

Treball de Fi de Màster
Màster en Enginyeria Química

**Desenvolupament d'un *soft sensor* a partir de
dades del tractament foto-Fenton en una planta
pilot**

MEMÒRIA

Autor: Patricia Vicente Núñez
Director: Moisès Graells Sobré
Convocatoria: Juny 2016



Escola Tècnica Superior
d'Enginyeria Industrial de Barcelona



Resum

Recentment, el procés de tractament foto-Fenton ha demostrat ser molt eficaç en la degradació d'una àmplia gama de contaminants orgànics, especialment farmaceutics i d'higiene personal, que són difícils de ser tractats per tractaments convencionals físics, biològics i/o químics. No obstant, degut a la complexitat del procés, la tasca de modelització continua siguent objecte de debat. Els models disponibles, més utilitzats, són útils pels estudis experimentals del procés a escala laboratori i per rangs concrets de condicions inicials. Tot i així, no existeixen models fiables disponibles per l'ús d'aquests en línia en una escala de planta industrial, ja que no hi ha suficient temps ni recursos pel mostreig i anàlisi fora de línia experimental.

Aquest treball investiga la modelització basada en dades del procés del tractament d'aigües foto-Fenton, que és un procés no lineal complex, pel que no existeix un model de principis bàsics útil pel seguiment i control del procés. Aquests models empírics es poden utilitzar com a *soft sensors* a fi supervisar o predir el progrés de la reacció, que és car per mesurar, com funció d'altres variables del sistema que es registren de forma contínua, automàtica i barata, lo que redueix el cost de la presa de mostres i anàlisis d'aquestes. La tècnica de modelat basada en dades utilitzada en aquest projecte és la *Support Vector Regression*. Un exemple de simulació, com és la cinètica de primer ordre, s'utilitza primer per plantejar el problema, avaluar i validar el mètode de modelitzat i comprovar la seva robustesa. A continuació, la metodologia s'aplica a una planta de tractament foto-Fenton per modelar i predir el progrés de la degradació del contaminant. Els resultats obtinguts es mostren prometedores inclús quan només alguns punts d'entrenament estan disponibles, lo que es tradueix en un gran estalvi de temps i cost en la part experimental.

Resumen

Recientemente, el proceso de tratamiento foto-Fenton ha demostrado ser muy eficaz en la degradación de una amplia gama de contaminantes orgánicos, especialmente farmacéuticos y de higiene personal, que son difíciles de ser tratados por tratamientos convencionales físicos, biológicos y/o químicos. Sin embargo, debido a la alta complejidad del proceso, la tarea de modelado sigue siendo objeto de debate. Los modelos disponibles, más utilizados, son útiles para los estudios experimentales del proceso a escala de laboratorio y para rangos concretos de condiciones iniciales. Aún así, no existen modelos fiables disponibles para el uso de estos en línea en una escala de planta industrial, ya que no hay suficiente tiempo ni recursos para el muestreo y análisis fuera de línea experimental.

Este trabajo investiga la modelización basado en datos del proceso de tratamiendo de aguas foto-Fenton, que es un proceso no lineal complejo, para el que no existe un modelo de principio básicos útil para el seguimiento y control del proceso. Estos modelos empíricos se pueden usar como *soft sensors* a fin de supervisar o predecir el progreso de la reacción, que es caro para medir, como función de otras variables del sistema que se registran de forma continua, automática y barata, lo que reduce el coste de toma de muestras y análisis. La técnica de modelado basada en datos utilizada en este proyecto es la *Support Vector Regression*. Un ejemplo de simulación, como es la cinètica de primer orden, se utiliza primero para plantear el problema, evaluar y validar el método de modelado y comprobar su robustez.. A continuación, la metodología se aplica a una planta piloto de tratamiento foto-Fenton para modelar y predecir el progreso de la degradación del contaminante. Los resultados obtendos se muestran prometedores incluso cuando solo algunos puntos de entrenamiento están disponibles, lo que se traduce en un gran ahorro de tiempo y coste en la parte experimental.

Abstract

Recently, the photo Fenton process has proved to be highly effective in of the degradation of a wide range of organic contaminants, especially pharmaceuticals and personal care products, that are difficult to be treated by conventional physical, biological and/or chemical treatments. However, due to the process high complexity, its modeling task still under dialectical debate. Even the available models are almost useful for the experimental studies of the process in a laboratory scale. And approximately, no models are available for the reliable online usage in an industrial or pilot plant scale, where it is necessary to monitor the processes, and no sufficient time and resources are available for the offline sampling and experimental analysis.

This work investigates data based modelling of the photo Fenton process, which is a complex nonlinear process, for which a first principle model useful for process monitoring and control is not available. These empirical models may be used as soft sensors in order to monitor/predict the process reaction's progress, which is expensive to measure, as a function of other online process variables, which are recorded continuously in an automatic and cheap way, so reducing expensive offline sampling and analysis. The data modelling technique used the Support Vector Regression. A simulation example is first used to state the problem, assess and validate the modelling approach and prove its robustness. Next, the methodology is applied to a photo-Fenton pilot plant to model and predict the reaction progress. The results show promising accuracy even when few training points are available, which results in huge savings of time and cost of the experimental work.

Sumari

RESUM	1
RESUMEN	3
ABSTRACT	5
SUMARI	7
1. INTRODUCCIÓ	9
1.1. Motivació	9
1.2. <i>Soft sensor</i>	11
1.3. Estat de l'art	14
1.4. Objectius del projecte	16
2. MÈTODES ANALÍTICS	17
2.1. Tècniques de modelat basades en dades	17
2.1.1. Kriging (OK)	17
2.1.2. <i>Artificial Neural Networks</i> (ANN)	19
2.1.3. Support Vector Regression (SVR)	20
2.2. Tècniques per validació de models	23
2.3. Programes per la modelització	23
2.3.1. R project	23
2.3.2. Python	24
2.3.3. MATLAB	24
3. ENFOC DE MODELAT	25
3.1. Casos d'estudi	27
3.1.1. Cinètica de primer ordre	27
3.1.2. Cabrera Reina	29
3.1.3. Aplicació d'un cas real: Planta Pilot	32
3.1.3.1. Material i mètode	32
3.1.3.2. Planta Pilot	32
4. RESULTATS I DISCUSSIÓ	37
4.1. Cinètica Primer Ordre	37
4.2. Cabrera Reina	40
4.3. Planta Pilot	42
5. AVALUACIÓ ECONÒMICA	47

5.1. Cost de Software	47
5.2. Amortitzacions	47
5.3. Cost de Personal	48
5.4. Cost Total.....	48
CONCLUSIONS	49
PROPOSTES DE MILLORA	51
AGRAÏMENTS	53
BIBLIOGRAFIA	55
Referències bibliogràfiques.....	55
Bibliografia complementària.....	58
ANNEX	59
<u>Codi de programació</u>	59
SVR (R project).....	59
Cinètica Primer Ordre (Matlab)	60
<u>Gràfica dels models</u>	61
Cinètica Primer Ordre	61
Entrenament	61
Validació	64
Cabrera Reina.....	65
Entrenament	65
Validation	67
Planta Pilot	68
Entrenament	68
Validació	69

1. Introducció

En aquest capítol es presenta una visió general sobre les qüestions ambientals relacionades amb l'escassetat de l'aigua i la contaminació, i la seva relació amb la indústria i la enginyeria. Es defineix què és un *soft sensor* i s'expliquen les aproximacions trobades a la literatura per la modelització del procés foto-Fenton. També es descriuen els objectius del treball.

1.1. Motivació

La població mundial supera els set milions d'habitants i s'estima que podria arribar als dotze milions en el 2100 (Gerland et al. 2014). Aquest augment de la població podria disparar el consum d'aigua, en especial en els països en vies de desenvolupament, on s'estima que succeeixi la major part de l'augment. Encara que més del 70% de la superfície de la Terra està coberta per aigua, només un 2.5% d'ella és aigua dolça, de la qual només és accessible un 0.26% (Shiklomanov 1998).

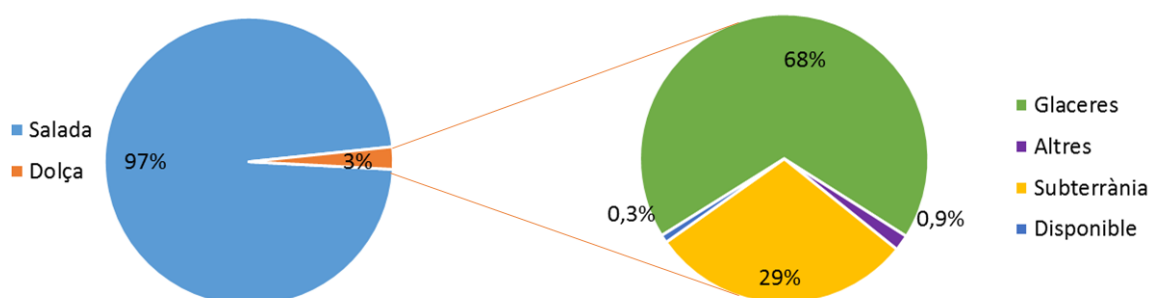


Figura 1. Distribució de l'aigua al planeta Terra. [Dades: Shiklomanov 1998]

El desenvolupament i la intensificació de l'activitat industrial des de la meitat del segle XIX, suposen un increment notable en l'ús d'aquesta, generant una gran quantitat de residus que en la seva gran majoria van a parar a l'aigua. L'aparició de lleis de regulació cada cop més estrictes, degut al compromís establert per la societat amb el medi ambient, obliga a les empreses a realitzar abocaments amb nivells de contaminació més baixos.

Durant l'última dècada, un nou grup de contaminants anomenat Contaminants Emergents (ECs), o recalcitrants, han sigut detectats en fonts naturals, aigües potables i residuals en ordre de ng-µg/L (Broséus et al. 2009; Petróvic et al. 2003). S'han dedicat una gran quantitat d'esforços a eliminar-los de les aigües residuals abans de que la seva acumulació es converteixi en un problema medioambiental real, degut a la seva alta

persistència i ecotoxicitat (Klamerth et al. 2010; Oller et al., 2011). Entre aquests contaminants emergents, es troben molts productes farmacèutics i d'higiene personal, que es coneixen per tenir la capacitat d'interferir en els processos naturals en el ecosistema, actuant com a disruptors endocrins o afectant a les bacteries utilitzades en els tractaments d'aigua convencionals (Homen et al. 2011). Aquests tractaments convencionals (biològics, físics o químics) es demostren que son incapaçs de fer front de manera eficient a aquesta classe de contaminants recalcitrants. Els Processos d'Oxidació Avançada (POA) apareixen com una alternativa molt eficaç a aquest processos, ja que poden degradar una amplia gama de compostos orgànics, per lo que es poden utilitzar abans o després dels tractaments convencionals amb el fi d'eliminar-los de les aigües residuals (Klamath et al., 2010; Pignatello et al., 2006). Un dels processos d'oxidació avançada més utilitzats és l'anomenat procés foto-Fenton o de fotocatalisi homogènia.

La naturalesa d'aquest procés és molt complexa, amb reaccions paral·leles completes i productes intermedis difícils d'identificar. Alguns models cinètics que s'han proposat, per a molècules molt simples, revelen que l'intent de resoldre rigorosament la degradació de molècules més gran i mescles és inassequible. En moltes fàbriques de l'indústria química, existeixen bases de dades amb històrics on hi ha emmagatzemades milers amb alta velocitat de mostreig. Moltes d'aquestes dades són dades de funcionament de la planta i variables de control, que poden ser fàcilment utilitzades per la identificació del sistema i el seu pronòstic. Gràcies aquesta informació, l'aplicació de tècniques de modelització basat en dades, com són els *soft sensors*, es mostren com una solució alternativa.

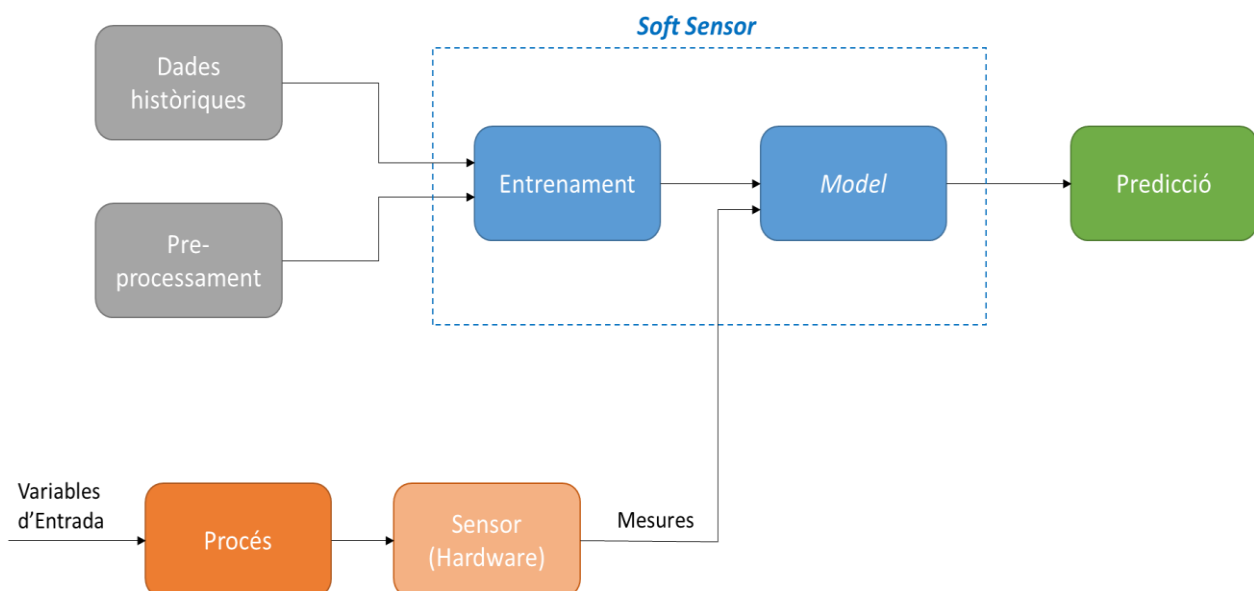


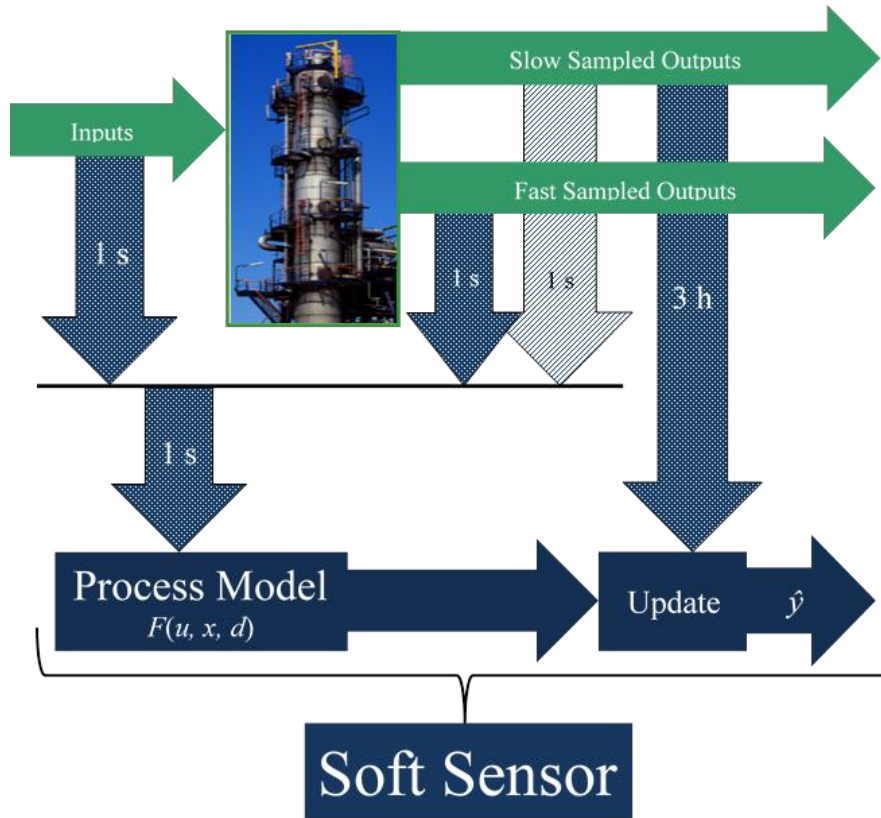
Figura 2. Diagrama de funcionament d'un *soft sensor*.

Aquestes tècniques es poden trobar que s'han utilitzat en els sistemes de reaccions complexes, principalment bioprocessos, per aprofitar de manera eficient mesures simples indirectes per la fiable monitorització del procés, després de l'adequada reducció de la dimensionalitat. Per sorprenent que pugui semblar, aquest enfoc és quasi inexistent en la literatura sobre els processos químics, en especial en el procés foto-Fenton.

Per tant, l'interès de l'indústria i el món acadèmic es centra en la busqueda de noves tècniques de modelat eficients, que permetin una major eficàcia en la generació del model i la reducció de costos dels processos.

1.2. *Soft sensor*

Els *soft sensors* són uns estimadors que infereixen i treuen conclusions a partir de les observacions del procés quan els sensors de hardware no estan disponibles, o no són aptes, tenint un paper auxiliar important en la validació del sensor quan el rendiment d'aquest disminueix, ja sigui per causa de l'envelliment o per l'acumulació de substàncies (Fortuna et al., 2007). Aquests poden ser aplicats per resoldre diferents problemes presents en el sistema industrial, formant part d'aquest el tractar amb dades parcials i errònies, i reconciliar dades des de múltiples dispositius de mesura diferent. A més, poden superar els problemes derivats de la falta de dades o la freqüència requerida. Els *soft sensors* són una valuosa eina amb una ampla aplicació en diferents camps industrials, incloent el petroli i el gas (Khatibisepehr et al., 2013), plantes químiques (Zamprogna et al., 2005) i la contaminació industrial i urbana, entre d'altres. En la Figura 3, es mostra una representació de la incorporació de la informació a *soft sensor*



Legend

Solid Green: Process Variables Blue, Hashed Arrows: Reliable Sensors
 Solid Blue: Soft Sensor Green, Hashed Arrows: Unreliable sensors

Figura 3. Esquema de funcionament d'un *soft sensor*. [Font: Yuri A.W. Shardt, 2012]

El disseny i implementació d'un *soft sensors* pot dividir-se en quatre apartats importants, que es mostren esquemàticament en la Figura 4 (Lin et al., 2007):

1) Reunir el coneixement, i comprensió, del procés a modelitzar

En aquest pas s'adquireix un coneixement profund del procés i dels objectius del disseny a través de la consulta a les persones pertinents, inclosos els operadors i gerents de planta.

2) Recopilació de dades i preprocessament previ

En aquest pas s'analitzen les dades per determinar la qualitat i utilitat de les mateixes pel disseny de *soft sensors*. Qüestions com la falta de dades o dades errònies, valors atípics i coneixement incomplet del procés, a vegades requereixen d'un retorn al pas 1 amb la finalitat d'obtenir la informació addicional que falta.

3) Crear el model del procés

En aquest pas, el model que representa el procés s'obté utilitzant qualsevol dels mètodes disponibles, que van des de l'anàlisi de regressió simple als complexos mètodes basats en dades. L'objectiu es desenvolupar un model que no només pugui predir amb exactitud els valors futurs del procés, si no que es la suficientment robust com per fer front a les variacions d'operació en la planta.

4) Validació del model del procés

Un cop el model ha sigut provat a fons fora de les instal·lacions, es necessita validar en línia, és a dir, en temps real, en la planta. Aquesta acció es realitzar per revelar nous problemes que des de la simulació no s'hagi detectat, com problemes en els valors d'entrada o un comportament inesperat en el rendiment, que han de ser tractats. En aquest punt, portar el model des de la solució obtinguda amb valors simulats a una solució que permeti la eficiència i efectivitat en línia en l'aplicació industrial, pot requerir un gran esforç. Això implica que molts cops el model haurà de ser simplificat o redissenyat degut a les demandes de càlcul del model.

Cap senyalar que aquest procés és un procés iteratiu que pot involucrar tornar de nou algun pas anterior, sobretot si hi ha problemes en l'obtenció de valors acceptables.

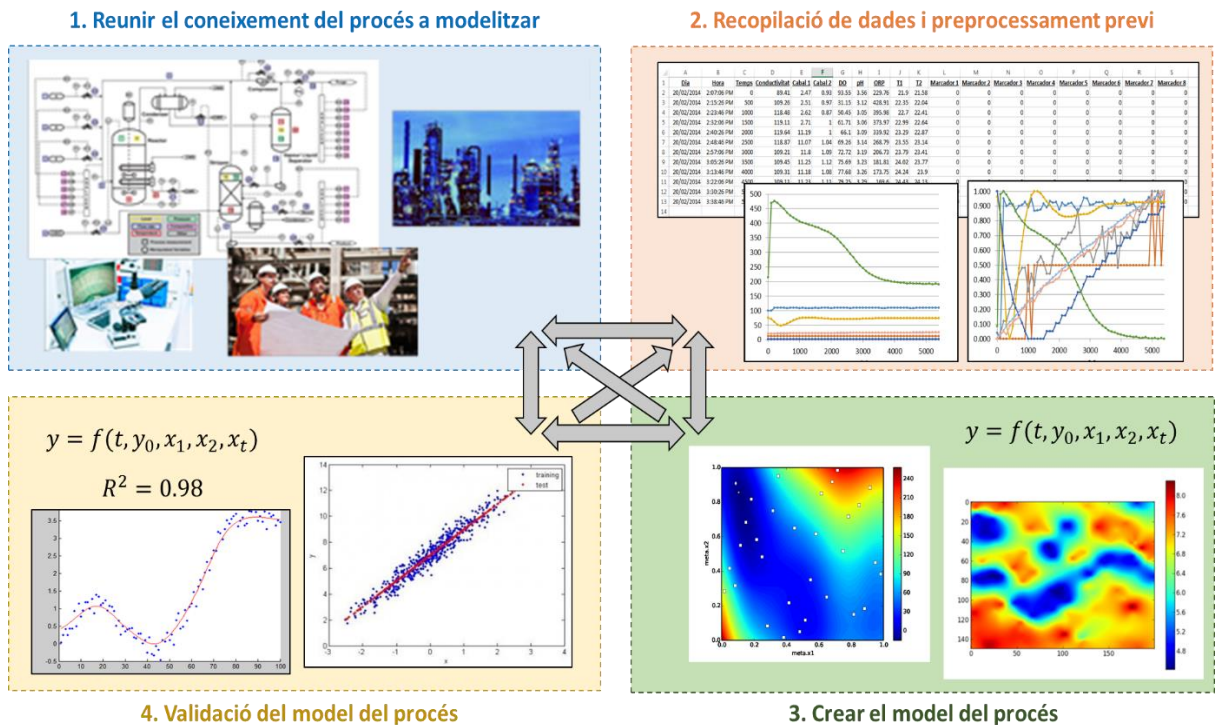
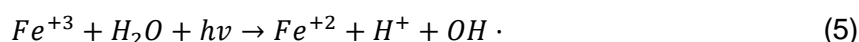
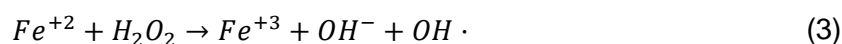


Figura 4. Esquema del disseny d'un *soft sensor*. [Imatges: Google]

Aquest projecte aborda el desenvolupament d'un *soft sensor* per una planta pilot, que treballa en lots, de tractament d'aigües residuals per un procés d'oxidació avançada, en concret el procés foto-Fenton (Pérez-Moya et al. 2008). Els Processos d'Oxidació Avançada (POA) són processos que es basen en la generació de radicals hidroxil altament oxidant ($\cdot\text{OH}$). Aquest radicals són capaços d'oxidar compostos orgànics no biodegradables per abstracció d'hidrogen, generant-se radicals orgànics lliures ($\cdot\text{R}$) els quals reaccionen a la seva vegada amb molècules d'oxigen formant un peroxiradical, iniciant-se una sèrie de reaccions de degradació oxidativa que poden conduir a la mineralització completa del compostos orgànics.



El procés foto-Fenton és un dels processos d'oxidació avançada més estudiats en les últimes dècades. Aquest procés utilitza una sal ferrosa (Fe^{+2}) i peròxid d'hidrogen (H_2O_2) per produir radicals hidroxil, així com una font de llum convenient que redueix de nou Fe^{+3} a Fe^{+2} , que en aquest sentit actua com a catalitzador i incrementa notablement la formació de hidroxil radicals. (Chacón et al, 2006; Tokumura et al, 2006)



La seqüenciació d'aquestes dos reaccions d'òxid-reducció permet la mineralització de contaminants orgànics a CO_2 i aigua (Tokumura et al., 2008).

Per tant el *soft sensor* esta destinat a ser utilitzats per la predicció en línia i el monitoreig del procés de degradació del contaminant d'estudi al llarg del temps mitjançant la mesura de quantitat de Carboni Orgànic Total (TOC), arribant a salvar així gran quantitat de temps, costos i esforç en la presa de mesures i anàlisi d'aquestes que es duen a terme de manera manual o *off-line*. El *soft sensor* que es desenvoluparà està basat en models basats en dades, els quals estan capacitats per utilitzar dades històriques mesurades de lots previs realitzats a la planta pilot. (Vicente, 2014)

1.3. Estat de l'art

Molts esforços s'han dedicat a entendre i predir el comportament del procés foto-Fenton,

però la complexitat del procés implica moltes hores de treball previ per la identificació completa dels subproductes involucrats i els mecanismes de degradació, tot i que les etapes més importants ja hagin estat resoltes.

El foto-Fenton és un procés no lineal que està format per fenòmens complexos, com per exemple la cinètica de la reacció, la transferència de massa i calor,...; i per les interaccions complexes entre les seves variables multidimensionals, lo que porta a un gran repte en el camp de la modelització. Com a resultat, els mecanismes d'oxidació pel reactiu Fenton encara estan en debat. No obstant, diferents enfoccs han estat suggerits per modelar el sistema de tractament d'aigües foto-Fenton. Alguns dels treballs proposen models analítics o basats en principis bàsics (FPM) (Cabrera Reina et al., 2012; Farias et al., 2009). Aquest models, descriuen l'evolució del sistema en termes de la cinètica de la degradació de les molècules orgàniques més simples de les variables observables, tals com el Carboni Orgànic Total (TOC). Tot i així, la complexa naturalesa i no linealitat del procés fan als models de principis bàsics incapaços d'incloure tots els mecanismes implicats de forma general. Claudio et al. 1999, descriu amb més detall com aquesta tasca ha resultat ser extremadament difícil, inclús quan es treballa amb les condicions experimentals més favorables i les geometries dels reactor més simples. D'altra banda, en una planta industrial o pilot, moltes altres variables afecten, o interactuen, en el sistema que els models de principis bàsics no tenen en compte, fet que complica l'ús d'aquests models per la monitorització i control del procés aquesta escala.

També s'han utilitzat models empírics de superfície de resposta o de regressió basats en resultats obtinguts d'experiments (Pérez-Moya et al., 2008). A pesar de la seva importància en proporcionar models simples de regressió capaços d'il·lustrar les relacions fundamentals entre les variables bàsiques que afecten el procés de degradació, aquests models de regressió tenen tendència a simplificar excessivament el comportament complex no lineal del sistema, mostrant grans limitacions per correlacionar un gran nombre de variables de procés (Shokry et al. 2014). Pel contrari, la modelització de sistemes complexos multivariants no lineals requereix de tècniques de modelat més avançades i sofisticades, les quals són capaces de capturar el comportament dels processos no lineals, i correlacions un major nombre de variables del sistema, com poden ser les xarxes neuronals o la *support vector machine*. Aquests models no lineals basats en dades, també anomenats models substitutius o metamodels, han sigut àmpliament utilitzats en moltes àrees de coneixement com a solució alternativa eficient per la generació de models predictius precisos basats en la simulació o en dades experimentals estalviant un enorme quantitat de temps i cost en treball experimental o inclús en complexes simulacions utilitzant sofisticats models de principis bàsics. Addicionalment, aquest models son capaços de correlacions variables de procés de diferents entitats, per exemple, químiques, físiques o mecàniques: o inclús correlacionar variables de diferents processos els quals interactuen, proporcionant eines de modelat

flexibles i exhaustives.

1.4. Objectius del projecte

L'objectiu d'aquest projecte és desenvolupar un soft sensor per quantificar el progrés d'una variable costosa, o *off-line*, a través de la correlació de les variable de monitoreig o *on-line*, les qual són mesures barates.

Per tal de poder aconseguir aquest propòsit, es plantegen els següents objectius:

- Comparar i seleccionar diferents tècniques de modelat basat en funció de la seva capacitat.
- Proposar un enfoc de modelat basat en dades eficient per a la caracterització de sistemes amb reacció química.
- Aplicar i validar l'enfoc a diferents casos d'estudi.
- Discutir la viabilitat de la seva aplicació en sistemes complexos amb reacció química.

2. Mètodes Analítics

En aquest apartat es presenta una visió general d'algunes de les tècniques, basades en dades, utilitzades per la modelització de processos i de les eines en les que aquestes es poden implementar. L'elecció de les diferents tècniques i eines, s'ha feta en base a l'ús generalitzat d'aquestes en l'enginyeria (Simpson et al. 2001).

2.1. Tècniques de modelat basades en dades

Gran part dels treballs que es porten a terme en el camp de l'enginyeria, avui en dia, requereixen de l'anàlisi i execució de codis complexos i computacionalment cars, com per exemple el models que representen la dinàmica de fluids. Malgrat el continu augment en les capacitats i velocitats dels processadors dels ordinadors, la gran quantitat de temps i costos computacionals d'execució de models complexos es continua mantenint. Una manera de superar aquest problemàtica és generar una aproximació del sistema complex a analitzar que descriu el seu comportament amb una precisió suficient però a un cost computacional molt més baix. Tals aproximacions a vegades reben el nom de metamodels, ja que es podria dir que proporcionen un "model del model" (Kleijnen, 1987)

Una manera senzilla d'explicar matemàticament el funcionament d'un metamodel seria suposant que el procés a analitzar està aproximat a una funció $y = f(x)$, on x representa el vector de les variables d'entrada i y el vector de les variables de sortida, ambdues referides al sistema real; i $f(x)$ és una funció que simula un procés de naturalesa complexa. L'aproximació eficient feta per un metamodel estaria representada per la funció $\hat{y} = g(x)$, de tal manera que la sortida del sistema real quedaria definida per l'equació $y = \hat{y} + \varepsilon$, on ε inclou tant els error de l'aproximació com els aleatoris (Clarke et al. 2005).

Actualment existeixen un gran nombre de metamodels per aproximar $f(x)$ a $g(x)$, tal com el *Kriging*, les xarxes neuronal artificial o la *Support Vector Regression*, les qual es detallarà el funcionament a continuació.

2.1.1. Kriging (OK)

El metamodel Kriging han estat desenvolupat en el camp de la geoestadística, i és un mètode d'interpolació òptim pel que els valors interpolats són obtinguts a partir d'una regressió pel procés Gaussià i ponderats d'acord amb els valors de covariància més rellevants que representa una mesura d'incertesa sobre el valor de la predicció. En les últimes dos dècades, aquest mètode ha demostrat tenir una alta capacitat per la modelització i simulació de sistemes complexos, altament no lineals i multimodals, com la

dinàmica de fluids o la aeronàutica, utilitzant un nombre relativament baix de punts d'entrenament comparat amb altres tècniques. Aquesta capacitat de proporcionar models precisos de tals sistemes, es deriva del fet de que és capaç de combinar alhora el modelat global i el local, a través de l'estimació d'una tendència general del sistema i de la funció de correlació respectivament. Recentment, el Kriging ha estat utilitzat per la modelització i l'optimització de processos químics i sistemes complexos.

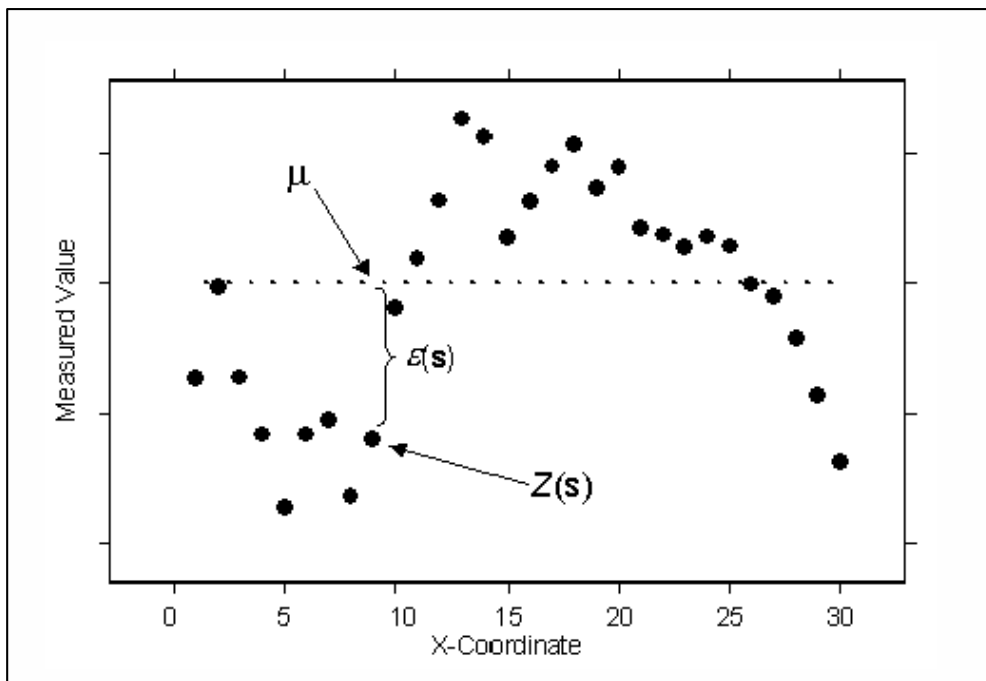


Figura 5. Exemple de funcionament del Kriging en una dimensió [Font: <https://desktop.arcgis.com/es/arcmap/latest/extensions/geostatistical-analyst/understanding-ordinary-kriging.htm>, Maig 2016].

Donat un conjunt de dades d'entrenament $[x_i, y_i]$, $i = 1, 2, \dots, n$; $x \in R^k$, $y \in R$ (on k és el número de variables de sortida), el model Kriging adopta un procés estocàstic en el que l'error en el valor predictiu és funció de les variables d'entrada x . El predictor $\hat{y}(x)$ està compost per un terme polinòmic $f(x) = \mu$, i una desviació $Z(x)$ d'aquest polinòmic. Per tant, $\hat{y}(x) = f(x) + Z(x)$, on $Z(x)$ és un procés estocàstic Gaussià amb un valor esperat de zero ($E(Z(\theta)) = 0$), i una covariància entre x_i , x_j com $cov(Z(x_i), Z(x_j)) = \sigma^2 R(x_i, x_j)$, sent σ^2 la variància del procés i $R(x_i, x_j)$ la funció de correlació $R(x_i, x_j) = \exp\left(-\sum_{l=1}^k \xi_l |x_{i,l} - x_{j,l}|^{p_l}\right) + \delta_{i,j} \lambda$, on δ_{ij} és la delta de Kronecker i λ , la constant de regularització que permet al predictor del Kriging revertir dades amb soroll i no passar a través d'elles sobre-ajustant el model (Azman et al. 2007). Per poder estimar els valors de μ , σ^2 , ξ_l , p_l i λ , la funció de probabilitat de les dades observades $[y]_{n \times 1}$ ha de ser maximitzada. El predictor del Kriging (Eq. 6) s'obté portant la funció de probabilitat augmentada a través dels valors d'entrenament originals i el nou punt

interpolat (x_{new}, y_{new}) . En l'equació 1, r és un vector de dimensions $nx1$ de correlació $R(x_{new}, x_i)$ entre el punt que s'ha de predir i els punts d'entrenament. La variància del predictor ve donada per l'equació 7.

$$\hat{y}(x_{new}) = \mu + r^T R^{-1}(Y - 1\mu) \quad (6)$$

$$\hat{\sigma}^2(x_{new}) = \hat{\sigma}^2(1 + \lambda - r^T R^{-1}r + (1 - r^T R^{-1}r)^2 / (1^T R^{-1}1)) \quad (7)$$

L'ajust del metamodel Kriging aconseguit, es fa mitjançant l'obtenció dels paràmetres òptims $[\mu, \sigma^2, \xi_i, P_i]$ a través de maximitzar la funció concentrada de la probabilitat logarítmica, que en la pràctica és un problema d'optimització computacionalment difícil. Això és degut al mal acondicionament de la matriu de correlació $R_{n \times n}$ i a la complicada naturalesa de la funció concentrada de la probabilitat logarítmica que es torna plana al voltant del punt òptim.

2.1.2. Artificial Neural Networks (ANN)

La Xarxa Neuronal Artificial, o *Artificial Neural Networks*, és un mètode eficient i àmpliament conegut que s'utilitza per la modelització i aproximació de sistemes no lineals. En el món de l'enginyeria és el mètode generalment més utilitzat per la modelització i identificació de sistemes (Z. K. Nagy, 2007). El seu desenvolupament s'inspira en el sistema nerviós del cos humà. Bàsicament, la ANN és una estructura d'unitats bàsiques de processament no lineal, les quals es denominen neurones. Aquestes neurones es col·loquen formant una estructura amb un número específic de capes que s'interconnecten gracies a certs pesos, tal i com s'observa a la Figura 6.

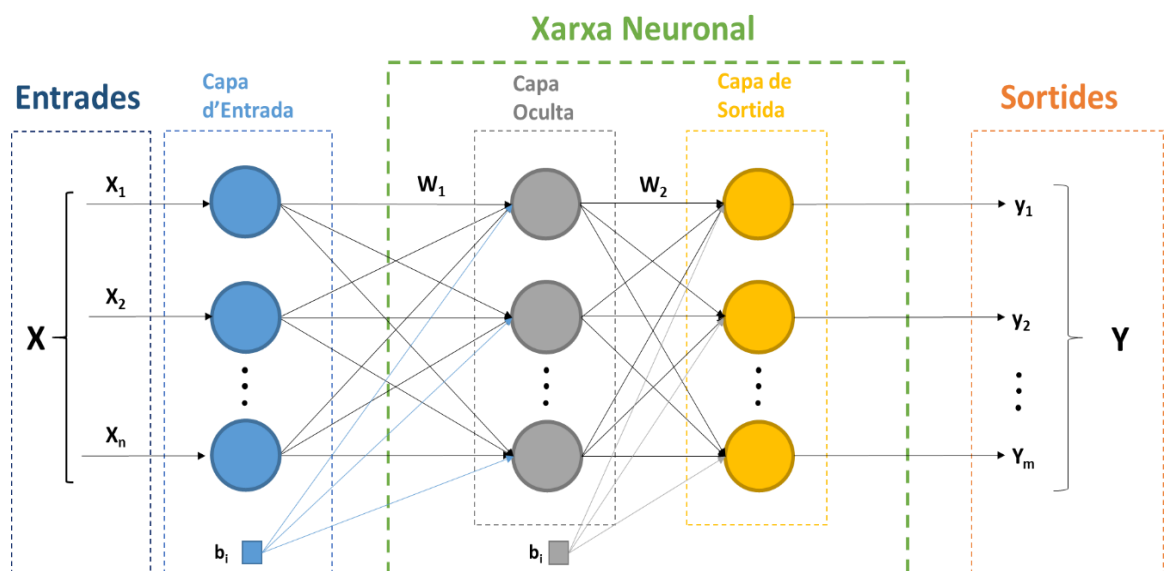


Figura 6. Esquema d'una xarxa neuronal artificial.

Donat un conjunt de dades d'entrenament d'entrada-sortida, la xarxa es entrenada de manera òptima mitjançant l'aplicació de la funció d'activació γ a la suma ponderada entre el vector d'entrada $\underline{x} = (x_1, x_2, \dots, x_n)^T$ i el de pesos $\underline{w} = (w_1, w_2, \dots, w_d)^T$, equivalent a les connexions sinàptiques en una neurona real, més un biaix (b_i), obtenint-se la següent expressió:

$$y = \gamma \cdot \left(\sum_{i=1}^n w_i \cdot x_i + b_i \right) \quad (8)$$

on γ és una funció no lineal. La funció proposada posseeix una sortida binària coneguda com la funció de tot o res que equival a la funció esglaó unitària donada per

$$\gamma(z) = U(z) = \begin{cases} 1 & z \geq 0 \\ 0 & z < 0 \end{cases} \quad (9)$$

Quan es consideren neurones amb respostes de processament gradual, llavors es poden utilitzar funcions d'activació de forma lineal $\gamma(z) = z$ o de forma sigmoïdal com la funció logística $\gamma(z) = 1/(1 + e^{-z})$ o la tangent hiperbòlica $\gamma(z) = \tanh(z)$.

La funció de l'entrenament de la xarxa neuronal és minimitzar la funció cost, que es pot definir com la suma dels errors quadràtics entre el valor predits per la xarxa i el valors objectiu introduïts per l'entrenament.

A pesar de que mostra capacitats amb gran potencial, l'ANN té algunes limitacions específiques com poden ser la gran quantitat de temps i esforç que s'ha d'invertir en la selecció de la configuració de la xarxa: número de capes ocultes, número de neurones per cada capa, el tipus de funció de transferència a fer servir, l'algoritme d'entrenament i inclús l'error/cost de la funció (T. Masters, 1993).

2.1.3. Support Vector Regression (SVR)

La *Support Vector Regression* (SVR) és un mètode que es basa en el càlcul d'una funció de regressió lineal en un espai de característiques d'alta dimensionalitat, on les dades d'entrada s'assignen a través d'una funció no lineal. La capacitat del sistema és controlada per paràmetres que no depenen de la dimensionalitat de l'espai de característiques (Basak et al. 2007). Recentment, la SVR s'està utilitzant en moltes aplicacions en el camp de l'enginyeria de processos químics mostrant prometedores capacitats (Jain et al. 2007).

Donat un conjunt de n dades d'entrenament d'entrada-sortida $[x_i, y_i]$, $i=1, 2, \dots, n$, $x \in R^k$, $y \in R$, SVR mapeja l'espai original de les dades d'entrada en un espai de característiques de dimensions superiors, generalment a través d'una funció kernel de tipus Gaussià no lineal

$\phi(x_i, x_j) = \exp(-\|x_i, x_j\|^2 / 2\sigma^2)$. En l'espai característic el problema comença en la identificació o determinació de la superfície lineal òptima $f(x) = \mu + w^T \Phi(x)$, la més plana, que s'ajusti a les dades a través de la minimització dels pesos del vector $w^2 \in R^k$, on μ una base o biaix. Per tal de garantir un major rendiment, SVR permet especificar els marges o un tub amb radi ε , dins del qual s'accepten o es toleren els errors en les dades de la mostra. En conseqüència, aquest problema d'optimització (minimització) es sotmet algunes limitacions: l'error en el valor predit de cada una de les dades de l'entrenament ha de ser com a màxim igual a valor de ε . En altres paraules, lo que es busca és la funció $f(x)$ con la màxima desviació d' ε per totes les dades del conjunt d'entrenament. Per permetre o tolerar valors atípics, les dades que tenen un error de predicció més gran que el valor de ε es penalitzen utilitzant lo que s'anomena funció de pèrdues ε -sensible. El problema pot ser expressat de la manera següent:

$$\text{Min } \frac{1}{2} \|w\|^2 + \frac{c}{n} \sum_{i=1}^n \xi_i + \xi_i^* \quad (10)$$

Sotmès a:

$$y_i - \mu - w x_i \leq \varepsilon + \xi_i \quad (11)$$

$$\mu + w x_i - y_i \leq \varepsilon + \xi_i^* \quad (12)$$

$$\xi_i, \xi_i^* \geq 0 \quad (13)$$

On ξ_i , ξ_i^* son variables de folgança, i C és un factor de penalització que controla l'equilibri entre la quantitat en la que les desviacions majors als marges (ε) son tolerades i el error de l'entrenament. En la Figura 7 és pot observar gràficament.

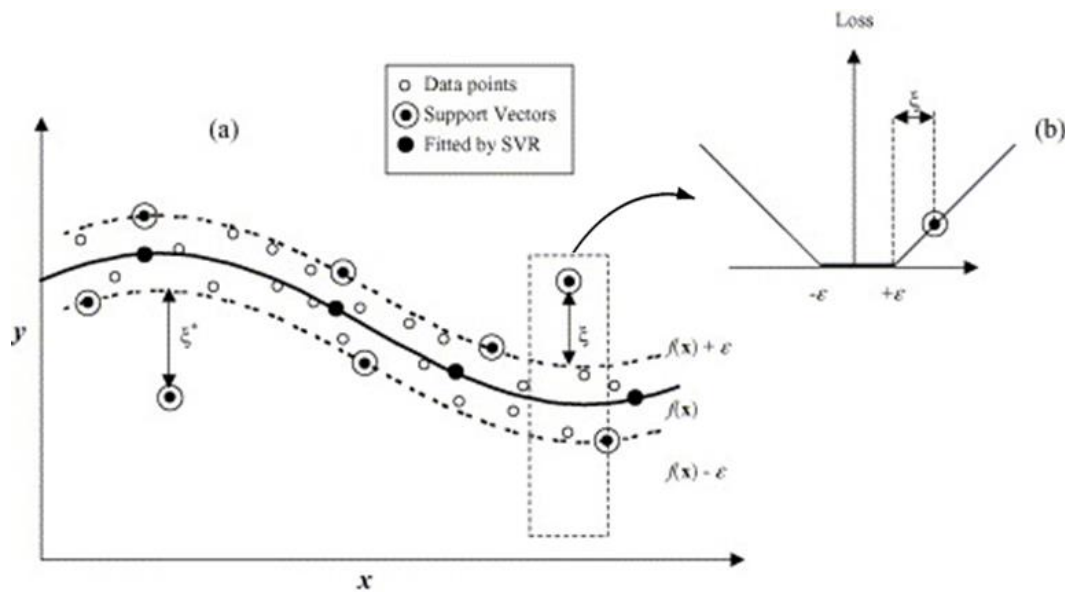


Figura 7. (a) Representació esquemàtica de la SVR utilitzant la funció de pèrdua ϵ -insensible, (b) la funció de pèrdua ϵ -insensible lineal en la que el paràmetre C determina la pendent. [Font: Desai et al. 2006]

Només els punts fora de la regió ombrejada contribueixen al cost en la mitja, ja que les desviacions es penalitzen de forma lineal. En la majoria dels casos, el problema d'optimització es poden reformular en forma de doble problema mitjançant la introducció de multiplicadors de Lagrange α_i, α_i^* per cada una de les restriccions. Els valors dels multiplicadors de Lagrange son determinats resolent el doble problema d'optimització utilitzant Programació Quadràtica (QP). Un cop els valors de α_i, α_i^* son determinats, el pesos òptims w de la funció original $f(x)$ es calcula μ como es mostra a continuació:

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \Phi(x_i) \quad (14)$$

$$\mu = \frac{1}{n} \sum_{i=1}^n y_i - \Phi(x_i)^T w \quad (15)$$

$$f(x_{new}) = \mu + \sum_{i=1}^n (\alpha_i - \alpha_i^*) \Phi(x_{new}, x) \quad (16)$$

Els vector d'entrenament amb els multiplicador de Lagrange amb valors diferent de zero s'anomenen com *support vectors* i representen els marges de les fronteres del tub (Jain et al., 2007).

A pesar del potencial que presenta aquest mètode, la SVR presenta certs desavantatges com poden ser que si el número de característiques és molt més gran que el número de mostres, el mètode es probable que convergeixi en un model amb males prediccions.

2.2. Tècniques per validació de models

Una forma comú per validar si un model basat en dades esta ajustat és utilitzant un conjunt diferent de dades d'entrada-sortida $[\mathbf{\theta}_v]_{n_v \times k}$ $[\mathbf{y}_v]_{n_v \times 1}$, llavors el metamodel s'utilitza per predir la sortida del conjunt de validació \hat{y}_i i aquesta és comparada amb el valor real de la sortida y_{vi} . A continuació, l'error quadràtic mitjà (RMSE), el percentatge de l'error relatiu (RPE) i el coeficient de correlació (CC) es calculen com a mesura de precisió de la predicció. L'RMSE i el RPE son mesures directes, ja que estan mesurant el valor mig de la desviació de la predicció a partir del valor real.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_1^{*i} - \hat{y}_1^{*i})^2} \quad (17)$$

$$CC = r_{y, \hat{y}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}; \bar{y} = \frac{\sum_{i=1}^n y_i}{n}; \bar{\hat{y}} = \frac{\sum_{i=1}^n \hat{y}_i}{n} \quad (18)$$

2.3. Programes per la modelització

Existeixen diferents programes informàtics, tant de software privat com de software lliure, especialment dissenyats per resoldre problemes en l'àmbit de l'enginyeria de processos i l'estadística. Abasten des de les necessitats elementals per l'obtenció d'estadística descriptiva i gràfiques per l'anàlisi, com mètodes avançats de modelització i anàlisi multivariant. En aquest apartat es parlarà de tres del programes més utilitzats del moment per la modelització de processos com son R project, Python i MATLAB.

2.3.1. R project

R és un llenguatge i entorn de programació amb codi obert per l'anàlisi estadístic i generació de gràfics. Es tracta de un software lliure que proporciona un amplia varietat de tècniques estadístiques (models lineals i no lineals, test estadístics, anàlisi de series temporals, algoritmes de classificació i agrupament, etc) i gràfiques. Es tracta d'un llenguatge de programació que permet als usuaris definir les seves pròpies funcions en el cas que es considerés necessari. Per algoritmes computacionals exigents és possible desenvolupar

biblioteques en C o C++ que es carreguen dinàmicament. A més, es poden ampliar les seves capacitats amb la possibilitat de carregar diferents paquets o biblioteques amb finalitats específiques ja siguin de càlcul o gràfiques. *R project* permet generar fàcilment gràfics incloent símbols i fórmules matemàtiques quan sigui necessari. També pot utilitzar-se com eina de càlcul numèric, camp en el que pot ser tan eficaç com altres eines específiques, i pot integrar-se a distintes bases de dades facilitant la utilització des de llenguatges de programació interpretats.

2.3.2. Python

Python és un llenguatge de programació d'alt nivell de codi obert que ha guanyat en popularitat a un ritme constant els últims anys. Desenvolupat ara fa uns deu anys per Guido van Rossum, va ser dissenyat per ser fàcil de llegir i simple de realitzar la seva implementació, un llenguatge orientat a l'ensenyament desenvolupat als anys 80. No obstant això, Python també va ser creat per solucionar problemes reals i pren una àmplia varietat de característiques de llenguatges de programació tals com C++, Java, Modula-3 i Schema, i es pot executar en diferents sistemes operatius com Mac, Windows i sistemes Unix. Degut a això, una de les característiques més remarcables de Python es la seva gran capacitat d'atreure desenvolupadors de programari professional, científics, investigadors, artistes i educadors.

2.3.3. MATLAB

MATLAB, abreviatura de laboratori de matrius (*MATrix LABoratory*, en anglès), és una eina de software matemàtic amb un llenguatge de alt nivell i entorn interactiu pel càlcul numèric, la visualització i la programació. El llenguatge, les eines i les funcions matemàtiques incorporades permeten explorar diversos enfocaments. Entre les seves prestacions bàsiques es troben la manipulació de matrius, la representació de dades i funcions, la implementació d'algoritmes, la creació de interfases d'usuari i la comunicació amb programes en altres llenguatges i altres dispositius hardware. A més, es poden ampliar les seves capacitats amb els diferents paquets existents anomenats *toolboxes*.

Mitjançant MATLAB, és possible descriure i analitzar dades; desenvolupar algoritmes i crear models utilitzant l'estadística i tècniques de *Machine Learning*. Es poden emprar l'estadística descriptiva i gràfics per l'anàlisi exploratori de dades, distribucions de probabilitat per l'ajust d'aquestes i dur a terme proves d'hipòtesis. Algoritmes de regressió i classificació permeten fer inferències a partir de les dades i construir models predictius.

3. Enfoc de modelat

El problema tractat en aquest projecte és un procés per lots sense entrades o pertorbacions externes al llarg de la síntesi o producció d'aquest. Aquest sistema està format per dos tipus de variables. Una d'elles, la variable que controla el progrés de la reacció $y(t)$, o variable *off-line*, és extremadament costosa de mesurar i l'anàlisi de les seves mostres es fa en períodes de temps relativament grans. L'altre tipus de variable $x(t)$, anomenada variable *on-line*, es registra de manera automàtica i continua, amb temps de mostreig molt petits, cosa que fa que el cost d'obtenció d'aquests valors sigui mínim. Per tant, el que es vol es obtenir el valor de la variable de mesura costosa en funció de les variables *on-line*, on les mesures son barates, per tal de poder estalviar en costos, temps de mostreig i anàlisi de les mostres. Això s'aconsegueix a través del desenvolupament d'un *soft sensor*. Aquests es poden fer servir tant per predir el progrés de la reacció en qualsevol moment al llarg de la síntesi del procés per lots como per vigilar si aquest s'està desenvolupant de manera adequada.

Per poder desenvolupar el *soft sensor* el sistema es caracteritza de la següent manera:

$$y(t) = f[x(t), y(0), x(0)] \quad (19)$$

El qual prediu les mesures actuals de la variable *off-line* ($y(t)$) en funció de les condicions inicials tant de les variables del sistema $y(0)$ com d'un nombre establert de variables *on-line* $x(0)$; i els valors actuals d'aquelles variables *on-line* que han estat escollides $x(t)$.

Per dur a terme aquest enfoc, la tècnica per modelitzar el comportament del procés que s'ha escollit és la Support Vector Regression (SVR). Aquesta tècnica presenta la solució per mitjà de la divisió del conjunt d'entrenament en petits subconjunts de punts, cosa que dona un enorme avantatge computacional respecte les altres tècniques presentades.

L'eina escollida per implementar-la ha estat R project. Avui en dia R project, és probablement un dels entorns més utilitzats per les universitats per a la investigació estadística, i per empreses, com Google, Bank of americà o Shell pel tractament de dades i generació de patrons. És una eina que s'utilitza amb facilitat sense haver de tenir coneixement elevats de programació informàtica, fet que facilita poder realitzar complexos anàlisi estadístics sense conèixer els secrets de la computació. A més és un software lliure. Això ens permet treballar en qualsevol ordinador i lloc de treball, no només dintre de la universitat perquè no es necessiten llicències.

3.1. Casos d'estudi

Per tal de poder il·lustrar l'enfoc plantejat es desenvolupen tres casos d'estudi. La complexitat dels quals va de menor a major.

3.1.1. Cinètica de primer ordre

Moltes reaccions químiques tenen lloc en una conjunt d'etapes en que els reactius es transformen en productes a través de múltiples etapes de reaccions elementals, o de primer ordre, seqüencials. En aquest cas d'estudi, es considera el següent esquema de reacció en sèrie, que es coneix com una de les reaccions més simples que es poden trobar en la enginyeria química:



En aquest esquema, la concentració del reactiu A decau per formar l'intermedi B, que a la seva vegada sofreix un decaïment subseqüent resultant la formació del producte C. Això implica una sèrie de reaccions de primer ordre, les quals les expressions de velocitat per cada una de les espècies es poden descriure seguint les següents equacions:

$$\frac{dC_A}{dt} = -k_1 \cdot C_A \quad (21)$$

$$\frac{dC_B}{dt} = k_1 \cdot C_A - k_2 \cdot C_B \quad (22)$$

$$\frac{dC_C}{dt} = k_2 \cdot C_B \quad (23)$$

En la Figura 8, es pot observar l'evolució de les concentració dels components a través del temps en la síntesi del procés per lots.

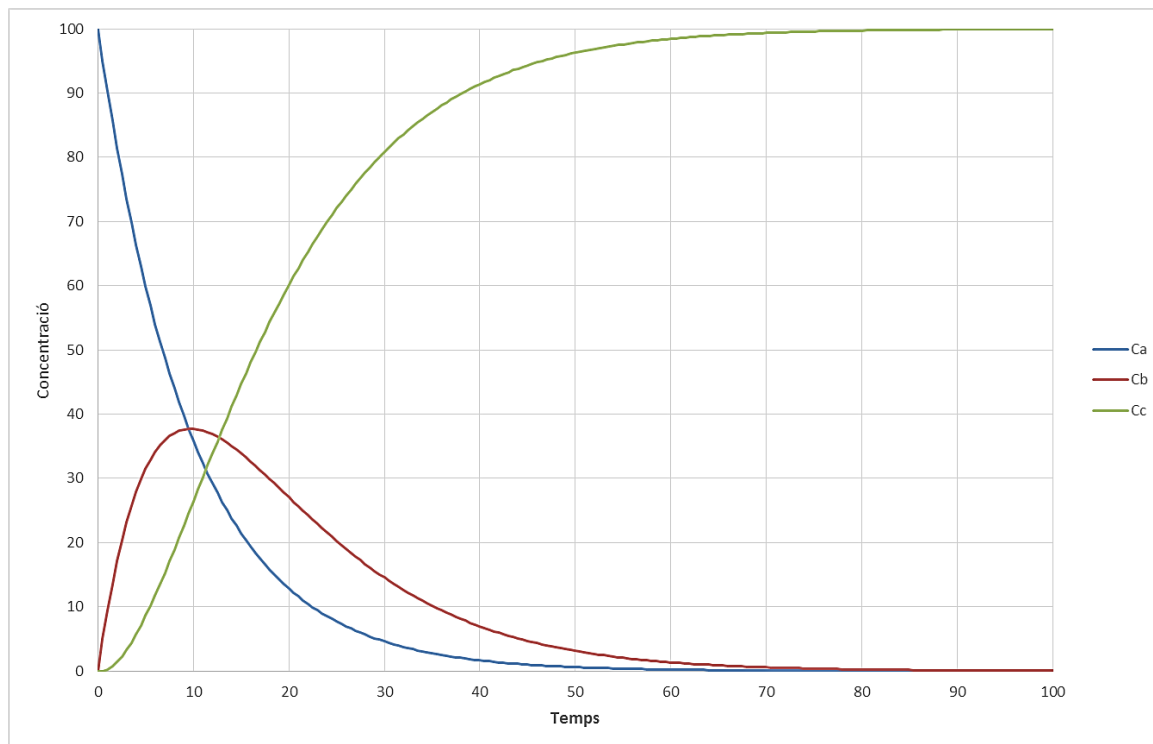


Figura 8. Representació del progrés de la reacció de primer ordre seqüencial.

Per poder realitzar l'enfoc plantejat, es considera que la concentració C_C del producte C és la variable *off-line*, aquella on les mesures i l'anàlisi són costosos. Mentre que les concentracions C_A i C_B del reactiu i producte, A i B, són les variables que es mesuren i registren de manera automàtica pels sensors a cada segon del procés.

El temps total emprat per sintetitzar el procés per lots és de 30 minuts sense tenir en compte perturbacions en el sistema. Per tant, durant l'execució del procés, es calculen 8 punts de mostreig per la concentració de C_C , a temps específics, que simulen la variable *off-line* als quals se'ls hi afegeix una quantitat elevada de soroll $N(\mu=0, \sigma=0,3)$ per imitar els errors experimentals que es poden cometre al realitzar la mesura. Pel contrari, les concentracions de C_A i C_B es calculen cada segon emulant l'enregistrament automàtic de les variables *on-line*, al que se les afegeix certa quantitat de soroll $N(\mu=0, \sigma=0,03)$.

Per poder generar el model, es simulen 124 processos per lots amb valors de concentracions inicials diferents $[C_A(0), C_B(0), C_C(0)]$ seguint un procediment de mostreig Hammersley dins dels límits $[14:20, 0:2, 0:2]$, i la forma anteriorment descrita.

3.1.2. Cabrera Reina

Aquest cas d'estudi es basa en el model dinàmic i semi-empíric desenvolupat per Cabrera Reina per simular la reacció de degradació de contaminants utilitzant com a tractament el procés foto-Fenton (Cabrera Reina et al.,2012).

Aquest model es capaç de simular la degradació del contraminant a través de la mesura de carboni orgànic total que queda en el sistema (TOC), l'evolució del peròxid d'hidrogen i del oxigen emprant reaccions simplificades del procés foto-Fenton.

El model proposat per Cabrera Reina assumeix 9 processos i 8 estats: dos espècies de salt ferrosa (Fe^{+2} i Fe^{+3}), peròxid d'hidrogen (H_2O_2), radicals que es formen a partir del peròxid d'hidrogen (R), oxigen dissolt, i tres estats per la matèria orgànica, dos parcialment oxidats (MX_1 , MX_2) i un estat aquesta no ha estat degradada (M)

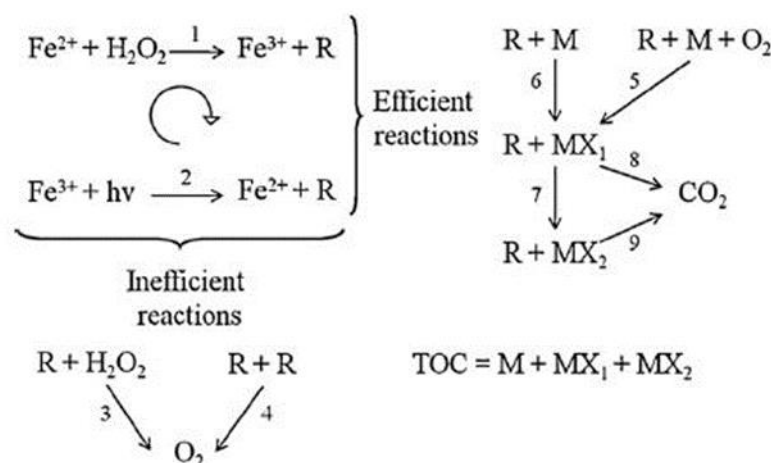


Figura 9. Esquema de les reaccions proposat per Cabrera Reina. [Font: Cabrera Reina et al.,2012]

Com s'observa en la **Figura 9. Esquema de les reaccions proposat per Cabrera Reina. [Font: Cabrera Reina et al.,2012]** Figura 9, Cabrera Reina proposa un cicle de reaccions que defineixen la reacció foto-Fenton. Primer, es creen uns radicals lliures (R) que degraden la matèria orgànica (M). La degradació de M esta formada per dos passos responen a l'observació que es fa de les dades experimentals, pel que MX_1 és un producte intermedi del qual no es coneix la seva composició i MX_2 és l'etapa final pel producte orgànic degradat. S'ha de destacar l'aparició de reaccions ineficients que relentitzen el procés. Aquestes reaccions han de tenir-se en compte degut a que la reacció 4 és de segon ordre, pel que la producció d'oxigen pot produir-se molt ràpidament.

Per tant aquest model es basa en el balanç de massa de les espècies presents a l'esquema per un procés per lots. El principi de la conservació es pot escriure de la següent manera:

$$\left\{ \begin{array}{c} \text{Acumulació} \\ \text{de masa component } j \end{array} \right\} = \left\{ \begin{array}{c} \text{Rati de masa} \\ \text{de } j \text{ entrada} \end{array} \right\} - \left\{ \begin{array}{c} \text{Rati de masa} \\ \text{de } j \text{ sortida} \end{array} \right\} + \left\{ \begin{array}{c} \text{Rati de masa} \\ \text{de } j \text{ produïda} \end{array} \right\} \quad (24)$$

Com el sistema es tracta de un procés per lots no existeixen ni entrades ni sortides en el sistema, per lo tant l'equació 24 es pot simplificar de la següent manera

$$\left\{ \begin{array}{c} \text{Acumulació} \\ \text{de masa component } j \end{array} \right\} = \left\{ \begin{array}{c} \text{Rati de masa} \\ \text{de } j \text{ produïda} \end{array} \right\} \quad (25)$$

$$\frac{dn_j}{dt} = \left(\sum_{i=1}^{NB} v_j \cdot R_j \right) \cdot V \quad (26)$$

On R_j és el rati de producció del component j per cada reacció i ; V és el volum, nb és el nombre total de reaccions on j es produït, n_j és el número de mols i v_j són els coeficients estequiomètrics del component j en la reacció.

La densitat es manté constant al llarg del procés, considerant que $n_j = C_j \cdot V$, el balanç de massa és pot escriure en termes de concentració:

$$\frac{dC_j}{dt} = \sum_{i=1}^{NB} v_j \cdot R_j \quad (27)$$

Tots les reaccions del procés són considerades com reaccions elementals o simples, ja que son produïdes en un sol pas, per tant R_j pot ser expressada de la següent manera:

$$R_j = k_j \cdot \prod_{i=1}^{NR} [C_i]^{\text{Ordre reacció del reactiu } i \text{ en la reacció } j} \quad (28)$$

On k_j és la constant cinètica de la reacció j i NR és el nombre total de reactius.

Aquest balanç de massa és eficient per tots els components menys per l'oxigen dissolt. Per fer el seu balanç de matèria s'ha de tenir en compte la llei de Henry per descriure la transferència de l'oxigen a l'aire on la seva concentració és més elevada que la concentració màxima d'oxigen dissolt. Per tant el balanç de massa de l'oxigen queda de la següent forma:

$$\frac{d[O_2]}{dt} = \sum_{i=1}^{NB} (v_j \cdot R_j) + K_L \cdot a \cdot (O_2^* - [O_2]) \quad (29)$$

On K_L representa la constant de Henry, a és l'àrea de la bombolla i O_2^* representa la saturació d'oxigen respecte l'aire. En aquest cas l'àrea de la bombolla es considera constant, per tant una nova constant es obtinguda K_{La} .

Les expressions cinètiques per aquest model de reacció són:

$$r_1 = k_1 \cdot [Fe^{+2}] \cdot [H_2O_2] \quad (30)$$

$$r_2 = k_2 \cdot [Fe^{+3}] \cdot [I] \quad (31)$$

$$r_3 = k_3 \cdot [R] \cdot [H_2O_2] \quad (32)$$

$$r_4 = k_4 \cdot [R] \cdot [R] \quad (33)$$

$$r_5 = k_5 \cdot [M] \cdot [R] \cdot [O_2] \quad (34)$$

$$r_6 = k_6 \cdot [M] \cdot [R] \quad (35)$$

$$r_7 = k_7 \cdot [MX1] \cdot [R] \quad (36)$$

$$r_8 = k_8 \cdot [MX1] \cdot [R] \quad (37)$$

$$r_9 = k_9 \cdot [MX2] \cdot [R] \quad (38)$$

I els balanços de massa dels components del model són:

$$\frac{d[Fe^{+2}]}{dt} = -r_1 + r_2 \quad (39)$$

$$\frac{d[Fe^{+3}]}{dt} = r_1 - r_2 \quad (40)$$

$$\frac{d[H_2O_2]}{dt} = -r_1 - r_3 \quad (41)$$

$$\frac{d[R]}{dt} = r_1 + r_2 - r_3 - 2 \cdot r_4 - r_5 - r_6 - r_7 - r_8 - r_9 \quad (42)$$

$$\frac{d[M]}{dt} = -r_5 - r_6 \quad (43)$$

$$\frac{d[MX1]}{dt} = r_5 + r_6 - r_7 - r_8 \quad (44)$$

$$\frac{d[MX2]}{dt} = r_7 - r_9 \quad (45)$$

$$\frac{d[O_2]}{dt} = g_1 \cdot r_3 + g_2 \cdot r_4 - c_1 \cdot r_5 + K_{La} \cdot (O_2^* - O_2) \quad (46)$$

On g_1 , g_2 i c_1 són els coeficients estequiomètrics relacionats amb el balanç de l'oxigen.

Per poder realitzar l'enfoc plantejat, es considera com variable *off-line* la degradació de la matèria orgànica (TOC), és a dir, aquella on les mesures i l'anàlisi són costosos, i com a variable *on-line*, la mesura de oxigen dissolt.

Per poder generar el model, es simulen 22 processos per lots amb valors de concentracions inicials diferents [TOC(0), H₂O₂(0)] dins dels límits [9.33:7.33, 35.3:45.3], utilitzant la simulació desenvolupada per Gerard Campanyà basant-se en el model descrit (Campanyà, 2013). L'interval de temps entre dades serà de 15 minuts, per tal de simular el sistema com si fos una planta real.

3.1.3. Aplicació d'un cas real: Planta Pilot

Aquest cas d'estudi tracta sobre una planta pilot fotoquímica que treballa en modo per lots, en la qual es dur a terme un procés de tractament d'aigua basat en la reacció foto-Fenton on el paracetamol és el contaminant a tractar.

3.1.3.1. Material i mètode

El paracetamol (acetaminofeno o 4-amidofenol, PCT des d'aquest moment) és un analgèsic, antiinflamatori i antipirètic, considerat com l'analgèsic no opioide més popular venut en Espanya en els últims anys (Martínez Bueno et al., 2012) i el contaminant més investigat en el processos d'oxidació avançada.

Els experiments es duen a terme utilitzant PCT al 98% de puresa de marca Aldrich per preparar les mostres en aigua destil·lada. El reactiu Fenton utilitzat son H₂O₂ 33% w/v (Panreac) y Fe₂SO₄·7H₂O (Merck). El rendiment del procés, és a dir, el progrés de la reacció s'avalua mitjançant l'extracció d'aliquotes de l'experiment que es dur a terme en la planta pilot i mesurant la concentració de carboni orgànic total (TOC) utilitzant l'analitzador Shimadzu V_{CHS/CSN} TOC.

3.1.3.2. Planta Pilot

La planta pilot fotoquímica és una planta de procés a escala reduïda o preindustrial, pel tractament d'aigües residuals mitjançant tractament d'oxidació avançada. Aquestes aigües poden procedir de qualsevol indústria la qual generi una gran carga orgànica, contaminants tòxics i/o recalcitrants, els quals son resistents als tractaments convencionals d'aigües residuals.



Figura 10. Planta pilot fotoquímica on es duen a terme els experiments.

La planta pilot (Figura 11) esta formada per un dipòsit de vidre, un foto-reactor tubular equipat amb una làmpada d'ultraviolat (Philips actínica BL TL-DK 36W / 10 1SL) i un sistema de bombeig que permet mantenir la recirculació del sistema constant. El volum total que es tracta en el procés per lots és e 15L, bombejant a $12\text{L}\cdot\text{min}^{-1}$ per garantir que la mescla perfecte.

Les condicions de procés dels experiments són iguals per cada un d'ells. El temps total de durada de l'experiment és de 90 minuts i la irradiació té lloc durant tot aquest temps. El pH inicial és fixat a un valor de 3 ± 0.1 mitjançant l'addició d'àcid clorhídric al 37% (J. T. Baker), valor que és considerat òptim pel procés Fenton (Pignatello et al., 2006). La salt de ferro (II) s'introdueix a una concentració de 10 mg/L, mentre que les concentracions inicials de H_2O_2 varien en un rang de 99 a 791 mg / L.

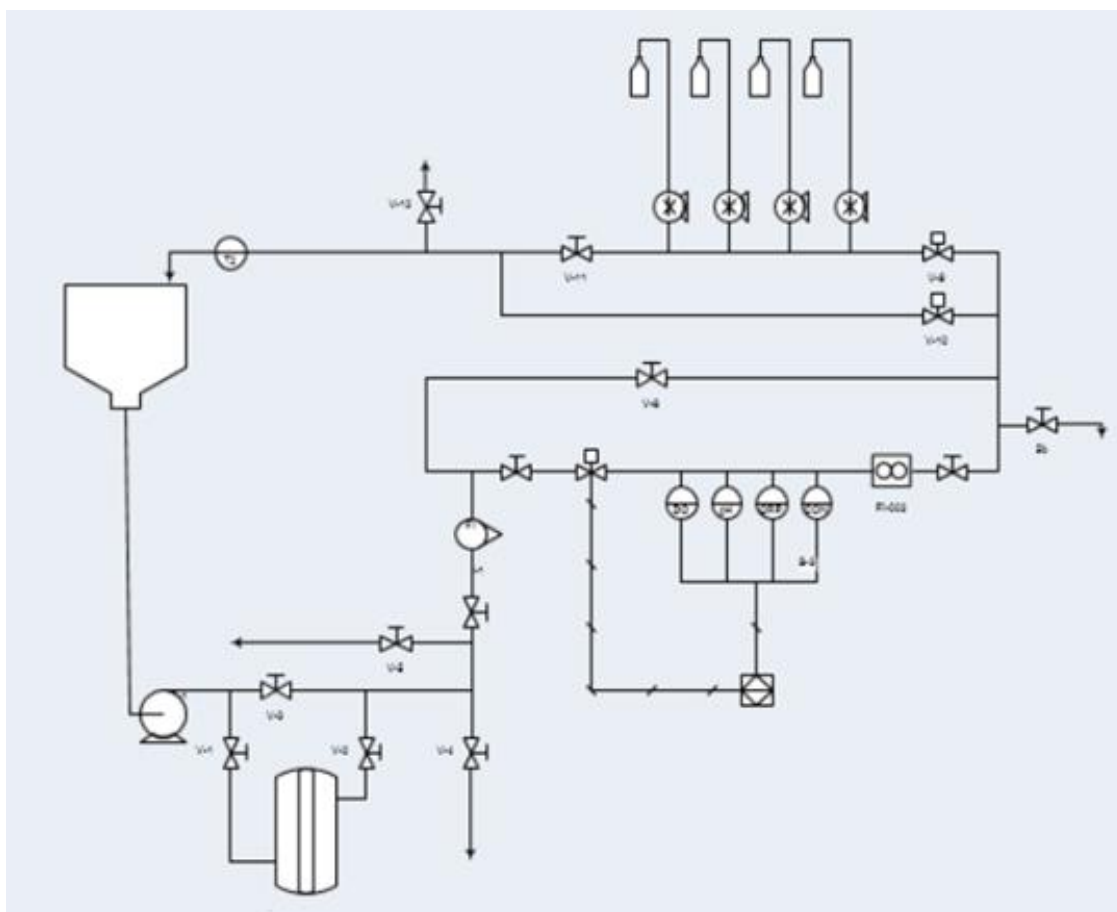


Figura 11. Diagrama de la planta pilot de tractament foto-Fenton.

Les mostres procedents de la mescla de la reacció s'extreuen en intervals de temps regulars de 15 minuts durant el transcurs del procés per lots de 90 minuts. Aquestes són portades a l'analitzador TOC on es mesurarà el progrés de la reacció, tenint un total de 7 mesures per lot realitzat. Per lo tant, aquesta mesura serà la que representarà la variable *off-line*. Per altre banda, el sistema SCADA integrat a la planta permet registrar de manera senzilla i automàtica els valors de temperatura i potencial redox, necessaris per determinar el bon funcionament de la reacció que es du a terme en la planta. Cada una de les dues variables es mesuren de forma *on-line* a cada segon de l'experiment proporcionant un total de 5400 mesures.

Pel desenvolupament del model es seleccionen 12 lots amb diferents concentracions inicials de contaminant, o TOC, i de H_2O_2 , els valors dels quals es mostren per cada un dels experiments en la Taula 1.

Taula 1. Conjunt d'experiments seleccionats per la generació del model.

Lot	$[TOC]_0$ (mg/L)	$[H_2O_2]_0$ (mg/L)
1	21	99
2	21	99
3	21	198
4	42	99
5	42	198
6	42	198
7	42	395

8	42	395
9	84	395
10	84	395
11	84	791
12	84	791

En la taula anterior es pot veure com algunes de les concentracions, i per tant els experiments, estan duplicades. Això es va fer pensant en la modelització posterior, ja que aquestes repeticions permetrien minimitzar els efectes del soroll provocat pels errors humans a l'hora de l'entrenament, i avaluar la sensibilitat del metamodel en el conjunt de validació.

4. Resultats i Discussió

Aquest apartat es mostraran els resultats obtinguts en el desenvolupament del *soft sensor* per cada un dels casos d'estudi plantejats, així com s'explicarà com s'han gestionat les dades per generar-los.

4.1. Cinètica Primer Ordre

Per poder generar el model que simuli la cinètica de la concentració del producte C, es divideixen els 124 lots generats en dos subgrup. El primer subgrup, o d'entrenament, està format per 24 lots, mentre que el segon grup, o de validació del model, està format per els 100 lots restants.

Un pas important abans de l'aplicació de l'enfoc de modelat, per la generació del *soft sensor*, és el suavitzat de les dades *on-line* dels lots (C_A , C_B) utilitzant una tècnica de mesura mòbil amb un marc de temps de 50 segons, amb la finalitat de disminuir l'efecte del soroll afegit. Després d'això, els 8 punts d'entrenament entrada-sortida, és a dir, *on-line-off-line*; es recullen per cada un dels 24 lots que representen la matriu d'entrenament.

Un cop es tenen les dades processades, i per tal d'obtenir el metamodel, es construeix una matriu amb els valors necessaris pel seu entrenament per cada un dels lots que el formen. Aquesta matriu consta de 5 variables d'entrada: les condicions inicials de les substàncies A, B i C (aquest valor es mantindrà constant al llarg de tota la matriu) i les variables *on-line* que representen el progrés de les substàncies A i B al llarg del temps d'experiment.; i una variable de sortida, que representa l'evolució del producte C al llarg del temps. Els valors de la variable de sortida són introduïts en el model com valors objectius, o *targets*, per tal de que aquest pugui trobar la correlació entre les variables d'entrada donades i el seu valor de sortida (Equació 47).

$$C_c(t) = f[c_A(0), c_B(0), c_c(0), c_A(t), c_B(t)] \quad (47)$$

Per tant el metamodel es entrena utilitzant una única matriu formada pels 24 lots, que dona un total de 192 punts d'entrenament entrada-sortida. Després d'això, el model obtingut es utilitza per predir el valor del progrés de la substància C (variable de sortida) dels 24 lots utilitzats per entrenar. Aquest pas donarà una primera aproximació que permetrà valorar si el model construït representa el procés que es vol modelar. En la Figura 12, es poden observar els valors de C predits per algun dels lots entrenats. La resta de lots queden recollits en l'annex 1.

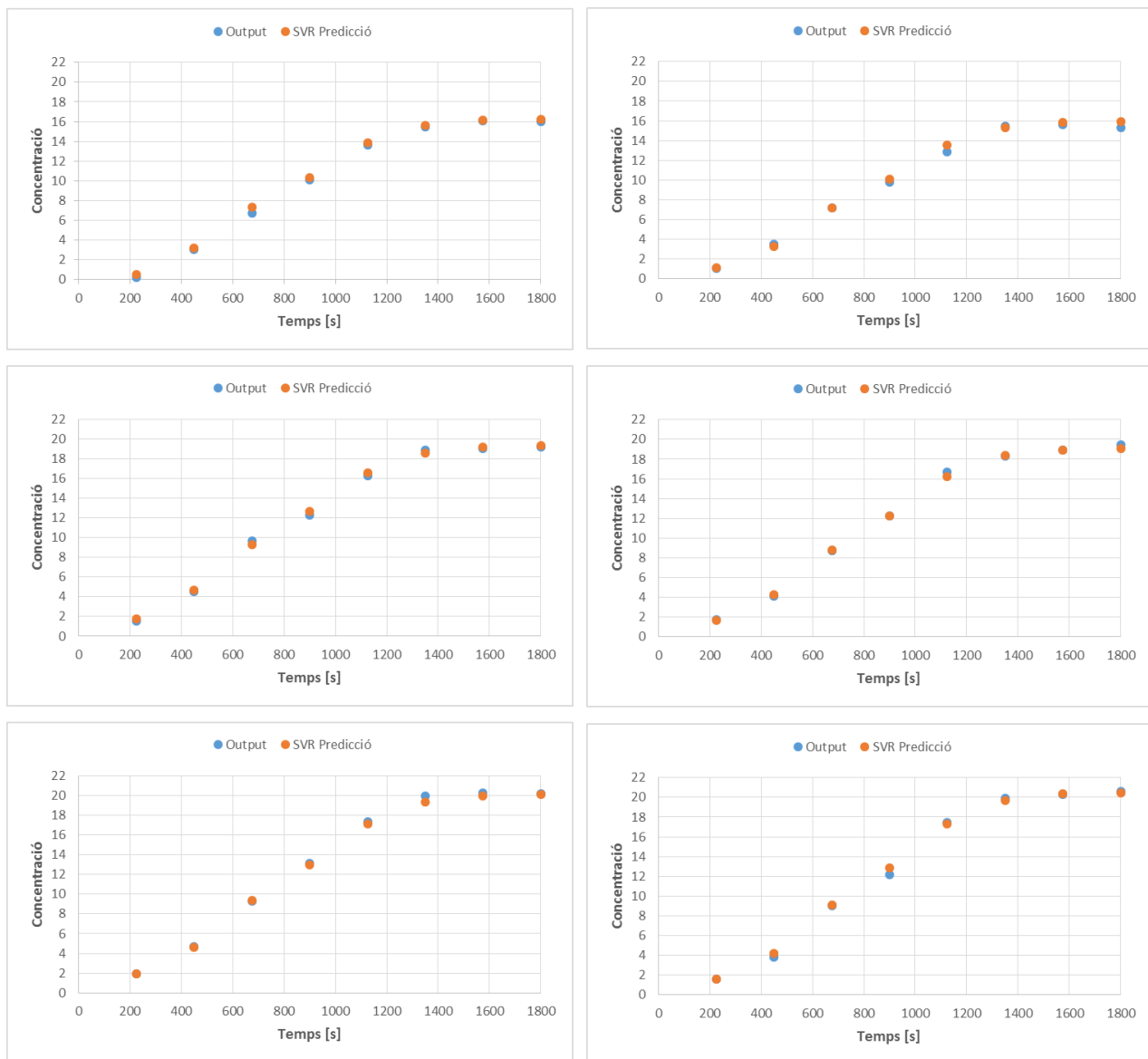


Figura 12. Predicció d'una representació dels 24 lots utilitzats per l'entrenament.

A la figura anterior, es pot observar com els valors de les prediccions de les concentracions inicials de C per tots els lots, valors situats entre 0 i 2, són més exactes que els estimats per la resta dels set punts durant el mateix. Aquest fet es deu a que la informació que s'aporta sobre la concentració inicial del sistema al metamodel es molt més gran que la que s'aporta sobre l'evolució del procés. Per cada conjunt de 8 punts recopilats per lot, la informació sobre les concentracions inicials existeix en cada un dels punts.

Tot i així, encara que la resta de punts predits presentin certa exactitud respecte al valor real i que el model obtingut mantingui la tendència del sistema pel qual es va entrenar, no es pot donar per vàlid fins que no es validi amb un conjunt de dades que no pertanyin al conjunt amb el qual s'ha entrenat. Per això, s'utilitzen el 100 lots del subconjunt de validació anteriorment separats. A l'igual que succeïa en el cas de l'entrenament, cada un dels lots han de ser manipulats per tal de obtenir una matriu de les mateixes característiques que la matriu de variables d'entrada amb la qual s'ha entrenat. Per tant la matriu que s'introdueix en la validació

està formada per les 5 variables d'entrada amb un valor per cada segon del temps total.

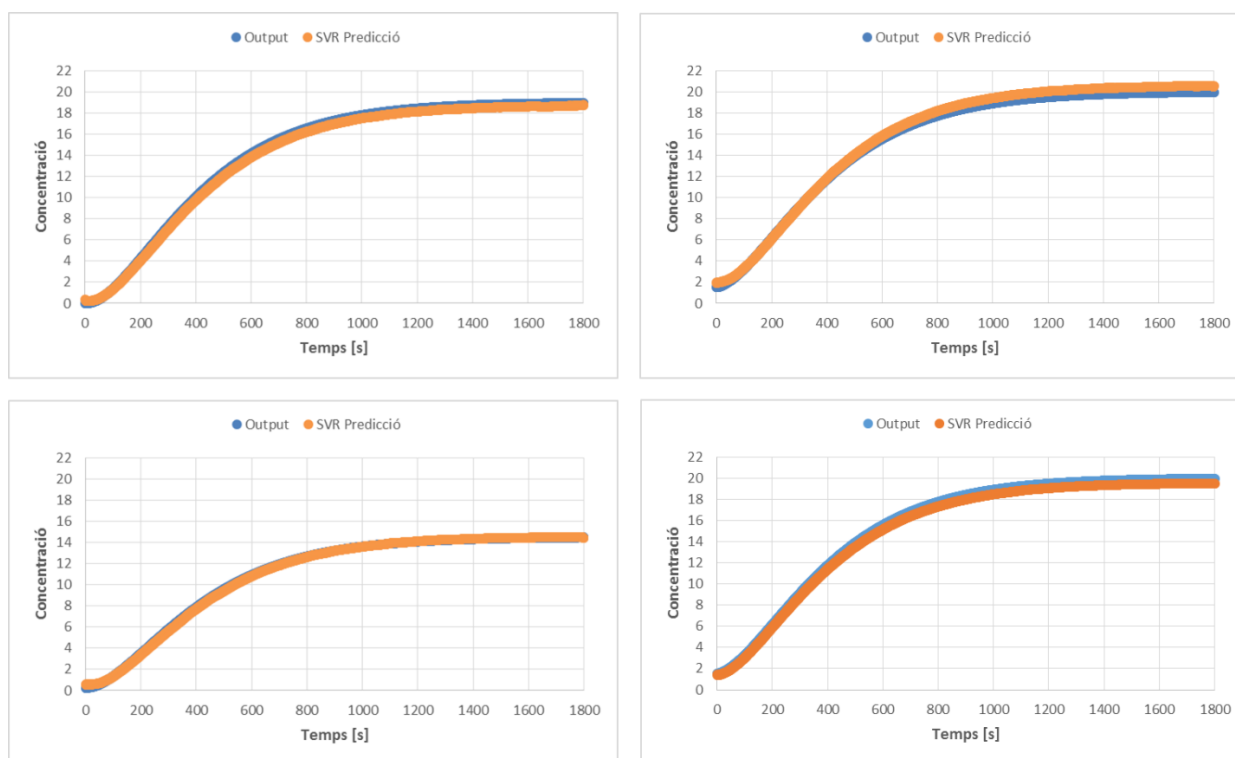


Figura 13. Predicció de C per a 4 lots de validació.

En la Figura 13, es demostra que el metamodel és capaç de predir de forma continua el progrés del producte C al llarg de la síntesi del lots, com si fos una variable *on-line*, amb una precisió molt alta, capturant així el comportament real del procés d'estudi sense tenir sobreajust en els valors d'entrenament.

Taula 2. RMSE, RPE i CC per les prediccions obtingudes.

	RMSE			RPE [%]			CC		
	Entren.	Valid.	Tot	Entren.	Valid.	Tot	Entren.	Valid.	Tot
Dades Mesurades (8 punts)	0.371	0.436	0.424	21.34	12.83	14.48	0.9985	0.9980	0.9981
Comportament Sistema Real	0.231	0.355	0.335	2.575	4.049	3.764	0.9993	0.9982	0.9984

En la Taula 2 es pot observar com la SVR és capaç de predir la concentració del producte C amb una precisió molt alta i una alta correlació ja que en tots el casos el valor es superior a 0.99. Es pot comprovar com les mesures de precisió calculades respecte al sistema real son relativament millors que les obtingudes respecte les dades experimentals mesurades. Això demostra que el *soft sensor* desenvolupats son capaços d'identificar el comportament real del procés, malgrat que ha sigut entrenat amb dades amb soroll per representar lo que seria l'error experimental.

4.2. Cabrera Reina

Per poder generar el model que simuli el procés foto-Fenton amb dades obtingudes de la simulació feta amb el model proposat per Cabrera Reina, es divideixen els 22 lots generats en dos subgrup. El primer subgrup, o d'entrenament, està format per 13 lots o experiments, mentre que el segon grup, o de validació del model, està format per els 9 experiments restants.

Tal i com succeïa en l'apartat anterior, en el de la cinètica de primer ordre, el *soft sensor* està dissenyat per modelar el valor de la variable *off-line*, que en aquest cas és tracta del valor de la degradació del contaminant quantificat a partir de la quantitat de Carboni Orgànic Total (TOC). Aquesta modelització es portarà a terme a partir de correlacionar el valor de la variable de sortida (TOC) amb el valor de les concentració inicials de TOC, peròxid d'hidrogen i oxigen dissolt; i valors mesurats en línia, que descriuen l'evolució del sistema, per l'oxigen dissolt (Equació 48)

$$TOC(t) = f[TOC(0), H_2O_2(0), DO(0), DO(t)] \quad (48)$$

El metamodel es entrenat utilitzant una única matriu formada pels 13 experiment, que dona un total de 117 punts d'entrenament entrada-sortida. Després d'això, el model obtingut es utilitzat per predir el valors de la degradació del contaminant (TOC) dels 13 experiments utilitzats per entrenar. En la Figura 14, es poden observar els valors de TOC predits per alguns dels experiments entrenats. La resta d'experiments queden recollits en l'annex.

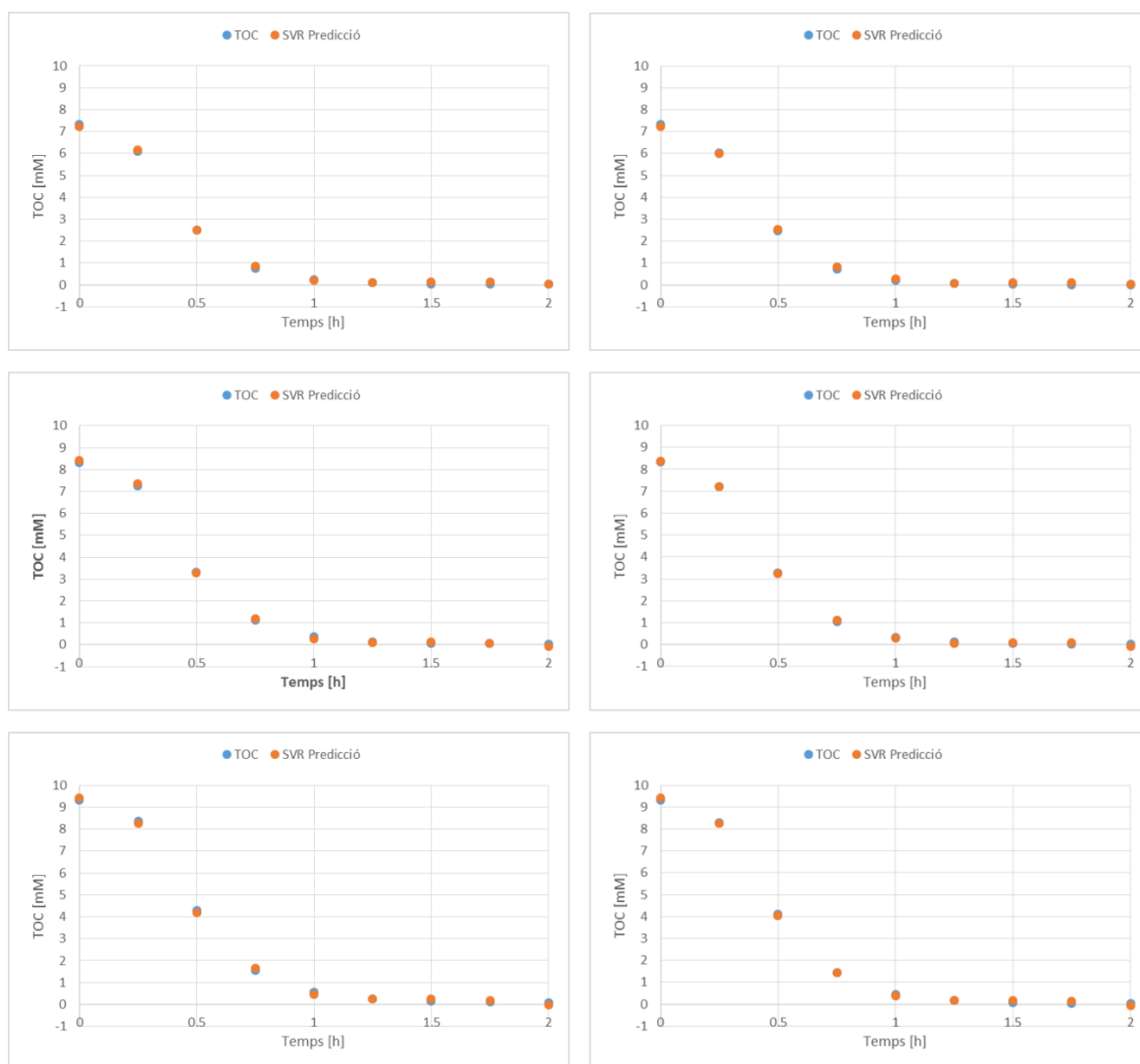


Figura 14. Predicció d'una representació dels 13 lots utilitzats per l'entrenament.

A la figura anterior, es pot observar com els valors de les prediccions de les concentracions de la degradació del contaminant, pels 9 punts d'entrenament per cada experiment, són molt acurades sent l'error comés entre el valor simulat, introduït com valor objectiu, i el valor de la predicció de un 0.4 %.

Tot i l'obtenció de resultats tant bons, no es pot donar per vàlid el metamodel fins que no es validi amb un conjunt de dades que no pertanyin al conjunt amb el qual s'ha entrenat, tal i com passava amb l'exemple anterior. Per la seva validació, s'utilitzen els 9 experiments del subconjunt de validació anteriorment separats.

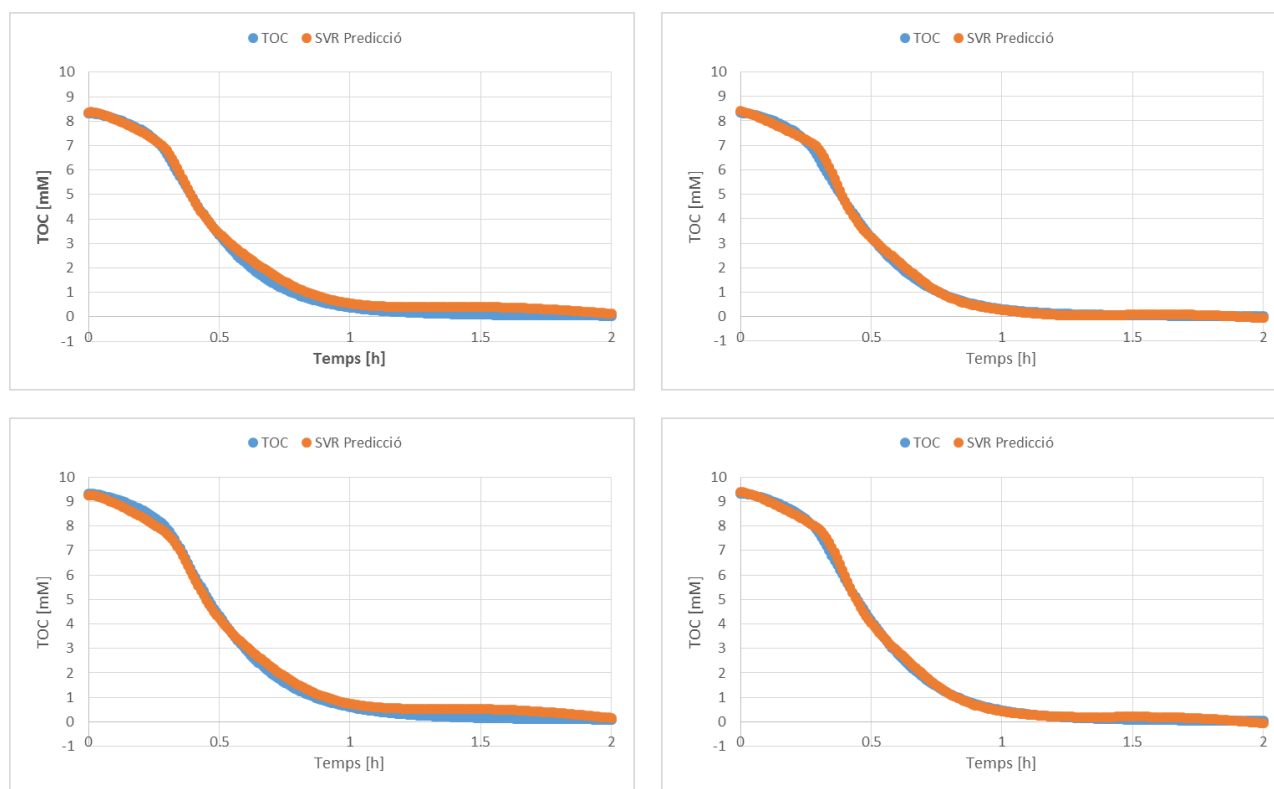


Figura 15. Predicció de la degradació del contaminant per a 4 experiments de la validació.

En la Figura 15, es pot observar com el metamodel generat és capaç de predir la tendència del sistema amb una precisió acceptable, tot i no ser la precisió mostrada en l'entrenament. Es pot observar però com en tots els experiments, als primers 20 minuts aproximadament de predicció, presenten la mateixa tendència rectilínia i no la curvatura que descriu el sistema. Tal i com s'ha observat el sistema es capaç de predir amb una molt bona precisió els punts del sistema introduïts mentre que aquelles zones on no té dades la SVR fa una interpretació del comportament reduint al mínim possible l'error però sense tenir en compte la dinàmica del sistema. Això fa pensar que possiblement el metamodel necessiti més informació per predir la dinàmica del procés, com pot ser introduir les derivades del sistema a la matriu d'entrenament.

4.3. Planta Pilot

Aquest cas d'estudi pretén generar un soft sensor, a partir d'un cas real. Aquest simularà el comportament de la degradació del contaminant utilitzant la tècnica de tractament d'aigües foto-Fenton en una planta pilot. Les dades aportades són 12 experiments que es divideixen en dos subgrups. El primer subgrup, o d'entrenament, està format per 7 dels experiments, mentre que el segon grup, o de validació del model, està format per 5 dels experiments restants. En la Taula 3, es pot observar quin dels experiments s'han destinat a cada un d'aquests.

Taula 3. Divisió del conjunt d'experiment per l'entrenament i la validació.

Lot	[TOC] ₀ [mg/L]	[H ₂ O ₂] ₀ [mg/L]	Entrenament	Validació
1	21	99		•
2	21	99	•	
3	21	198		•
4	42	99		•
5	42	198	•	
6	42	198	•	
7	42	395	•	
8	42	395	•	
9	84	395		•
10	84	395		•
11	84	791	•	
12	84	791	•	

Amb la finalitat de reduir el soroll del sensor, les dades en línia de temperatura i redox es suavitzen utilitzant la tècnica de mitja mòbil amb un marc de temps de 60 segons. A continuació, es recullen 7 punts entrada-sortida per cada un dels experiments que formaran part de la matriu d'entrenament.

Tal i com succeïa en l'apartat anterior, en el de dades perfectes de Cabrera Reina, la variable *off-line*, és el valor de la degradació del contaminant quantificat a partir de la quantitat de Carboni Orgànic Total (TOC) present a la mostra mesurada per l'analitzador. Aquesta modelització es portarà a terme a partir de correlacionar el valor de la variable de sortida (TOC) amb el valor de les concentració inicials de temperatura, redox i TOC; i valors mesurats en línia, que descriuen l'evolució del sistema, per la temperatura i el redox (Equació 49)

$$TOC(t) = f[TOC(0), T(0), Redox(0), T(t), Redox(t)] \quad (49)$$

Per tant el metamodel es entrenat utilitzant una única matriu formada pels 7 experiment, que dona un total de 49 punts d'entrenament entrada-sortida. Després d'això, el model obtingut es utilitzat per predir el valors de la degradació del contaminant (TOC) dels 7 experiments utilitzats per entrenar. En la Figura 16, es poden observar els valors de TOC predits per alguns dels experiments entrenats. La resta d'experiments queden recollits en l'annex.

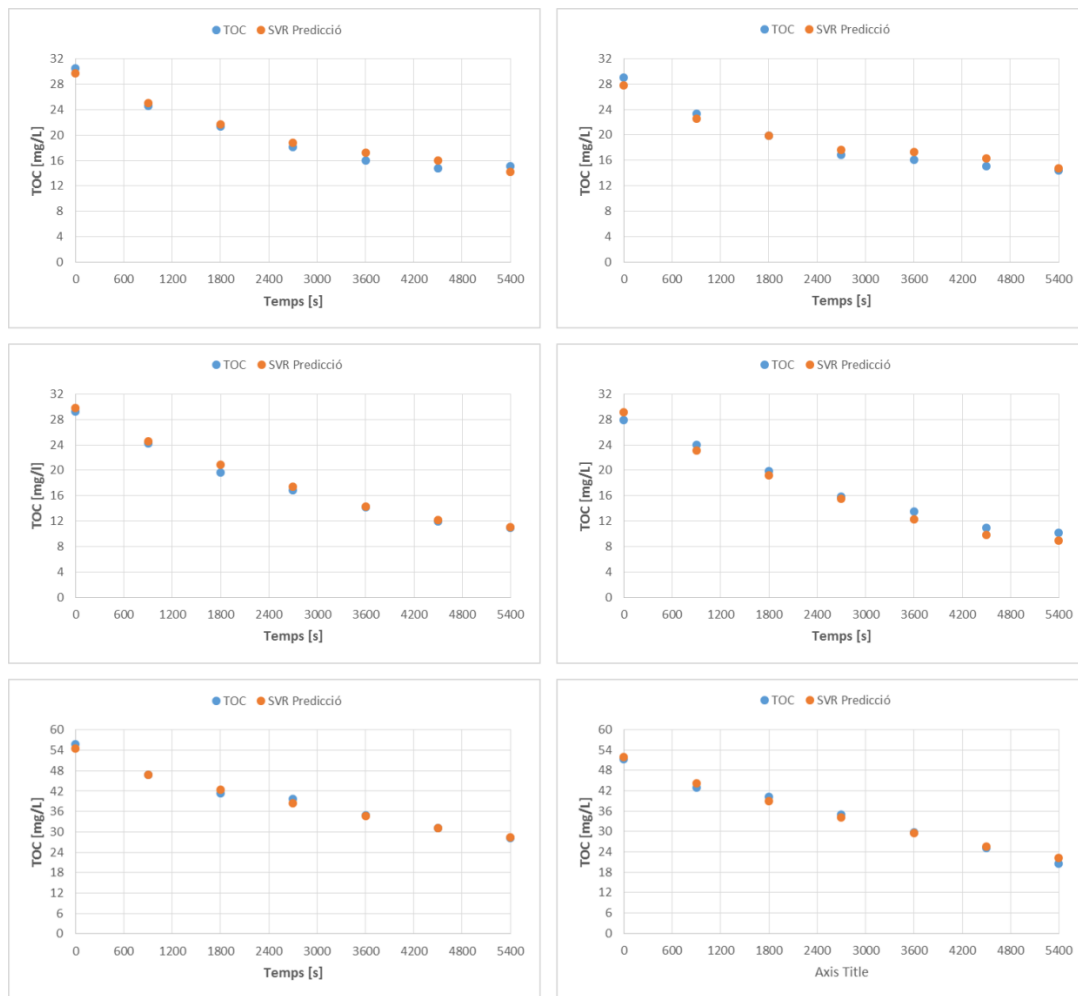


Figura 16. Predicció d'una representació dels experiments utilitzats per l'entrenament.

Encara que el model obtingut mantingui la tendència del sistema pel qual es va entrenar, no es pot donar per vàlid fins que no es validi amb un conjunt de dades que no pertanyin al conjunt amb el qual s'ha entrenat, a l'igual que s'ha fet amb els casos.

El metamodel es valida pels 5 experiments del subconjunt de validació anteriorment separats. La predicció que s'obté es a partir de validar el model amb el conjunt de dades pertanyent a la totalitat del experiment i no només amb els 7 punts per experiment amb els quals s'estava validant. Per tant, el *soft sensor* desenvolupat és capaç de permetre el seguiment de la degradació del contaminant per tractament foto-Fenton al llarg del temps que dura l'experiment, el qual és pràcticament impossible.

A la Figura 17, s'observa com tot i seguir la tendència del sistema l'error ha augmentat respecte a la predicció en l'entrenament. A l'igual que passava amb Cabrera Reina, el sistema presenta un lleuger error en la simulació dels punts inicials de l'experiment.

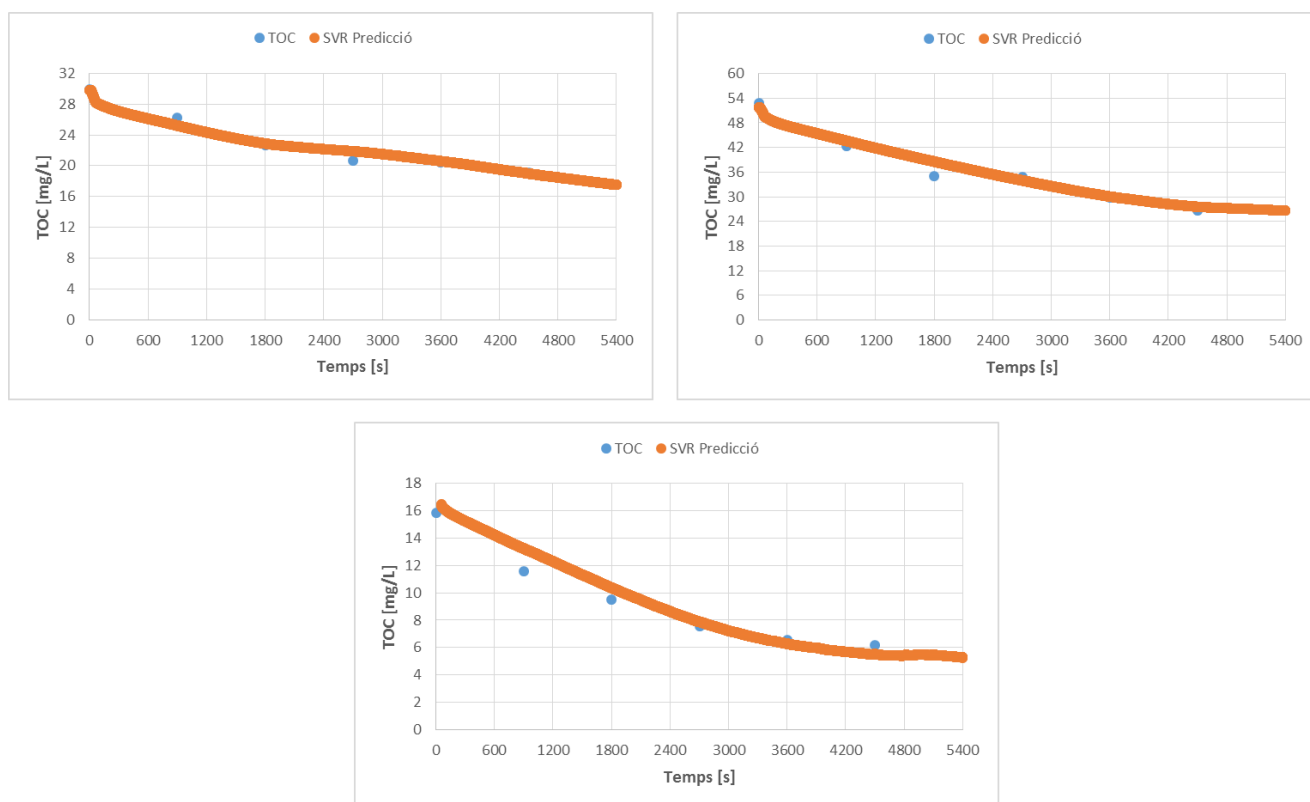


Figura 17. Predicció de la degradació del contaminant per a 3 experiments de la validació.

Si es té en compte l'avaluació de la precisió numèrica, s'observa com el metamodel té una precisió molt alta si es té en compte amb el número limitat de dades disponibles (49 punts d'entrenament).

Taula 4. . RMSE, RPE i CC per les prediccions obtingudes

	RMSE			RPE %			CC		
	Entren.	Valid.	Tot	Entren.	Valid.	Tot	Entren.	Valid.	Tot
Planta Pilot	0.8541	2.021	1.459	4.384	6.927	5.444	0.9976	0.9906	0.9939

És difícil de fer una valoració precisa sobre el rendiment del metamodel en aquest cas, ja que el comportament real del procés encara no es coneix perfectament com en el cas d'estudi de la cinètica de primer ordre i per tant, l'única manera de poder validar la precisió es comparant la predicció dels metamodels amb les dades mesurades que ja tenen errors. Pel cas de la planta pilot el coneixement sobre el procés, l'avaluació visual i les mesures de validació numèrica han de participar per la validació dels resultats.

Tot i així, i a l'igual que passava amb la simulació de dades perfecte, el que sembla passar es que la manca d'informació fa que la predicció de la dinàmica no sigui tot lo precisa que es pretenia per tant això sembla indicar que s'ha d'introduir més informació al metamodel en l'entrenament, la suficient perquè s'introdueixi informació sobre el comportament.

5. Avaluació Econòmica

En aquest apartat es detallen tots els costos ocasionats durant la realització d'aquest projecte. Els costos totals es dividiran en amortitzacions dels equips utilitzats, costos de software utilitzats i costos de personal en funció de les hores dedicades.

5.1. Cost de Software

A continuació es detallen els software utilitzats per duu a terme aquest projecte.

Taula 5.Desglossament del software emprat

Nom	Cost [€]
R project	0
Microsoft Office 2013	69.00
TOTAL	69.00

5.2. Amortitzacions

Per fer el càlcul de les amortitzacions s'han considerat el cost d'adquisició de tots els equipaments emprats durant aquest projecte.

El càlcul de les amortitzacions es realitza a partir de la següent fórmula:

$$Amortització = (Cost - Valor_{residual}) \cdot \dot{U}_{srecurs} / Vida_{útil} \quad (50)$$

El valor residual es considera zero, ja que els equips empleats s'esgoten fins que finalitza la seva vida útil, i per tant, no es poden vendre.

Taula 6.Cost dels equips utilitzats.

Equip	Preu [€]	Vida útil [any]	Ús del recurs	Cost Total [€]
Portàtil	850.00	5	0.9	153.00
TOTAL				153.00

5.3. Cost de Personal

El càlcul dels costos del personal es calculen com si el projecte fos realitzat per una empresa. Es tenen en compte els costos d'un enginyer, d'un consultor i un tècnic de laboratori.

Els costos de personal s'han calculat a partir de la següent fórmula:

$$Cost_{personal} = Recursos_{Humans} \cdot ((SBA + SS)/Hores_{anuals}) \quad (51)$$

Es consideraran un total de 1700 hores anuals. La part de la quota de la seguretat social (SS) assignada a la empresa és d'un 30% sobre el salari brut anual (SBA). Com que s'han diferenciat tres grups professionals, cada un de les quals pertanyen a una categoria de cotització diferent, es tindrà un valor de salari brut anual diferent per cada categoria.

Taula 7. Cost total de personal en el cas d'una empresa

Càrrec	Recursos Humans [h]	SBA [€]	SS [€]	Hores anuals [h]	Cost Total [€]
Enginyer	300	20000.00	6000.00	1700	4588.235
Consultor	30	28000.00	8400.00	1700	642.353
TOTAL					5230.588

5.4. Cost Total

Pel càlcul del costos totals es realitza la suma dels costos desglossats en els diferents apartats anteriors.

Taula 8. Cost Total del projecte

Cost Software	69.00
Amortitzacions	153.00
Cost Personal	5230.588
TOTAL	5452.588

Conclusions

En aquest projecte s'han assolit exitosament els objectius

- 1) S'han analitzat tres tècniques de modelat basat en dades: el *Kriging*, les xarxes neuronal i la *Support Vector Regression*. De la qual s'ha escollit la *Support Vector Regression* pel modelat degut la seva avantatge computacional respecte les altres tècniques.
- 2) S'ha determinat un enfoc de modelat que descriu el procés d'estudi. La forma que aquest presenta és:

$$y(t) = f[x_1(t), x_2(t), y(0), x_1(0), x_2(0)]$$

- 3) S'ha aplicat l'enfoc determinat per la generació d'un *soft sensor* a tres casos d'estudi:
 1. La inferència de C a partir les mesures d'A i B, utilitzant dades simulades provinent de un sistema caracteritzat per una cinètica de primer ordre.
 2. La inferència del TOC a partir de les mesures de d'Oxigen Dissolt utilitzant dades simulades del procés foto-Fenton seguint el model proposat per Cabrera Reina al 2012.
 3. La inferència del TOC a partir de les mesures de Temperatura i Redox en el procés que té lloc en una planta pilot de tractament d'aigua pel procés d'oxidació avançada foto-Fenton.
- 4) S'ha desenvolupat un *soft sensor* capaç de predir el comportament de la variable de sortida costosa d'un procés a partir de les dades mesurades on-line.

Les principals conclusions a les que s'ha arribat després d'haver implementat el *soft sensor* i analitzat els resultats són:

- 1) Que és possible desenvolupar un *soft sensor* capaç de predir el comportament d'un procés químic complex a partir de dades *on-line*, estalviant enormes quantitats de temps i costos.
- 2) Els metamodels generats per cada un dels casos, tenen una precisió en la predicció de la variable sortida:
 - Bona quan hi ha molts punts experimentals/simulats (mesures *on-line*)

- No tant acurada quan tenen un baix nombre de dades (mesures off-line).
- 3) La capacitat d'estimació de les variables no mesurades es milloraria, sobretot en els casos del procés foto-Fenton:
- Afegint més dades de la mesura *off-line*, implicant una reducció en la presa de mostres no sempre viable.
 - Afegint més informació al metamodel que indiqués la dinàmica del sistema, donant un lligam entre les dades introduïdes, com poden ser les derivades en el cas de tenir un model que el complementi.

Que suggereix el camí del treball futur.

Propostes de millora

Del resultat de realitzar aquest projecte sorgeixen un conjunt de propostes de millora que es podrien dur a terme. Tot seguit s'enumeren algunes de les idees a tenir en compte per la continuïtat d'aquest estudi i/o ampliació del coneixement en aquest camp:

- Obtenir un enfoc de modelat que sigui capaç de caracteritzar millor el sistema desitjat.
- Connectar la planta pilot via OPC per poder avaluar el *soft sensor* de manera on-line.
- Generar un algoritme que ens permeti fer diagnosi de falles de la planta pilot.

Agraïments

Vull mostrar el meu agraïment a diverses persones que han fet possible la realització d'aquest projecte.

En primer lloc vull agrair al professor Moisès Graells, director del projecte, per inculcar-me la seva passió pel món de la simulació de processos. Li agraeixo la seva dedicació al projecte i tots els consells donats quan les coses no sortien.

També vull agrair al professor Gerard Escudero, per la seva paciència i el seu cop de mà en l'enteniment dels models basats en dades. Sense la seva ajuda aquest projecte no podria haver-se realitzat.

A Ana Somoza i Joel Nevado, perquè aquest projecte no hauria estat igual de suportable sense els nostres moments al CIAO. A Gerard Campanyà, amic, col·lega i company de foto-Fenton.

Als meus pares i germà, que durant tot aquests anys m'han donat el seu suport i ha confiat en mi, inclús quan jo dubtava de mi mateixa. I per últim, vull agrair-li a Albert Carnero, per la seva paciència i suport en els meus moments d'estrès i deliri estadístic.

A tots i cadascun de vosaltres, gràcies.

Bibliografia

Referències bibliogràfiques

- [1] K. Azman, J. Kocijan, 2007, Application of Gaussian processes for black-box modelling of biosystems, *ISA Transactions* 46, 443–457.
- [2] Basak, D., Pal, S., Patranabis, D. Support Vector Regression. *Neural Information Processing*, 11, (2007), 203-224.
- [3] Broséus, R., Vincent, S., Aboulfd, K., Daneshvar, A., Sauvé, S., Barbeau, B., Prévost, M. Ozone oxidation of pharmaceuticals, endocrine, disruptors and pesticides during drinking water treatment. *Water Res.* 43, 2009, 4707–4717.
- [4] Cabrera Reina, A., Santos-Juanes, L., Garcia, J.L., Casas, J.L., Sanchez, J.S. 2012. Modelling photo-Fenton process for organic matter mineralization, hydrogen peroxide consumption and dissolved oxygen evolution, *Appl. Catal. B.* 119-120,132.
- [5] Campanyà, G. Simulation and control of a photo-Fenton plant, Treball Final de Grau, Universitat Politècnica de Catalunya.
- [6] Chacon, J.M., Leal, M.T., Sanchez, M. y Bandala, E. 2006. Solar photocatalytic degradation of azo-dyes by photo-Fenton process. *Dyes. Pigm.* 69, 144.
- [7] Clarke, S., Griebisch, J., Simpson, T. Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses. *Journal of Mechanical Design*, (2005), 127, 1077-1087.
- [8] Desai, K.; Badhe, Y.; Tambe, S; Kulkarni, B. Soft-sensor development for fed-batch bioreactors using support vector regression, *Biochemical Engineering Journal*, 27, 2006, 225-239.
- [9] A. Duran, J.M. Monteagudo, M. Mohedano. Neural networks simulation of photo-Fenton degradation of Reactive Blue 4. *Applied Catalysis B: Environmental* 65 (2006) 127–134
- [10] J. Farias, E.D. Albizzati, O.M. Alfano, 2009, Kinetic study of the photo-Fenton degradation of formic acid: Combined effects of temperature and iron concentration, *Cat.Today*, 144,117-123.
- [11] Fortuna, L., Graziani, S., Rizzo, A., Xibilia, M.G. Soft Sensors for Monitoring and

Control of Industrial Processes, Advances in Industrial Control, Springer, 2007.

- [12] P. Gerland, A. Raftery, H. Ševčíková, N. Li, D. Gu, T. Spoorenberg, L. Alkema, B. Fosdick, J. Chunn, N. Lalic, G. Bay, T. Buettner, G. Heilig, J. Wilmoth. World population stabilization unlikely this century. *Science* (2014, 234-237)
- [13] P. Jain, I. Rahman, B. D. Kulkarni, 2007, Development of a Soft Sensor for a Batch Distillation Column Using Support Vector Regression Techniques, *Chem.Eng.Res.Des.*, 85, 283-287.
- [14] Shima Khatibisepehr, Biao Huang, and Swanand Khare (2013). "Design of inferential sensors in the process industry: A review of Bayesian methods," *Journal of Process Control*, 23 (10), pp. 1575-1596.
- [15] Klammerth, N., Malato, S., Maldonado, M.I., Agüera, A., Fernández-Alba, A.R., 2010a. Application of photo-Fenton as a tertiary treatment of emerging contaminants in municipal wastewater. *Environ. Sci. Technol.* 44, 1792–1798.
- [16] Kleijnen, J. P. C., 1987, *Statistical Tools for Simulation Practitioners*, Marcel Dekker, New York.
- [17] Bao Lin, Bodil Recke, Jørgen K. H. Knudsen, and Sten Bay Jørgensen (2007). "A systematic approach for soft sensor development," *Computers and Chemical Engineering*, 31, pp. 419-425
- [18] Martínez Bueno, M.J., Gomez, M.J., Herrera, S., Hernando, M.D., Agüera, A.; Fernández-Alba, A.R. Occurrence and persistence of organic emerging contaminants and priority pollutants in five sewage treatment plants of Spain: Two years pilot survey monitoring. *Environmental Pollution* 164: 267 – 273 (2012)
- [19] T. Masters, 1993, *Practical Neural Network Recipes in C++*.
- [20] Z. K. Nagy, Model based control of a yeast fermentation bioreactor using optimally designed artificial neural networks, *Chemical Engineering Journal* 127, 2007, 95-109.
- [21] Oller I, Malato, S, Sánchez-Pérez JA. Combination of advanced oxidation processes and biological treatments for wastewater decontamination: a review. *Sci Total Environ* 2011;409(20):4141–66.
- [22] Perez-Moya, M. Graells, M., Buenestado, E., Gutierrez, E., Galindo, J., Mansilla, H.D. 2008. Modelling Approach to Fenton and Photo-Fenton Treatments. *J. Adv.Oxid. Technol.* 11:1, 1.

- [23] Petrovic, M., Gonzalez, S., Barcelo, D., 2003. Analysis and removal of emerging contaminants in wastewater and drinking water. *Trac-Trends Anal. Chem.* 22, 685-696.
- [24] Pignatello, J.J.; Oliveros, E.; MacKay, A. 2006. Advanced Oxidation Processes for Organic Contaminant Destruction Based on the Fenton Reaction and Related Chemistry. *Crit. Rev. in Environ. Sci. Technol.* 36(1), 1.
- [25] Sabine Gob , Esther Oliveros, Stefan H. Bossmann , Andre M. Braun , Roberto Guardani , Claudio A.O. Nascimento. Modeling the kinetics of a photochemical water treatment process by means of artificial neural networks *Chemical Engineering and Processing* 38 (1999) 373–382
- [26] Shardt, Y. Data Quality Assessment for Closed-Loop System Identification and Forecasting with Application to Soft Sensors. Doctoral Thesis, University of Alberta, 2012.
- [27] A.Shokry, A. Espuña, 2014. Applying Metamodels and Sequential Sampling for Constrained Optimization of Process Operations, *Lecture Notes in Computer Science*, 8468, 396-407.
- [28] I.A. Shiklomanov, *World water resources, A New Appraisal and Assessment for the 21st Century*, 1998. ISBN: 0070570612
- [29] Simpson, T. W., Peplinski, J., Koch, P. N., and Allen, J. K. , “Metamodels for Computer-Based Engineering Design: Survey and Recommendations,” *Eng. Comput.*, 2001, 17(2), pp. 129–150.
- [30] Tokumura, M., Tawfeek Znad, H., Kawase, Y. 2006. Modeling of an external light irradiation slurry photoreactor: UV light or sunlight-photoassisted Fenton discoloration of azo-dye Orange II with natural mineral tourmaline powder. *Chem. Eng. Sci.* 61, 6361.
- [31] Vicente, P. Modelització i Anàlisi basada en dades del tractament foto-Fenton en una Planta Pilot, Treball Final de Grau, Universitat Politècnica de Catalunya.
- [32] Yuri A.W. Shardt. Data Quality Assessment for Closed-Loop System Identification and Forecasting with Application to Soft Sensors, Doctoral Thesis, University of Alberta, 2012.
- [33] Eliana Zamprogna, Massimiliano Barolo, and Dale E. Seborg (2005). “Optimal selection of soft sensor inputs for batch distillation columns using principal component analysis,” *Journal of Process Control*, 15 (1), pp. 39-52.

Bibliografia complementària

- <https://desktop.arcgis.com/es/arcmap/latest/extensions/geostatistical-analyst/understanding-ordinary-kriging.htm>, Maig 2016
- <http://www.mathworks.es/products/matlab/>, Maig 2016
- <http://www.r-project.org/>, Maig 2016
- http://www.seg-social.es/Internet_1/Trabajadores/CotizacionRecaudaci10777/Basesytiposdecotiza36537/index.htm , Juny 2016
- <http://www.svms.org/regression/> , Juny 2016
- http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?_r=0 , Març 2016
- <http://blog.revolutionanalytics.com/2011/08/google-r-effective-ads.html>, Març 2016
- <http://stackoverflow.com/>, Gener 2016
- <http://www.inside-r.org/>, Abril 2016

Annex

Codi de programació

SVR (R project)

```
install.packages("kernlab")
library(kernlab)

#Carregar les dades
dt <- read.table("Training_data.csv", header=F, sep=";")
nc <- dim(dt) [2]
nf <- dim(dt) [1]
names(dt) <- sapply(1:6, function(n) paste('v',n,sep=""))

#Grup training
l <- 1:nf
ltr <- l[!is.na(dt[l,6])]

#Model
model<-ksvm(v6~., data=dt, kernel="rbfdot", kpar =
list(sigma=0.0033),C=750, type="eps-svr")
model

#Prediccions de TOC segons el model entrenat
res <- predict(model,dt[,-6])
#Unir prediccions amb dades inicals i guardar en txt
resultat <- data.frame(dt,res)
write.table(resultat,"Journal_Training_data_0.0033_rbf_c750.txt",row.name
=FALSE,sep="," )

#Exemple de codi de validació
dt <- read.table("Validation_1.csv", header=F, sep=";")
names(dt) <- sapply(1:5, function(n) paste('v',n,sep=""))
res <- predict(model,dt[,-6])
resultat <- data.frame(dt,res)
write.table(resultat,"Journal_Validation_1_0.0033_rbf_c750.txt",row.name=
FALSE,sep="," )
```

Cinètica Primer Ordre (Matlab)

```
function [dCdt] = simufirst( Ca, Cb, Cc, k1, k2, dt)

tt=100;

Ca(1)=Ca;

Cb(1)=Cb;

Cc(1)=Cc;

t(1)= 0;

nsteps=round(tt/dt)+1;

for n=2:nsteps

    Ca(n)= Ca(n-1)-dt*k1*Ca(n-1);

    Cb(n)= Cb(n-1)+ dt*(k1*Ca(n-1)-k2*Cb(n-1));

    Cc(n)= Cc(n-1)+ dt*k2*Cb(n-1);

    t(n)=t(n-1)+dt;

end

dCdt=[Ca; Cb; Cc];

plot(t,Ca,t,Cb,t,Cc)

xlabel('Time')

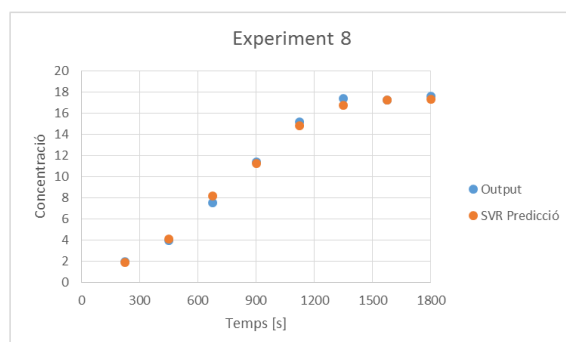
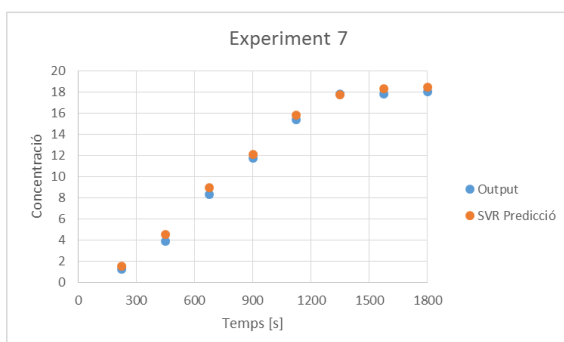
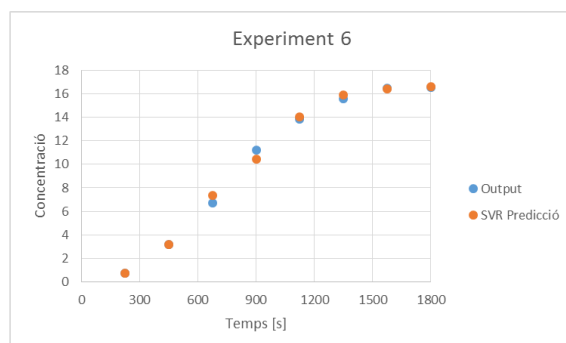
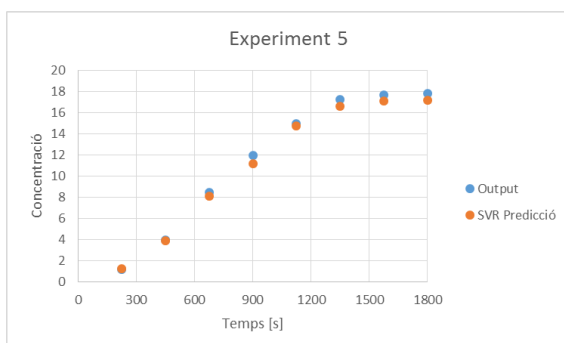
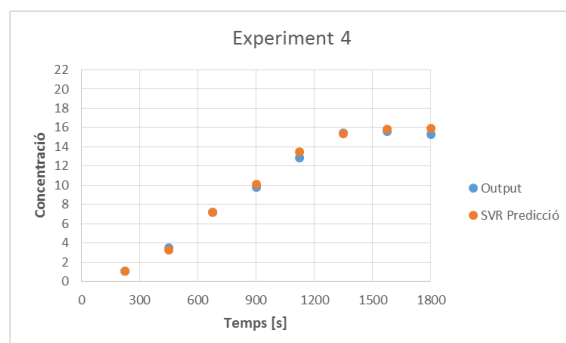
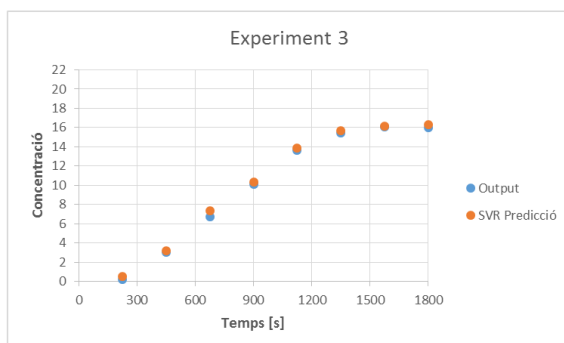
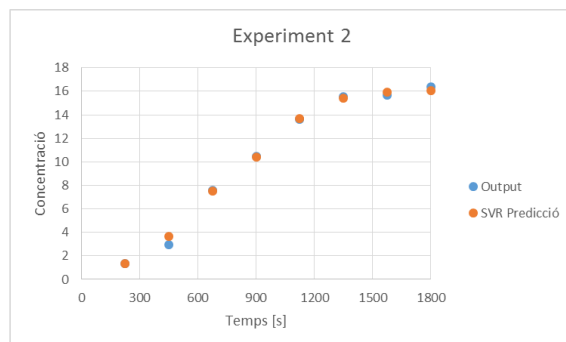
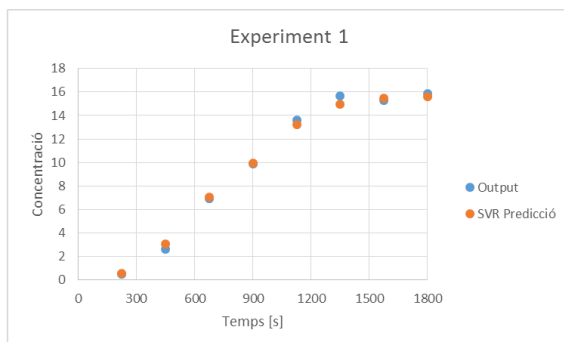
ylabel('Concentration')

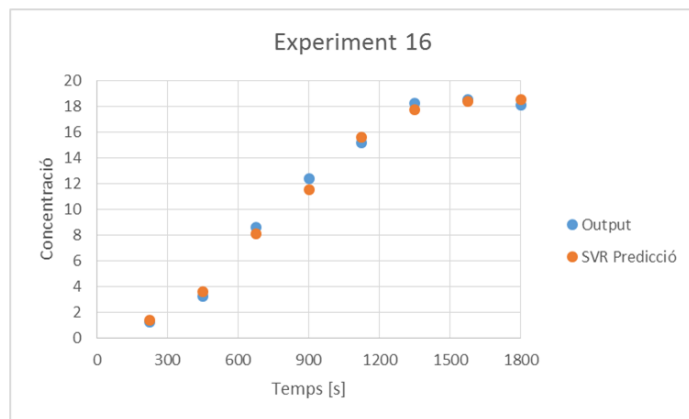
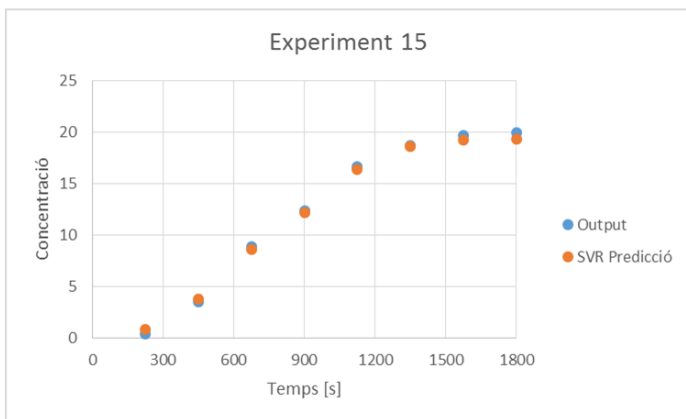
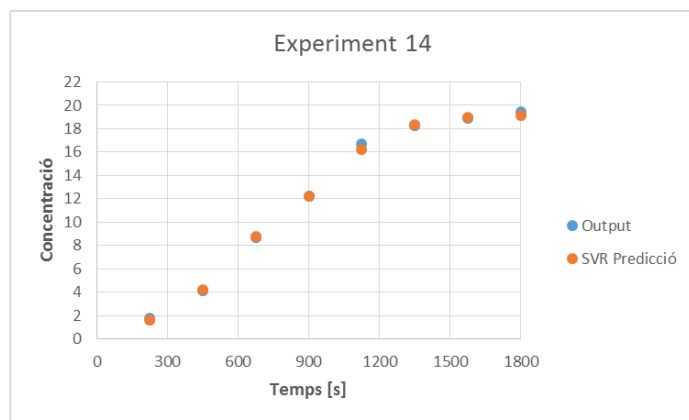
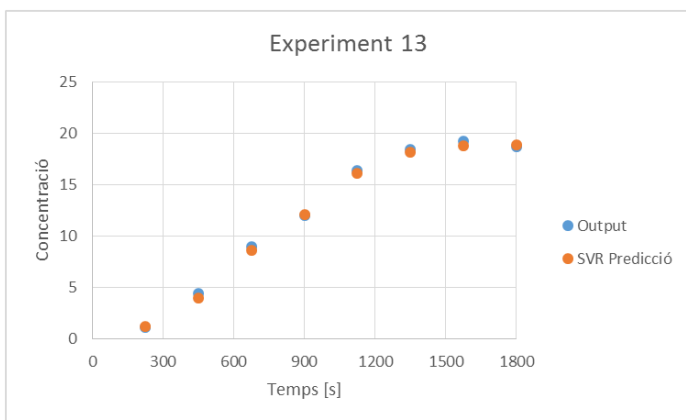
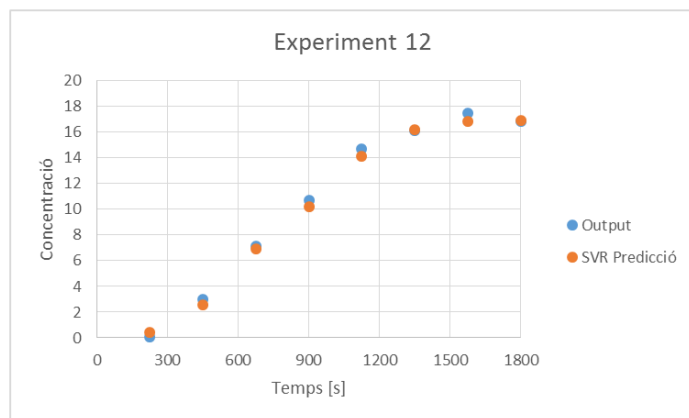
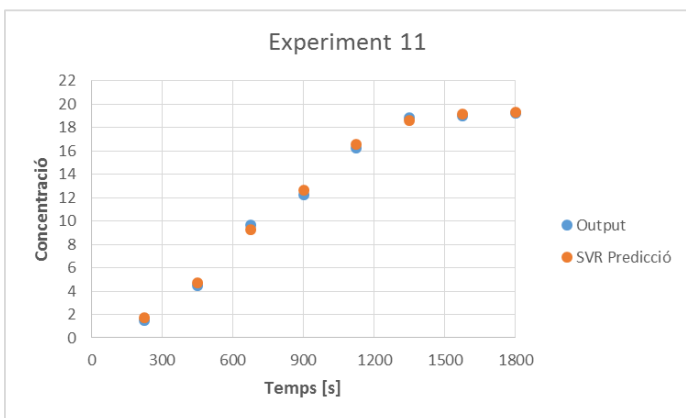
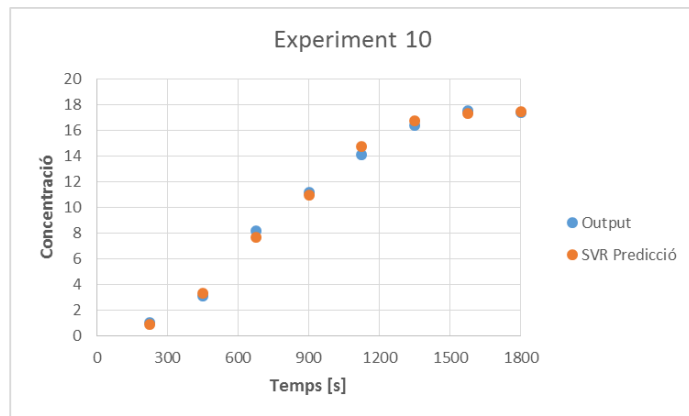
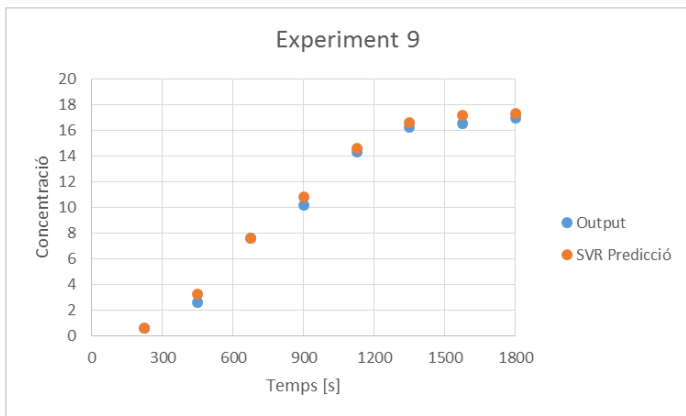
end
```

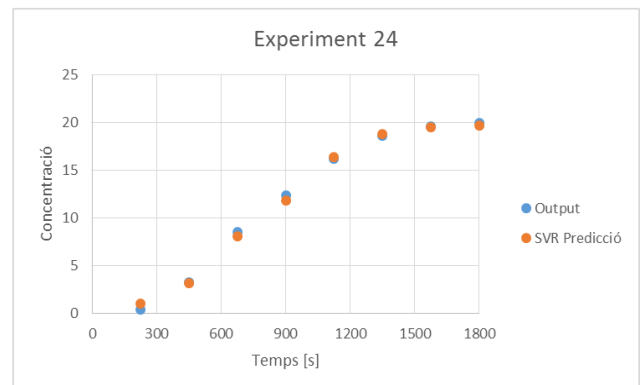
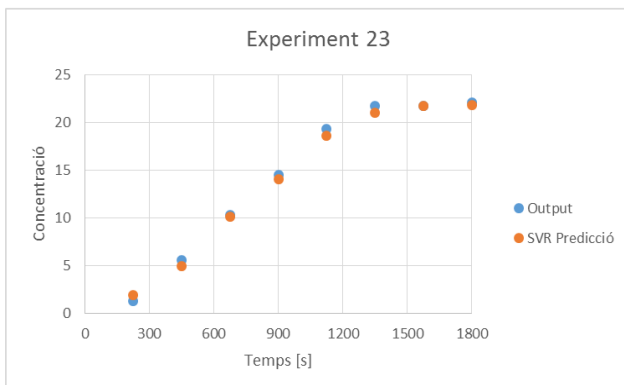
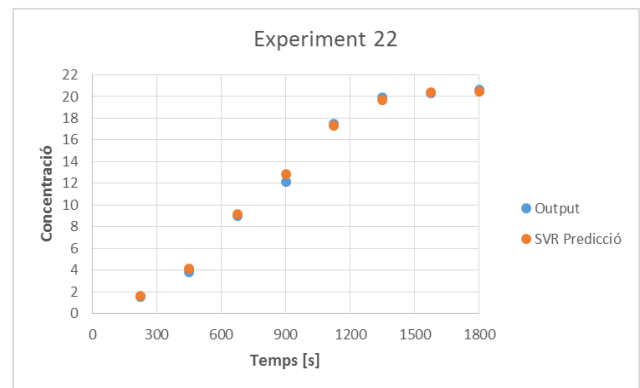
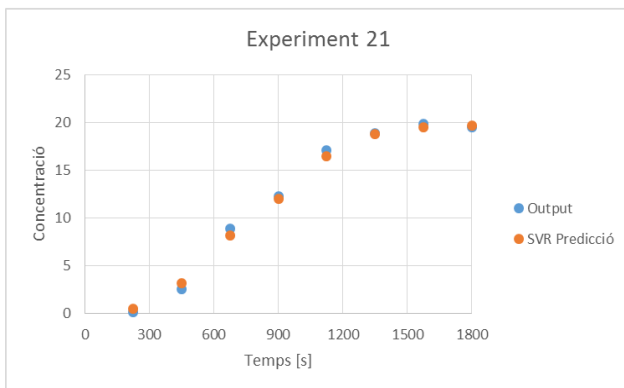
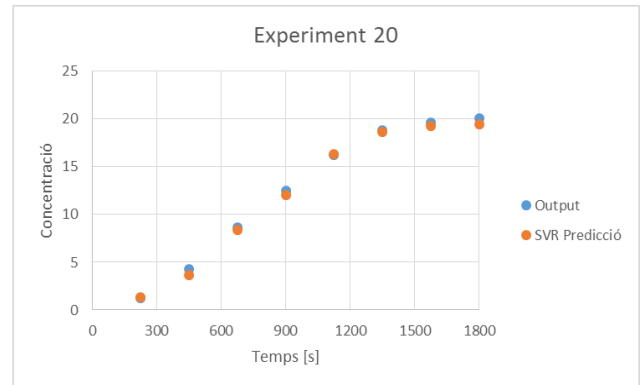
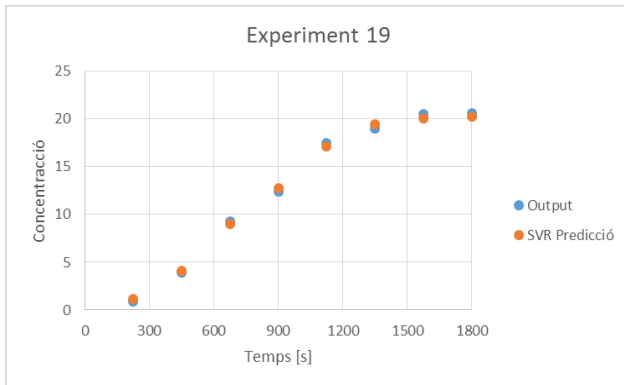
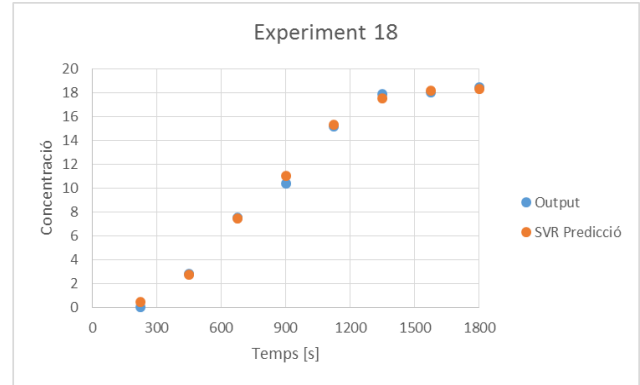
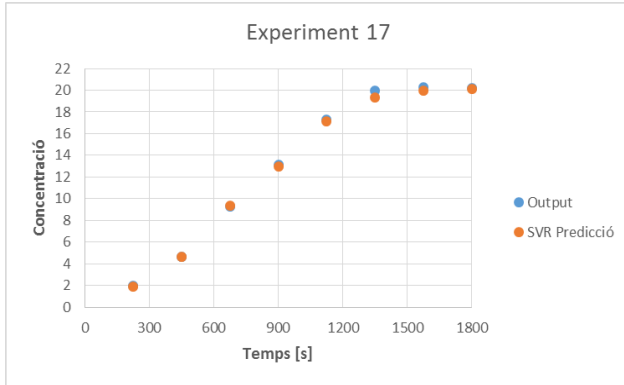
Gràfica dels models

Cinètica Primer Ordre

Entrenament





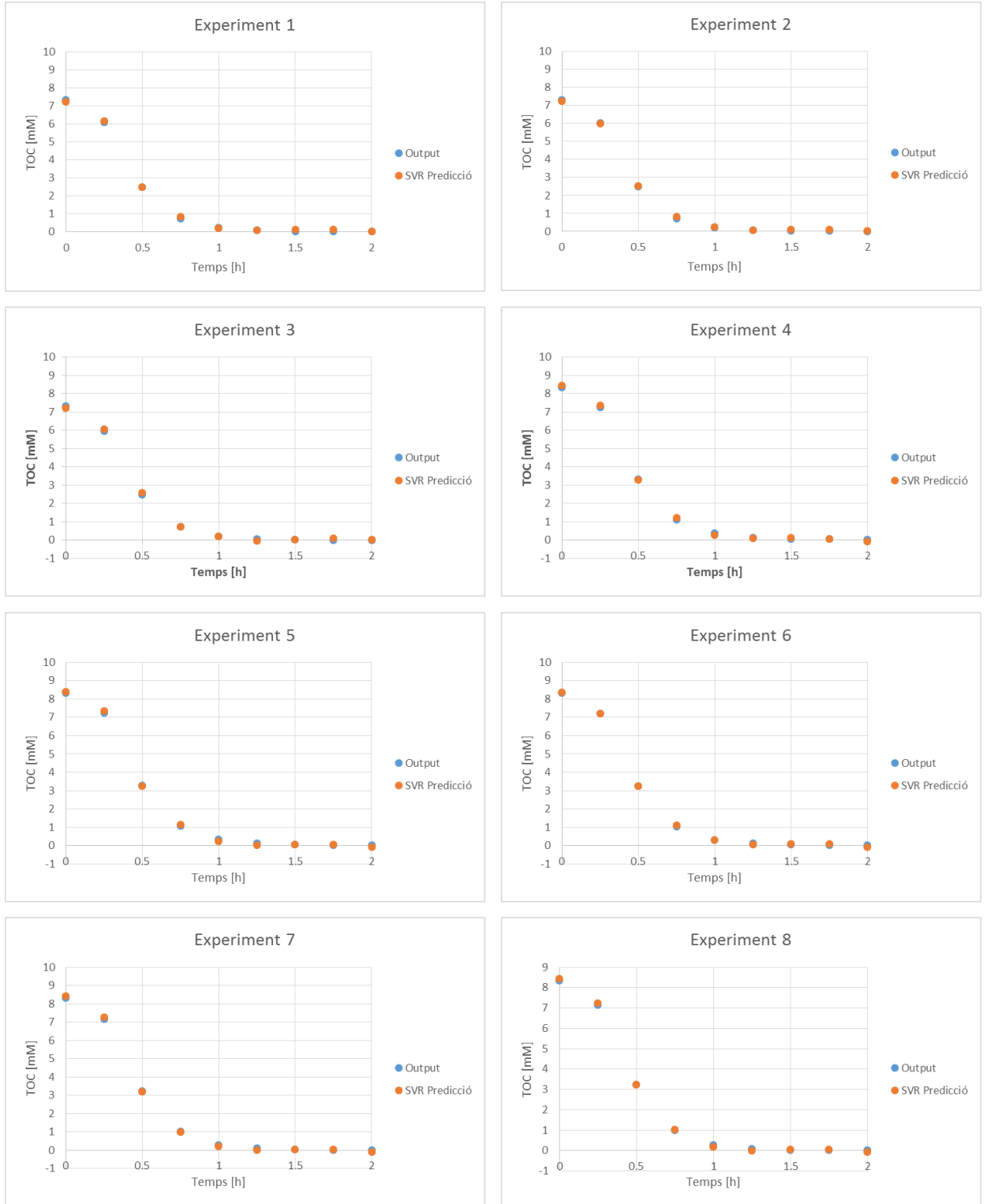


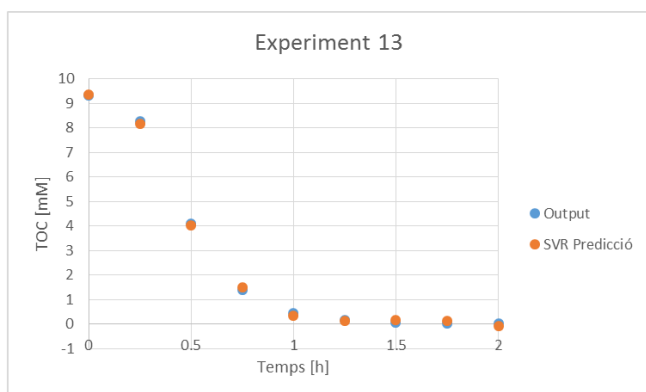
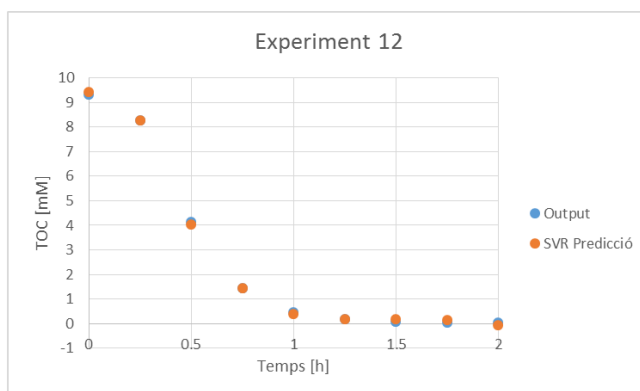
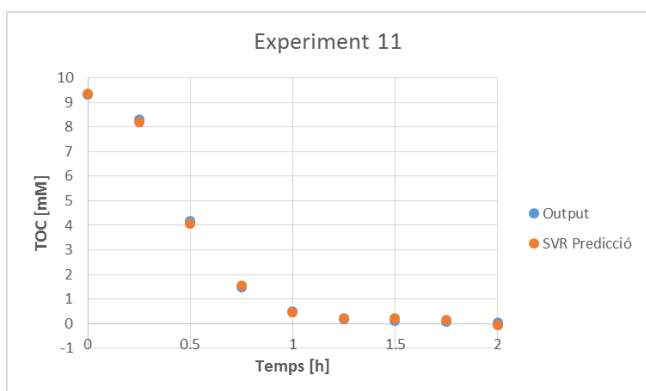
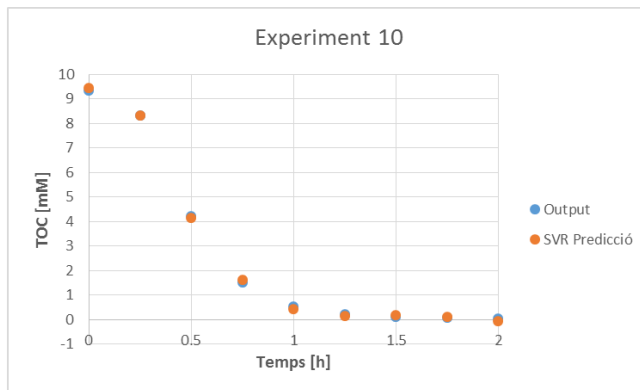
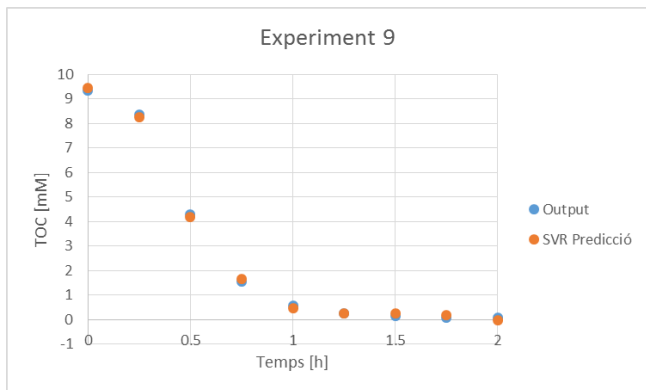
Validació

Les gràfiques pertanyents a la validació del model de cinètica de primer ordre es presenten en format digital. El nom del arxiu on es recullen té per nom: errors_calculation_Kinetics_first_order_24_100.

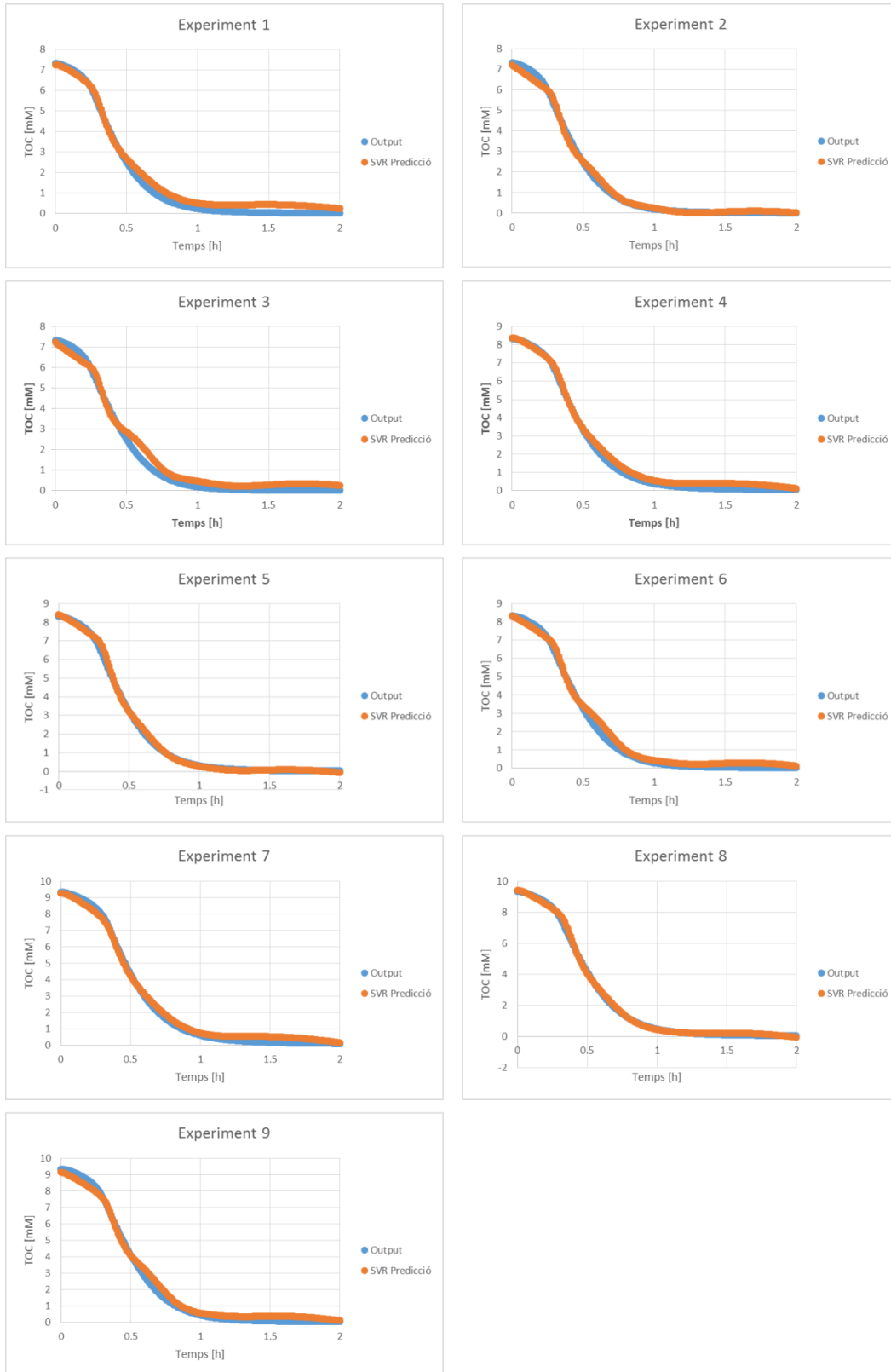
Cabrera Reina

Entrenament



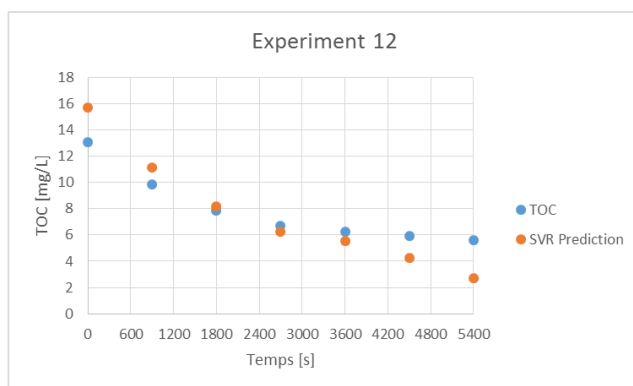
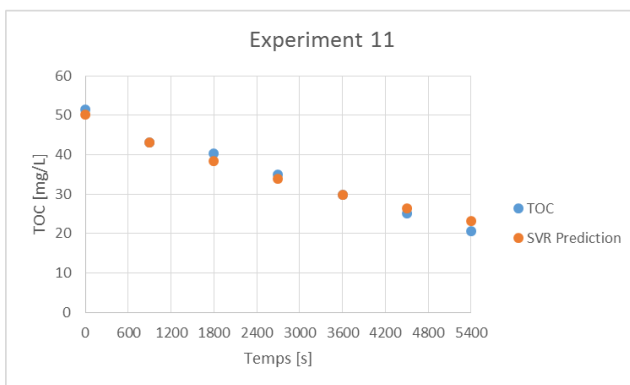
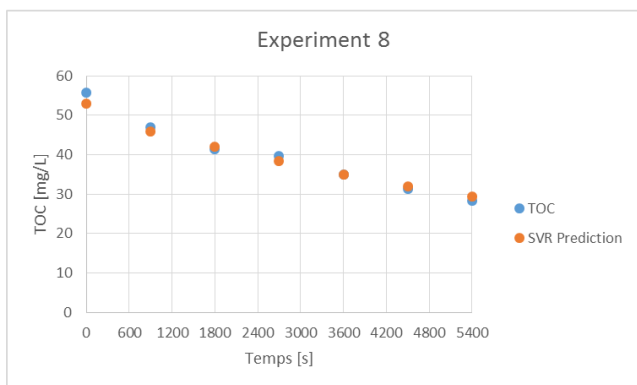
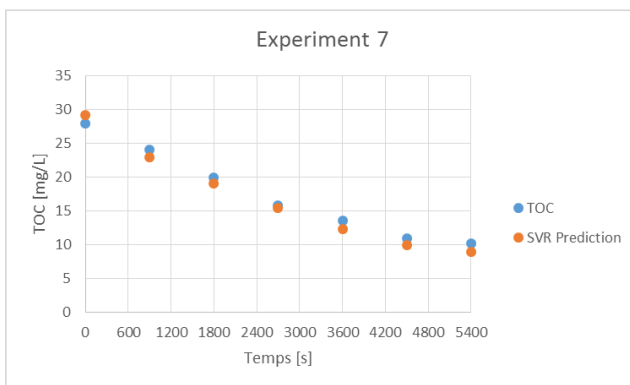
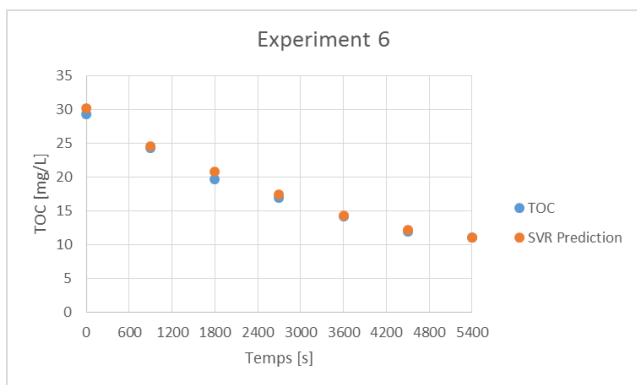
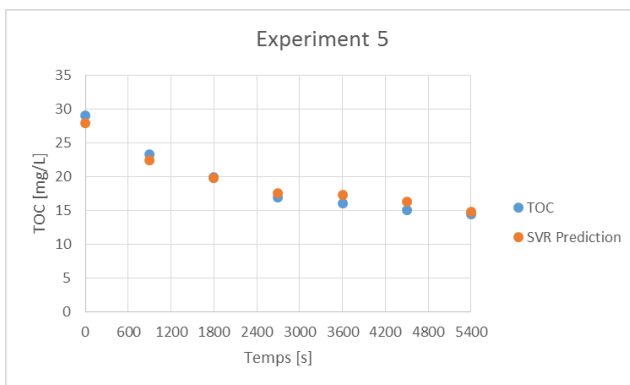
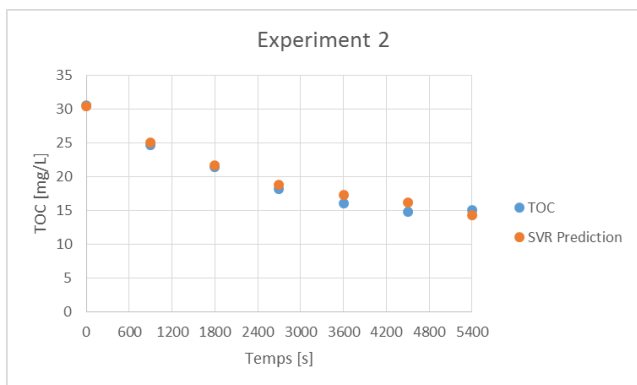


Validation



Planta Pilot

Entrenament



Validació

