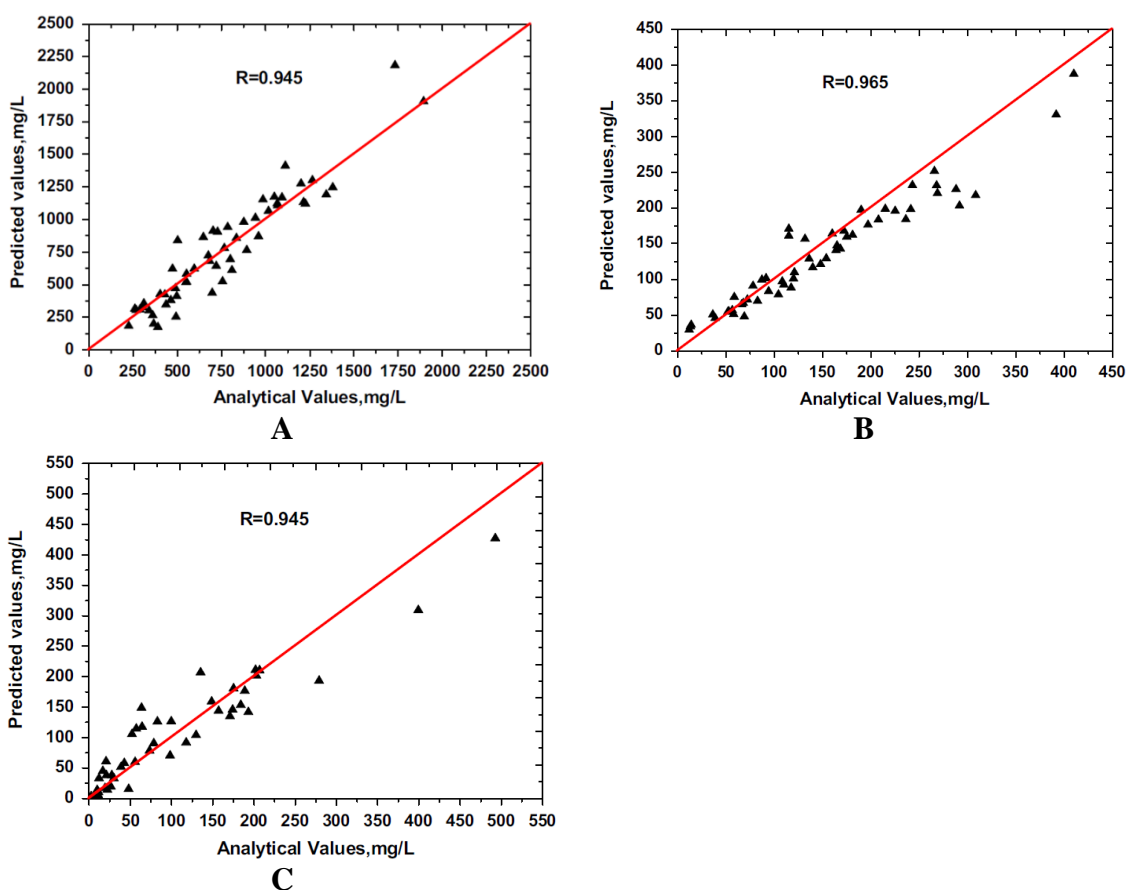


Figura 1.13. Estimación por medio de la técnica Boosting-IPW-PLS de (A) DQO, (B) SST y (C) Aceites y grasas [87].

Los resultados del estudio han demostrado una alta precisión de los modelos calculados, tal y como se muestra en la Figura 1.13, siendo mejor el ajuste para el caso de la DQO y SST (Figura 1.13 A-B), con un 94.5% y 96.5% respectivamente.

1.3.4.2 Determinación del nitrógeno total mediante redes neuronales y SVM

Los estudios llevados a cabo por los investigadores Hong Guo, Kwanho Jeong, Jiyeon Lim, Jeongwon Jo, Young Mo Kim, Jong-pyo Park, Joon Ha Kim y Kyung Hwa Cho [89] para el cálculo del nitrógeno total presente en el agua residual, han dejado patente la adecuación de las redes neuronales como técnica de análisis de conjuntos de datos no lineales. El estudio se llevó a cabo sobre el efluente de la planta de tratamiento de aguas residuales de Ulsan, Corea del Sur, durante un periodo de 10 meses. Las muestras de agua fueron analizadas en laboratorio y contrastadas con diferentes parámetros relativos a la calidad de las aguas y datos meteorológicos, aunque es necesario resaltar que los autores no han especificado qué parámetros de entrada han utilizado exactamente.

En el estudio, se desarrollaron modelos de aprendizaje automático basados en redes neuronales y máquinas de vectores de soporte para la predicción de la concentración de nitrógeno presente en el agua, donde se recurrió al software Matlab para la generación de dichos modelos.

En la Figura 1.14, se muestra una comparativa entre los niveles de nitrógeno total medidos en el agua con respecto a los valores estimados, tanto por medio de redes neuronales (Figura 1.14 A) como mediante Máquina de vectores de soporte (Figura 1.14 B).

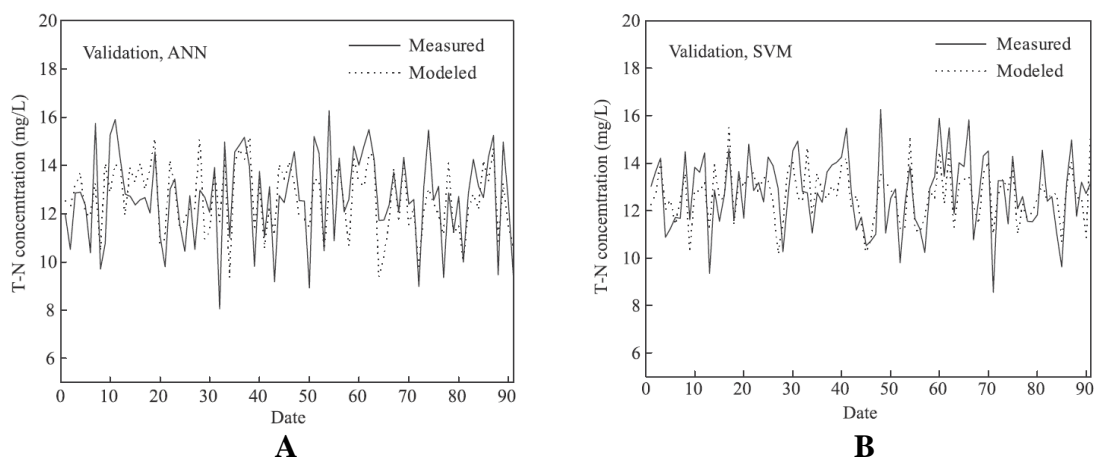


Figura 1.14. Comparativa entre los valores de nitrógeno total medidos con respecto a los estimados mediante (A) Redes neuronales y (B) Máquinas de vectores de soporte. [89]

Como se puede observar, la precisión de la predicción del modelo SVM ha sido muy similar al obtenido por medio de las redes neuronales, entando en ambos casos el ajuste entorno al 95%.

Esto denota las ventajas que este tipo de técnicas presentan respecto a otras basadas en análisis de carácter lineal para la caracterización de la carga contaminante a partir de mediciones indirectas.

1.3.4.2 Determinación de DQO, DBO₅ y SST a partir de las características de la red de captación urbana mediante técnicas de inteligencia artificial

En contraposición otros estudios para la caracterización de la carga contaminante de las aguas a partir de ciertos parámetros intrínsecos a las mismas, como podría ser la respuesta espectrofotométrica o la turbidez, el estudio llevado a cabo por investigadores de la Universidad de Nápoles [90] plantea una enfoque innovador, al utilizar como parámetros de entrada las características del punto de vertido.

El estudio persigue la determinación, por medio de redes neuronales de la DQO, DBO₅ y SST a partir de los siguientes parámetros de la zona de vertido:

- Área de la zona de drenaje
- Porcentaje de viviendas residenciales
- Porcentaje de áreas comerciales
- Porcentaje de área libre de restricciones para el paso del agua
- Precipitación

Para ello desarrollaron dos técnicas machine learning diferentes: Máquinas de vectores de soporte y programación genética. En la Tabla 1.1, se recoge el número de muestras utilizadas para la generación de los modelos, donde destaca el elevado número de datos utilizados para la fase de entrenamiento de los modelos en comparación a los utilizados para

la validación de los mismos, representando el 96%, 94% y 94% de los datos totales respectivamente, lejos de la más habitual proporción de 60%-40% o 70%-30%.

Tabla 1.1. Datos utilizados para la generación de modelos de Máquinas de vectores de soporte y programación genética en estudio. [90]

	Datos entrenamiento	Datos test	R^2
DQO	1108	42	78.4%
DBO₅	680	40	87.1%
SST	811	50	90.6%

En la Figura 1.15 se muestran los resultados logrados por el estudio en base a los parámetros de entrada utilizados. Como se puede observar, los resultados obtenidos por ambas técnicas (Máquinas de vectores de soporte y programación genética) son muy similares, con niveles de ajustes elevado, entre el 80 y 90%. No obstante, la desproporción entre el número de datos utilizados para entrenar los modelos con respecto los que han sido utilizados para validar, hace que estos presenten un ajuste tan elevado, restando cierta generalidad a los mismos.

A pesar de ello, este estudio constata uno de los aspectos claves del uso de técnicas no lineales basadas: La posibilidad de trabajar con datos de entradas de distinta naturaleza para la estimación de un cierto parámetro de salida.

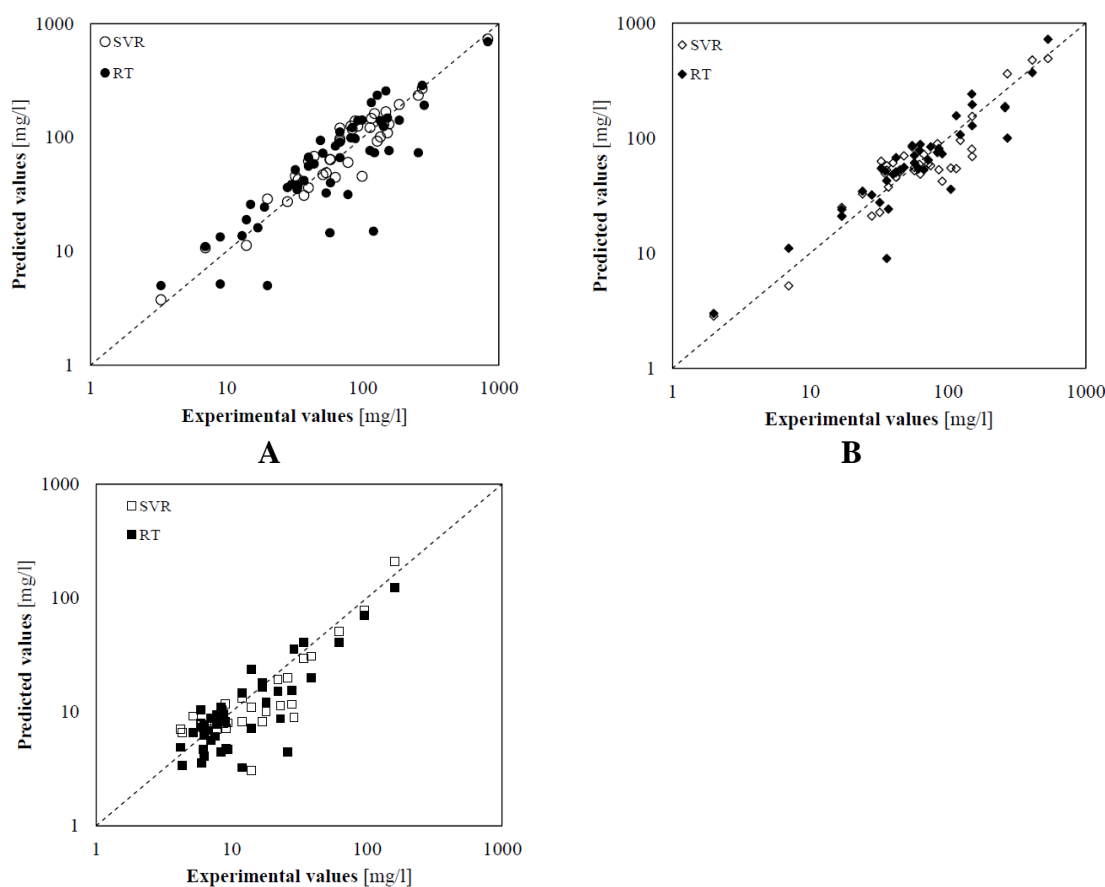


Figura 1.15. Comparativa de resultados obtenidos mediante máquinas de vector de soporte y programación genética para la estimación de: (A) DQO, (B) DBO₅ y (C) SST [90].

1.4 Equipos para la caracterización de la carga contaminante

Los estudios descritos en el epígrafe anterior, hacen uso de espectrofotómetros y turbidímetros (entre otros equipos) para poder estimar, por medio de una serie de modelos de cálculo, determinados parámetros característicos de la carga contaminante de las redes de saneamiento.

Dada las posibilidades que presenta el uso de técnicas como la espectrofotometría a la hora de conocer ciertos parámetros físico-químicos del agua, en la última década han surgido sistemas que, basados en esta técnica, son capaces de estimar parámetros que hasta entonces sólo eran posibles de medir en un laboratorio.

En este epígrafe, se mostrará alguno de los sistemas más relevantes que pueden encontrarse en el mercado para llevar a cabo una monitorización de la calidad de las aguas residuales en tiempo real.

1.4.1 Estación para agua residual S::can

Dentro del campo de la caracterización de la carga contaminante de las aguas residuales a partir del análisis espectrofotométrico, destaca la gama de sondas sumergibles de la multinacional S::can™ [91].

Se tratan de un conjunto de sondas capaces de analizar y calcular en tiempo real hasta 17 parámetros, como la DBO, DQO o SST entre otros. Estas sondas se conectan a un nodo central que permite la visualización de los niveles de carga contaminante al tiempo que envía de dicha información a la nube.

Las sondas son sumergidas en el medio que se desea monitorizar. Gracias a una apertura en su parte inferior, el agua es capaz de atravesar la sonda, donde por medio de lámparas de xenón, se lleva a cabo un análisis espectrofotométrico que sirve de base para la estimación de la carga contaminante. La sonda dispone un mecanismo de auto limpieza basado en un cepillo rotatorio que se encarga de limpiar la zona de medición para dejarla lista para la siguiente medida.

Las sondas pueden operar tanto dentro del espectro visible como ultravioleta, más concretamente entre 190-750nm, gracias a la lámpara de xenón, lo que lleva aparejado altos consumos energéticos que limita su manejo por batería, debiendo estar conectado a un nodo central para su alimentación y manejo.

En la Figura 1.16, se muestra un montaje de sondas de la marca S::can, la cual ha sido extraída de su sitio web oficial: <https://www.s-can.at/es> [91].



Figura 1.16. Estación para agua residual S::can™. [91]

Como se puede observar, el sistema precisa de múltiples sondas para llevar a cabo la caracterización de un total de 17 parámetros, siendo la información procesada, no por la sonda, sino por un nodo central, lo cual encarece el conjunto.

A pesar de la calidad de sus materiales y amplia caracterización de las aguas, su elevado coste y consumos energéticos derivados del uso de lámparas de xenón, limita su uso a ciertos puntos donde existe una conexión a la red eléctrica, especialmente en las zonas próximas a las estaciones de tratamiento de aguas residuales.

1.4.2 Sistema de espectrómetro UV-Vis Roice Water™

Otra sonda que con características similares a la antes descrita es la sonda de espectrofotometría UV-Vis de la marca Roice Water™ [92], con un rango de longitudes de onda que abarca desde 200 a 720nm, por medio de una lámpara de xenón.

En la Figura 1.17 se muestra una vista del sistema extraída del sitio web del fabricante. El sistema consta de un panel central que se encarga de controlar y recibir la información de las sondas, las cuales se conectan al mismo, pudiendo soportar hasta un máximo de 200, según las necesidades del usuario.

De forma similar al sistema S::can [91], este equipo es capaz de detectar de forma simultánea hasta un máximo de 16 parámetros, tales como el amonio, nitrato o SST, entre otros. Esta información es enviada a la nube para un tiempo real para proveer de una monitorización remota de lo que sucede en la red. Así mismo, este hace uso de un cepillo

rotatorio para la limpieza del equipo, no obstante, este mecanismo podría presentar dificultades a la hora de eliminar la totalidad de los sedimentos que pueden quedarse adheridos al equipo, por lo que es necesario la limpieza periódica del equipo por parte del personal encargado de su manejo.



Figura 1.17. Sistema de espectrómetro UV-Vis Roice Water™. [92]

Al igual que el sistema S::can [91], el uso de lámparas de xenón para la generación del espectro de trabajo limita su uso con batería, debido a sus elevados consumos energéticos, de ahí que requiera ser conectado a un nodo central que a su vez se encuentra conectado a la red para poder ser alimentado, lo que limita la portabilidad del conjunto.

1.4.3 Analizador de aguas residuales VWA-3000UV de la empresa Vasthi Engineers Private Limited TM

Se trata de una sonda de características análogas a las anteriores. Por medio de una lámpara de xenón, el sistema, consistente en un espectrofotómetro, es capaz de llevar a cabo un análisis de las aguas en el rango de los 200 a 800nm.

En la Figura 1.18 imagen del analizador de aguas residuales, el cual consta de un panel central que recibe y procesa la información y una única sonda.

El sistema puede estimar parámetros como la DQO, DBO o la concentración de nitratos. La principal diferencia con respecto a los anteriores sistemas, es que esta sonda precisa un periodo de mantenimiento más espaciado en el tiempo, siendo necesario llevar a cabo su limpieza del equipo cada 24 meses.

Sistema compacto, pero supone un alto coste su instalación y mantenimiento para su implantación en la red de saneamiento.

No obstante, el sistema repite las mismas limitaciones que los otros equipos antes descritos, es decir, alto consumo energético asociado al uso de lámparas de xenón y limitada portabilidad dada su dependencia de un nodo central que debe ser conectado a la red.



Figura 1.18. Analizador de aguas residuales VWA-3000UV, de la empresa Vasthi Engineers Private Limited TM. [93]

1.4.4 Sistema ISA con sonda espectrómetro UV-Vis de GO-SystemelektronikTM

El sistema ISA (innovador analizador espectral inteligente, por su siglas en inglés) de la marca GO-SystemelektronikTM [94] está formado por un panel central y una sonda de análisis espectrofotométrico que es capaz de operar entre los 200 y 710nm mediante el uso de lámparas de xenón, de forma análoga a los anteriores.

En la Figura 1.19 se muestra el sistema la cual ha sido extraída de la web de su fabricante [94]. A pesar de estar basada en el uso de lámparas de xenón, este equipo es capaz de analizar de forma simultánea hasta un máximo de 99 parámetros, como la DQO, la concentración de nutrientes entre otros.

Otro aspecto característico de esta sonda, es su sistema de limpieza, el cual se basa en aire comprimido, en lugar de usar un cepillo rotatorio como los sistemas anteriores.

Aunque este sistema presenta grandes ventajas en comparación a los sistemas anteriores como la capacidad de analizar un número mucho más extenso de parámetros, este posee las mismas limitaciones relativas al uso de lámparas de xenón antes descritas.

Esto unido al coste del equipo y a la necesidad de utilizar un elemento central para la conexión y procesamiento de la información de las sondas, limita las posibilidades de ubicar el equipo en algún punto alejado de la propia estación de depuración, dado que este panel central precisa estar conectado a la red para operar, al igual que los sistemas antes descritos.



Figura 1.19. Sistema ISA con sonda espectrómetro UV-Vis de GO-Systemelektronik™. [94]

1.5 Líneas de actuación de la tesis doctoral

Los equipos descritos en el epígrafe anterior, presentan ventajas respecto a los sistemas tradicionales de caracterización de las aguas, donde por medio de mediciones indirectas son capaces de estimar parámetros que hasta la fecha sólo eran posible medir en un laboratorio, permitiendo obtener un estudio de la calidad de las aguas en un corto periodo de tiempo y sin necesidad de emplear químicos o someterlas a pretratamiento.

La totalidad de los sistemas analizados, como el equipo desarrollado por S::can [91] o RoyceWater [92] basan su funcionamiento en el uso de lámparas de xenón para generar el espectro de trabajo, el cual abarca desde el espectro ultravioleta hasta el visible. Aunque este tipo de lámparas son muy comunes en equipos de espectrofotometría, estas tienen aparejado un elevado consumo energético durante su fase de encendido, además de presentar elevadas temperaturas de operación, superiores a los 200°C [95].

Las ventajas que este tipo de lámparas proporcionan a la hora de generar el espectro de trabajo, se convierte en su principal inconveniente a la hora de desarrollar los equipos que operen de forma autónoma, dado que su funcionamiento por batería no resulta viable, tan siquiera si se alimentara mediante energía solar. Esto supedita los equipos a tener que estar próximos a puntos de conexión eléctrica, lo que reduce la portabilidad.

Otra de las limitaciones de los sistemas presentados es el precio, sobre todo si tenemos en cuenta que el análisis se limita al punto donde el equipo es emplazado, por lo que una implantación masiva en la red de saneamiento se traduciría en un desembolso económico muy significativo. Por ello, abogar por sistemas de bajo coste resulta una tendencia que cada vez va ganando más terreno en el campo de la monitorización.

El presente trabajo de investigación busca desarrollar un equipo que pueda contribuir a solventar muchas de estas limitaciones, buscando un enfoque más simple, donde por medio de tecnología LED como fuente de generación del espectro de trabajo se pueda desarrollar dispositivos de bajo consumo y coste que puedan ser emplazados en la red de saneamiento para operar de forma autónoma, tomando las muestras, analizándolas y transmitiendo dicha información a la nube a través de diferentes modos de comunicación, al tiempo que garantice una adecuada limpieza del equipo tras cada análisis con el fin de alargar su vida útil y evitar desviaciones en las mediciones.

Un aspecto clave del presente trabajo de investigación, más allá del desarrollo de equipos para la caracterización de las aguas, es la generación de los modelos estadísticos que permitan estimar los diversos parámetros contaminantes a partir de la respuesta espectral de la muestra [41].

Estos modelos permiten reducir el tiempo de análisis y reducen la necesidad de emplear sondas adicionales o reactivos para caracterizar las aguas, lo que permite obtener equipos más sencillos, de menores costes y dimensiones.

Para la generación de modelos de estimación se hace necesario llevar a cabo una campaña experimental sobre un volumen de muestras mucho mayor al empleado en los estudios descritos en el presente estudio y que exista heterogeneidad entre las mismas, es decir, que se incluyan muestras con diferentes niveles de carga contaminante, con el fin de generar modelos que sean adecuados para cualquier tipo de agua.

Capítulo 2

Espectrofotómetro Desktop

2.1 Introducción

Conocer cómo evoluciona la carga contaminante de las aguas residuales que circulan por la red de saneamiento a lo largo del tiempo resulta de vital importancia para una correcta gestión de las mismas [96].

La presente tesis doctoral se ha enmarcado en el desarrollo de sistemas que permitan determinar la carga contaminante de las redes de saneamiento a partir de la respuesta espectrofotométrica de la muestra. Estos se han concretado en el cálculo de modelos matemáticos de estimación, así como en el desarrollo de diferentes equipos:

- **Espectrofotómetro Desktop**, el cual se describirá en el presente capítulo, y ha sido ideado como un espectrofotómetro de sobremesa automatizado para el análisis de muestras introducidas de forma manual, operando en el rango de 380-700nm [97, 98]. Este equipo sirve de complemento a los análisis basados en la transmitancia de la luz ya presentes en el laboratorio de las plantas de tratamiento de aguas residuales y así poder proporcionar una estimación rápida de la carga contaminante de las aguas.
- **Espectrofotómetro RGB**, que presenta unas dimensiones mucho más reducidas, así como un menor consumo, estando también orientado como espectrofotómetro de longitud de onda reducida de sobremesa controlado por medio de un PC. Su rango de operación es más limitado (510-645nm), siendo adecuada para aplicaciones que no precisen de un análisis multiespectral completo, es decir, de todo el espectro visible, como podría ser el estudio de la clorofila [99].
- **Espectrofotómetro In-Sewer**, una versión autónoma que es capaz de operar desde el interior de la red de saneamiento incluyendo la toma de muestras de agua de forma automática, su análisis y la transmisión de los resultados a la nube. Así mismo, también es capaz de almacenar muestras analizadas, útil para realizar contrastes en laboratorio. Este equipo será detallado en el Capítulo 3.

CAPÍTULO SUJETO A SECRETO
INDUSTRIAL. PÁGINAS 33-64

Capítulo 3

Espectrofotómetro In-Sewer

3.1 Introducción

El objetivo principal de la presente tesis doctoral es la caracterización en tiempo real de la carga contaminante presente en la red de saneamiento, mediante el desarrollo de un equipo portátil que pueda ser emplazado en la propia red, con capacidad de operar de manera autónoma.

Para poder alcanzar este objetivo, inicialmente se desarrolló una versión de sobremesa (Espectrofotómetro Desktop, descrita en el capítulo anterior) que, como se mostrará en los próximos capítulos, solventa la problemática inherente a los equipos de espectrofotometría basados en lámparas incandescentes y que, al mismo tiempo, presenta una nueva metodología para la selección de longitudes de onda sin necesidad de elementos ópticos tales como redes de difracción o monocromadores, reduciendo de forma significativa las dimensiones y consumos del equipo.

A pesar del aceptable rendimiento de este equipo, su diseño exige la intervención del usuario para la introducción manual de las muestras para su análisis, así como su control mediante un software para PC el cual se describe en el Capítulo 4. Más allá de la necesidad de intervención humana para su operación, el Espectrofotómetro Desktop presentaba ciertas características de diseño que no lo hacían apto para operar en un ambiente hostil donde puede estar sometido a golpes y vibraciones que pueden provocar el desajuste de sus partes móviles y donde la existencia de gases corrosivos, como el ácido sulfhídrico, pueden además afectar a su durabilidad.

En base a estas consideraciones de diseño, en el presente capítulo se describe en detalle el diseño e implementación de una versión que puede trabajar de forma autónoma en la red de saneamiento, tomando muestras de manera automática, analizándolas y transmitiendo los resultados de la caracterización a la nube.

Con la intención de permitir almacenar las muestras, además de caracterizar la carga contaminante, el Espectrofotómetro In-Sewer ha sido diseñado para ser capaz de almacenar en su interior un máximo de 31 probetas de 2.5ml ó 18 de 14ml según el tipo de cargador

CAPÍTULO SUJETO A SECRETO
INDUSTRIAL. PÁGINAS 65-94

Capítulo 4

Desarrollo Software

4.1 Introducción

Como se ha comentado, el objetivo principal del presente trabajo de investigación, es el desarrollo de equipos para la caracterización de la carga contaminante basado en la espectrofotometría. A partir de dicho desarrollo, y a partir de los datos de contaminación caracterizados en diversos puntos del sistema de saneamiento, puede surgir la posibilidad de poder disponer un sistema de alerta temprana, que permita conocer en tiempo real la existencia de valores anómalos en diferentes puntos de la red, con el fin de lograr una actuación más coordinada que permita optimizar la capacidad de tratamiento de los sistemas de saneamiento y así mejorar la preservación del medioambiente y luchar contra los vertidos no autorizados.

Los equipos descritos en capítulos anteriores tienen un carácter instrumental, encargados de obtener el espectro de absorción de las muestras analizadas, el cual, aunque guarda relación con la carga contaminante de las muestras y puede servir de orientación sobre cómo de contaminada está el agua desde un punto de vista cualitativo, no proporcionan por sí solos información de carácter cuantitativo, es decir, no indica el valor de DQO o DBO₅, únicamente absorbancia y transmitancia.

Para lograr esta “conversión”, se han desarrollado unas series de modelos de estimación, los cuales son descritos en el Capítulo 7 [41], que analizan la información aportada por los equipos, para proporcionar con una alta precisión una estimación de la carga contaminante, más concretamente: DQO, DBO₅, SST, P, NT, NO₃-N.

Para poder generar estos modelos, era necesario recopilar un elevado número de muestras [151], de las que se dispusiera tanto de la respuesta espectral, como de la carga contaminante medida por la planta de tratamiento, con el fin de encontrar la correlación entre ambos.

Para este fin, se desarrolló un servidor web, que no sólo permite el almacenamiento de los datos antes mencionados, sino que proporciona una interfaz integral para poder conocer la carga contaminante en distintos puntos de la red, de una forma sencilla a partir de la

implementación de los modelos del Capítulo 7 [41], permitiendo conocer cómo evolucionan los contaminantes a lo largo del tiempo.

Todas las herramientas mostradas en el presente capítulo fueron utilizadas por el personal que realiza sus labores en las depuradoras que han participado en la toma de muestras usadas en la presente tesis. Los trabajos han consistido tanto en el manejo de los equipos así como el almacenamiento de los parámetros analíticos medidos en laboratorio.

En el presente capítulo se tratará de forma pormenorizada todos los aspectos relativos a la web desarrollada, la cual es accesible a través de la siguiente dirección: <http://espectrofotometro.ingeniatic.com/>. Para ello se describirán las distintas secciones, así la lógica subyacente, con el fin de que sirvan de guía para el desarrollo de sistemas análogos al implementado.

Así mismo, se describirá los aspectos más importantes de la aplicación para PC desarrollada para el manejo de los equipos mediante conexión USB, así como el envío de la información, especialmente en el caso del Espectrofotómetro Desktop. Aunque también es posible enviar el análisis llevado a cabo por la versión In-Sewer por este método, además de por redes IoT y TCP/IP [138].

4.2 Disposiciones generales

Antes de comenzar con la descripción de las distintas secciones que componen la web y su operativa interna, se hace necesario tratar un aspecto, que aunque puede parecer banal, condiciona enormemente el desarrollo: La tecnología de programación.

Cualquier desarrollo, por muy innovador que sea en su inicio, puede verse rápidamente desfasado si la tecnología utilizada para su implantación se queda obsoleta. Por ese motivo, el servidor fue desarrollado utilizando dos de las tecnologías con mayor proyección en el mercado: Laravel [152, 153] y ReactJs [154].

Laravel es un framework de código abierto diseñado para la creación de aplicaciones y servicios webs basados en PHP. Laravel representa el back-end del servidor, es decir, la lógica subyacente del sistema donde se realiza el control del flujo de información. Por defecto Laravel, emplea una tecnología llamada Vue.js [155] que constituye el front-end del sistema, es decir, la interfaz de usuario.

Sin embargo, Vue.js ha empezado a perder fuerza dentro del desarrollo de interfaces de usuarios, a favor de otras alternativas como ReactJS, una librería desarrollada por Facebook que proporciona un enfoque distinto, más simple y con mayor flexibilidad de desarrollo.

Fruto de esta combinación se ha desarrollado la interfaz mostrada en la Figura 4.1, donde se puede observar el panel de control con las distintas secciones que lo componen.

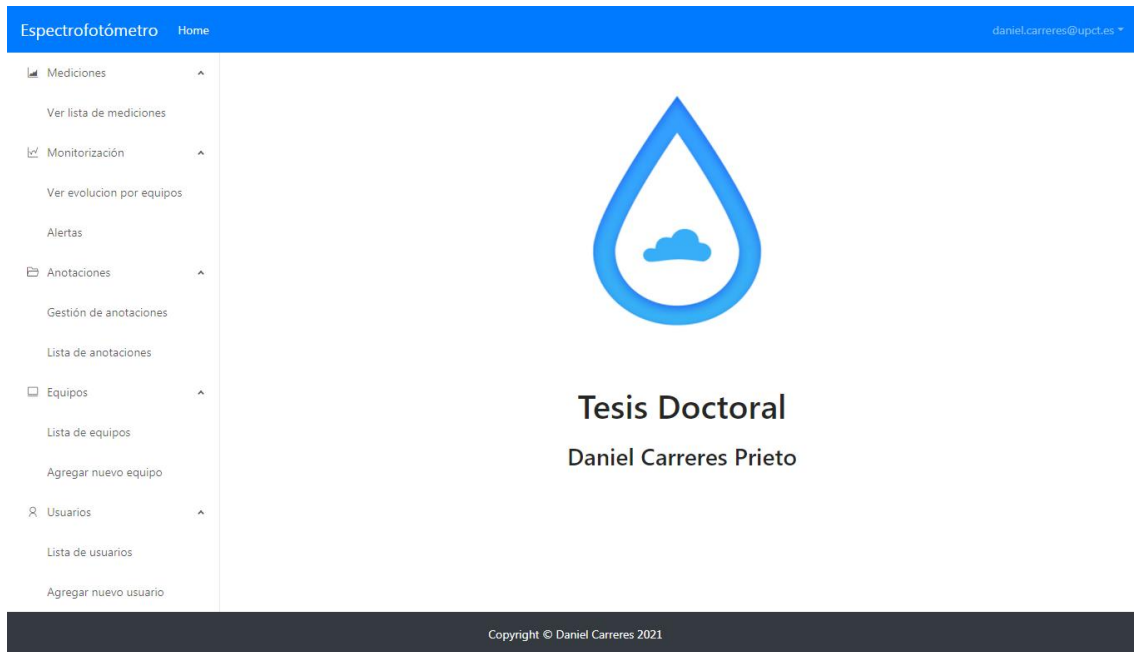


Figura 4.1. Panel de control servidor web.

Como se puede observar, la web dispone de las siguientes secciones:

1. Mediciones

Recoge la lista de muestras analizadas por los distintos equipos, las cuales contienen tanto la respuesta espectral y los niveles de carga contaminante medidos en laboratorio así como la estimación obtenida por medio de los modelos desarrollados.

2. Monitorización

Permite observar la evolución de la carga contaminante medida por equipo para un periodo de tiempo determinado, a partir de los modelos de correlación (Capítulo 7), así como las alertas que han sido registradas a consecuencia de valores de contaminación más altos de los esperados.

3. Anotaciones

Con el fin de mejorar la operativa diaria, se hizo necesario incluir una sección en la que se pudiera agregar anotaciones por parte de los operadores de planta.

4. Equipos

Posibilita tanto la visualización de los distintos equipos registrados, indicándose a qué usuario/usuarios pertenecen así como la posibilidad de ver las mediciones asociadas a cada equipo.

También permite la creación, modificación y eliminación de los mismos.

5. Usuarios

De forma análoga al anterior, esta sección permite visualizar la lista de usuarios dados de alta en el sistema, así como visualizar los equipos asociados a los mismos, además de poder crear, modificar o eliminar dichos usuarios.

4.3 Control de acceso

La seguridad de los datos es un aspecto clave en cualquier tipo de servicio web, y en especial en uno que debe manejar información confidencial de forma segura, como son los datos de la carga contaminante de las aguas.

El sistema ha sido desarrollado para albergar a dos tipos de usuarios:

- **Administradores**
Poseen un control completo del sistema, pudiendo controlar la creación y edición de nuevos usuarios y equipos, así como poder visualizar todas las mediciones registradas por todos los equipos, sin que sea necesario su asignación expresa.
- **Invitados**
Su control se limita a los equipos asignados, y por ende a las mediciones generadas por dichos equipos. No pueden crear nuevos usuarios y equipos, pero sí pueden acceder de forma completa a los informes de caracterización de la carga contaminante así como la evolución de la misma por equipo, de forma análoga a los Administradores.

Para implementar este sistema de credenciales, Laravel dispone de un sistema de acreditación llamado Passport [152, 153], que es un paquete que permite integrar un servidor OAuth2 [156, 157] en el proyecto. OAuth 2.0 es un protocolo para el intercambio de información entre el servidor web y el usuario, que elimina la necesidad de tener que compartir la identidad del mismo, aspecto por el que es utilizado en los sistemas de credenciales de acceso.

OAuth2.0 es un protocolo de autorización que permite a tanto a usuarios como a sistemas, acceder a contenidos propiedad de un usuario, sin que éstos tengan que manejar ni conocer las credenciales del mismo, es decir, su usuario y contraseña.

Este sistema se basa en el empleo de tokens [158], claves alfanuméricas generadas de forma aleatoria durante el alta de un nuevo usuario. Cada uno de estos tokens se asocia con el ID único de cada usuario, de forma que no se envían las credenciales del usuario en ninguna operación.

Las credenciales de cada uno de los usuarios registrados en el sistema, se almacenan en la tabla “users” de la base de datos, tal y como se observa en la Figura 4.2, donde cada usuario dispone un campo “id” único e irrepetible, además del campo “email” que será uno de las credenciales de acceso junto con la contraseña, la cual se encuentra encriptada en la base de datos por seguridad.

Dado el carácter sensible de los datos de los usuarios, y bajo el amparo de la Ley Orgánica 3/2018, Protección de los datos personales y garantías de los digitales [159], los datos mostrados en el presente capítulo relativos a datos de naturaleza personal (correo electrónico y contraseña), corresponderán a datos generados de forma sintética por Laravel para preservar la confidencialidad de los mismos.

			id	name	email	password	
<input type="checkbox"/>	Editar	Copiar	Borrar	35	Rey Deckow	reilly.wanda@example.net	\$2y\$10\$gNxTzVhv61xh47iriSh7iewjABwY/mpO.fJyFPi5i4Z...
<input type="checkbox"/>	Editar	Copiar	Borrar	14	Adah Cassin	monahan.neil@example.net	\$2y\$10\$gNxTzVhv61xh47iriSh7iewjABwY/mpO.fJyFPi5i4Z...
<input type="checkbox"/>	Editar	Copiar	Borrar	6	Marcia Lubowitz	henderson05@example.org	\$2y\$10\$gNxTzVhv61xh47iriSh7iewjABwY/mpO.fJyFPi5i4Z...
<input type="checkbox"/>	Editar	Copiar	Borrar	8	Prof. Riley Rath I	boyle.chaz@example.com	\$2y\$10\$gNxTzVhv61xh47iriSh7iewjABwY/mpO.fJyFPi5i4Z...
<input type="checkbox"/>	Editar	Copiar	Borrar	40	Mr. Sammie Balistreni II	alverta.prosacco@example.com	\$2y\$10\$gNxTzVhv61xh47iriSh7iewjABwY/mpO.fJyFPi5i4Z...
<input type="checkbox"/>	Editar	Copiar	Borrar	5	Alexandre Watsica	lgrady@example.net	\$2y\$10\$gNxTzVhv61xh47iriSh7iewjABwY/mpO.fJyFPi5i4Z...
<input type="checkbox"/>	Editar	Copiar	Borrar	19	Jared Corkery	lempi39@example.net	\$2y\$10\$gNxTzVhv61xh47iriSh7iewjABwY/mpO.fJyFPi5i4Z...
<input type="checkbox"/>	Editar	Copiar	Borrar	12	Jay Padberg	dock52@example.com	\$2y\$10\$gNxTzVhv61xh47iriSh7iewjABwY/mpO.fJyFPi5i4Z...
<input type="checkbox"/>	Editar	Copiar	Borrar	33	Prof. Brandon Cremin	tromp.rosamond@example.net	\$2y\$10\$gNxTzVhv61xh47iriSh7iewjABwY/mpO.fJyFPi5i4Z...

Figura 4.2. Tabla “users” de la base de datos con muestra de credenciales.

Laravel se basa en la creación de controladores [152, 153] para la realización de distintas tareas, como es el registro de nuevos usuarios o la validación de su acceso a la web. Para su implementación, se ha desarrollado un controlador llamado “*RegisterController*”, el cual posee dos métodos o servicios: “register” y “create”.

El primer servicio, es el que realiza una comprobación previa de los datos del usuario que se desea introducir en el sistema antes de crearlo en la base de datos, mientras que el segundo método es el que crea a dicho usuario.

Espectrofotómetro Home daniel.carreres@upct.es

Agregar nuevo usuario

* Nombre y Apellidos: Mariano García ✓

* Correo Electrónico: mariano@email.com ✓

* Contraseña: contraseña ✓

* Repetir contraseña: contraseña ✓

* Organización: Universidad Politécnica de Cartagena (UPCT) ✓

Nivel de permisos: Invitado

Observación: Introduce alguna información relevante sobre el usuario.

Asignación de equipos:

1/6 elementos

Buscar aquí

- Equipo 2 (Desktop)
- Equipo 3 (Desktop)
- Equipo 5-DINTEL (Desktop)
- Espectrofotometro RGB (Desc...
- FAKE EQUIPMENT (Sewer)
- Desktop 6 (Desktop)

3 elementos

Buscar aquí

- Espectro In-Sewer 1 (Sewer)
- Equipo 1 (Desktop)
- Equipo 4 (Desktop)

[CREAR NUEVO USUARIO](#)

Copyright © Daniel Carreres 2021

Figura 4.3. Interfaz para la creación de nuevos usuarios.

En la Figura 4.3 se muestra la interfaz desarrollada para la creación de nuevos usuarios. En ella se deben especificar los siguientes parámetros:

- **Nombre y apellidos**
- **Dirección de correo electrónico**
Es utilizada como usuario para acceder al sistema, así como para el uso de los equipos desarrollados
- **Contraseña**
- **Organización de pertenencia**
Condiciona la asignación de roles.
- **Tipo de usuario**
Pudiendo ser Administrador, con acceso a todos los equipos y control sobre todos los usuarios, o Invitado, limitado a los equipos asignados y con poder de edición limitado.
- **Asignación de equipos**
Dado que la web es utilizada por diversas empresas y laboratorios, se hace necesario definir qué equipos son asignados a cada usuario, con el fin de garantizar la privacidad de las mediciones.

4.4 Información sobre las muestras

Tal y como se ha comentado con anterioridad, los usuarios con la categoría de “Invitado”, sólo tienen acceso a las mediciones realizadas por los equipos asignados, mientras que los Administradores pueden acceder a todos los equipos sin necesidad de que les sean asignados de forma expresa.

Los equipos, tanto la versión de escritorio (Espectrofotómetro Desktop), como la versión autónoma (Espectrofotómetro In-Sewer), envían el análisis espectral de las muestras a la web, las cuales pueden ser visualizadas desde la sección “Lista de mediciones” tal y como se observa en la Figura 4.5.

The screenshot displays the 'Lista de Mediciones' interface. The table contains the following data:

ID	Muestra	Equipo	Fecha	Hora	Ver Muestra	Eliminar
913	TLL2-30092020-FERR-B15-E1	(1) Equipo 1	09/02/2021	15:28:44	Ver Muestra	Eliminar
912	TLL2-30092020-FERR-B14-E1	(1) Equipo 1	09/02/2021	15:17:27	Ver Muestra	Eliminar
911	TLL2-30092020-FERR-B12-E1	(1) Equipo 1	09/02/2021	15:00:21	Ver Muestra	Eliminar
910	TLL2-30092020-FERR-B8-E1	(1) Equipo 1	09/02/2021	13:08:22	Ver Muestra	Eliminar
909	TLL2-30092020-FERR-B9-E1	(1) Equipo 1	09/02/2021	12:55:00	Ver Muestra	Eliminar
908	TLL2-30092020-FERR-B5-E1	(1) Equipo 1	09/02/2021	12:42:29	Ver Muestra	Eliminar
907	TLL2-30092020-FERR-B3-E1	(1) Equipo 1	09/02/2021	12:28:01	Ver Muestra	Eliminar
906	210211050	(8) Desktop 6	05/02/2021	22:34:45	Ver Muestra	Eliminar
905	210211040	(8) Desktop 6	05/02/2021	22:24:09	Ver Muestra	Eliminar
904	210211000	(8) Desktop 6	05/02/2021	22:12:54	Ver Muestra	Eliminar
903	210210990	(8) Desktop 6	05/02/2021	21:59:52	Ver Muestra	Eliminar
902	210210970	(8) Desktop 6	05/02/2021	21:44:36	Ver Muestra	Eliminar
901	210210950	(8) Desktop 6	05/02/2021	21:29:18	Ver Muestra	Eliminar
900	210208860	(8) Desktop 6	04/02/2021	23:06:32	Ver Muestra	Eliminar
899	210208850	(8) Desktop 6	04/02/2021	21:58:33	Ver Muestra	Eliminar
898	210208830	(8) Desktop 6	04/02/2021	21:47:36	Ver Muestra	Eliminar
897	210208820	(8) Desktop 6	04/02/2021	21:36:10	Ver Muestra	Eliminar
896	210208800	(8) Desktop 6	04/02/2021	21:14:51	Ver Muestra	Eliminar
895	210208780	(8) Desktop 6	04/02/2021	21:00:05	Ver Muestra	Eliminar
894	TLL6-22012021-FERR-B22-E1	(1) Equipo 1	04/02/2021	13:37:19	Ver Muestra	Eliminar

Figura 4.5. Lista de muestras analizadas por los equipos desarrollados.

Como se puede observar, cada registro dispone de la siguiente información:

- **Identificador de la muestra (ID)**
Este es único y permite acceder a la información contenida en el mismo, el nombre asignado a la misma
- **Nombre de la muestra**
Durante las campañas de medición, se solicitaron a las entidades de saneamiento que emplearan una nomenclatura que ayudara la identificación de las muestras para poder agregar con posterioridad los valores analíticos medidos en sus laboratorios.
En esa designación, se indica aspectos como la fecha, el tipo de agua analizada o la EDAR/laboratorio que ha llevado la medición, así como el identificador del equipo utilizado para este fin.
- **Equipo**
Identificación y nombre del equipo utilizado para su análisis, para poder saber cómo se distribuye la contaminación a lo largo de la red.
- **Fecha y hora**
- **Ver Muestra/Eliminar**
Desde esta sección se puede acceder a vista en detalle de la muestra seleccionada, lo cual será descrito a lo largo del presente capítulo.
Dado la relevancia de la información contenida en el servidor, en lo relativo a la opción de eliminar muestras en el servidor, era necesario evitar que los datos pudieran ser eliminados por accidente. Por ese motivo, la web no elimina ningún dato del servidor, únicamente se introduce en la base de datos la fecha de eliminación del registro. Esto hace que el servidor obvie el registro, no mostrándose ni en la web ni en la respuesta proporcionada por los servicios, pero dando la posibilidad de recuperarlo en caso de ser necesario.

Toda la operativa interna del servidor se sustenta en servicios webs, que permiten tanto al propio servidor como a aplicaciones externas acceder a los datos contenidos en el mismo. En la Figura 4.6 se muestra servicio tipo GET desarrollado para este fin así como la respuesta obtenida.

```

GET http://espectrofotometro.ingeniatic.com/api/v1/measures?page=1

Authorization Headers (3) Body Pre-request Script Tests

Pretty Raw Preview JSON

[
  {
    "id": 912,
    "user_id": 21,
    "equipment_name": "(1) Equipo 1",
    "name_sample": "TLL2-30092020-FERR-B14-E1",
    "deleted_at": "",
    "created_at": "2021-02-09 14:17:27",
    "updated_at": "2021-02-09 14:17:27"
  },
  {
    "id": 911,
    "user_id": 21,
    "equipment_name": "(1) Equipo 1",
    "name_sample": "TLL2-30092020-FERR-B12-E1",
    "deleted_at": "",
    "created_at": "2021-02-09 14:00:21",
    "updated_at": "2021-02-09 14:00:21"
  },
  {
    "id": 910,
    "user_id": 21,
    "equipment_name": "(1) Equipo 1",
    "name_sample": "TLL2-30092020-FERR-B8-E1",
    "deleted_at": "",
    "created_at": "2021-02-09 12:08:22",
    "updated_at": "2021-02-09 12:08:22"
  },
  {
    "id": 909,
    "user_id": 21,
  }
]

```

Figura 4.6. Respuesta servicio web para la recuperación de la lista de muestras almacenadas.

En este punto es necesario destacar que los únicos datos que se suben de forma automática a la web son las mediciones realizadas por los equipos, mientras que los datos analíticos medidos en los laboratorios de las plantas de tratamiento son suministrados con posterioridad, debido a que esta caracterización precisa de numerosos ensayos. La información es suministra por las plantas en un fichero Excel, el cual es subido de forma automática al servidor por medio de una aplicación desarrollada, la cual se muestra en la Figura 4.7.

Fecha	Tipo	Q. Entr.	pH	CONDUCT.	SST	DBO5	DOO	NH3-N	NO3-N	NTK	Ntotal	Ptotal
03/02/20	BRUTA	24554.0	7.46	2730.0	368.0	540.0	848.0	58.0	0.6	0.0	74.0	9.8
04/02/20	BRUTA	24004.0	7.54	2560.0	368.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
05/02/20	BRUTA	23984.0	7.5	2530.0	296.0	560.0	896.0	60.0	0.5	0.0	60.0	9.6
06/02/20	BRUTA	23219.0	7.53	2310.0	900.0	0.0	866.0	0.0	0.0	0.0	0.0	0.0
07/02/20	BRUTA	23310.0	7.6	2500.0	280.0	500.0	774.0	71.0	0.5	0.0	83.0	10.4
10/02/20	BRUTA	22417.0	7.49	2670.0	708.0	760.0	1403.0	66.0	0.9	0.0	88.0	11.4
11/02/20	BRUTA	23984.0	7.56	2710.0	436.0	600.0	1038.0	68.0	0.9	0.0	90.0	11.1
13/02/20	BRUTA	23531.0	7.56	2510.0	380.0	560.0	936.0	67.0	1.3	0.0	91.0	10.3
17/02/20	BRUTA	24444.0	7.54	2590.0	516.0	660.0	1108.0	69.0	1.0	0.0	86.0	10.1
18/02/20	BRUTA	22776.0	7.53	2630.0	404.0	640.0	993.0	67.0	0.5	0.0	85.0	10.9
20/02/20	BRUTA	22943.0	7.03	2480.0	548.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
21/02/20	BRUTA	23105.0	7.82	2670.0	400.0	600.0	1021.0	63.0	0.3	0.0	87.0	10.3
24/02/20	BRUTA	23137.0	7.16	2400.0	904.0	920.0	1816.0	71.0	1.1	0.0	90.0	9.6
25/02/20	BRUTA	23152.0	7.65	2330.0	472.0	0.0	1231.0	0.0	0.0	0.0	0.0	0.0
26/02/20	BRUTA	23205.0	7.49	2390.0	300.0	560.0	814.0	68.0	0.4	0.0	89.0	11.2
27/02/20	BRUTA	23137.0	7.57	2420.0	284.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
28/02/20	BRUTA	23320.0	7.6	2400.0	320.0	540.0	849.0	69.0	0.7	0.0	86.0	10.5

MUESTRA EDAR = MUESTRA WEB ID

28/02/20 - BRUTA = (495) 28/02/20 Bruta-E2 495

Anterior Recargar Subir datos Siguiente

Figura 4.7. Aplicación para la subida de datos analíticos medidos en laboratorio.

Esta aplicación descarga la lista completa de muestras almacenadas en el servidor para poder realizar un proceso de casación con los datos asociados, con los identificadores contenidos en el fichero Excel. Cuando existe una coincidencia, el programa muestra las posibles opciones de casación dejando al usuario la elección final. El motivo de no realizar este proceso de forma automática es que no siempre los identificadores de las muestras analizadas por los equipos coinciden con las contenidas en los ficheros suministrados por las plantas, o más bien no de forma completa.

Cuando se utiliza el Espectrofotómetro Desktop, el usuario debe introducir además de la muestra en el equipo, el identificador de la misma, tal y como se muestra en la Figura 4.8.



Figura 4.8. Aplicación de control de los equipos desarrollados.

Dado que en ocasiones este identificador se ha introducido con erratas, un proceso automático de casación no sería capaz de llevarlo a cabo de forma correcta, dado que no encontraría todas las evidencias. Por ello, se optó por un sistema semi-automático que precise de la validación del usuario para operar.

Las muestras almacenadas en el servidor pueden ser visualizadas al pulsar en la opción “Leer Muestra”.

Espectrofotómetro Home David Carreres@hapt.es

Mediciones
 Ver lista de mediciones

Monitorización
 Ver evolución por equipos

Alertas

Anotaciones
 Gestión de anotaciones
 Lista de anotaciones

Equipos
 Lista de equipos
 Agregar nuevo equipo

Usuarios
 Lista de usuarios
 Agregar nuevo usuario

Detalles: 02/03/20 Tratada-E2 (ID: 496)

Operaciones

DETALLES

ID Muestra: 02/03/20 Tratada-E2 Equipo: (2) Equipo 2

Fecha: 02/03/2020 Hora: 08:23:18

Comentario:
 Añade un comentario sobre la muestra

Guardar Detalles de Muestra

GRÁFICAS

Transmitancia Absorbancia

Transmitancia
 Ganancia: 500

PARÁMETROS ANALÍTICOS

Agua

DQO: 64 DBO5: 10 COT: Número SST: 28 SSV: Número

NKT: 0 NH4: Número P: 2.2 SS: Número PH: 7.63

OD: Número SD: Número HS: Número S2: Número H2S: Número

Cond: 2280 S.seg: Número NT: 53 N-NO3: 4.8 PO4: Número

SO42: Número Metales Pesados: Número Hidrocarburos totales: Número Turbidez: Número

Biológicos

MLSS: 0 MLVSS: 0 V30: 0 IVF: 0 T: Número

Cond: Número

Guardar Parámetros Analíticos

ESTIMACION DE CARGA CONTAMINANTE

Demanda Química de Oxígeno (DQO) Demanda Bioquímica de Oxígeno 5 días (DBO5)

Referencia (Lab): Referencia (Lab): DQO: 64

Algoritmo Genético

Sólidos Totales en Suspensión (SST) Fósforo (P)

Referencia (Lab)

Algoritmo Genético

Nitrógeno Total (NT) Nitrógeno Nitrato (NO3-N)

Referencia (Lab)

Algoritmo Genético

FICHEROS ADJUNTOS

Informe Laboratorio.pdf Punto de registro PNG

SUBIR FICHEROS

Seleccionar archivo Ningún archivo seleccionado

Subir Fichero

Copyright © David Carreres 2021

Figura 4.9. Vista en detalle de una muestra contenida en el servidor.

Tal y como se muestra en la Figura 4.9, la vista en detalle de las muestras se divide en las siguientes secciones:

- **Detalles**
Comprende el ID o nombre la muestra dado por la planta de tratamiento, así como la fecha y hora y el equipo encargado de realizar ese estudio. Además incluye un cuadro de comentarios para agregar anotaciones sobre la muestra en caso de ser necesario.
- **Gráficas**
Permite visualizar la respuesta espectral de las muestras obtenida por los equipos desarrollados, mostrando tanto la curva de transmitancia como la de absorbancia, esta última generada de forma automática a partir de la primera.
El usuario puede visualizar el valor de cada uno de los puntos de la gráfica con sólo colocar el cursor sobre ella, siendo posible visualizar múltiples gráficas de forma simultánea, aspecto de gran relevancia para hacer una comparativa de la respuesta espectral para distintos valores de ganancia.
- **Parámetros Analíticos**
Contiene los valores de caracterización de las muestras llevados a cabo en los laboratorios de las plantas de tratamiento, los cuales son subidos por medio de la aplicación mostrada en la Figura 4.7.
Dado que el objetivo de los equipos desarrollados es estimar estos valores de carga contaminante a partir de la respuesta espectral, valores medidos en laboratorio sólo se encuentran disponibles en las muestras utilizadas para la generación de los modelos de estimación que se describe en el Capítulo 7.
- **Estimación carga contaminante**
A partir de la respuesta espectral y mediante el uso de los modelos de estimación (Capítulo 8), el servidor proporciona una caracterización de las muestras de forma automática, así como una comparativa con los valores medidos en laboratorio (Parámetros analíticos) en caso de estar disponibles.
- **Ficheros adjuntos**
La web permite la subida y almacenamiento de cualquier otro tipo de archivo, como ficheros PDF o imágenes que permite completar la información de las muestras.

La información mostrada en esta sección es suministrada por un servicio web tipo GET que recibe como único parámetro el ID de la muestra que se desea visualizar. En la Figura 4.10 se muestra el servicio desarrollado para este fin así como la respuesta de la muestra recogida en la Figura 4.9.

A lo largo del transcurso de la presente tesis doctoral, el número de longitudes de onda soportadas por los equipos ha ido variando a medida que se profundizaba en el desarrollo. Debido a esto, la información sobre la respuesta espectral analiza por los equipos es almacenada en el servidor como una variable tipo JSON, tal y como se muestra en la figura

4.10. Esto confiere flexibilidad al sistema para soportar cualquier rango de trabajo que posea el equipo, sin necesidad de realizar cambios en su programación.

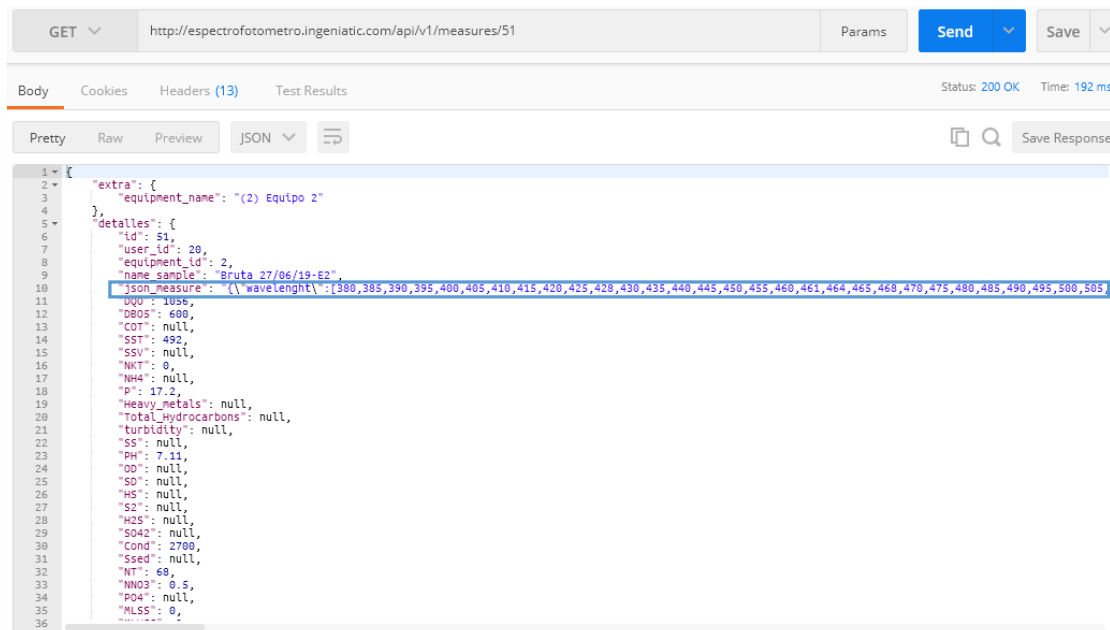


Figura 4.10. Servicio web para la recuperación de los datos de una muestra en Postman.

Entre las secciones antes expuestas, la más relevante es la estimación de la carga contaminante a partir de la respuesta espectral de las muestras. Para llevar a cabo esta estimación, a lo largo del presente trabajo de investigación se han desarrollado diferentes modelos de cálculo basados en técnicas distintas:

- **Modelos de regresión multivariable:** Válidos para la estimación de DQO, DBO₅ y SST en muestras de agua bruta.
- **Modelos basados en algoritmo genético:** Capaces de estimar con precisión los valores de DQO, DBO₅, SST, P, NT, NO₃-N, siendo válido tanto para muestras de agua bruta como tratada.

Gracias a estos modelos, la web proporciona al usuario una estimación muy precisa de la carga contaminante de las muestras a analizadas con cualquiera de los equipos de forma instantánea. Esto es de vital importancia si tenemos en cuenta que la obtención de este tipo de parámetros en laboratorio está supeditada a dilatados periodos de análisis con diferentes técnicas, equipos y procesos, lo que imposibilita llevar a cabo un estudio en tiempo real y; por ende; limita la velocidad de respuesta del operador ante eventualidades como valores anormales de contaminación a la entrada de la planta.

Como se puede observar en la Figura 4.11, el servidor no sólo proporciona una caracterización de las muestras desde un punto de vista cuantitativo, sino que realiza una comparativa con respecto a los valores medidos en laboratorio (cuando esta información se encuentra disponible), así con respecto a los distintos modelos desarrollados.

Tal y como será mostrado en posteriores capítulos, los modelos generados durante el transcurso de la tesis doctoral, permiten obtener una alta precisión sin necesidad de someter

a las muestras a ninguna manipulación, de ahí la utilidad de este sistema. Dado que los modelos de regresión lineal desarrollados sólo son válidos para muestras de agua bruta, y únicamente para el cálculo de la DQO, DBO₅ y SST, en la Figura 4.11 se muestra una comparativa de los modelos basados en algoritmo genético (válido para todas las aguas y todos los parámetros) y regresión multivariable, con respecto al valor de referencia en una muestra de agua bruta.

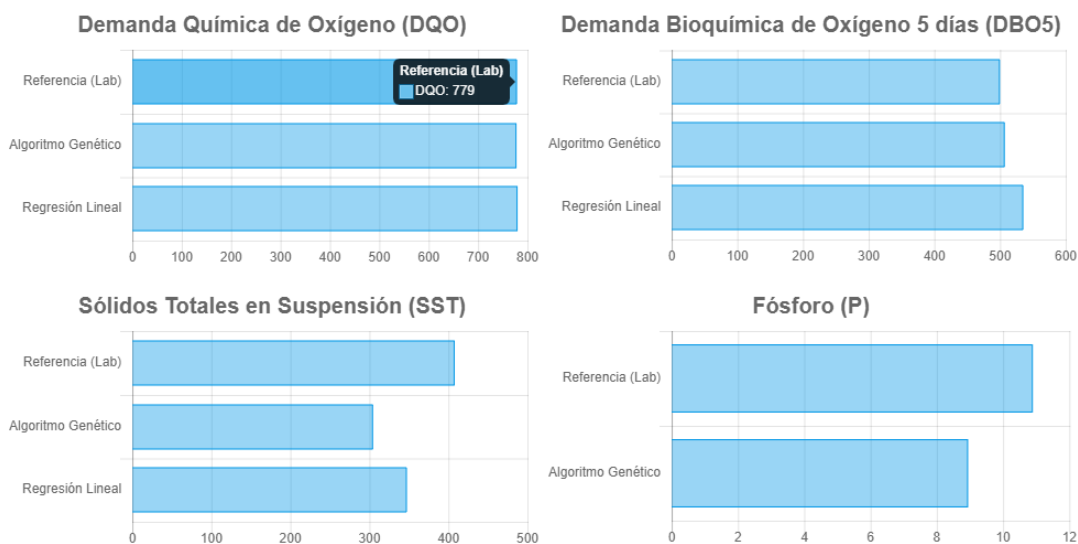


Figura 4.11. Estimación de parámetros contaminantes de una muestra de agua bruta almacenada en el servidor, acompañada de una comparativa entre modelos y los valores medidos en laboratorio.

4.5 Evolución de la carga contaminante

Además de poder conocer la carga contaminante de una muestra, también es relevante poder saber cómo evoluciona en cada punto concreto de la red.

La variación de los parámetros a lo largo del tiempo proporciona una comprensión más detallada de qué sucede en la red de saneamiento, posibilitando una acción mejor coordinada, tanto para ajustar los niveles de depuración de la planta de tratamiento, como para detectar vertidos no autorizados al medio, o reducir los efectos de los desbordamientos durante episodios de lluvia.

Con el fin de servir como una herramienta de detección temprana, se desarrolló la sección denominada “Monitorización”, la cual consta de dos subsecciones claramente diferenciadas:

1. Ver evolución por equipos

Los equipos, tanto en su versión Desktop como In-Sewer, permiten obtener información de la evolución de la carga contaminante en el punto en el que se instalan.

Los sistemas tradicionales de caracterización no permiten tener un conocimiento de cómo evoluciona la contaminación en tiempo real, por lo que

disponer de una gráfica que muestre la variación de los distintos parámetros a lo largo del tiempo, es clave para una respuesta mejor coordinada.

2. Alertas

Más allá de conocer la carga contaminante de las muestras desde un punto de vista cuantitativo, que puede ser más o menos preciso con respecto a los valores obtenidos en laboratorio, lo realmente relevante es saber si un cierto parámetro es o no superior a los máximos exigidos por la administración, es decir, disponer de una información de tipo cuantitativo.

Por ese motivo, el sistema muestra una lista de las anomalías detectadas en las muestras con una clasificación de color en función de la gravedad de las mismas:

- Amarillo: Desviación inferior al 10%.
- Naranja: Desviación inferior al 30%
- Rojo: Desviación superior al 30%

Estos umbrales han sido definidos en base a la propia experiencia de las plantas de tratamiento.

Esto permite al usuario conocer de una forma sencilla la existencia de anomalías como la existencia de vertidos no autorizados, permitiendo una respuesta mucha más rápida y mejor coordinada a la hora de abordar esta problemática.

Con el fin de mostrar con mayor claridad las peculiaridades de esta sección, trataremos de forma separada cada uno de los dos subsecciones antes descritas.

La Figura 4.12 muestra la subsección “Ver evolución por equipo”. En esta se recogen la lista de dispositivos a los que el usuario tiene acceso, donde se permite visualizar la designación del equipo, el tiempo, la fecha y hora de su creación, así como tres opciones:

1. Evolución

Permite visualizar la variación temporal de los distintos contaminantes a lo largo del tiempo.

2. Editar

Posibilita la modificación de los parámetros que rigen el equipo así como su asignación a usuarios (Figura 4.13)

3. Eliminar

Espectrofotómetro Home daniel.carreres@upct.es

Mediciones
Ver lista de mediciones

Monitorización
Ver evolucion por equipos

Alertas

Anotaciones
Gestión de anotaciones
Lista de anotaciones

Equipos
Lista de equipos
Agregar nuevo equipo

Usuarios
Lista de usuarios
Agregar nuevo usuario

Lista de equipos

ID	Nombre	Tipo	Fecha	Hora	
1	Equipo 1	Desktop	27/12/2018	15:41:35	Evolución Editar Eliminar
2	Equipo 2	Desktop	27/12/2018	15:41:55	Evolución Editar Eliminar
3	Equipo 3	Desktop	27/12/2018	15:42:01	Evolución Editar Eliminar
4	Equipo 4	Desktop	27/12/2018	15:42:05	Evolución Editar Eliminar
5	Equipo 5-DINTEL	Desktop	02/05/2019	21:11:55	Evolución Editar Eliminar
6	Espectrofotometro RGB	Desconocido	13/10/2019	22:21:27	Evolución Editar Eliminar
7	FAKE EQUIPMENT	Sewer	10/10/2020	01:08:04	Evolución Editar Eliminar
8	Desktop 6	Desktop	07/11/2020	01:20:13	Evolución Editar Eliminar
9	Espectro In-Sewer 1	Sewer	06/01/2021	03:02:17	Evolución Editar Eliminar

Copyright © Daniel Carreres 2021

Figura 4.12. Lista de equipos asociados al usuario sobre los que se puede consultar la evolución de la carga contaminante en el tiempo.

Espectrofotómetro Home daniel.carreres@upct.es

Mediciones
Ver lista de mediciones

Monitorización
Ver evolucion por equipos

Alertas

Anotaciones
Gestión de anotaciones
Lista de anotaciones

Equipos
Lista de equipos
Agregar nuevo equipo

Usuarios
Lista de usuarios
Agregar nuevo usuario

Desktop 6 (ID: 8)

Nombre del equipo: ✓

Tipo de equipo: ✓

Estado del equipo: Activo

Observación:

Asignación de usuarios:

15 elementos

Buscar aquí

- Daniel Carreres Prieto (UPCT)
- Juan Tomás García Bermejo...
- Juan Suardíaz Muro (UPCT)
- Fernando Cerdán Cartagen...
- Depuradora 1 (HIDROGEA)
- Depuradora 2 (HIDROGEA)
- Depuradora 3 (HIDROGEA)

1 elemento

Buscar aquí

- EDAR Munuera (OTROS)

[EDITAR EQUIPO](#)

Copyright © Daniel Carreres 2021

Figura 4.13. Interfaz para la edición de equipos.

La Figura 4.14 muestra cómo ha evolucionado la carga contaminante a lo largo del tiempo para el equipo designado como “Desktop 2”. Cada una de las gráficas se basa en los valores estimados por el sistema a partir de la respuesta espectral de cada muestra, la cual se acompaña (cuando está disponible) de los valores obtenidos de forma analítica en laboratorio.

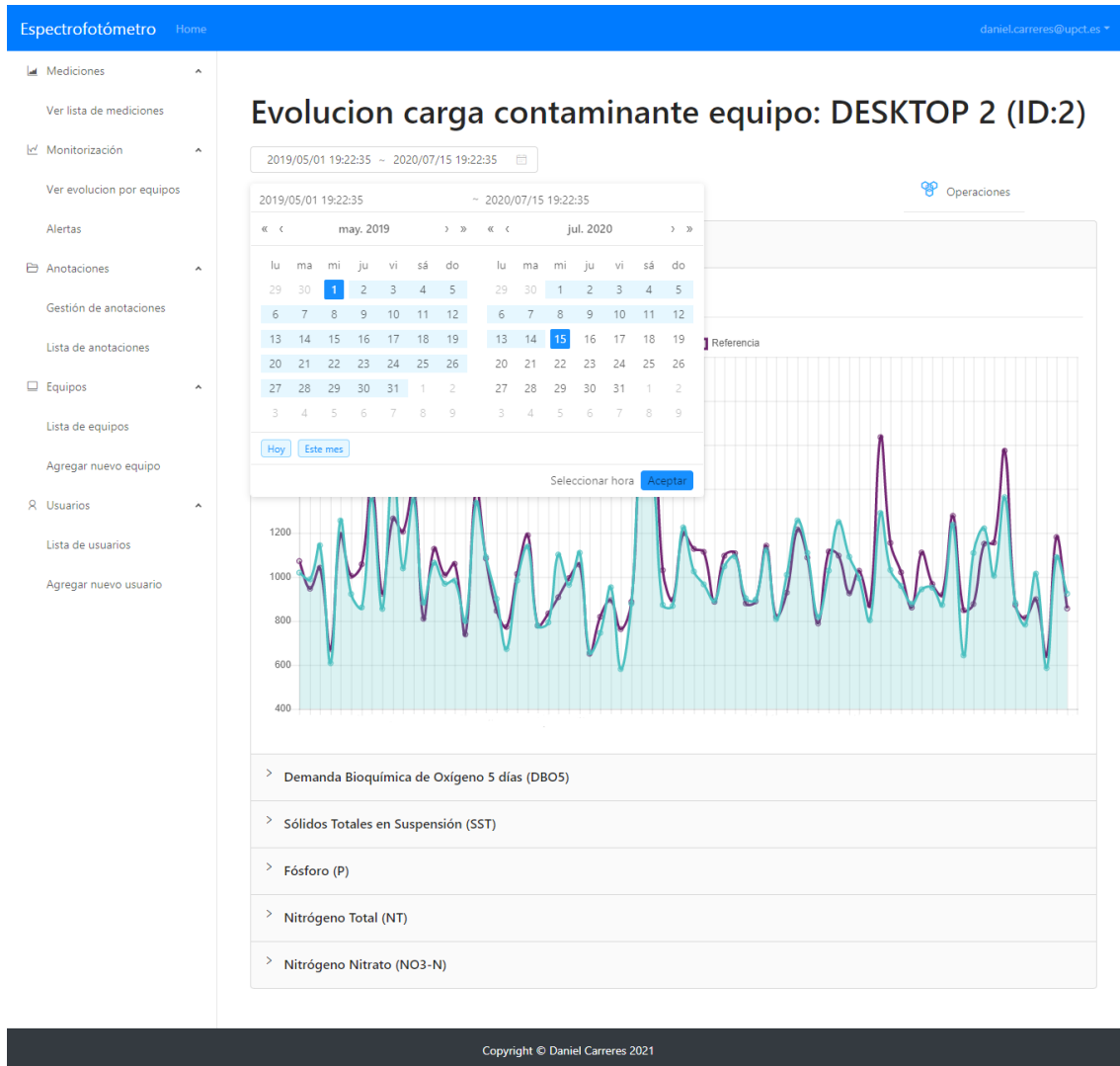


Figura 4.14. Evolución de la carga contaminante por equipo.

Dado que los equipos de sobremesa han sido utilizados para analizar tanto muestras de agua bruta como agua tratada, la web realiza una discriminación entre ambos tipos de muestras en base la información contenida en su designación (nombre de la muestra) para generar dos gráficas separadas en dos pestañas.

El usuario puede seleccionar el intervalo de visualización por medio del panel situado en la parte superior de la interfaz.

El otro elemento crucial del sistema desarrollado durante el presente trabajo de investigación es el conocer, desde un punto de vista cualitativo, la existencia de muestras que no superan con los niveles máximos establecidos por la legislación española.

Para ello, la subsección denominada “Alertas”, permite, tal como se observa en la Figura 4.15, mostrar el listado de muestras cuya carga contaminante ha superado los umbrales establecidos.

Lista de alertas

ID	Alerta	Muestra	Equipo
1	SST 1032mg/l	17/03/20 Bruta-E2 (ID:539)	(2) Equipo 2
2	SST 844mg/l	16/03/20 Bruta-E2 (ID:534)	(2) Equipo 2
3	DQO 1460mg/l	03/12/19 Bruta-E2 (ID:380)	(2) Equipo 2
4	SST 892mg/l	19/11/19 Bruta-E2 (ID:352)	(2) Equipo 2
5	DQO 1636mg/l	29/10/19 Bruta-E2 (ID:326)	(2) Equipo 2
6	SST 956mg/l	07/10/19 Bruta-E2 (ID:284)	(2) Equipo 2
7	SST 768mg/l	01/08/19 Bruta-E2 (ID:114)	(2) Equipo 2
8	SST 1404mg/l	26/07/19 Bruta-E2 (ID:131)	(2) Equipo 2
9	DQO 740mg/l	02/12/19 Bruta-E22 (ID:375)	(2) Equipo 2
10	SST 728mg/l	18/10/19 Bruta-E2 (ID:310)	(2) Equipo 2
11	DQO 1266mg/l	25/09/19 Bruta-E2 (ID:270)	(2) Equipo 2
12	SST 744mg/l	27/08/19 Bruta-E2 (ID:215)	(2) Equipo 2
13	SST 780mg/l	26/08/19 Bruta-E2 (ID:210)	(2) Equipo 2
14	DQO 720mg/l	18/11/19 Bruta-E2 (ID:347)	(2) Equipo 2
15	DBOS 720mg/l	29/08/19 Bruta-E2 (ID:218)	(2) Equipo 2
16	DBOS 720mg/l	27/08/19 Bruta-E2 (ID:215)	(2) Equipo 2
17	SST 616mg/l	24/08/19 Bruta-E2 (ID:616)	(2) Equipo 2
18	SST 536g/l	22/08/19 Bruta-E2 (ID:202)	(2) Equipo 2
19	SST 588mg/l	19/09/19 Bruta-E2 (ID:259)	(2) Equipo 2
20	SST 512mg/l	07/08/19 Bruta-E2 (ID:176)	(2) Equipo 2

Copyright © Daniel Carreres 2021

Figura 4.15. Lista de alertas detectadas.

Cada uno de los registros indica cual es el motivo de la alerta, la muestra a la que pertenece así como el equipo que ha realizado el análisis, aspecto crucial para detectar posibles fallos en el proceso.

Además, se muestra un código de color en función de en qué porcentaje ha superado el valor máximo fijado por la administración para ese tipo de agua. Tal y como se puede observar, las alertas son ordenadas de mayor a menor gravedad y por fecha.

El usuario puede acceder a la muestra que ha dado lugar a la alerta, así como cambiar la categoría de la alerta o incluso eliminarla.

4.6 Aplicación de escritorio

Una vez descrito los aspectos más relevantes de la web, resta abordar las principales características de la aplicación desarrollada para la gestión de los equipos desde el PC a través de USB, en especial la versión Desktop, aunque también es posible el manejo de la versión In-Sewer.

Aunque en la Figura 4.8 se ha mostrado una vista de la aplicación durante el proceso de medición, en esta sección se describirá con mayor detalle el resto de elementos que la constituyen.

La aplicación ha sido desarrollada en JAVA con el fin de garantizar su uso en cualquier sistema operativo [161]. Además, con el fin de facilitar su despliegue, la aplicación no precisa de instalación, siendo un autoejecutable.

En un inicio, el único cometido de la aplicación era visualizar la respuesta espectral de las muestras analizadas por los equipos y el envío de la información al servidor. Para ello, el usuario debe introducir sus credenciales (que son los mismos que para acceder a la web) lo que permite llevar a cabo una trazabilidad completa de las muestras. Sin embargo, a lo largo del desarrollo sus funcionalidades se han visto incrementadas.

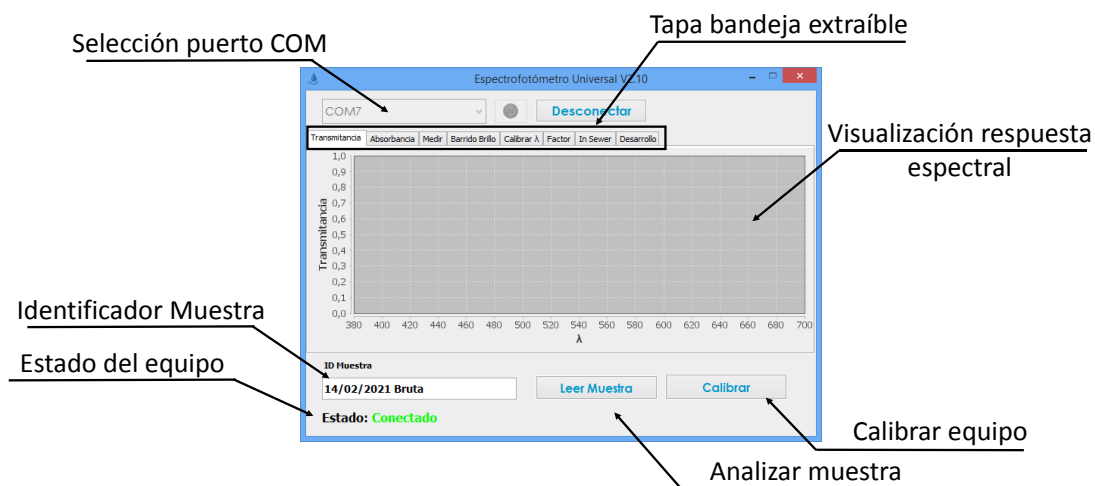


Figura 4.16. Partes de la aplicación para PC para el control de los equipos.

En la Figura 4.16, se muestra las distintas partes que componen la pantalla principal. En su parte superior se encuentra la opción de seleccionar el puerto comunicación del equipo, así como distintas pestañas:

1. Transmitancia/Absorbancia

Ambas permiten visualizar la curva de absorbancia y transmitancia generada durante el análisis llevado a cabo por el equipo. También permite exportar la imagen así como mostrar una vista en detalle de una determinada zona.

2. Medir

Permite obtener el valor de transmitancia y absorbancia a una determinada longitud de onda concreta.

3. Barrido Brillo

Es utilizado durante la fase de calibración del equipo. Dado que no es adecuado seleccionar un nivel de brillo (corriente aplicada a cada LED) común a todos ellos, esta sección permite definir los niveles más adecuados para cada diodo instalado en el equipo.

4. Calibrar λ /Factor

Estas opciones han sido utilizadas durante la fase de generación de los modelos para equiparar la respuesta espectral del equipo con respecto a los valores obtenidos mediante un equipo basado en lámparas incandescente. Estos modelos son descritos en los Capítulos 5 y 6.

5. In-Sewer/Desarrollo

Proporciona ciertas funcionalidades específicas para el equipo In-Sewer (Figura 4.17), así como otras aplicables también a la versión de escritorio, como es la definición del identificador único del equipo (Pestaña Desarrollo).

En la pestaña “In-Sewer” se posibilita el control de todos los aspectos de este tipo de equipo, pudiendo activar cada uno de los elementos, como puede ser la bomba de agua, así como la cantidad de muestra adquirida o el control del sistema de selección de longitudes de onda.

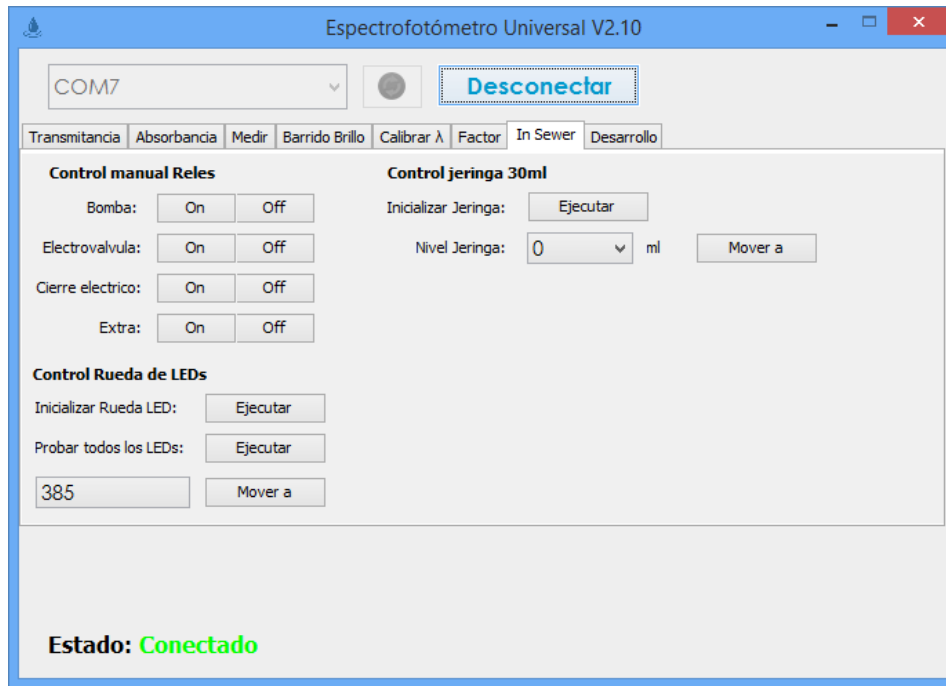


Figura 4.17. Pestaña “In-Sewer” de la aplicación para PC.

4.7 Conclusiones

El servidor descrito en el presente capítulo representa, no sólo un complemento indispensable a los equipos desarrollados, sino que proporciona información de gran utilidad para una adecuada gestión de la contaminación de las aguas.

Su función va más allá del simple almacenamiento de las muestras analizadas por los distintos equipos, sino que proporciona una interfaz integral para la gestión de los distintos puntos de la red de una forma simple y rápida.

Al estar los modelos de estimación en la web en lugar de en los equipos, se hace mucho más sencillo poder actualizarlos o introducir otros nuevos basados en diferentes procesos de análisis, aspecto difícil de gestionar si estos se encontraran instalados en los propios equipos, dado que obligaría la actualización periódica de los mismos.

El sistema alarmas implementado, permite conocer en todo momento la existencia de valores anómalos en los niveles de la carga contaminante, aspecto de gran utilidad a la hora de llevar a cabo la gestión de la red.

En definitiva, todos los desarrollos mostrados en el presente capítulo juegan un papel fundamental en la lucha contra los vertidos no autorizados y para lograr una respuesta temprana.

Capítulo 5

Calibración de LED de ancho espectral fijo

5.1 Introducción

Los equipos descritos en capítulos anteriores basan su funcionamiento en el uso de tecnología LED como fuente para la generación del espectro de trabajo. No obstante, la respuesta espectral que se obtiene con este tipo de fuente de luz, difiere de la que se obtendría con el uso de lámparas incandescentes y redes de difracción, por ello es necesario un proceso de calibración para corregir estas desviaciones.

En el presente capítulo se describirá de forma pormenorizada los modelos de cálculo generados para este fin, los cuales a su vez permiten “aislar” una longitud de onda concreta de las múltiples emitidas por los diodos, aspecto clave del sistema de selección de longitudes de onda descrita en los Capítulos 2 y 3 y que se abordará en el presente capítulo.

Para ilustrar esta necesidad de generar modelos de calibración, consideraremos el LED VAOL-5EUV8T4 [162], que posee un pico en la longitud de onda de 385nm y cuyas características técnicas más relevantes extraídas de las especificaciones del fabricante se recogen en la Tabla 5.1.

Tabla 5.1. Características LED VAOL-5EUV8T4. [162]

Parámetro	Valor		
Tensión reversa	5V		
Corriente media	30mA		
Intensidad lumínica máxima	100 mcd		
Longitud de onda característica	420nm		
Longitud de onda pico	Min	Med	Max
	380	385	-

Donde la longitud de onda característica, es aquella detectada por el ojo humano, y por lo tanto es un valor subjetivo y por ende carece de relevancia en la aplicación que nos ocupa. Por su parte, la longitud de onda pico alude a aquella emitida con mayor intensidad.

De esto último se deduce que si existe una longitud de onda máxima o pico, significa que debe de existir otras longitudes de onda que son emitidas en menor nivel de intensidad. Esto constituye la respuesta espectral del LED, la cual se recoge en la Figura 5.1.

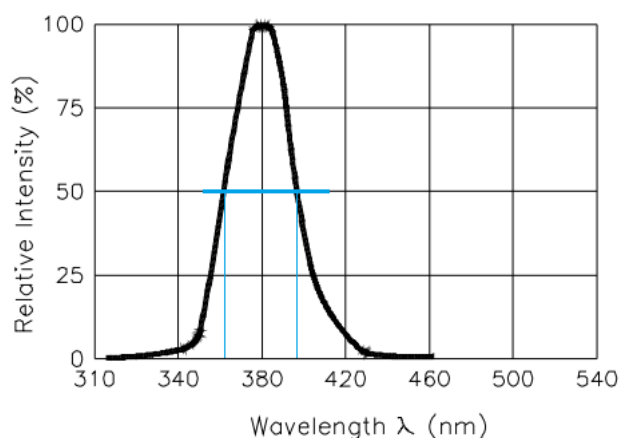


Figura 5.1. Espectro de emisión del LED VAOL-5EUV8T4 extraído del datasheet del fabricante. [162]

Como se puede observar, el LED no emite una única longitud de onda, sino que por el contrario, este emite en múltiples longitudes de onda con diferentes niveles de intensidad siguiendo una distribución normal.

No obstante, las longitudes de onda “útiles”, se reducen a las comprendidas en el “Ancho Espectral” del LED [103], que se definen como aquellas cuyo nivel de intensidad está por encima del 50%, donde en el caso de la Figura 5.1, comprenden entre los 360 y 400nm.

Con el fin de mostrar el efecto de estas longitudes de onda que son emitidas de forma simultánea en la respuesta espectral, la Figura 5.2 muestra los valores de transmitancia medidos tanto con el LED [162] (gráfica naranja y azul) a diferentes niveles de brillo (determinado por la corriente suministrada, aspecto que se describirá de forma más extensa a lo largo del capítulo) y la respuesta obtenida con el equipo comercial V-5000 Vis a 385nm [104] (gráfica de color negro) sobre diferentes muestras descritas en la Tabla A1 del Anexo A.

Como se puede observar en la Figura 5.2, a valores bajos de transmitancia, los resultados proporcionados por el LED son muy similar al obtenido con el equipo comercial. No obstante, las discrepancias entre ambas fuentes de luz se acentúan cuanto mayor es el valor de transmitancia de la muestra, siendo los valores registrados por el LED un 14-20% más altos que los proporcionados por el equipo comercial, especialmente cuanto mayor es el nivel de brillo.

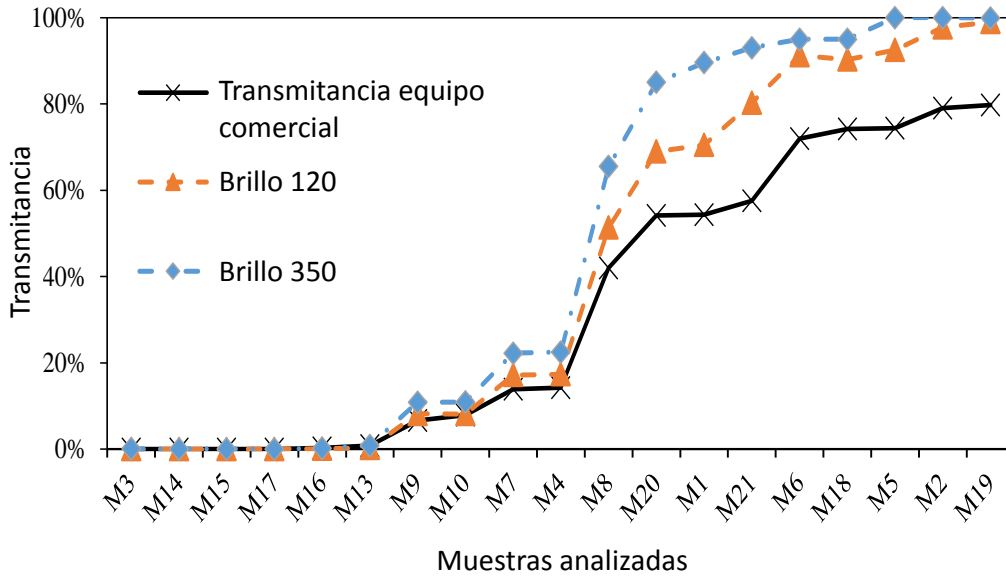


Figura 5.2. Comparación de los niveles de transmitancia a 385nm medidos con un equipo comercial basado en lámparas incandescentes y el LED VAOL-5EUV8T4 con dos brillos diferentes. [162]

Para comprender la causa de esta desviación, debemos tener en cuenta cómo funciona un equipo comercial basado en lámpara: En un equipo comercial, gracias a los elementos ópticos que incorpora, se garantiza que una única (385nm) o reducido número de longitudes de onda atraviesen la muestra, por lo que el valor de transmitancia registrado corresponde al efecto que tiene esa única longitud de onda, sin embargo, esto no es así en el LED. A pesar de que la longitud de onda pico del LED sea 385nm, en realidad lo que se mide es el efecto, no sólo de los 385nm, sino también de todas las demás longitudes de onda que están siendo emitidas, principalmente las correspondientes al ancho espectral, lo cual hace que los valores de transmitancia sean mayores a los esperados.

Otro aspecto a tener en cuenta es el nivel de intensidad a la que emite el LED. Este nivel de brillo puede determinarse mediante fotómetros [163], pero también es posible encontrar un orden de magnitud en la cantidad de corriente que recibe el LED [164, 165], el cual se tratará en detalle a lo largo del presente capítulo.

Cuando el LED de ancho espectral fijo se enciende con un cierto nivel de brillo, este emite su longitud de onda pico a ese valor máximo de intensidad, pero también emite otras longitudes de onda a menor nivel (Figura 5.1). Si se duplica el nivel de brillo (corriente), ese efecto multiplicador se verá sobre todas las longitudes de onda, por lo que al aumentar el brillo, mayor cantidad de luz logrará pasar a través de la muestra y por lo tanto mayor será el valor de transmitancia registrado. Sin embargo, si el nivel de brillo es muy bajo, el efecto de las longitudes de onda adicionales se verá atenuado, haciendo que los resultados se aproximen cada vez más a los que se obtendrían si el LED sólo emitiera a su longitud de onda pico: 385nm.

Esto podría llevarnos a pensar, que sería preferible establecer un nivel de brillo muy bajo para corregir esta desviación en los valores de transmitancia, sin embargo esta plantea otro problema: Si el nivel de brillo es demasiado bajo, la cantidad de luz que logrará atravesar

una muestra de blanco (agua destilada) será bajo, por lo que no estaríamos aprovechando todo el rango de detección del sensor (fotodiodo), lo cual restaría precisión en las mediciones si estas son demasiado turbias, es decir, si tiene valores de transmitancia bajos. Si por el contrario, el nivel de brillo es demasiado alto, provocaría la saturación del sensor, haciendo imposible la lectura.

Gracias a las investigaciones llevadas a cabo [29], se ha desarrollado un modelo que no sólo permite corregir el efecto de las desviaciones ocasionadas por las diferentes longitudes de onda, sino que también permite “seleccionar” una longitud de onda concreta dentro del ancho espectral de LED a partir de la atenuación o amplificación del valor de transmitancia medido en función de la distancia a la longitud de onda pico, y todo ellos sin necesidad de utilizar elementos ópticos como monocromadores o lentes de difracción.

5.2 Calibración de LED de ancho espectral fijo

Los LED de ancho espectral fijo están diseñados para emitir en ciertas regiones del espectro electromagnético, principalmente del espectro visible, el cercano al ultravioleta y el cercano al infrarrojo, regiones clave para la caracterización de la materia orgánica [166] e inorgánica [167, 168]. Aunque es posible encontrar LED que emitan en la región infrarroja y ultravioleta, estos presentan mayor un coste económico que los que trabajan en el espectro visible y además, el número de longitudes de onda pico disponibles en el mercado es muy limitado.

Por ello, el estudio se ha centrado en diodos que operan dentro del espectro visible, los cuales se recogen en la Tabla A2 del Anexo A.

5.2.1 Selección del nivel de brillo

Tal y como se ha comentado, el nivel de brillo de cada LED es determinante a la hora de realizar el análisis de las muestras.

Cada LED tiene diferentes capacidades lumínicas (Intensidad Luminosa) determinadas por su proceso de fabricación, de tal manera que dos LEDs diferentes, alimentados por la misma tensión y corriente pueden emitir a diferentes intensidades de brillo.

En capítulos anteriores, se ha descrito el control de los LED (Rueda de LED) desde un punto de vista electrónico. Aunque la circuitería utilizada para su control difiera entre el Espectrofotómetro Desktop (TLC5940RHBR [149]) y el Espectrofotómetro In-Sewer (PCA9685PW [150]), el principio de funcionamiento es el mismo: Generar una señal cuadrada de ancho variable (PWM), con una resolución de 16bits (es decir un número entre 0 y 4095) y una corriente máxima suministrada a la carga de 40mA a un valor numérico de 4095.

Gracias al software desarrollado, el cual ha sido descrito en el Capítulo 4, es posible calcular la variación de los valores de transmitancia obtenidos sobre distintas muestras a diferentes niveles de brillo (corriente suministrada a cada LED) de forma automática. En la Figura 5.3, se muestran los valores de transmitancia obtenidos para una longitud de onda concreta (470nm) sobre diferentes muestras, a distintos niveles de corriente aplicada al LED de trabajo (brillo), siendo este valor de transmitancia calculado en base a un valor de

referencia medido a cada uno de esos niveles de corriente (brillo), el cual se recoge como “Blanco” en la Figura 5.3.

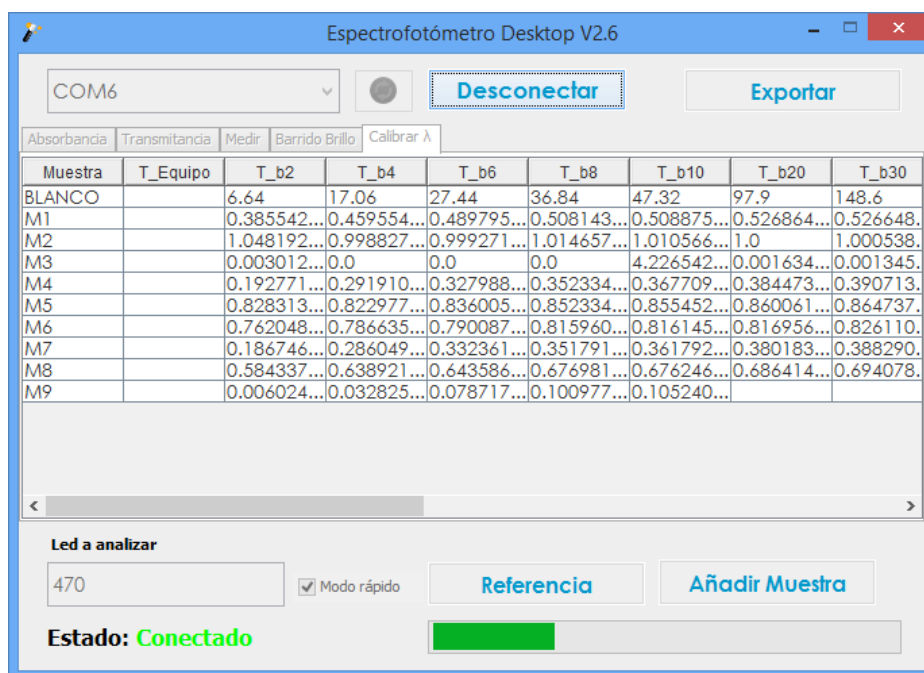


Figura 5.3. Vista del software desarrollado para la gestión de los equipos descritos en los Capítulos 2 y 3, durante el proceso de obtención de los valores de transmitancia para distintos niveles de brillo.

Teniendo en cuenta que la transmitancia es $T = \frac{I}{I_0}$ [53], donde I es la cantidad de luz que atraviesa la muestra (que incide sobre el fotodiodo) e I_0 es la cantidad de luz que incide sobre la muestra, la cual puede ser considerada como la cantidad de luz que atraviesa una muestra de agua destilada, se hace necesario determinar el valor de I_0 para cada nivel de brillo. Una vez determinado los valores de referencia, el software a través del Espectrofotómetro Desktop, calcula de forma automática los valores de transmitancia de las muestras introducidas de forma manual en el equipo.

Si consideramos por ejemplo la muestra de colorante alimentario Rojo 40 (Rojo allura AC) [169] diluida al 50% recogida en la Tabla A1 Anexo A bajo la designación M1, a diferentes niveles de brillo y analizada con un LED de una determinada longitud de onda de pico (por ejemplo, 385nm), se observa, tal y como se muestra en la Figura 5.4, que los valores de transmitancia se mantienen prácticamente constantes entre el 68% y el 70%, lo que corresponde al nivel de brillo obtenido mediante una corriente de 0,073 a 2,051mA. Este valor, aunque estable, es superior al valor de referencia medido con el equipo comercial, 54,35%.

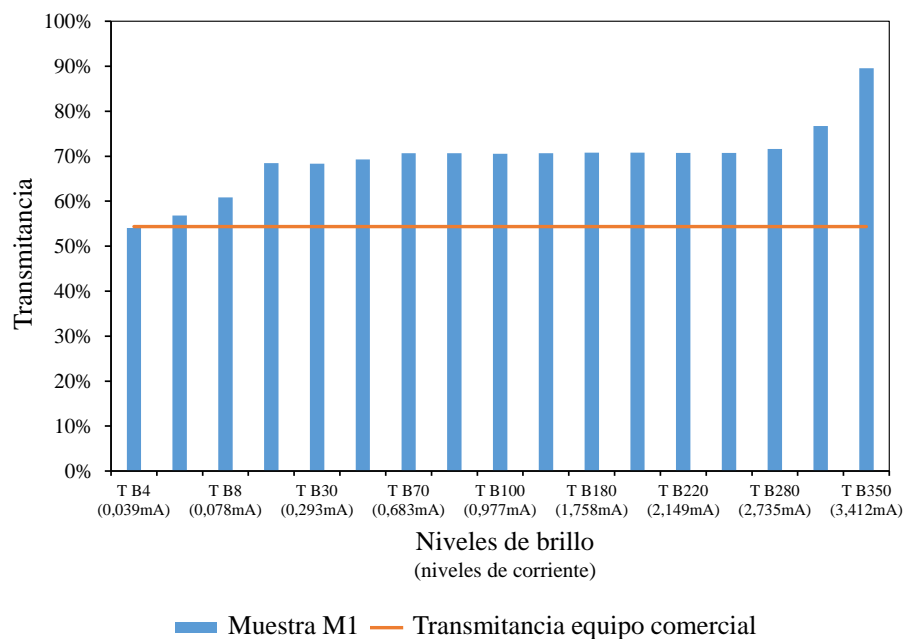


Figura 5.4. Efecto del nivel de brillo en los valores de transmitancia registrados en la muestra M1 por medio de un LED de 385nm de longitud de onda pico.

Esto plantea la pregunta de cuál es el nivel de brillo (corriente) que debe escogerse. En valores bajos de luminosidad (baja corriente), el efecto de las longitudes de onda adicionales es despreciable. Esto se observa con claridad para el nivel de brillo designado como T B4, el cual correspondería a una corriente aplicada de $0,039mA \left(\frac{4 \cdot 40mA}{4095} = 0.039mA \right)$, donde el valor de la transmitancia fue del 54%, bastante cercano al esperado de 54,35%. Sin embargo, aunque los resultados son cercanos, se considera que un nivel de brillo tan bajo no es adecuado para las operaciones de espectrofotometría. Si la muestra tuviera una mayor turbidez, los fotones no podrían atravesarla. Por lo tanto, el uso de un nivel de corriente dentro de los valores estabilizados de transmitancia es la mejor opción, aunque los resultados sean mayores de lo esperado.

Pero si el nivel de brillo es excesivo, se produciría la saturación del sensor, perdiendo resolución. Por ello, aunque es posible seleccionar cualquier nivel de corriente inferior a aquella que provoca la saturación del sensor (cuando se introduce una muestra de agua destilada), las pruebas llevadas a cabo han concluido que el valor de corriente suministrada a cada LED (nivel de brillo), debe ser aquél que permita registrar un valor de luminosidad del orden del 50 al 70% de la resolución máxima del sensor.

En la Tabla A3 del Anexo A, se recoge tanto los niveles de brillo seleccionados para cada uno de los diodos utilizados (Tabla A2 Anexo A) así como el nivel de tensión registrado por el sensor (fotodiodo), donde en ningún caso se alcanza el valor de tensión de saturación de 5V del sensor.

5.2.2 Calibración de la longitud de onda pico

Las pruebas llevadas a cabo, han demostrado que los valores de transmitancia registrados con un LED de una determinada longitud de onda pico, y los valores obtenidos por un equipo comercial basado en lámparas incandescente a esa misma longitud de onda pico, guardan algún tipo de proporcionalidad. Esto se observa con claridad en la Figura 5.5.

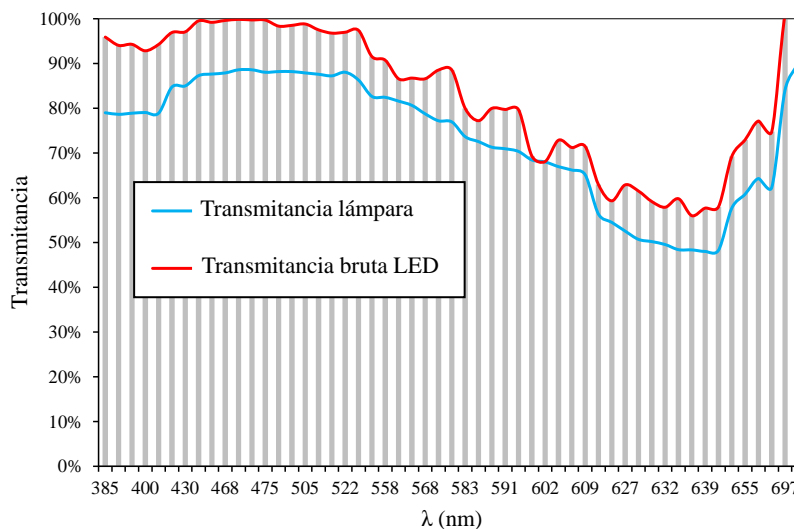


Figura 5.5. Comparación de valores de transmitancia entre 385 y 700nm, medidos con 59 LEDs de diferentes longitudes de onda pico (rojo) y con un equipo comercial basado en lámparas incandescentes (azul).

En esta figura se ilustra el análisis realizado sobre una muestra M2 consistente en colorante alimentario azul (Tabla A1 Anexo A), a través de 59 LEDs con λ_{pico} comprendida entre 385 y 700nm. Cada uno de los LED, ha sido representado por medio de una línea vertical ubicándolo en su longitud de onda pico con el fin de clarificar cómo se ha conformado el espectro de trabajo y la imputación de cada parte del mismo, utilizando cada LED para modelizar una única longitud de onda, su longitud de onda pico.

Las pruebas llevadas a cabo para establecer el ajuste entre los valores de transmitancia medidos con el equipo comercial y con el equipo desarrollado, han concluido que la relación entre ambos resultados es de tipo lineal, tal y como se muestra en la Figura 5.6, consistente en un diagrama de dispersión donde se correlacionan los valores de transmitancia obtenidos con el equipo comercial con respecto los medidos con el LED de 470nm de pico, a un nivel de brillo 40, para 19 muestras diferentes (Tabla A1 Anexo A, donde las muestras de leche M11 y M12, no fueron utilizadas dado que toda la luz era absorbida, tanto con el LED como con el equipo comercial).

En todos los casos analizados, la bondad del ajuste ha sido superior al 98%, denotando que sí es posible aislar el efecto de la longitud de onda pico sin necesidad de utilizar elementos ópticos.

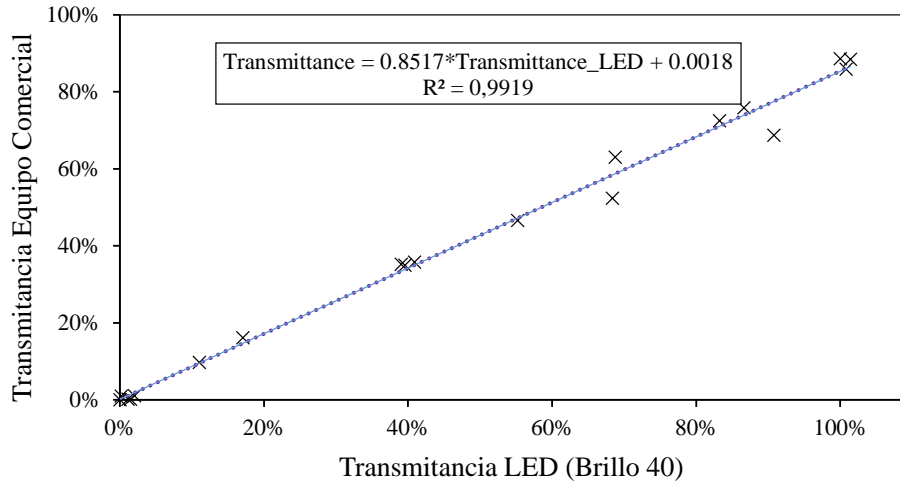


Figura 5.6. Diagrama de dispersión entre los valores de transmitancia medidos con un equipo basado en lámparas incandescentes y los obtenidos por un LED de 470nm de pico y a un nivel de brillo 40 sobre 19 muestras.

Esta corrección se muestra en la Figura 5.7, donde se recogen los valores de transmitancia medidos con el LED tras la aplicación del modelo mostrado en la Figura 5.6 (naranja) junto con los valores de referencia medidos con el equipo comercial (azul), siendo muy similares los valores obtenidos mediante ambos sistemas. Con el fin de facilitar la exposición de los resultados, las muestras se han ordenado de forma creciente, según los valores de transmitancia obtenidos con el equipo comercial.

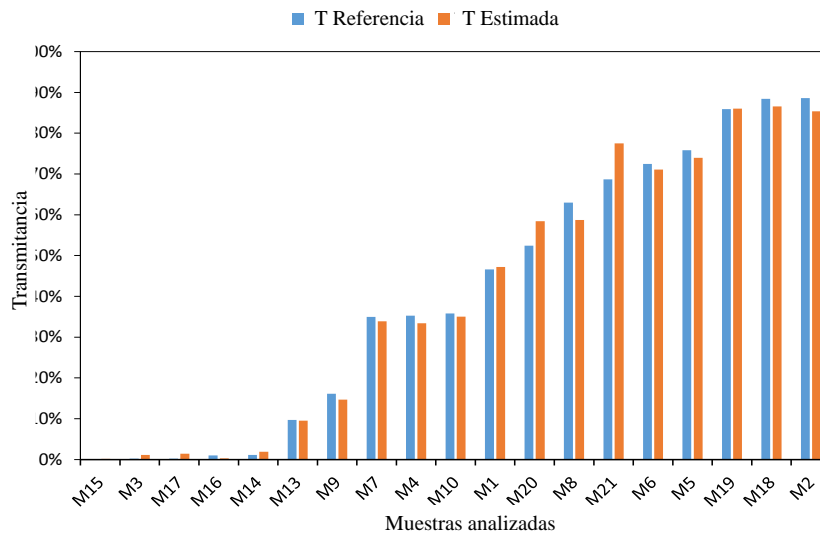


Figura 5.7. Comparación entre los valores de transmitancia a 470nm medidos con un equipo comercial y los obtenidos con un LED de 470nm de pico tras aplicar el modelo de calibración sobre 19 muestras.

Este proceso de correlación entre los valores de transmitancia medidos con cada uno de los 34 LEDs y los obtenidos con el equipo comercial basado en lámparas incandescentes se ha repetido para las distintas longitudes de onda pico recogidas en la Tabla A2 del Anexo A, cuyos diagramas de dispersión se recogen en el Anexo B. Dado que muchas longitudes

de onda pueden ser modelizadas por más de un LED, algunas tablas recogidas en el Anexo B poseen más de una ecuación de calibración, donde se ha resaltado en aquella que ha presentado el mejor ajuste, y por ende, la que ha sido utilizada a lo largo del desarrollo.

5.2.3 Extensión del rango de trabajo

El ancho espectral del LED siempre ha sido uno de los principales factores limitantes a la hora de desarrollar equipos de espectrofotometría debido al efecto de las múltiples longitudes de onda.

Aunque se ha mostrado que es posible compensar su efecto, el hecho de tener que utilizar un LED para modelizar una única longitud de onda, requeriría de un enorme número de luces LED y esto no parece ser lo más eficiente, ya que llevaría aparejado un aumento del coste, dimensiones y tiempo de análisis del equipo. Por ello, gracias a las investigaciones llevadas a cabo, se ha logrado desarrollar una metodología de calibración, que permite aislar y aprovechar, otras longitudes de onda emitidas, más allá de la longitud de onda de pico. Es decir, a continuación se va a mostrar cómo cada luz LED sea capaz de reproducir la transmitancia de una franja superior a la de estrictamente su longitud de onda pico.

Las pruebas realizadas mostraron que la porción del espectro que puede utilizarse para ampliar el rango de trabajo de cada una de las luces LED de un espectrofotómetro está definida por aquellas longitudes de onda comprendidas entre 10 y 20nm por encima del valor de pico en el límite superior, así como el límite inferior del ancho espectral (LSWL), con una bondad en el ajuste superior al 96%.

De una forma gráfica, estos límites se recogen en la Figura 5.8, donde se ha tomado, a modo de ejemplo, un LED de 470nm de longitud de onda de pico, delimitándose en color rojo la porción del espectro útil.

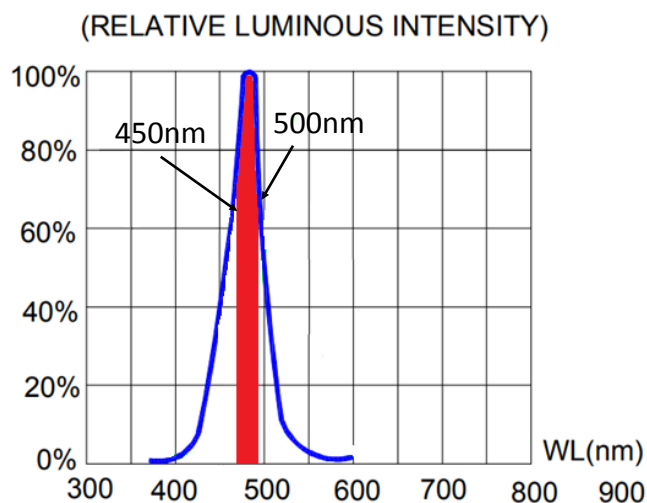


Figura 5.8. Representación del rango de trabajo útil de un LED de longitud de onda pico de 470nm.

Aunque en esa región del ancho espectral permite obtener ajustes muy precisos, las pruebas llevadas a cabo han demostrado que es preferible utilizar una región algo más estrecha, que aunque se reduzca el número de longitudes de onda “disponibles”, el ajuste

mejora de forma significativa por encima del 98%, tal y como se muestra en la Ecuación 5.1.

$$\begin{cases} \lambda_{LSWL} < \lambda_{peak} < \lambda_{peak} + 10\sim 20nm, & 0.96 < R^2 < 0.99 \\ \lambda_{LSWL} + 10nm < \lambda_{peak} < \lambda_{peak} + 10\sim 20nm, & 0.98 < R^2 < 0.99 \end{cases} \quad (5.1)$$

Con el fin de demostrar la adecuación de esta región a la hora de modelizar las longitudes de onda, analizaremos el efecto en las correlaciones para los valores extremos definidos en la Ecuación (5.1). Para ello, consideraremos un LED con 470nm de pico y un ancho espectral útil entre 450 y 500nm.

En primer lugar, analizaremos la viabilidad de modelizar la longitud de onda de 450nm, que representa el límite inferior del ancho espectral del LED de 470nm, y por ende, la situación más desfavorable, dado que cuanto más alejada esté la longitud de onda a analizar con respecto a la de pico, peor es el ajuste.

Estos valores se comparan en la Figura 5.9 con los obtenidos por el equipo comercial basado en lámparas a 450nm con el fin de mostrar la precisión del ajuste.

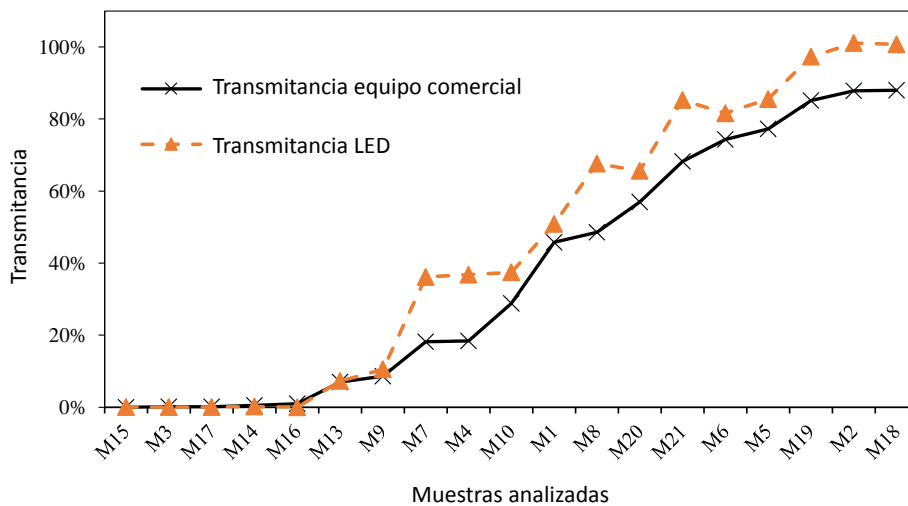


Figura 5.9. Comparación de valores de transmitancia obtenida a 450nm con el equipo comercial y un LED de longitud de onda pico de 470nm.

Como se observa en la Figura 5.9, la tendencia es similar en ambas gráficas, aunque presenta algunos desajustes para ciertas muestras. Estas irregularidades se observan con mayor claridad en el diagrama de dispersión de la Figura 5.10, donde ciertos puntos tienden a alejarse de la recta de tendencia.

No obstante, el ajuste es superior al 97%, lo que demuestra que es posible modelizar con éxito el límite inferior del ancho espectral. Si se seleccionara otra longitud de onda por debajo de este límite, la calidad de las correlaciones se vería muy afectadas, cayendo su nivel de correlación de forma abrupta.

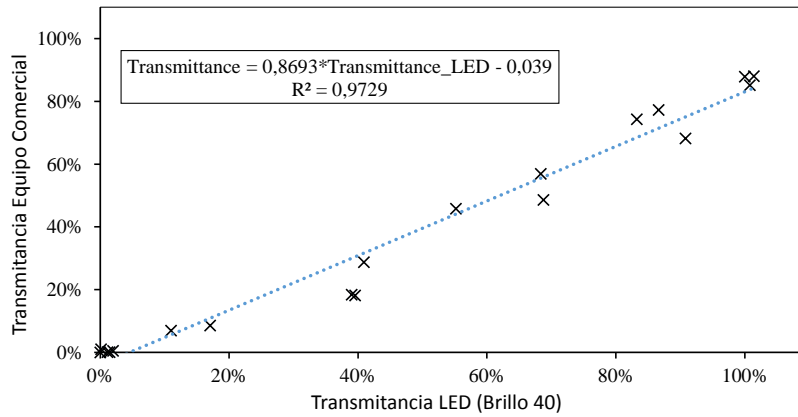


Figura 5.10. Diagrama de dispersión entre el valor de referencia de transmitancia a 450nm con el LED de 470nm de pico y un cierto nivel de brillo (T40).

Como es de esperar, las longitudes de onda comprendidas entre este límite inferior de la anchura espectral (450nm en el caso del ejemplo) y el valor de pico, también pueden calibrarse mediante el LED de 470nm, consiguiendo una mejora del ajuste a medida que nos acercamos a la longitud de onda de pico.

Por otro lado, debemos tener en cuenta que una misma longitud de onda puede ser caracterizada por más de un LED, siempre y cuando esta se encuentre dentro de la zona designada en la Ecuación 5.1.

Con el fin de determinar si es viable o no modelizar el límite superior del ancho espectral, en la Figura 5.11 se muestra una comparativa de los valores de transmitancia medidos a 500nm con un equipo comercial basado en lámparas incandescentes, y el LED de 470nm.

Así mismo, se realiza una comparación con los valores de transmitancia que se obtendrían con un LED con longitud de onda pico de 500nm (SSL-LX5093UEGC, [170]) con el fin de mostrar el error que se obtendría en comparación a otro LED que emitirá en una zona más cercana.

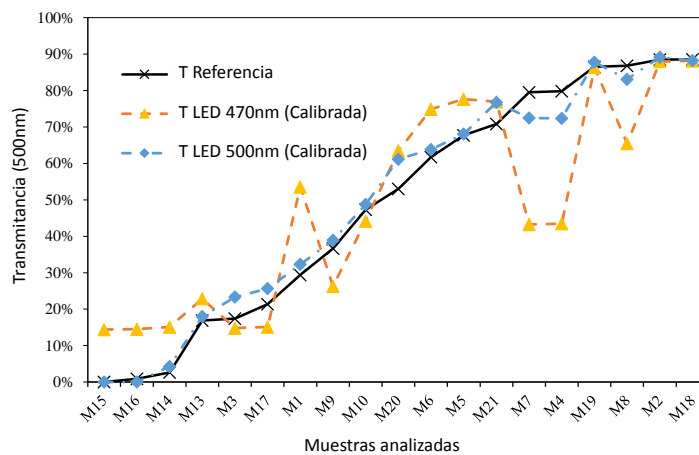


Figura 5.11. Comparación de los valores de transmitancia obtenida a 500nm con el equipo comercial y el prototipo desarrollado a partir de un LED con una longitud de onda de pico de 470nm y 500nm.

Como era de esperar, la modelización con el LED de 470nm ha presentado una peor correlación (74.68%), frente al 98.97% obtenido con el LED de 500nm. El LED de 470nm representa un empeoramiento del 24%. Esto se debe a que 500nm se encuentra en el límite superior del ancho espectral del LED de 470nm, es decir, a 30nm de su longitud de onda pico, y por tanto se espera que proporcione una peor respuesta que un LED con una longitud de onda de pico más cercana a 500nm.

Este resultado demuestra que no podemos utilizar un LED cuya longitud de onda de pico esté a más de 30nm de la que pretendemos calibrar. Sin embargo, si la longitud de onda a calibrar está a 20nm o menos con respecto al valor de pico del LED, es posible obtener una alta correlación. Prueba de ello es el LED con una longitud de onda de pico de 428nm, que puede explicar los 445nm con una precisión del 99% (Tabla 5.2).

Esto demuestra que existe una capacidad limitada en cuanto a las longitudes de onda “aprovechables” del ancho espectral de un LED. Este proceso de experimentación nos ha permitido definir los límites recogidos en la Ecuación (5.1) antes presentada.

Gracias a este proceso de calibración se han obtenido diferentes modelos de calibración para las longitudes de onda comprendidas entre 380-700nm, obtenidas a partir de 34 LEDs distintos, frente a los 59 utilizados en un inicio, gracias a la extensión del rango de trabajo. Estos modelos se recogen en la Tabla 5.2, donde se indica, no sólo el modelo calculado, sino también la bondad del ajuste, así como el LED utilizado para este fin.

En el Anexo B, se recogen de forma pormenorizada todos los modelos de calibración utilizados para modelizar las longitudes de onda comprendidas entre 380 y 700nm, donde se indica la longitud de onda pico del LED con la que ha sido calibrada, ecuación de calibración, así como su ajuste.

Tabla 5.2. Modelos de calibración calculados para longitudes de onda comprendidas entre 380-700nm.

λ	$\lambda_{Pico LED}$	Ecuación calibración	$\overline{R^2}$	λ	$\lambda_{Pico LED}$	Ecuación calibración	$\overline{R^2}$
380	385	$T = 0,7715x + 0,0057$	0,9952	558	558	$T = 0,8955x - 0,0069$	0,9926
385		$T = 0,7898x + 0,0053$	0,9974	560		$T = 0,8895x - 0,0093$	0,9913
390	390	$T = 0,8113x + 0,0028$	0,9986	565	570	$T = 0,8907x - 0,0098$	0,9865
395		$T = 0,8169x + 0,0023$	0,9989	574		$T = 0,9053x - 0,0223$	0,9958
400	400	$T = 0,8196x + 0,0015$	0,9982	574	574	$T = 0,914x - 0,0242$	0,9933
405		$T = 0,825x + 0,0001$	0,9983	575		$T = 0,9114x - 0,0279$	0,9948
410	405	$T = 0,8251x - 0,0026$	0,9956	580	583	$T = 0,9134x - 0,0189$	0,9851
415		$T = 0,8337x - 0,0019$	0,9941	583		$T = 0,9194x - 0,0241$	0,9919
420	405	$T = 0,8382x + 0,0018$	0,9916	585	586	$T = 0,9207x - 0,0287$	0,9937
425		$T = 0,8416x + 0,0057$	0,9897	586		$T = 0,889x - 0,0001$	0,9824
428	428	$T = 0,9004x - 0,0418$	0,9892	590	594	$T = 0,8828x - 0,0053$	0,9951
430		$T = 0,8985x - 0,0399$	0,9906	591		$T = 0,8828x - 0,006$	0,9958
435	428	$T = 0,9x - 0,0391$	0,9911	594	594	$T = 0,8805x - 0,0062$	0,9963
440		$T = 0,8973x - 0,0342$	0,9917	595		$T = 0,8806x - 0,0085$	0,9954
445	461	$T = 0,891x - 0,0286$	0,9924	600	607	$T = 0,8705x - 0,006$	0,9928
450		$T = 0,8728x - 0,0214$	0,9918	602		$T = 0,8687x - 0,0062$	0,9902
455	461	$T = 0,879x - 0,0214$	0,9941	605	605	$T = 0,9086x - 0,0133$	0,9868
460		$T = 0,8738x - 0,0115$	0,9973	607		$T = 0,9106x - 0,0171$	0,9886
461	461	$T = 0,8665x - 0,0039$	0,9986	609	609	$T = 0,876x + 0,0015$	0,9897
464		$T = 0,8621x + 0,0034$	0,9981	610		$T = 0,8841x - 0,0046$	0,9919
465	468	$T = 0,8747x - 0,0117$	0,9972	615	622	$T = 0,8958x - 0,0213$	0,9924
468		$T = 0,8753x - 0,0002$	0,9978	620		$T = 0,896x - 0,0098$	0,9906
470	468	$T = 0,8685x + 0,0092$	0,9953	622	624	$T = 0,9025x - 0,0171$	0,9935
475		$T = 0,864x + 0,0169$	0,9875	624		$T = 0,8973x - 0,0114$	0,9912
480	500	$T = 0,8537x + 0,0265$	0,9628	625	627	$T = 0,9013x - 0,0154$	0,9924
485		$T = 0,9198x - 0,0737$	0,9354	627		$T = 0,9797x - 0,0748$	0,9914
490	500	$T = 0,9053x - 0,0367$	0,9673	630	631	$T = 0,9081x - 0,0195$	0,9923
495		$T = 0,9197x - 0,0305$	0,9778	631		$T = 0,9069x - 0,0202$	0,9919
500	505	$T = 0,9254x - 0,0177$	0,99	632	636	$T = 0,9081x - 0,0196$	0,9916
505		$T = 0,9061x - 0,0149$	0,9896	635		$T = 0,946x - 0,0551$	0,9944
510	515	$T = 0,9084x - 0,0368$	0,9708	636	639	$T = 0,9455x - 0,0549$	0,9944
515		$T = 0,9134x - 0,0227$	0,9714	639		$T = 0,9546x - 0,0637$	0,9919
520	521	$T = 0,9191x - 0,0386$	0,9737	640	639	$T = 0,952x - 0,0587$	0,9907
521		$T = 0,9213x - 0,0313$	0,9751	642		$T = 0,934x - 0,0464$	0,9947
522	532	$T = 0,8969x - 0,0193$	0,9757	645	660	$T = 0,9226x - 0,0335$	0,9951
530		$T = 0,9061x - 0,0113$	0,9627	660		$T = 0,9749x - 0,0592$	0,9824
532	557	$T = 0,9128x - 0,0031$	0,9671	700	700	$T = 0,9402x - 0,0191$	0,9788
550		$T = 0,8831x - 0,0216$	0,9689				
555	557	$T = 0,8852x - 0,0158$	0,989				
557		$T = 0,8904x - 0,0106$	0,9938				

5.2.4 Generación de ecuación genérica de calibración

Las longitudes de onda contenidas dentro de una cierta región del ancho espectral del LED (Ecuación 5.1), guardan una correlación de tipo lineal con respecto a los valores que se obtendrían en un equipo comercial, es decir, sin el efecto de las restantes longitudes de onda emitidas [29].

Esto permite seleccionar la longitud de onda de trabajo de una forma simple y rápida, pero exige calcular la recta de calibración específica para cada longitud de onda. Esto resta versatilidad y limita las posibilidades de ampliar el espectro, dado que precisa de la recolección de nuevos valores de transmitancia y su comparación con los valores esperados en un equipo comercial.

No obstante, gracias a las investigaciones llevadas a cabo, es posible encontrar una expresión genérica que nos permita ajustar cada una de las longitudes de onda deseadas, independientemente del LED utilizado, de una forma simple y rápida y con una alta precisión (siempre que se encuentre contenido en el rango de la Ecuación 5.1)

Para ilustrar el procedimiento seguido para obtener las ecuaciones genéricas de calibración, se han representado en la Figura 5.12 todas las líneas de calibración recogidas en la Tabla 5.2.

Para una mejor exposición de los resultados, cada una de las líneas de calibración se ha representado con el color característico de su longitud de onda.

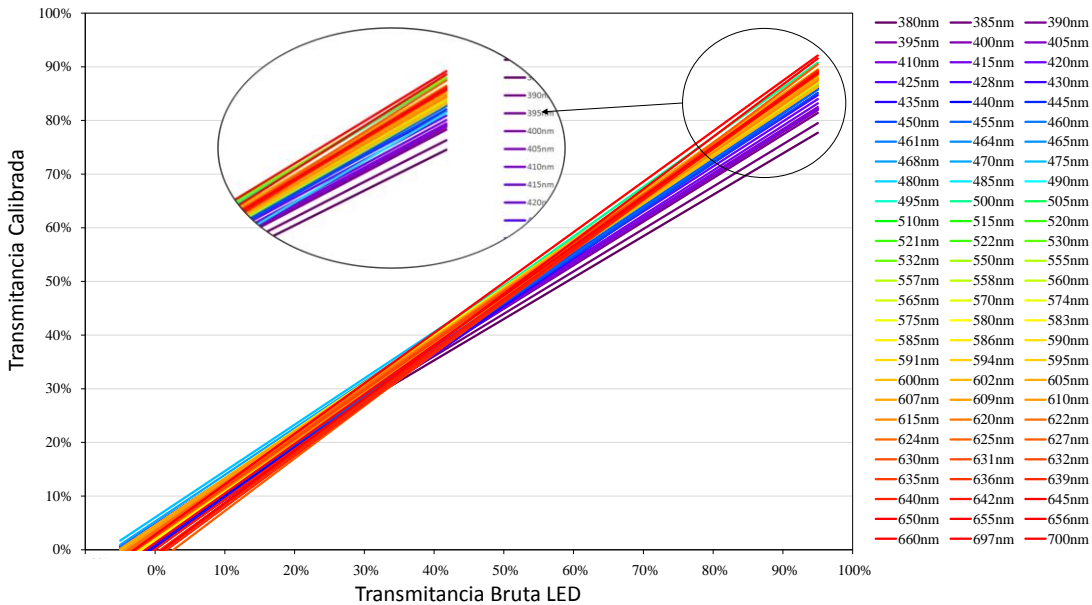


Figura 5.12. Rectas de calibración entre 380 y 700nm.

Las líneas de calibración tienen una pendiente que cambia en función de la longitud de onda. En la vista ampliada de la Figura 5.12, se puede observar que éstas se distribuyen de forma muy similar a como lo hace el espectro visible. Esto se aprecia con mayor claridad en la zona marcada. Sin embargo, la forma en que evoluciona la pendiente cambia dependiendo del rango o franja de longitudes de onda que se analice.

Para entender el comportamiento de las líneas de calibración (Figura 5.12), la Figura 5.13 muestra el valor de la transmitancia calibrada en un único punto de dicha línea, que es aquel en el que todos los LEDs proporcionan un valor de transmitancia bruta (antes de la calibración) del 50% para una determinada muestra, según el eje de abscisa de la Figura 5.12. Gráficamente, se observa cómo este punto se puede considerar, de forma aproximada, como un lugar de intersección común a todas las rectas de calibración. Este lugar donde las pendientes de las rectas de calibración cambian, permite alcanzar ecuaciones genéricas de calibración, tal y como se muestra a continuación.

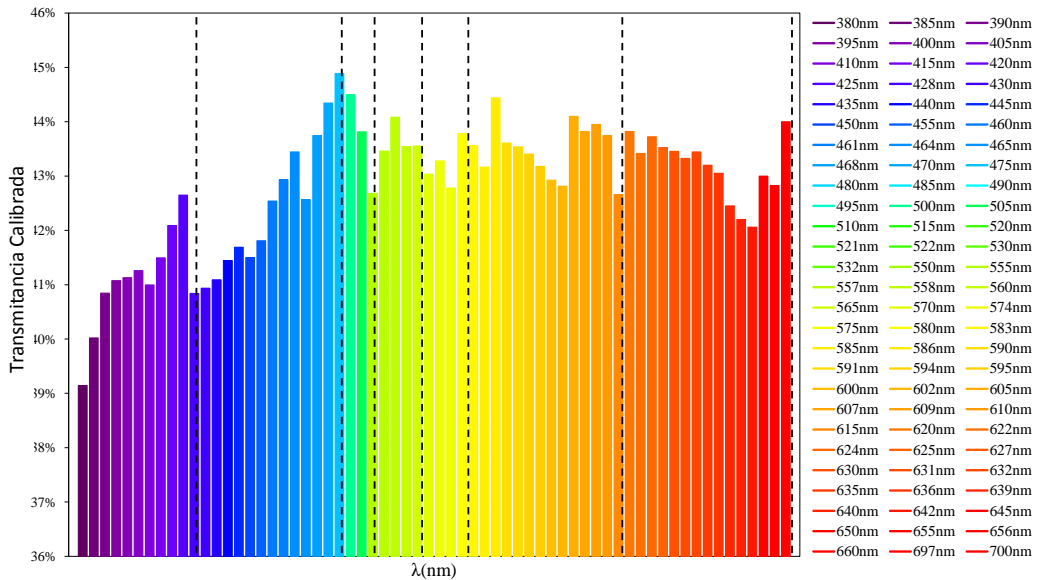


Figura 5.13. Valores de transmitancia calculados en el punto de inflexión de las líneas de calibración de la Figura 5.12. Transmitancia LED 50%.

Como podemos ver en la Figura 5.13, la representación de los valores de transmitancia en el punto de inflexión (50%), parece formar ciertos grupos de cambio de pendiente de la recta de calibración. Cada línea vertical representa el valor de transmitancia a una longitud de onda concreta, la cual ha sido representada con su color característico para facilitar la exposición de los resultados.

Si nos fijamos en las longitudes de onda que delimitan cada grupo, observamos que éstas se corresponden con los diferentes grupos de colores que componen el espectro de la luz visible cuyos límites espectrales se muestran en la Figura 5.14.

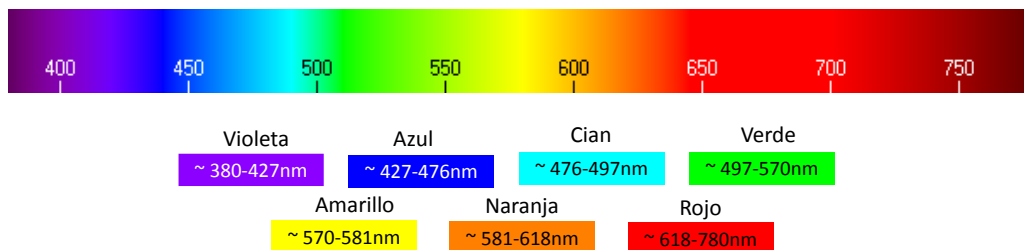


Figura 5.14. Espectro de color de la luz visible.

A partir de estos grupos, hemos determinado las expresiones que relacionan la línea de calibración de la Figura 5.12 con la longitud de onda. Con el fin de simplificar la exposición

de los resultados, a continuación se muestra únicamente el procedimiento para determinar la ecuación genérica de calibración para el grupo de color "Azul", es decir, entre 427 y 480nm, a intervalos de 5nm, cuyas ecuaciones de calibración se recogen en la ya citada Tabla 5.2.

Al tratarse de modelos lineales [171], los únicos parámetros que necesitamos estimar es la pendiente m y la ordenada en el origen n . Mediante un gráfico de dispersión, es posible obtener la línea de tendencia que relaciona ambos parámetros con respecto al valor de λ , tal y como se muestra en la Figura 5.15.

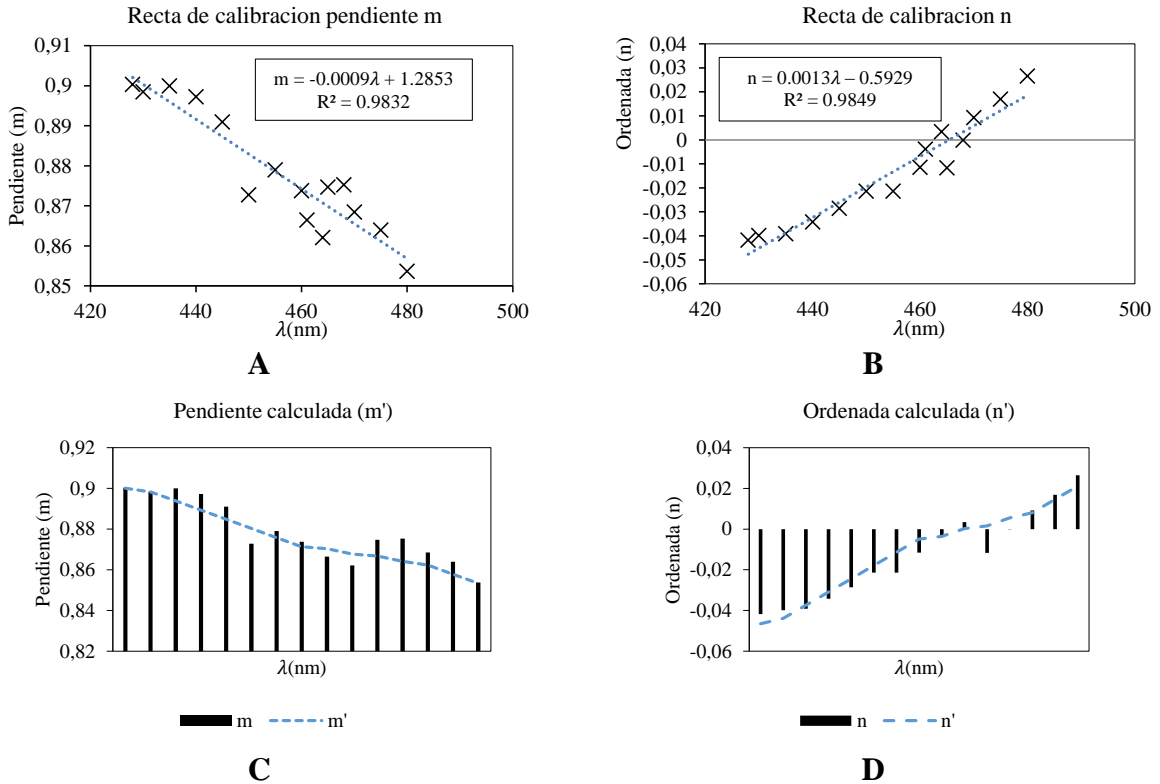


Figura 5.15. Cálculo de las líneas de tendencia de la pendiente m y de la ordenada en el origen n en función de λ . **A** y **B** son los gráficos de dispersión utilizados para poder calcular la recta de regresión (modelo) de la pendiente m y la ordenada en el origen n . **C** y **D** muestran los valores de m y n calculados mediante ese modelo.

Tanto la pendiente m como la ordenada en el origen n , pueden definirse como una recta cuya pendiente depende de la longitud de onda que pretendemos calcular (Figuras 5.15 A y B). Como se observa en las Figuras 5.15 C y D, los valores de m' y n' calculados con este modelo (representados con la línea discontinua) mediante el procedimiento de calibración descrito en el presente capítulo, son muy próximos a los valores obtenidos empíricamente, m y n (barras verticales). A partir de estos datos, es posible determinar la expresión que permite correlacionar los resultados de transmitancia obtenidos por el equipo comercial con respecto a los valores obtenidos por un LED, según la longitud de onda.

En la Ecuación (5.2), se muestran las ecuaciones genéricas de calibración para cada uno de los grupos de colores que se muestran en las Figuras 5.13 y 5.14, siendo λ la longitud de onda que se desea modelizar y T el valor de transmitancia medida con el LED utilizado.

$$\begin{aligned}
 y_{380-427nm} &= (0.0013\lambda + 0.2941) * T \\
 y_{427-480nm} &= (-0.0009 * \lambda + 1.2853) * T + (0.0013 * \lambda - 0.5929) \\
 y_{480-497nm} &= (-0,00001 * \lambda + 0.9198) * T + (0.0043 * \lambda - 2.1638) \\
 y_{497-570nm} &= (-0.0004 * \lambda + 1.1273) * T + (0.0002 * \lambda - 0.1245) \quad (5.2) \\
 y_{570-581nm} &= (0.0578 * \lambda - 32.296) * T + (0.0221 * \lambda + 12.652) \\
 y_{581-618nm} &= (-0.018 * \lambda + 11.677) * T + (-0.0032 * \lambda + 1.8929) \\
 y_{618-700nm} &= (0.0006 * \lambda + 0.5462) * T + (-0.0001 * \lambda + 0.0277)
 \end{aligned}$$

Es importante destacar que aunque el color rojo es hasta 780nm, sólo hemos realizado la calibración hasta 700nm, ya que no se disponía de ningún diodo que emitiera en esas longitudes de onda durante la fase de desarrollo.

5.3 Resultados del proceso de calibración

Utilizando el procedimiento de calibración descrito a lo largo del presente capítulo, se ha comprobado que es posible conseguir resultados muy cercanos a los obtenidos utilizando equipos comerciales basados en lámparas incandescentes.

En este epígrafe, se exponen los valores de transmitancia medidos para diferentes tipos de muestras y concentraciones, junto con los valores proporcionados por el equipo comercial con el fin de que sirvan como referencia para comprobar la precisión de los resultados.

Como se puede observar en la Figura 5.16, existe un alto nivel de correlación con los valores obtenidos por las lámparas incandescentes. Con el fin de mostrar el efecto del proceso de calibración, cada una de las gráficas se acompaña de los valores de transmitancia obtenidos por los LEDs antes del proceso de calibración, con el fin de mostrar la desviación en los resultados, los cuales han sido representados por la gráfica de color verde, mientras que los valores tras calibración se muestran en color rojo.

Si presta atención a las Figuras 5.16 B, C, D y E, se observa que para las longitudes de onda entre 510 y 550nm (que corresponden al espectro de color verde) existen ciertas irregularidades.

Dicha irregularidad se debe a que, tras el proceso de calibración, el ajuste fue del 96%, mientras que en el resto de longitudes de onda se obtuvieron valores superiores al 98%.

Este problema se debe probablemente a que los LEDs utilizados para explicar esta región del espectro no tienen la longitud de onda pico indicada por el fabricante. Este problema sólo se ha observado en los LEDs utilizados para ese rango de longitudes de onda, cuyas longitudes de onda pico según el fabricante son 515, 531, 532 y 550nm, respectivamente.

En cada uno de los gráficos de resultados, se ha delimitado mediante líneas verticales el rango de longitudes de onda explicado por cada uno de los 34 LEDs utilizados tras el

proceso de calibración. Cada línea vertical representa a un LED con su longitud de onda de pico.

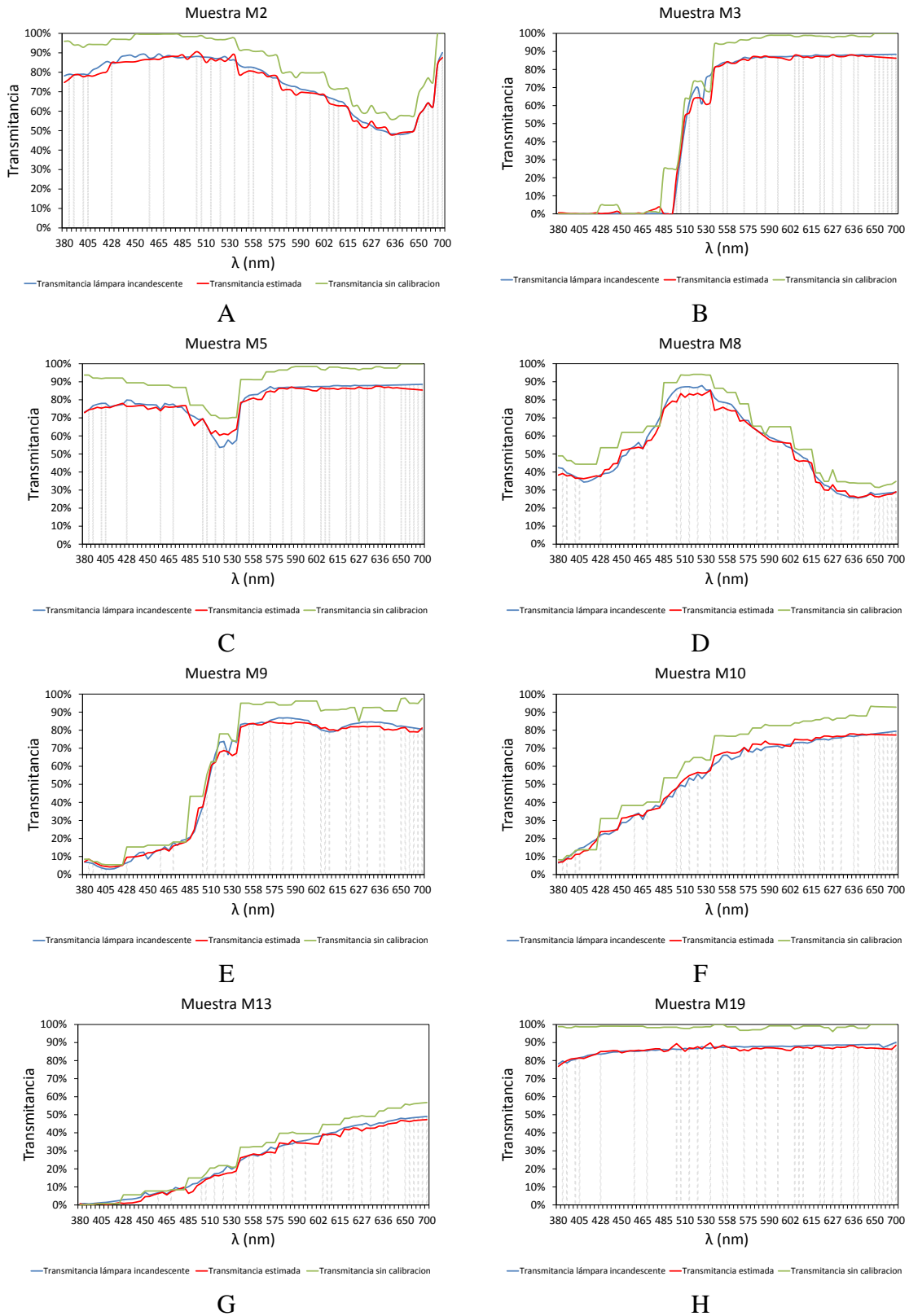


Figura 5.16. Comparativa de los valores de transmitancia antes (verde), después (rojo) con respecto al equipo comercial (azul) para las siguientes muestras: (A) M2, (B) M3, (C) M5, (D) M8, (E) M9, (F) M10, (G) M13 y (H) M19.

Las muestras utilizadas para llevar a cabo el proceso de calibración son representativas de los valores de transmitancia que pueden presentar las muestras de aguas residuales.

En una muestra de aguas residuales, las respuesta espectral en cuanto a la transmitancia tiene una tendencia creciente a medida que nos acercamos a la región infrarroja del espectro, tal y como se muestra en la Figura 5.17.

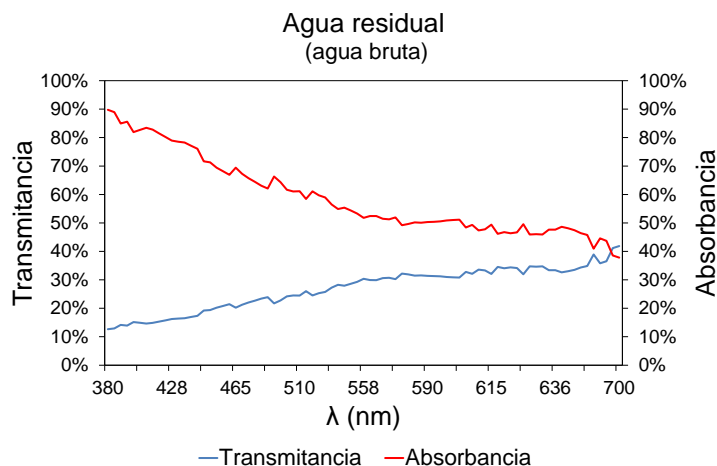


Figura 5.17. Respuesta espectral de muestra de agua residual.

A la vista de los resultados, se concluye que sí es posible obtener mediante el uso de tecnología LED, una respuesta comparable a la obtenida por equipos comerciales basados en lámparas incandescentes, tras un proceso de calibración.

Con el fin de mostrar la adecuación de los resultados, en la Tabla 5.3 muestra el error medio cuadrático (RMSD) [172], y el índice de error, Er, de cada uno de las muestras reflejadas en la 5.16. Estos indicadores han sido calculados por medio de las Ecuaciones (5.3) y (5.4).

$$RMSD = \sqrt{\frac{1}{n} \sum_i^n (T_{medida_i} - T_{calculada_i})^2} \quad (5.3)$$

$$Er(\%) = \left\| \frac{\sum_i^n (T_{medida_i} - T_{calculada_i})}{\sum_i^n T_{medida_i}} * 100 \right\| \quad (5.4)$$

Donde Er es el índice de error (%); n es el número de longitudes de onda (en nuestro caso, toma el valor de 81); T_{medida_i} y $T_{calculada_i}$ son los valores de transmitancia obtenidos mediante el equipo comercial [104] y el equipo descrito en el Capítulo 2.

En todos los casos se obtuvo un RMSD muy cercano. Esto demuestra que los valores de transmitancia obtenidos mediante equipos comerciales y los obtenidos mediante el proceso de calibración están muy próximos entre sí.

Prueba de ello son los valores de error, Er. Para todas las muestras, el error ha sido siempre inferior al 5%, y en casi todos los casos fue inferior al 1%.

Tabla 5.3. RMSD e Índice de Error.

Muestra	RMSD	Er (%)
M2	0,021	1,451
M3	0,028	0,962
M5	0,024	0,990
M8	0,027	2,135
M9	0,023	1,473
M10	0,019	1,426
M13	0,017	4,921
M19	0,0130	0,986

5.4 Conclusiones

En este capítulo, se ha descrito de forma pormenorizada que es posible utilizar la tecnología LED como un reemplazo viable de las lámparas incandescentes para el desarrollo de equipos de espectrofotometría, con un error inferior al 5% en todos los casos

Esta metodología propuesta permite:

- i)* Resultados comparables a los proporcionados por los equipos comerciales basados en lámparas incandescentes, sin utilizar ningún tipo de elementos ópticos como lentes o redes de difracción.
- ii)* Ampliación del rango de trabajo de cada LED, cubriendo todo el espectro visible utilizando un reducido número de diodos. Utilizando sólo 34 LEDs, es posible cubrir entre 380-700nm con una resolución de 5nm.
- iii)* Reducción de las dimensiones, costes y tiempos de muestreo del equipo, aspectos vitales para el desarrollo de sistemas autónomos de bajo coste diseñados para medir en cualquier tipo de entorno.

Esto abre la puerta al desarrollo de nuevos sistemas basados en esta tecnología, abaratando el coste de los equipos a la vez que reduciendo su tamaño y consumo, permitiendo la creación de equipos autónomos que puedan funcionar con baterías en cualquier entorno.

Capítulo 6

Calibración de LED RGB

6.1 Introducción

En el capítulo anterior se ha descrito de forma pormenorizada cómo el uso de LEDs de ancho espectral fijo puede ser un reemplazo viable para el desarrollo de equipos de espectrofotometría, describiendo cómo es posible seleccionar una longitud de onda específica sin necesidad de emplear elementos ópticos [29].

No obstante, dentro de la categoría de LED [173], encontramos a los conocidos como LED RGB, que no son más que encapsulados que albergan en su interior tres pequeños diodos que emiten a longitudes de onda pertenecientes a las regiones del rojo, verde y azul [31].

La importancia de este tipo de LEDs, radica, no en la respuesta espectral obtenida cuando cada uno de ellos actúa de forma individualizada, dado que eso es algo análogo a lo ya mostrado en el Capítulo 5, sino en la posibilidad de modelizar la respuesta de una porción mucho más amplia del espectro a partir de la combinación de las longitudes de onda del 460 (azul), 525 (verde) y 625nm (rojo) [174].

A lo largo de la tesis se ha ido mostrando cómo el elemento en torno al que gira todo el presente trabajo de investigación es el LED, el cual condiciona las dimensiones así como el coste y consumo de los equipos. El poder implantar un elevado número de dispositivos de caracterización en la red, exigen que los equipos sean lo más pequeños y económicos posibles, por ese motivo, se llevó a cabo un estudio para determinar la viabilidad de reemplazar estos LEDs de ancho espectral fijo, por un único diodo RGB (LED RGB).

El uso de este tipo de tecnología en procesos de medición no es algo nuevo [175, 176]. Hoy en día, se utilizan en los procesos de medición molecular, como los desarrollados para mediciones de absorbancia molecular en procesos de análisis químico. Por ejemplo, se han desarrollado para la determinación de la concentración de microalgas y la biomasa dentro de un fotobiorreactor, presentando un error del 2% [177], la determinación de nitrógeno y hierro o incluso el fósforo [178] en el agua.

No obstante, este tipo de diodos presentan una problemática análoga a los LEDs de ancho espectral fijo [29, 179, 180], por lo que es necesario llevar a cabo un proceso de calibración que permita equiparar los resultados proporcionado por un LED RGB [31], con lo que se obtendrían con un equipo basado en lámparas incandescentes comercial. Tal y como se ha descrito en capítulos anteriores, esta comparativa se ha realizado con el espectrofotómetro V-5000 Vis [104].

La idea que se persigue verificar a partir del uso de este tipo de diodos es la de comparar y calibrar la transmitancia de un equipo comercial con la respuesta producida en la muestra cuando diferentes longitudes de onda de distintas regiones del espectro visible inciden sobre la misma, dado que combinaremos las tres longitudes de onda incluidas en el LED RGB.

Por ello, en el presente capítulo, se tratará de forma pormenorizada dicho proceso de calibración para correlacionar los valores de transmitancia obtenidos como consecuencia de la combinación de diferentes longitudes de onda, demostrando su alto desempeño dentro del rango de 510-645nm [31].

Aunque este rango de trabajo es inferior al mostrado en el capítulo anterior (380-700nm) por medio de 34 LEDs, debemos tener en cuenta que este nuevo rango de emisión (510-645nm), ha podido ser generado con un único LED que dispone de tres diodos.

No todas las aplicaciones que utilizan la técnica de la espectrometría requieren el uso de un análisis multiespectral completo, sino sólo algunos puntos espectrales específicos, como es el caso de la medición de la cantidad de ozono en el agua [181].

Por ello, el uso de este tipo de diodos puede representar una excelente fuente de luz para ciertos entornos.

6.2 Resultados preliminares

Antes de comenzar con la exposición de los modelos obtenidos para equiparar los resultados, se hace necesario introducir algunos aspectos relativos a los LED RGB.

El desarrollo de equipos de espectrofotometría basado en este tipo de diodos siempre se ha enfrentado la misma problemática: El limitado rango de emisión de longitudes de onda, dado que sólo emite en tres porciones muy estrechas del espectro visible, más concretamente 435-485nm (azul), 500-550nm (verde) y 610-650nm (rojo), según una distribución normal, tal y como se muestra en la Figura 6.1, relativa a la respuesta espectral del LED RGB HV-5RGB25 [182], un diodo de cátodo común de 5mm de diámetro.

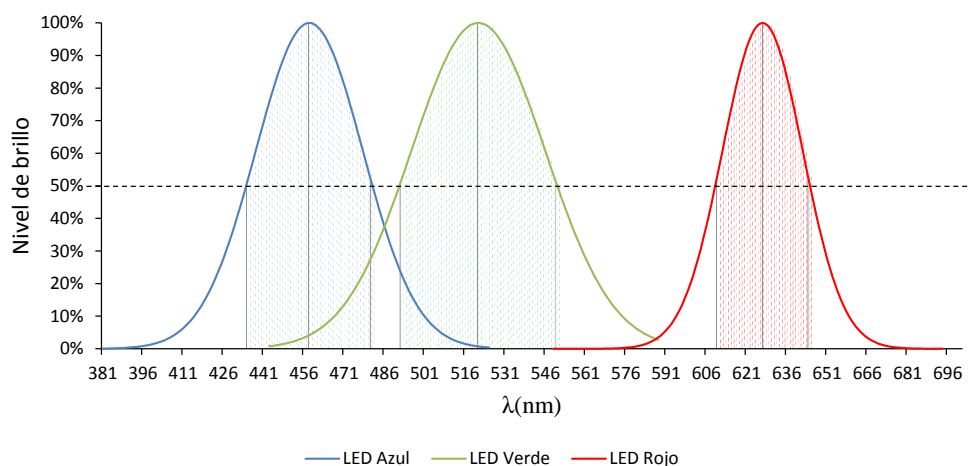


Figura 6.1. Espectro de emisión del LED RGB HV-5RGB25. [182]

Como se puede observar, cada uno de los LEDs emite ciertas longitudes de onda simultáneamente. Esto significa que en caso de ser utilizados como fuente de luz en un análisis espectrofotométrico, las muestras se verían atravesadas por más de una longitud de onda simultáneamente, lo que causará discrepancias con respecto a los resultados proporcionados por el equipo comercial, dando lugar a valores de transmitancia mayores, de forma análoga a lo ocurrido con los LEDs de ancho espectral fijo descritos en el capítulo anterior.

Si observamos cada LED de forma individualizada, estos parecen cubrir una región muy extensa del espectro visible, por ejemplo el LED azul emite entre 381 a 581nm, es decir 200 nm longitud de onda, sin embargo, no todas estas resultan “adecuadas” para su uso en aplicaciones como la que nos ocupa.

En la Figura 6.1 se ha resaltado una zona de cada espectro de emisión, lo que se conoce como "ancho espectral"[103], donde en caso del LED azul sería entre 435 y 485nm. Así pues, las longitudes de onda situadas por debajo de ese umbral tienen una baja influencia en las muestras y por lo tanto no “aprovechables” para el análisis de las muestras, tal y como se ha constatado durante la presente tesis doctoral [31].

Un paso previo a cualquier proceso de calibración es conocer cuál es la respuesta obtenida antes de su implementación, es decir, cual es el valor de transmitancia obtenido cuando cada LED individual (rojo, verde o azul) se enciende por separado.

Para ello, se ha hecho uso del LED RGB el HV-5RGB25 antes citado [182] para analizar las muestras de agua contenidas en la Tabla A4 presente en el Anexo A, las cuales abarcan tanto muestras de aguas residuales con diferentes niveles de carga contaminante, como colorantes alimentarios entre otras sustancias. De las cuales, las 20 primeras (S1-S20) fueron utilizadas para el proceso de calibración y obtención de los modelos de cálculo, mientras que las restantes fueron utilizadas para el proceso de validación de los mismos. Para llevar a cabo el estudio, se reprodujeron las muestras utilizadas en el Capítulo 5 a las que se agregaron otras nuevas. Para evitar confusión, las muestras en el presente capítulo se han denominado con la letra *S* seguida de un número, a diferencia de la letra *M* empleada en el capítulo anterior.

Como se puede observar, las muestras estudiadas presentan una gran heterogeneidad, con el fin de asegurar que los modelos obtenidos sean válidos, no sólo con muestras de agua residual, sino también con cualquier otro tipo de líquidos.

De estas, 9 muestras pertenecen a aguas residuales de la ciudad de Cartagena, las cuales son tanto de origen urbano como industrial. Con el fin de generar modelos más polivalentes, estas muestras fueron recogidas en diferentes puntos de la planta de tratamiento de aguas residuales del Cabezo Beaza, más concretamente: a la entrada de la planta de tratamiento (S21, S43 y S44), a la salida del decantador primario (S22, S45 y S46) y en la salida de la planta (S23, S47 y S48).

Con el fin de generar una mayor variedad de muestras, se utilizaron los siguientes colorantes alimentarios: E105 (S3), E133 (S3 y S4), E124 (S10, S13 y S14), E100 (S16) y E133 (S19).

Ante de enunciar el proceso de calibración y los modelos obtenidos, se hace necesario describir los diferentes elementos utilizados durante este proceso, con el fin de que sirva como referencia en caso de querer reproducir los resultados obtenidos. En la Figura 6.2 [31], se incluye un esquema general del proceso llevado a cabo para la obtención de los modelos de calibración.

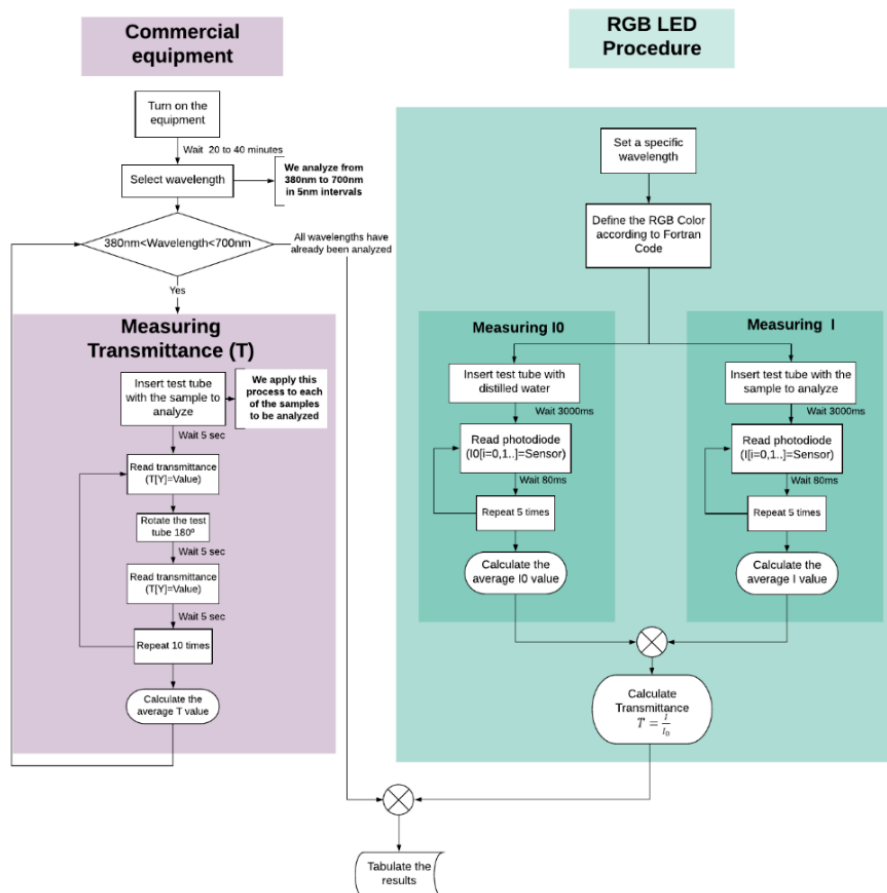


Figura 6.2. Esquema para la calibración de LED RGB. [31]

Como punto de partida, centraremos el estudio en cada uno de los LEDs individuales que componen el LED RGB por separado, con el fin de determinar su comportamiento de forma aislada, es decir, sin la interferencia de los otros LEDs.

Si tomamos el LED verde (500 a 550nm, y 525nm de pico) como ejemplo, la Figura 6.3 muestra que los valores de transmitancia alcanzados son más altos que los proporcionados por la lámpara incandescente a 525nm para las muestras detalladas en la Tabla 6.1.

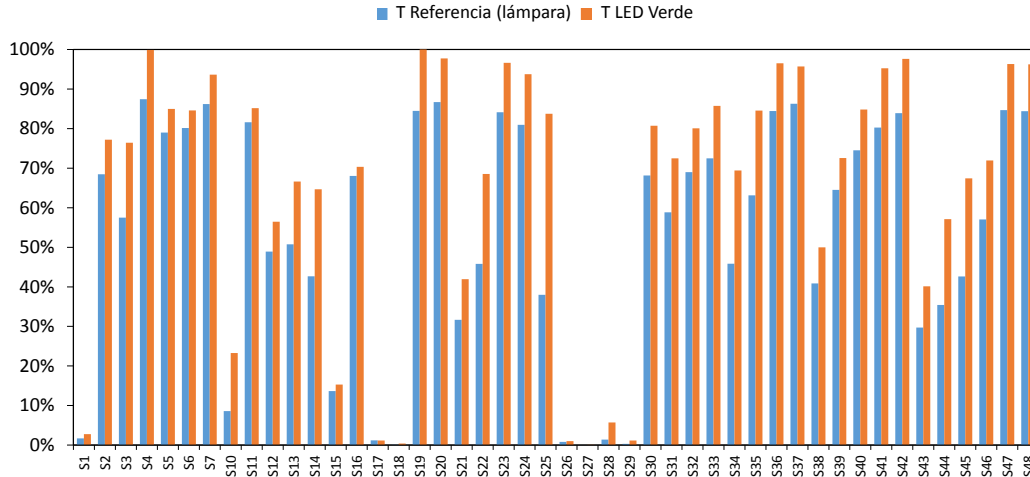


Figura 6.3. Comparativa entre los valores de transmitancia medidos con el LED verde del RGB lámpara incandescente a 525nm.

Aunque el LED verde tiene una longitud de onda máxima de 525nm, también emite otras longitudes de onda de forma simultánea (de 500 a 550nm). Esto aumenta la cantidad de luz que pasa a través de la muestra, y en consecuencia, se obtienen valores de transmitancia superiores a los obtenidos por el equipo comercial.

Este comportamiento es análogo al observado con los LEDs rojo y azul (Figuras 6.4 y 6.5), excepto que las diferencias fueron mucho más significativas en el caso del LED azul.

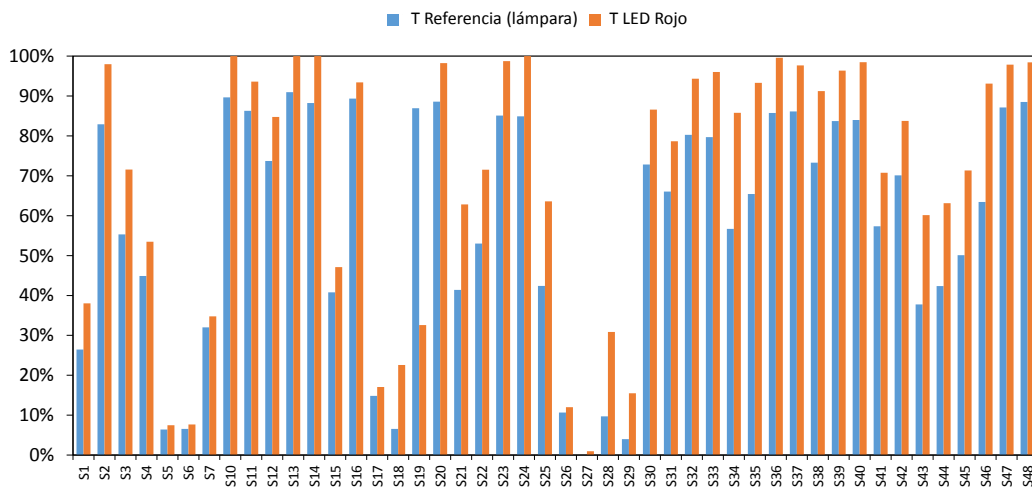


Figura 6.4. Comparativa entre los valores de transmitancia medidos con el LED rojo y la lámpara incandescente a 625nm.

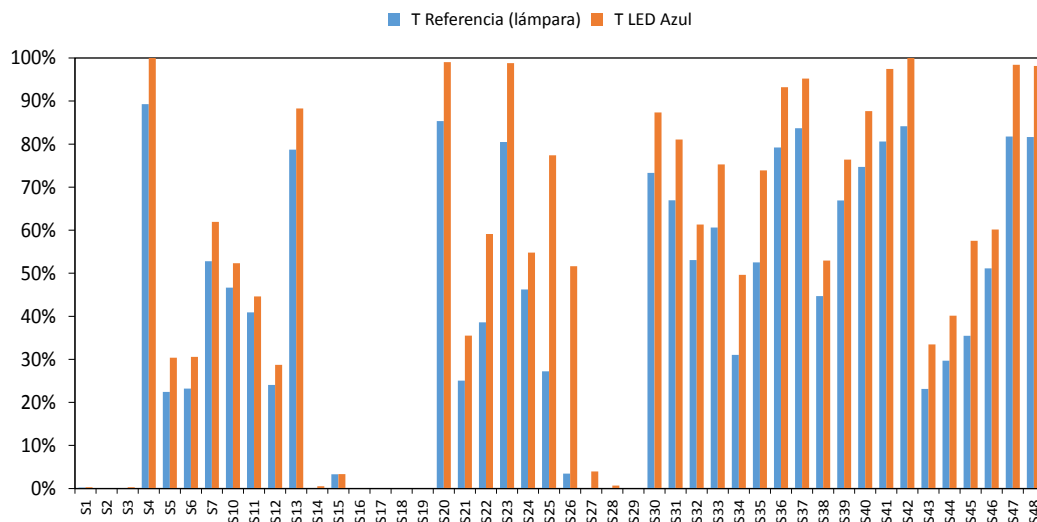


Figura 6.5. Comparativa entre los valores de transmitancia medidos con el LED azul y la lámpara incandescente a 460nm.

En vista de los resultados, se hace patente la necesidad de eliminar el efecto de las longitudes de onda adicionales. Las investigaciones llevadas a cabo concluyeron que existe una relación lineal entre los valores de transmitancia proporcionados por el LED RGB cuando utiliza cada uno de sus LED (rojo, verde y azul) de forma individualizada, y los valores que se obtendrían con equipos comerciales que trabajan con la misma longitud de onda pico de dichos LEDs.

Sin embargo, las pruebas han demostrado que no sólo se puede calibrar la longitud de pico, sino también todas las longitudes de onda emitidas dentro de un cierto rango. No obstante, debemos tener en cuenta que a medida que la longitud de onda que buscamos modelizar se aleje de la de pico, la precisión del modelo se reduce, por lo que no todas las longitudes de onda pueden ser modelizadas con éxito.

Para determinar qué porción del espectro de emisión puede utilizarse para aplicaciones espectrofotométricas, las Figuras 6.6-6.9 muestran el grado de correlación entre los niveles de transmitancia obtenidos con una lámparas incandescentes y los obtenidos por cada uno de los diodos individuales (rojo, verde y azul) que lo componen, actuando de forma independiente. Para ello se seleccionaron varias longitudes de onda seleccionadas el azar,

Tal y como se observa, la bondad del ajuste mejora a medida que la longitud de onda a modelizar se aproxime a la longitud de onda pico de cada LED, aspecto que era lógico dado que su efecto es mayor que el resto.

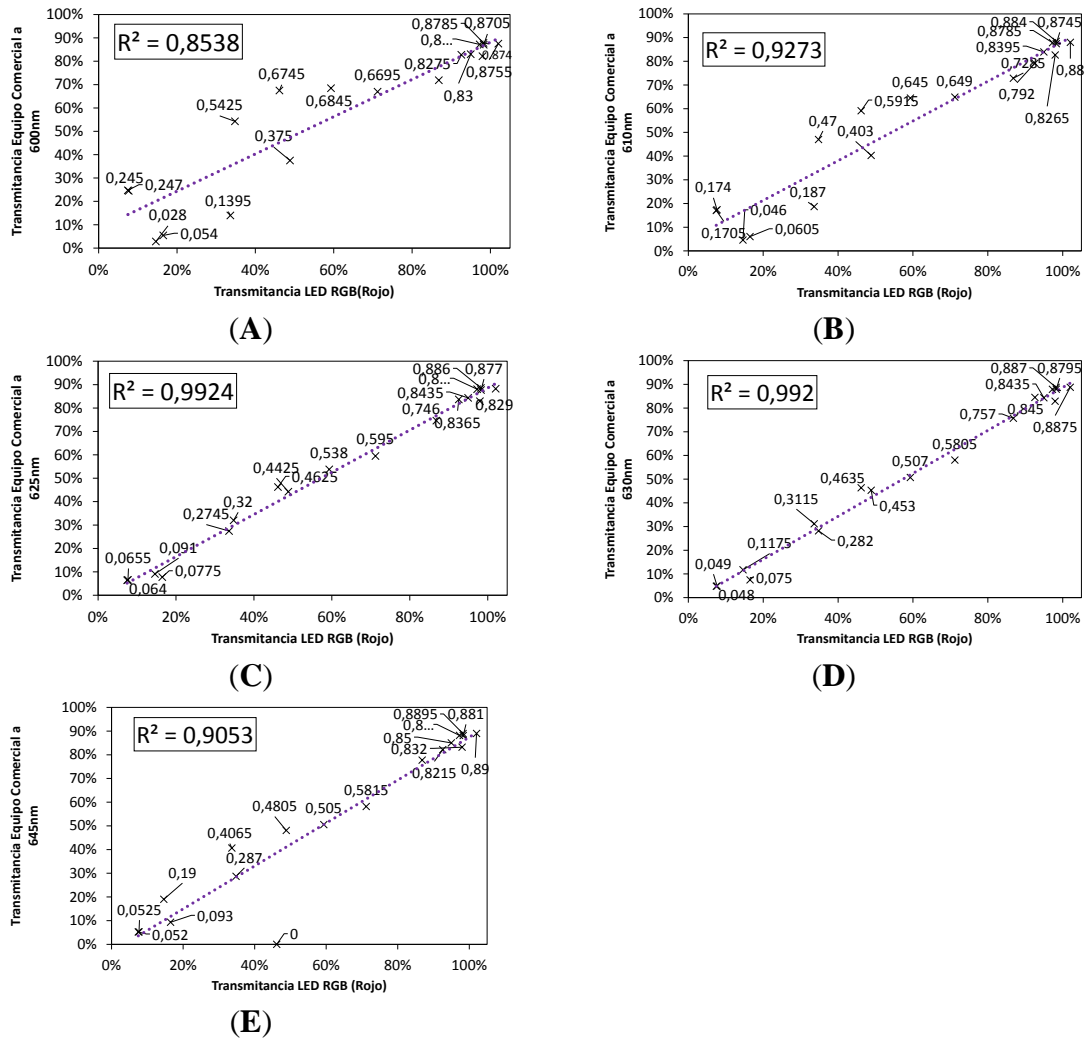


Figura 6.6. Comparativa del ajuste para el espectro de emisión del LED rojo: (A) 600nm, (B) 610nm, (C) 625nm, (D) 630nm, y (E) 645nm.

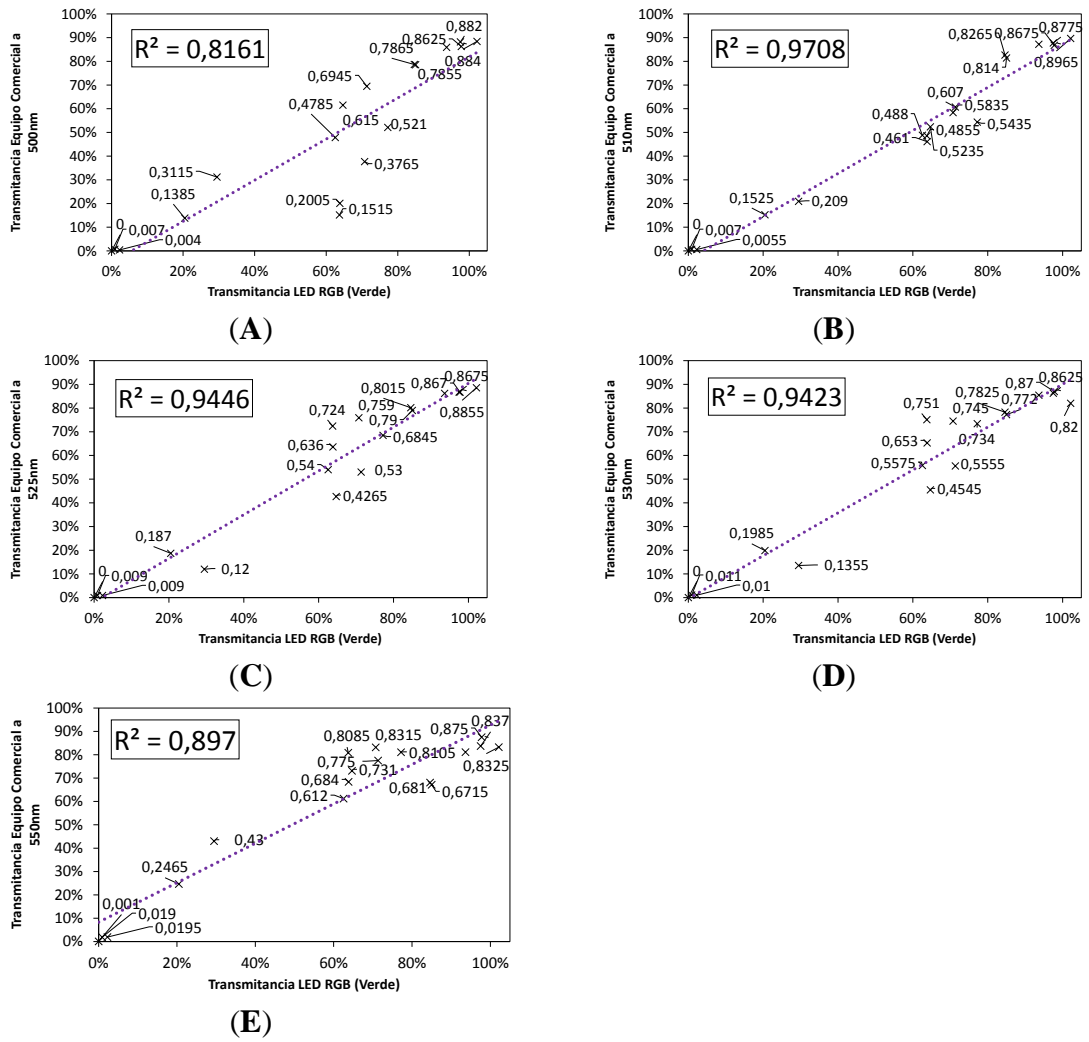


Figura 6.7. Comparativa del ajuste para el espectro de emisión del LED verde: (A) 500nm, (B) 510nm, (C) 525nm, (D) 530nm, y (E) 550nm.

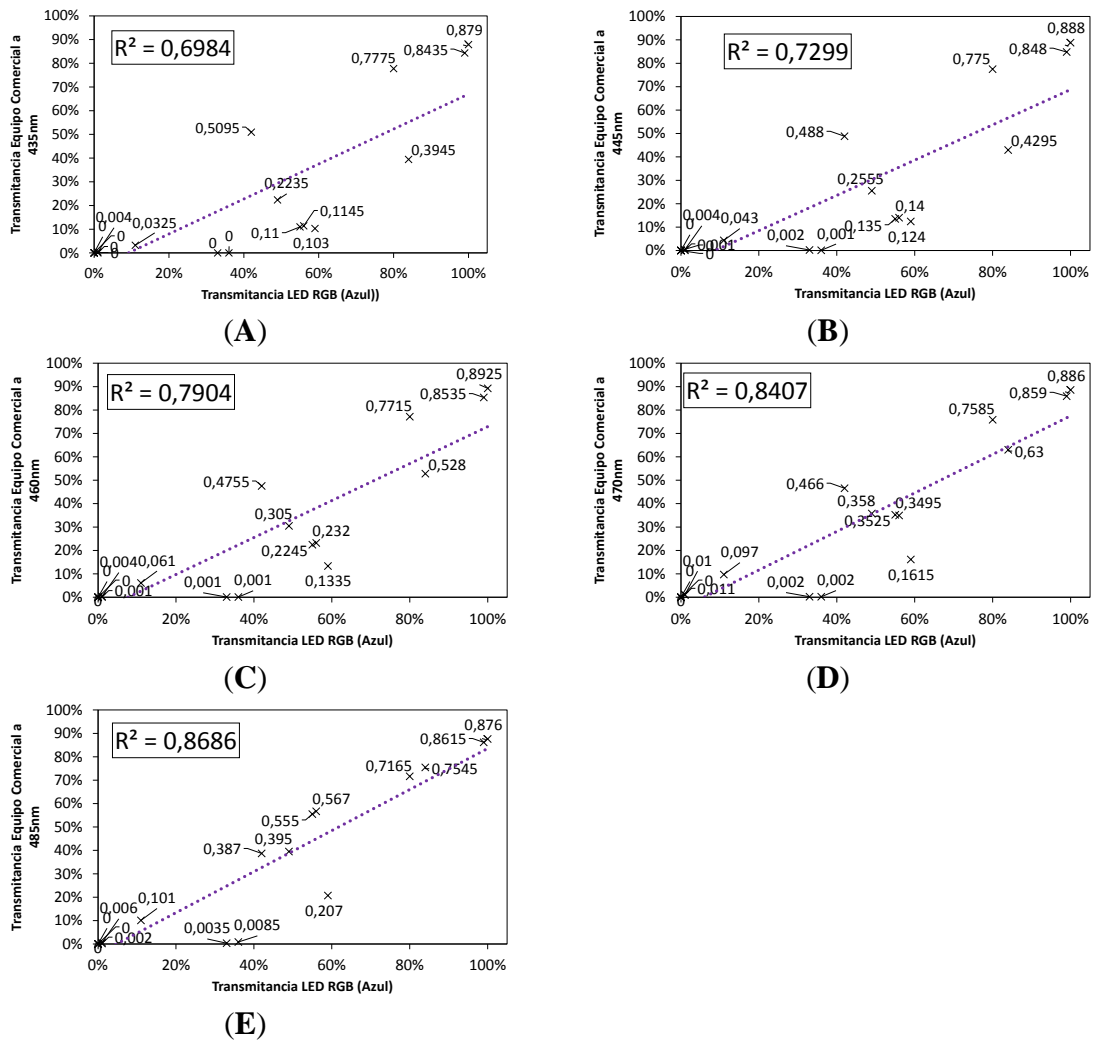


Figura 6.8. Comparativa del ajuste para el espectro de emisión del LED azul: (A) 435nm, (B) 445nm, (C) 460nm, (D) 470nm, y (E) 485nm.

Como se puede observar, el LED rojo presentó el mejor rendimiento, con un ajuste superior al 90%, seguido de cerca por el LED verde. Por otro lado, el LED azul presentó la mayor desviación, con un ajuste bastante bajo, lo que impide su uso para aplicaciones de espectrofotometría. Como era de esperar, los mejores ajustes se obtuvieron en la longitud de onda máxima y disminuyeron a medida que nos alejábamos de ese punto. Para comprender la causa del este bajo rendimiento en el caso del diodo azul, la Figura 6.9 muestra el grado de correlación entre los valores de transmitancia obtenidos por un LED de ancho espectral fijo con una longitud de onda de pico de 460nm, que es la longitud de onda pico del LED azul contenido en el LED RGB utilizado (Figura 6.8 D).

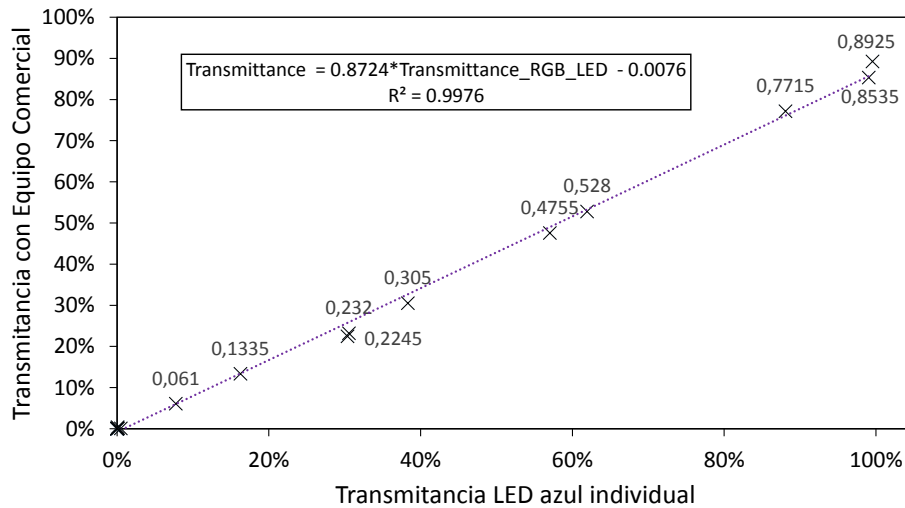


Figura 6.9. Ajuste del LED azul, con longitud de onda máxima es de 460nm.

Como se puede observar, las diferencias son notables. El LED azul produce un ajuste de más del 99%, a diferencia del LED RGB, que sólo alcanza el 79% como se muestra en la Figura 6.8 C.

Estos atípicos resultados obedecen a la calidad del semiconductor utilizado durante la fabricación del LED azul contenido en el LED RGB [183, 184]. Para poder comprenderlo, debemos describir por qué un LED emite luz. Este es el resultado del salto de electrones de la capa N a la capa P. El tamaño de este salto o “bandgap” [185, 186] es el que determina el tipo de luz que será emitida (longitudes de onda)

Un salto pequeño, producirá luz tendente al color rojo (alrededor de 1,91eV), mientras que un “bandgap” grande (2,64eV), producirá longitudes de onda más próximas al ultravioleta [187, 188]. Por lo tanto cuanto mayor sea este salto, más azul será la luz emitida por el diodo, tal y como se ilustra en la Figura 6.10.

La mayoría de los LEDs individuales de luz azul, constan de un sustrato de 1µm de zafiro, en el que crecen capas alternas de nitrato de galio, indio y aluminio. Estos elementos adicionales son clave para aumentar la eficiencia y el brillo azul. Además, con el uso del aluminio, es posible hacer incluso desarrollar versiones que sean capaz de emitir en el espectro ultravioleta.

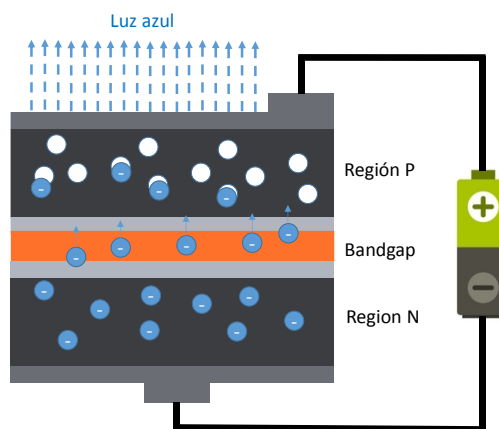


Figura 6.10. Diagrama de bandas del LED azul. Elaboración propia.

Es necesario mencionar que la generación del color azul mediante tecnología LED, siempre ha representado un desafío durante años, de hecho las investigaciones de Isamu Akasaki, Hiroshi Amano y Shuji Nakamura sobre el desarrollo de este tipo de diodos, les valió el Premio Nobel de Física en 2014 [189, 190].

A pesar que el LED azul contenido en el RGB tenga la misma composición antes descrita, el problema con la luz azul generada radica en la calidad de sus materiales. Debemos tener en cuenta que este tipo de diodos están diseñado fundamentalmente con fines visuales, es decir, que no existe criticidad a la hora de generar una luz “más o menos azul” [191].

Estos se basan en la combinación de colores primarios para crear la ilusión óptica de que ciertos colores están siendo generados. Este efecto óptico significa que no es necesario diseñar LEDs RGB que generen un "azul puro", sino algo que nuestro ojo perciba como azul [192] (en combinación con el rojo y el verde).

Para ilustrar este punto, imaginemos una pantalla de ordenador en la que se está proyectando el color naranja. El hecho de ver ese color, nos podría hacer pensar que la longitud de onda emitida por la misma sería de 615nm, que es la longitud de onda típica de lo que percibimos como naranja. El problema es que lo que percibimos como naranja en la pantalla del ordenador, en realidad no lo es, sino que es la combinación de rojo, verde y azul (píxeles) a distintos niveles de intensidad, que estimula unos fotoreceptores en nuestros ojos llamados conos, los cuales a un fopigmento llamado opsina que es sensible a las longitudes de onda del rojo, verde y azul (de ahí que estos sean los colores primarios) y que nuestro cerebro interpreta como un cierto tono de naranja [193, 194].

Por esta razón, la calidad de los sustratos utilizados y la pureza de los cristales son inferiores a los LEDs individuales, dado que se fundamentan en un efecto óptico.

Por lo tanto, aunque un LED RGB hace uso de tres diferentes diodos individuales, que emiten en una gran parte del espectro visible, sólo los LEDs rojo y verde parecen ser adecuados para la aplicación que nos ocupa. Además, de éstos, sólo una porción del espectro entre las siguientes longitudes de onda es adecuada para aplicaciones espectrofotométricas: 510 a 550nm (LED verde) y 610 a 645nm (LED rojo), como se presenta en la Figura 6.11.

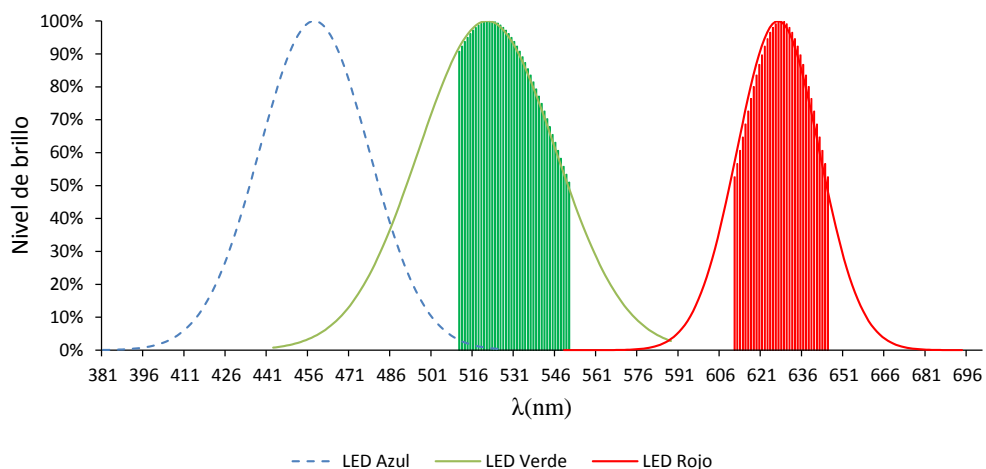


Figura 6.11. Espectro de emisión útil del LED RGB HV-5RGB25. [182]

Para poder realizar análisis fuera de ese rango, por ejemplo, entre 380 y 500nm, este tipo de diodos no sería adecuados, y por lo tanto, debería combinarse con un LED azul individual (con una longitud de onda pico de 460nm) para sustituir al que ya está presente en el LED-RGB. Este LED individual permitiría realizar mediciones fuera del rango de 510-645nm como se muestra en la Figura 6.9.

6.3 Extensión del rango de trabajo

En vista de los resultados obtenidos, se concluye que sí es posible calibrar la longitud de onda de cada uno de esos diodos individuales. No obstante, se puede plantear la pregunta de si es posible aprovechar el resto de longitudes de onda de una forma análoga a los LEDs de ancho espectral fijo.

Debemos tener en cuenta que el valor de transmitancia de una muestra, es la respuesta de las propiedades físicas y químicas de las sustancias presentes en la misma a una determinada longitud de onda. La mera combinación de dos o tres colores diferentes (grupos de longitudes de onda) nunca proporcionará la misma respuesta espectral que se alcanzaría con una específica longitud de onda. Es decir, no se pueden crear longitudes de onda artificiales simplemente combinando colores, aunque el ojo humano perciba lo contrario. Por ejemplo, la combinación del azul y el amarillo produce el verde, pero esto no significa que se estén emitiendo las longitudes de onda de 500 a 550nm. Es sólo un efecto óptico [193].

A pesar de lo anterior, para tratar de extender el rango de trabajo de un LED RGB más allá de sus longitudes de onda de pico, se opta por combinar los diodos individuales para genera una respuesta en las muestras que pueda correlacionarse con la obtenida a una cierta longitud de onda medida con un equipo basado en lámparas incandescentes.

Para lograr ese objetivo se implementó un sistema de generación visual de colores desarrollado por Dan Bruton de la Universidad de Texas A&M [195], que permite aproximar una cierta longitud de onda (desde el punto de vista visual) a una determinada combinación de rojo, verde y azul, tal y como se muestra la Figura 6.12.

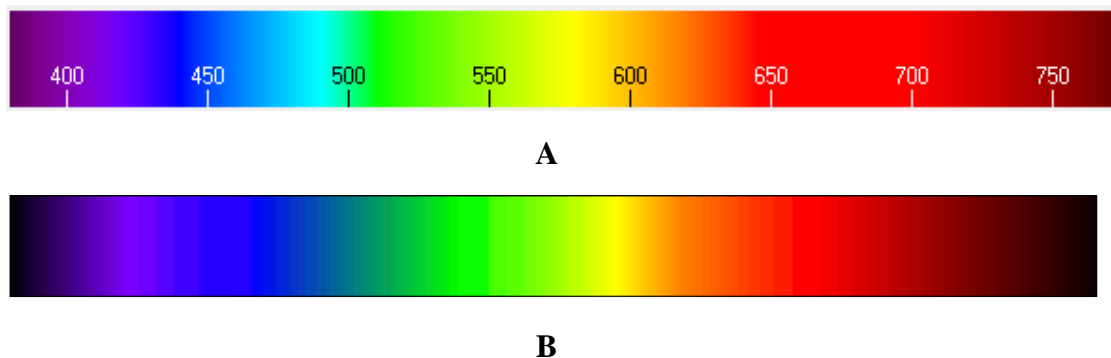


Figura 6.12. Espectro visible comparativo basado en: (A) Ecuaciones de equivalencia RGB – λ [195], (B) Espectro visible referencia [31].

Desde un punto de vista de renderización del color, ambos presentan una estructura muy similar.

Las pruebas llevadas a cabo demostraron que estas combinaciones de RGB proporcionadas por el modelo de renderizado de color de Dan Bruton, guardan una relación de tipo lineal con los valores de transmitancia obtenidos a las longitudes de onda que pretendemos “simular”. Esto se abordará con detalle a lo largo del presente capítulo, pero antes es necesario describir cómo el espectro de la Figura 6.12 A ha sido generado, dado que éste representa la base para la selección del espectro de trabajo.

La determinación de la correlación entre RGB y la longitud de onda Lambda (λ) se lleva a cabo mediante varios conjuntos de ecuaciones agrupadas por regiones espectrales [195], como se presenta en la Ecuación (6.1), donde R' , G' , y B' son los valores adoptados del rojo, verde y azul en el rango de 0-1, respectivamente, sin considerar el nivel de brillo, denominado con la letra Y:

$$\begin{aligned}
 (380nm \leq \lambda < 440nm) &\rightarrow \begin{cases} R' = -\left(\frac{\lambda - 440}{440 - 380}\right) \\ G' = 0 \\ B' = 1 \end{cases} \\
 (440nm \leq \lambda < 490nm) &\rightarrow \begin{cases} R' = 0 \\ G' = \left(\frac{\lambda - 440}{490 - 440}\right) \\ B' = 1 \end{cases} \\
 (490nm \leq \lambda < 510nm) &\rightarrow \begin{cases} R' = 0 \\ G' = 1 \\ B' = -\left(\frac{\lambda - 510}{510 - 490}\right) \end{cases} \\
 (510nm \leq \lambda < 580 nm) &\rightarrow \begin{cases} R' = \left(\frac{\lambda - 510}{580 - 510}\right) \\ G' = 1 \\ B' = 0 \end{cases} \\
 (580nm \leq \lambda < 645nm) &\rightarrow \begin{cases} R' = 1 \\ G' = -\left(\frac{\lambda - 645}{645 - 580}\right) \\ B' = 0 \end{cases} \\
 (645nm \leq \lambda < 781nm) &\rightarrow \begin{cases} R' = 1 \\ G' = 0. \\ B' = 0 \end{cases}
 \end{aligned} \tag{6.1}$$

El nivel de brillo, Y, es otro factor a tener en cuenta al calcular las correlaciones. Se define según el espectro que deseamos calcular (Ecuación (6.2)) [195]:

$$\begin{aligned}
 (380nm \leq \lambda < 420nm) &\rightarrow Y = 0.3 + 0.7 * \frac{(\lambda - 380)}{420 - 380} \\
 (420nm \leq \lambda < 701nm) &\rightarrow Y = 1 \\
 (701nm \leq \lambda < 781nm) &\rightarrow Y = 0.3 + 0.7 * \frac{(780-\lambda)}{780-700}.
 \end{aligned} \tag{6.2}$$

Para poder realizar la variación de la corriente, se utilizó un microcontrolador de 8bits, que puede proporcionar $2^8 - 1 = 256$ diferentes valores, es decir, entre 0 y 255, y por lo tanto, el máximo rango de trabajo en cuanto a nivel de brillo o intensidad con el que el LED funcionará es $I = 255$. Sin embargo, este procedimiento sería escalable a otras resoluciones. Por ejemplo, un driver de corriente de 10bits haría que tomara el valor de 1023.

Por último, las expresiones que permiten determinar la correlación entre λ y el valor RGB se muestran en la Ecuación (6.3) [195], siendo Gamma un coeficiente empírico al que se ha dado el valor de 0.8:

$$\begin{cases} (R' = 0) \rightarrow R = 0 \\ (R' \neq 0) \rightarrow R = I * (R' * Y)^{Gamma} \end{cases} \begin{cases} (G' = 0) \rightarrow G = 0 \\ (G' \neq 0) \rightarrow G = I * (G' * Y)^{Gamma} \end{cases} \begin{cases} (B' = 0) \rightarrow B = 0 \\ (B' \neq 0) \rightarrow B = I * (B' * Y)^{Gamma} \end{cases} \quad (6.3)$$

Donde R, G y B son los valores actuales necesarios para poder lograr un color específico a través de un LED RGB.

En la Tabla A5 (Anexo A), se recoge para las longitudes de onda comprendidas entre 380 y 780nm, las diferentes combinaciones de RGB obtenidas por medio de las ecuaciones anteriores. Tal y como se puede apreciar en la Tabla A5, salvo para las longitudes de onda comprendidas entre 645 a 700nm, existe una combinación única de RGB que permite generar una cierta longitud de onda desde un punto de vista visual. Sin embargo, la existencia de 75 valores cuyo valor RGB es el mismo, presentará una problemática que comentaremos más adelante en mayor profundidad.

Con el fin de facilitar la reproducción de los datos contenidos en la Tabla A5, se presenta el código desarrollado en C para su generación (Código 6.1).

```
void waveLengthToRGB(double Wavelength){
    double Gamma = 0.8;
    double IntensityMax = 255;
    double factor;
    double Red,Green,Blue;

    if((Wavelength >= 380) && (Wavelength<440)){
        Red = -(Wavelength - 440) / (440 - 380);
        Green = 0.0;
        Blue = 1.0;
    }else if((Wavelength >= 440) && (Wavelength<490)){
        Red = 0.0;
        Green = (Wavelength - 440) / (490 - 440);
        Blue = 1.0;
    }else if((Wavelength >= 490) && (Wavelength<510)){
        Red = 0.0;
        Green = 1.0;
        Blue = -(Wavelength - 510) / (510 - 490);
    }else if((Wavelength >= 510) && (Wavelength<580)){
        Red = (Wavelength - 510) / (580 - 510);
        Green = 1.0;
```

```

    Blue = 0.0;
}else if((Wavelength >= 580) && (Wavelength<645)){
    Red = 1.0;
    Green = -(Wavelength - 645) / (645 - 580);
    Blue = 0.0;
}else if((Wavelength >= 645) && (Wavelength<781)){
    Red = 1.0;
    Green = 0.0;
    Blue = 0.0;
}else{
    Red = 0.0;
    Green = 0.0;
    Blue = 0.0;
}

if((Wavelength >= 380) && (Wavelength<420)){
    factor = 0.3 + 0.7*(Wavelength - 380) / (420 - 380);
}else if((Wavelength >= 420) && (Wavelength<701)){
    factor = 1.0;
}else if((Wavelength >= 701) && (Wavelength<781)){
    factor = 0.3 + 0.7*(780 - Wavelength) / (780 - 700);
}else{
    factor = 0.0;
};

int rgb[3];
rgb[0] = Red==0.0 ? 0 : (int) round(IntensityMax * pow(Red *
factor, Gamma));
rgb[1] = Green==0.0 ? 0 : (int) round(IntensityMax * pow(Green
* factor, Gamma));
rgb[2] = Blue==0.0 ? 0 : (int) round(IntensityMax * pow(Blue *
factor, Gamma));

setColourRgb(rgb[0],rgb[1],rgb[2]);
}

```

Código 6.1. Función en C para el cálculo del valor RGB en función del valor de la longitud de onda.

Con el fin de proporcionar una representación más clara de la combinación de RGB para las distintas longitudes de onda, los datos de la Tabla A5 se recogen en la Figura 6.13.

Al ser el LED RGB controlado por un microcontrolador de 8bits, el rango de valores digitales que el convertidor analógico digital puede tomar es entre 0 y 255. Sin embargo, para aclarar la exposición de los resultados, el nivel de brillo que se muestra en la Figura 6.13 se expresó en función de la corriente aplicada a cada LED, en mA, teniendo en cuenta que el nivel más alto de intensidad del LED (255) se alcanza a 20mA.

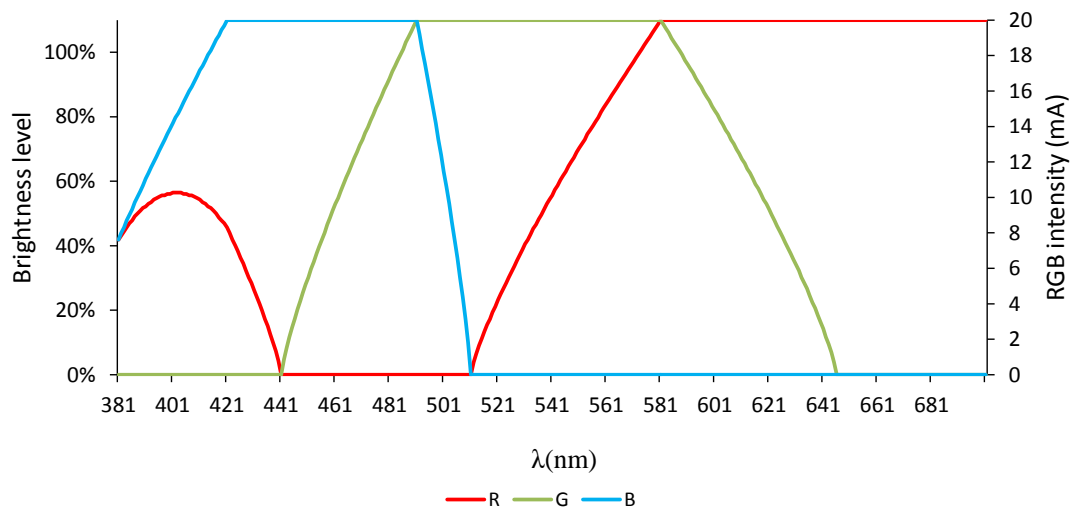


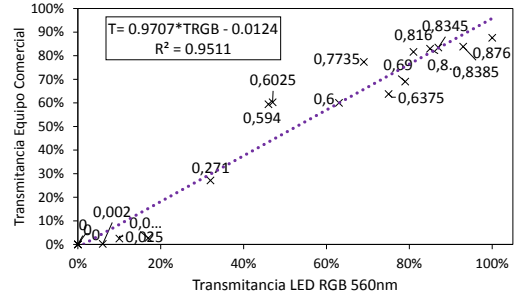
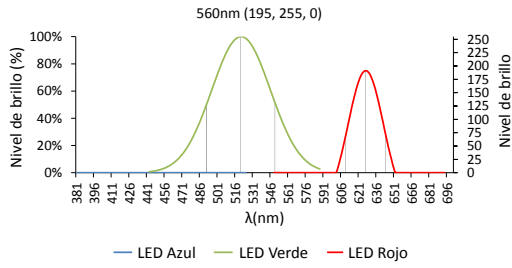
Figura 6.13. Valores heurísticos RGB para las longitudes de onda visibles según la representación del color.

Tal y como se ha descrito con anterioridad, los LEDs rojo y verde son capaces de explicar por sí mismos (sin combinación con otros) las longitudes de onda comprendidas entre 510 y 550nm y 610 y 645nm con una gran precisión, pero no son capaces de modelar otras regiones como 550-610nm, 645-700nm, y 380-550nm debido a que, o bien no emiten en esas regiones, o se alejan demasiado de sus longitudes de onda pico.

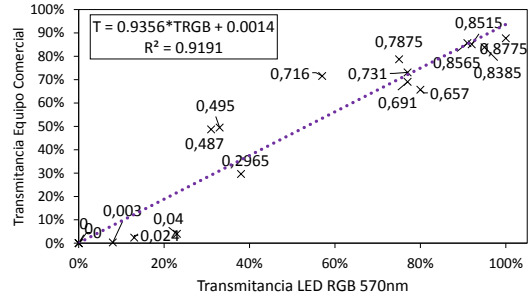
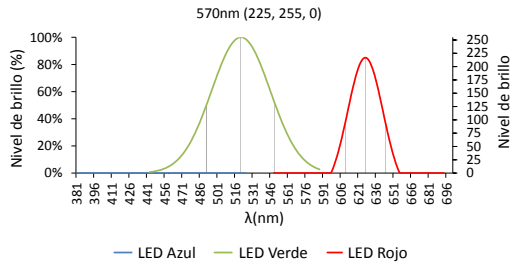
Aunque la combinación de colores nunca puede generar una “longitud de onda artificial”, sí se ha observado que permiten producir una respuesta en las muestras que puede correlacionarse con la longitud de onda que se pretende “simular”, aunque sólo dentro de un cierto rango, mucho más limitado que con LEDs de ancho espectral fijo.

Las pruebas realizadas demostraron que las longitudes de onda entre 550 y 610nm pueden ser ajustadas con la combinación de LEDs según estas ecuaciones.

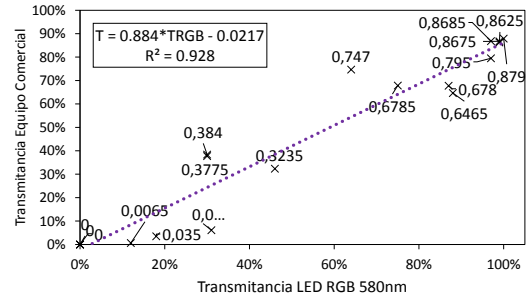
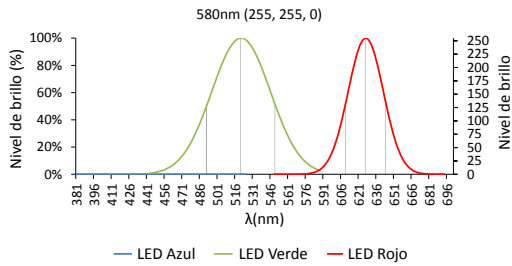
Para ilustrar esta afirmación con mayor claridad, en la Figura 6.14 se muestra el espectro de emisión de diferentes longitudes de onda simuladas, es decir, qué se está emitiendo cuando se desea renderizar una determinada longitud de onda (izquierda), y a la derecha, la correlación entre los valores de transmitancia que proporciona dicha combinación con respecto a los obtenidos con el uso de lámparas incandescentes, para las siguientes longitudes de onda: 560, 570, 580, 590 y 600nm y para las 20 muestras diseñadas en la Tabla A3 (S1-S20).



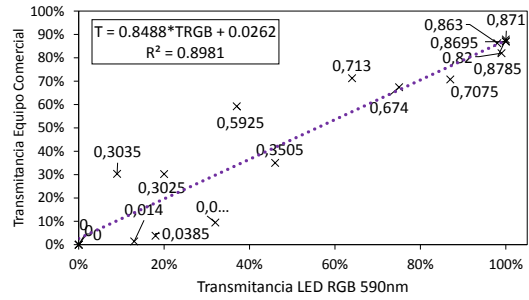
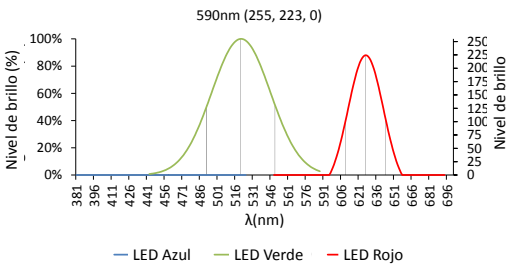
A



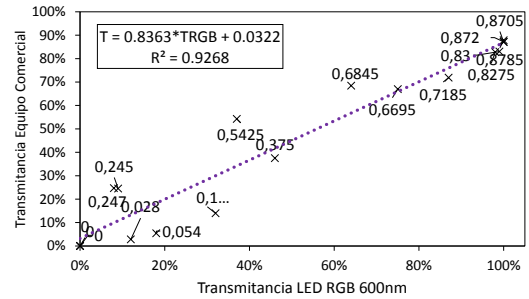
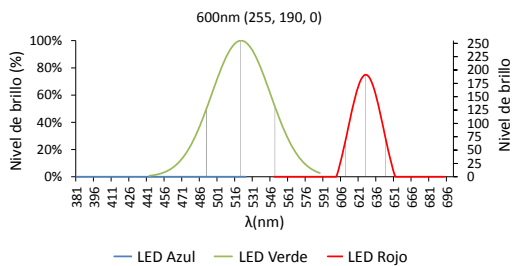
B



C



D



E

Figura 6.14. Combinación de longitudes de onda y correlación para: (A) 560, (B) 570, (C) 580, (D) 590, y (E) 600nm.

Como se puede observar, en todos los casos se ha obtenido un ajuste superior al 90%.

Con el fin de corroborar la validez de los modelos presentados en la figura anterior, la Figura 6.15 muestra una comparativa entre los valores de transmitancia obtenidos con el LED RGB y el equipo basado en lámparas incandescentes a 580nm.

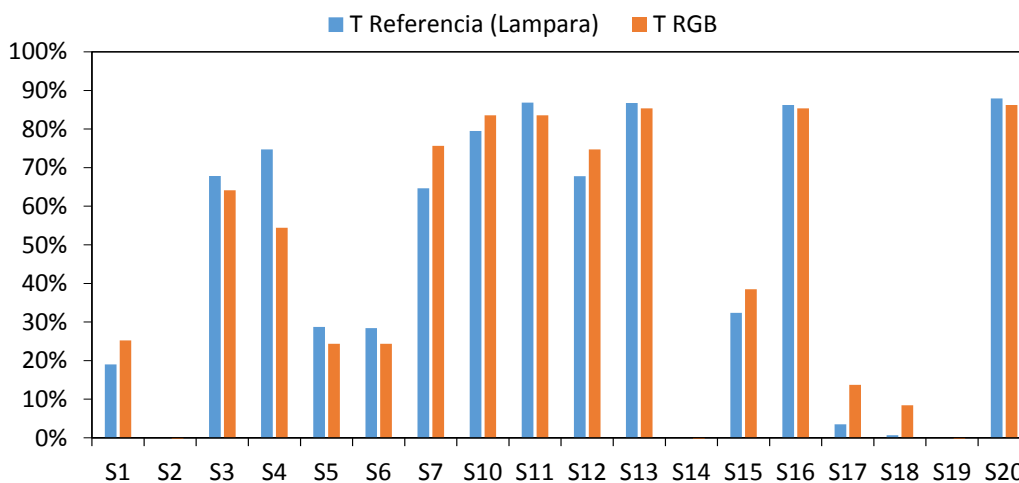


Figura 6.15. Comparativa entre el valor de transmitancia registrado con el LED RGB y el obtenido con la lámpara incandescente, a 580nm para 20 muestras analizadas.

Estos resultados tienen una gran trascendencia. A pesar que los LEDs no emiten longitudes de onda entre 550 y 610nm, la combinación de sus espectros es capaz de producir una respuesta en las muestras que puede relacionarse con los valores de transmitancia esperados en los equipos comerciales (referencia) siguiendo una distribución normal.

Sin embargo, las restantes longitudes de onda (mayores de 645nm y menores de 510nm) no pudieron ser modeladas con los LEDs trabajando ni de forma individual, ni tampoco con ninguna combinación de los mismos.

Para entender este comportamiento, realizamos el análisis de tres longitudes de onda cualesquiera, tomadas en diferentes regiones de color: 670, 505 y 460nm, los resultados se muestran en la Figura 6.16.

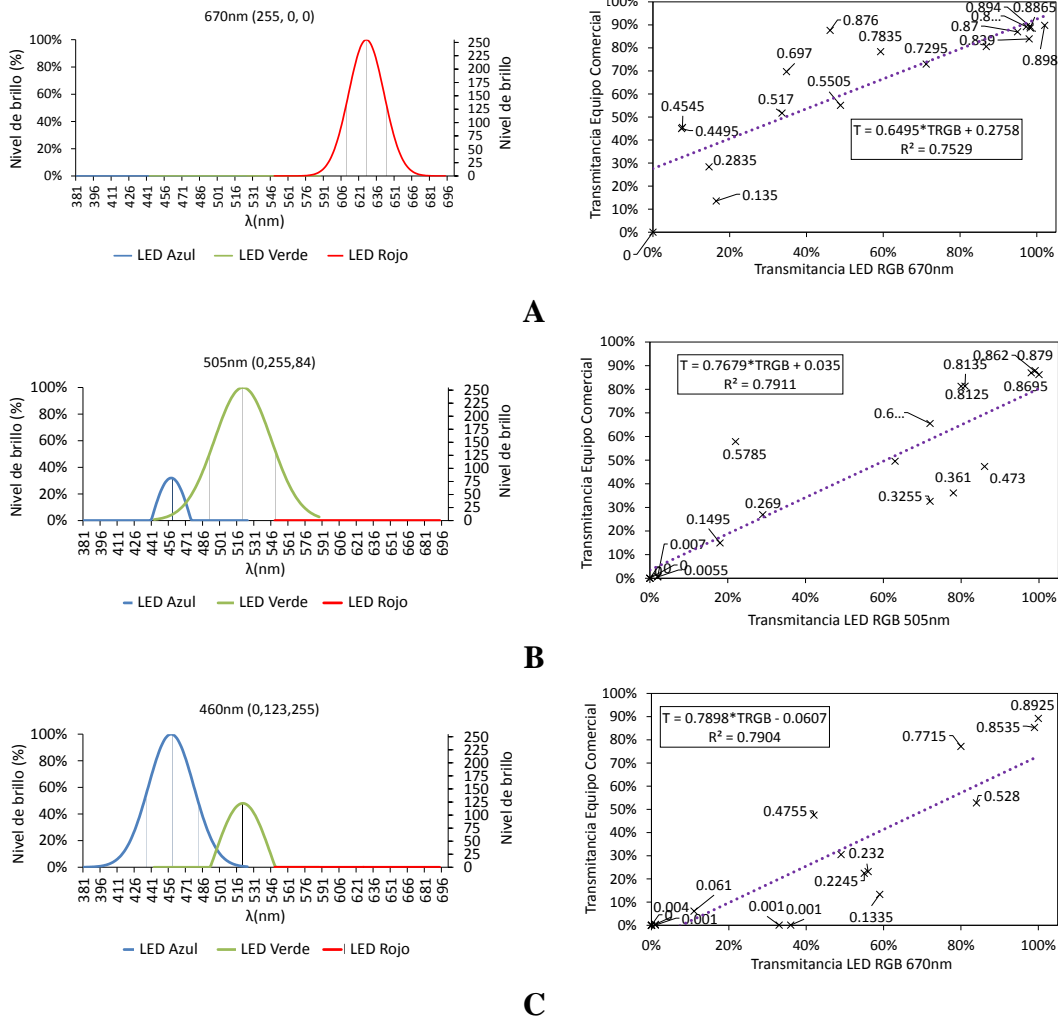


Figura 6.16. Combinación de longitudes de onda y correlación para: (A) 670, (B) 505, y (C) 460nm.

Como podemos ver, el ajuste es inferior al 80% en todos los casos. En 670nm (Figura 6.16 A), el diagrama de emisión muestra que sólo el LED rojo está encendido, es decir, no hay combinación de longitudes de onda pero, además, como 670nm está fuera del ancho espectral del LED rojo (Figura 6.6), la correlación es muy pobre.

Algo diferente parece explicar la falta de precisión de ajuste en el rango de 460 a 505nm. En este rango, hay una combinación de LEDs verdes y azules. Aunque la longitud de onda está en el ancho espectral cubierto por los LEDs verdes y azules, la presencia del LED azul lastra los resultados, debido a la baja calidad de sus materiales utilizados durante su construcción.

Por lo tanto, podemos concluir que un solo LED RGB, puede modelar sólo el espectro entre 510 y 645nm, con un de ajuste superior al 90%, y sin necesidad de utilizar ningún elemento óptico, de forma análoga a los LEDs de ancho espectral fijo descritos en el capítulo anterior.

6.4 Resultados del proceso de calibración

Para mostrar la idoneidad de los modelos de calibración, se muestran los resultados de los análisis realizados en el rango de 510 a 645nm para diferentes muestras. Para clarificar la presentación de los resultados, distinguimos entre las muestras utilizadas durante el proceso de calibración (Figura 6.17) y las utilizadas para comprobar la eficacia de los modelos (Figura 6.18), es decir, muestras que no han sido utilizadas durante el proceso de calibración (Tabla A3 de S21 a S48).

Como se puede ver, los gráficos muestran en la Figura 6.17 un buen nivel de ajuste dentro de la zona de 510-645nm. Fuera de esa zona, los valores de transmitancia difieren de los proporcionados por el equipo de referencia (lámpara incandescente).

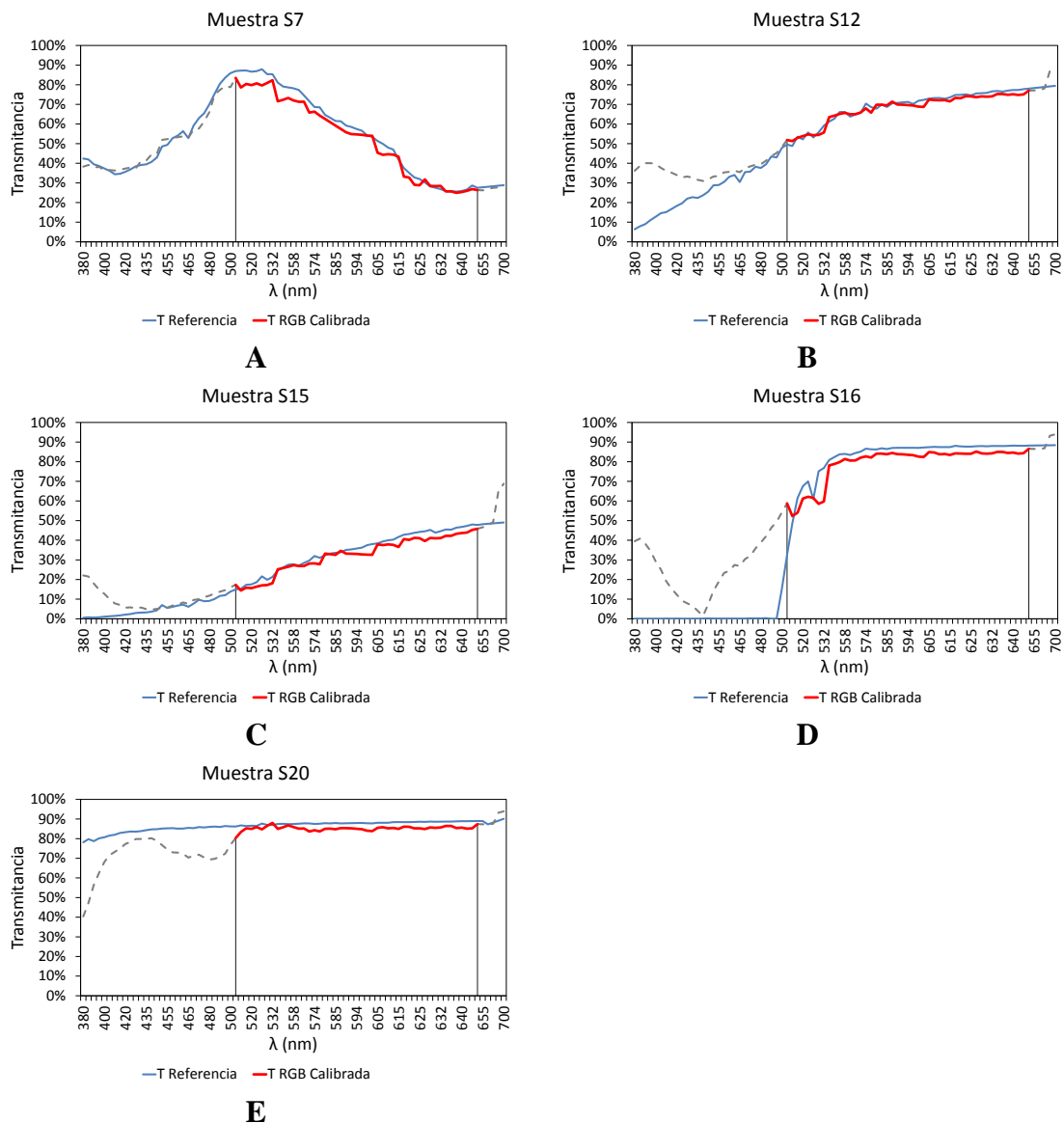


Figura 6.17. Comparativa entre los valores de transmitancia medidos con un equipo comercial basado en la lámpara incandescente (azul) y el dispositivo LED RGB desarrollado (rojo) después del proceso de calibración llevado a cabo dentro de 510-645nm, para las siguientes muestras: **(A)** S7, **(B)** S12, **(C)** S15, **(D)** S16, y **(E)** S20.

Sin embargo, para mostrar la idoneidad del modelo con muestras no utilizadas durante el proceso de calibración, la Figura 6.18 muestra los resultados obtenidos con muestras de aguas residuales (S21-S23 y S43-S48), correspondientes a diferentes puntos de una planta de tratamiento de aguas residuales, específicamente: Entrada de la planta de tratamiento de aguas residuales (S21, S43 y S44), en la salida del decantador primario (S22, S45 y S46) y en la salida de la planta de tratamiento (S23, S47 y S48).

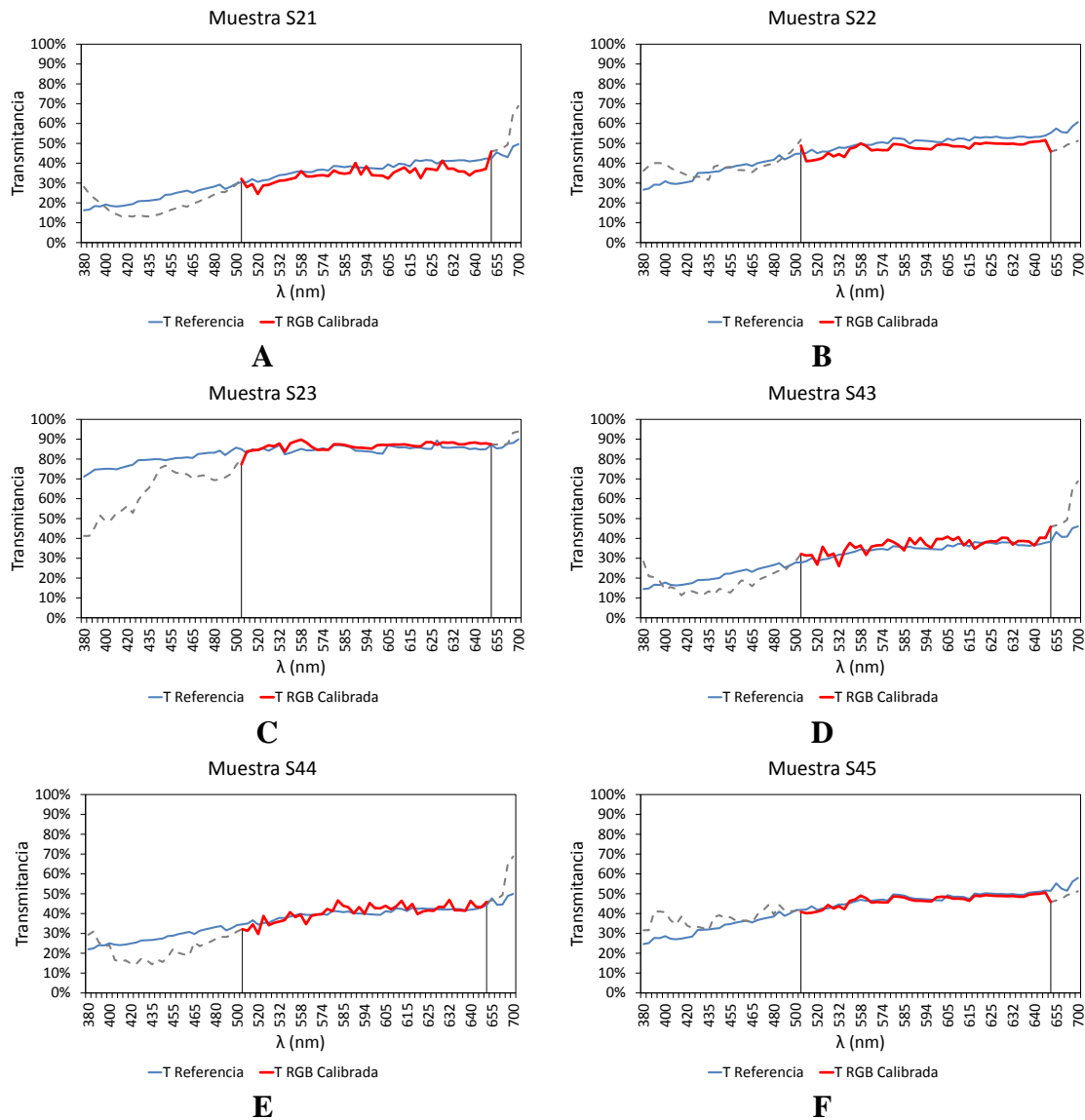


Figura 6.18. Comparativa entre los valores de transmitancia medidos con un equipo comercial basado en la lámpara incandescente (azul) y el dispositivo LED RGB desarrollado (rojo) después del proceso de calibración en el rango de 510-645nm, para las siguientes muestras: (A) S21, (B) S22, (C) S23, (D) S43, (E) S44, (F) S45, (G) S46, (H) S47, y (I) S48.

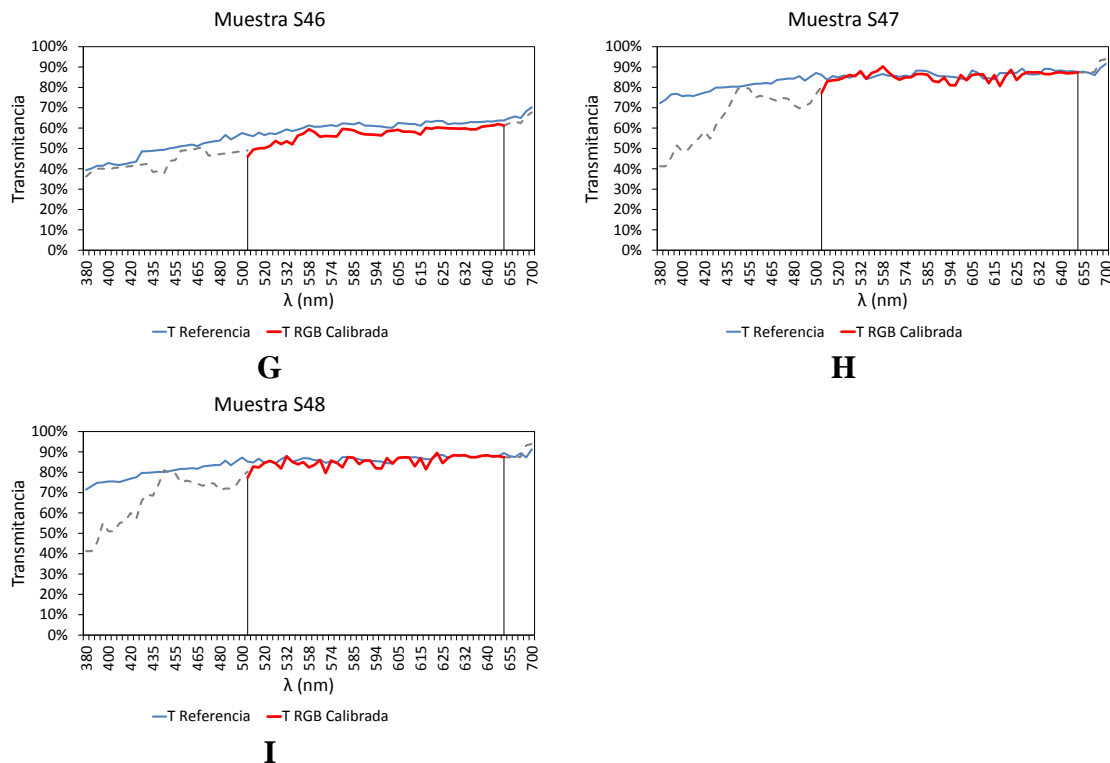


Figura 6.18 Continuación. Comparativa entre los valores de transmitancia medidos con un equipo comercial basado en la lámpara incandescente (azul) y el dispositivo basado en LED RGB desarrollado (rojo) después del proceso de calibración en el rango de 510-645nm, para las siguientes muestras: (A) S21, (B) S22, (C) S23, (D) S43, (E) S44, (F) S45, (G) S46, (H) S47, y (I) S48.

Como puede observarse en la Figura 6.19, entre 510nm y 645nm, todas las muestras analizadas presentan una alta precisión con respecto al equipo comercial basado en lámparas incandescentes.

Aunque el LED RGB no es capaz de modelar todo el espectro visible, proporciona un método sencillo para llevar a cabo un análisis espectrofotométrico sin el uso de elementos ópticos. La Tabla 6.1, proporciona más información sobre los tipos de aguas residuales que son adecuados para su análisis mediante un LED RGB, mostrando sus principales parámetros contaminantes:

Tabla 6.1. Características de las aguas residuales mostradas en la Figura 6.18.

Parámetros	Agua bruta (S21)	Decantador Primario (S22)	Agua tratada (S23)
DQO	763 mg/l	475 mg/l	52 mg/l
DBO ₅	500 mg/l	310 mg/l	9 mg/l
SST	304 mg/l	88 mg/l	14 mg/l
Fósforo (P)	9.1 mg/l	7.2 mg/l	2.5 mg/l
Nitrógeno total (NT)	74 mg/l	74 mg/l	16.6 mg/l
NO ₃ -N	0.5 mg/l	0.3 mg/l	10.3 mg/l
PH	7.59	7.5	7.56
Conductividad	2770 μS/cm	2590 μS/cm	2580 μS/cm
	Agua bruta (S43)	Agua bruta (S44)	Decantador Primario (S45)
DQO	1275 mg/l	908 mg/l	727 mg/l
DBO ₅	720 mg/l	480 mg/l	460 mg/l
SST	624 mg/l	558 mg/l	151 mg/l
Fósforo (P)	8.7 mg/l	9,4 mg/l	12.5 mg/l
Nitrógeno total (NT)	75 mg/l	59 mg/l	89 mg/l
NO ₃ -N	0.6 mg/l	0.8 mg/l	0.3 mg/l
PH	7.48	7.14	7.24
Conductividad	2590 μS/cm	2630 μS/cm	2600 μS/cm
	Decantador Primario (S46)	Agua tratada (S47)	Agua tratada (S48)
DQO	732 mg/l	47 mg/l	46 mg/l
DBO ₅	470 mg/l	6 mg/l	5 mg/l
SST	106 mg/l	11 mg/l	13 mg/l
Fósforo (P)	12.2 mg/l	0.8 mg/l	1.4 mg/l
Nitrógeno total (NT)	73 mg/l	59 mg/l	19.3 mg/l
NO ₃ -N	0.5 mg/l	3.8 mg/l	10.8 mg/l
PH	7.13	7.69	7.45
Conductividad	2930 μS/cm	2340 μS/cm	2170 μS/cm

Para cuantificar la precisión de los resultados proporcionados por el LED RGB después del proceso de calibración, se calcularon varios indicadores estadísticos: El error cuadrático medio [172], y el índice de error, Er, a través de las ecuaciones (6.4) y (6.5):

$$RMSD = \sqrt{\frac{1}{n} \sum_i^n (T_{medido_i} - T_{calculada_i})^2} \quad (6.4)$$

$$Er(\%) = \frac{\sum_i^n (T_{medido_i} - T_{calculado_i})}{\sum_i^n T_{medido_i}} * 100 \quad (6.5)$$

Donde Er es el índice de error (%); n es el número; y T_{medido} y $T_{calculado}$ son los valores de transmitancia obtenidos a través del equipo comercial [104] y nuestro diseño basado en la tecnología LED RGB, respectivamente.

El Tabla 6.2 muestra tanto el índice de error como el valor RMSD de las muestras que se recogen en la Figura 6.17 (utilizadas durante el proceso de calibración) y en la Figura 6.18 (muestras de prueba de aguas residuales), sólo para el espectro entre 510 y 645nm.

Para todas las muestras, el nivel de error fue siempre inferior al 6%, incluso en el caso de las muestras de aguas residuales. Esto demuestra que la tecnología LED RGB puede utilizarse para aplicaciones espectrofotométricas, logrando una respuesta bastante cercana al equipo comercial y permitiendo el desarrollo de equipos más rentables.

Tabla 6.2. Error cuadrático medio (RMSD) e índice de error entre 510-645nm.

	Muestra	RMSD	Er (%)
Muestras de calibración	S4	0.038	2.453
	S7	0.042	3.415
	S12	0.017	0.920
	S15	0.028	5.252
	S16	0.051	4.173
	S20	0.028	1.691
Muestras de prueba	S21	0.038	5.688
	S22	0.031	3.595
	S23	0.0231	-1.125
	S43	0.0297	-3.163
	S44	0.0257	-1.009
	S45	0.0125	0.8946
	S46	0.0408	3.7038
	S47	0.0217	0.5203
	S48	0.0222	0.7365

6.5 Conclusiones

En este capítulo se ha estudiado la idoneidad de los LED RGB como un reemplazo viable a las lámparas incandescentes.

Más allá de la posibilidad de modelizar la longitud de onda pico que componen los diodos que constituyen el LED RGB, se ha demostrado que la combinación de longitudes de onda puede producir una respuesta en las muestras que guarda relación con la longitud de onda que se pretende simular desde un punto de vista visual, aunque con un rango más limitado, entre 510 a 645nm pero con un índice de error inferior al 6% en todos los casos.

Esta metodología de calibración posibilita el desarrollo de equipos más sencillos y baratos para la caracterización de las aguas residuales, permitiendo así la implantación masiva de la red y sin necesidad de elementos ópticos o de someter a las muestras a pretratamientos.

Aunque no es posible utilizar la tecnología RGB para desarrollar equipos de espectrofotometría de espectro visible completo (380-700nm), el rango de 510 a 645nm

resulta adecuado para muchas aplicaciones, como es el cálculo de la medición de la cantidad de ozono en el agua [181].

Este tipo de sistemas puede ser muy útil para la detección de descargas no autorizadas, ya que el sistema puede situarse en cualquier entorno, vigilando las 24 horas del día. Por lo tanto, es una herramienta para combatir el fraude y puede contribuir a la protección del medio ambiente.

Capítulo 7

Modelos de estimación de la carga contaminante

7.1 Introducción

Los equipos y estudios presentados en capítulos anteriores permiten llevar a cabo un análisis espectral de las muestras de agua, es decir, medir la variación de sus valores de transmitancia y absorbancia a distintas longitudes de onda.

Aunque esta respuesta espectral guarda relación con la carga contaminante de las muestras, los equipos no son capaces por sí solos de determinar desde un punto de vista cuantitativo parámetros como la DQO o SST.

Para ello, es necesario el uso de modelos de estimación que, a partir de la información proporcionada por el equipo (información espectrofotométrica), a determinadas longitudes de onda, puedan estimar este tipo de parámetros sin necesidad de usar reactivos químicos o someter las muestras a pretratamientos.

A lo largo del presente capítulo, se describe los diferentes modelos generados para la estimación de los siguientes parámetros contaminantes a partir de la información espectrofotométrica: Demanda química de oxígeno (DQO), Demanda biológica de oxígeno a cinco días (DBO₅), Sólidos suspendidos totales (SST), Fósforo (P), Nitrógeno total (TN) y el Nitrógeno Nítrico (NO₃-N).

Para el desarrollo de los modelos de estimación [41], se han llevado a cabo una campaña de muestreo en la planta de tratamiento de aguas residuales de Cabezo Beaza (Región de Murcia, España), donde se realizaron unas 1.200 determinaciones analíticas en el laboratorio, pertenecientes a unas 200 muestras, durante el período comprendido entre junio de 2019 y abril de 2020, siendo las muestras tomadas tanto a la entrada de la planta (agua bruta), como a la salida de la misma (agua tratada), con el fin de generar modelos que permitan estimar de forma integral, cualquier tipo de agua presente en la misma.

Para la generación de estos modelos, se emplearon dos técnicas analíticas diferentes: Modelos de regresión lineal multivariable [43, 44] y Modelos basados en Algoritmos

Genéticos [196, 197], cuyas características y ámbito de aplicación serán abordadas a lo largo del capítulo.

7.2 Materiales y métodos

El elemento clave de cualquier análisis estadístico son los datos. La forma en que esos son obtenidos y sus características intrínsecas, condicionan la forma en la que estos pueden ser procesados [198].

En este epígrafe se abordará tanto la campaña experimental llevada a cabo para la obtención de los mismos, describiendo la forma en la que las muestras de agua residual han sido obtenidas, así como las características que dichas muestras presentan y que condicionaran el tipo de análisis estadístico a llevar a cabo.

7.2.1 Campaña experimental

Para poder determinar la carga contaminante de las aguas a partir de su información espectrofotométrica, no sólo es necesario conocer su respuesta espectral, sino también los resultados analíticos de dichas aguas llevados a cabo en laboratorio con el fin de encontrar las correlaciones entre ambos.

Por ese motivo, la campaña experimental fue llevada a cabo en la planta de tratamiento de Cabezo Beaza (Cartagena, Murcia), durante un periodo de 11 meses, donde las muestras fueron tomadas en dos puntos diferentes de muestreo:

- Agua residual afluyente a la entrada de la EDAR: **Agua bruta**
- Agua a la salida del decantador secundario y antes de llegar al tratamiento terciario: **Agua tratada**.

Respondiendo a los requerimientos de las campañas de muestreo de inspección por parte de la administración supervisora (Administración de Aguas Residuales de la Región de Murcia, ESAMUR [119]), las muestras se integraron, es decir, se tomaron homogéneamente durante 24 horas en un volumen de 5 litros, mediante una muestra acumulada de 200ml/h, la cual era recolectada cada día en torno a las 7:00 y analizadas en el laboratorio poco después.

Las pruebas realizadas en el laboratorio de la EDAR se ajustaron a los Métodos Estándar (SM) [199] y a la Organización Internacional de Normalización (ISO) [200], tal y como se describe en la Tabla 7.1. Los métodos fueron desarrollados por los miembros del Comité de Métodos Estándar (SMC) con la publicación mutua de la Asociación Americana de Salud Pública (APHA) [201], la Asociación Americana de Obras Hidráulicas (AWWA) [202] y la Federación del Medio Ambiente del Agua (WEF) [203].

Tabla 7.1. Variables analizadas durante la campaña de muestreo.

Variable	Descripción	Norma/Procedimiento de análisis
DBO ₅	Método respirométrico	SM 5210 D
DQO	Método del dicromato con Espectroscopia UV-Vis	ISO 6060:1989
SST	Sólidos sedimentables	SM 2540 F
NO ₃ ⁻ N	Método espectrométrico del dimetilfenol	ISO 7890-1
NT	Digestión con persulfato Espectroscopia UV-Vis	SM 4500-NC
TP	Método del ácido ascórbico complementado con espectroscopia UV-Vis	SM 4500-P B

Una vez las muestras eran caracterizadas por el laboratorio de la planta de tratamiento, esa información era introducida en un fichero Excel que la planta remitía el último día del mes, el cual era posteriormente cargada en el servidor web descrito en el Capítulo 4.

	Q. Entr.	pH	CONDUCT.	SST	DBO ₅	DQO	NH ₃ -N	NO ₃ -N	NTK	Ntotal	Ptotal
Día	(m ³ /día)	(Ud.)	(μS/cm)	(mg/l)	(mg/l)	(mg/l)	(mg/l)	(mg/l)	(mg/l)	(mg/l)	(mg/l)
1	21.341	7,27	2880	768	720	1460	56	0,4		68	9,9
2	20.925										
3	17.313	7,16	2980	664	780	1422	62	0,5		71	10,1
4	17.830										
5	18.615	7,36	3010	300	480	820	56	0,3		69	9,3
6	18.718	7,28	2890	324							
7	18.718	7,27	2950	512	720	1128	58	0,4		67	8,7
8	18.808										
9	18.017	7,37	2950	340	540	810	53	0,4		68	8,7
10	18.288										
11	15.213										
12	19.638	7,05	2010	768	620	1329	56	0,3		56	8
13	21.999	7,28	26	408	500	779	47	0,4		52	10,9
14	21.449	7,32	2890	340	600	873	48	0,4		71	8,8

Figura 7.1. Ejemplo de caracterización de muestras de agua a la entrada de la planta de tratamiento proporcionada por el laboratorio (Fichero Excel).

En la Figura 7.1 se muestra un ejemplo de esos ficheros Excel, donde sólo se muestran los parámetros relativos al agua de entrada para mayor claridad en la exposición de los datos. Para cada día del mes, se recoge la información tanto del agua de entrada de la planta (Bruta) como a la salida de la misma (Tratada), indicándose:

- Caudal
- Conductividad
- PH
- Sólidos en Suspensión Totales (SST)
- DQO
- DBO₅
- NO₃-N
- Nitrógeno Total
- Fósforo Total

De forma simultánea a la caracterización llevada a cabo por el laboratorio, se tomaba una muestra sin alterar de 2.5ml en una probeta estándar de espectrofotometría [100], la cual era introducida en el Espectrofotómetro Desktop descrito en el Capítulo 2, el cual, previa

introducción de la designación de la muestra por parte del usuario (que luego es utilizada para la casación con los datos proporcionados en el fichero Excel), procedía al análisis y envío de la información al servidor web.

Las muestras analizadas por el equipo no fueron sometidas a ningún pretratamiento, con el fin de poder generar modelos que permitan estimar la carga contaminante de muestras sin necesidad de tener que llevar a cabo ningún pretratamiento, con el fin de poder desarrollar un sistema que permitiera obtener una caracterización más rápida y de forma más sencilla.

A final de cada mes, la información contenida en el fichero Excel proporcionado por la planta de tratamiento, era subido de forma automática a la web por medio de la aplicación mostrada en la Figura 4.7. En base a los identificadores de las muestras definidos por la planta de tratamiento, la aplicación asociaba los resultados de la caracterización llevada a cabo por el laboratorio, con los datos de espectrofotometría medidos por el Espectrofotómetro Desktop.

Al término de la campaña de medición, se procedió a la descarga de los datos en un fichero CSV para su posterior análisis estadístico, obteniendo cerca de 200 muestras caracterizadas.

En la Figura 7.2 se muestra una porción del fichero CSV generado por la web, donde se indica para cada muestra:

- Identificador
- Parámetros contaminantes antes citados (DQO, DBO₅, SST, etc.)
- Valores de transmitancia y absorbancia para cada una de las 81 longitudes de ondas que utiliza el Espectrofotómetro Desktop para llevar a cabo el análisis entre 380-700nm.

ID	Nombre Muestra	DQO	DBO5	SST	NKT	P	PH	Cond	NT	NNO3	Transmitancia λ380	Transmitancia λ385
43	soluble bruta 26/06/19-E2	947	600	324	0	8.7	7.31	2750	76	0.7	0.635099	0.649628
44	dec. 1 26/06/19-E2	613	440	106	0	9.6	7.29	2700	75	0.7	0.335268	0.342685
46	tratada 26/06/19-E2	62	12	17	0	1.1	7.44	2330	15.4	7.9	0.696237	0.712217
50	tratada 27/06/19-E2	53	4	18	0	0.7	7.43	2330	14.3	8.5	0.687337	0.703106
51	Bruta 27/06/19-E2	1056	600	492	0	17.2	7.11	2700	68	0.5	0.143822	0.146698

Figura 7.2. Fichero CSV con la información espectrofotométrica medida con el Espectrofotómetro Desktop y los parámetros contaminantes medidos en el laboratorio de la planta de tratamiento.

7.2.2 Características de los datos

Las características inherentes al tipo de agua a analizar condiciona de forma significativa el tipo de estudio a llevar a cabo. Como es de esperar, las características del agua son muy diferentes en función del punto de recogida de las mismas, debido a que las concentraciones de materia orgánica e inorgánica varían de forma significativa entre ellas.

Esta carga contaminante guarda relación con la respuesta espectral de las muestras. Dado que la información espectrofotométrica va a ser utilizada como información de partida para la estimación de la carga contaminante, se hace necesario como paso previo al análisis de los datos, observar la relación entre ambos.

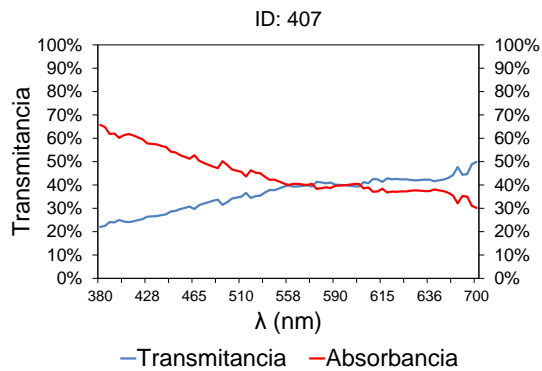
Con el fin de exponer con mayor claridad las características de los distintos tipos de agua analizadas y sus condicionantes durante la generación de los modelos, a continuación se va a proceder a exponer las características intrínsecas de las mismas a partir del análisis muestral llevado a cabo por el equipo desarrollado (Espectrofotómetro Desktop) así como por el laboratorio de la planta de tratamiento.

7.2.2.1 Agua bruta

En la Figura 7.3 se presentan cuatro muestras de agua tomadas mediante una muestra integrada durante 24 horas a la entrada de la planta de tratamiento, las cuales corresponden a diferentes periodos de tiempo.

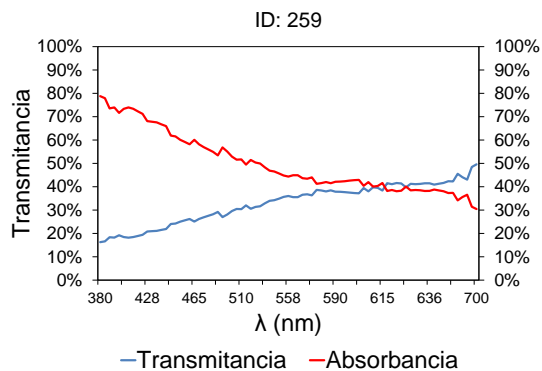
Como cabría esperar, a mayor carga contaminante, la curva de transmitancia será cada vez más baja, debido a que al existir un mayor número de sustancias en el agua (tanto orgánicas como inorgánicas), éstas representan un mayor obstáculo al paso de la luz. [204].

En general, los análisis llevados a cabo, han concluido que cuanto mayor es la carga de contaminantes, más fácil es encontrar correlaciones entre los datos de transmitancia/absorbancia y los contaminantes medidos en el laboratorio de las plantas de tratamiento de aguas residuales.



Contaminantes	Valor
DQO	763 mg/l
DBO ₅	500 mg/l
SST	304 mg/l
Fósforo (P)	9.1 mg/l
Nitrógeno Total (NT)	74 mg/l
NO ₃ -N	0.5 mg/l
PH	7.59
Conductividad	2770 μS/cm

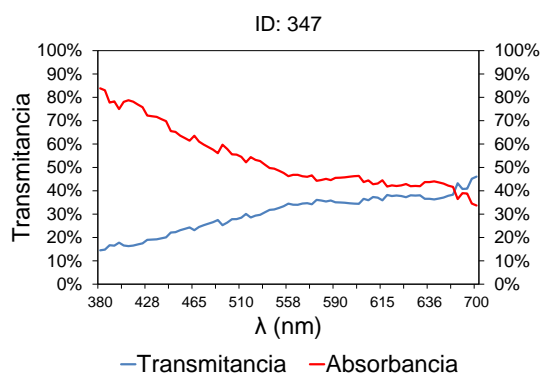
A



Contaminantes	Valor
DQO	908 mg/l
DBO ₅	480 mg/l
SST	558 mg/l
Fósforo (P)	9.4 mg/l
Nitrógeno Total (NT)	59 mg/l
NO ₃ -N	0.8 mg/l
PH	7.14
Conductividad	2630 μS/cm

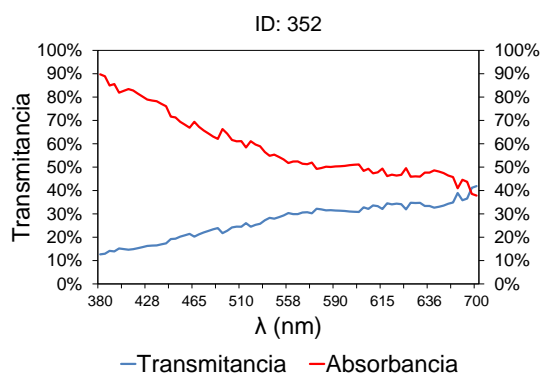
B

Figura 7.3. Datos espectrofotométricos y parámetros de contaminantes para muestras agua bruta.



Contaminantes	Valor
DQO	1275 mg/l
DBO ₅	720 mg/l
SST	624 mg/l
Fósforo (P)	8.7 mg/l
Nitrógeno Total (NT)	75 mg/l
NO ₃ -N	0.6 mg/l
PH	7.48
Conductividad	2590 μS/cm

C



Contaminantes	Valor
DQO	1108 mg/l
DBO ₅	240 mg/l
SST	892 mg/l
Fósforo (P)	5.6 mg/l
Nitrógeno Total (NT)	56 mg/l
NO ₃ -N	0.4 mg/l
PH	7.4
Conductividad	1894 μS/cm

D

Figura 7.3 Continuación. Datos espectrofotométricos y parámetros de contaminantes para muestras agua bruta.

No obstante, la cantidad de luz que logra atravesar la muestra varía en función de la longitud de onda irradiada. En zonas próximas al violeta, la transmitancia es más baja que en longitudes de onda próximas al rojo. Entre 380nm y 700nm, los gráficos muestran una pendiente ascendente, que es mucho más pronunciada entre 380 y 558nm, para luego volverse más horizontal.

Como puede observarse en la Figura 7.3, todas las muestras tienen el mismo gráfico de transmitancia a partir de 558nm, y en ningún caso el valor de transmitancia supera el 50%. Por lo tanto, para poder analizar la correlación entre los valores de transmitancia y los parámetros del contaminante, debemos prestar atención a la región comprendida entre 380 y 558nm que son las que presenta mayor variación espectral. Este aspecto es de gran relevancia, ya que limita la aplicación de técnicas de regresión lineal a un número más reducido de parámetros (longitudes de onda), limitando las capacidades de estimación del modelo, aspecto que se agrava aún más en el caso del agua tratada tal y como se comentará a lo largo del presente capítulo.

Tal y como se aprecia en las Figuras 7.3 A-B, una pequeña variación en el valor de transmitancia a 380nm, implica una gran variación en los valores de DQO, DBO₅, SST y TN, no siendo tan evidente en los demás parámetros. El hecho de pasar de un 20% de transmitancia (Figura 7.3 A) a un 17% (Figura 7.3 B) en 380nm, ha hecho que la DQO [204] aumente desde los 763mg/l a 908mg/l, es decir, un incremento del 20%. Sin embargo, la DBO₅ apenas ha cambiado, pero la cantidad de SST ha experimentado un aumento muy significativo, del orden del 80%, pasando de 304mg/l a 558mg/l.

Sin embargo, a un 15% de transmitancia (Figura 7.3 C a 380nm), el aumento es de más del 40% tanto para la DQO como para la DBO₅ respecto a los casos anteriores, con pocos cambios en los SST. A valores de transmitancia más bajos (Figura 7.3 D), la DQO deja de experimentar cambios significativos, al contrario de los SST que sí los experimenta.

Esta variación en la pendiente se debe a que la materia orgánica presente en el agua, tiende a ser más sensible a este tipo de longitudes de onda, por lo que suelen absorberla en mayor proporción que el resto. En longitudes de onda bajas, cercanas a los 380nm, la materia orgánica absorbe más radiación y por lo tanto menos luz puede pasar a través de la muestra (menor transmitancia). Así mismo, si la muestra fuera irradiada por longitudes inferiores a 380nm, la absorción sería aún mayor y, por lo tanto, la transmitancia sería menor, hasta un cierto punto.

A medida que las longitudes de onda se alejan de la zona ultravioleta/azul, la materia orgánica absorbe menos luz y el cambio de transmitancia es menos significativo. Por lo tanto, los parámetros que dependen del comportamiento biológico de la materia aumentan cuanto mayor es la cantidad de luz absorbida a determinadas longitudes de onda, como es el caso de la DQO, DBO₅, SST, entre otros, ya que cuanto mayor es la cantidad de organismos presentes en el agua, mayor es la cantidad de luz absorbida (en un determinado rango de longitudes de onda) y, por lo tanto, menor es la luz que pasa a través de las muestras.

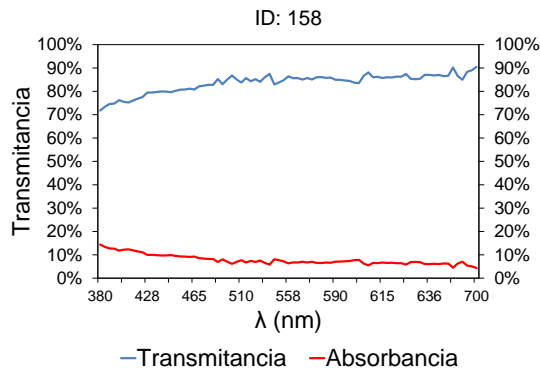
7.2.2.2 Agua tratada

En el caso del agua tratada (Figura 7.4 A-D), es decir, el agua obtenida a la salida de la planta de tratamiento, presenta unos valores de transmitancia mucho mayores que los que se muestran en la Figura 7.3, dado que la carga contaminante es baja tanto en términos de DQO, DBO₅ [205] como de SST, por lo que la luz logra pasar con mayor facilidad a través de las muestras en un rango más amplio de longitudes de onda.

El agua a la salida de la planta de tratamiento tiene un nivel de transmitancia muy alto, cercano al 90% entre 445 y 700nm, donde se comporta de forma horizontal, a diferencia del agua bruta (Figura 7.3) donde los valores de transmitancia parecían estabilizarse a partir de 558nm.

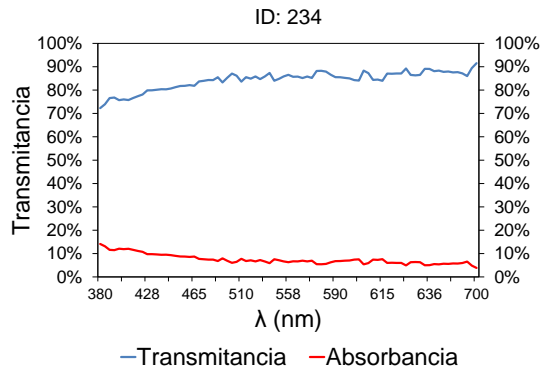
Además, los cambios en la pendiente de la transmitancia (gráfica azul), sólo son evidentes en la zona cercana al ultravioleta/azul, dado que es donde la materia orgánica es más sensible.

Esta propiedad es la responsable de la dificultad de llevar a cabo modelos de regresión lineal, dado que al basar estos su funcionamiento en los cambios de pendiente, resulta de difícil aplicación en el análisis de este tipo de aguas, tal y como se describirá de forma pormenorizada en el siguiente epígrafe.



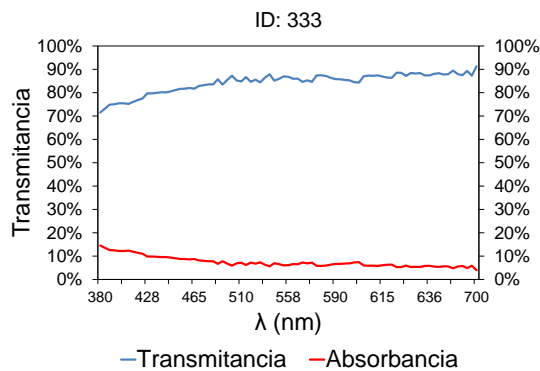
Contaminantes	Valor
DQO	52 mg/l
DBO ₅	9 mg/l
SST	14 mg/l
Fósforo (P)	2.5 mg/l
Nitrógeno Total (NT)	16.6 mg/l
NO ₃ -N	10.3 mg/l
PH	7.56
Conductividad	2580 μS/cm

A



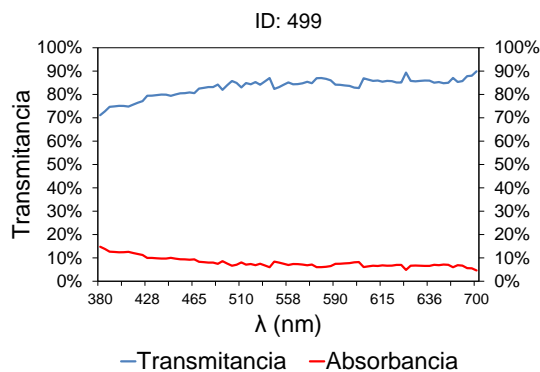
Contaminantes	Valor
DQO	40 mg/l
DBO ₅	0 mg/l
SST	7 mg/l
Fósforo (P)	1.3 mg/l
Nitrógeno Total (NT)	24 mg/l
NO ₃ -N	16.4 mg/l
PH	7.5
Conductividad	2570 μS/cm

B



Contaminantes	Valor
DQO	46 mg/l
DBO ₅	5 mg/l
SST	13 mg/l
Fósforo (P)	1.4 mg/l
Nitrógeno Total (NT)	19.3 mg/l
NO ₃ -N	10.8 mg/l
PH	7.45
Conductividad	2170 μS/cm

C



Contaminantes	Valor
DQO	47 mg/l
DBO ₅	6 mg/l
SST	11 mg/l
Fósforo (P)	0.8 mg/l
Nitrógeno Total (NT)	59 mg/l
NO ₃ -N	3.8 mg/l
PH	7.69
Conductividad	2340 μS/cm

D

Figura 7.4. Datos espectrofotométricos y parámetros de contaminantes para las muestras tratadas.

7.3 Modelos estadísticos de estimación de la carga contaminante

Debido a las características intrínsecas de cada tipo de agua, se han desarrollado modelos de estimación basados en dos técnicas diferentes: una en modelos de **regresión lineal** y otra en **algoritmos genéticos**.

Los modelos de regresión, han presentado una buena estimación para la carga contaminante contenida en al agua bruta, pero un bajo rendimiento en el agua tratada, debido a que este tipo de modelos se basan en la variabilidad de los datos. En una muestra de agua bruta, la respuesta espectral dentro del espectro visible sigue (por lo general) una respuesta (transmitancia) en forma de pendiente ascendente, mientras que en al agua tratada, al ser menor la cantidad de contaminantes presentes en la misma, la luz logra pasar con mayor facilidad en un mayor número de longitudes de onda, por lo que la respuesta tiende a ser más horizontal, sin fluctuaciones significativas [41]. Por ese motivo, los modelos de regresión lineal no han resultado válidos para realizar estimaciones en este tipo de aguas.

Con el fin de solventar esta problemática, se hacía necesario implementar una técnica de procesamiento de datos que permitiera extraer patrones contenidos en los mismos, aunque las fluctuaciones fueran mínimas, como sucede en el agua tratada. Por ese motivo, se han desarrollado modelos basados en algoritmo genético [196, 197, 206], donde el carácter evolutivo de los mismos, hacen que puedan acabar detectando estas variaciones tras la experiencia de modelos de generaciones anteriores.

A continuación se describe ambos tipos de modelos de cálculo, ilustrando sus características y limitaciones.

7.3.2 Modelos de regresión lineal

Como primera aproximación a la generación de modelos de estimación, se ha optado por el desarrollo de modelos de regresión lineal multivariable (MLR), debido a su simplicidad de cálculo y ejecución, aspecto clave si tenemos en cuenta que estos modelos deben ser embebidos en los prototipos desarrollados, los cuales han sido descritos en las secciones anteriores.

Estos modelos han sido calculados mediante el software SPSS Statistics desarrollado por IBM, utilizando un ajuste basado en mínimos cuadrados parciales [207]. De forma sintetizada, el proceso seguido para la estimación de los modelos se recoge en la Figura 7.5.

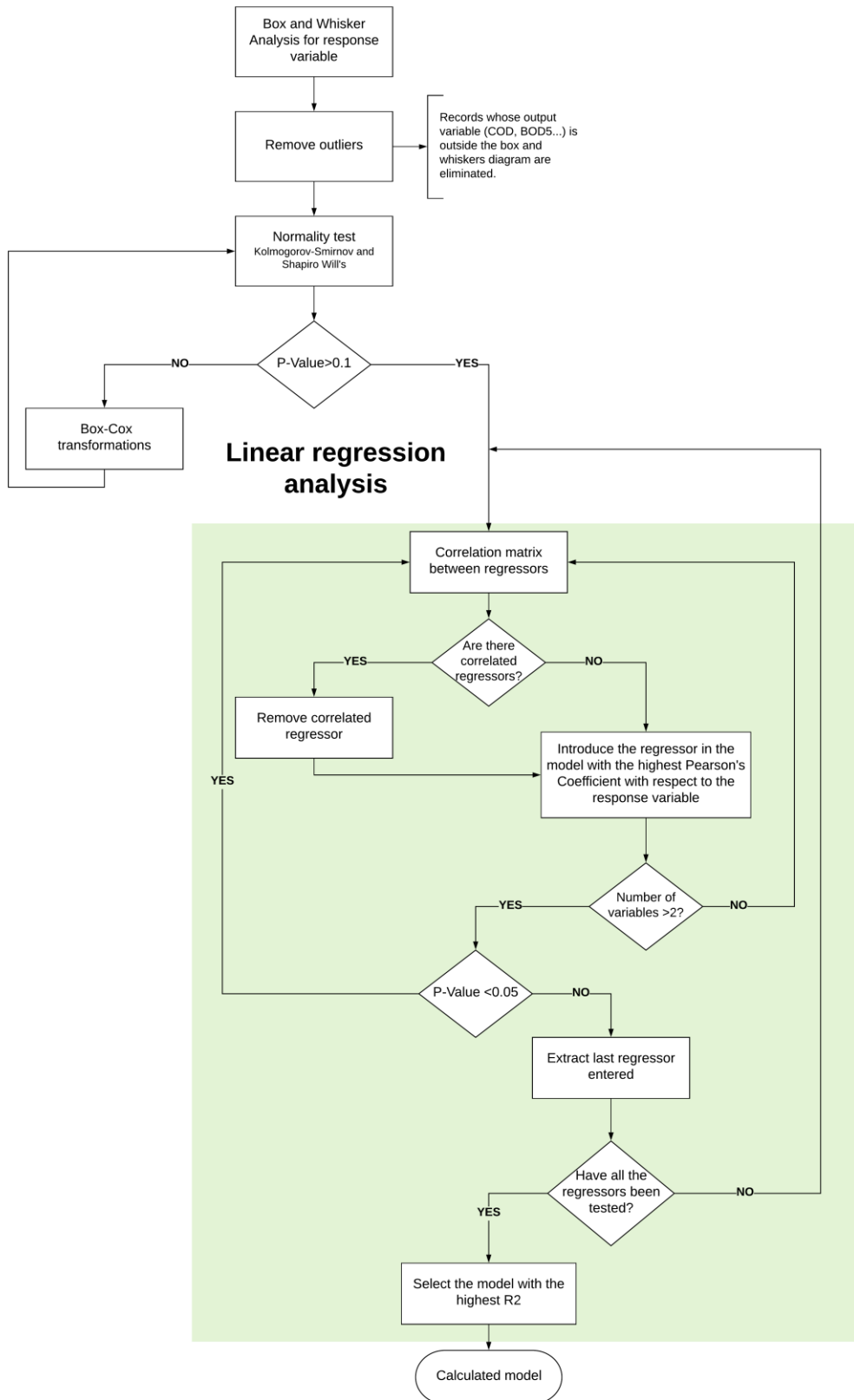


Figura 7.5. Esquema de cálculo de los modelos de regresión lineal.

Tal y como se muestra en la Figura 7.5, como paso previo a cualquier análisis estadístico, los datos han sido preprocesados para eliminar la existencia de datos atípicos que puedan lastrar los modelos, lo cual ha llevado a cabo mediante un diagrama de cajas y bigotes.

Una vez que los datos siguen una distribución normal, se utilizan las herramientas de SPSS para realizar el análisis, seleccionando la opción de cálculo "Paso a paso" [208], que permite modelos de cálculo más optimizados. Por lo tanto, las longitudes de onda (regresores) presentes en cada uno de los modelos fueron seleccionados automáticamente por SPSS de acuerdo con la siguiente metodología: El proceso comienza introduciendo el regresor cuyo P-Valor es el más alto dentro del rango 0,05 y 0,1 (criterios de entrada y salida, respectivamente). En la siguiente interacción, SPSS reintroduce el regresor con el P-Valor más alto (dentro del rango) y luego reevalúa el modelo para comprobar si alguno de los regresores introducidos ya no es significativo y/o si hay multicolinealidad en el modelo, es decir, que hay regresores correlacionados entre sí en el modelo. Este proceso se repite con todas las combinaciones posibles de regresores.

Una vez que SPSS ha calculado los modelos, se seleccionan aquellos cuyo coeficiente de determinación R-cuadrado (\bar{R}^2) es mayor y se comprueba que el modelo no incluye variables correlacionadas, comprobando que el factor de inflación de la varianza es menor que 7.

La condición *sine qua non* para poder implementar modelos de regresión lineal, es que estos sigan una distribución normal, es decir, que las pruebas de normalidad de Kolmogorov-Smirnov y Shapiro-Wilk [62, 63] proporcionen un P-Valor mayor que 0,1 (con un intervalo de confianza del 90%). En caso contrario, se hace necesario llevar a cabo una transformación de los datos con el fin de lograr que estos sigan una distribución normal, pudiendo optar por realizar la transformación Box-Cox [209] que proporciona una orientación sobre el tipo de transformación que se debe implementar para normalizar los datos. No obstante, no siempre es posible.

En un inicio, los esfuerzos fueron encaminados a la obtención de un modelo único que permitiera estimar cada uno de los parámetros contaminantes tanto en el agua bruta como tratada. Sin embargo, el grado de variabilidad entre los dos subconjuntos de datos es tan alto, que el dataset resultante no sólo no sigue una distribución normal, sino que tampoco es posible su normalización, tal y como se muestra en la Figura 7.6, a pesar de haber eliminado los valores atípicos.

Cada uno de los histogramas que se muestran en la Figura 7.6 contienen dos zonas diferenciadas: Una zona (derecha) que tiene una distribución aproximadamente normal, que corresponde a los datos del agua bruta, y por otro lado (izquierda), una clase o clases dominantes en cuanto a frecuencia, que corresponde a las muestras de agua tratada. Esto es especialmente visible en la Figura 7.6 A, B, C y F.

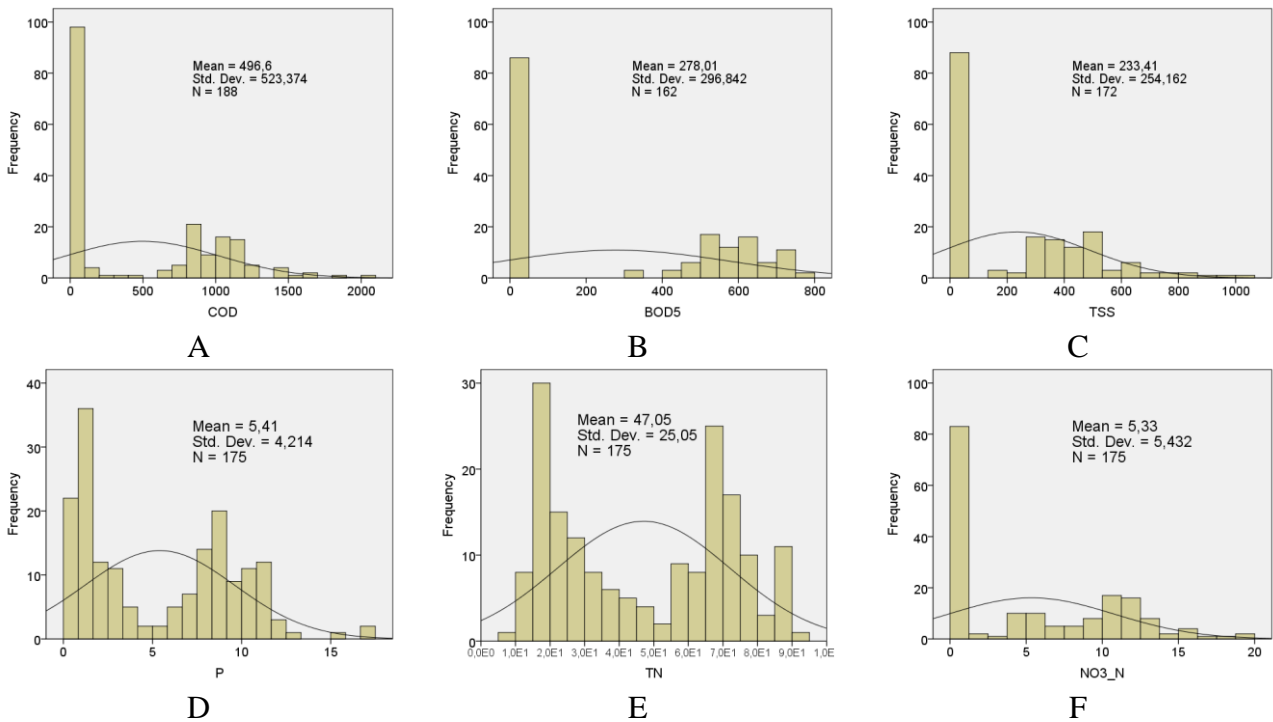


Figura 7.6. Histograma con combinación de muestras de agua bruta y tratada. (A) DQO, (B) DBO₅, (C) SST, (D) P, (E) NT and (F) NO₃-N.

Las muestras de agua con niveles de DQO y DBO₅ inferiores a 55mg/l y 15mg/l (agua tratada), muestran una respuesta espectral casi horizontal dentro del rango de 380-700nm (Figura 7.4). Esto hace que cambios en la DQO y DBO₅ medida en el laboratorio (por debajo de esos umbrales), no se vea reflejado en cambios significativos de la respuesta espectral. Es decir, si una muestra tiene un valor de DQO de 22mg/l y otra tiene un valor de 44mg/l, a pesar de que una duplique el valor de la otra, la respuesta espectral será prácticamente la misma, lo que hace que la variabilidad de los datos de las muestras de agua tratada sea mucho más significativa que en las muestras de agua bruta, en las que pequeños cambios en los niveles de los parámetros contaminantes se traducen en un cambios mucho más significativos en la respuesta espectral.

Por esta razón, los modelos de regresión lineal no son válidos para el agua tratada, ya que es imposible la normalización de los mismos.

Otros parámetros, como es el caso del nitrógeno total (Figura 7.6 E), permite observar que cómo los subconjuntos de datos (agua bruta y tratada) presentan distribuciones normales de forma individualizadas, no siendo así cuando se combinan entre sí.

Por lo tanto, hemos centrado la aplicación de estos modelos sólo en el agua bruta, donde los datos sí siguen una distribución normal.

Sin embargo, seguir una distribución normal no significa que los modelos calculados sean capaces de estimar la carga de contaminación con buena precisión. De hecho, los estudios realizados han demostrado que sólo es posible aplicar este tipo de modelo para el cálculo de la DQO, DBO₅ y SST relativos al agua bruta. El resto de los parámetros, es decir: P, TN y NO₃-N no pueden ser calculados con este método, ya que en la mayoría de los casos, o bien los datos no podían ser estandarizados o el modelo resultante tenía un bajo

nivel de correlación (inferior al 50%), no resultando válido para el objetivo de este trabajo de investigación.

La información relativa a las pruebas de normalidad llevadas a cabo para el agua bruta se recoge en la Tabla 7.2.

Tabla 7.2. Pruebas de normalidad relativas al agua bruta.

Parámetros	Muestras totales	Atípicos	Muestras válidas	Test de normalidad	
				Kolmogorov-Smirnov	Shapiro-Wilk
DQO	108	7	101	0.2*	0.376
DBO ₅	108	22	86	0.2*	0.172
SST	108	16	92	0.2*	0.043
P	108	2	106	0.2*	0.468
PH	108	1	107	0.2*	0.715
NO3-N	108	11	97	0	0

* Es el límite inferior de la significación

Los diferentes modelos de regresión lineal obtenidos para el cálculo de la DQO, DBO₅ y SST para las aguas residuales (agua bruta) se muestran a continuación, donde para poder validar los modelos calculados, los datos se dividieron en dos grupos: Datos de entrenamiento (utilizados para el desarrollo de los modelos) y datos de prueba (para la validación de los mismos). Estos datos se dividieron de forma aleatoria, con proporciones del 66% y el 34%, respectivamente.

7.3.2.1 Demanda Química de Oxígeno (DQO)

En la Ecuación (7.1) se muestra al modelo de regresión lineal calculado para el caso de la demanda química de oxígeno correspondiente al agua bruta. Este modelo ha presentado una bondad en el ajuste del 77.4% para los datos de entrenamiento. De las 108 muestras de agua bruta iniciales, 101 fueron utilizadas tras la eliminación de datos atípicos, y de estas, 69 fueron empleadas para desarrollar el modelo, mientras que el resto fueron usadas para el proceso de validación del mismo.

$$\begin{aligned}
 DQO_{(mg/l)} &= c_0 - c_1 \cdot T_{380} + c_2 \cdot T_{580} - c_3 \cdot T_{555} + c_4 \cdot T_{521} \\
 \overline{R^2} &= 77,4\%
 \end{aligned}
 \tag{7.1}$$

$$\begin{aligned}
 c_0 &= 844.247657 \\
 c_1 &= 1752.845 \\
 c_2 &= 5665.418 \\
 c_3 &= 7189.785 \\
 c_4 &= 8046.775
 \end{aligned}$$

Las estimaciones proporcionadas por el modelo son recogidas en la Figura 7.7, donde se muestra una comparativa entre los valores de DQO obtenidos en la planta de tratamiento de aguas residuales (azul), los valores de DQO proporcionados por el modelo que se muestra en la Ecuación (7.1), tanto con los datos de entrenamiento (rojo) como con los datos de prueba (amarillo), es decir, datos que no fueron utilizados para calcular el modelo.

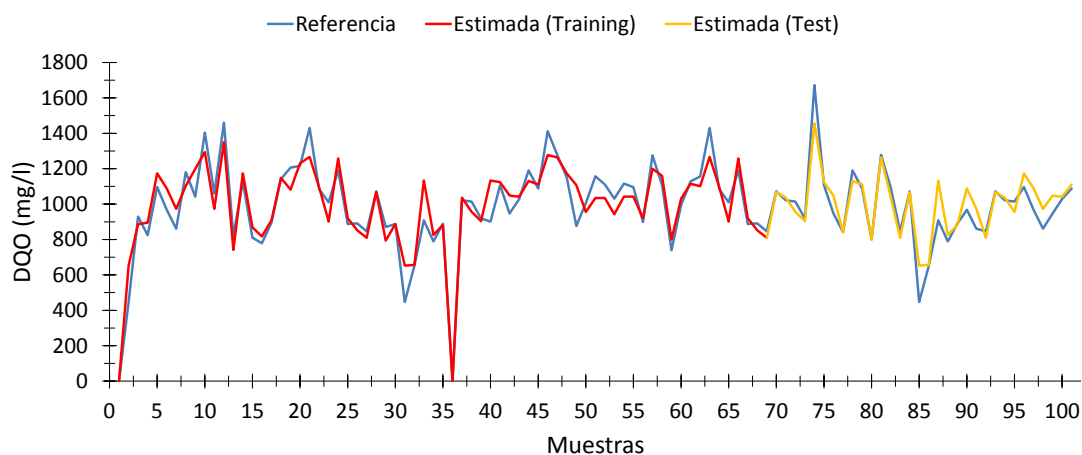


Figura 7.7. Comparación entre los valores de DQO medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos por el modelo de regresión lineal multivariante de la Ecuación (7.1), para el agua bruta.

Como se puede observar, las estimaciones se ajustan de una forma adecuada a los valores de referencia, aunque con ciertas discrepancias en los valores extremos.

Cada uno de los regresores utilizados en el modelo es capaz de explicar por sí mismo una parte de la demanda química de oxígeno. En la Figura 7.8 se muestra un diagrama de dispersión que denota el grado de correlación entre cada uno de los regresores utilizados en el modelo (Ecuación (7.1)) con respecto a la DQO.

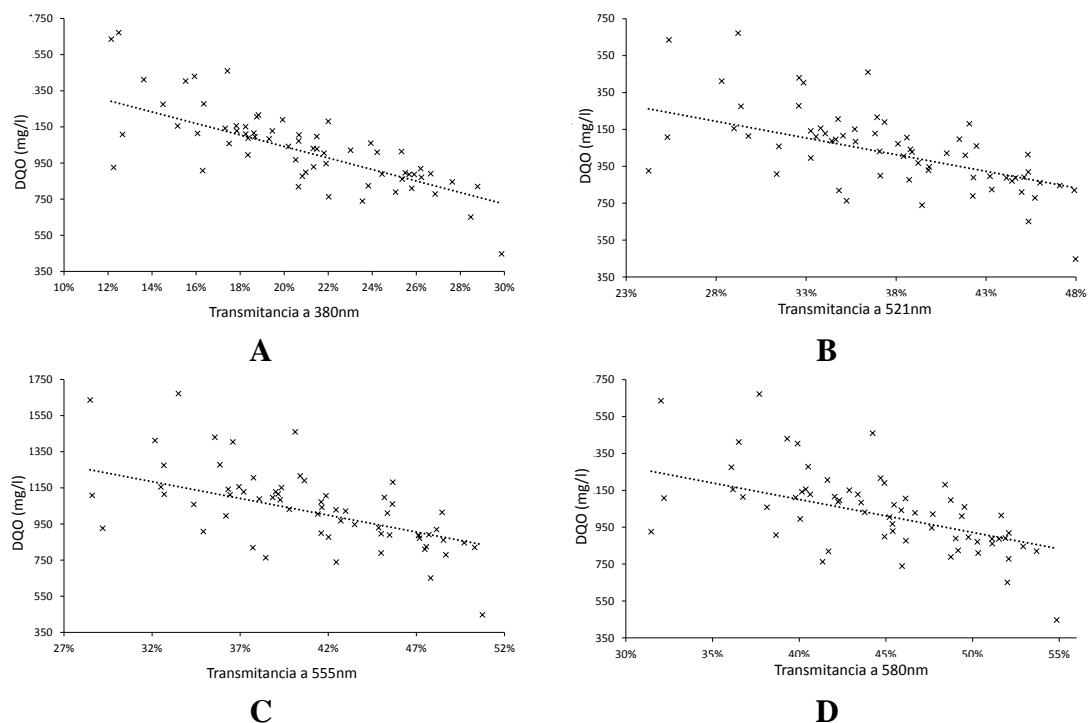


Figura 7.8. Gráfica de dispersión de las variables utilizadas en el modelo MLR para calcular la DQO con respecto a: (A) Transmittancia a 380nm, (B) Transmittancia a 521nm, (C) Transmittancia a 555nm y (D) Transmittancia a 580nm.

Como se puede observar, todos los regresores presentan una correlación similar con la variable respuesta, lo que denota que todos presentan una importancia similar en el cálculo del modelo.

7.3.2.2 Demanda biológica de oxígeno a 5 días (DBO₅)

De forma análoga a la DQO, la DBO₅ presenta un modelo de regresión lineal que involucra pocos regresores tal y como se muestra en la Ecuación (7.2). Este modelo proporcionó una bondad de ajuste del 61,9% para los datos de entrenamiento, más bajo comparado con el modelo anterior. El número de muestras utilizadas para el cálculo del modelo fue de 86, de un total de 108 muestras, después de eliminar los valores atípicos, de modo que se utilizaron 70 muestras en el desarrollo del modelo, mientras que las muestras restantes se utilizaron para su validación.

$$\begin{aligned}
 DBO_{5 (mg/l)} &= c_0 - c_1 \cdot A_{515} + c_2 \cdot A_{425} - c_3 \cdot T_{555} & c_0 &= 2171.855 \\
 \overline{R^2} &= 61.9\% & c_1 &= 7898.15 \\
 & & c_2 &= 4755.737 \\
 & & c_3 &= 2906.184
 \end{aligned}
 \tag{7.2}$$

En la Figura 7.9 se muestra una comparación entre los valores de DBO₅ proporcionados por el modelo de la Ecuación (7.3) (rojo, el conjunto de datos utilizado para construir el modelo, y amarillo, el conjunto de datos de prueba) y los valores obtenidos en la planta de tratamiento de aguas residuales en laboratorio (azul).

Como se observa, el modelo presenta desviaciones en la mayoría de las muestras, lo cual denota la limitación de este tipo de técnicas.

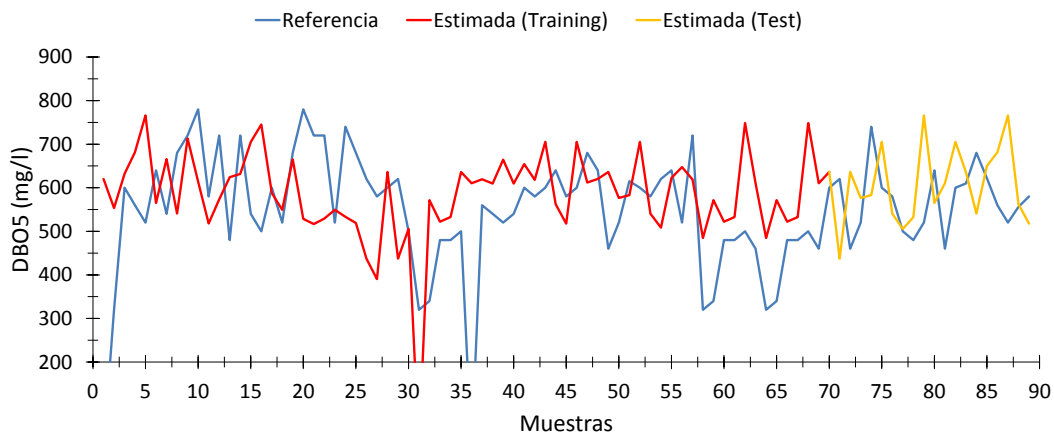


Figure 7.9. Comparación entre los valores de DBO₅ medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos por el modelo de regresión lineal multivariante de la Ecuación (7.2), para el agua bruta.

De forma análoga al caso anterior, en la Figura 7.10 se muestran las correlaciones entre los regresores y la variable respuesta.

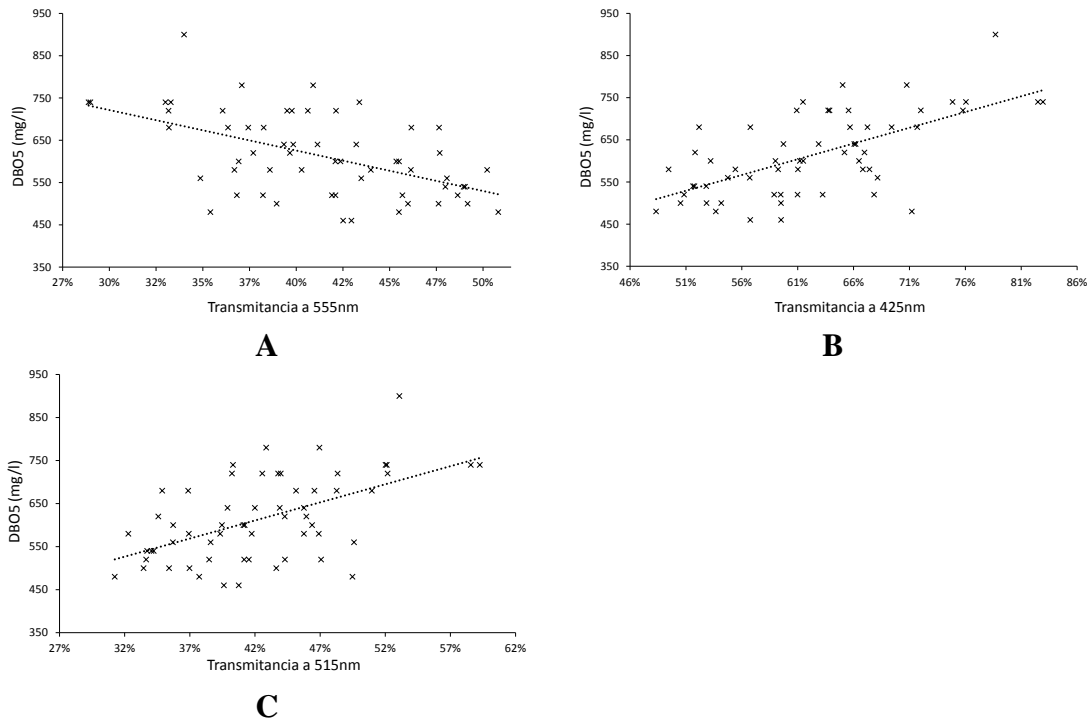


Figura 7.10. Diagrama de dispersión de las variables utilizadas en el modelo de la ecuación (7.2) para calcular la DBO₅ con respecto a: (A) Transmittancia a 555nm, (B) Absorbancia a 425nm y (C) Absorbancia a 515nm.

7.3.2.3 Sólidos suspendidos totales (SST)

En la Ecuación (7.3), se muestra el modelo de regresión lineal calculado para la estimación de los SST, el cual proporciona una bondad de ajuste del 72,2% para los datos de entrenamiento. El número de muestras utilizadas para el cálculo del modelo fue de 92, de un total de 108 muestras, después de eliminar los valores atípicos, de modo que se utilizaron 69 muestras en el desarrollo del modelo, mientras que las muestras restantes se utilizaron para probarlo.

$$SST_{(mg/l)} = c_0 - c_1 \cdot T_{380} - c_2 \cdot A_{656}$$

$$\overline{R^2} = 72.2\%$$

$$c_0 = 2428.586$$

$$c_1 = 5060.755 \quad (7.3)$$

$$c_2 = 2928.048$$

La Figura 7.11 muestra una comparación entre los valores de SST proporcionados por el modelo en la Ecuación (7.3) (rojo, para los datos utilizados para construir el modelo, y amarillo, para los datos de prueba), así como los valores obtenidos en la planta de tratamiento de aguas residuales (azul). Como puede verse, en términos generales, el modelo se ajusta bastante bien a los valores de referencia.

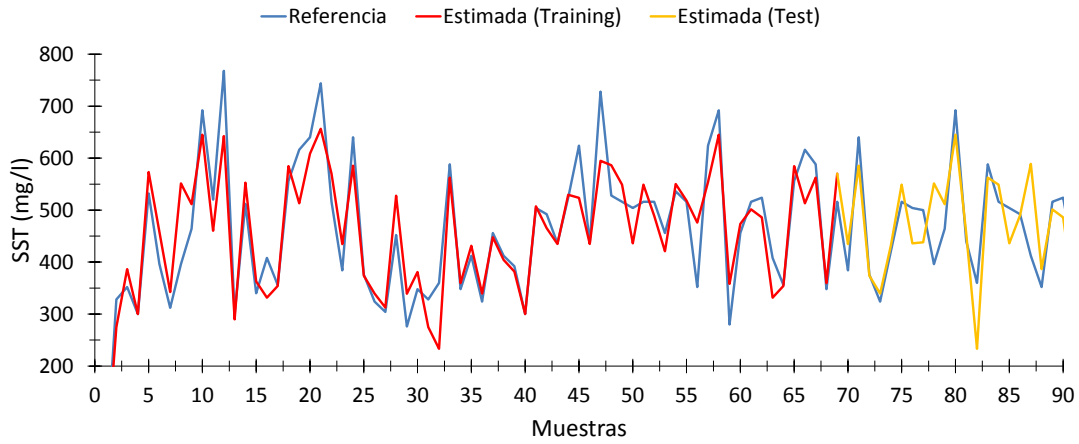


Figura 7.11. Comparación entre los valores de SST medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos por el modelo de regresión lineal, para las aguas residuales (agua bruta).

Donde las correlaciones entre los regresores utilizados en el modelo de la Ecuación (7.3) y los SST se muestran en la Figura 7.12.

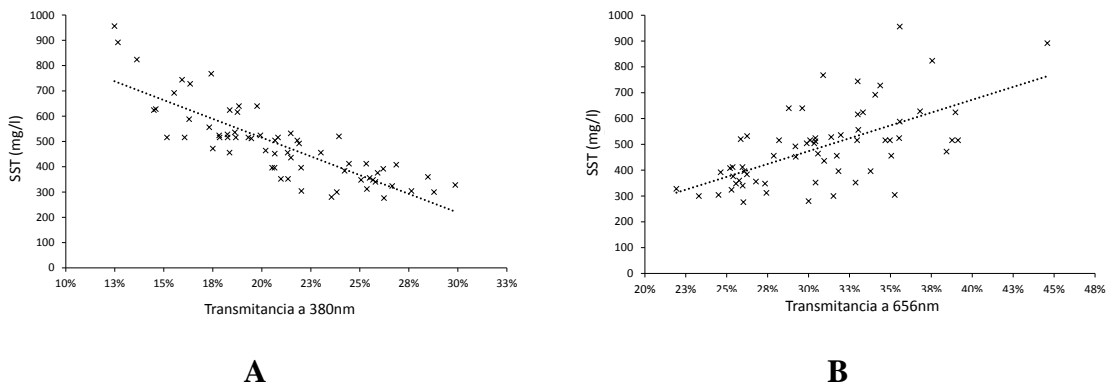


Figura 7.12. Diagrama de dispersión de las variables utilizadas en el modelo de la Ecuación (7.3) para calcular los SST con respecto a (A) Transmittancia a 380nm, (B) Absorbancia a 656nm.

7.3.3 Modelos basados en algoritmo genético

Aunque los modelos de regresión lineal descritos en la sección anterior son capaces de proporcionar una estimación de la carga contaminante, estos presentan una serie de limitaciones: Son válidos para un reducido número de parámetros contaminantes (DBO y SST), su uso se limita al agua bruta y su ajuste no es muy elevado.

Los modelos de regresión presentan dificultades para modelar el comportamiento de muestras con bajos niveles de DQO y DBO₅, es decir, DQO inferior a 55 mg/l y DBO₅ inferior a 15 mg/l, que son los valores típicos del efluente de salida de la planta de tratamiento (agua tratada). Esto se debe a que las fluctuaciones de transmitancia/absorbancia en el espectro visible cercano a la luz ultravioleta son menos significativas que las observadas en las aguas residuales. Esto se observa en la Figura 7.4, donde el gráfico de transmitancia se asemeja a una línea casi horizontal.

Dado que el objetivo del presente trabajo de investigación es el desarrollo de modelos que permitan estimar un amplio número de parámetros contaminantes con una alta precisión, se hizo necesario implementar un nuevo tipo de técnica de cálculo que superara las limitaciones intrínsecas de los modelos de regresión.

Por ese motivo, se desarrollaron una serie de modelos basados en algoritmo genético, más concretamente, basados en regresión simbólica [210, 211]. Este tipo de técnicas permite extraer patrones de los datos aunque la variación en los mismos sea mínima. Además, su carácter evolutivo permite alcanzar modelos más robustos. Los modelos de cálculos fueron desarrollados en Python utilizándolas siguientes librerías: TensorFlow [212], NumPy [213, 214] y gpLearn [215].

La regresión simbólica funciona a través de un sistema de "árboles" [216, 217] compuestos de nodos interconectados. Cada uno de esos nodos puede estar compuesto por una variable (valores de transmitancia y absorbancia para cada una de las 81 longitudes de onda, es decir, un total de 162 variables) u operadores/funciones (suma, resta, división, multiplicación, funciones trigonométricas, etc.)

El proceso de encontrar un modelo que correlacione las variables de entrada con las de salida se basa en un proceso evolutivo. Como punto de partida, se han utilizado tanto modelos de regresión lineal calculados para los parámetros del contaminante, como funciones inicializadas aleatoriamente basadas en ciertas restricciones de longitud y tipo de operadores.

Los modelos calculados, han sido obtenidos a partir de 100 generaciones, en las que en cada generación se crean 1000 árboles diferentes, con una tasa de mutación del 15% por el método de intercambio de subárboles [216, 217].

Cada árbol se genera a partir de un nodo de adición, del que se deriva un número aleatorio de nodos, que pueden ser constantes, variables y operaciones. Los nodos que consisten en operaciones, tendrán nuevos nodos descendentes, que pueden ser de nuevo, constantes, variables u operaciones. El proceso de ramificación del árbol continúa hasta que todas las terminaciones son constantes o variables, o se supera la longitud total y/o la profundidad del árbol.

Cada árbol generado aleatoriamente se prueba con los datos de entrada (absorbancia/transmitancia) para comprobar cuán cerca está la variable de respuesta (por ejemplo, la DQO) de los valores calculados por la EDAR. Los árboles más cercanos a este resultado serán mutados (combinados) para generar otros 1000 árboles, y el proceso se repite hasta que se completen 100 generaciones.

La regresión simbólica (algoritmo genético), se basa en el desarrollo de una red neuronal, que debe ser entrenada y probada. Como datos de entrenamiento, se ha utilizado el 66% de los datos de entrada, tomado al azar, y el 34% restante de los datos, se ha utilizado para probar la validez de los modelos calculados tomados al azar también.

Es importante señalar que las expresiones proporcionadas por el algoritmo genético son sólo una de las muchas soluciones posibles. Las fórmulas generadas por el algoritmo pueden tener una extensión variable, incluir todo tipo de operaciones, tantas aritméticas y/o trigonométricas, exponenciales o logarítmicas, así como más o menos parámetros.

Gracias a las investigaciones llevadas a cabo, se ha desarrollado un total de 6 modelos de estimación basados en algoritmos genéticos, para el cálculo de: Demanda química de oxígeno (DQO), Demanda biológica de oxígeno a cinco días (DBO₅), Sólidos suspendidos totales (SST), Fósforo (P), Nitrógeno total (TN) y el Nitrógeno Nítrico (NO₃-N). Todos ellos con un alto nivel de precisión y un nivel de error inferior al 4%.

En las siguientes subsecciones se muestran los resultados obtenidos por los algoritmos, para cada uno de los parámetros analizados: DQO, DBO₅, TSS, P, TN y NO₃-N. A cada uno de ellos le sigue su fórmula de correlación, así como una comparación con los valores esperados de los parámetros contaminantes.

7.3.3.1 Demanda Química de Oxígeno (DQO)

En la Ecuación (7.4), se muestra el modelo obtenido mediante algoritmo genético para la estimación de la DQO a partir de los datos espectrofotométricos se muestra. Este modelo presenta un coeficiente de Pearson medio del 90,95%, con un ajuste similar en los datos de entrenamiento (95,07%) y los datos de prueba (90,93%). En total, se utilizaron para el cálculo 188 muestras de 196, tomadas en diferentes días, las cuales embarcaban tanto muestras de agua bruta como tratada. El modelo óptimo se alcanzó en el número de generación 84 de un máximo de 100, tal y como se muestra en la Figura 7.13.

$$DQO_{(mg/l)} = \left[\left((c_0 \cdot A_{594} - c_1 \cdot A_{557}) + (c_2 \cdot A_{380} - c_3 \cdot A_{521}) \right) + \left((c_4 \cdot A_{425} - c_5 \cdot A_{575}) + (c_6 \cdot A_{445} - c_7 \cdot A_{520}) \right) \right] \cdot c_8 + c_9 \quad (7.4)$$

$c_0 = 2.4268$
 $c_1 = 2.7910$
 $c_2 = 2.5317$
 $c_3 = 2.6341$
 $c_4 = 2.3278$
 $c_5 = 2.6879$
 $c_6 = 2.4569$
 $c_7 = 2.7717$
 $c_8 = 1191.8$
 $c_9 = -263.45$

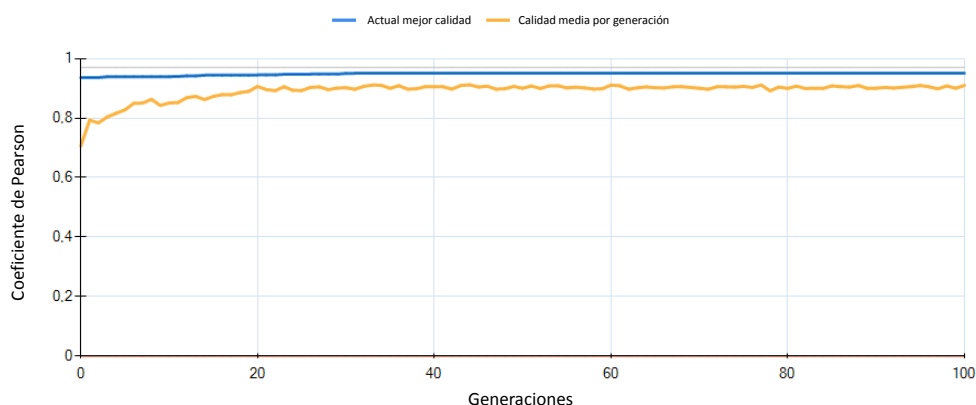


Figura 7.13. Curva de aprendizaje para el cálculo de DQO mediante algoritmo genético.

Este modelo se basa en ocho longitudes de onda para su cálculo: 380, 425, 445, 520, 521, 570, 575 y 594nm, más específicamente a partir de los datos de absorbancia. Sin embargo, no todas las variables (longitudes de onda) son igualmente relevantes. Como se muestra en la Tabla 7.3, 380, 425, 445 y 594nm son las variables más relevantes, con un factor de impacto (porcentaje en que la variable respuesta es explicada dentro del modelo por un determinado regresor) cercano al 17%, mientras que las variables restantes se sitúan en torno al 5%.

Tabla 7.3. Índices de impacto de las variables utilizadas para el modelo de DQO basado en algoritmo genético.

Variable	Impacto
A380 nm	17.021%
A425 nm	17.652%
A445 nm	17.691%
A520 nm	5.670%
A521 nm	5.417%
A570 nm	5.067%
A575 nm	4.966%
A594 nm	17.015%

Es importante señalar que las longitudes de onda con mayor índice de impacto fueron las pertenecientes a la zona violeta del espectro visible (380-450nm), cercanas a la zona ultravioleta, lo que era de esperar, ya que la materia orgánica es mucho más sensible a esas longitudes de onda. Además, se observó también que las longitudes de onda cercanas al rojo mostraban una mayor interacción con las muestras, lo cual sugiere que el uso de longitudes de onda más cercanas al espectro infrarrojo proporcionaría una mejor caracterización.

Con el fin de mostrar la alta precisión del modelo generado, el cual es válido tanto para muestras de agua bruta como tratada, la Figura 7.14 muestra una comparativa entre las estimaciones obtenidas (rojo para los datos de entrenamiento y amarillo para los de prueba), y los valores de referencia medidos en laboratorio.

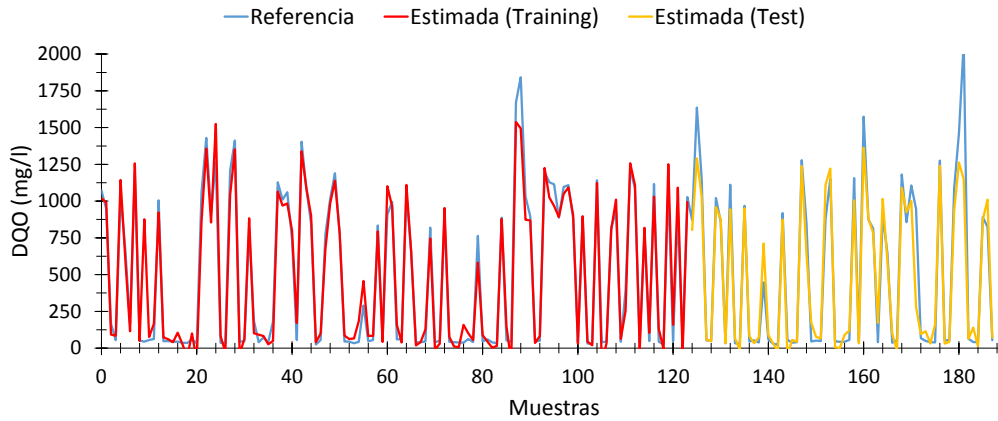


Figura 7.14. Comparación entre los valores de DQO medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.4).

Como se puede observar, valores muy altos de DQO (>1600 mg/l) hacen que las estimaciones tienden a ser más bajas de lo esperado. Sin embargo, los resultados podrían ser adecuados para proporcionar un sistema de alerta temprana donde no se requiere una precisión absoluta, sino únicamente avisar cuando se supere un cierto umbral. Sin embargo, para valores bajos de DQO, el modelo es capaz de proporcionar una estimación bastante acertada a partir de datos espectrofotométricos, lo cual no era posible con modelos de regresión lineal.

En la Tabla 7.4 se muestran 15 registros aleatorios, que incluyen los valores de absorbancia obtenidos para cada una de las variables utilizadas en el modelo (Ecuación (7.4)), así como los valores de DQO esperados (Referencia) y los calculados por el modelo (Estimados).

Tabla 7.4. Ejemplo de estimación DQO de 15 muestras aleatorias.

Abs. λ380 nm	Abs. λ425 nm	Abs. λ445 nm	Abs. λ520 nm	Abs. λ521 nm	Abs. λ570 nm	Abs. λ575 nm	Abs. λ594 nm	DQO	
								Referencia	Estimada
0.6849	0.6150	0.5584	0.4286	0.4190	0.3686	0.3712	0.3532	1072.000	1019.642
0.6595	0.5925	0.5068	0.4089	0.3997	0.3387	0.3409	0.3317	947.000	989.649
0.1468	0.1029	0.0970	0.0712	0.0664	0.0646	0.0645	0.0693	54.000	84.371
0.6948	0.6092	0.5478	0.4206	0.4112	0.3523	0.3548	0.3478	1042.000	1143.886
0.6514	0.5852	0.5439	0.4505	0.4405	0.3854	0.3882	0.3758	672.000	608.425
0.2238	0.1776	0.1539	0.1310	0.1256	0.1138	0.1139	0.1141	167.000	114.573
0.7042	0.6147	0.5676	0.4148	0.4055	0.3554	0.3579	0.3456	1192.000	1256.427
0.1438	0.1103	0.0973	0.0742	0.0693	0.0704	0.0702	0.0723	52.000	48.711
0.7268	0.6568	0.6107	0.4691	0.4587	0.4033	0.4063	0.3914	1206.000	1039.073
0.8657	0.7763	0.7350	0.5606	0.5481	0.4599	0.4638	0.4446	1412.000	1352.363
0.1487	0.1088	0.0945	0.0753	0.0704	0.0674	0.0672	0.0699	53.000	56.831
0.1556	0.1181	0.0992	0.0794	0.0745	0.0804	0.0804	0.0842	50.000	46.730
0.1363	0.0949	0.0807	0.0639	0.0591	0.0628	0.0626	0.0663	33.000	33.484
0.1504	0.1096	0.0917	0.0741	0.0693	0.0702	0.0701	0.0662	44.000	33.788
0.1525	0.1159	0.0948	0.0770	0.0721	0.0735	0.0734	0.0717	56.000	42.996

Se puede observar que los resultados calculados son muy similares a los esperados, incluso cuando el nivel de DQO es bajo. El modelo obtenido mediante el algoritmo genético es capaz de estimar con precisión los valores de DQO a partir de los datos proporcionados por el espectrofotómetro.

7.3.3.2 Demanda biológica de oxígeno a 5 días (DBO₅)

Para calcular el modelo para la DBO₅, se utilizaron 162 de un total de 196 muestras después de eliminar los valores atípicos -un número menor que antes- debido a dos aspectos: la existencia de valores atípicos y las mediciones en las que no se disponía de datos sobre la DBO₅. El modelo calculado se muestra en la Ecuación (7.5). Este modelo mostró un ajuste Pearson media del 90,71% (datos de entrenamiento) y del 90% para los datos de prueba (88,23% de media). Además, el modelo es válido tanto para muestras de agua con altos niveles de contaminación (agua cruda) como con bajos niveles de contaminación (agua tratada). El modelo óptimo se logró en la generación 98.

$$DBO_5 (mg/l) = \left[\frac{c_0 \cdot A_{574} \cdot c_1 \cdot T_{585}}{c_2 \cdot A_{655} - c_3 \cdot T_{415}} \cdot \frac{c_4 \cdot T_{585} \cdot c_5 \cdot A_{445}}{c_6 \cdot A_{655} - c_7 \cdot T_{415}} \right] \cdot c_8 + c_9 \tag{7.5}$$

$c_0 = 2.0733$
 $c_1 = 1.3974$
 $c_2 = -1.0226$
 $c_3 = 1.2453$
 $c_4 = 1.3974$
 $c_5 = -0.1356$
 $c_6 = -1.0226$
 $c_7 = 1.2453$
 $c_8 = -10078$
 $c_9 = -18.784$

Este modelo se basa en cinco longitudes de onda para su cálculo: 415, 445, 574, 585 y 655nm. Sin embargo, no todas las variables (longitudes de onda) son igualmente relevantes. Como puede verse en la Tabla 7.5, las longitudes de onda más relevantes son las más cercanas a la zona del violeta [218], aunque las longitudes de onda cercanas al rojo tienen un nivel de importancia similar, aunque en menor proporción.

Tabla 7.5. Índices de impacto de las variables utilizadas para el modelo de DBO₅ basado en algoritmo genético.

Variable	Impacto
T415 nm	21.211%
A445 nm	18.522%
A574 nm	18.030%
T585 nm	13.924%
A655 nm	15.312%

La Figura 7.15 muestra las estimaciones del algoritmo genético. Como puede observarse, el ajuste es alto, lo que denota la validez del modelo.

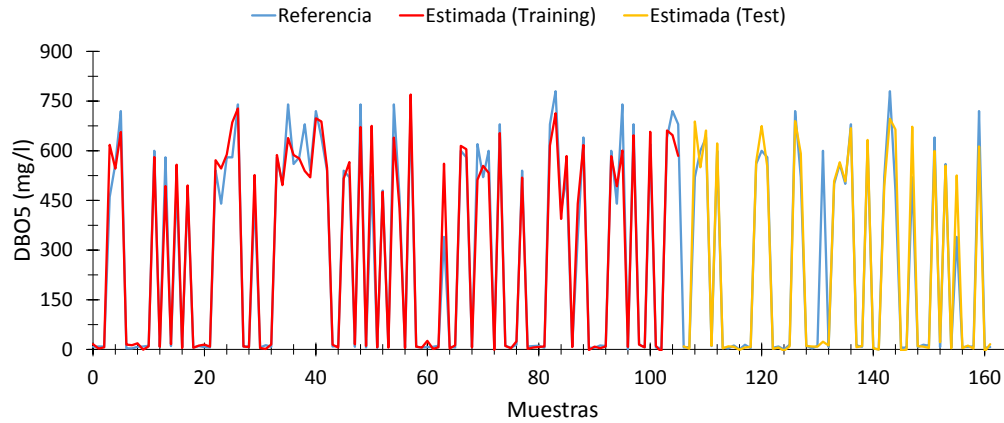


Figura 7.15. Comparación entre los valores de DBO₅ medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.5).

En la Tabla 7.6 se muestran 15 registros aleatorios, que incluyen los valores de absorbancia obtenidos para cada una de las variables utilizadas en el modelo (Ecuación (7.5)), así como los valores de DBO₅ esperados (Referencia) y los calculados por el modelo (Estimados).

Tabla 7.6. Ejemplo de estimación DBO₅ de 15 muestras aleatorias.

Trans. λ415 nm	Trans. λ585 nm	Abs. λ445 nm	Abs. λ574 nm	Abs. λ655 nm	DBO ₅	
					Referencia	Estimado
0.76629	0.87322	0.10162	0.06002	0.05524	9.000	6.420
0.24403	0.45556	0.55325	0.34426	0.27607	460.000	618.008
0.27323	0.48592	0.49108	0.31802	0.25519	560.000	546.023
0.22075	0.42956	0.58969	0.36959	0.29248	720.000	656.963
0.20025	0.38834	0.61710	0.43081	0.35095	600.000	581.177
0.75852	0.85732	0.09879	0.06732	0.06093	7.000	7.909
0.30964	0.52375	0.45107	0.29201	0.23394	580.000	493.581
0.74203	0.84762	0.11473	0.07572	0.07354	10.000	15.819
0.28582	0.50567	0.46952	0.30787	0.23439	500.000	558.014
0.77097	0.87022	0.08342	0.06775	0.05189	7.000	4.297
0.28026	0.48173	0.47856	0.32273	0.26617	480.000	495.409
0.76758	0.86484	0.09159	0.06533	0.06259	6.000	5.035
0.74707	0.85266	0.09923	0.07717	0.07499	10.000	11.608
0.74225	0.85333	0.10753	0.07400	0.05520	7.000	14.561
0.75661	0.86687	0.09744	0.06849	0.06876	6.000	8.296

7.3.3.3 Sólidos suspendidos totales (SST)

La Ecuación (7.6) muestra el modelo calculado para el total de sólidos en suspensión. Este modelo presentaba un ajuste Pearson media del 87,47% (94,67% con los datos de entrenamiento y 90% con los datos de prueba). En total, se utilizaron 172 de las 196 muestras, después de eliminar los valores atípicos.

$$SST_{(mg/l)} = \frac{c_0 \cdot A_{574} + c_1 \cdot T_{558}}{c_2 \cdot T_{485}} \cdot \left(\frac{c_3 \cdot T_{565}}{c_4 \cdot T_{380}} + c_5 \cdot T_{632} \right) \cdot c_6 + c_7$$

$$c_0 = -0.062545$$

$$c_1 = 2.6249$$

$$c_2 = 3.4131$$

$$c_3 = 2.5468$$

$$c_4 = 3.3361$$

$$c_5 = 0.46423$$

$$c_6 = 782.89$$

$$c_7 = -779.73$$
(7.6)

El modelo utiliza seis variables (longitudes de onda): transmitancia a 380, 485, 558, 565 y 632nm y absorbancia a 574nm. Sin embargo, las más relevantes son 380 y 485nm, como se muestra en la Tabla 7.7. Asimismo, observamos que a medida que las longitudes de onda se acercan al espectro infrarrojo, el peso relativo de estas variables disminuye significativamente, como es el caso de 632nm.

Tabla 7.7. Índices de impacto de las variables utilizadas para el modelo de SST basado en algoritmo genético.

Variable	Impacto
T380 nm	34.059%
T485 nm	34.975%
A574 nm	0.018%
T558 nm	11.166%
T565 nm	10.600%
T632 nm	1.552%

Esto demuestra que las partículas en suspensión son mucho más sensibles a las longitudes de onda cercanas al violeta que a otras.

La Figura 7.16 muestra los resultados obtenidos con el modelo calculado. Como se puede ver, el ajuste es adecuado, incluso a altos niveles de SST.

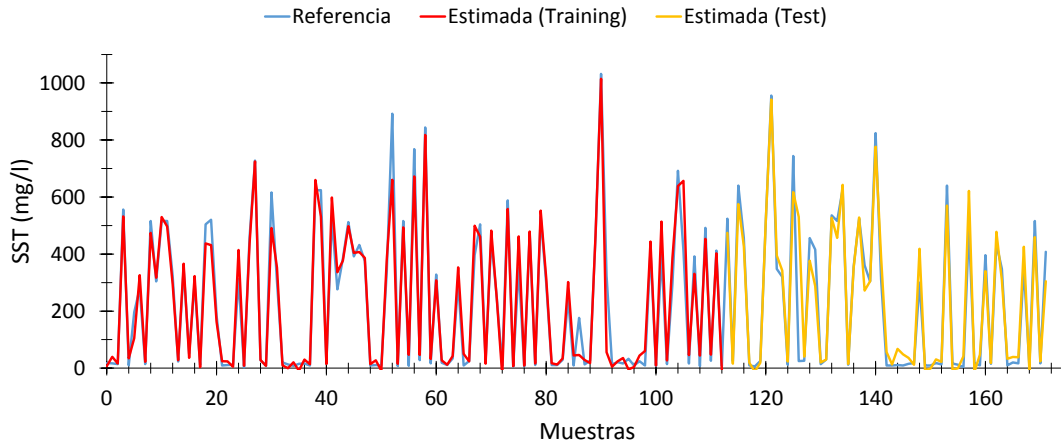


Figura 7.16. Comparación entre los valores de SST medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.6).

En la Tabla 7.8 se muestran 15 registros aleatorios, que incluyen los valores de absorbancia obtenidos para cada una de las variables utilizadas en el modelo (Ecuación (7.6)), así como los valores de SST esperados (Referencia) y los calculados por el modelo (Estimados).

Tabla 7.8. Ejemplo de estimación SST de 15 muestras aleatorias.

Trans. $\lambda 380 \text{ nm}$	Trans. $\lambda 485 \text{ nm}$	Trans. $\lambda 558 \text{ nm}$	Trans. $\lambda 565 \text{ nm}$	Trans. $\lambda 632 \text{ nm}$	Abs. $\lambda 574 \text{ nm}$	SST	
						Referencia	Estimado
0.718	0.852	0.864	0.856	0.853	0.067	14.000	16.465
0.173	0.317	0.381	0.376	0.426	0.413	556.000	532.238
0.711	0.858	0.870	0.863	0.880	0.061	10.000	34.512
0.449	0.607	0.644	0.638	0.672	0.196	196.000	105.120
0.235	0.379	0.438	0.432	0.472	0.350	280.000	326.214
0.701	0.835	0.848	0.840	0.861	0.070	14.000	22.257
0.186	0.328	0.392	0.387	0.441	0.398	516.000	474.119
0.276	0.442	0.502	0.497	0.545	0.292	304.000	317.133
0.208	0.360	0.439	0.434	0.486	0.344	516.000	529.615
0.193	0.337	0.405	0.400	0.465	0.369	516.000	496.820
0.266	0.422	0.478	0.472	0.531	0.304	324.000	295.014
0.664	0.805	0.822	0.815	0.820	0.087	23.000	27.927
0.196	0.334	0.387	0.382	0.423	0.394	348.000	366.015
0.684	0.850	0.836	0.829	0.874	0.086	5.000	6.355
0.725	0.857	0.862	0.855	0.867	0.064	6.000	8.570

7.3.3.4 Fósforo (P)

El modelo calculado para P se muestra en la Ecuación (7.7). El modelo presentaba un ajuste Pearson media del 74,01% (74,28% con los datos de entrenamiento y 78,33% con los datos de prueba), elevándose el óptimo en la generación 38 de un máximo de 100. En total, se utilizaron 175 datos para su cálculo.

$$P_{(mg/l)} = \left[\left(\frac{c_0 \cdot T_{430}}{c_1 \cdot T_{585}} - c_2 \cdot T_{650} \right) \cdot (c_3 \cdot T_{425} - c_4 \cdot T_{585}) \cdot (c_5 \cdot T_{450} - c_6 \cdot T_{650}) \cdot c_7 + c_8 \right] \quad (7.7)$$

$c_0 = 1.53$
 $c_1 = 0.8773$
 $c_2 = 1.1618$
 $c_3 = 1.5294$
 $c_4 = 0.8773$
 $c_5 = 2.2034$
 $c_6 = 1.1618$
 $c_7 = -40.766$
 $c_8 = 9.0573$

El modelo calculado utiliza cinco longitudes de onda: 425, 430, 450, 585 y 650nm. Una vez más, las longitudes de onda más representativas son las más cercanas a la zona violeta, como se muestra en la Tabla 7.9. Al igual que en el modelo de sólidos en suspensión totales, el peso de las longitudes de onda disminuye a medida que se acerca a la porción infrarroja del espectro.

Tabla 7.9. Índices de impacto de las variables utilizadas para el modelo del P basado en algoritmo genético.

Variable	Impacto
T425 nm	19.057%
T430 nm	19.035%
T450 nm	18.130%
T585 nm	12.355%
T650 nm	7.024%

El modelo generado para estimar los niveles de fósforo a partir de los datos espectrofotométricos tiene un ajuste menor en comparación con los modelos anteriores y presenta inexactitudes sistemáticas para concentraciones más altas. El modelo calculado sólo pudo estimar con precisión los valores P inferiores o iguales a 9 mg/l. Esta característica puede observarse en la Figura 7.17, donde los valores estimados nunca son superiores a ese valor.

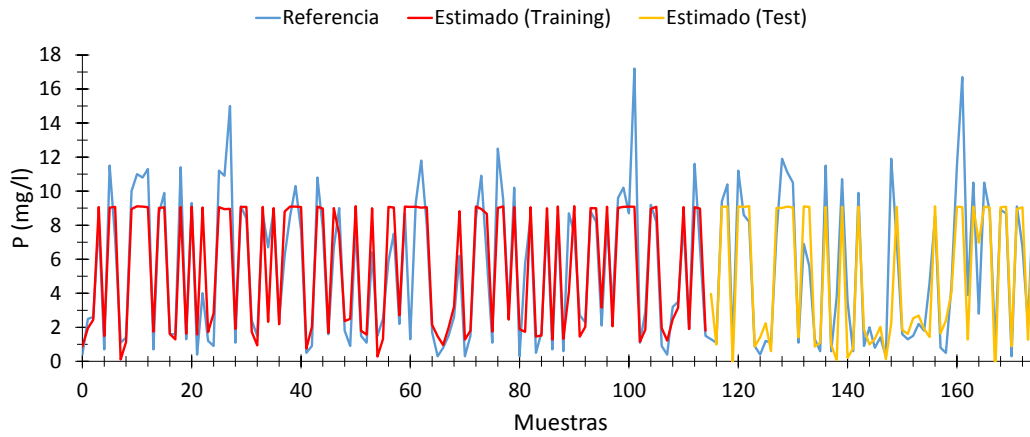


Figura 7.17. Comparación entre los valores de P medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.7).

Con el fin de clarificar este punto, la Tabla 7.10 muestran 15 registros aleatorios, que incluyen los valores de absorbancia obtenidos para cada una de las variables utilizadas en el modelo (Ecuación (7.7)), así como los valores de P esperados (Referencia) y los calculados por el modelo (Estimados).

Tabla 7.10. Ejemplo de estimación P de 15 muestras aleatorias.

Trans. $\lambda 425$ nm	Trans. $\lambda 430$ nm	Trans. $\lambda 450$ nm	Trans. $\lambda 585$ nm	Trans. $\lambda 650$ nm	P	
					Referencia	Estimado
0.76348	0.78803	0.79807	0.84547	0.86287	0.400	0.879
0.75775	0.77838	0.78847	0.86838	0.86515	2.600	2.443
0.24537	0.26567	0.29388	0.43165	0.48103	8.200	9.064
0.76757	0.79379	0.80362	0.85966	0.87870	0.700	1.494
0.29621	0.33102	0.35410	0.50567	0.54240	11.500	9.028
0.25278	0.27821	0.30349	0.44776	0.49307	7.200	9.070
0.79540	0.81032	0.79494	0.86041	0.86512	1.100	0.103
0.76358	0.83164	0.79458	0.87531	0.86525	1.400	1.136
0.30329	0.33745	0.36199	0.49713	0.54051	10.000	8.951
0.22378	0.24929	0.27953	0.44088	0.47647	11.000	9.106
0.17840	0.19534	0.22203	0.35499	0.38754	10.800	9.088
0.25701	0.28659	0.31936	0.44796	0.48733	11.300	9.057
0.75361	0.79287	0.79741	0.86315	0.86191	0.700	1.743
0.29637	0.32694	0.35853	0.49760	0.55641	8.700	9.008
0.17356	0.18391	0.21443	0.35122	0.41531	9.900	9.050

7.3.3.5 Nitrógeno total (NT)

El modelo para el Nitrógeno Total (TN) se muestra en la Ecuación (7.8). Este modelo tuvo una bondad de ajuste Pearson media del 79,93% (85,91% con los datos de entrenamiento y 85,91% con los datos de prueba), habiendo sido calculada a partir de 175 muestras del total de 196 después de eliminar los valores atípicos. El óptimo se elevó en la generación 87.

$$\begin{aligned}
 NT_{(mg/l)} = & [(c_0 \cdot T_{655} - c_1 \cdot T_{585}) \cdot (c_2 \cdot T_{640} - c_3 \cdot T_{510}) \\
 & \cdot (c_4 \cdot T_{557} - c_5 \cdot T_{585}) \cdot (c_6 \cdot T_{640} - c_7 \cdot T_{500}) \cdot c_8 + c_9]
 \end{aligned}
 \tag{7.8}$$

$$\begin{aligned}
 c_0 &= 1.3315 \\
 c_1 &= 0.85214 \\
 c_2 &= 2.1725 \\
 c_3 &= 1.6762 \\
 c_4 &= 1.4023 \\
 c_5 &= 0.85214 \\
 c_6 &= 2.1725 \\
 c_7 &= 1.6605 \\
 c_8 &= -1271.6 \\
 c_9 &= 83.172
 \end{aligned}$$

El modelo utiliza seis longitudes de onda: 500, 510, 557, 585, 640 y 655nm. Como se muestra en la Tabla 7.11, las longitudes de onda más representativas utilizadas para calcular el contenido de nitrógeno del agua fueron las más cercanas al infrarrojo. Esto ya se ha destacado en [219], donde el nitrógeno tiene una mayor correlación con las longitudes de onda cercanas al espectro infrarrojo [220].

Tabla 7.11. Índices de impacto de las variables utilizadas para el modelo del NT basado en algoritmo genético.

Variable	Impacto
T500 nm	7.254%
T510 nm	6.639%
T557 nm	19.502%
T585 nm	9.426%
T640 nm	18.765%
T655 nm	19.714%

La figura 7.18 muestra los resultados proporcionados por el modelo descrito en (7.8). Como puede verse, la fórmula funciona bien dentro de un cierto rango de valores de nitrógeno entre 20 y 75mg/l, pero empeora ligeramente fuera de ese rango, aunque no de manera significativa.

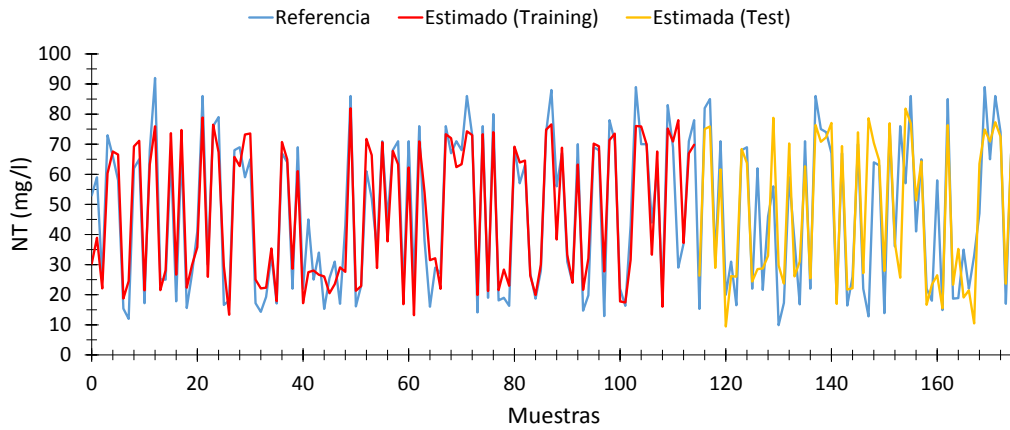


Figura 7.18. Comparación entre los valores de NT medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.8).

Con el fin de clarificar este punto, la Tabla 7.12 muestran 15 registros aleatorios, que incluyen los valores de absorbancia obtenidos para cada una de las variables utilizadas en el modelo (Ecuación (7.8)), así como los valores de NT esperados (Referencia) y los calculados por el modelo (Estimados).

Tabla 7.12. Ejemplo de estimación NT de 15 muestras aleatorias.

Trans. $\lambda 500 \text{ nm}$	Trans. $\lambda 510 \text{ nm}$	Trans. $\lambda 557 \text{ nm}$	Trans. $\lambda 585 \text{ nm}$	Trans. $\lambda 640 \text{ nm}$	Trans. $\lambda 655 \text{ nm}$	NT	
						Referencia	Estimada
0.82757	0.80397	0.82113	0.85584	0.86375	0.86940	25.000	22.087
0.44955	0.45871	0.50539	0.52375	0.55827	0.58352	73.000	60.330
0.36769	0.38270	0.43492	0.44776	0.48237	0.51313	66.000	67.600
0.36531	0.37489	0.42788	0.45556	0.48742	0.52958	58.000	66.548
0.86872	0.84108	0.87294	0.86815	0.88365	0.87850	15.400	18.737
0.32128	0.32217	0.37064	0.39510	0.43704	0.48072	65.000	71.143
0.85905	0.83739	0.85368	0.84908	0.87566	0.86752	17.200	21.488
0.43282	0.43977	0.48290	0.49760	0.53349	0.56887	68.000	63.238
0.28615	0.29335	0.33406	0.36080	0.38741	0.43757	92.000	75.947
0.86767	0.84051	0.86005	0.87977	0.89105	0.86902	25.000	21.501
0.85770	0.82966	0.84810	0.85738	0.86480	0.86812	25.000	28.133
0.33833	0.34440	0.38470	0.41056	0.43200	0.48097	66.000	73.640
0.85727	0.82356	0.83882	0.85370	0.86860	0.86295	17.800	26.685
0.27823	0.27440	0.34322	0.34476	0.37727	0.42773	68.000	74.717
0.87407	0.84157	0.89042	0.85226	0.87565	0.87850	17.100	17.848

7.3.3.6 Nitrato Nitrógeno (NO₃-N)

Para la modelización del NO₃-N, se han obtenido la expresión mostrada en Ecuación (7.9), la cual presenta un ajuste dado por el coeficiente de Pearson medio del 81.26% (81,26% con los datos de entrenamiento y 83,46% con los datos de prueba). En total, se utilizaron 175 muestras para el cálculo de 196 después de eliminar los valores atípicos, alcanzándose el óptimo en la generación 81 de 100.

$$NO_{3-N} (mg/l) = \left(\frac{c_0 \cdot A_{560} \cdot c_1 \cdot A_{560}}{c_2 \cdot A_{607} \cdot c_3 \cdot A_{624}} \cdot \frac{c_4 \cdot A_{428} \cdot c_5 \cdot T_{385}}{c_6 \cdot A_{607} \cdot c_7 \cdot A_{645}} \cdot c_8 + c_9 \right) \quad (7.9)$$

$c_0 = 2.2576$
 $c_1 = 2.2576$
 $c_2 = -0.53193$
 $c_3 = 1.5017$
 $c_4 = 0.66989$
 $c_5 = 2.277$
 $c_6 = -0.53193$
 $c_7 = -0.50608$
 $c_8 = -0.010536$
 $c_9 = -0.12637$

El modelo calculado, hacer uso de las siguientes seis longitudes de onda: 385, 428, 560, 607, 624 y 645nm. Las pruebas mostraron que el NO₃-N tiene una mayor correlación con longitudes de onda cercanas a los 600nm, como se muestra en la Tabla 7.13.

Tabla 7.13. Índices de impacto de las variables utilizadas para el modelo del NO₃-N basado en algoritmo genético.

Variable	Impacto
T385nm	8.257%
A428nm	14.054%
A560nm	16.613%
A607nm	20.243%
A624nm	11.210%
A645nm	11.924%

La Figura 7.19 muestra los resultados obtenidos para diferentes muestras de agua. Teniendo en cuenta que la escala vertical de la figura se muestra en intervalos de 2mg/l, las discrepancias entre los valores calculados y los valores de referencia no son significativas.

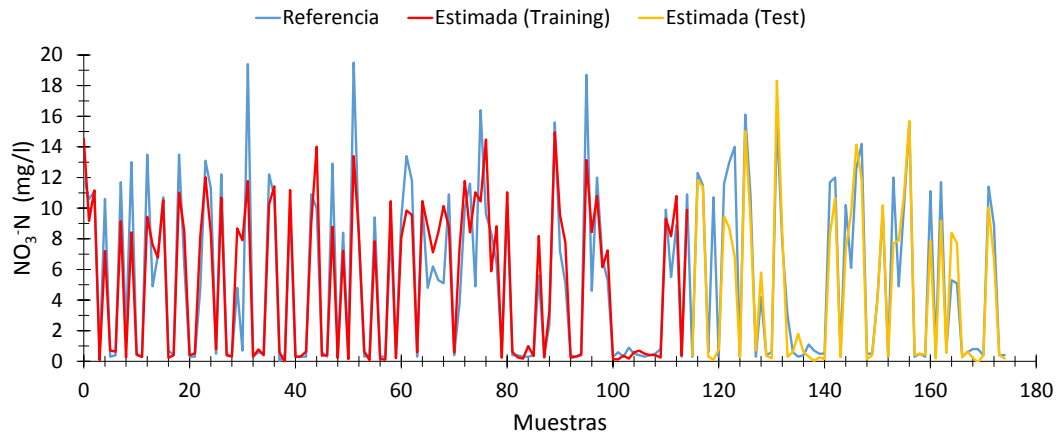


Figura 7.19. Comparación entre los valores de NO₃-N medidos por la planta de tratamiento de aguas residuales y los valores calculados a partir de los datos espectrofotométricos, según la Ecuación (7.9).

En la Tabla 7.12 se muestran 15 casos elegidos al azar, en los que se puede observar el alto grado de similitud entre los datos proporcionados por el modelo y los valores calculados en las plantas de tratamiento de aguas residuales.

Tabla 7.12. Ejemplo de estimación NO₃-N de 15 muestras aleatorias.

Trans. λ385 nm	Abs. λ428 nm	Abs. λ560 nm	Abs. λ607 nm	Abs. λ624 nm	Abs. λ645 nm	NO ₃ -N	
						Reference	Estimated
0.7427	0.0939	0.0825	0.0578	0.0595	0.0621	12.000	14.538
0.7051	0.0972	0.0631	0.0563	0.0634	0.0557	10.600	9.169
0.7358	0.0981	0.0582	0.0540	0.0502	0.0562	11.100	11.164
0.1233	0.8244	0.5345	0.4948	0.4640	0.4615	0.300	0.085
0.7273	0.1079	0.0723	0.0685	0.0719	0.0631	10.600	7.199
0.2677	0.4623	0.3142	0.2838	0.2679	0.2599	0.300	0.703
0.2635	0.4870	0.3112	0.2908	0.2743	0.2673	0.400	0.637
0.7500	0.0940	0.0678	0.0610	0.0636	0.0562	11.700	9.151
0.7243	0.0738	0.0653	0.0527	0.0624	0.0580	6.300	8.508
0.2279	0.5592	0.3987	0.3626	0.3448	0.3307	0.300	0.388
0.2560	0.4910	0.3422	0.3139	0.3004	0.2917	0.300	0.523
0.2824	0.4625	0.3037	0.2798	0.2573	0.2484	0.500	0.790
0.7362	0.1000	0.0694	0.0621	0.0601	0.0538	12.200	10.685
0.2062	0.5632	0.3706	0.3378	0.3253	0.3150	0.500	0.393
0.1768	0.6745	0.4240	0.3851	0.3599	0.3498	0.300	0.311

7.4 Comparación de modelos

Una vez descrito los diferentes modelos calculados, se hace necesario establecer una comparativa entre los mismos para poder determinar su adecuación para cada escenario de estudio.

Para poder llevar a cabo la comparación entre los diferentes tipos de modelos (modelos de regresión lineal y modelos basados en algoritmo genético), se analizaron los siguientes parámetros: Error cuadrático medio (RMSD) [172] y el índice de error, Er, a través de las Ecuaciones 7.10 y 7.11:

$$RMSD = \sqrt{\frac{1}{n} \sum_i^n (X_{referencia_i} - X_{Estimada_i})^2} \quad (7.10)$$

$$Er(\%) = \frac{\sum_i^n (X_{referencia_i} - X_{Estimada_i})}{\sum_i^n X_{referencia_i}} * 100 \quad (7.11)$$

Donde n es el número de muestras, X_Referencia y X_Estimada son los valores de los parámetros contaminantes (DQO, DBO₅, SST, P, TN y NO₃-N) obtenidos por los métodos analíticos utilizados por la planta de tratamiento de aguas residuales y por los modelos de cálculo, respectivamente.

Tabla 7.13. Error cuadrático medio e índice de error (Er) de los diferentes modelos calculados.

Modelo	Número de muestras	Parámetro	RMSD	Er (%)
Modelos de regresión multivariable (Sólo para el agua bruta)	101	DQO	95,910	-0.096%
	86	DBO ₅	134,372	-5.540%
	92	SST	62,197	0.295%
	188	DQO	137,048	-2.374%
Algoritmo genético (Agua bruta y tratada)	162	DBO ₅	69,051	-0.173%
	172	SST	67,159	0.621%
	175	P	2,037	-2.634%
	175	NT	11,783	0.147%
	175	NO ₃ -N	2,323	-3.928%

Tal y como se puede observar en la Tabla 7.13, el error que presenta la DQO proporcionado por el modelo de regresión lineal (-0,096%) es inferior al obtenido por el algoritmo genético (-2,374%). No obstante, hay que tener en cuenta, que el tipo y número de muestras utilizadas es diferente. Los modelos basados en algoritmo genético han sido calculados teniendo en cuenta el agua tratada, por lo que, si bien es cierto que los modelos de regresión presentan un mejor rendimiento que el obtenido por el algoritmo genético, su uso se limita al agua bruta.

Por lo tanto, si las muestras que se desean analizar su valor de DQO son muestras de agua bruta, los modelos de regresión serían la mejor opción. Sin embargo, si deseásemos realizar el estudio en ambos tipos de agua (bruta y tratada), se debería recurrir al uso del algoritmo genético, el cual ha presentado un rendimiento más alto en el resto de parámetros.

En cuanto a la contribución de las longitudes de onda en el proceso de estimación, no todas las tienen el mismo peso en los modelos, tal y como se muestra en la Tabla 7.14. En términos generales, las longitudes de onda más cercanas al violeta tienen una mayor significación, lo cual es comprensible si se tiene en cuenta que la materia orgánica reacciona más a este tipo de radiación que a otras longitudes de onda. Por otra parte, las longitudes de onda cercanas al infrarrojo también tienen una mayor importancia en el cálculo de parámetros inorgánicos como el nitrógeno total.

Además, aunque en términos generales los modelos obtenidos por medio de algoritmos genéticos presentan un mejor rendimiento, es necesario destacar que estos utilizan un mayor número de variables (longitudes de onda) que los de regresión lineal. Esto implica un mayor tiempo de análisis y un aumento de la carga del sistema así como del precio del equipo ya que se necesitan más diodos emisores de luz. Por lo tanto, la elección del modelo dependerá de la aplicación.

En la Tabla 7.14 se muestran las diferentes longitudes de onda utilizadas para el cálculo de los seis parámetros del contaminante, donde cada celda muestra el grado de importancia de esa longitud de onda en su cálculo, acompañado de un código de colores, para mayor claridad: verde (alta relevancia), azul (media-alta relevancia), naranja (media-baja relevancia) y rojo (baja relevancia). Los coeficientes mostrados se determinaron automáticamente por el software SPSS y gpLearn a partir del valor P de las variables introducidas en los diferentes modelos.

Tabla 7.14. Resumen de la relevancia de las longitudes de onda en los diferentes modelos de estimación calculados.

	Modelo de regresión lineal			Algoritmo genético					
	DQO	DBO ₅	SST	DQO	DBO ₅	SST	P	NT	NO ₃ N
380 nm	32.83%		30.21%	17.02%		34.06%			
385 nm									8.26%
415 nm					21.21%				
425 nm		26.48%	29.33%	17.65%			19.05%		
428 nm									14.05%
430 nm							19.03%		
445 nm				17.69%	18.52%				
450 nm							18.13%		
485 nm						34.97%			
500 nm								7.25%	
510 nm								6.63%	
515 nm		18.78%							
520 nm				5.67%					
521 nm	14.79%			5.41%					
555 nm	14.20%	12.41%							
557 nm								19.50%	
558 nm						11.16%		9.42%	
560 nm									16.61%
565 nm						10.60%			
570 nm				5.06%					
574 nm					18.03%	≈ 0%			
575 nm				4.96%					
580 nm	13.28%								
585 nm					13.92%		12.36%		
594 nm				17.01%					
607 nm									20.24%
624 nm									11.21%
632 nm						1.55%			
640 nm								18.76%	
645 nm									11.92%
650 nm							7.02%		
655 nm					15.31%			19.71%	
656 nm			11.87%						

Tal y como se puede apreciar en la tabla anterior, los parámetros contaminantes relacionados con la materia orgánica, como la DQO y la DBO₅, muestran una mayor interacción con las longitudes de onda cercanas al violeta y en menor medida con las longitudes de onda del orden de 500-550 nm (verde).

Por otra parte, los parámetros más relacionados con la materia inorgánica como el nitrógeno total (NT) son más sensibles a las longitudes de onda cercanas al infrarrojo (IR); de hecho, el NT se calcula mediante técnicas de NIRS (near infrareds) [221].

En este punto es necesario hacer una comparación entre las longitudes de onda presentes en ambos modelos. Como se muestra en la Tabla 7.14, las longitudes de onda seleccionadas por ambas metodologías son similares, especialmente en la zona ultravioleta, donde, por ejemplo, la longitud de onda de 380 nm está presente tanto para la DQO como para la DBO₅ en ambos tipos de modelos. Por otra parte, es necesario tener en cuenta que los modelos

basados en algoritmo genético son válidos tanto para el agua bruta como para el agua tratada y, por lo tanto, es lógico pensar que el número de longitudes de onda utilizadas es mayor que el necesario para modelar sólo el agua bruta. A pesar de ello, existen similitudes entre ambos tipos de modelos.

7.5 Conclusiones

Los resultados mostrados en el presente capítulo son de suma importancia para el desarrollo de equipos de bajo coste que permitan estimar la carga contaminante de las redes de saneamiento a partir de mediciones indirectas.

Las pruebas realizadas determinaron que los modelos calculados mediante algoritmos genéticos son capaces de obtener estimaciones válidas principalmente para cinco de los contaminantes en estudio (DQO, DBO₅, TSS, TN y NO₃-N), incluyendo tanto las aguas residuales como las tratadas en los ajustes, con una tasa de error inferior al 4% en todos los modelos. En el caso de los modelos de regresión, su adecuación se limita a la DQO y los SST, mientras que la DBO₅ presenta un ajuste muy bajo, no adecuado para los fines que se persiguen en el presente trabajo de investigación. A diferencia de los modelos basados en algoritmo genético, los modelos de regresión presentaron mejores tasas de error, inferior al 0,5% para la DQO y la SST, sin embargo, los modelos de regresión se limitan a muestras de agua bruta.

La variabilidad de las muestras de aguas residuales, dificulta que este tipo de modelos puedan ser aplicados, dado que es difícil desarrollar un modelo de regresión que sea capaz de funcionar tanto para el agua bruta como para el agua tratada

Estos resultados proporcionan una visión más clara del efecto que cada una de las longitudes de onda (380-700nm) tienen en la estimación de cada uno de los parámetros contaminantes. Como se muestra en la Tabla 7.14, las longitudes de onda que tienen mayor efecto en el cálculo son las correspondientes a las zonas de UV (o cercano al mismo) y de infrarrojo cercano (600-700nm), con una relevancia (impacto) del 17-20% en el cálculo del modelo, mientras que la zona entre 500 y 600nm es la menos relevante, con un impacto de alrededor del 5%, aunque con algunas excepciones, como el SST (alrededor del 10%).

La DQO, la DBO₅, la SST y el P dependen principalmente de la zona UV para su cálculo, representando (en el caso de los modelos calculados con el algoritmo genético) alrededor del 52%, 40%, 70% y 40%, respectivamente. Por otro lado, NT y NO₃-N dependen principalmente de la zona IR.

El utilizar las muestras sin filtrar para el cálculo de los modelos de estimación mostrados en el presente capítulo, permiten llevar a cabo la caracterización de las muestras sin necesidad de someterlas a ningún pretratamiento, simplificando el proceso de análisis por parte de los equipos descritos en los Capítulos 2 y 3.

Esta investigación puede servir como punto de partida para la futura vigilancia continua en tiempo real de todo el sistema de saneamiento, que incluye el despliegue de equipos más simples, más pequeños y más rentables para el estudio de la carga contaminante en las redes de aguas residuales, capaces de obtener información valiosa a partir de los modelos

estadísticos basados en la espectrofotometría y proporcionar una alerta temprana. Esta distribución de este equipo a lo largo de las redes puede ser especialmente útil durante los episodios de lluvia, cuando la carga contaminante de las redes de saneamiento tiende a aumentar y representa un peligro para el medio ambiente. Por lo tanto, disponer de información rápida sobre este tipo de parámetros es esencial para prevenir y reducir los desastres ambientales.

Capítulo 8

Conclusiones, aportaciones y desarrollos futuros

8.1 Conclusiones y aportaciones

El presente trabajo de investigación se ha centrado en el desarrollo de equipos y modelos estadísticos para la caracterización de la carga contaminante de las aguas residuales a partir del análisis espectrofotométrico de las mismas, sin necesidad del uso de reactivos químicos o someter a las muestras a pretratamientos.

Se ha buscado un enfoque diferente a los sistemas existentes en el mercado a la hora de llevar a cabo este análisis de la calidad de las aguas. Para ello se ha redefinido el concepto tradicional de equipo de espectrofotometría basado en el uso de elementos ópticos y lámparas incandescentes, por otro que hace uso de tecnología LED para la generación del espectro de trabajo, logrando un diseño más compacto, de menor consumo y bajo coste.

Gracias a los trabajos de investigación llevados a cabo a lo largo de la tesis, se ha diseñado un sistema que permite, no sólo equiparar con una alta precisión los resultados del análisis espectral obtenidos con tecnología LED con respecto a equipos basados en lámparas incandescentes, sino que también permite seleccionar una longitud de onda específica de las múltiples emitidas por los diodos, sin necesidad de recurrir a elementos ópticos como monocromadores o redes de difracción, lo que ha permitido alcanzar un diseño más simple y robusto, haciéndolo menos sensible a golpes y vibraciones.

La tesis doctoral ha culminado con el desarrollo de dos equipos, que al término de redacción de la presente memoria, se encuentran en proceso de protección bajo patente y/o modelo de utilidad, los cuales han recibido el sobrenombre de **Espectrofotómetro Desktop** y **Espectrofotómetro In-Sewers**. El primero, orientado como un equipo de sobremesa, proporciona una estimación de la carga contaminante de las muestras de agua introducidas en el laboratorio, siendo alimentado y controlado directamente desde el PC. El segundo, es un equipo autónomo que, por medio de una bomba, es capaz de captar, caracterizar y almacenar las muestras de agua, al tiempo que puede realizar el envío de los resultados del análisis a la nube.

Más allá del desarrollo físico de los equipos, uno de los principales pilares en la que todo este proceso de caracterización se sustenta son los diferentes modelos de estimación generados a partir de las campañas de muestreos llevadas a cabo. Siendo destacable el uso de algoritmos genéticos para la obtención de modelos para la estimación de la DQO, DBO₅, SST, P, NT y NO₃⁻N, tanto en muestras de agua bruta como tratada, con una alta precisión.

Todo este desarrollo se ha completado con un servidor web que permite el almacenamiento y procesamiento de la información obtenido por los equipos, la cual sirve a su vez de sistema de alerta temprana ante la existencia de niveles de contaminación de las aguas superiores a los esperados.

8.2 Desarrollos futuros

Como línea futura de desarrollo y enmarcado en el proyecto “**Nuevo sistema de monitorización inteligente de microorganismos y contaminantes emergentes en redes de saneamiento**” obtenido en la convocatoria Retos-Colaboración 2019 publicada por el Ministerio de Ciencia e Innovación a razón de los resultados obtenidos en la presente tesis doctoral, se plantea el estudio de contaminantes emergentes, por medio de fluorescencia UV-Vis como complemento a la espectrofotometría.

Así mismo, el proyecto se enmarca en la aplicación de técnicas de microscopía óptica, donde en combinación con modelos de inteligencia artificial aplicadas al campo de la visión por ordenador, se busca un mejor análisis de la composición de las aguas desde el punto de vista cuantitativo y cualitativo, aspecto clave para detectar la presencia de contaminantes emergentes en las aguas de una forma simple y rápida, mejorando la detección de patógenos en las mismas.

Anexo A

Tabla A1. Muestras utilizadas para la generación de modelos lineales para correlacionar la respuesta espectral de los LEDs con los resultados obtenidos con una lámpara incandescente.

Designación	Sustancia	Disolución
M0	Agua destilada	100%
M1	Colorante alimentario rojo 40	50%
M2	Colorante alimentario azul	50%
M3	Colorante alimentario amarillo	40%
M4	Detergente	50%
M5	Colorante alimentario rojo 40	75%
M6	Colorante alimentario rojo 40	55%
M7	Detergente	65%
M8	Detergente	75%
M9	Aceite	100%
M10	Vinagre	90%
M11	Leche	100%
M12	Leche	50%
M13	Café soluble	75%
M14	Café soluble	50%
M15	Vino rojo	100%
M16	Vino rojo	50%
M17	Colorante alimentario azul y amarillo	20%-80%
M18	Colorante alimentario azul	30%
M19	Agua marina	100%
M20	Té	80%
M21	Suavizante para ropa	40%

Tabla A2. LED de ancho espectral fijo utilizados para la generación del espectro de trabajo de los equipos desarrollados en Capítulo 2 y 3.

ID	λ_{Pico} (nm)	Ancho Espectral		Referencia	Datasheet
		Límite inferior (nm)	Límite superior (nm)		
1	385	360	400	VAOL-5EUV8T4	Link
2	390	388	392	UV5TZ-390-15	Link
3	395	393	398	UV5TZ-395-15	Link
4	400	398	400	UV5TZ-400-15	Link
5	405	395	415	VAOL-5EUV0T4	Link
6	428	400	440	4304H6	Link
7	461	440	480	151053BS04500	Link
8	468	456	480	LTL2P3TBK5	Link
9	500	475	520	SSL-LX5093UEGC	Link
10	505	487	518	LTL2V3TCYK2	Link
11	515	497	531	WP7113ZGCK	Link
12	521	505	536	HLMP-HM74-34CDD	Link
13	532	506	544	LTL2P3TGZ2KS	Link
14	557	543	575	WP7113PGC	Link
15	558	500	570	HLMP-C615-G0001	Link
16	574	585	602	TLCYG5100	Link
17	583	550	610	521-9466F	Link
18	586	565	615	521-9271F	Link
19	594	587	600	HLMP-EL1A-Z1KDD	Link
20	605	590	620	SLI-580DT3F	Link
21	607	581	623	WP7113NC	Link
22	609	600	618	HLMP-EJ15-SV000	Link
23	622	612	632	TLCR5200	Link
24	624	597	648	C503B-RBN-CX0Y0AA1	Link
25	627	600	650	WP57EYW	Link
26	631	623	645	151053RS03000	Link
27	636	625	648	SSL-LX5093SIC	Link
28	639	630	650	LTL2H3KRK	Link
29	650	635	663	HLMP-4101	Link
30	655	650	675	WP7113SRC/DU	Link
31	656	645	665	VAOL-5GAE4	Link
32	660	648	675	LTL-4268-H3	Link
33	697	650	743	LTL-4213	Link
34	700	655	730	SSL-LX5093HD-TR	Link

Tabla A3. Niveles de brillo seleccionados para cada LED de la Tabla A2.

LED λ_{pico} (nm)	Nivel de brillo		Cantidad de luz medida	
	Resolución (16bits)	Corriente (A)	Resolución (8bits)	Medición Sensor (V)
385	50	0,488	583	2,847
390	30	0,293	544	2,656
395	30	0,293	509	2,485
400	40	0,391	615	3,003
405	40	0,391	677	3,306
428	2090	20,415	499	2,437
461	80	0,781	527	2,573
468	60	0,586	557	2,720
500	50	0,488	517	2,524
505	110	1,074	531	2,593
515	70	0,684	529	2,583
521	210	2,051	523	2,554
532	50	0,488	502	2,451
557	4000	39,072	237	1,157
558	4000	39,072	208	1,016
574	670	6,545	499	2,437
583	4000	39,072	340	1,660
586	4000	39,072	114	0,557
594	70	0,684	536	2,617
600	2030	19,829	502	2,451
602	4000	39,072	93	0,454
605	150	1,465	534	2,607
607	1330	12,991	503	2,456
609	240	2,344	502	2,451
622	160	1,563	531	2,593
624	60	0,586	506	2,471
627	4000	39,072	102	0,498
631	60	0,586	549	2,681
636	130	1,270	516	2,520
639	240	2,344	505	2,466
650	90	0,879	562	2,744
655	190	1,856	502	2,451
656	410	4,005	501	2,446
660	260	2,540	502	2,451
697	4000	39,072	241	1,177
700	4000	39,072	461	2,251

Tabla A4. Muestras analizadas para la calibración y validación de los modelos de regresión lineal utilizados para correlacionar la respuesta espectral de los LED RGB con los resultados obtenidos con una lámpara incandescente.

	Designación	Sustancia	Disolución
Muestras de calibración	S0	Agua destilada	100%
	S1	Vino tinto	50%
	S2	Té	80%
	S3	Colorante alimentario amarillo y azul	20%–80%
	S4	Colorante alimentario azul	50%
	S5	Detergente de lavadora	50%
	S6	Detergente de lavadora	65%
	S7	Detergente de lavadora	75%
	S8	Leche	100%
	S9	Leche	50%
	S10	Colorante alimentario rojo	50%
	S11	Aceite de cocina	100%
	S12	Vinagre	90%
	S13	Colorante alimentario rojo	75%
	S14	Colorante alimentario rojo	55%
	S15	Café soluble	75%
	S16	Colorante alimentario amarillo	40%
	S17	Café soluble	50%
	S18	Vino tinto	100%
	S19	Colorante alimentario azul	30%
S20	Agua de mar	100%	
Muestras de prueba	S21	Aguas residuales urbanas. Agua bruta. Entrada EDAR	100%
	S22	Aguas residuales urbanas. Decantador Primario	100%
	S23	Aguas residuales tratadas. Salida de EDAR	100%
	S24	Aceite de oliva	100%
	S25	Cacao en polvo	5%
	S26	Cacao en polvo	30%
	S27	Cacao en polvo	55%
	S28	Café soluble	10%
	S29	Café soluble	30%
	S30	Cloruro de cetilpiridinio	50%
	S31	Cloruro de cetilpiridinio	100%
	S32	Cerveza	100%
	S33	Cerveza	50%
	S34	Agua de oliva	100%
	S35	Agua de oliva	50%

Tabla A4 Continuación. Muestras analizadas para la calibración y validación de los modelos de regresión lineal utilizados para correlacionar la respuesta espectral de los LED RGB con los resultados obtenidos con una lámpara incandescente.

	Designación	Sustancia	Disolución
Muestras de prueba	S36	Vino blanco	100%
	S37	Vino blanco	50%
	S38	Vino rosado	100%
	S39	Vino rosado	50%
	S40	Vino rosado	30%
	S41	Surfactantes anfotéricos	100%
	S42	Surfactantes anfotéricos	50%
	S43	Agua bruta. Entrada EDAR	100%
	S44	Agua bruta. Entrada EDAR	100%
	S45	Aguas residuales urbanas. Decantador primario	100%
	S46	Aguas residuales urbanas. Decantador primario	100%
	S47	Aguas residuales tratadas. Salida de EDAR	100%
	S48	Aguas residuales tratadas. Salida de EDAR	100%

Tabla A5. Equiparación Lambda-RGB pasado en ecuaciones empíricas.

Lambda	R	G	B	Lambda	R	G	B
380	97	0	97	581	255	252	0
381	100	0	102	582	255	249	0
382	103	0	106	583	255	246	0
383	106	0	111	584	255	242	0
384	109	0	115	585	255	239	0
385	111	0	119	586	255	236	0
386	114	0	124	587	255	233	0
387	116	0	128	588	255	230	0
388	118	0	132	589	255	226	0
389	120	0	136	590	255	223	0
390	121	0	141	591	255	220	0
391	123	0	145	592	255	217	0
392	124	0	149	593	255	213	0
393	126	0	153	594	255	210	0
394	127	0	157	595	255	207	0
395	128	0	161	596	255	203	0
396	129	0	165	597	255	200	0
397	129	0	169	598	255	197	0
398	130	0	173	599	255	193	0
399	130	0	177	600	255	190	0
400	131	0	181	601	255	187	0
401	131	0	185	602	255	183	0
402	131	0	188	603	255	180	0

Tabla A5 Continuación. Equiparación Lambda-RGB pasado en ecuaciones empíricas.

Lambda	R	G	B	Lambda	R	G	B
403	131	0	192	604	255	176	0
404	130	0	196	605	255	173	0
405	130	0	200	606	255	169	0
406	129	0	204	607	255	166	0
407	129	0	207	608	255	162	0
408	128	0	211	609	255	159	0
409	127	0	215	610	255	155	0
410	126	0	219	611	255	152	0
411	124	0	222	612	255	148	0
412	123	0	226	613	255	145	0
413	121	0	230	614	255	141	0
414	120	0	233	615	255	137	0
415	118	0	237	616	255	134	0
416	116	0	241	617	255	130	0
417	113	0	244	618	255	126	0
418	111	0	248	619	255	123	0
419	109	0	251	620	255	119	0
420	106	0	255	621	255	115	0
421	102	0	255	622	255	111	0
422	97	0	255	623	255	107	0
423	93	0	255	624	255	103	0
424	89	0	255	625	255	99	0
425	84	0	255	626	255	95	0
426	80	0	255	627	255	91	0
427	75	0	255	628	255	87	0
428	70	0	255	629	255	83	0
429	66	0	255	630	255	79	0
430	61	0	255	631	255	75	0
431	56	0	255	632	255	70	0
432	51	0	255	633	255	66	0
433	46	0	255	634	255	62	0
434	40	0	255	635	255	57	0
435	35	0	255	636	255	52	0
436	29	0	255	637	255	48	0
437	23	0	255	638	255	43	0
438	17	0	255	639	255	38	0
439	10	0	255	640	255	33	0
440	0	0	255	641	255	27	0
441	0	11	255	642	255	22	0
442	0	19	255	643	255	16	0

Tabla A5 Continuación. Equiparación Lambda-RGB pasado en ecuaciones empíricas.

Lambda	R	G	B	Lambda	R	G	B
443	0	27	255	644	255	9	0
444	0	34	255	645	255	0	0
445	0	40	255	646	255	0	0
446	0	47	255	647	255	0	0
447	0	53	255	648	255	0	0
448	0	59	255	649	255	0	0
449	0	65	255	650	255	0	0
450	0	70	255	651	255	0	0
451	0	76	255	652	255	0	0
452	0	81	255	653	255	0	0
453	0	87	255	654	255	0	0
454	0	92	255	655	255	0	0
455	0	97	255	656	255	0	0
456	0	102	255	657	255	0	0
457	0	108	255	658	255	0	0
458	0	113	255	659	255	0	0
459	0	118	255	660	255	0	0
460	0	123	255	661	255	0	0
461	0	127	255	662	255	0	0
462	0	132	255	663	255	0	0
463	0	137	255	664	255	0	0
464	0	142	255	665	255	0	0
465	0	146	255	666	255	0	0
466	0	151	255	667	255	0	0
467	0	156	255	668	255	0	0
468	0	160	255	669	255	0	0
469	0	165	255	670	255	0	0
470	0	169	255	671	255	0	0
471	0	174	255	672	255	0	0
472	0	178	255	673	255	0	0
473	0	183	255	674	255	0	0
474	0	187	255	675	255	0	0
475	0	192	255	676	255	0	0
476	0	196	255	677	255	0	0
477	0	200	255	678	255	0	0
478	0	205	255	679	255	0	0
479	0	209	255	680	255	0	0
480	0	213	255	681	255	0	0
481	0	218	255	682	255	0	0
482	0	222	255	683	255	0	0

Tabla A5 Continuación. Equiparación Lambda-RGB pasado en ecuaciones empíricas.

Lambda	R	G	B	Lambda	R	G	B
483	0	226	255	684	255	0	0
484	0	230	255	685	255	0	0
485	0	234	255	686	255	0	0
486	0	239	255	687	255	0	0
487	0	243	255	688	255	0	0
488	0	247	255	689	255	0	0
489	0	251	255	690	255	0	0
490	0	255	255	691	255	0	0
491	0	255	245	692	255	0	0
492	0	255	234	693	255	0	0
493	0	255	224	694	255	0	0
494	0	255	213	695	255	0	0
495	0	255	203	696	255	0	0
496	0	255	192	697	255	0	0
497	0	255	181	698	255	0	0
498	0	255	169	699	255	0	0
499	0	255	158	700	255	0	0
500	0	255	146	701	253	0	0
501	0	255	135	702	251	0	0
502	0	255	123	703	250	0	0
503	0	255	110	704	248	0	0
504	0	255	97	705	246	0	0
505	0	255	84	706	244	0	0
506	0	255	70	707	242	0	0
507	0	255	56	708	241	0	0
508	0	255	40	709	239	0	0
509	0	255	23	710	237	0	0
510	0	255	0	711	235	0	0
511	9	255	0	712	233	0	0
512	15	255	0	713	232	0	0
513	21	255	0	714	230	0	0
514	26	255	0	715	228	0	0
515	31	255	0	716	226	0	0
516	36	255	0	717	224	0	0
517	40	255	0	718	222	0	0
518	45	255	0	719	220	0	0
519	49	255	0	720	219	0	0
520	54	255	0	721	217	0	0
521	58	255	0	722	215	0	0

Tabla A5 Continuación. Equiparación Lambda-RGB pasado en ecuaciones empíricas.

Lambda	R	G	B	Lambda	R	G	B
522	62	255	0	723	213	0	0
523	66	255	0	724	211	0	0
524	70	255	0	725	209	0	0
525	74	255	0	726	207	0	0
526	78	255	0	727	206	0	0
527	82	255	0	728	204	0	0
528	86	255	0	729	202	0	0
529	90	255	0	730	200	0	0
530	94	255	0	731	198	0	0
531	97	255	0	732	196	0	0
532	101	255	0	733	194	0	0
533	105	255	0	734	192	0	0
534	108	255	0	735	190	0	0
535	112	255	0	736	188	0	0
536	115	255	0	737	186	0	0
537	119	255	0	738	185	0	0
538	123	255	0	739	183	0	0
539	126	255	0	740	181	0	0
540	129	255	0	741	179	0	0
541	133	255	0	742	177	0	0
542	136	255	0	743	175	0	0
543	140	255	0	744	173	0	0
544	143	255	0	745	171	0	0
545	146	255	0	746	169	0	0
546	150	255	0	747	167	0	0
547	153	255	0	748	165	0	0
548	156	255	0	749	163	0	0
549	160	255	0	750	161	0	0
550	163	255	0	751	159	0	0
551	166	255	0	752	157	0	0
552	169	255	0	753	155	0	0
553	173	255	0	754	153	0	0
554	176	255	0	755	151	0	0
555	179	255	0	756	149	0	0
556	182	255	0	757	147	0	0
557	185	255	0	758	145	0	0
558	189	255	0	759	143	0	0
559	192	255	0	760	141	0	0
560	195	255	0	761	138	0	0

Tabla A5 Continuación. Equiparación Lambda-RGB pasado en ecuaciones empíricas.

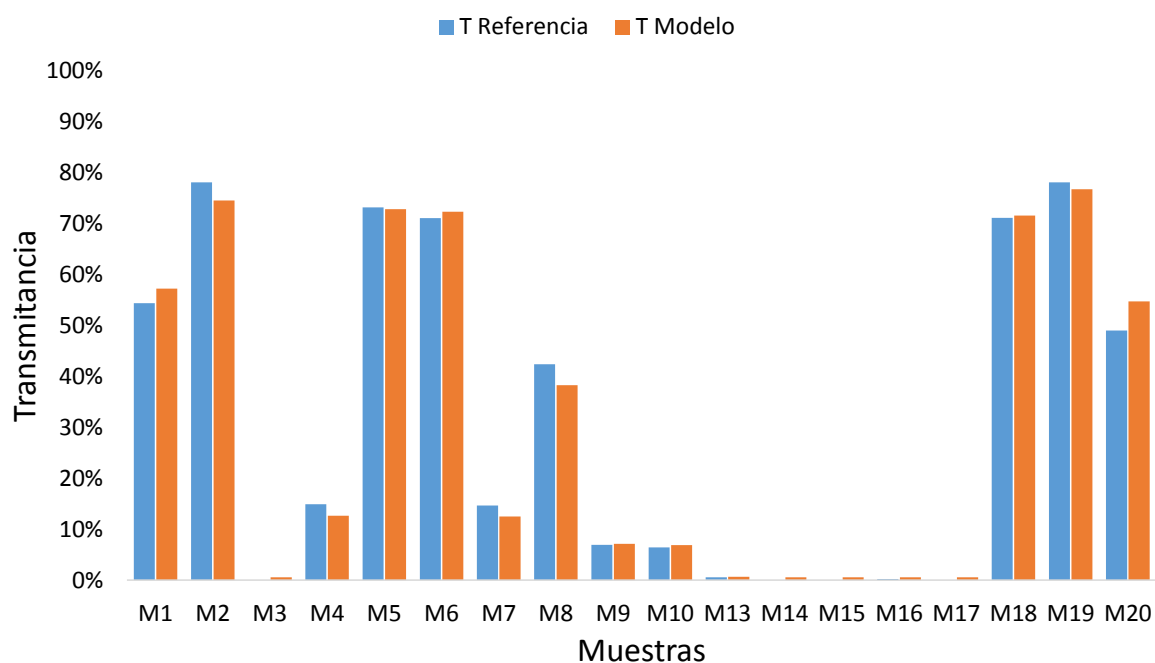
Lambda	R	G	B	Lambda	R	G	B
561	198	255	0	762	136	0	0
562	201	255	0	763	134	0	0
563	204	255	0	764	132	0	0
564	207	255	0	765	130	0	0
565	210	255	0	766	128	0	0
566	213	255	0	767	126	0	0
567	216	255	0	768	124	0	0
568	219	255	0	769	122	0	0
569	222	255	0	770	119	0	0
570	225	255	0	771	117	0	0
571	228	255	0	772	115	0	0
572	231	255	0	773	113	0	0
573	234	255	0	774	111	0	0
574	237	255	0	775	109	0	0
575	240	255	0	776	106	0	0
576	243	255	0	777	104	0	0
577	246	255	0	778	102	0	0
578	249	255	0	779	100	0	0
579	252	255	0	780	97	0	0
580	255	255	0				

Anexo B

380nm

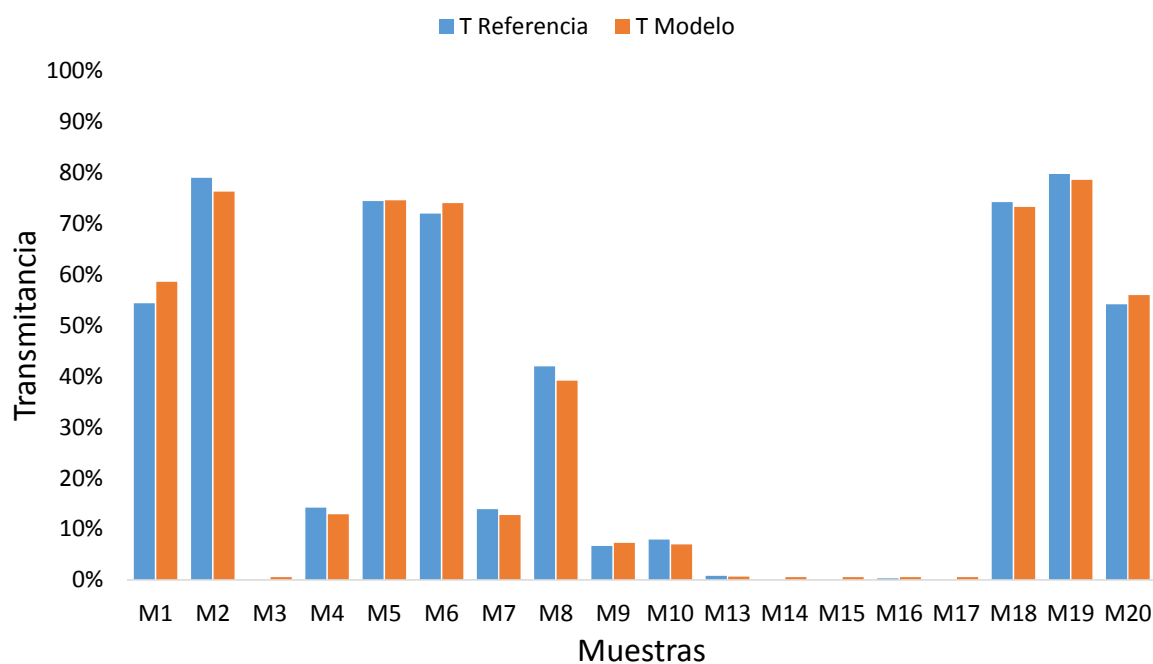
λ_{picoLED}	R^2	Ecuación
385nm	0,9952	$T = 0,7715x + 0,0057$

Gráfica resultado del ajuste



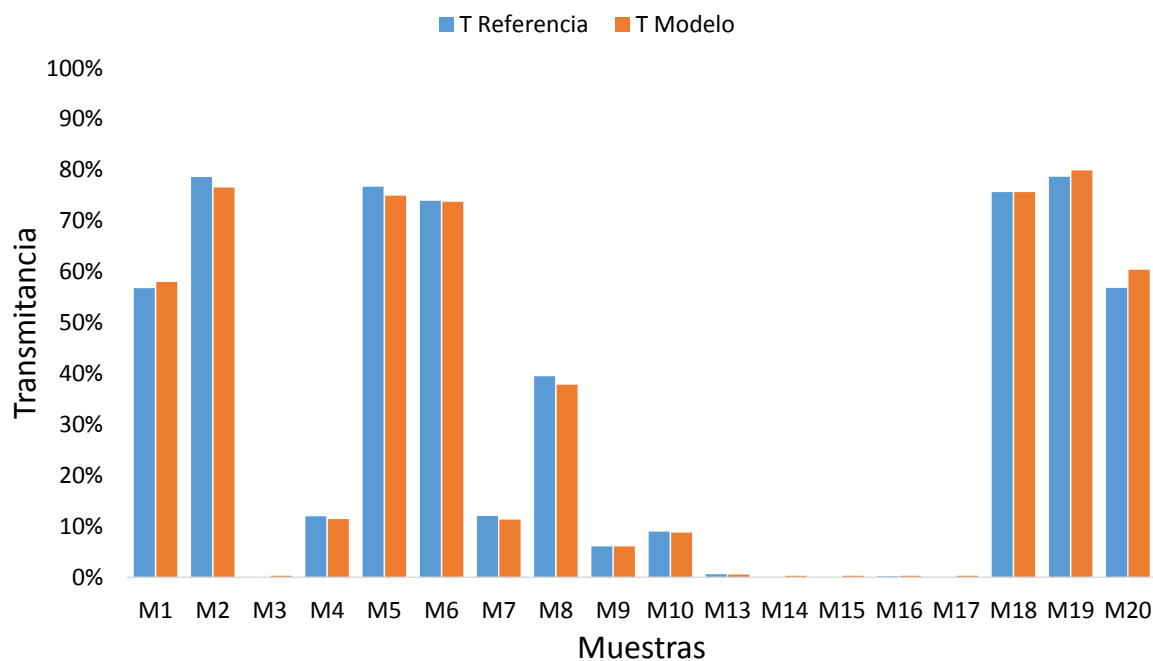
385nm

λ_{picoLED}	R^2	Ecuación
385nm	0,9974	$T = 0,7898x + 0,0053$

Gráfica resultado del ajuste

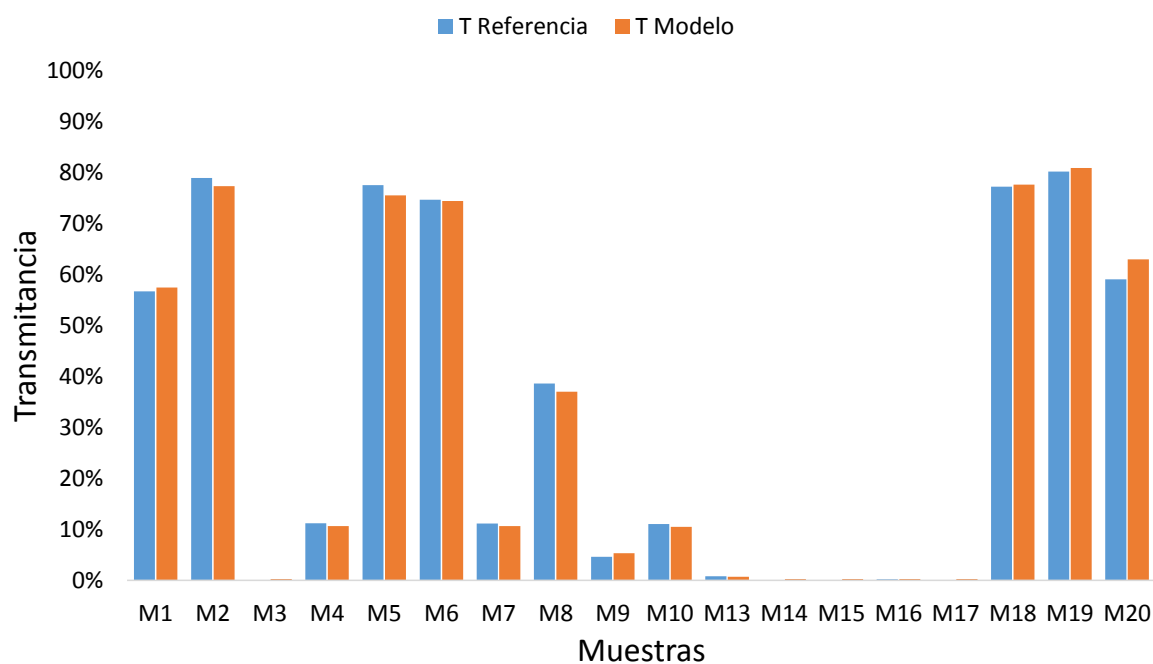
390nm

LED	R^2	Ecuación
390nm	0,9986	$T = 0,8113x + 0,0028$

Gráfica resultado del ajuste

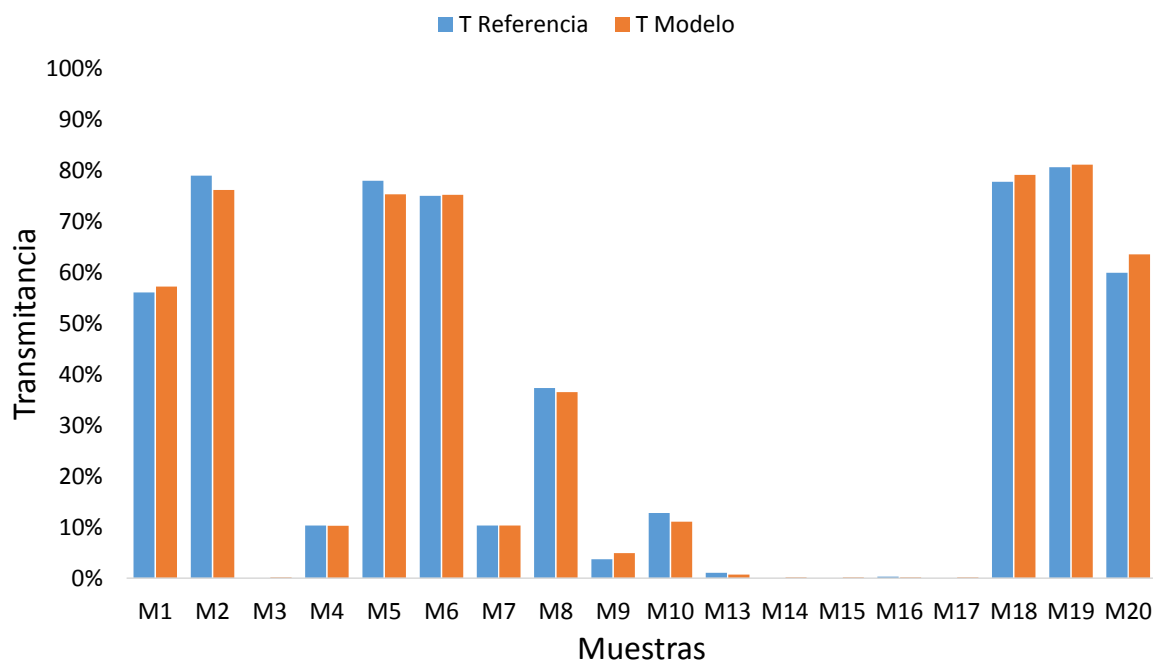
395nm

LED	R ²	Ecuación
390nm	0,9989	$T = 0,8169x + 0,0023$

Gráfica resultado del ajuste

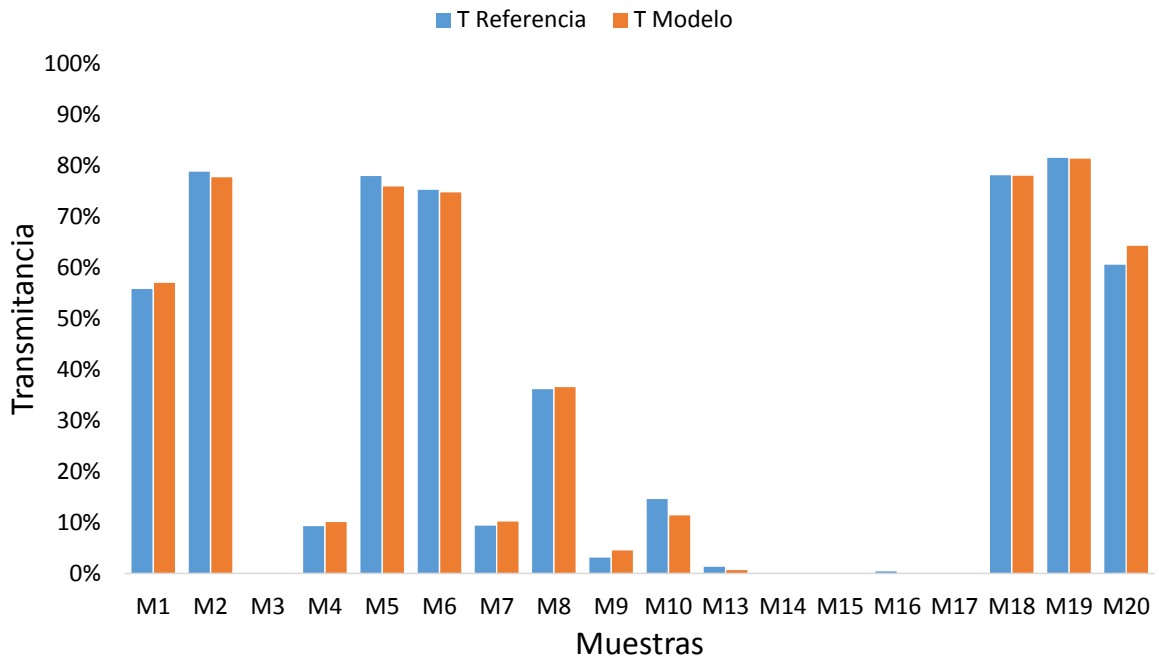
400nm

LED	R ²	Ecuación
400nm	0,9982	$T = 0,8196x + 0,0015$

Gráfica resultado del ajuste

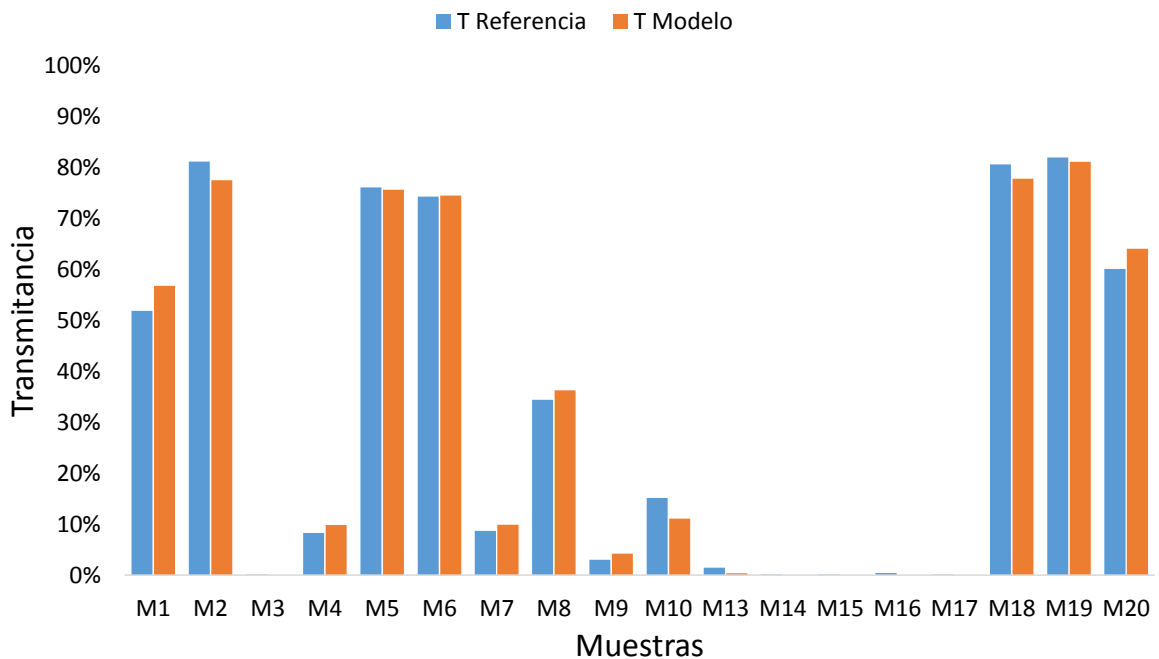
405nm

LED	R ²	Ecuación
405nm	0,9983	$T = 0,825x + 0,0001$

Gráfica resultado del ajuste

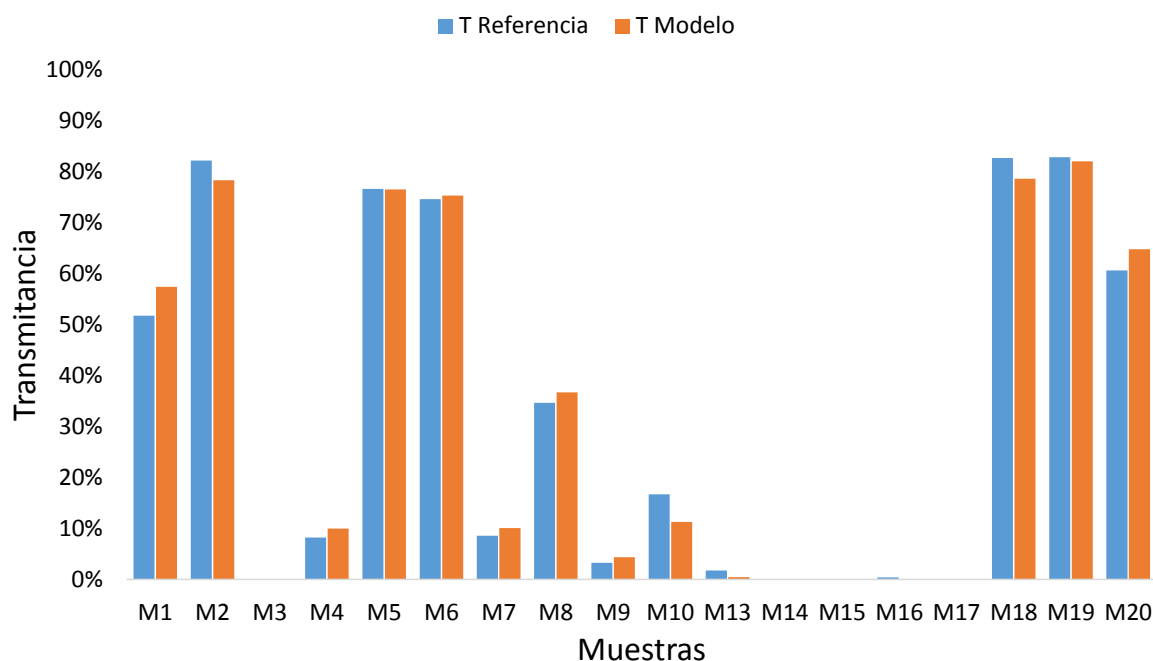
410nm

LED	R ²	Ecuación
405nm	0,9956	$T = 0,8251x - 0,0026$

Gráfica resultado del ajuste

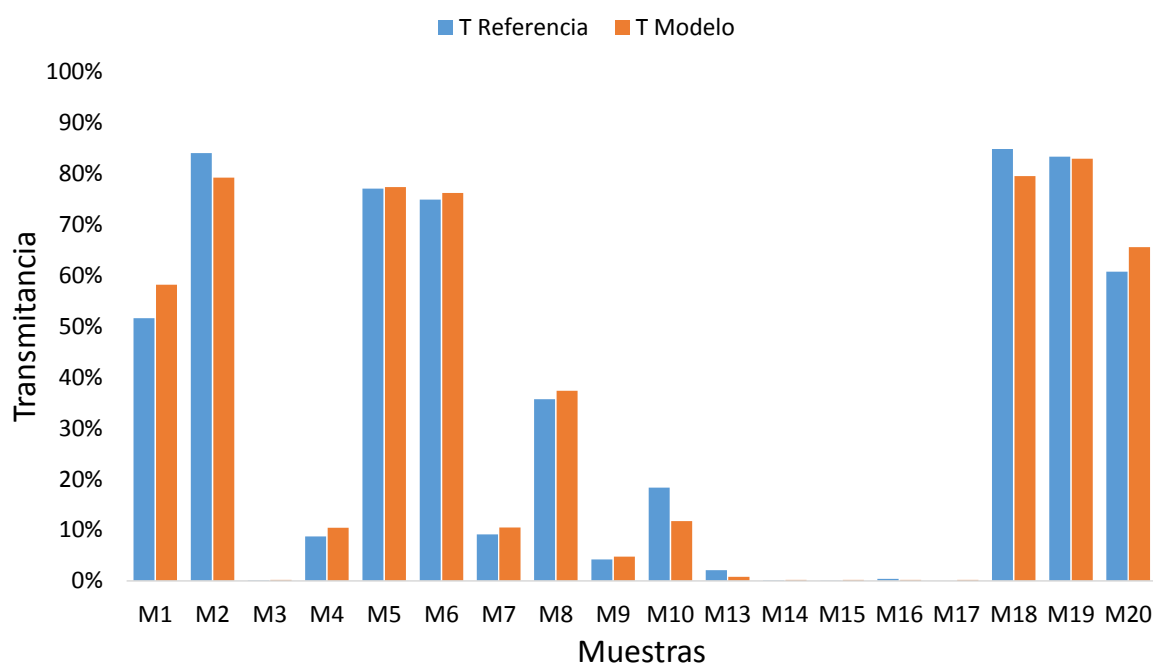
415nm

LED	R ²	Ecuación
405nm	0,9941	$T = 0,8337x - 0,0019$

Gráfica resultado del ajuste

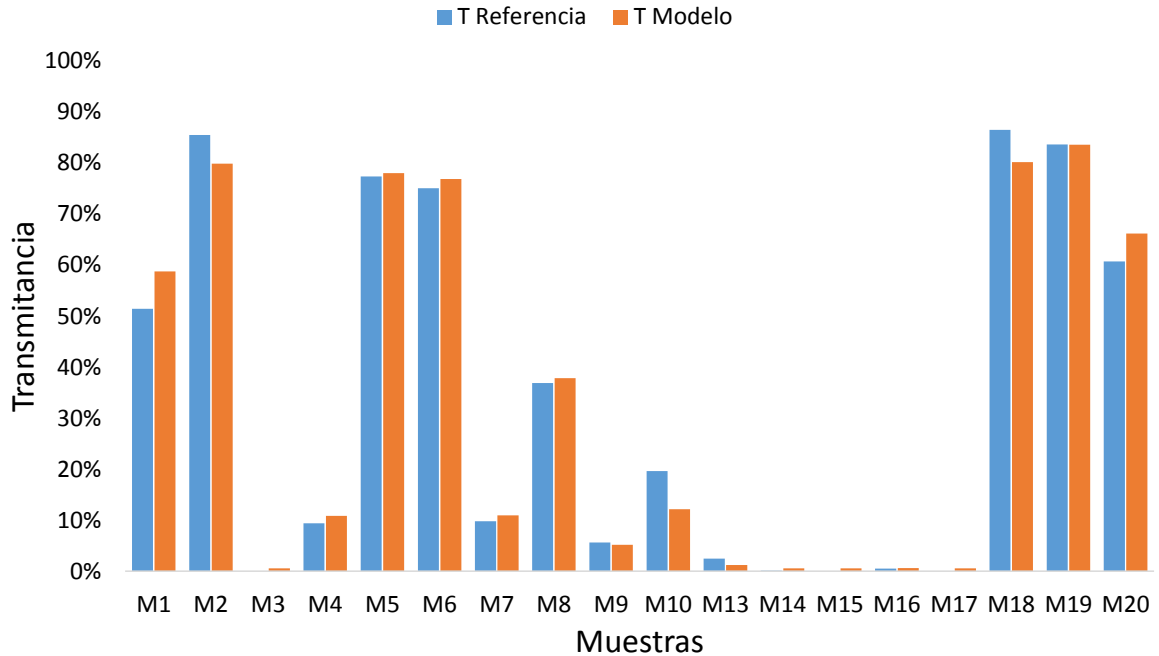
420nm

LED	R ²	Ecuación
405nm	0,9916	$T = 0,8382x + 0,0018$
428nm	0,9859	$T = 0,8917x + 0,051$

Gráfica resultado del ajuste

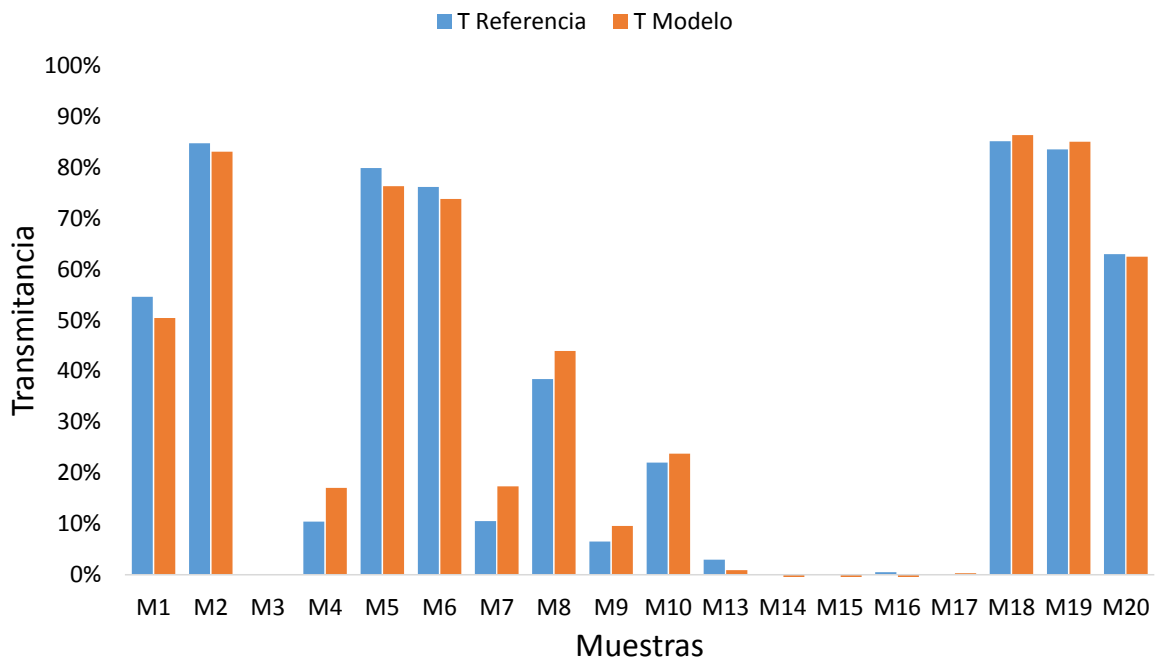
425nm

LED	R ²	Ecuación
405nm	0,9897	$T = 0,8416x + 0,0057$
428nm	0,9879	$T = 0,8973x + 0,0481$

Gráfica resultado del ajuste

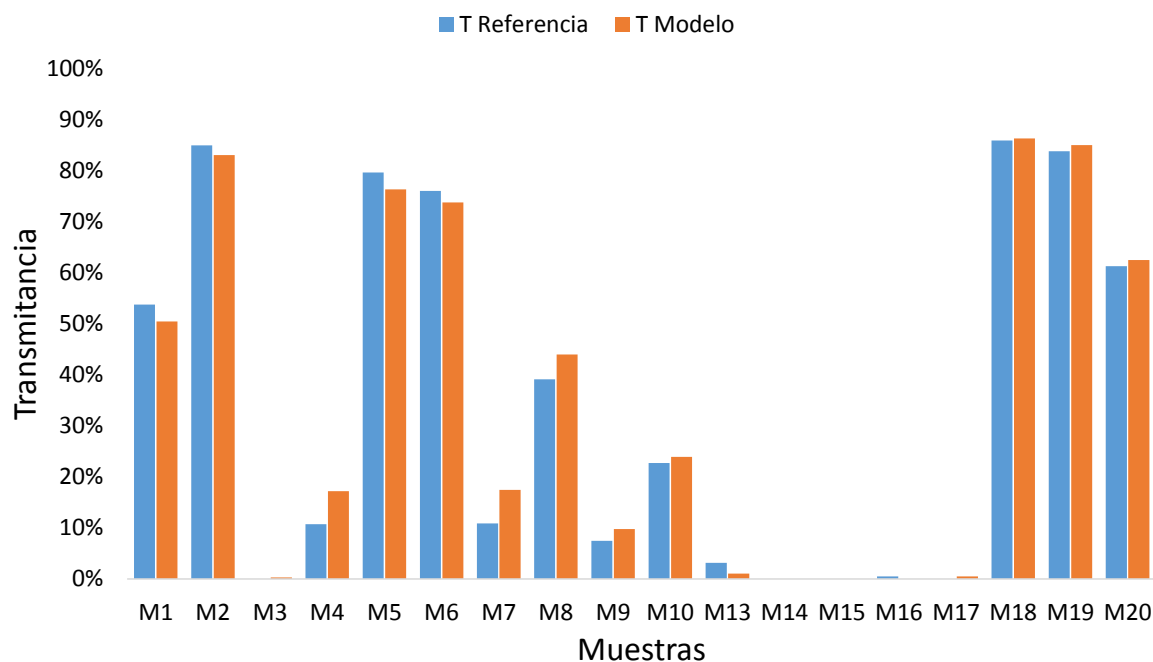
428nm

LED	R ²	Ecuación
428nm	0,9892	$T = 0,9004x - 0,0418$

Gráfica resultado del ajuste

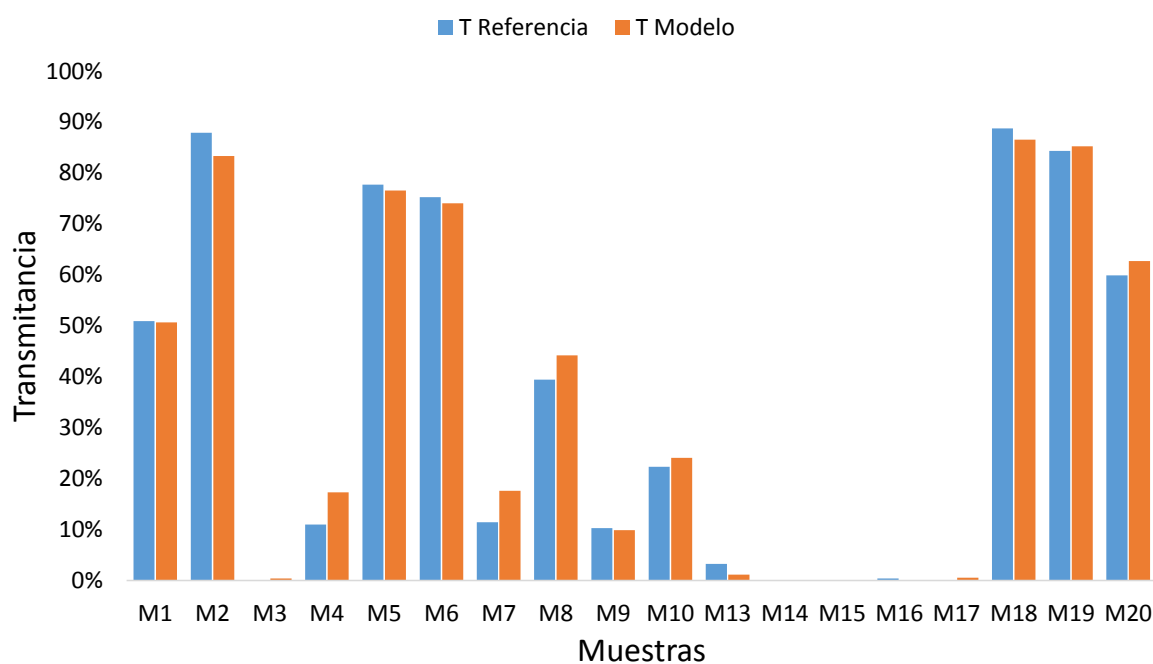
430nm

LED	R ²	Ecuación
405nm	0,9897	$T = 0,8416x + 0,0014$
428nm	0,9906	$T = 0,8985x - 0,0399$
430nm	0,9908	$T = 0,8942x - 0,0388$

Gráfica resultado del ajuste

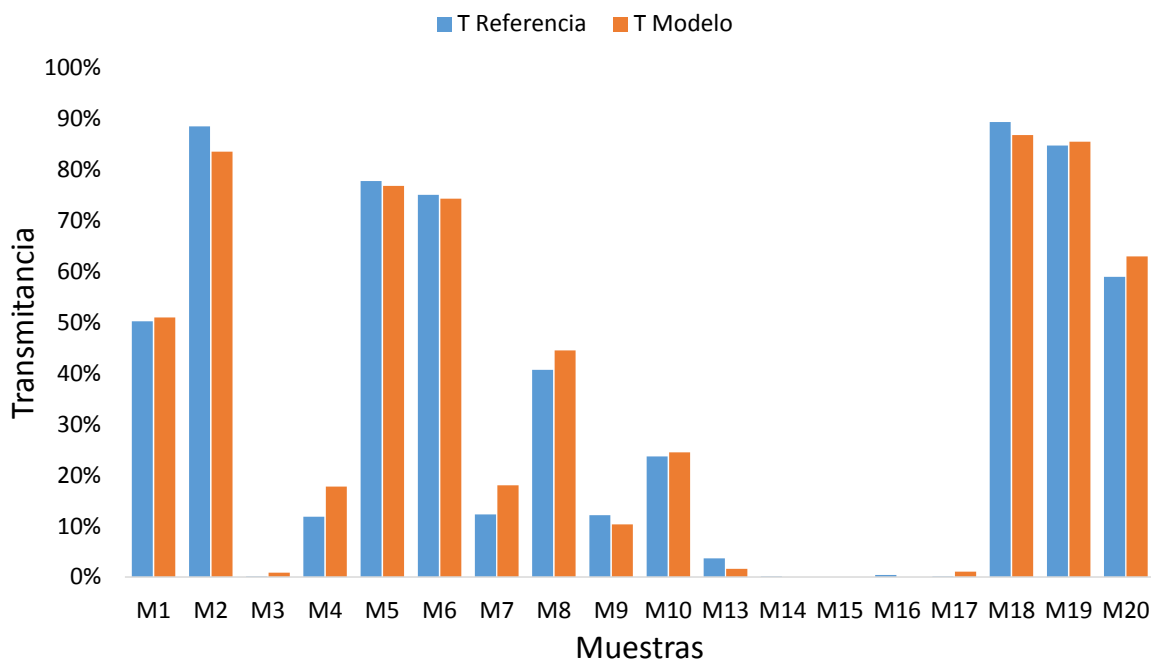
435nm

LED	R ²	Ecuación
428nm	0,9911	$T = 0,9x - 0,0391$
430nm	0,9914	$T = 0,8958x + 0,0379$

Gráfica resultado del ajuste

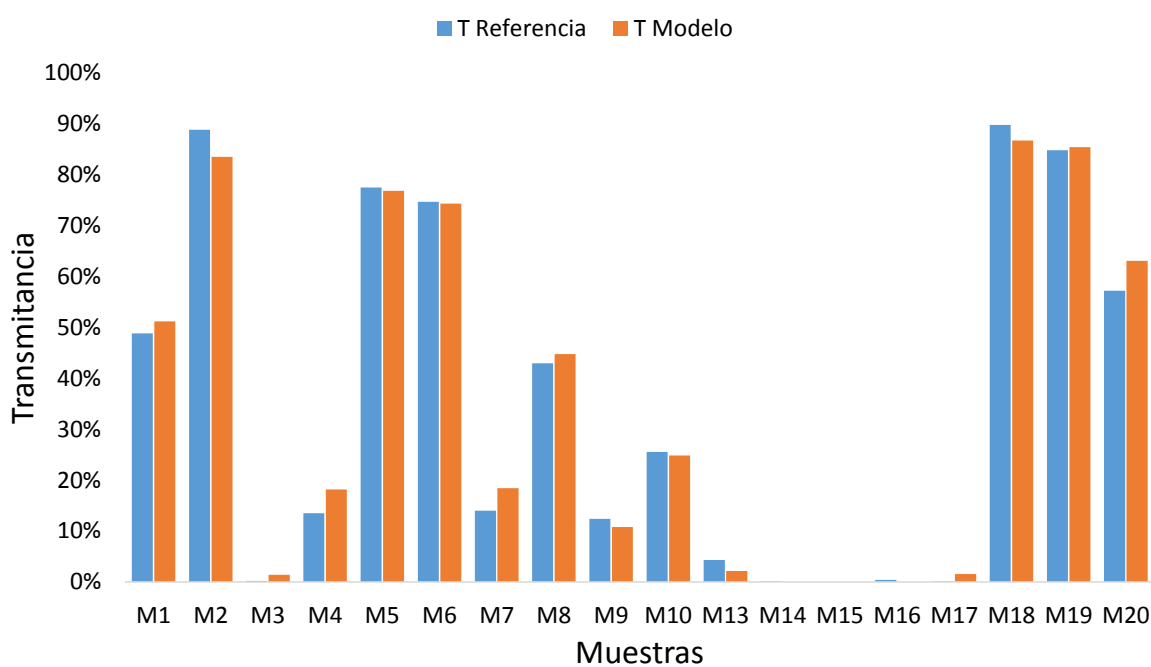
440nm

LED	R ²	Ecuación
428nm	0,9917	$T = 0,8973x - 0,0342$
430nm	0,9919	$T = 0,8931x - 0,0331$

Gráfica resultado del ajuste

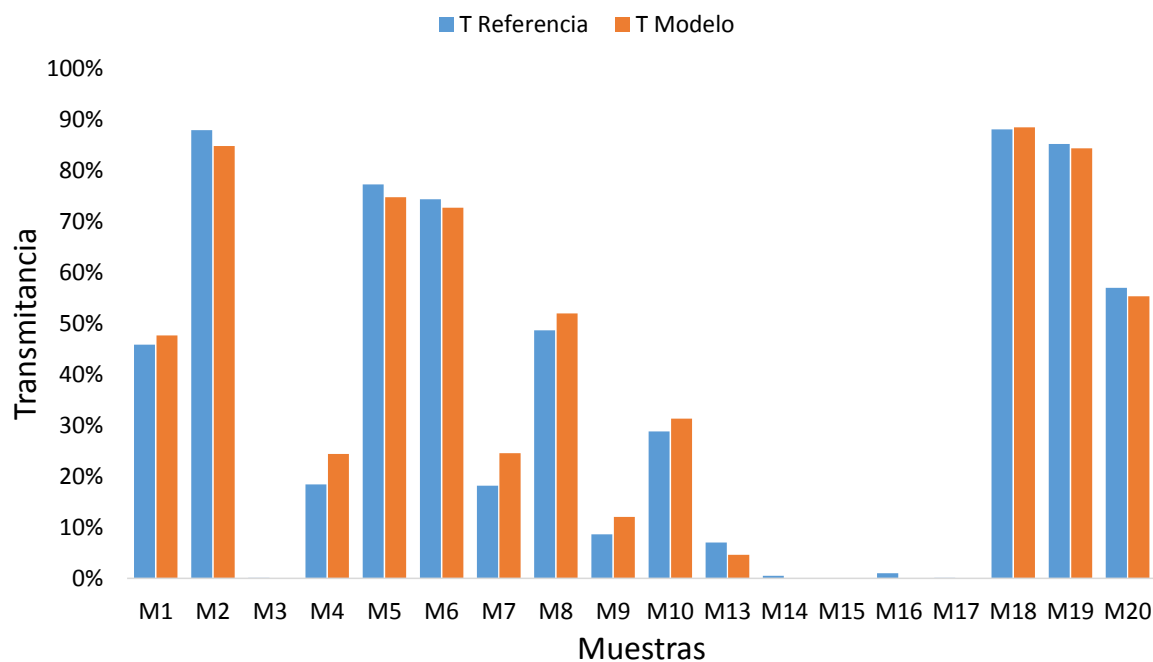
445nm

LED	R ²	Ecuación
428nm	0,9924	$T = 0,891x - 0,0286$
430nm	0,9927	$T = 0,8868x - 0,0274$
461nm0	0,9814	$T = 0,8832x - 0,0323$

Gráfica resultado del ajuste

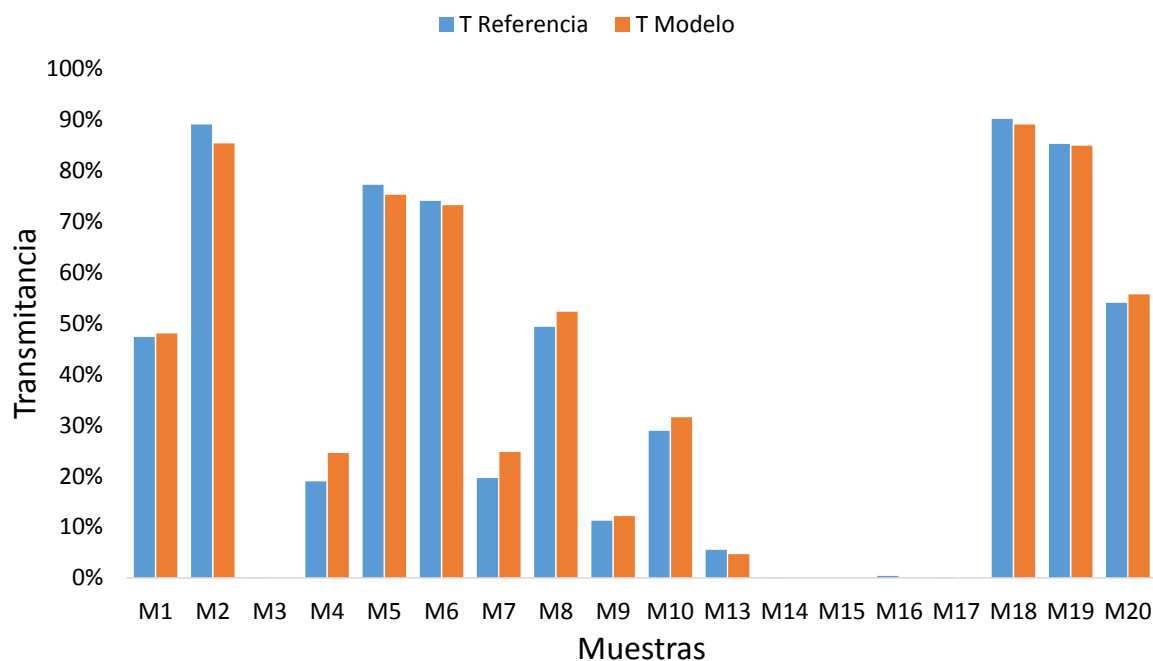
450nm

LED	R ²	Ecuación
461nm	0,9918	T = 0,8728x - 0,0214

Gráfica resultado del ajuste

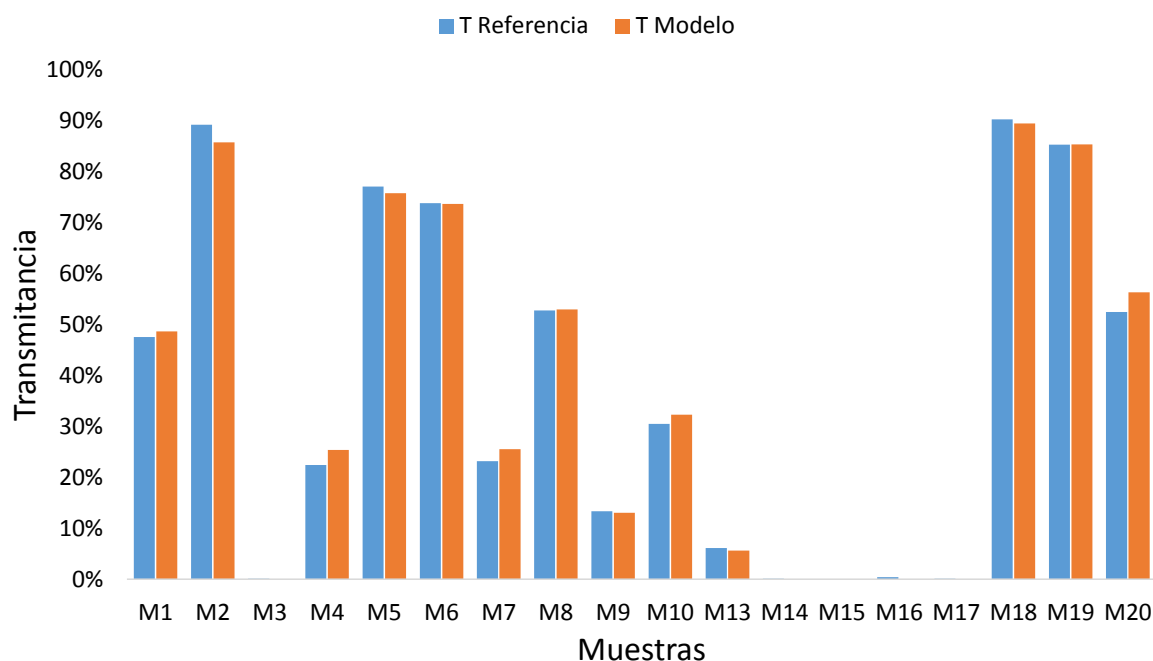
455nm

LED	R ²	Ecuación
461nm	0,9941	T = 0,879x - 0,0214

Gráfica resultado del ajuste

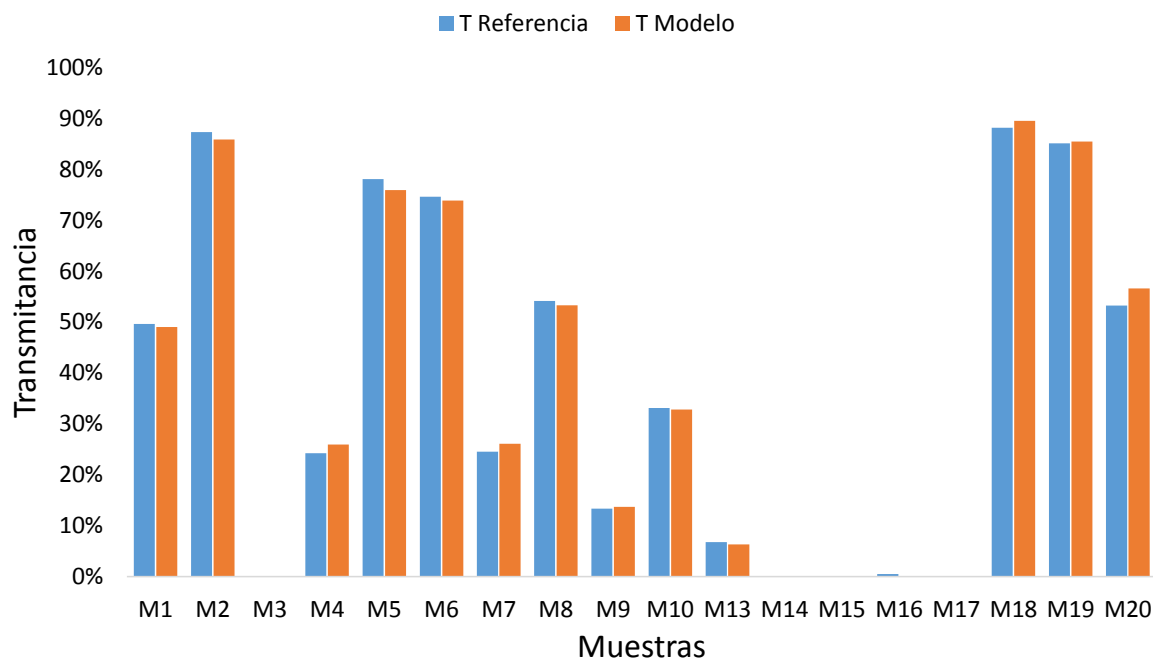
460nm

LED	R ²	Ecuación
461nm	0,9973	T = 0,8738x - 0,0115

Gráfica resultado del ajuste

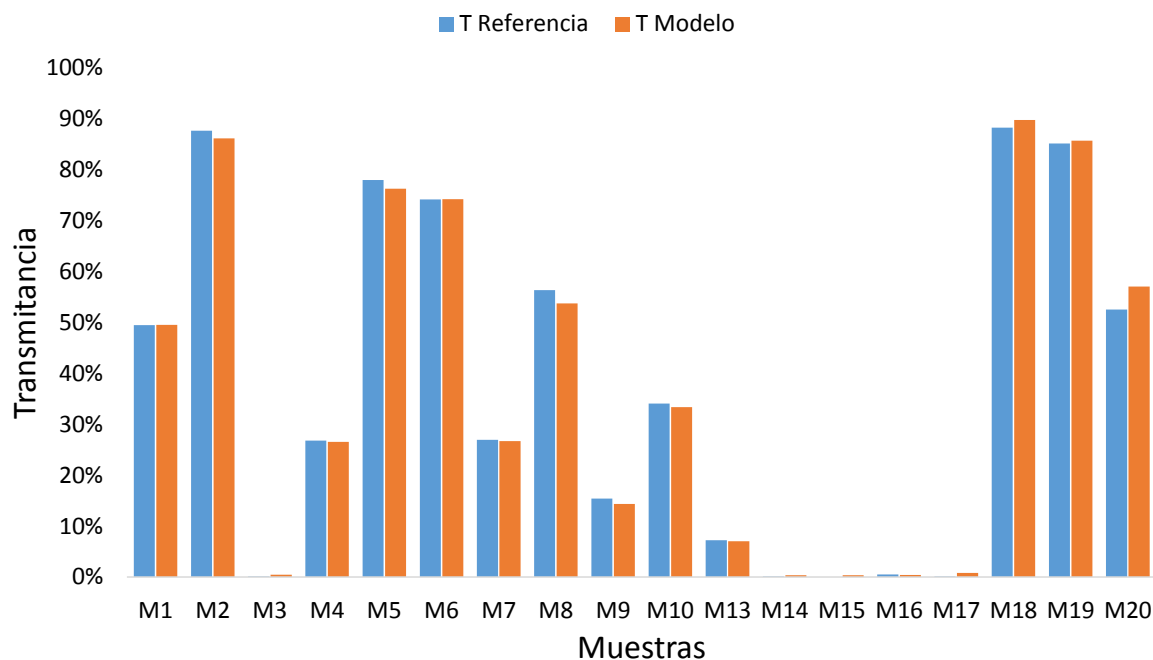
461nm

LED	R ²	Ecuación
461nm	0,9986	T = 0,8665x - 0,0039

Gráfica resultado del ajuste

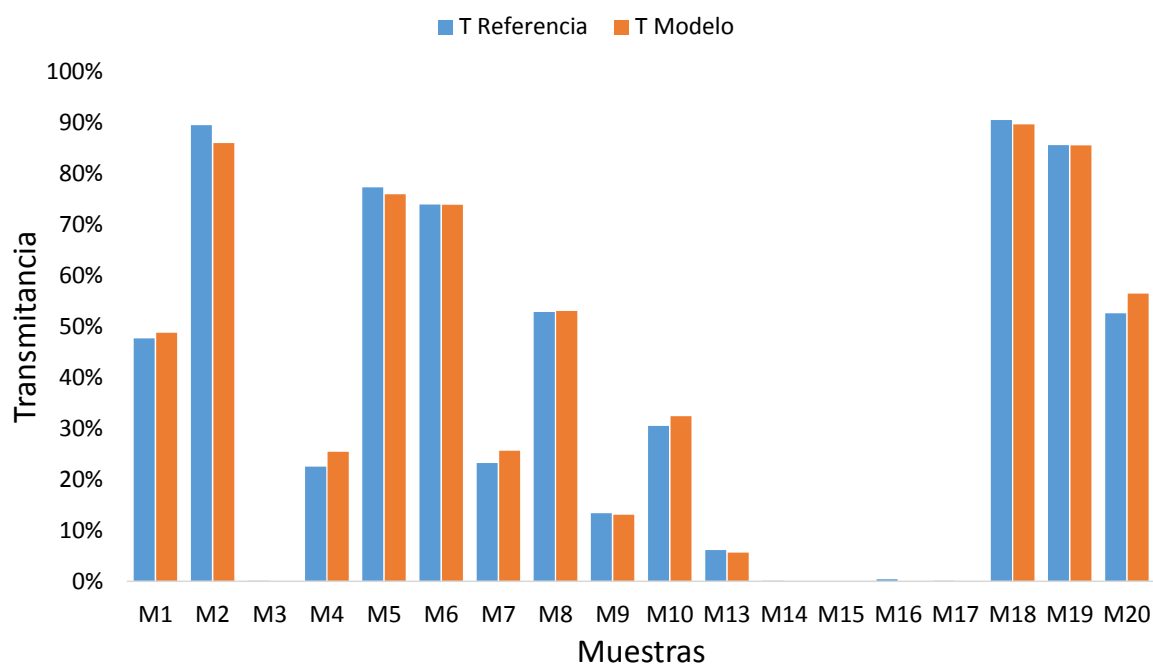
464nm

LED	R ²	Ecuación
461nm	0,9981	$T = 0,8621x + 0,0034$
464nm	0,9977	$T = 0,8749x - 0,0064$

Gráfica resultado del ajuste

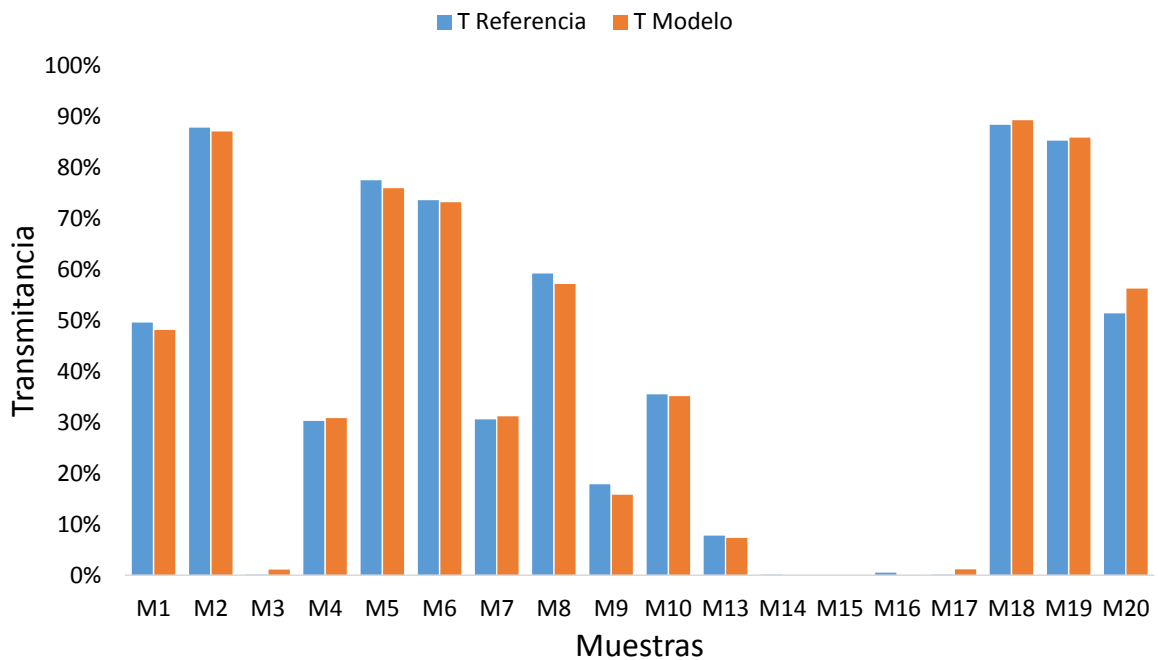
465nm

LED	R ²	Ecuación
461nm	0,9972	$T = 0,8747x - 0,0117$
464nm	0,994	$T = 0,8864x - 0,021$

Gráfica resultado del ajuste

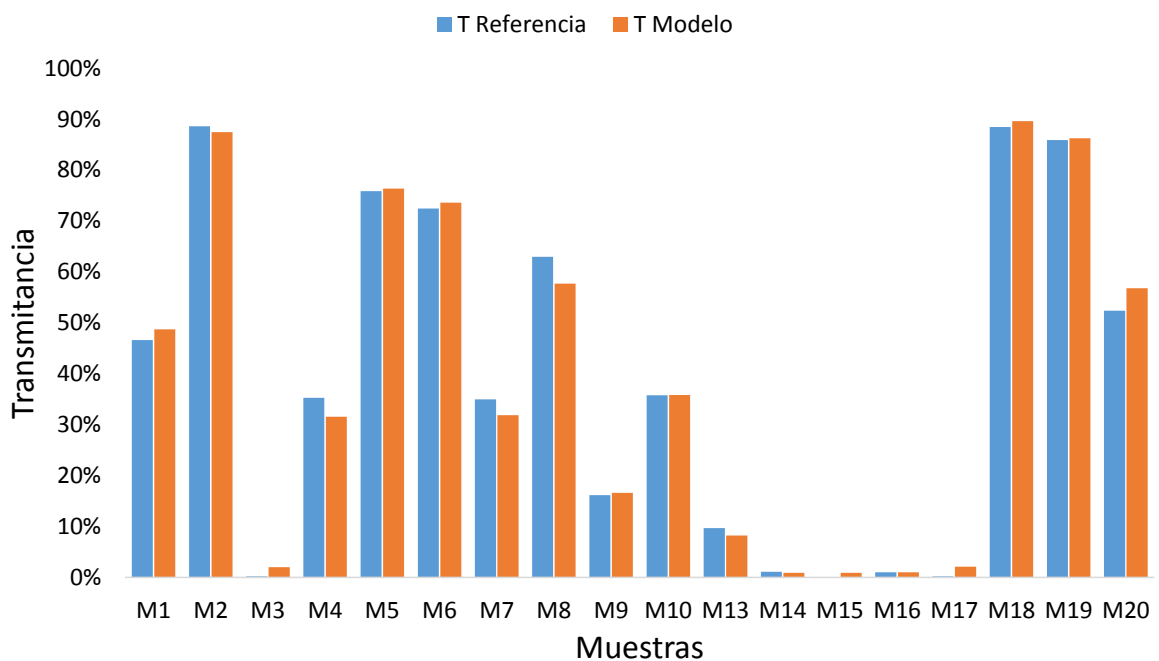
468nm

LED	R ²	Ecuación
468nm	0,9978	$T = 0,8753x - 0,0002$

Gráfica resultado del ajuste

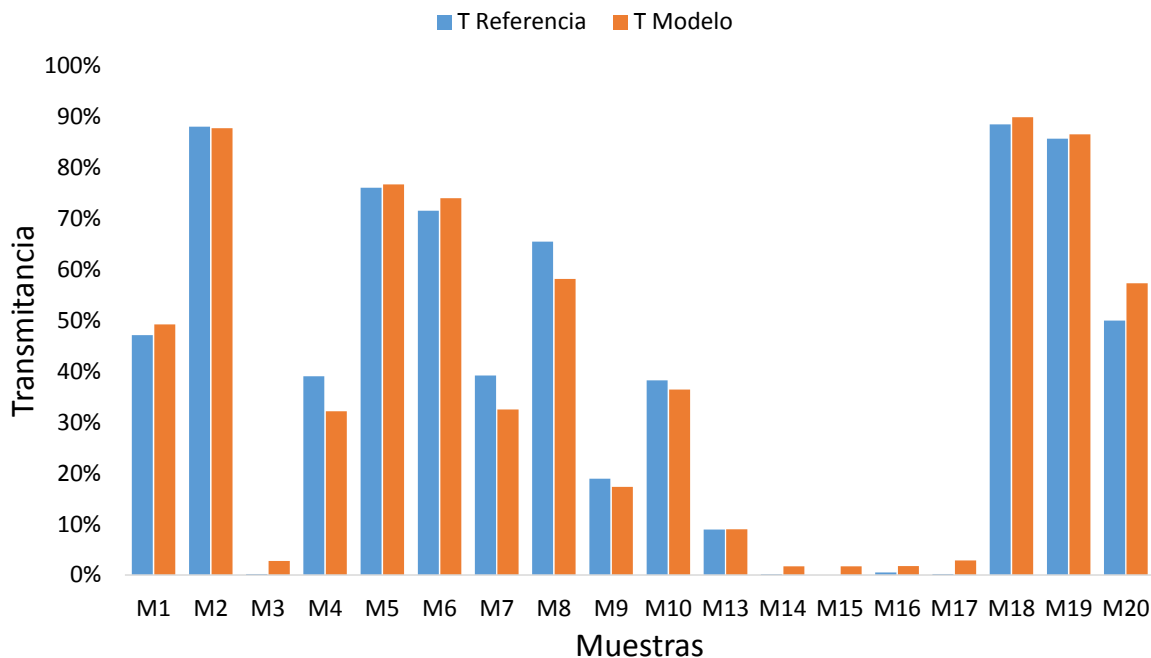
470nm

LED	R ²	Ecuación
468nm	0,9953	$T = 0,8685x + 0,0092$
470nm	0,9949	$T = 0,859x + 0,0102$

Gráfica resultado del ajuste

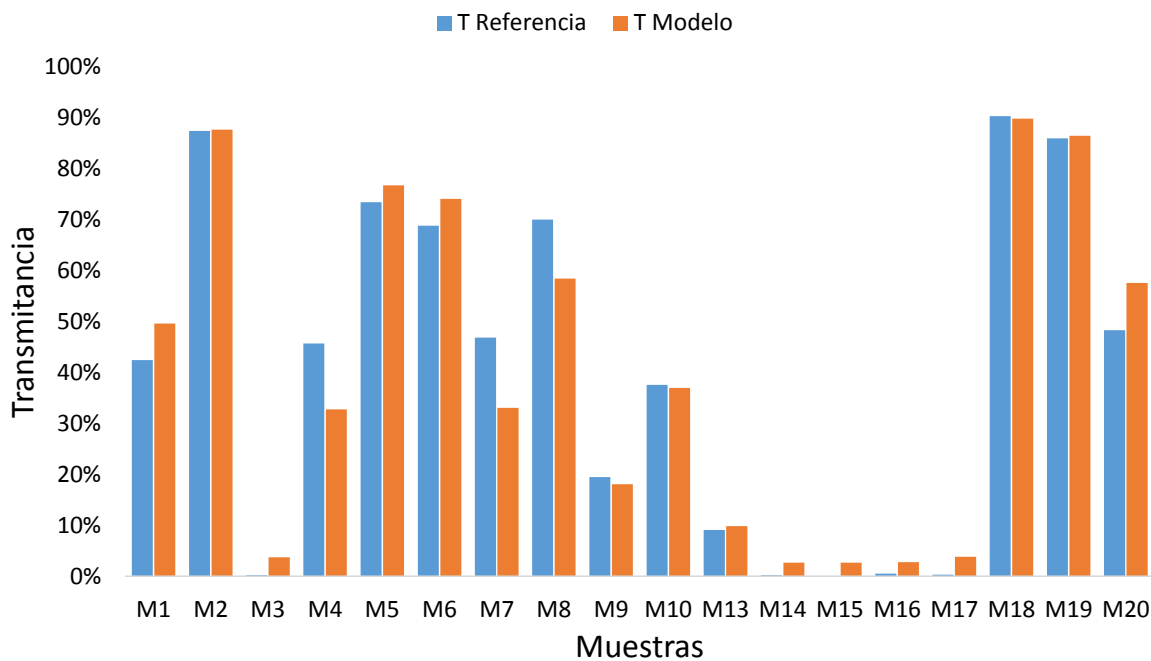
475nm

LED	R ²	Ecuación
468nm	0,9875	$T = 0,864x + 0,0169$
475nm	0,9851	$T = 0,8552x + 0,0204$

Gráfica resultado del ajuste

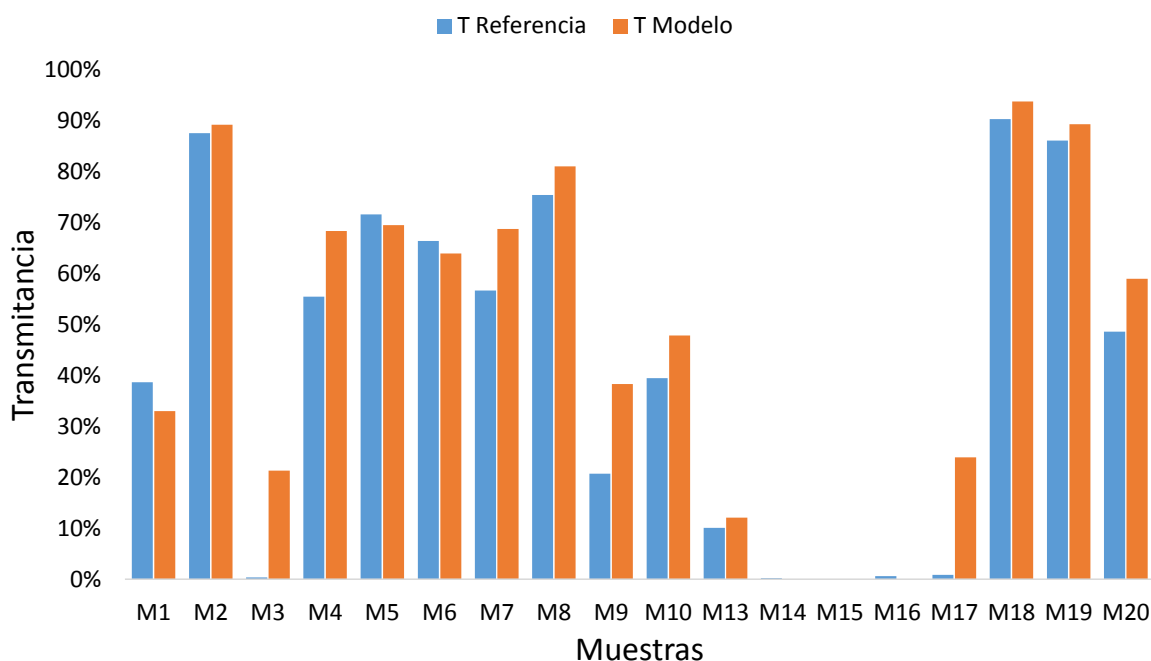
480nm

LED	R ²	Ecuación
468nm	0,9628	$T = 0,8537x + 0,0265$
475nm	0,9591	$T = 0,8444x + 0,0302$

Gráfica resultado del ajuste

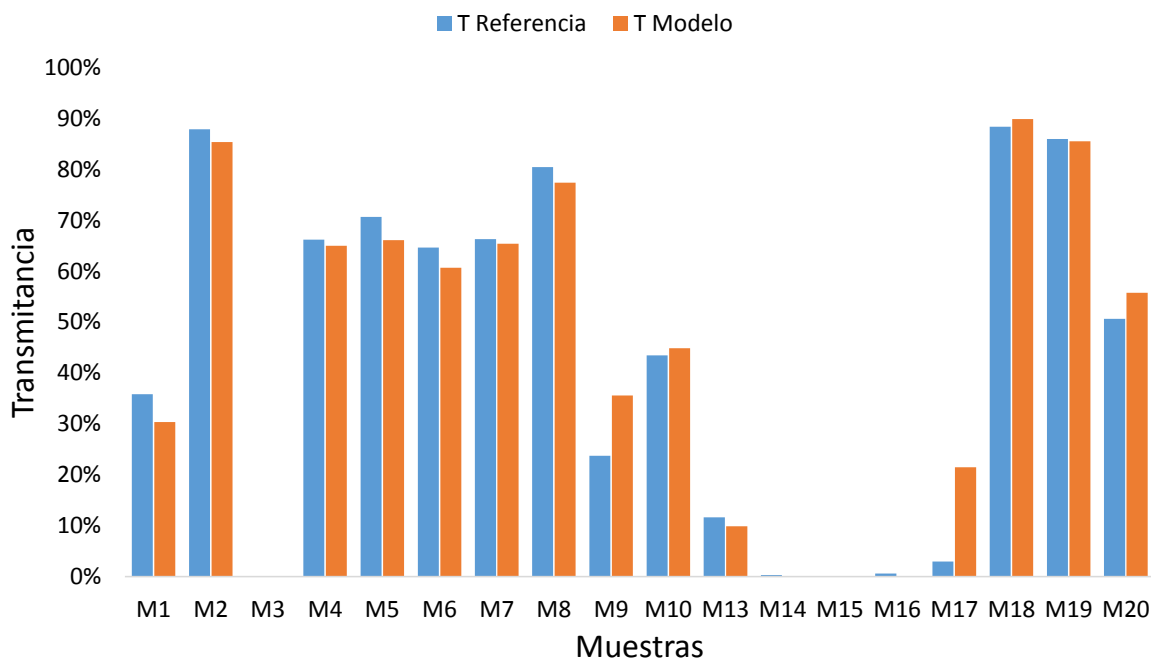
485nm

LED	R ²	Ecuación
500nm	0,9354	T = 0,9198x - 0,0737

Gráfica resultado del ajuste

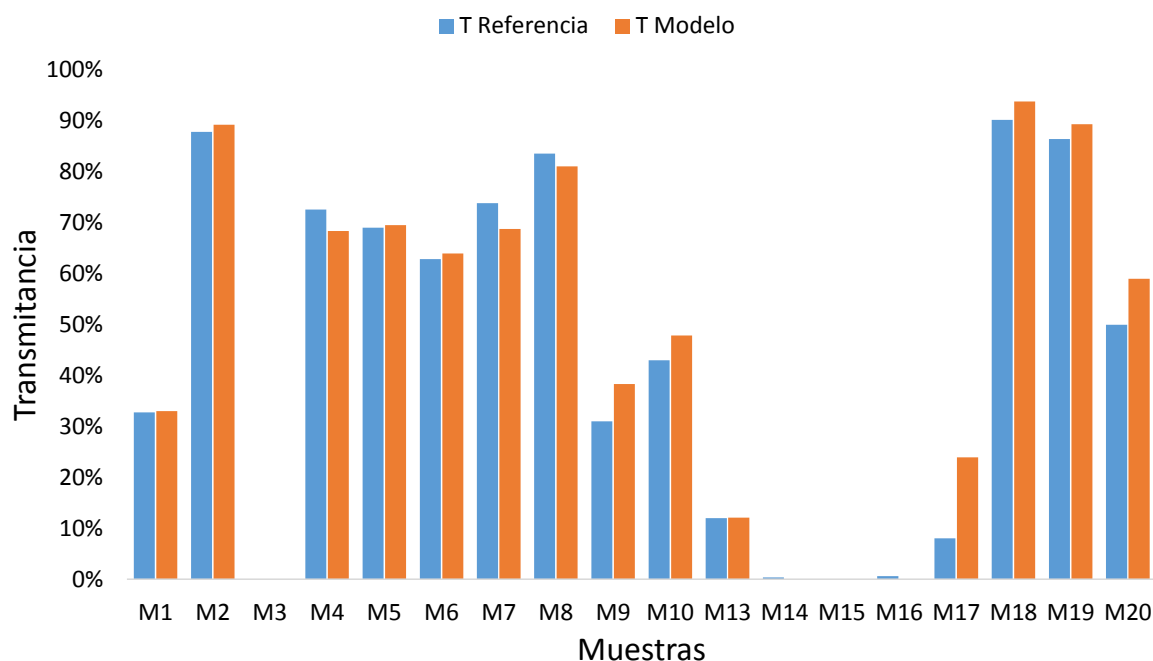
490nm

LED	R ²	Ecuación
500nm	0,9673	T = 0,9053x - 0,0367

Gráfica resultado del ajuste

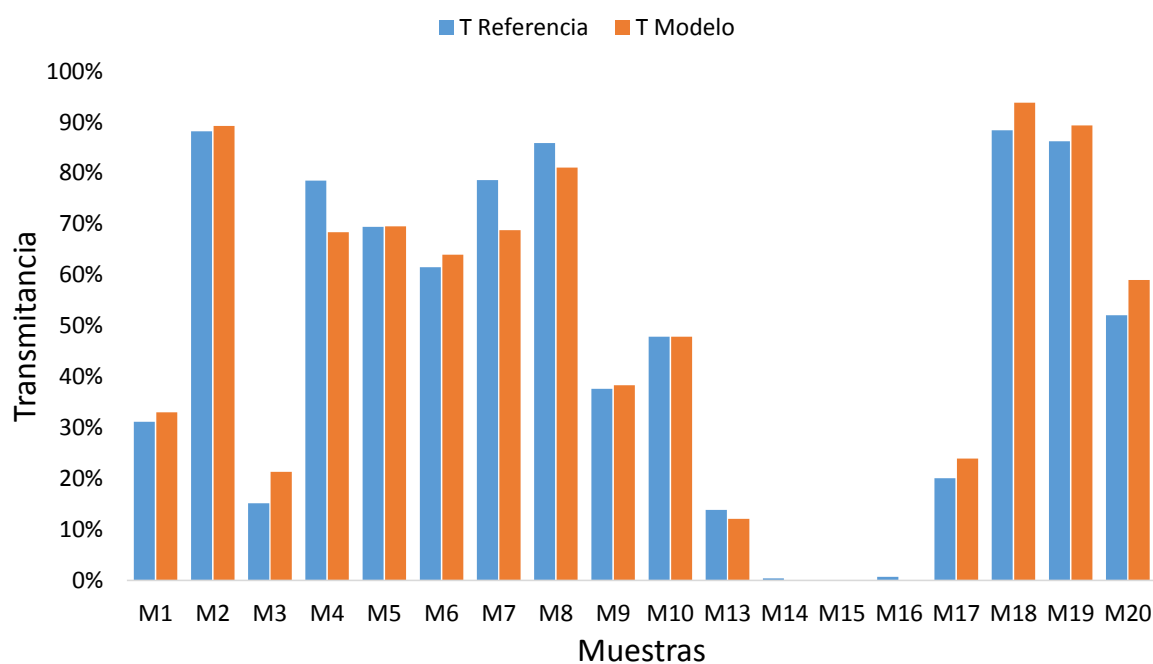
495nm

LED	R ²	Ecuación
500nm	0,9778	$T = 0,9197x - 0,0305$

Gráfica resultado del ajuste

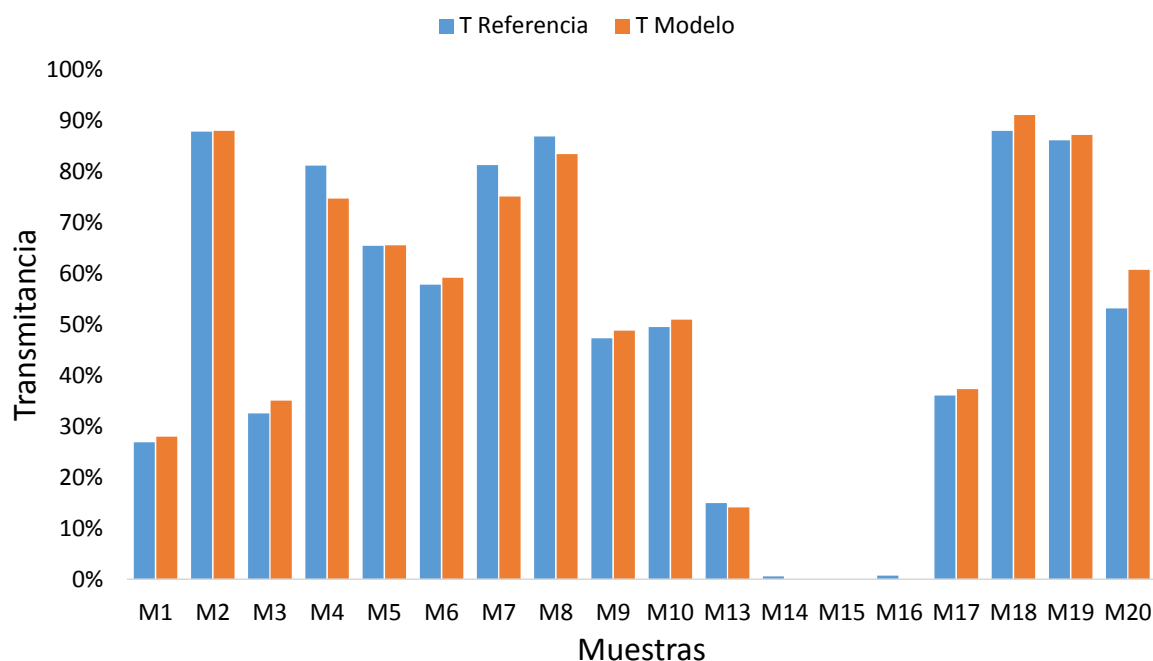
500nm

LED	R ²	Ecuación
500nm	0,99	$T = 0,9254x - 0,0177$
501nm	0,9787	$T = 0,915x - 0,0176$

Gráfica resultado del ajuste

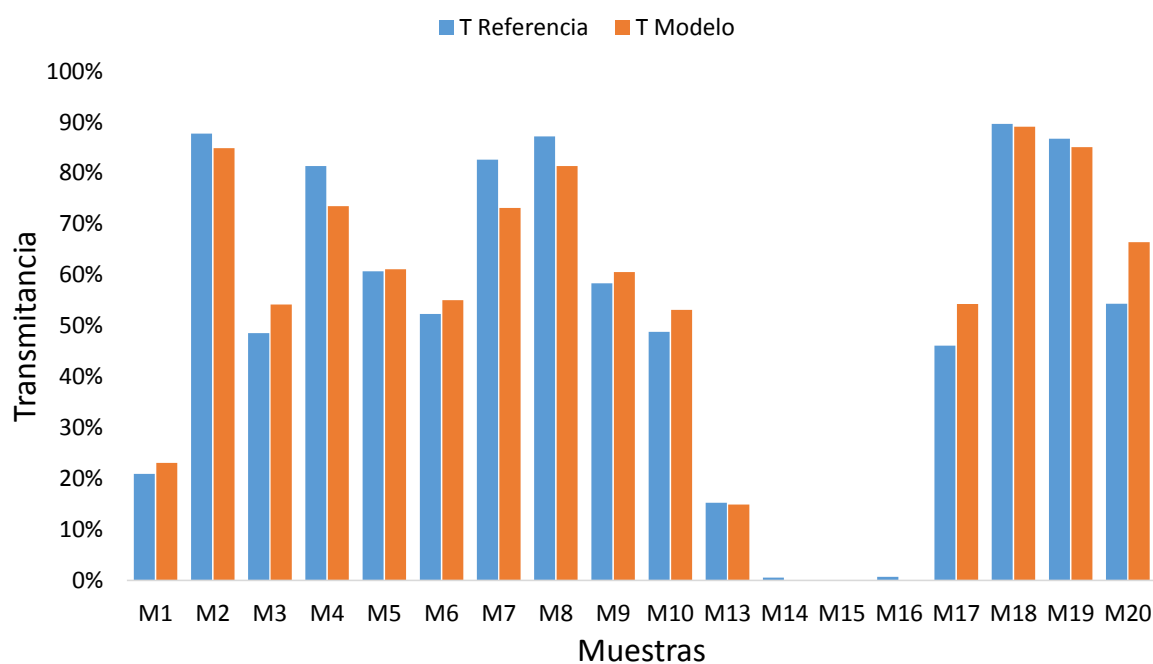
505nm

LED	R ²	Ecuación
505nm	0,9896	$T = 0,9061x - 0,0149$

Gráfica resultado del ajuste

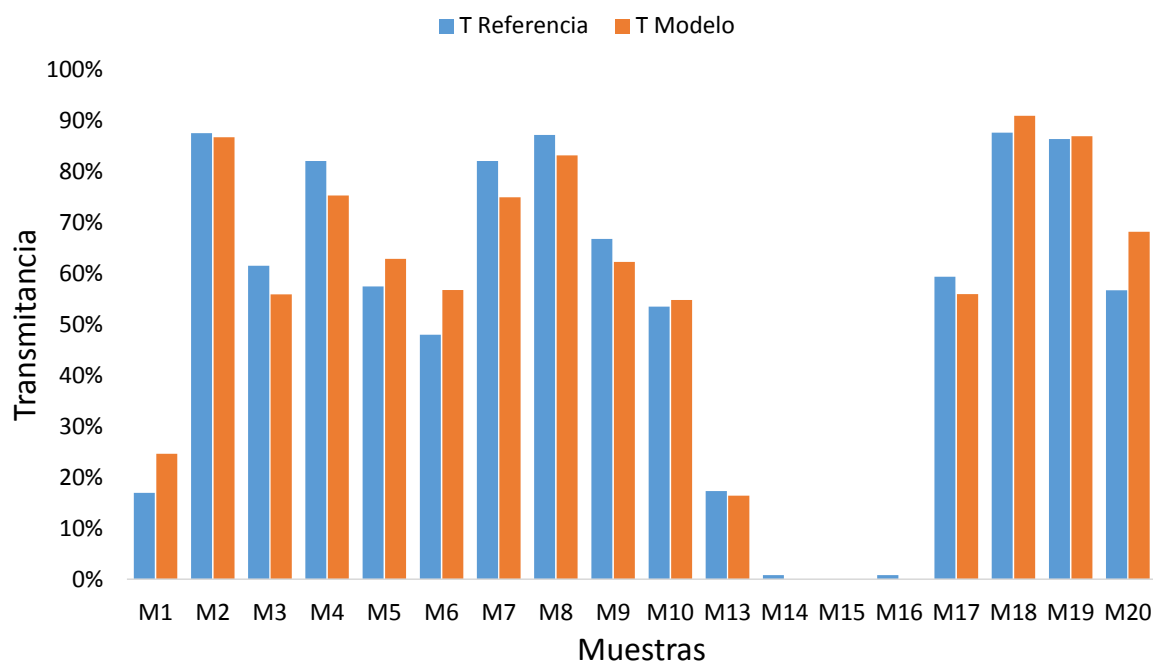
510nm

LED	R ²	Ecuación
510nm	0,9708	$T = 0,9084x - 0,0368$
515nm	0,9708	$T = 0,9084x - 0,0368$

Gráfica resultado del ajuste

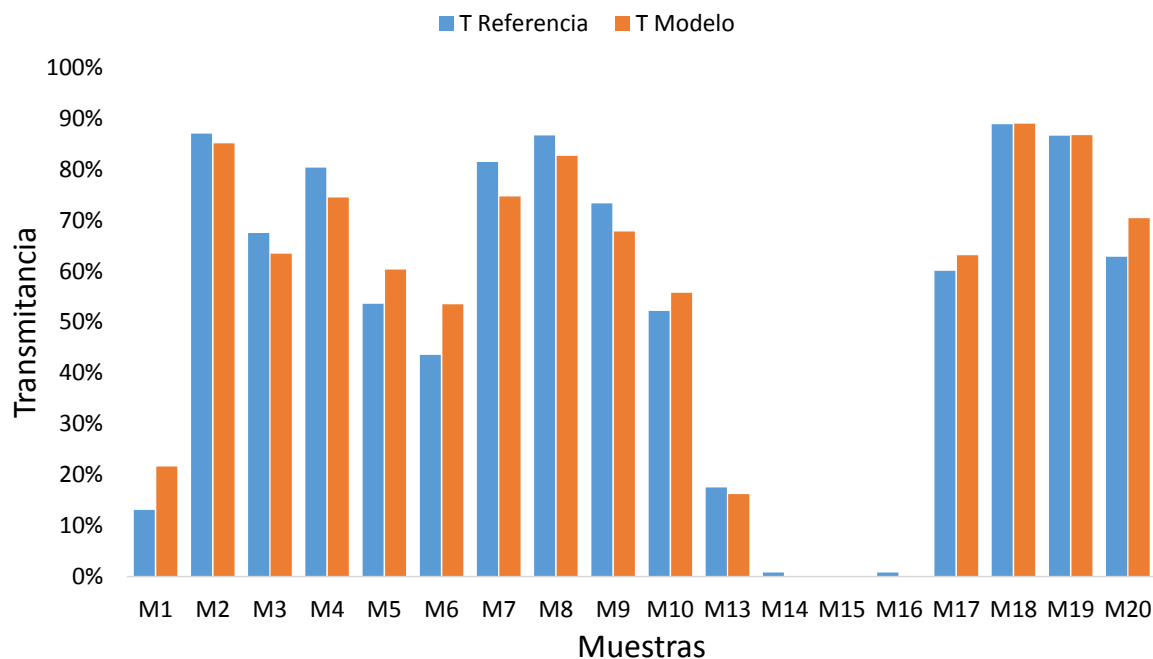
515nm

LED	R ²	Ecuación
515nm	0,9714	$T = 0,9134x - 0,0227$

Gráfica resultado del ajuste

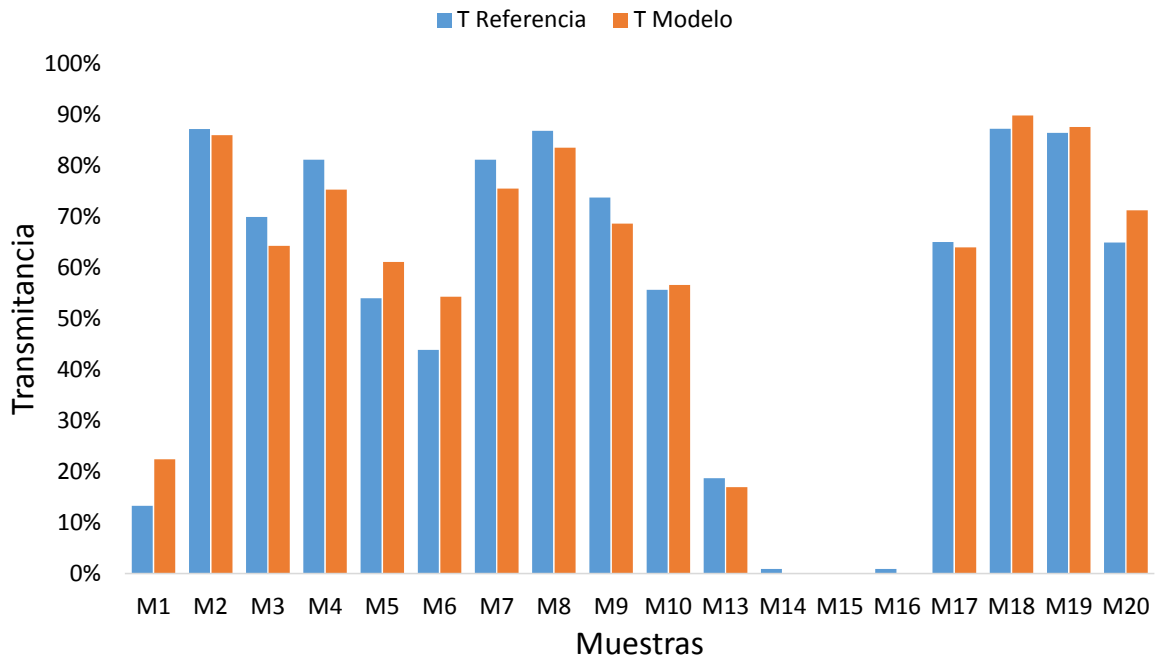
520nm

LED	R ²	Ecuación
521nm	0,9737	$T = 0,9191x - 0,0386$

Gráfica resultado del ajuste

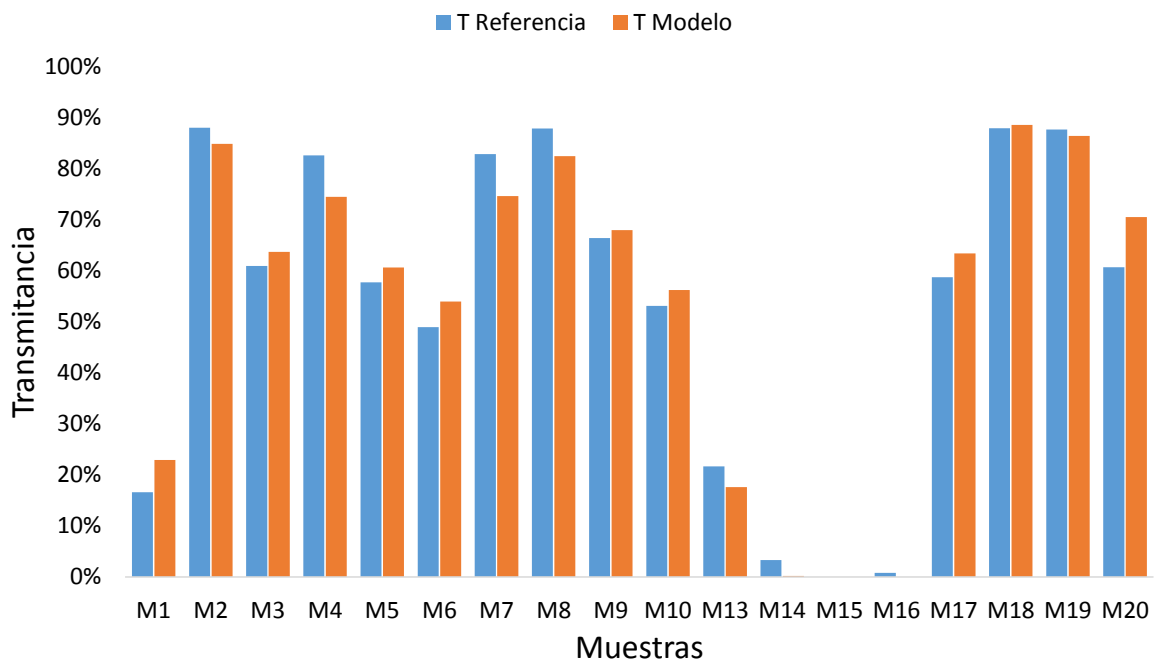
521nm

LED	R ²	Ecuación
521nm	0,9751	T = 0,9213x - 0,0313

Gráfica resultado del ajuste

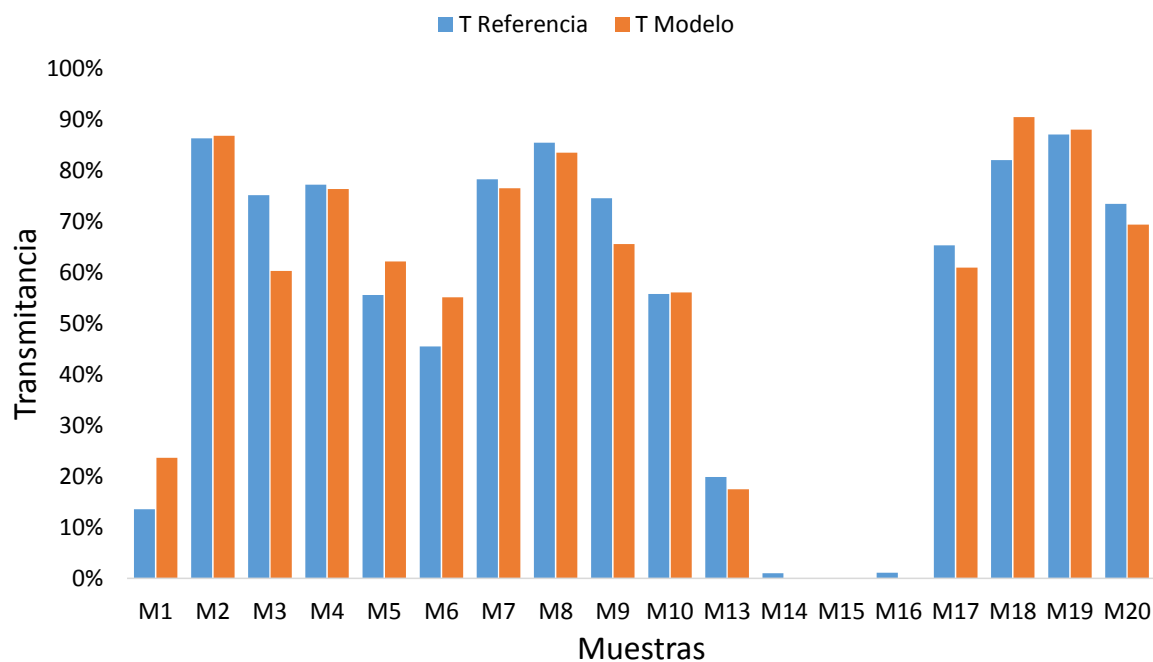
522nm

LED	R ²	Ecuación
521nm	0,9757	T = 0,8969x - 0,0193
522nm	0,9741	T = 0,9076x - 0,0117

Gráfica resultado del ajuste

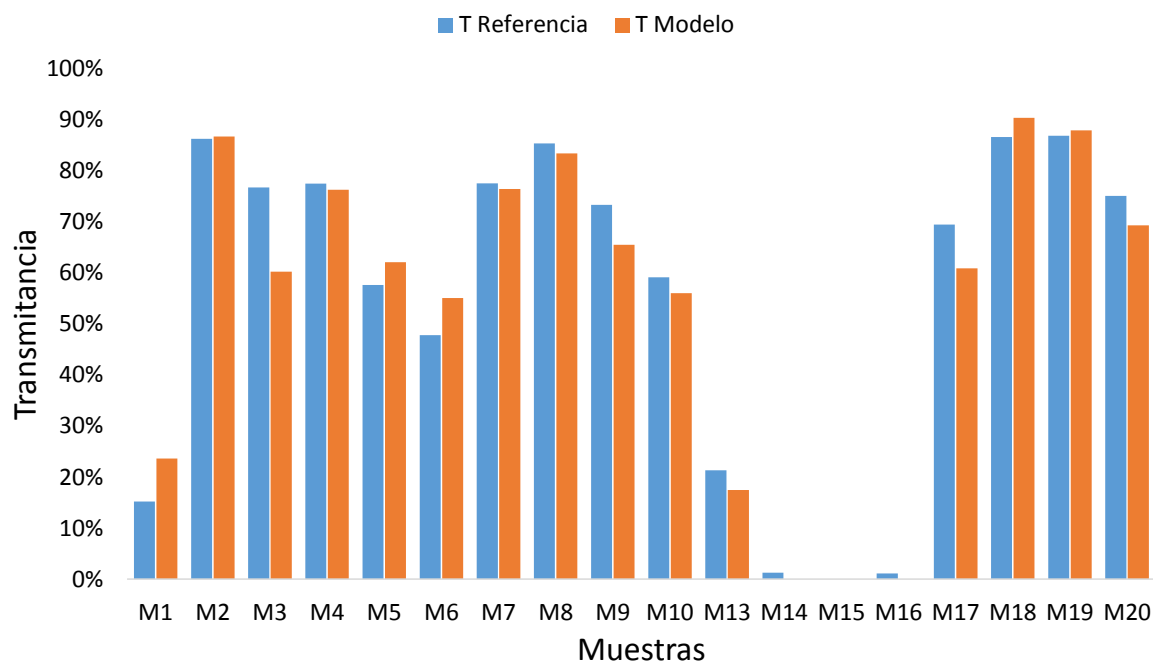
530nm

LED	R ²	Ecuación
532nm	0,9627	T = 0,9061x - 0,0113

Gráfica resultado del ajuste

532nm

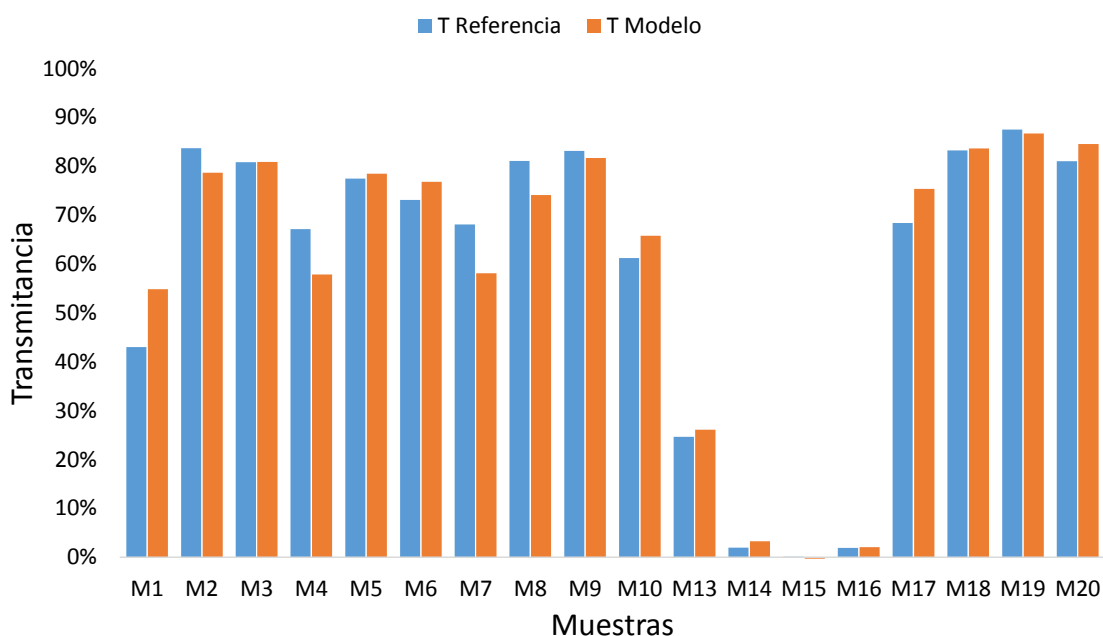
LED	R ²	Ecuación
532nm	0,9671	T = 0,9128x - 0,0031

Gráfica resultado del ajuste

550nm

LED	R ²	Ecuación
557nm	0,9689	$T = 0,8831x - 0,0216$
558nm	0,9543	$T = 0,8824x - 0,0146$

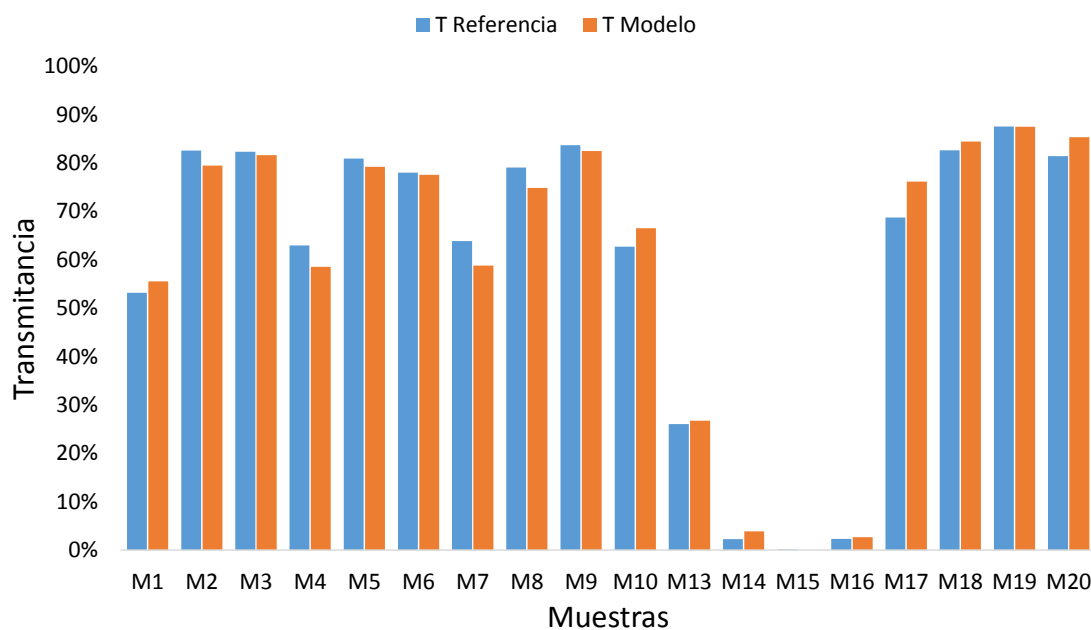
Gráfica resultado del ajuste



555nm

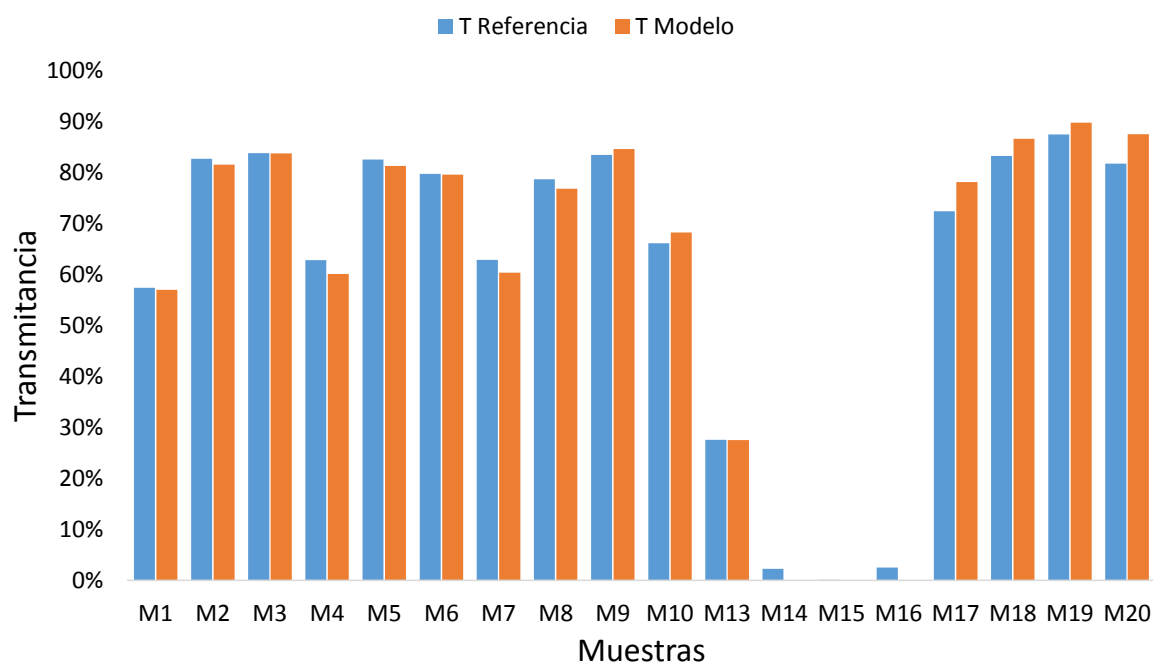
LED	R ²	Ecuación
557nm	0,989	$T = 0,8852x - 0,0158$
558nm	0,983	$T = 0,8885x - 0,0115$

Gráfica resultado del ajuste



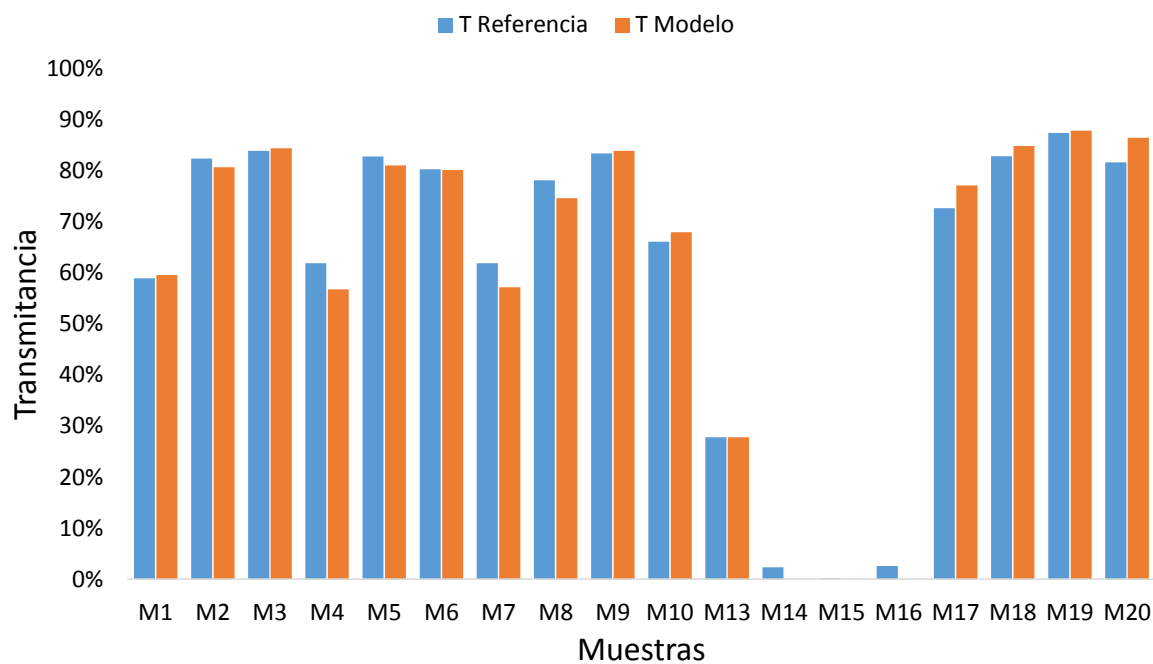
557nm

LED	R ²	Ecuación
557nm	0,9938	T = 0,8904x - 0,0106

Gráfica resultado del ajuste

558nm

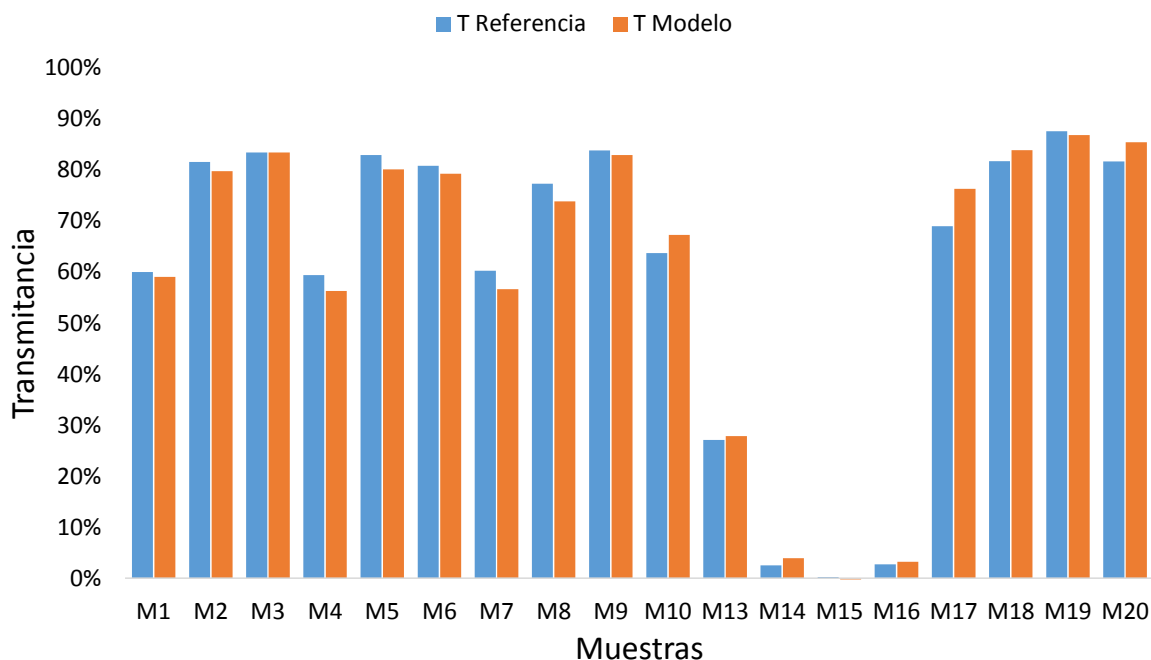
LED	R ²	Ecuación
558nm	0,9926	T = 0,8955x - 0,0069

Gráfica resultado del ajuste

560nm

LED	R ²	Ecuación
557nm	0,9913	$T = 0,8833x - 0,0116$
558nm	0,9913	$T = 0,8895x - 0,0093$

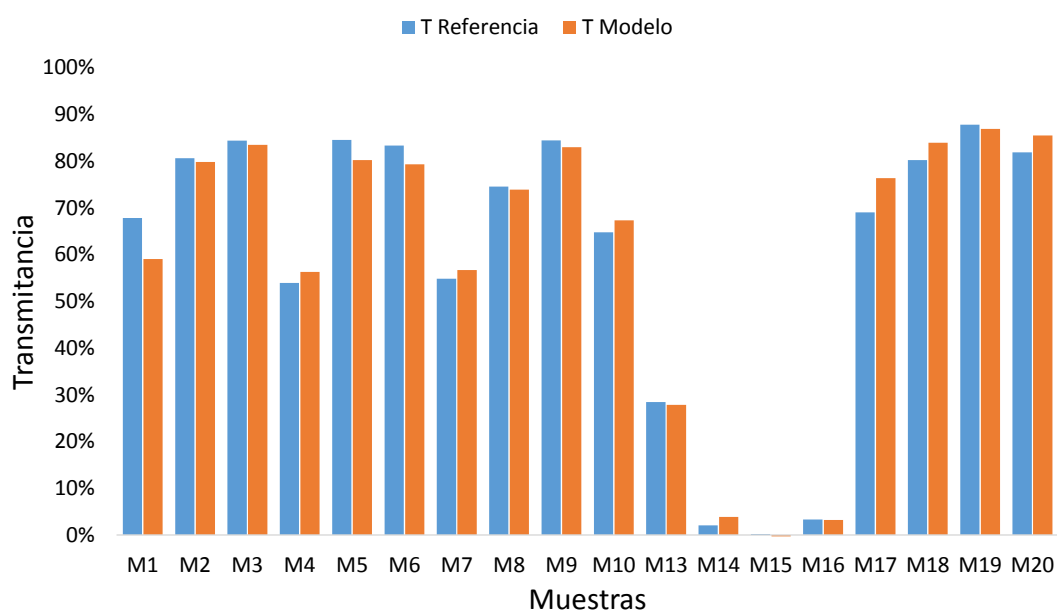
Gráfica resultado del ajuste



565nm

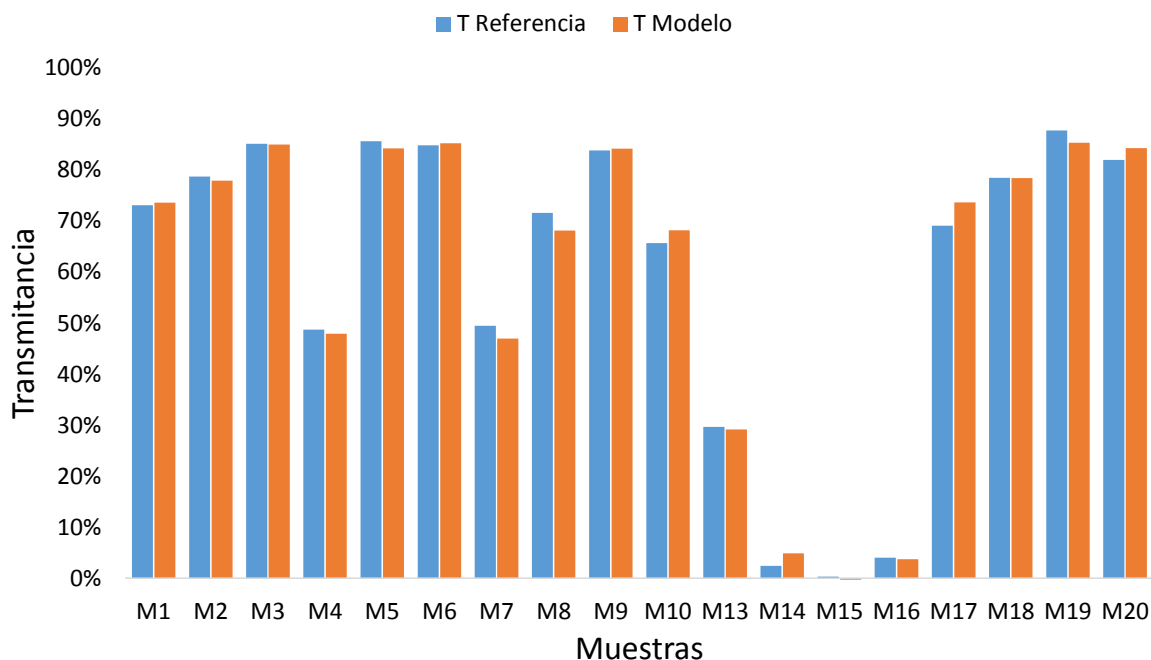
LED	R ²	Ecuación
558nm	0,9865	$T = 0,8907x - 0,0098$
574nm	0,9851	$T = 0,9002x - 0,0156$
575nm	0,9866	$T = 0,9054x - 0,0265$

Gráfica resultado del ajuste



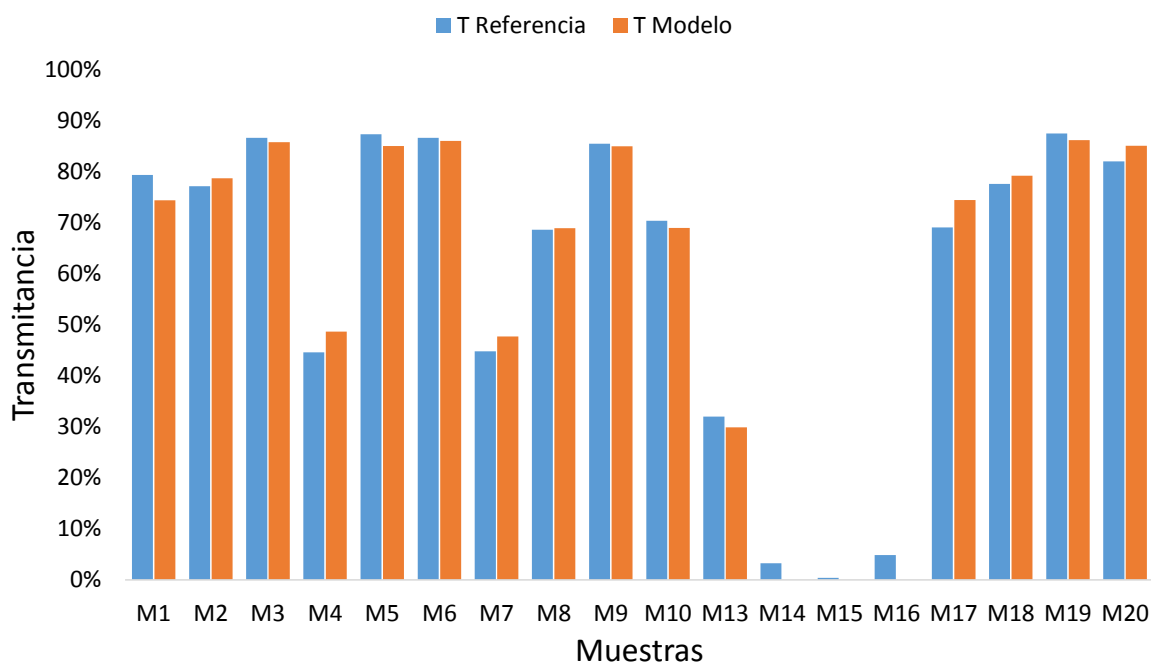
570nm

LED	R ²	Ecuación
574nm	0,9958	$T = 0,9053x - 0,0223$
575nm	0,995	$T = 0,9095x - 0,0323$

Gráfica resultado del ajuste

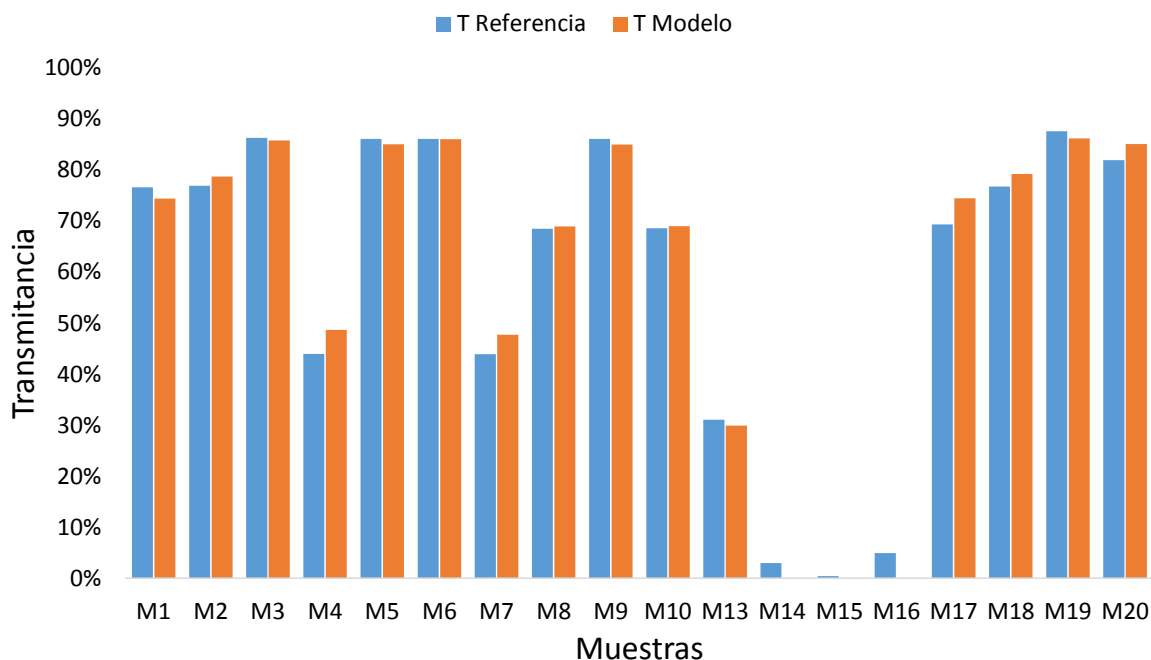
574nm

LED	R ²	Ecuación
574nm	0,9933	$T = 0,914x - 0,0242$

Gráfica resultado del ajuste

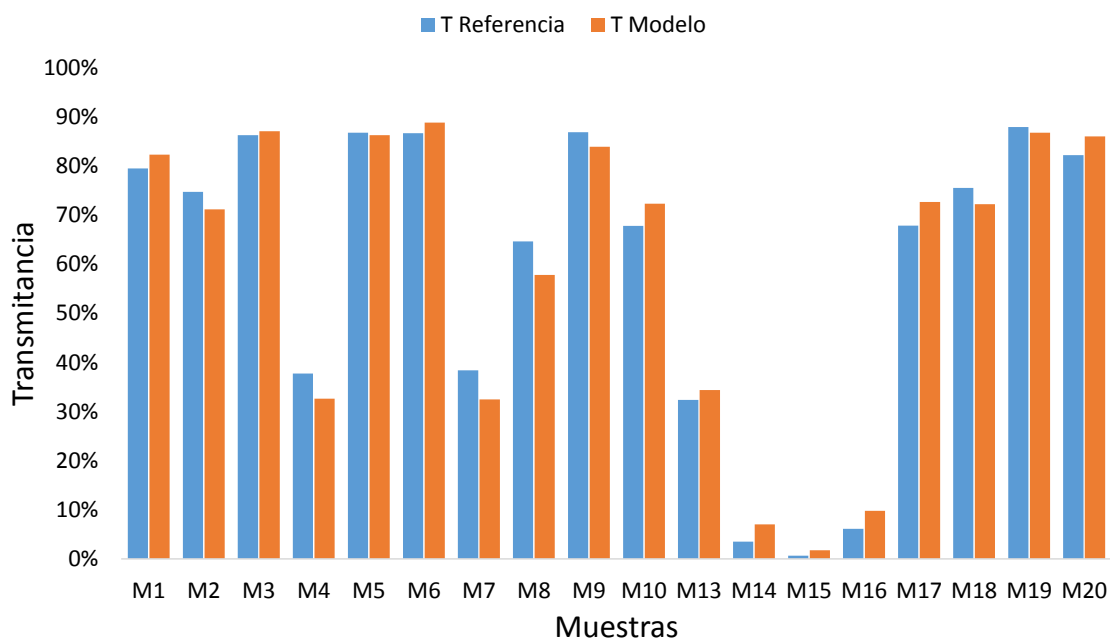
575nm

LED	R ²	Ecuación
574nm	0,9948	$T = 0,9114x - 0,0279$
575nm	0,9919	$T = 0,9145x - 0,0374$

Gráfica resultado del ajuste

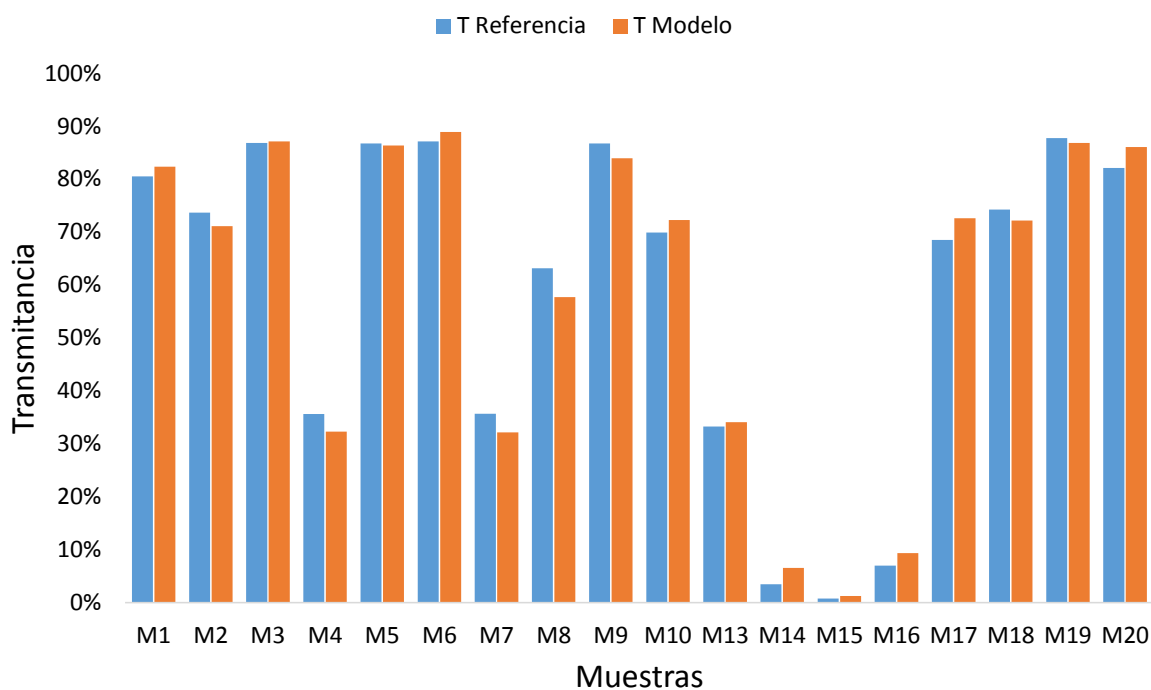
580nm

LED	R ²	Ecuación
583nm	0,9851	$T = 0,9134x - 0,0189$
575nm	0,9757	$T = 0,9118x - 0,0426$

Gráfica resultado del ajuste

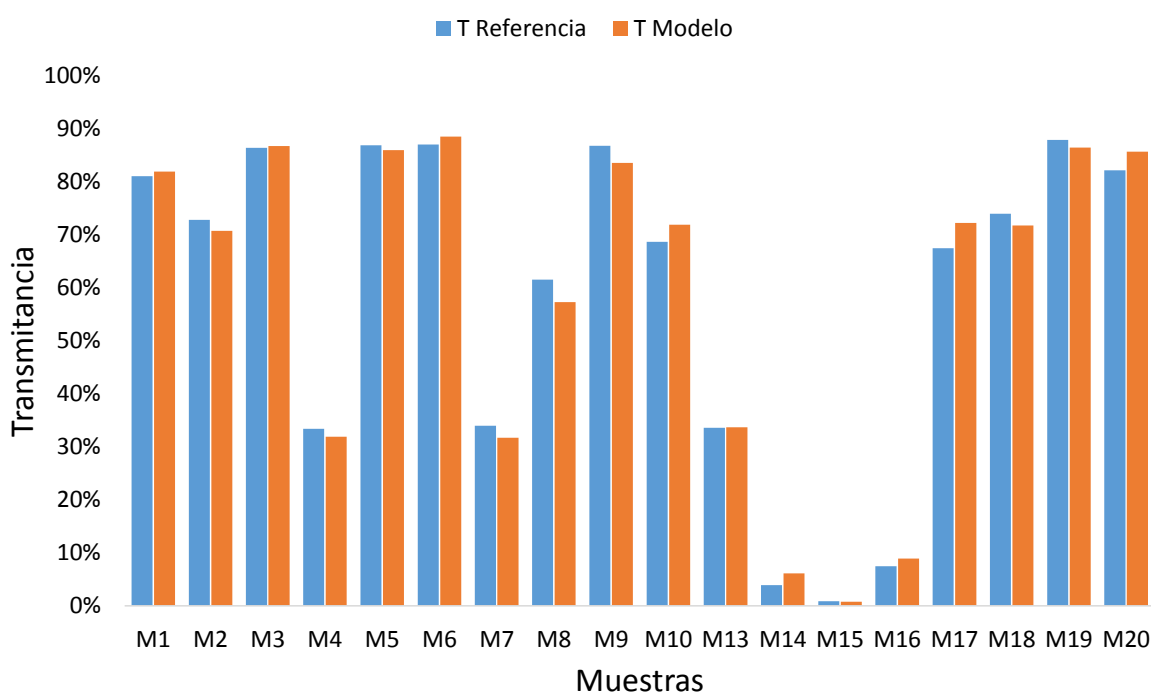
583nm

LED	R ²	Ecuación
583nm	0,9919	$T = 0,9194x - 0,0241$

Gráfica resultado del ajuste

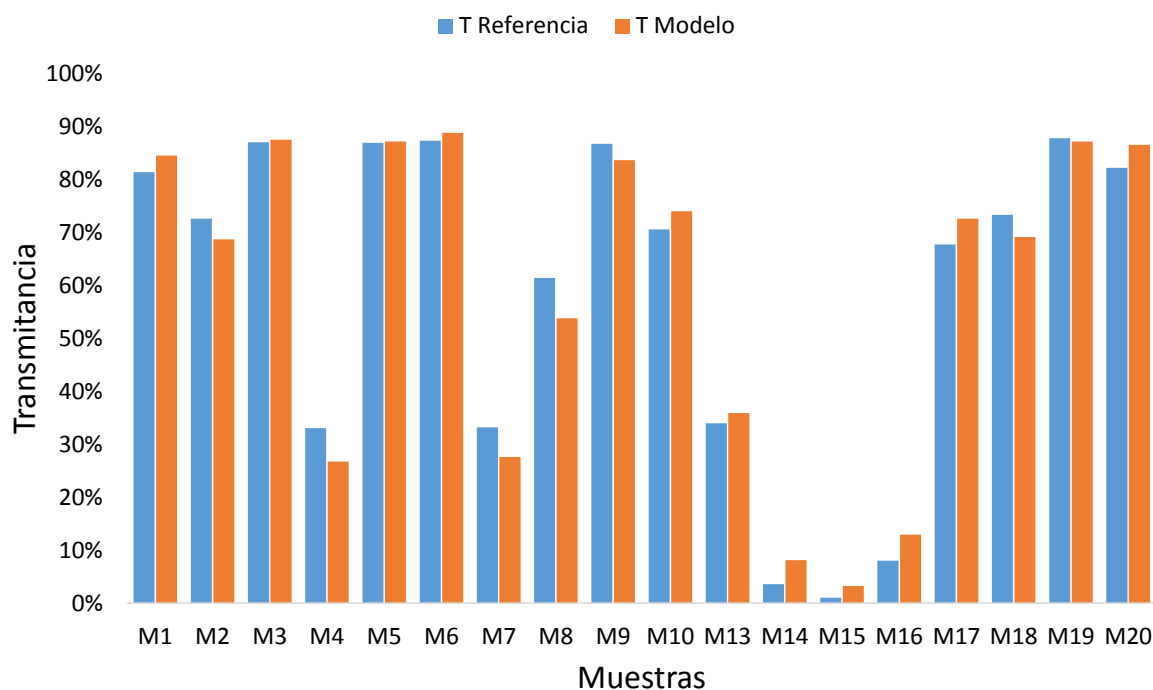
585nm

LED	R ²	Ecuación
583nm	0,9937	$T = 0,9207x - 0,0287$

Gráfica resultado del ajuste

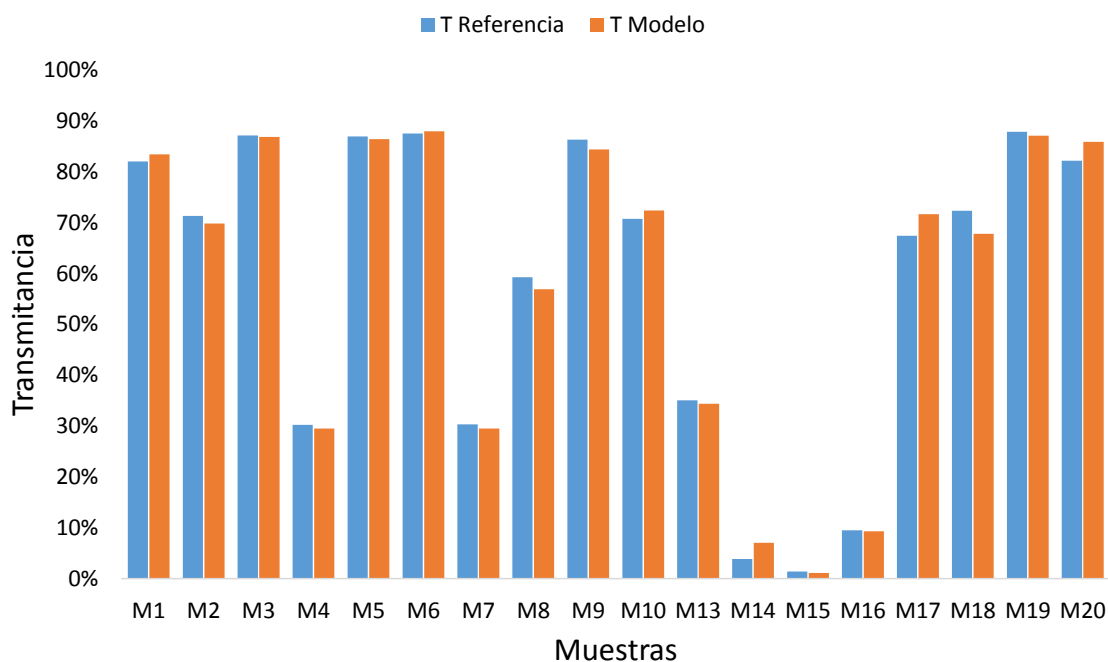
586nm

LED	R ²	Ecuación
586nm	0,9824	$T = 0,889x - 0,0001$

Gráfica resultado del ajuste

590nm

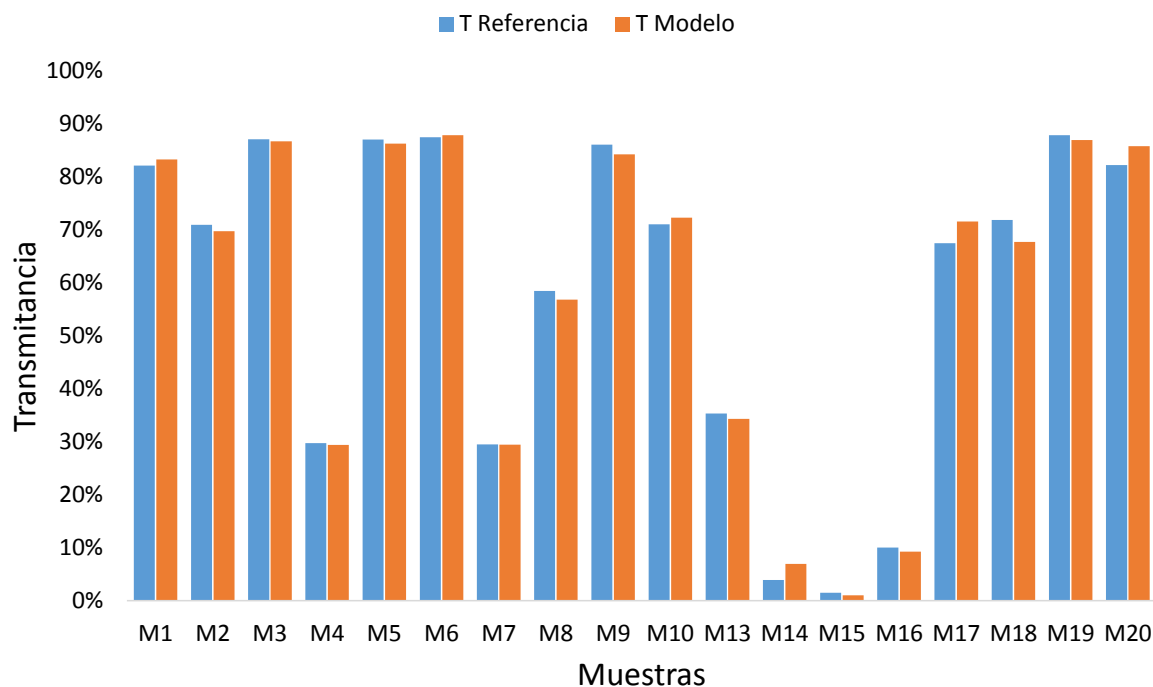
LED	R ²	Ecuación
590nm	0,9951	$T = 0,8793x - 0,0063$
594nm	0,9951	$T = 0,8828x - 0,0053$

Gráfica resultado del ajuste

591nm

LED	R^2	Ecuación
591nm	0,995	$T = 0,8839x - 0,0086$
594nm	0,9958	$T = 0,8828x - 0,006$

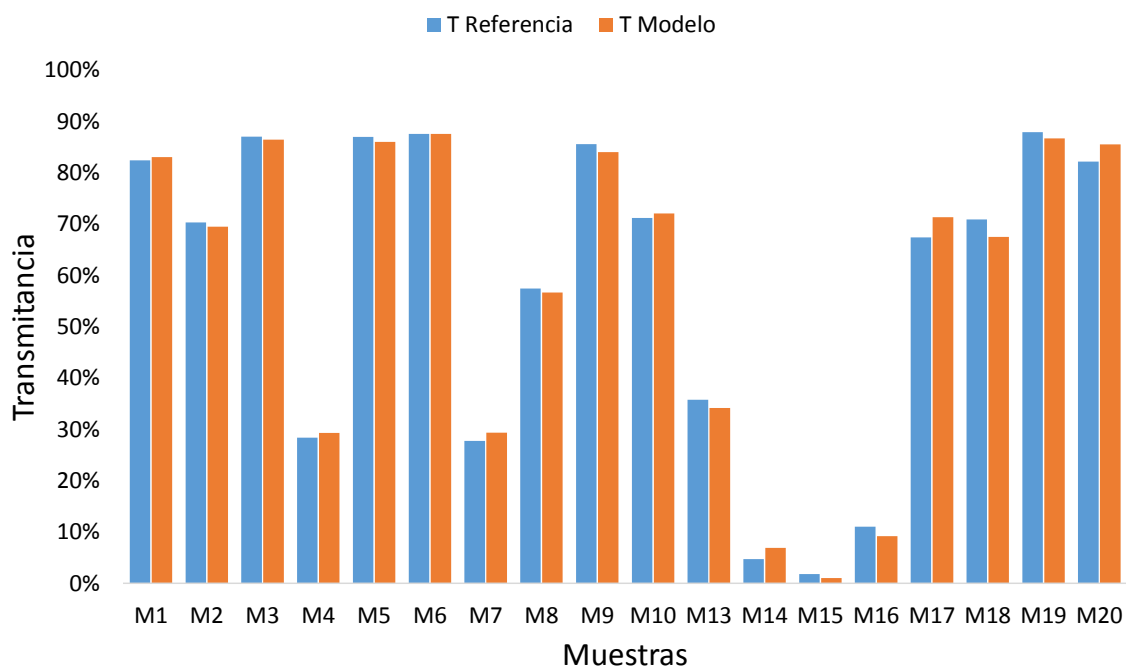
Gráfica resultado del ajuste



594nm

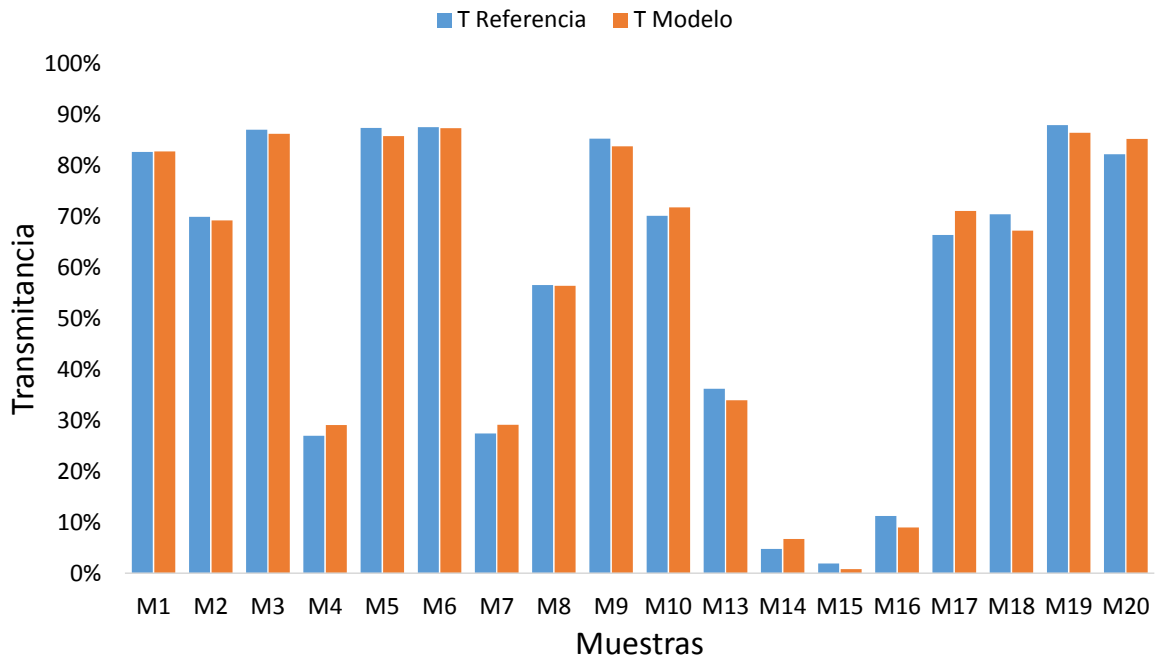
LED	R^2	Ecuación
594nm	0,9963	$T = 0,8805x - 0,0062$

Gráfica resultado del ajuste



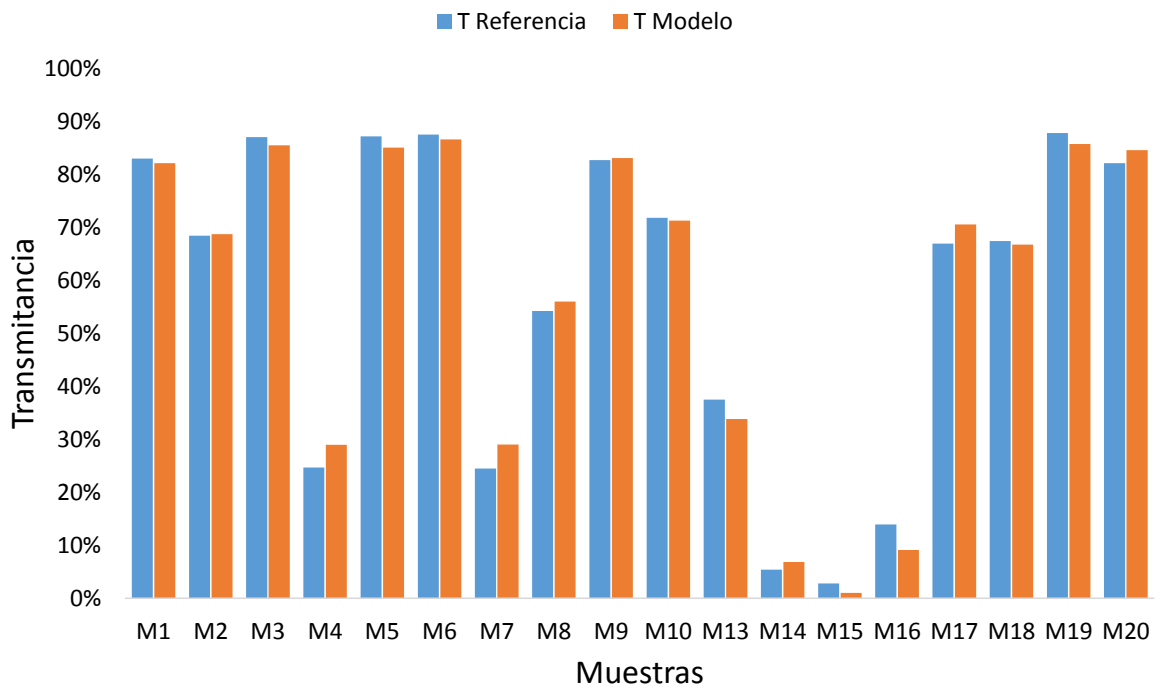
595nm

LED	R ²	Ecuación
594nm	0,9954	T = 0,8806x - 0,0085

Gráfica resultado del ajuste

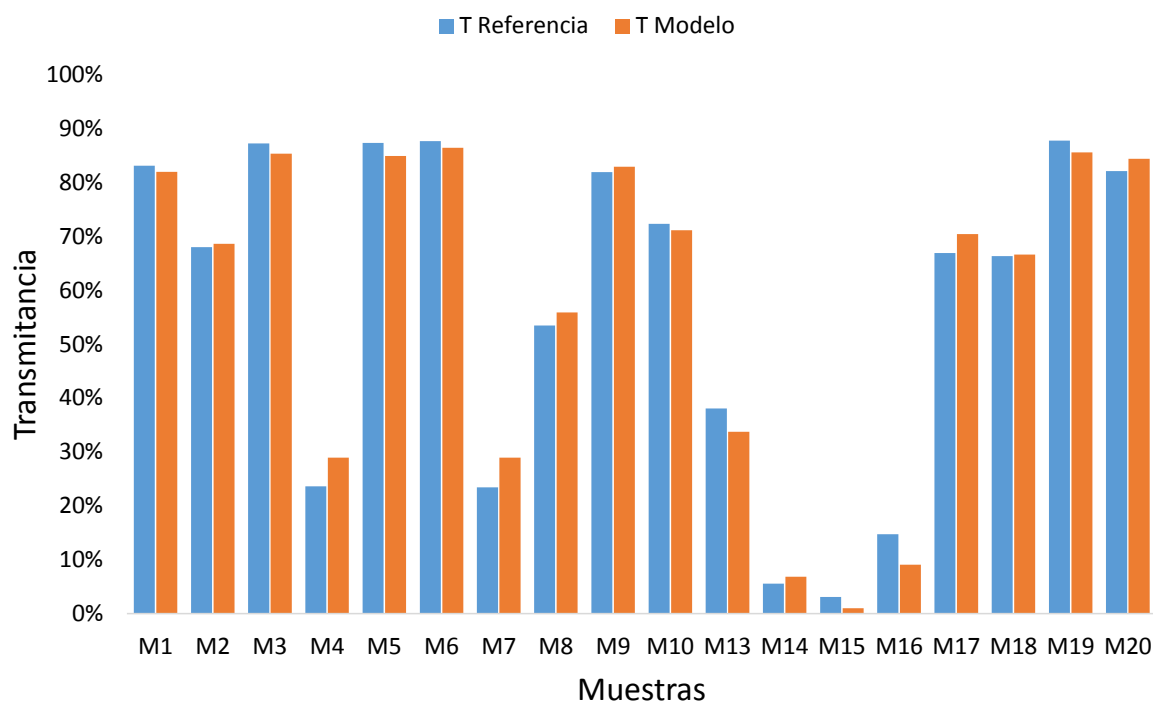
560nm

LED	R ²	Ecuación
594nm	0,9928	T = 0,8705x - 0,006

Gráfica resultado del ajuste

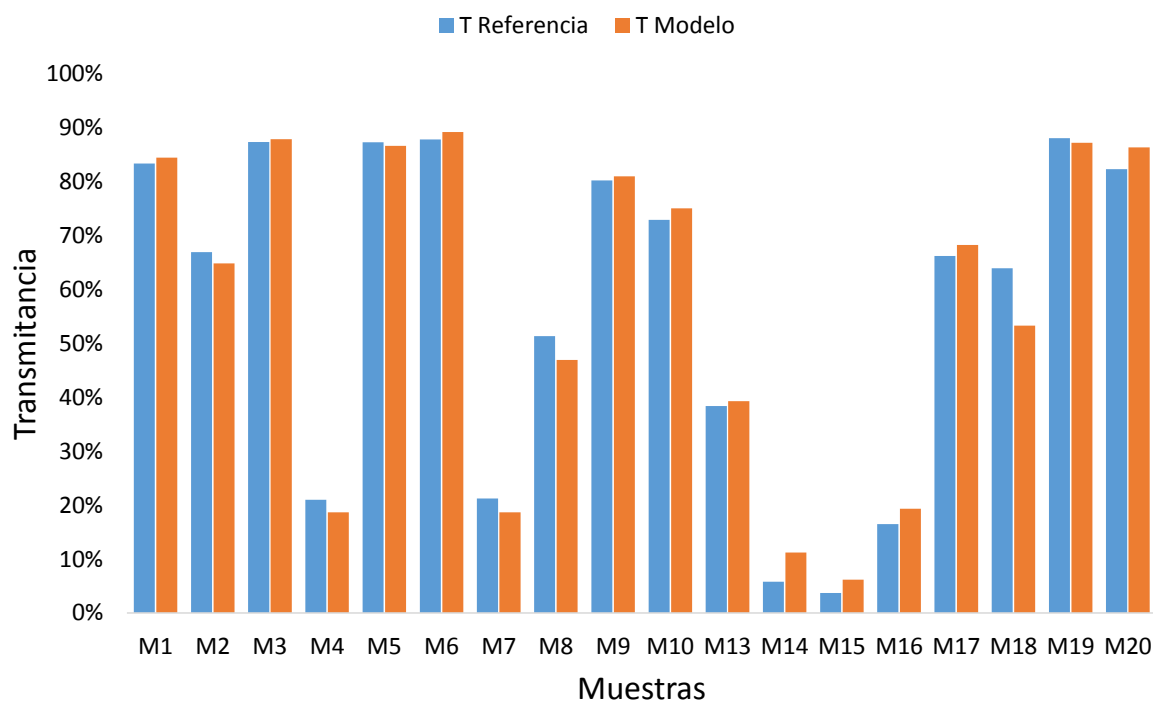
562nm

LED	R ²	Ecuación
594nm	0,9902	$T = 0,8687x - 0,0062$

Gráfica resultado del ajuste

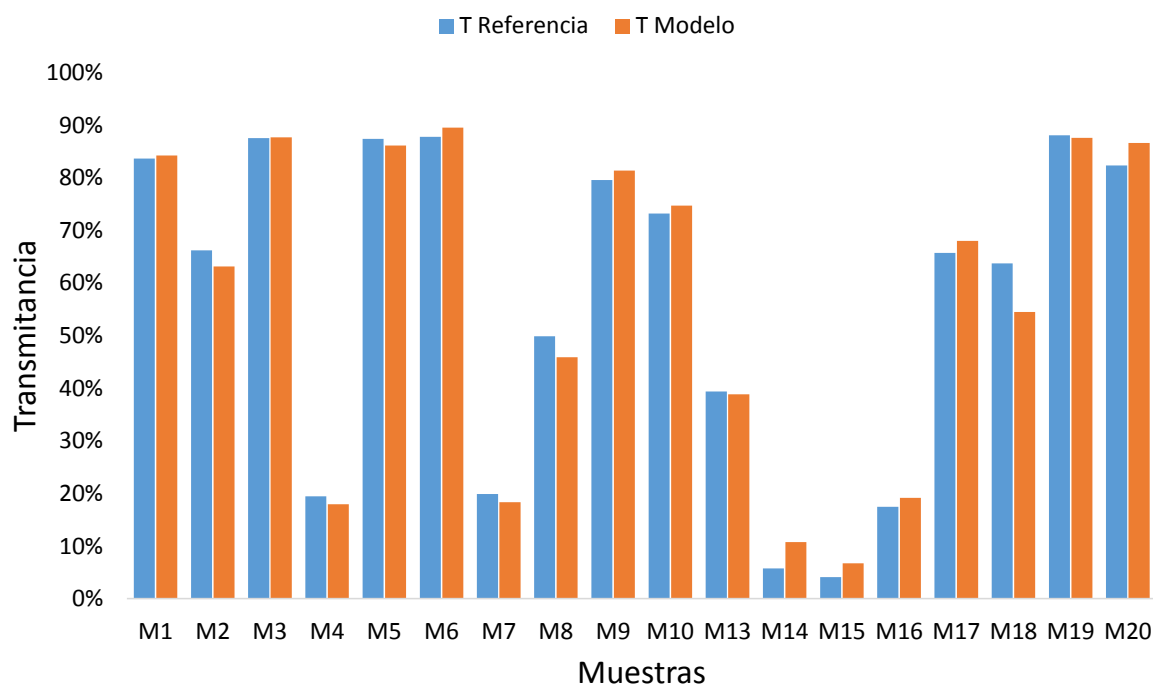
605nm

LED	R ²	Ecuación
605nm	0,9868	$T = 0,9086x - 0,0133$

Gráfica resultado del ajuste

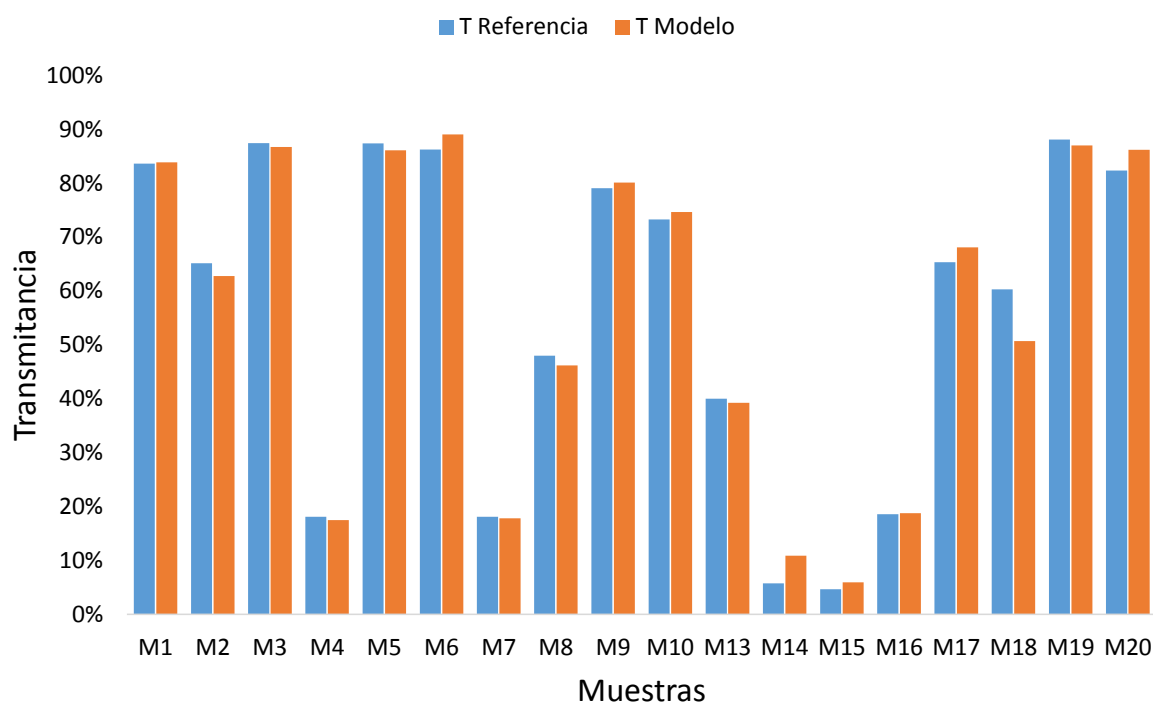
607nm

LED	R ²	Ecuación
607nm	0,9886	T = 0,9106x - 0,0171

Gráfica resultado del ajuste

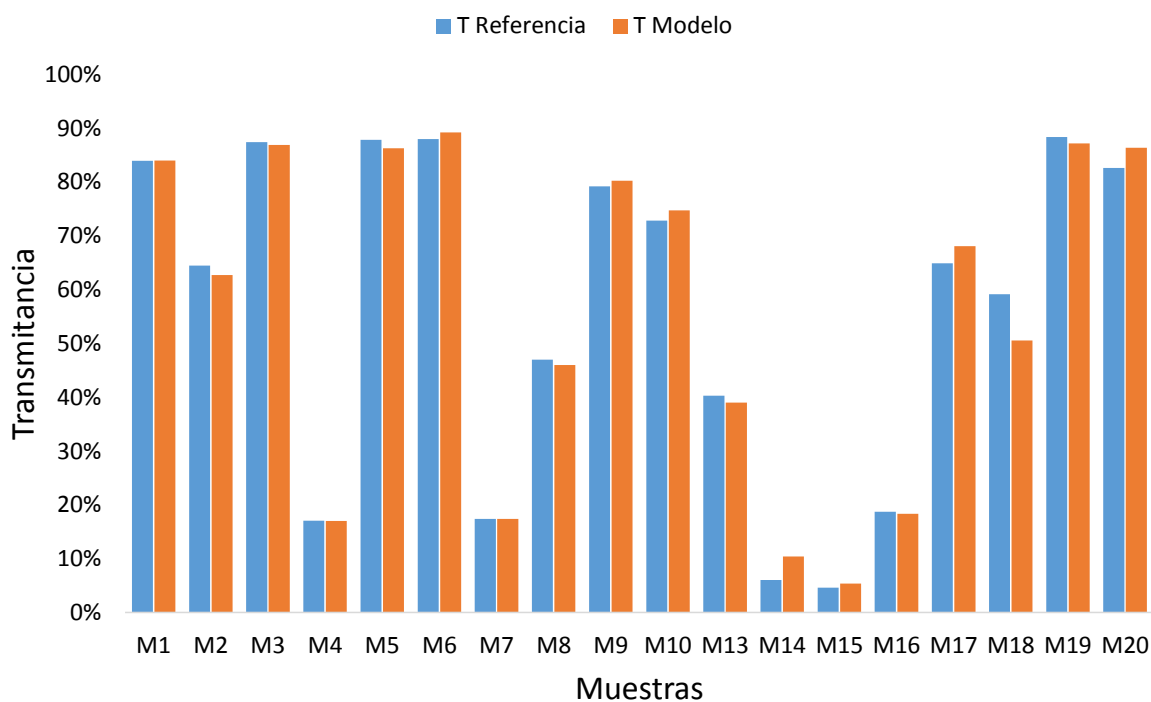
609nm

LED	R ²	Ecuación
609nm	0,9897	T = 0,876x + 0,0015

Gráfica resultado del ajuste

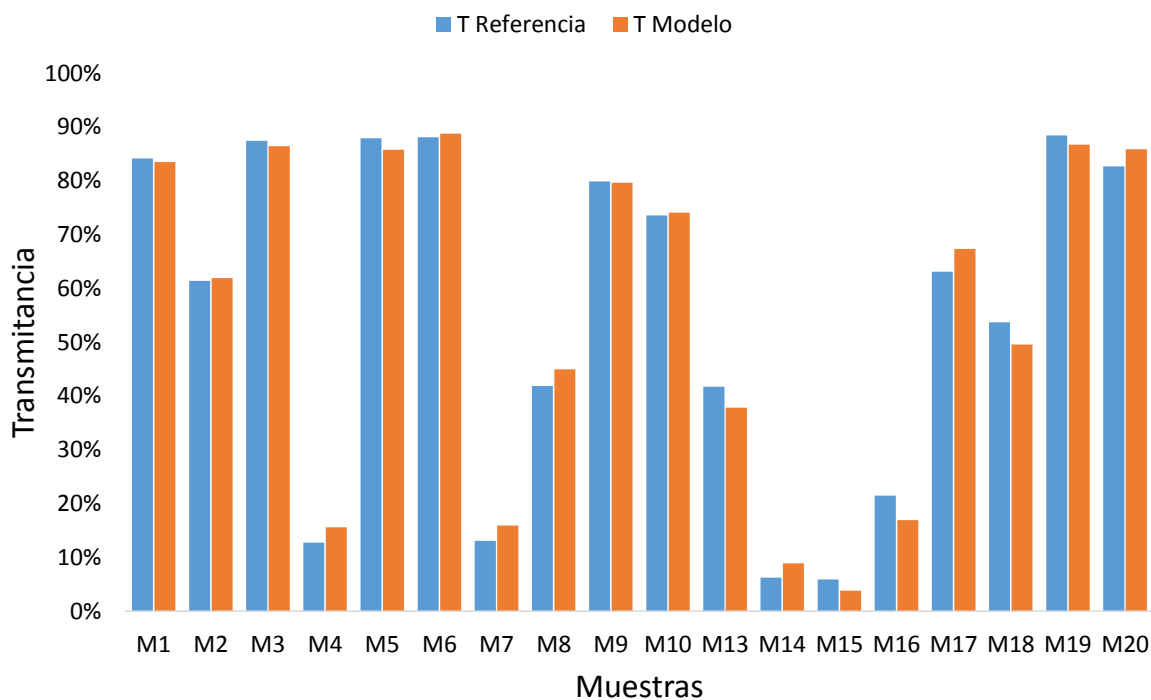
610nm

LED	R ²	Ecuación
609nm	0,9919	T = 0,8841x - 0,0046

Gráfica resultado del ajuste

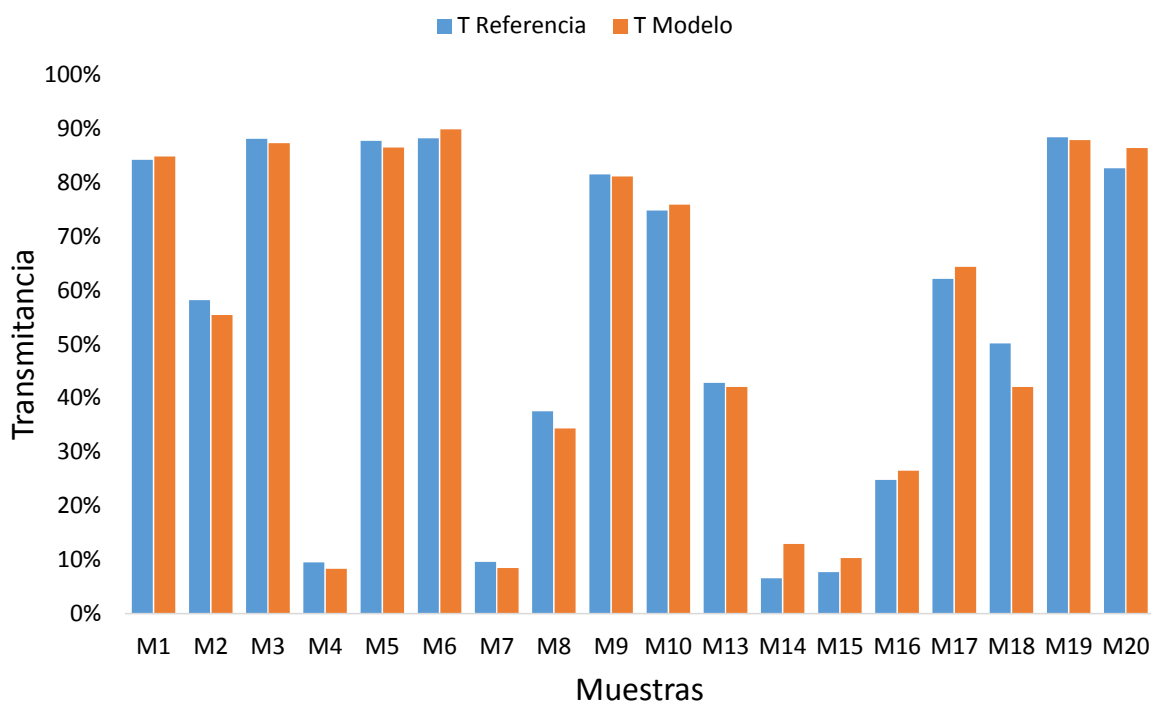
615nm

LED	R ²	Ecuación
609nm	0,9924	T = 0,8958x - 0,0213

Gráfica resultado del ajuste

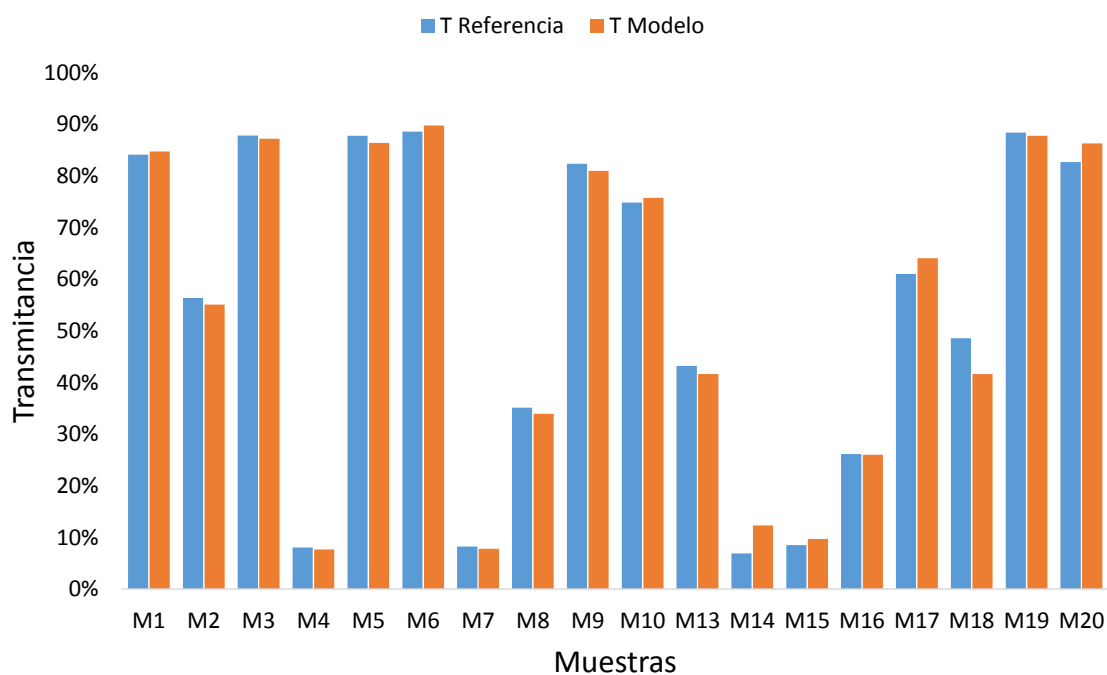
620nm

LED	R ²	Ecuación
622nm	0,9906	$T = 0,896x - 0,0098$

Gráfica resultado del ajuste

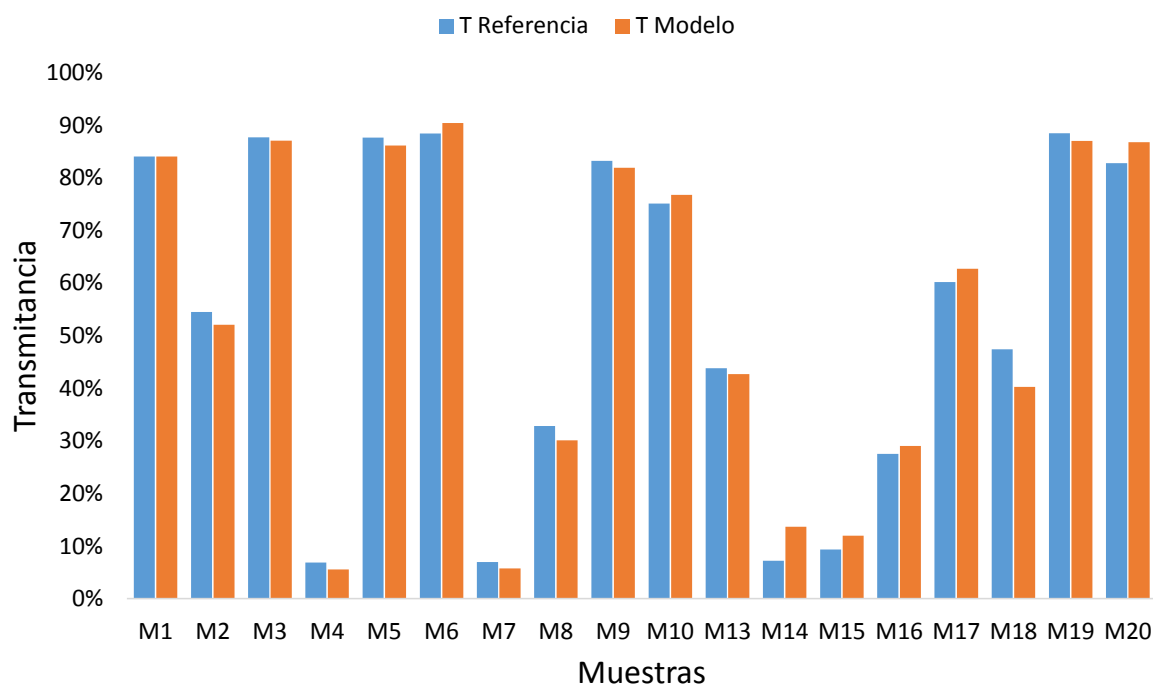
622nm

LED	R ²	Ecuación
622nm	0,9935	$T = 0,9025x - 0,0171$

Gráfica resultado del ajuste

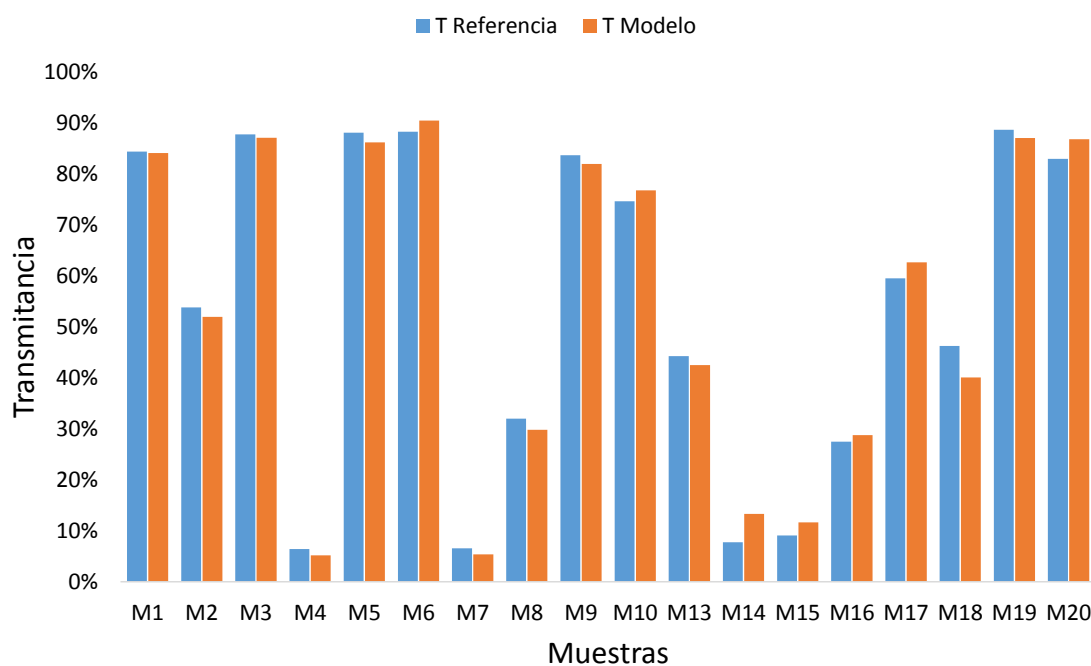
624nm

LED	R ²	Ecuación
624nm	0,9912	$T = 0,8973x - 0,0114$

Gráfica resultado del ajuste

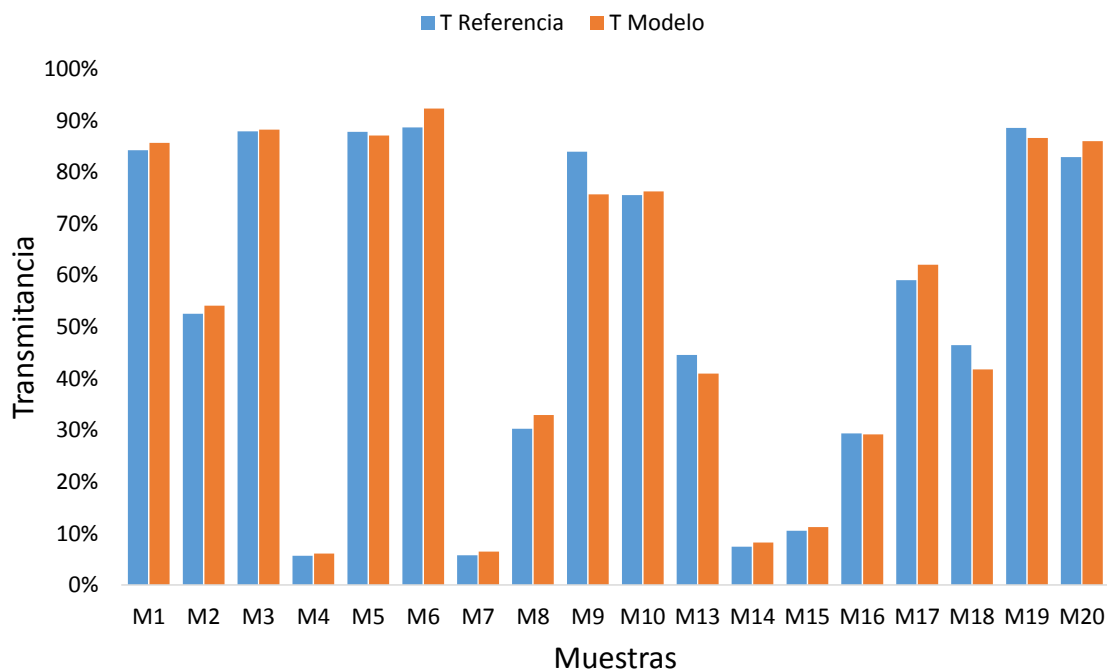
625nm

LED	R ²	Ecuación
624nm	0,9924	$T = 0,9013x - 0,0154$
630nm	0,9942	$T = 0,9271x - 0,0394$

Gráfica resultado del ajuste

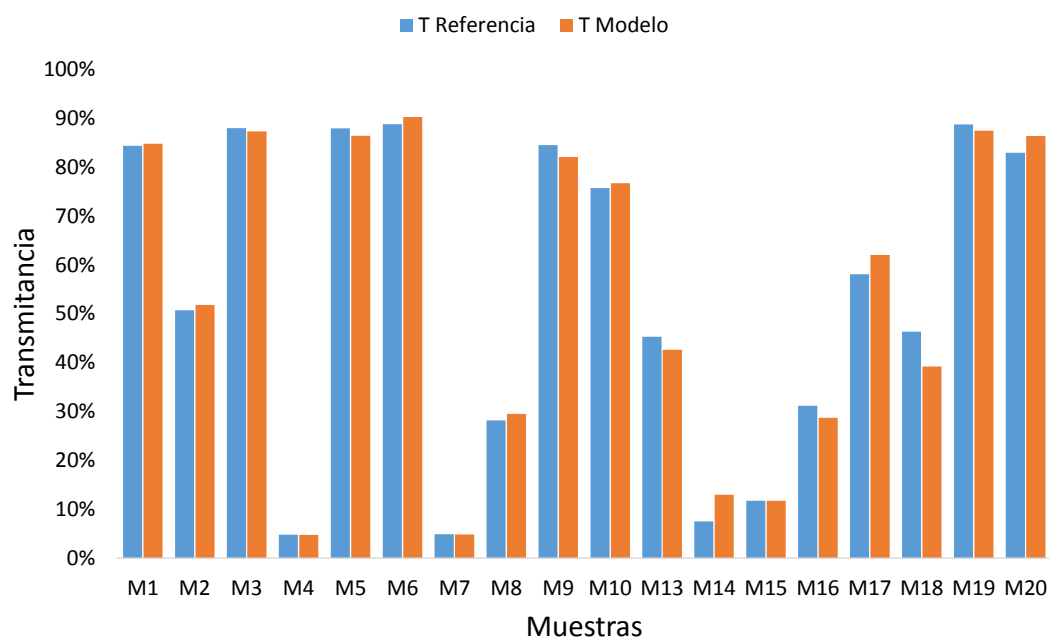
627nm

LED	R ²	Ecuación
624nm	0,9914	$T = 0,9797x - 0,0748$
630nm	0,9937	$T = 0,9295x - 0,041$

Gráfica resultado del ajuste

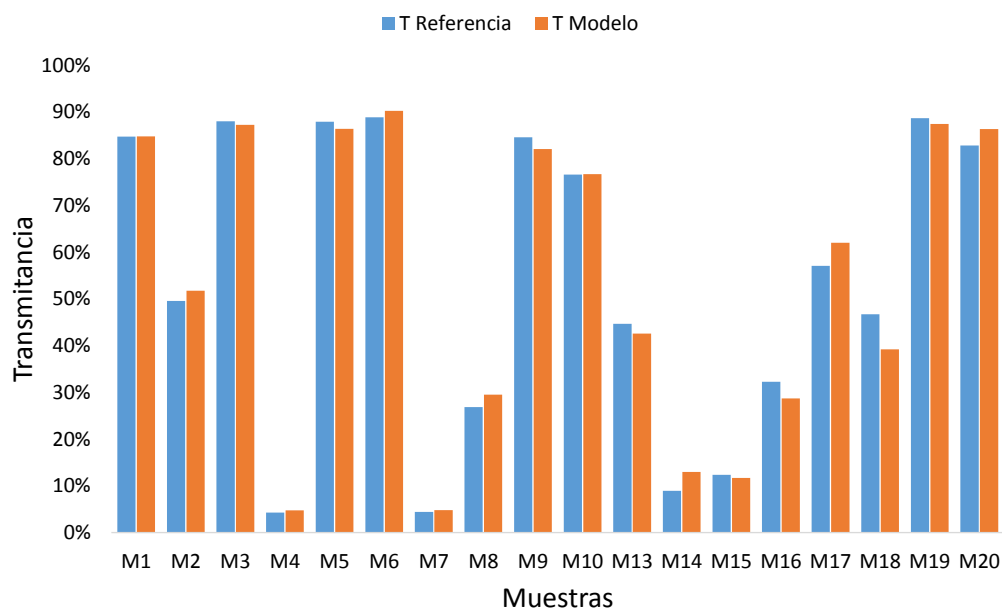
631nm

LED	R ²	Ecuación
631nm	0,9923	$T = 0,9081x - 0,0195$
630nm	0,9913	$T = 0,9317x - 0,0432$

Gráfica resultado del ajuste

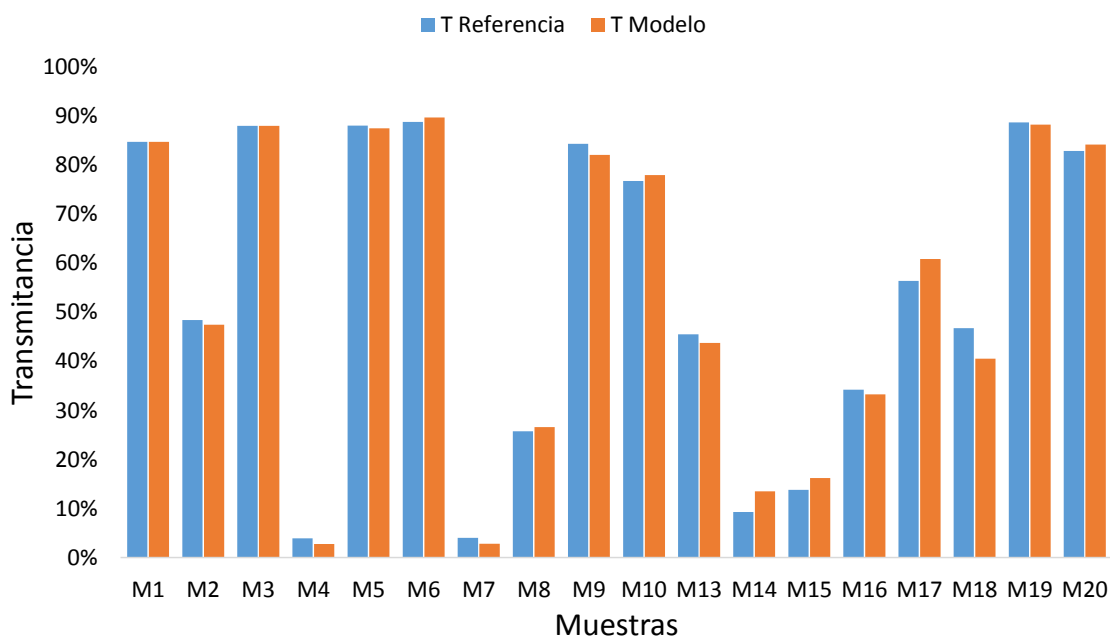
632nm

LED	R ²	Ecuación
630nm	0,9895	$T = 0,9312x - 0,0431$
631nm	0,9916	$T = 0,9081x - 0,0196$
632nm	0,9886	$T = 0,9211x - 0,0356$

Gráfica resultado del ajuste

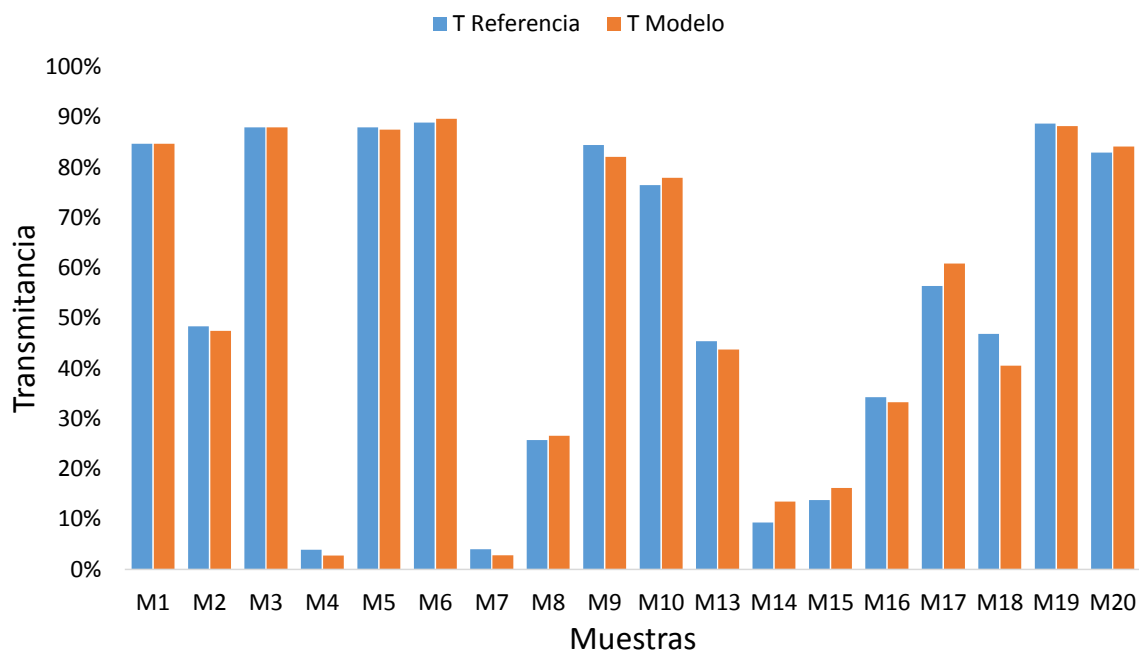
635nm

LED	R ²	Ecuación
630nm	0,9858	$T = 0,9272x - 0,0399$
635nm	0,988	$T = 0,9245x - 0,0395$
636nm	0,9944	$T = 0,946x - 0,0551$

Gráfica resultado del ajuste

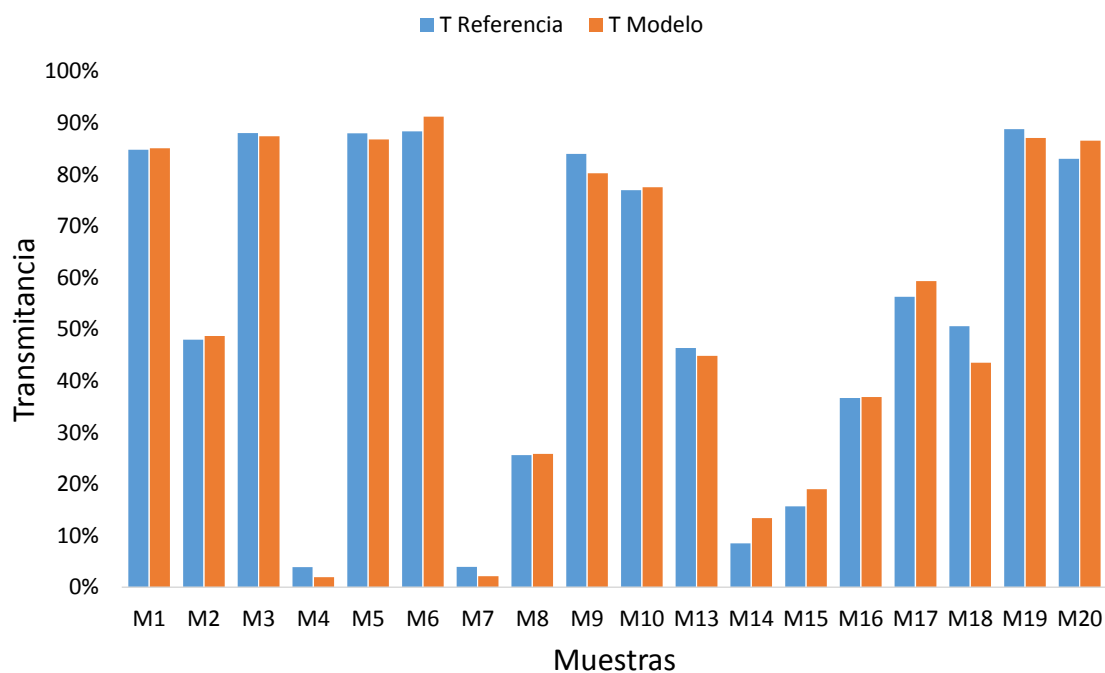
636nm

LED	R ²	Ecuación
630nm	0,9858	$T = 0,9268x - 0,0398$
636nm	0,9944	$T = 0,9455x - 0,0549$

Gráfica resultado del ajuste

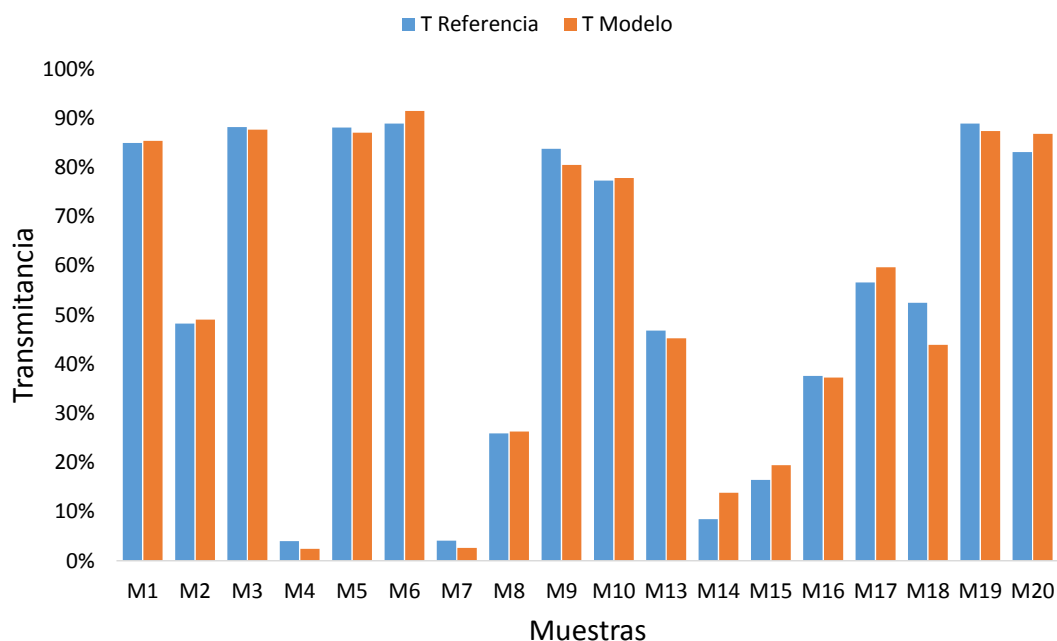
639nm

LED	R ²	Ecuación
630nm	0,9779	$T = 0,9176x - 0,0298$
639nm	0,9919	$T = 0,9546x - 0,0637$

Gráfica resultado del ajuste

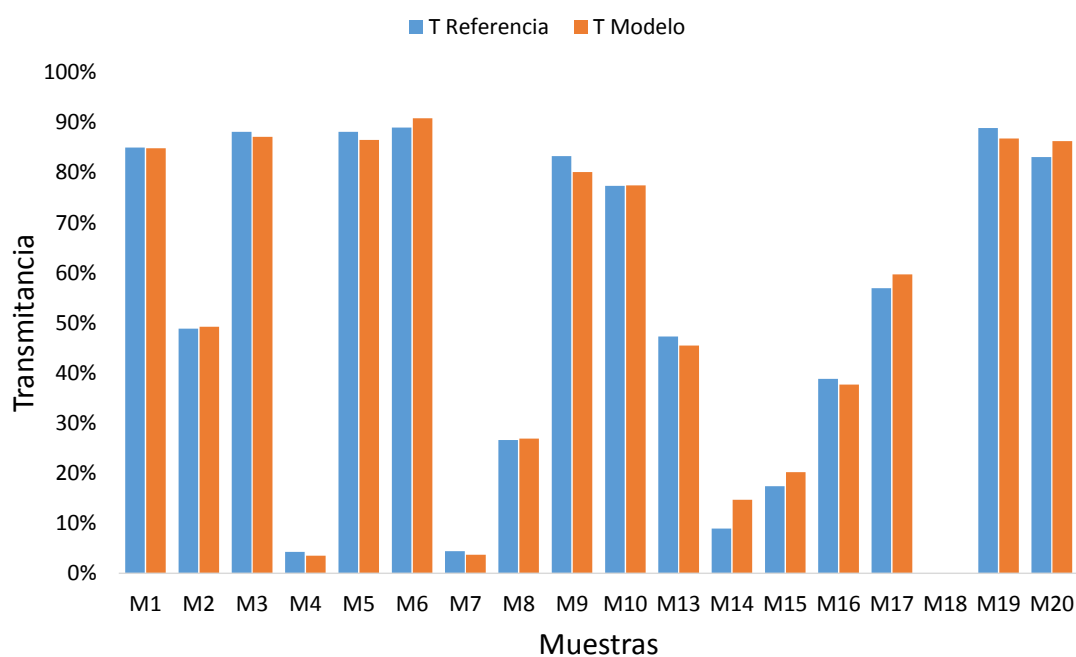
640nm

LED	R ²	Ecuación
630nm	0,9749	$T = 0,9142x - 0,0243$
639nm	0,9907	$T = 0,952x - 0,0587$
640nm	0,9805	$T = 0,8889x + 0,0008$

Gráfica resultado del ajuste

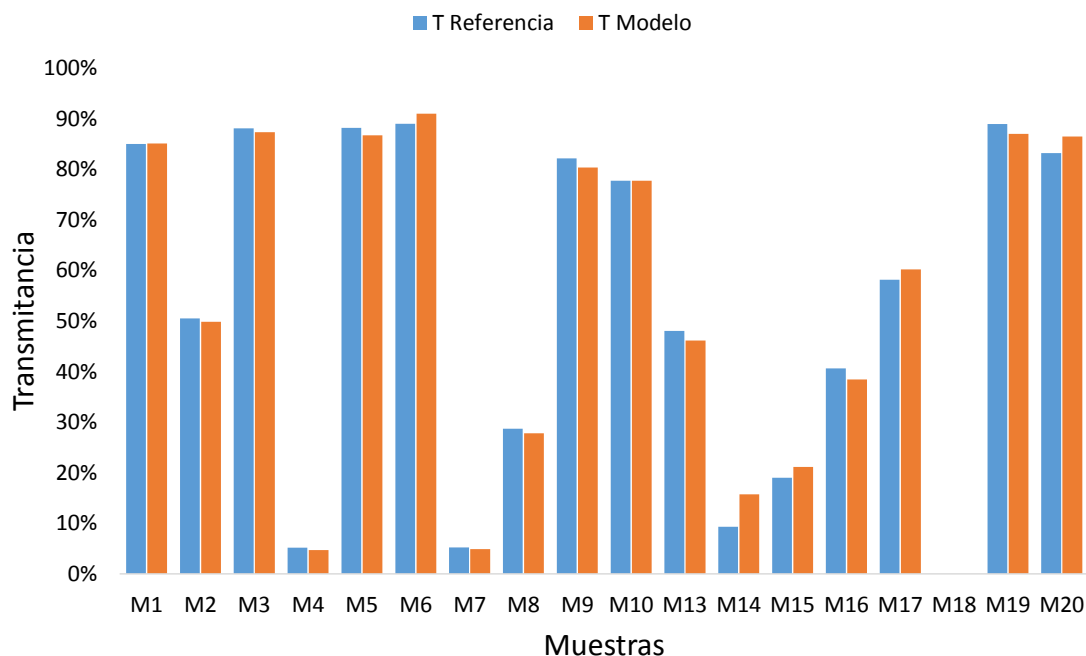
642nm

LED	R ²	Ecuación
639nm	0,9947	$T = 0,934x - 0,0464$
640nm	0,9905	$T = 0,8903x - 0,0042$

Gráfica resultado del ajuste

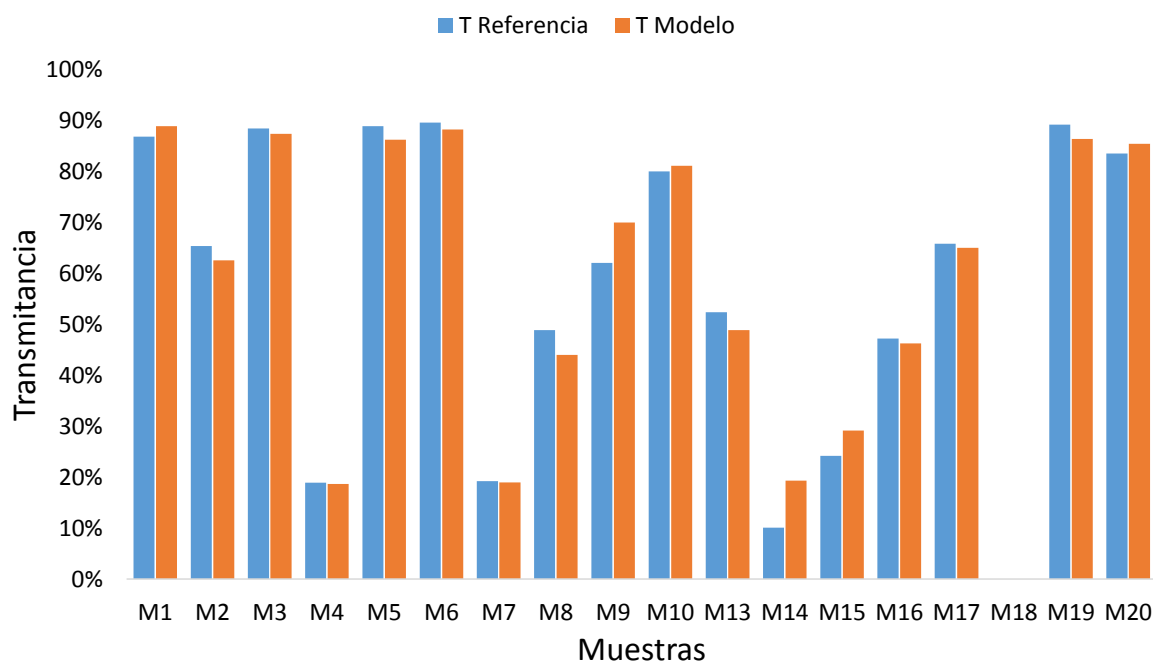
645nm

LED	R ²	Ecuación
639nm	0,9951	$T = 0,9226x - 0,0335$
640nm	0,9898	$T = 0,8788x + 0,0085$
645nm	0,9555	$T = 1,0346x - 0,1439$

Gráfica resultado del ajuste

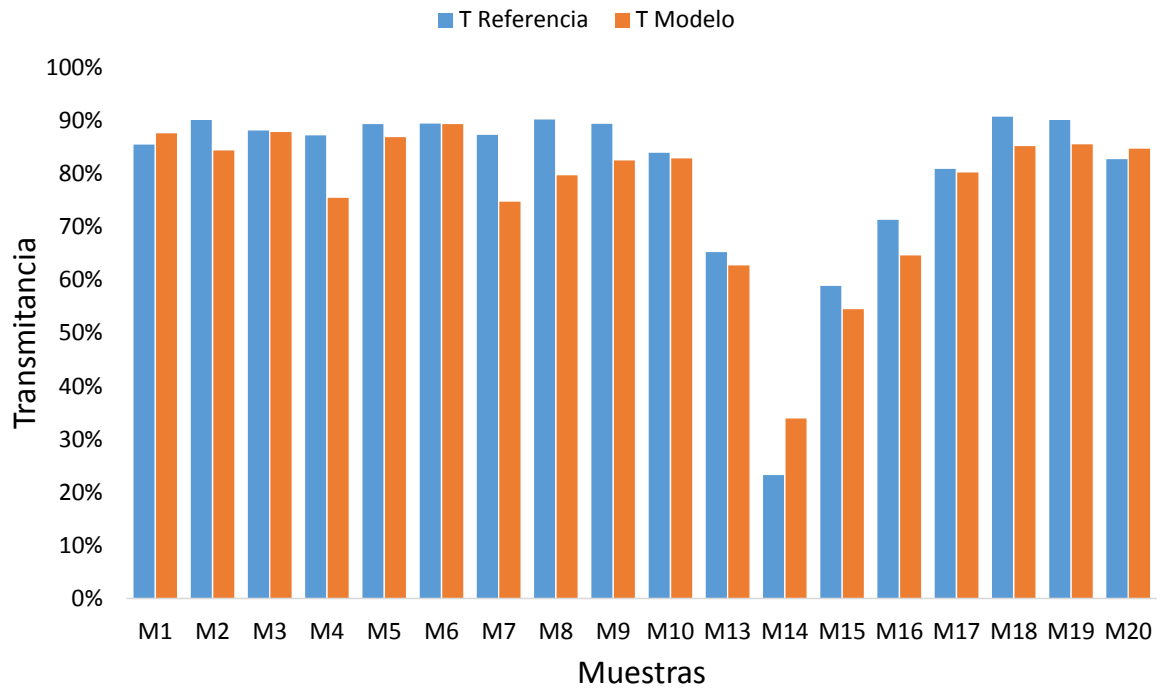
660nm

LED	R ²	Ecuación
660nm	0,9824	$T = 0,9749x - 0,0592$

Gráfica resultado del ajuste

700nm

LED	R ²	Ecuación
700nm	0,9788	$T = 0,9402x - 0,0191$

Gráfica resultado del ajuste

Bibliografía

1. Puertas, J.; Suárez, J.; Anta, J. Gestión de las Aguas Pluviales. Implicaciones en el Diseño de los Sistemas de Saneamiento y Drenaje Urbano. In Monografía M98; CEDEX: Madrid, Spain, 2008.
2. Ward, S.; Butler, D. Compliance with the Urban Waste Water Treatment Directive: European Union City Responses in Relation to Combined Sewer Overflow Discharges; Thames Water: London, UK, 2009.
3. Naves, J.; Anta, J.; Suárez, J.; Puertas, J. Hydraulic, wash-off and sediment transport experiments in a full-scale urban drainage physical model. *Sci. Data* 2020, 7, 1–13, doi:10.1038/s41597-020-0384-z.
4. Anta, J.; Pena, E.; Suarez, J.; Cagiao, J. A BMP selection process based on the granulometry of runoff solids in a separate urban catchment. *Water SA* 2007, 32, 419–428, doi:10.4314/wsa.v32i3.5268.
5. Bourgeois, W.; Burgess, J.; Stuetz, R. On-line monitoring of wastewater quality: A review. *J. Chem. Technol. Biotechnol.* 2001, 76, 337–348, doi:10.1002/jctb.393.
6. Melendez-Pastor, I.; Almendro-Candel, M.B.; Pedreno, J.N.; Gómez, I.; Lillo, M.G.; Hernández, E.I. Monitoring Urban Wastewaters' Characteristics by Visible and Short Wave Near-Infrared Spectroscopy. *Water* 2013, 5, 2026–2036, doi:10.3390/w5042026.
7. Brzezińska, A.; Zawilski, M.; Sakson, G. Assessment of pollutant load emission from combined sewer overflows based on the online monitoring. *Environ. Monit. Assess.* 2016, 188, 1–10, doi:10.1007/s10661-016-5461-6.
8. Chen, B.; Wu, H.; Li, S.F.Y. Development of variable pathlength UV–vis spectroscopy combined with partial-least-squares regression for wastewater chemical oxygen demand (COD) monitoring. *Talanta* 2014, 120, 325–330, doi:10.1016/j.talanta.2013.12.026.
9. Gondal, M. A., Hussain, T. Determination of poisonous metals in wastewater collected from paint manufacturing plant using laser-induced breakdown spectroscopy. *Talanta*, 2017, 71, 73-80.
10. Ferree, M.; Shannon, R.D. Evaluation of a second derivative UV/visible spectroscopy technique for nitrate and total nitrogen analysis of wastewater samples. *Water Res.* 2001, 35, 327–332, doi:10.1016/s0043-1354(00)00222-0.
11. Qin, X.; Gao, F.; Chen, G. Wastewater quality monitoring system using sensor fusion and machine learning techniques. *Water Res.* 2012, 46, 1133–1144, doi:10.1016/j.watres.2011.12.005.
12. Lepot, M.; Torres, A.; Höfer, T.; Caradot, N.; Gruber, G.; Aubin, J.-B.; Bertrand-Krajewski, J.-L. Calibration of UV/Vis spectrophotometers: A review and comparison of different methods to estimate TSS and total and dissolved COD concentrations in sewers, WWTPs and rivers. *Water Res.* 2016, 101, 519–534, doi:10.1016/j.watres.2016.05.070.

13. Wolf, C.; Gaida, D.; Stuhlsatz, A.; Ludwig, T.; McLoone, S.F.; Bongards, M. Predicting organic acid concentration from UV/vis spectrometry measurements— A comparison of machine learning techniques. *Trans. Inst. Meas. Control.* 2011, 35, 5–15, doi:10.1177/0142331211403797.
14. Drolc, A.; Vrtovšek, J. Nitrate and nitrite nitrogen determination in waste water using on-line UV spectrometric method. *Bioresour. Technol.* 2010, 101, 4228–4233, doi:10.1016/j.biortech.2010.01.015.
15. Korshin, G.V.; Sgroi, M.; Ratnaweera, H. Spectroscopic surrogates for real time monitoring of water quality in wastewater treatment and water reuse. *Curr. Opin. Environ. Sci. Heal.* 2018, 2, 12–19, doi:10.1016/j.coesh.2017.11.003.
16. Hoppe, H.; Messmann, S.; Giga, A.; Gruening, H. A real-time control strategy for separation of highly polluted storm water based on UV–Vis online measurements— From theory to operation. *Water Sci. Technol.* 2011, 63, 2287–2293, doi:10.2166/wst.2011.164.
17. Brito, R.S.; Pinheiro, H.M.; Ferreira, F.; Matos, J.S.; Lourenço, N.D. In situ UV-Vis spectroscopy to estimate COD and TSS in wastewater drainage systems. *Urban Water J.* 2013, 11, 261–273, doi:10.1080/1573062x.2013.783087.
18. Launay, M. A., Dittmer, U., Steinmetz, H. Organic micropollutants discharged by combined sewer overflows—characterisation of pollutant sources and stormwater-related processes. *Water research*, 2016, 104, 82-92.
19. Mesquita, D.; Quintelas, C.; Amaral, A.L.P.D.; Ferreira, E.C. Monitoring biological wastewater treatment processes: Recent advances in spectroscopy applications. *Rev. Environ. Sci. Bio/Technol.* 2017, 16, 395–424, doi:10.1007/s11157-017-9439-9.
20. Fleischmann, N.; Langergraber, G.; Weingartner, A.; Hofstaedter, F.; Nusch, S.; Maurer, P. On-Line and in-Situ Measurement of Turbidity and COD in Wastewater Using UV-Vis Spectrometry. Available online: <https://www.s-can.at/> (accessed on 30 September 2020).
21. Hochedlinger, M.; Hofbauer, P.; Wandl, G.; Meyer, S.; Rauch, W.; Kroiss, H.; Heindl, M. Online UV-Vis measurements—The basis for future pollution based sewer real time control in Linz. In *Proceedings of the 2nd International IWA Conference on Sewer Operation and Maintenance*, Vienna, Austria, 26–28 October 2006.
22. Van den Broeke, J. On-line and in-situ UV/vis spectroscopy. *AWE Int.* 2007, 55–59.
23. Gruber, G.; Bertrand-Krajewski, J.L.; Beneditis, J.D.; Hochedlinger, M.; Lettl, W. Practical aspects, experiences and strategies by using UV-Vis sensors for long-term sewer monitoring. *Water Pract. Technol.* 2006, 1.
24. Sarraguça, M.C.; Paulo, A.; Alves, M.M.; Dias, A.; Lopes, J.; Ferreira, E.C. Quantitative monitoring of an activated sludge reactor using on-line UV-visible and near-infrared spectroscopy. *Anal. Bioanal. Chem.* 2009, 395, 1159–1166, doi:10.1007/s00216-009-3042-z.
25. Fogelman, S.; Zhao, H.; Blumenstein, M.; Zhang, S. Estimation of oxygen demand levels using UV–vis spectroscopy and artificial neural networks as an effective tool for real-time, wastewater treatment control. In *Proceedings of the 1st Australian*

- Young Water Professionals Conference, Sydney, Australia, 15–17 February 2006, pp. 15–17.
26. Del Río Cambeses, H. Estudio de los Flujos de Contaminación Movilizados en Tiempo de Lluvia y Estrategias de Gestión en un Sistema de Saneamiento y Drenaje unitario de una Cuenca Urbana Densa de la España Húmeda. Ph.D. Thesis, Universidade da Coruña, A Coruña, Spain, 2011.
 27. García, J.T.; Espín-Leal, P.; Viguera-Rodríguez, A.; Castillo, L.G.; Carrillo, J.M.; Martínez-Solano, P.; Nevado-Santos, S. Urban Runoff Characteristics in Combined Sewer Overflows (CSOs): Analysis of Storm Events in Southeastern Spain. *Water* 2017, 9, 303, doi:10.3390/w9050303.
 28. García, J.T.; Espín-Leal, P.; Viguera-Rodríguez, A.; Carrillo, J.M.; Castillo, L.G. Synthetic Pollutograph by Prediction Indices: An Evaluation in Several Urban Sub-Catchments. *Sustainability* 2018, 10, 2634, doi:10.3390/su10082634.
 29. Carreres-Prieto, D.; García, J.T.; Cerdán-Cartagena, F.; Suardíaz, J. Spectroscopy Transmittance by LED Calibration. *Sensors* 2019, 19, 2951, doi:10.3390/s19132951.
 30. Benavides, M.; Mailier, J.; Hantson, A.-L.; Muñoz, G.; Vargas, A.; Van Impe, J.F.; Wouwer, A.V. Design and Test of a Low-Cost RGB Sensor for Online Measurement of Microalgae Concentration within a Photo-Bioreactor. *Sensors* 2015, 15, 4766–4780, doi:10.3390/s150304766.
 31. Carreres-Prieto, D., García, J. T., Cerdán-Cartagena, F., & Suardiaz-Muro, J. (2020). Performing Calibration of Transmittance by Single RGB-LED within the Visible Spectrum. *Sensors*, 20(12), 3492.
 32. Bozhynov, V.; Soucek, P.; Barta, A.; Urbanova, P.; Bekkozhayeva, D. Visible Aquaphotomics Spectrophotometry for Aquaculture Systems. In *Proceedings of the Agreement Technologies*; Springer Science and Business Media LLC: Berlin, Germany, 2018; pp. 107–117.
 33. Wego, A. Accuracy simulation of an LED based spectrophotometer. *Optik* 2013, 124, 644–649, doi:10.1016/j.ijleo.2012.01.005.
 34. Schnable, J.G.; Grochowski, P.J.; Wilhelm, L.; Harding, C.; Kiefer, M.; Orr, R.S. Portable LED-array Vis–NIR spectrophotometer/nephelometer. *Field Anal. Chem. Technol.* 1998, 2, 21–28, doi:10.1002/(sici)1520-6521(1998)2:13.3.co;2-7.
 35. Rocha, F.R.P.; Martelli, P.B.; Reis, B.F. Simultaneous in-line concentration for spectrophotometric determination of cations and anions. *J. Braz. Chem. Soc.* 2004, 15, 38–42, doi:10.1590/s0103-50532004000100008.
 36. Venugopalan, H. UVC LEDs enable cost-effective spectroscopic instruments. *Laser Focus World* 2015, 51, 81–85.
 37. De la Torre, C.; Muñoz, R.; Pérez, M.A. A new, low-cost, on-line RGB colorimeter for wine industry based on optical fibers. In *Proceedings of the XIX IMEKO World Congress*, Lisbon, Portugal, 6–11 September 2009; pp. 6–11.
 38. Sampedor, Óscar; Salgueiro, J.R. Turbidimeter and RGB sensor for remote measurements in an aquatic medium. *Measurements* 2015, 68, 128–134, doi:10.1016/j.measurement.2015.02.049.
 39. Lima, M.B.; Andrade, S.I.; Neta, M.S.S.; Barreto, I.S.; Almeida, L.F.; De Araújo, M.C.U. A Micro-Flow-Batch Analyzer using Webcam for Spectrophotometric

- Determination of Ortho -phosphate and Aluminium(III) in Tap Water. *J. Braz. Chem. Soc.* 2014, 25, 898–906, doi:10.5935/0103-5053.20140059.
40. Suzuki, Y.; Aruga, T.; Kuwahara, H.; Kitamura, M.; Kuwabara, T.; Kawakubo, S.; Iwatsuki, M. A simple and portable colorimeter using a red-green-blue light-emitting diode and its application to the on-site determination of nitrite and iron in river-water. *Anal. Sci.* 2004, 20, 975–977, doi:10.2116/analsci.20.975.
 41. Carreres-Prieto, D., García, J. T., Cerdán-Cartagena, F., & Suardiaz-Muro, J. (2020). Wastewater Quality Estimation Through Spectrophotometry-Based Statistical Models. *Sensors*, 20(19), 5631.
 42. Agenda 2030. <https://www.agenda2030.gob.es/objetivos/>
 43. Ngo, T. H. D., & La Puente, C. A. (2012, April). The steps to follow in a multiple regression analysis. In *SAS Global forum* (Vol. 2012, pp. 1-12).
 44. Nesvizhskii, A. I., Keller, A., Kolker, E., & Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Analytical chemistry*, 75(17), 4646-4658.
 45. Abdi, H. (2003). Partial least square regression (PLS regression). *Encyclopedia for research methods for the social sciences*, 6(4), 792-795.
 46. Abdi, H., & Williams, L. J. (2013). Partial least squares methods: partial least squares correlation and partial least square regression. In *Computational toxicology* (pp. 549-579). Humana Press, Totowa, NJ.
 47. Wu, Z., Tran, A., Rincon, D., & Christofides, P. D. (2019). Machine learning-based predictive control of nonlinear processes. Part I: theory. *AIChE Journal*, 65(11), e16729.
 48. Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4), 18-28.
 49. Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer Science & Business Media.
 50. Wang, L. (Ed.). (2005). *Support vector machines: theory and applications* (Vol. 177). Springer Science & Business Media.
 51. Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, 20(1), 61-80.
 52. Michalski, R. S. (2000). Learnable evolution model: Evolutionary processes guided by machine learning. *Machine learning*, 38(1), 9-40.
 53. Swinehart, D. F. (1962). The beer-lambert law. *Journal of chemical education*, 39(7), 333.
 54. Katipamula, S., Reddy, T. A., & Claridge, D. E. (1998). Multivariate regression modeling.
 55. Zou, K. H., Tuncali, K., & Silverman, S. G. (2003). Correlation and simple linear regression. *Radiology*, 227(3), 617-628.
 56. Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3), 370-374.
 57. Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician*, 36(3a), 158-160.
 58. Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92-107.

-
59. Craney, T. A., & Surlles, J. G. (2002). Model-dependent variance inflation factor cutoff values. *Quality Engineering*, 14(3), 391-403.
60. Multicolinealidad. <https://www.uv.es/uriel/material/multicolinealidad3.pdf>
61. O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, 41(5), 673-690.
62. Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2(1), 21-33.
63. Hanusz, Z., & Tarasińska, J. (2015). Normalization of the Kolmogorov–Smirnov and Shapiro–Wilk tests of normality. *Biometrical Letters*, 52(2), 85-93.
64. West, B. J., & Shlesinger, M. (1990). The noise in natural phenomena. *American Scientist*, 78(1), 40-45.
65. Manly, B. F. (1976). Exponential data transformations. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 25(1), 37-42.
66. Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics, London*, 22(1), 23-36.
67. Granados, R. M. (2016). Modelos de regresión lineal múltiple. *Granada, España: Departamento de Economía Aplicada, Universidad de Granada*.
68. Bertrand-Krajewski, J. L. (2004). TSS concentration in sewers estimated from turbidity measurements by means of linear regression accounting for uncertainties in both variables. *Water Science and Technology*, 50(11), 81-88.
69. Turbidímetro Endress+Hauser CUS 1
https://portal.endress.com/dla/5000566/4111/000/00/BA117CEN_1196.pdf
70. NF EN 27027 Calidad del agua. Determinación de la turbiedad.
71. Bertrand-Krajewski, J. L., Barraud, S., & Chocat, B. (2000). Need for improved methodologies and measurements for sustainable management of urban water systems. *Environmental impact assessment review*, 20(3), 323-331.
72. Bersinger, T., Pigot, T., Bareille, G., & Le Hecho, I. (2013). Continuous monitoring of turbidity and conductivity: a reliable, easy and economic tool for sanitation management. *WIT Transactions on Ecology and the Environment*, 171, 151-162.
73. Hur, J., & Cho, J. (2012). Prediction of BOD, COD, and total nitrogen concentrations in a typical urban river using a fluorescence excitation-emission matrix with PARAFAC and UV absorption indices. *Sensors*, 12(1), 972-986.
74. Brito, R. S., Pinheiro, H. M., Ferreira, F., Matos, J. S., & Lourenço, N. D. (2014). In situ UV-Vis spectroscopy to estimate COD and TSS in wastewater drainage systems. *Urban Water Journal*, 11(4), 261-273.
75. Espín Leal, P. (2016). Estudio de los flujos de contaminación transportados por un sistema de saneamiento y drenaje unitario en tiempo de lluvia para la ciudad de Murcia.
76. DaCosta, R. S., Andersson, H., & Wilson, B. C. (2003). Molecular Fluorescence Excitation–Emission Matrices Relevant to Tissue Spectroscopy. *Photochemistry and photobiology*, 78(4), 384-392.
77. Harshman, R. A., & Lundy, M. E. (1994). PARAFAC: Parallel factor analysis. *Computational Statistics & Data Analysis*, 18(1), 39-72.

78. Morales, G. M. A., & Domínguez, A. J. C. (2009). La regresión por mínimos cuadrados parciales: orígenes y evolución. In *Historia de la probabilidad y la estadística (IV)* (pp. 441-448). Universidad de Huelva.
79. Wold, H. (1975). Path models with latent variables: The NIPALS approach. In *Quantitative sociology* (pp. 307-357). Academic Press.
80. Qiong, Y. A. N. G., Zhenyao, L. I. U., & Jidong, Y. A. N. G. (2009). Simultaneous determination of chemical oxygen demand (COD) and biological oxygen demand (BOD5) in wastewater by near-infrared spectrometry. *Journal of Water Resource and Protection*, 2009.
81. Espectrofotómetro U-4100 UV-Vis-NIR https://www.hitachi-hightech.com/file/us/pdf/library/literature/U-4100_General_Brochure.pdf
82. Martens, H. (2001). Reliable and relevant modelling of real world data: a personal account of the development of PLS regression. *Chemometrics and intelligent laboratory systems*, 58(2), 85-95.
83. S. Weisberg, "Applied Linear Regression [M]," New York: John Wiley & Sons. Vol. 168, 1985
84. Han, S., Qubo, C., & Meng, H. (2012, June). Parameter selection in SVM with RBF kernel function. In *World Automation Congress 2012* (pp. 1-4). IEEE.
85. Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In *Neural networks for perception* (pp. 65-93). Academic Press.
86. Zelinka, I., Oplatkova, Z., & Nolle, L. (2005). Analytic programming—Symbolic regression by means of arbitrary evolutionary algorithms. *International Journal of Simulation: Systems, Science and Technology*, 6(9), 44-56.
87. Qin, X., Gao, F., & Chen, G. (2012). Wastewater quality monitoring system using sensor fusion and machine learning techniques. *Water research*, 46(4), 1133-1144.
88. Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical methods in medical research*, 22(3), 278-295.
89. Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y. M., Park, J. P., ... & Cho, K. H. (2015). Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *Journal of Environmental Sciences*, 32, 90-101.
90. Granata, F., Papirio, S., Esposito, G., Gargano, R., & De Marinis, G. (2017). Machine learning algorithms for the forecasting of wastewater quality indicators. *Water*, 9(2), 105.
91. S::can™. <https://www.s-can.at>
92. Sonda de espectrofotometría UV-Vis de la marca Rycewater™. <https://roycewater.com.au/>
93. Sonda VWA-3000UV de la empresa Vasthi Engineers Private Limited™.
94. ISA system with UV-Vis spectrometer Probe. <https://www.go-sys.de/>
95. MacIsaac, D., Kanner, G., & Anderson, G. (1999). Basic physics of the incandescent lamp (lightbulb). *The physics teacher*, 37(9), 520-525.
96. [REDACTED]
97. [REDACTED]

-
98. [Redacted]
[Redacted]
 99. [Redacted]
[Redacted]
 100. [Redacted]
[Redacted]
[Redacted]
 101. [Redacted]
[Redacted]
[Redacted]
 102. [Redacted]
[Redacted]
[Redacted]
 103. [Redacted]
[Redacted]
[Redacted]
 104. [Redacted]
 105. [Redacted]
[Redacted]
 106. [Redacted]
[Redacted]
[Redacted]
 107. [Redacted]
[Redacted]
 108. [Redacted]
[Redacted]
 109. [Redacted]
 110. [Redacted]
[Redacted]
 111. [Redacted]
[Redacted]
 112. [Redacted]
[Redacted]
[Redacted]
 113. [Redacted]
[Redacted]
 114. [Redacted]
[Redacted]
 115. [Redacted]
 116. [Redacted]
[Redacted]
 117. [Redacted]
[Redacted]
 118. [Redacted]

119. [Redacted]
[Redacted]
120. [Redacted]
[Redacted]
121. [Redacted]
[Redacted]
122. [Redacted]
[Redacted]
[Redacted]
123. [Redacted]
[Redacted]
124. [Redacted]
[Redacted]
125. [Redacted]
[Redacted]
[Redacted]
126. [Redacted]
[Redacted]
[Redacted]
127. [Redacted]
[Redacted]
[Redacted]
128. [Redacted]
[Redacted]
129. [Redacted]
[Redacted]
130. [Redacted]
131. [Redacted]
[Redacted]
132. [Redacted]
[Redacted]
133. [Redacted]
134. [Redacted]
[Redacted]
[Redacted]
[Redacted]
135. [Redacted]
[Redacted]
136. [Redacted]
[Redacted]
[Redacted]
[Redacted]
[Redacted]
137. [Redacted]
[Redacted]

138. [REDACTED]
[REDACTED]
139. [REDACTED]
140. [REDACTED]
[REDACTED]
141. [REDACTED]
[REDACTED]
[REDACTED]
142. [REDACTED]
[REDACTED]
[REDACTED]
143. [REDACTED]
[REDACTED]
[REDACTED]
144. [REDACTED]
[REDACTED]
145. [REDACTED]
146. [REDACTED]
147. [REDACTED]
[REDACTED]
148. [REDACTED]
[REDACTED]
149. [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
150. [REDACTED]
[REDACTED]
151. Asuero, A. G., Sayago, A., & Gonzalez, A. G. (2006). The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, 36(1), 41-59.
152. Bean, M. (2015). *Laravel 5 essentials*. Packt Publishing Ltd.
153. Stauffer, M. (2019). *Laravel: Up & Running: A Framework for Building Modern PHP Apps*. O'Reilly Media.
154. Fedosejev, A. (2015). *React.js essentials*. Packt Publishing Ltd.
155. Filipova, O. (2016). *Learning Vue.js 2*. Packt Publishing Ltd.
156. Richer, J., Sanso, A., & Glazer, I. (2017). *OAuth 2 in Action*. Shelter Island: Manning Publications.
157. Sucasas, V., Mantas, G., Radwan, A., & Rodriguez, J. (2016, May). An OAuth2-based protocol with strong user privacy preservation for smart city mobile e-Health apps. In *2016 IEEE International Conference on Communications (ICC)* (pp. 1-6). IEEE.
158. Ljungstrand, P., Redström, J., & Holmquist, L. E. (2000, April). WebStickers: using physical tokens to access, manage and share bookmarks to the Web. In *Proceedings of DARE 2000 on Designing augmented reality environments* (pp. 23-31).

159. Rodríguez, J. B. (2019). La Ley Orgánica 3/2018 como marco embrionario de garantía de los derechos digitales laborales. Claves para un análisis sistemático. *Trabajo y derecho: nueva revista de actualidad y relaciones laborales*, (54), 49-78.
160. Postman. The Collaboration Platform for API Development. <https://www.postman.com/>
161. Ishizaki, K., Takeuchi, M., Kawachiya, K., Suganuma, T., Gohda, O., Inagaki, T., ... & Nakatani, T. (2003, October). Effectiveness of cross-platform optimizations for a Java just-in-time compiler. In *Proceedings of the 18th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications* (pp. 187-204).
162. Datasheet LED VAOL-5EUV8T4. <https://www.mouser.es/datasheet/2/423/VAOL-5EUV8T4-1064927.pdf>
163. Laybourn, P. J. R., Dakin, J. P., & Gambling, W. A. (1970). A photometer to measure light scattering in optical glass. *Opto-electronics*, 2(1), 36-42.
164. Svilainis, L. (2008). LED brightness control for video display application. *Displays*, 29(5), 506-511.
165. Ye, Z., Greenfeld, F., & Liang, Z. (2008, June). Design considerations of a high power factor SEPIC converter for high brightness white LED lighting applications. In *2008 IEEE Power Electronics Specialists Conference* (pp. 2657-2663). IEEE.
166. Lichtenthaler, H. K., & Buschmann, C. (2001). Chlorophylls and carotenoids: Measurement and characterization by UV-Vis spectroscopy. *Current protocols in food analytical chemistry*, 1(1), F4-3.
167. Mateo, M. P., Nicolas, G., & Yanez, A. (2007). Characterization of inorganic species in coal by laser-induced breakdown spectroscopy using UV and IR radiations. *Applied Surface Science*, 254(4), 868-872.
168. Suárez-García, F., Martínez-Alonso, A., Llorente, M. F., & Tascón, J. M. D. (2002). Inorganic matter characterization in vegetable biomass feedstocks. *Fuel*, 81(9), 1161-1169.
169. Castillo, S. A. P. Validación de la metodología de identificación y cuantificación de Colorantes Artificiales (Rojo Allura FD& C rojo No. 40 (16,035) y Amarillo Ocaso FD&C amarillo No. 6 (15,985)) en mermeladas de fresa en presentación de tarro de vidrio 290g distribuidas en la Ciudad Capital de Guatemala.
170. LED SSL-LX5093UEGC con longitud de onda pico de 500nm. <https://www.mouser.es/datasheet/2/244/SSL-LX5093UEGC-1137511.pdf>
171. Searle, S. R., & Gruber, M. H. (1971). *Linear models* (Vol. 10). New York: Wiley.
172. Coutsiás, E. A., Seok, C., & Dill, K. A. (2004). Using quaternions to calculate RMSD. *Journal of computational chemistry*, 25(15), 1849-1857
173. Uchida, Y., & Taguchi, T. (2005). Lighting theory and luminous characteristics of white light-emitting diodes. *Optical Engineering*, 44(12), 124003.
174. Narendran, N., & Deng, L. (2002, November). Color rendering properties of LED light sources. In *Solid State Lighting II* (Vol. 4776, pp. 61-67). International Society for Optics and Photonics.
175. Suzuki, Y., Aruga, T., Kuwahara, H., Kitamura, M., Kuwabara, T., Kawakubo, S., & Iwatsuki, M. (2004). A simple and portable colorimeter using a red-green-

- blue light-emitting diode and its application to the on-site determination of nitrite and iron in river-water. *Analytical sciences*, 20(6), 975-977.
176. Gunawan, A. I., Pratama, A. E., Dewantara, B. S. B., Puspitasari, I., & Setyastuti, T. A. (2019, September). A Study for Estimation of Bio Organism Content in Aquaculture Pond Based on Image Color and Light Intensity. In *2019 International Electronics Symposium (IES)* (pp. 650-654). IEEE.
177. Benavides, M., Mailier, J., Hantson, A. L., Muñoz, G., Vargas, A., Van Impe, J., & Vande Wouwer, A. (2015). Design and test of a low-cost RGB sensor for online measurement of microalgae concentration within a photo-bioreactor. *Sensors*, 15(3), 4766-4780.
178. Knochen, M., Rodríguez-Silva, J. C., & Silva-Silva, J. (2020). Exploitation of reaction mechanisms for sensitivity enhancement in the determination of phosphorus by sequential injection analysis. *Talanta*, 209, 120589.
179. Qu, X., Wong, S. C., & Chi, K. T. (2007, November). Color control system for RGB LED light sources using junction temperature measurement. In *IECON 2007-33rd Annual Conference of the IEEE Industrial Electronics Society* (pp. 1363-1368). IEEE.
180. Muthu, S., Schuurmans, F. J., & Pashley, M. D. (2002). Red, green, and blue LEDs for white light illumination. *IEEE Journal of selected topics in quantum electronics*, 8(2), 333-338.
181. Bader, H., & Hoigné, J. (1981). Determination of ozone in water by the indigo method. *Water research*, 15(4), 449-456.
182. RGB-LEDs HV-5RGB25. <https://www.digikey.com>
183. Fang, T., Zhang, F., Yuan, S., Zeng, H., & Song, J. (2019). Recent advances and prospects toward blue perovskite materials and light-emitting diodes. *InfoMat*, 1(2), 211-233.
184. Ziegler, G., Lanig, P., Theis, D., & Weyrich, C. (1983). Single crystal growth of SiC substrate material for blue light emitting diodes. *IEEE Transactions on Electron Devices*, 30(4), 277-281.
185. Takahashi, K., Yoshikawa, A., & Sandhu, A. (2007). Wide bandgap semiconductors. *Fundamental properties and Modern Photonic and Electronic Devices*.
186. Yoder, M. N. (1996). Wide bandgap semiconductor materials and devices. *IEEE Transactions on Electron Devices*, 43(10), 1633-1636.
187. Yamada, M., Mitani, T., Narukawa, Y., Shioji, S., Niki, I., Sonobe, S., ... & Mukai, T. (2002). InGaN-based near-ultraviolet and blue-light-emitting diodes with high external quantum efficiency using a patterned sapphire substrate and a mesh electrode. *Japanese Journal of Applied Physics*, 41(12B), L1431.
188. Coe-Sullivan, S., Steckel, J. S., Kim, L., Bawendi, M. G., & Bulovic, V. (2005, March). Method for fabrication of saturated RGB quantum dot light-emitting devices. In *Light-Emitting Diodes: Research, Manufacturing, and Applications IX* (Vol. 5739, pp. 108-115). International Society for Optics and Photonics.
189. Amano, H., Kito, M., Hiramatsu, K., & Akasaki, I. (1989). P-type conduction in Mg-doped GaN treated with low-energy electron beam irradiation (LEEBI). *Japanese Journal of Applied Physics*, 28(12A), L2112.
190. Nanishi, Y. (2014). The birth of the blue LED. *Nature Photonics*, 8(12), 884-886.

191. Muthu, S., Schuurmans, F. J., & Pashley, M. D. (2002, October). Red, green, and blue LED based white light generation: issues and control. In *Conference Record of the 2002 IEEE Industry Applications Conference. 37th IAS Annual Meeting (Cat. No. 02CH37344)* (Vol. 1, pp. 327-333). IEEE.
192. Young, A.T. Rendering Spectra. <http://cort.as/-SvpK>
193. Sharpe, L. T., Stockman, A., Jägle, H., & Nathans, J. (1999). Opsin genes, cone photopigments, color vision, and color blindness. *Color vision: From genes to perception*, 3-51.
194. Kelber, A. (2016). Colour in the eye of the beholder: receptor sensitivities and neural circuits underlying colour opponency and colour perception. *Current Opinion in Neurobiology*, 41, 106-112.
195. Bruton, D. Colour Rendering Algorithm. <http://www.physics.sfasu.edu/astro/color/spectra.html>
196. Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and computing*, 4(2), 65-85.
197. Mirjalili, S. (2019). Genetic algorithm. In *Evolutionary algorithms and neural networks* (pp. 43-55). Springer, Cham.
198. Shoshani, A. (1982, September). Statistical Databases: Characteristics, Problems, and some Solutions. In *VLDB* (Vol. 82, pp. 208-222).
199. 4500-CL G. METODO ESTANDAR. https://www.nemi.gov/methods/method_summary/7431/
200. Organización Internacional de Normalización (ISO). <https://www.iso.org/>
201. Asociación Americana de Salud Pública (APHA). <https://www.apha.org/>
202. la Asociación Americana de Obras Hidráulicas (AWWA)
203. Federación del Medio Ambiente del Agua (WEF)
204. Coutsiias, E. A., Seok, C., & Dill, K. A. (2004). Using quaternions to calculate RMSD. *Journal of computational chemistry*, 25(15), 1849-1857.
205. Parra, L., Rocher, J., Escrivá, J., & Lloret, J. (2018). Design and development of low cost smart turbidity sensor for water quality monitoring in fish farms. *Aquacultural Engineering*, 81, 10-18.
206. Yang, J., & Honavar, V. (1998). Feature subset selection using a genetic algorithm. In *Feature extraction, construction and selection* (pp. 117-136). Springer, Boston, MA.
207. Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.
208. Mundry, R., & Nunn, C. L. (2009). Stepwise model fitting and statistical inference: turning noise into signal pollution. *The American Naturalist*, 173(1), 119-123.
209. Osborne, J. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation*, 15(1), 12.
210. Augusto, D. A., & Barbosa, H. J. (2000, November). Symbolic regression via genetic programming. In *Proceedings. Vol. 1. Sixth Brazilian Symposium on Neural Networks* (pp. 173-178). IEEE.
211. Davidson, J. W., Savic, D. A., & Walters, G. A. (2003). Symbolic and numerical regression: experiments and applications. *Information Sciences*, 150(1-2), 95-117.
212. Abadi, M.; Barham, P., Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Kudlur, M. Tensorflow: A system for large-scale machine learning. In *Proceedings of the 12th*

- {USENIX} symposium on operating systems design and implementation ({OSDI} 16), Savannah, GA, USA, 2–4 November 2016, pp. 265–283.
213. Oliphant, T.E. A Guide to NumPy. Available online: <https://ecs.wgtn.ac.nz/foswiki/pub/Support/ManualPagesAndDocumentation/numpybook.pdf>
214. Van Der Walt, S.J.; Colbert, S.C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* 2011, 13, 22–30, doi:10.1109/MCSE.2011.37.
215. GPLEARN. <https://gplearn.readthedocs.io/en/stable/>
216. Cha, S. H., & Tappert, C. C. (2009). A genetic algorithm for constructing compact binary decision trees. *Journal of pattern recognition research*, 4(1), 1-13.
217. Palmer, C. C., & Kershenbaum, A. (1994, June). Representing trees in genetic algorithms. In *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence* (pp. 379-384). IEEE.
218. Brookman, S. Estimation of biochemical oxygen demand in slurry and effluents using ultra-violet spectrophotometry. *Water Res.* 1997, 31, 372–374, doi:10.1016/s0043-1354(96)00250-3.
219. Fredin, L.; Nelander, B.; Ribbegård, G. Infrared spectrum of the water dimer in solid nitrogen. I. Assignment and force constant calculations. *J. Chem. Phys.* 1977, 66, 4065–4072, doi:10.1063/1.434478.
220. Dalal, R.C.; Henry, R.J. Simultaneous Determination of Moisture, Organic Carbon, and Total Nitrogen by Near Infrared Reflectance Spectrophotometry. *Soil Sci. Soc. Am. J.* 1986, 50, 120–123, doi:10.2136/sssaj1986.03615995005000010023x.
221. Dalal, R.C.; Henry, R.J. Simultaneous Determination of Moisture, Organic Carbon, and Total Nitrogen by Near Infrared Reflectance Spectrophotometry. *Soil Sci. Soc. Am. J.* 1986, 50, 120–123, doi:10.2136/sssaj1986.03615995005000010023x.