

Memòria de pràctiques

Reconciliació de la base de dades europea de sistemes de sanejament pel desenvolupament de models de contaminants aquàtics

Oriol Canal Pujol

Biotecnologia

Curs 2019-2020

Pràctiques en empresa

Tutor acadèmic: Manuel Poch Espallargues

Tutor de l'empresa: Lluís Corominas

ÍNDIX

ICRA	3
MODEL EUROPEU DE SISTEMES DE SENAJAMENT	3
TASQUES	5
METODOLOGIA	6
Introducció al projecte:	6
Estratègies per solucionar els errors:	6
Transferència dels canvis de la base de dades del servidor local al global	9
AVALUACIÓ INDIVIDUAL DE LES PRÀCTIQUES	9
AGRAÏMENTS	9

ICRA

L'ICRA (*Institut Català de Recerca de l'Aigua*) és un institut de recerca que va ser creat el 26 d'octubre de 2006 pel govern de la Generalitat de Catalunya en el marc del Programa de Centres de Recerca de Catalunya (CERCA). ICRA conta amb l'impuls del Departament d'Empresa i Coneixement de la Generalitat de Catalunya (DECO), l'Agència Catalana de l'Aigua (ACA) i de la Universitat de Girona (UdG). Es troba situat al Parc Científic i Tecnològic de la Universitat de Girona, Edifici H2O. És un centre multidisciplinari de la investigació de l'aigua i el seu pla de recerca s'estructura en tres grans àrees de recerca: recursos i ecosistemes, qualitat de l'aigua i tecnologies i avaluació.

La investigació que es porta a terme a ICRA tracta sobre tots els aspectes relacionats amb l'aigua, en especial aquells que tenen a veure amb el seu ús racional i l'efecte que pot provocar l'activitat humana sobre els recursos hídrics dels quals es disposen.

MODEL EUROPEU DE SISTEMES DE SANEJAMENT

ICRA està desenvolupant un model europeu que permet descriure el transport i l'eliminació de contaminants de l'aigua des del seu origen en les llars, passant pels col·lectors, les depuradores i fins que arriben a les xarxes fluvials. Amb aquest es poden definir noves estratègies de gestió i noves polítiques en el sector de l'aigua. Per fer això ICRA està desenvolupant un aplicatiu informàtic d'un mapa geoespacial on hi ha tota la informació necessària sobre les xarxes de sanejament.

Aquest model s'ha desenvolupat en l'àrea de recursos i ecosistemes en col·laboració amb l'àrea de tecnologies i avaluació. La primera versió del model es va publicar el 2019 (Carme Font, 11 gener 2019) i actualment s'està treballant amb una versió millorada que inclogui més detall en la part de sanejament, que incorpori les dades de sanejament europees de la base de dades UWWTD).

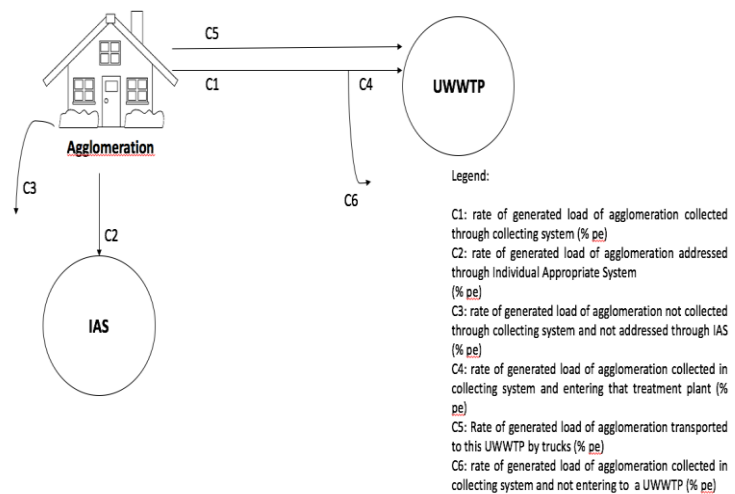
El model que es vol arribar a crear és un mapa que es separa en parcel·les de 7x7 km i a cada una d'aquestes parcel·les hi ha la informació sobre les aigües residuals que es generen en aquella zona i el seu tractament. S'utilitza el programa QGIS que permet l'assignació de diferents capes d'informació sobre el mapa per tal de poder acabar obtenint un mapa geoespacial del tractament de les aigües residuals europees.

Per tal de realitzar aquest model es parteix de la base de dades de l'european environment agency on es recull la informació de la recol·lecció, el tractament i el punt de descàrrega d'aigua urbana i industrial (Agency, 2019). L'objectiu d'aquesta base de dades és la protecció del medi ambient d'efectes adversos que poden estar produïts per la descàrrega de l'aigua residual.

Aquesta base de dades consta de 12 taules de les quals s'han utilitzat 5 d'aquestes per tal de realitzar el model que són les següents:

- T_Agglomerations: conté la informació de les diferents aglomeracions. Ens proporciona el nom, les coordenades, la càrrega generada i diferent informació sobre la aglomeració en la qual es recull l'aigua residual.
- T_UWWTPs: Conté la informació de les plantes de tractament, ens dona informació sobre el tractament, la localització, l'activitat la càrrega d'entrada
- T_UWWTP_Agglo: Taula que ens connecta les aglomeracions amb les diferents depuradores.

El que es vol aconseguir de la informació extreta en les taules esmentades anteriorment es mostra resumit a la imatge 1. La aglomeració genera una càrrega d'aigua residual. Aquesta aigua residual pot ser recollida per un sistema col·lector (C1). Aquesta aigua recollida pels sistemes col·lectors pot anar a parar a la depuradora (C4) o bé que no acabi arribant a la depuradora i per tant no sigui tractada (C6).



Excepcionalment l'aigua residual pot ser transportada mitjançant camions a les depuradores (C5) o pot ser tractada per sistemes de sanejaments individuals (C2). Finalment també trobem aigua residual que no és recollida (C3).

Il·lustració 1: Esquema de les dades aportades per les taules T_Agglomerations, T_UWWTPs i T_UWWTP_Agglo.

- T_DischargePoints: Taula que conté informació individual dels punts de descàrrega de les diferents plantes de tractament.
- T_UWWTPs_emission_load: Conté informació addicional de l'entrada i la descàrrega de matèria orgànica i dels nutrients de la planta de tractament.

TASQUES

Treballant en la generació d'aquest model, s'ha identificat que hi ha diferents errors en la informació que conté la base de dades proporcionada per la unió europea. Aquests errors podrien afectar posteriorment a la realització del model i per tant és necessària la detecció i la solució d'aquests.

La meua funció dins el projecte ha estat la verificació de les dades utilitzades que es troben a les taules esmentades anteriorment per aconseguir minimitzar els errors a l'hora de realitzar el model. Un cop detectats aquests problemes, s'han dissenyat diferents estratègies i s'ha escollit la més adient per la solució dels errors tenint en compte les característiques del model.

Per tant l'objectiu principal de les pràctiques en empresa ha sigut plantejar com solucionar els problemes detectats a la base de dades per així posteriorment poder introduir els canvis convenients a la base de dades de l'ICRA mitjançant comandes SQL.

Els objectius individuals de les pràctiques han sigut: saber com es treballa en un grup de recerca, saber analitzar críticament situacions complexes, adquirir coneixements informàtics, saber com es treballa amb base de dades i tenir un primer contacte amb l'àmbit professional.

La motivació que em va fer escollir aquestes pràctiques és el fet sobretot de treballar conjuntament amb un programador informàtic llicenciat en biotecnologia. El fet de voler centrar els meus estudis acadèmics en els pròxims anys en la bioinformàtica ha fet que hagi vist com a una oportunitat molt convenient per la meua formació acadèmica el fet de treballar en un projecte que es basa en la utilització de bases de dades per la creació d'un model. Així doncs vaig veure com una oportunitat aquestes pràctiques ja que tocaven conjuntament temes biològics com són les depuradores i a més es treballava a partir d'una base de dades que sol ser un fet molt usual en la bioinformàtica

Per tal de poder detectar i solucionar els errors de la base de dades s'ha treballat conjuntament amb en Lluís Bosch (programador informàtic) i en Lluís Corominas (investigador científic de l'àrea de tecnologies i avaluació). S'ha realitzat una reunió setmanalment amb en Lluís Corominas amb l'objectiu de presentar-li les propostes de solucions dels errors de la base de dades per així trobar la forma idònia per solucionar-los i aproximadament dos reunions setmanals amb el Lluís Bosch que conjuntament amb ell s'han introduït les comandes necessàries a les bases de dades per tal de modificar-les i corregir-les.

METODOLOGIA

Les pràctiques realitzades es podrien desglossar en 3 apartats diferents on a cada un d'ells m'ha permès desenvolupar-me en diferents coneixements:

Introducció al projecte:

Per començar a treballar amb les bases de dades, primer necessitava conèixer en què consistia el projecte i tenir el coneixement d'on provenien i què significaven les dades de la base de dades. És per això que se'm van proporcionar diferents documents on s'explica en què consisteix el projecte, en quin estat es trobava i diferents articles on podria ser útil l'aplicació del projecte. Això em va ajudar a entendre quina era la funció del projecte i quins eren els paràmetres més rellevants per posteriorment, sabent la finalitat i els passos que es seguirien per realitzar el model, poder analitzar críticament els errors que s'havien trobat en la base de dades i així poder proposar les solucions més adequades.

A més també vaig aprendre el funcionament de les comandes SQL, que em va permetre poder realitzar diferents comandes a la base de dades per tal d'extreure'n la informació que he cregut necessària en cada moment.

També a partir de la web de la base de dades de la unió europea vaig aprendre a interpretar tota la informació que ens proporcionava la base de dades. A més vaig realitzar recerca d'informació sobre la legislació que estava sotmesa el tractament de les aigües residuals (Europeas, 1991) .

Finalment es va instal·lar el programari necessari de forma local en el meu servidor per tal de poder accedir a la base de dades creada per l'ICRA i així poder realitzar els canvis necessaris per solucionar els errors trobats en la base de dades de forma local.

Estratègies per solucionar els errors:

A l'inici de l'estada de pràctiques se'm va proporcionar la base de dades utilitzada per l'ICRA amb un total de 8977 errors detectats que es poden separar en diferents grups segons la seva causa. Aquests s'han solucionat manualment en els casos en que el número d'errors no fos molt gran o semiautomàticament a partir de comandes SQL. Per tal de solucionar-ho semiautomàticament es necessita la utilització d'una estratègia que es pugui seguir per tots els errors, és a dir que per solucionar els errors semiautomàticament es necessita seguir una pauta de comandes que pugui ser útil per la resolució de tots els errors d'un determinat tipus. A continuació s'explicaran de què es tracten i quines estratègies s'han determinat per solucionar-los:

1. **Aglomeracions sense coordenades** (21 errors): Al saber el nom de l'aglomeració, aquestes s'han buscat les seves coordenades en el google maps i s'han introduït en la base de dades.

2. **Depuradores sense coordenades** (37 errors): Per aquestes depuradores s'ha mirat a la taula de connexions (T_UWWTP_Agglo), a quina aglomeració estava associada i se l'hi ha assignat la mateixa coordenada que l'aglomeració. Com que el model amb el qual es tractaran les dades treballa amb cel·les de 7x7 km, s'ha considerat que les depuradores no es troben a més de 7 km de l'aglomeració de la que rep l'aigua residual i per tant l'error que es genera al no introduir precisament la localització de la depuradora, no acabarà tenint repercussió en la realització del model.
3. **Depuradores que no es troben a la taula de connexions (T_UWWTP_Agglo)** (2443 errors): A partir de diferents cerques SQL es va observar que 2419 d'aquestes depuradores estaven inactives i 23 estaven temporalment inactives. Així doncs és normal que no estiguin connectats a cap aglomeració. Només es va trobar una aglomeració que està activa i no es troba a la taula de connexions. Per aquesta depuradora es va trobar que a la taula T_UWWTPs la depuradora tenia una aglomeració associada i es va crear la connexió entre l'aglomeració trobada i la depuradora.
4. **Aglomeracions que no es troben a la taula de connexions (T_UWWTP_agglo)** (4141 errors): Es va observar que 676 d'aquestes aglomeracions estaven inactives i 174 temporalment inactives. Aquestes no es van considerar com a error ja que si no estan actives és normal que no tinguin una depuradora associada. Tot i això hi ha 3291 aglomeracions actives que no tenen depuradora associada.
Es va observar que 2298 d'aquestes aglomeracions actives tenien un remark que ens informava de que tenien un sistema IAS. Es va veure que totes les aglomeracions que tenien un remark informant de que les aigües eren tractades per IAS tenien en comú que en el remark apareixia o bé la paraula IAS o sceptic. És per això que es van filtrar les depuradores que no es trobaven a la taula de connexions i que en el remark apareixia o bé la paraula IAS o sceptic. Amb aquestes aglomeracions es va crear una nova taula anomenada T_Agg_IAS que ens informa de les aglomeracions que tenen un sistema de sanejament IAS.
Per les 1070 aglomeracions restants no es sap la informació sobre com es tracta les aigües residuals generades. Segons la DOE, (REFERENCIAAR!) aglomeracions amb menys de 2000 habitants, no estan obligades a reportar el sistema de sanejament de les seves aigües i aquest pot ser un dels motius pels quals no es sap el tractament que reben les aigües residuals generades per aquestes aglomeracions. És per això que es va decidir crear una nova taula amb aquelles aglomeracions en les quals no es sap el tractament que reben les seves aigües.
D'aquesta forma per la posterior utilització de les dades pel model, es tenen les aglomeracions separades segons si les seves aigües residuals van a parar a una depuradora, a un IAS o bé no es sap la informació de com es tracta l'aigua residual.
5. **Depuradores sense punt de descàrrega associat** (1753 errors): A partir d'una comanda SQL es va determinar que totes les depuradores sense punt de descàrrega estaven o bé inactives (1751) o temporalment inactives (2). Per tant no es troben punts de descàrregues per aquestes depuradores pel fet de no estar en funcionament.

6. **Depuradores amb múltiples punts de descàrrega** (249 errors): Cada punt de descàrrega ens aporta informació sobre la localització d'aquest. Dels punts de descàrrega repetits es va veure que la majoria es localitzaven a pocs metres de distància entre ells. Es van eliminar primer els punts de descàrrega que contenien menys informació ja que eren inactius o bé que en el seu remark ens informaven que havien estat substituïts per un altre. Seguidament entre els que ens proporcionaven la mateixa informació i només diferien en pocs metres es van eliminar aquells els quals es trobaven més distants a la depuradora o bé que no es trobaven situats sobre el riu.

Es van trobar 6 excepcions els són depuradores que l'aigua residual va a parar a varis punts de descàrrega. Aquestes excepcions eren depuradores que descarregaven tant al riu com a dipòsits per a reutilització d'aigua pel rec de camps i per tant no podien ser assumits com a errors.

7. **Depuradores que es troben a la taula T_UWWTP_agglo però no es troben a la taula T_UWWTPs** (3 errors): són depuradores que es troben a la taula que connecten depuradores i aglomeracions però no es troben a la taula que ens proporciona la informació de la depuradora. Primer de tot es va mirar que l'error no vingués per un error de sintaxi del codi de la depuradora. Per comprovar que no fos així es va mirar a la taula de depuradores que no hi hagués un codi de la depuradora semblant al de la taula de connexions però no es va trobar cap cas.

Les 3 depuradores es troben a Alemanya i es va utilitzar un mapa de depuradores d'Alemanya creat per la Unió Europe (l'Eau, 2017). Es va mirar la localització de les aglomeracions i es va estimar a quina depuradora és més possible que desemboqués segons el relleu de la zona i la proximitat. Una vegada trobada la depuradora d'aquestes 3 aglomeracions es va substituir el codi de la depuradora que no es trobava pel de la nova depuradora.

8. **Distància entre aglomeració i depuradora o entre depuradora i punt de descàrrega superior als 30 km** (330 errors). Es va comprovar que les aglomeracions tinguessin les coordenades correctament i en cas de que no fos així es van introduir les noves coordenades de l'aglomeració.

Per les depuradores es va comprovar que la ubicació indicada correspongués a una depuradora i es va veure que en moltes ocasions no era així i la ubicació es trobava al mig del bosc o del mar. En els casos on la localització no corresponia a una depuradora es va introduir les coordenades de l'aglomeració ja que es va considerar que la ubicació real de la depuradora es trobava a menys de 7 km de l'aglomeració de la qual rebia l'aigua residual i que per tant no afectaria al model el fet d'introduir-hi les coordenades de l'aglomeració

Pel que fa als punts de descàrrega es va considerar que no era possible que la depuradora descarregués a més de 30 km. Per això es va canviar la ubicació d'aquells punts de descàrrega que es troben a més de 30 km de la depuradora per el punt del riu més pròxim a la depuradora.

Transferència dels canvis de la base de dades del servidor local al global

Un cop solucionats tots els problemes es transferiran tots els canvis generats a la base de dades local del meu servidor a la base de dades de l'ICRA per la posterior generació del model. Aquest pas encara s'ha de realitzar però es portarà a terme durant les pròximes setmanes on es tindrà una reunió amb la Carme Font juntament amb el Lluís Corominas per explicar-li com s'han solucionat els errors perquè així ella pugui continuar amb el desenvolupament del model.

AVALUACIÓ INDIVIDUAL DE LES PRÀCTIQUES

Un cop realitzades les pràctiques valoro molt positivament l'aprenentatge que he pogut obtenir. A part de veure com es treballa en un grup de recerca del que m'he sentit que en formava part, també he après a utilitzar diferents eines informàtiques que em poden servir en els futurs estudis bioinformàtics com podria ser el sistema de comandes SQL.

També he après a utilitzar altres eines informàtiques com el GitHub que et permet treballar en un projecte amb col·laboració de persones de tot el món per tal de poder fer un seguiment del treball i a més incorpora un sistema de control de versions que et permet guardar diferents versions de projectes, treballs i bases de dades. Així en el cas de que s'incorpori algun error en la base de dades, et permet poder obtenir la versió anterior a l'error i poder tornar a treballar a partir d'aquesta versió provocant que l'error no s'incorpori a la base de dades.

A més he après diferent legislació sobre el tractament d'aigües residuals i sobre quins tractaments han de tenir les depuradores segons l'aigua que reben o depenen del punt de descàrrega on desemboquen les aigües residuals.

També he pogut aplicar diferents coneixements obtinguts a la universitat com el plantejament de diferents estratègies per tal de la resolució dels problemes i poder realitzar un anàlisi crític per tal d'escollir l'estratègia més adient en cada cas.

AGRAÏMENTS

Agrair en Lluís Corominas i en Lluís Bosch pel tracte rebut durant la realització de les pràctiques i l'ajuda que m'han proporcionat en el cas de tenir qualsevol dubte. Agrair també a tots els membres d'ICRA per la disposició del material necessari per la realització de les pràctiques.

BIBLIOGRAFIA

- Agency, E. E. (7 / Octubre / 2019). *Waterbase - UWWTD: Urban Waste Water Treatment Directive – reported data*. Recollit de <https://www.eea.europa.eu/data-and-maps/data/waterbase-uwwtd-urban-waste-water-treatment-directive-6>
- Carme Font, F. B. (11 gener 2019). GLOBAL-FATE (version 1.0.0): A geographical information system (GIS)-based model for assessing contaminants fate in the. *Geoscientific model development*.
- Europeas, C. (21 / Maig / 1991). *Directiva del Consejo, de 21 de mayo de 1991, sobre el tratamiento de las aguas residuales urbanas*. Recollit de <https://www.boe.es/buscar/doc.php?id=DOUE-L-1991-80646>
- l'Eau, O. I. (2017). *European Commission urban waste water website: Germany*. Recollit de <https://uwwtd.eu/Germany/uwwtps/treatment>