

**Títol del treball:**

**Anàlisi de dades de potabilització d'aigua per estimar el risc  
microbiològic de l'aigua potable**

---

Estudiant: Ester Aguilera Moncunill

Grau en Ciències Ambientals

Correu electrònic: esaguilera29@gmail.com

Tutor: Hèctor Monclús

Cotutor\*: Lluís Godo

Empresa / institució: LEQUIA

Vistiplau tutor (i cotutor\*):

Nom del tutor: Hèctor Monclús Sales

Nom del cotutor\*: Lluís Godo Pla

Empresa / institució: LEQUIA

Correu(s) electrònic(s): hector.monclus@udg.edu;

lluis.godo@udg.edu

\*si hi ha un cotutor assignat

Data de dipòsit de la memòria a secretaria de coordinació:

## RESUM

Els brots de malalties transmises per l'aigua segueixen sent un problema de salut pública en el segle XXI sobretot en els països en desenvolupament, però també en els països desenvolupats.

Les estacions de tractament d'aigua potable (ETAP) s'encarreguen de garantir que l'aigua de consum tingui una bona qualitat tant microbiològica com fisicoquímica, de manera que, entre altres coses, minimitzen el risc de patir una malaltia infecciosa. No obstant, descompensacions en el tren de tractament, condicions meteorològiques extremes o contaminacions durant la distribució de l'aigua poden augmentar el risc microbiològic i provocar brots epidèmics.

Per quantificar el risc microbiològic es pot aplicar la metodologia QMRA (Quantitative Microbiological Risk Assessment). Aquesta consisteix en l'aplicació de quatre passos on s'exploren diversos aspectes com els tipus de microorganismes patògens, les característiques d'entrada de l'aigua, el funcionament del tren de tractament, o dades epidemiològiques de la població.

En el present treball s'analitza el risc microbiològic de les aigües tractades per l'ETAP del Ter seguint la metodologia QMRA.

Primerament s'ha desenvolupat un algoritme de classificació per estimar la concentració microbiològica d'entrada de l'ETAP segons les condicions ambientals monitoritzades diàriament. L'algoritme s'ha desenvolupat utilitzant grans volums de dades (dades diàries de sèries històriques de 20 anys), pel qual s'han necessitat eines de mineria de dades, o en anglès, *Data Mining*.

Finalment, les concentracions estimades d'entrada s'han utilitzat per calcular un dels paràmetres més comuns a l'hora de valorar el risc microbiològic: l'AVAD (Anys de Vida Ajustats per Discapacitat) o DALY en les seves sigles en anglès. Aquest paràmetre s'ha aplicat seguint les directrius dels manuals de la OMS i de MicroRisk. El DALY ha permès quantificar en un rang de valors el risc i visualitzar quan es donen les condicions necessàries en les quals es podria donar una situació perillosa.

El risc microbiològic de les aigües tractades per l'ETAP Ter ha resultat ser baix en la gran majoria de dies, però s'han detectat alguns escenaris on el risc superaria el llindar marcat per la OMS i, per tant, on la gestió ha de ser més acurada.

## RESUMEN

Los brotes de enfermedades transmitidas por el agua siguen siendo un problema de salud pública en el siglo XXI sobretodo en los países en desarrollo, pero también en los países desarrollados.

Las estaciones de tratamiento de agua potable (ETAP) se encargan de garantizar que el agua de consumo tenga una buena calidad tanto microbiológica como fisicoquímica, de manera que, entre otras cosas, se minimiza el riesgo de sufrir una enfermedad infecciosa. No obstante, descompensaciones en el tren de tratamiento, condiciones meteorológicas extremas o contaminaciones durante la distribución del agua pueden aumentar el riesgo microbiológico y provocar brotes epidémicos.

Para cuantificar el riesgo microbiológico se puede aplicar la metodología QMRA (Quantitative Microbiological Risk Assessment). Esta consiste en la aplicación de cuatro pasos donde se exploran diferentes aspectos como el tipo de microorganismo patógeno, las características de entrada del agua, el funcionamiento del tren de tratamiento, o datos epidemiológicos de la población.

En el presente trabajo se analiza el riesgo microbiológico de las aguas tratadas por la ETAP del Ter siguiendo la metodología QMRA.

Primeramente, se ha desarrollado un algoritmo de clasificación para estimar la concentración microbiológica de entrada de la ETAP según las condiciones ambientales monitorizadas diariamente. El algoritmo se ha desarrollado utilizando grandes volúmenes de datos (datos diarios de series históricas de 20 años), por el cual se ha necesitado usar herramientas de minería de datos, o en inglés, *Data Mining*.

Finalmente, las concentraciones estimadas de entrada se han usado para calcular uno de los parámetros más comunes cuando se valora el riesgo microbiológico: el AVAD (Años de Vida Ajustados por Discapacidad) o DALY en sus siglas en inglés. Este parámetro se ha aplicado siguiendo las directrices de los manuales de la OMS y de MicroRisk. El DALY ha permitido cuantificar en un rango de valores el riesgo y visualizar cuando se dan las condiciones necesarias en las cuales se podría dar una situación peligrosa.

El riesgo microbiológico de las aguas tratadas por la ETAP Ter ha resultado ser bajo en la mayoría de días, pero se han detectado algunos escenarios donde el riesgo superaría el lindar marcado por la OMS, y por lo tanto, donde la gestión tiene que ser más precisa.

## ABSTRACT

Outbreaks of waterborne diseases still are a public health problem in the 21st century, especially on developing countries, but also in developed countries.

Drinking Water Treatment Plants (DWTP) are in charge of guaranteeing that tap water has a good quality (in terms of both microbiological and physicochemical aspects), so that, among other things, they minimize the risk of suffering an infectious disease. Nevertheless, perturbations during the treatment process, extreme meteorological conditions or tap water pollution during its distribution can trigger microbiological risk and can provoke epidemic outbreaks.

To quantify microbiological risk QMRA (Quantitative Microbiological Risk Assessment) methodology can be applied. QMRA is a four steps methodology that explores several aspects as: type of pathogenic microorganisms, inlet water characteristics, treatment operation, epidemiological data, etc.

In the present report microbiological risk of tap water treated by DWTP Ter is analyzed according to the QMRA methodology.

First of all, a classification algorithm has been developed to estimate the microbiological concentration on the DWTP inlet by using different environmental conditions daily monitored. The algorithm has been developed using big volumes of data (daily data from historical series of 20 years), and for this reason it was necessary to use *Data Mining* tools.

Finally, estimated microbiological concentrations have been used to calculate one of the most common metrics used for quantifying microbiological risk: DALY (Disability Adjusted Life Years). This metric has been calculated following the WHO's and Microrisk's manuals directrices. DALY has allowed to quantify the risk in a range of values and has allowed to visualize under which conditions a dangerous situation could arise.

Microbiological risk calculus for tap water treated in DWTP Ter has been resulted to be low in most of the cases, but some scenarios where the risk would overcome the WHO's limit have been detected. Special focus should be paid on these scenarios, where water treatment should be more accurate.

## AGRAÏMENTS

Vull agrair especialment al meu tutor Lluís per tants divendres dedicats a tirar endavant la part pràctica d'aquest treball, i al meu tutor Hèctor per la confiança dipositada en mi; sense vosaltres aquest treball no seria el mateix.

Gràcies també a la meva família, en especial a la meva mare i en Raül, per estar sempre presents i animar-me a seguir endavant.

Per últim i no menys important, gràcies a les meves companyes de grau, que heu fet bonic aquest final d'etapa tant atípic.

# ÍNDEX

1	INTRODUCCIÓ .....	1
1.1	Funcionament d'una ETAP .....	1
1.2	Legislació vigent.....	3
1.3	Eliminació de microorganismes en les ETAPs .....	3
1.3.1	Indicadors microbiològics.....	3
1.3.2	Eliminació logarítmica .....	4
1.3.3	Operacions unitàries implicades en l'eliminació de microorganismes .....	5
1.4	Brots de malalties transmeses per l'aigua.....	6
1.5	Quantitative Microbiological Risk Assessment (QMRA) .....	6
1.5.1	Identificació del perill .....	7
1.5.2	Avaluació de l'exposició .....	8
1.5.3	Avaluació de l'efecte .....	8
1.5.4	Caracterització del risc (DALY).....	9
1.6	Motivació i hipòtesi d'estudi .....	9
2	OBJECTIVES .....	10
3	METODOLOGIA .....	11
3.1	Cas d'estudi .....	11
3.2	Obtenció de dades .....	11
3.3	Processament de dades .....	13
3.3.1	Variables.....	13
3.3.2	Correlacions entre variables.....	13
3.3.3	Algoritme Naive Bayes Classifier .....	14
3.3.4	Càlcul del risc microbiològic .....	14
3.4	Programari utilitzat.....	15
3.5	Planificació metodològica .....	15
4	RESULTATS I DISCUSSIÓ .....	16
4.1	Correlacions entre variables.....	16
4.1.1	PCA.....	16
4.1.2	Anàlisi de correlacions .....	17
4.2	Algoritme Naive Bayes Classifier .....	18
4.2.1	Definició dels escenaris segons la distribució de <i>Clostridium perfringens</i> .....	18
4.2.2	Model escollit .....	19
4.2.3	Aplicació de l'algoritme al cas d'estudi .....	22
4.3	Càlcul del risc microbiològic .....	23
4.3.1	Avaluació de l'exposició .....	24
4.3.2	Avaluació de l'efecte .....	25
4.3.3	Avaluació d'escenaris de tractament en el risc microbiològic .....	25
5	CONCLUSIONS.....	27
6	ÈTICA I SOSTENIBILITAT.....	28
7	BIBLIOGRAFIA.....	29

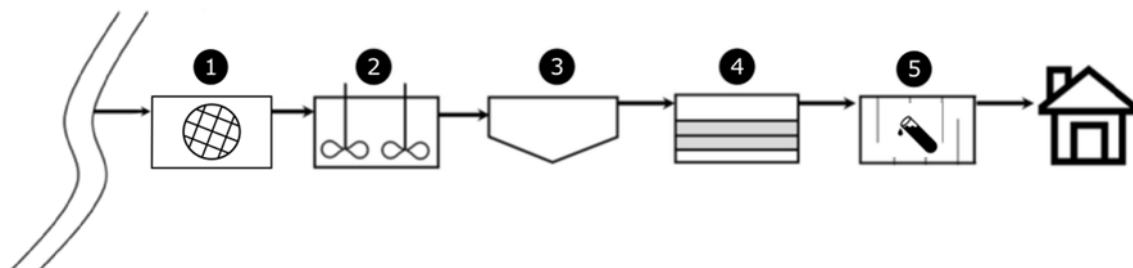
# 1 Introducció

L'aigua és un dels recursos naturals més importants pel manteniment de la vida en la Terra, i com a tal, garantir una bona qualitat i gestió d'aquest recurs en un context d'expansió demogràfica i de canvi climàtic és un dels grans reptes del segle XXI (Montaño, 2019).

## 1.1 Funcionament d'una ETAP

L'aigua captada del medi ambient és tractada en les Estacions de Tractament d'Aigua Potable (ETAP), on passa per diferents etapes abans que es consideri aigua apta pel consum humà. Per aconseguir-ho, cal eliminar-ne la matèria en suspensió, la matèria orgànica, nutrients, metalls, sals i organismes patògens (Comas, 2020).

Com que la composició de l'aigua d'entrada varia segons la font (rius, embassaments, mar, aigües subterrànies, etc.), els processos unitaris que es duen a terme en una ETAP no són sempre els mateixos. No obstant, hi ha un seguit d'operacions unitàries que estan presents àmpliament a la majoria d'ETAPs (Figura 1).



**Figura 1.** Operacions unitàries d'una ETAP convencional. 1: Reixes o tamisos. 2: Flocculació – Coagulació. 3: Decantació. 4: Filtració. 5: Desinfecció. Adaptat de: Godo-Pla (2020).

### 1-Reixes / Tamisos

Aquest és el primer procés unitari de l'ETAP, i consisteix en separar les matèries voluminoses arrossegades per l'aigua gràcies a la presència de reixes o tamisos (segons la mida de la llum) a l'entrada de l'ETAP.

### 2-Coagulació - flocculació

En aquesta operació unitària s'eliminen les partícules col·loïdals (d'un diàmetre de  $10^{-3}$  a  $1 \mu\text{m}$ ) que degut al seu baix pes no sedimenten. Amb l'addició d'un coagulant carregat positiva

o negativament s'aconsegueix desestabilitzar els col·loides, reduint les forces que les mantenen en suspensió.

A continuació s'afegeix un floculant que aglomera els col·loides desestabilitzats formant floculs que són prou grans i pesants com per sedimentar de forma molt més ràpida del que passaria sense l'addició de ningun reactiu. En ambdós casos l'aigua es troba en agitació per tal de posar en contacte més eficientment el coagulant i el floculant amb les partícules (Crittenden et al., 2012).

### 3-Decantació / Sedimentació

La decantació consisteix en deixar que les partícules en suspensió i els floculs de l'aigua provinent de la coagulació – floculació vagin sedimentant per acció de la gravetat.

### 4-Filtració

Durant aquest procés es fa passar l'aigua clarificada provinent de la decantació a través d'un medi porós que reté els possibles sòlids que encara quedin a l'aigua. La filtració es pot dur a terme amb diferents materials: filtres de sorra, carbó actiu, mòduls d'ultrafiltració i filtres de pressió, entre d'altres (IWA, 2021).

### 5-Desinfecció

Per últim, es duu a terme un procés de desinfecció per tal d'eliminar els possibles microorganismes que conté l'aigua i que podrien suposar un perill per a la salut. La desinfecció sol actuar típicament com una barrera final per tal d'assegurar que l'aigua està lliure de patògens (Asami et al., 2016).

Aquesta desinfecció es pot fer mitjançant processos químics, com l'addició de clor, o processos físics, com l'aplicació de raigs ultraviolats (UV). La desinfecció més habitual és la cloració ja que és una opció segura i econòmicament barata (Tandukar et al., 2020). Té l'inconvenient, però, que es creen secundàriament compostos organoclorats (DPBs, de l'anglès *disinfection by-products*) quan reaccionen amb la matèria orgànica, i aquests poden ser perjudicials per a la salut (Aguirre et al., 2004).



## 1.2 Legislació vigent

Els criteris sanitaris que han de complir les aigües de consum humà la marca la legislació vigent, que en el cas de l'Estat Espanyol, és el Real Decreto 140/2003, de 7 de febrer. El criteri de potabilitat de l'aigua depèn fonamentalment de l'ús al que se la destina (humà, industrial, agrícola, etc.). Una bona qualitat de l'aigua ha de reunir les següents característiques (Casero, 2008):

- No contenir substàncies ni microorganismes que siguin perillosos pels consumidors.
- No contenir substàncies que comuniquin sensacions sensorials desagradables pel consum (color, terbolesa, olor, sabor).

En quant a paràmetres biològics, el Real Decreto 140/2003 estableix que l'aigua potable ha de contenir:

- 0 UFC/100 mL d'*Escherichia coli*,
- 0 UFC/100 ml d'enterococs,
- 0 UFC/100mL de *Clostridium perfringens* (incloses les espores),
- 0 UFC/100mL de bacteries coliformes i
- 100 UFC/mL de microorganismes cultivables a 22°C.

A l'entrada de l'ETAP les concentracions de microorganismes no solen ser nul·les, i tenen una alta variabilitat segons les condicions ambientals. Això comporta que en l'ETAP s'hagin de fer tractaments de desinfecció per garantir el compliment del Real Decreto 140/2003. Les operacions unitàries de l'ETAP eliminen d'una forma més o menys efectiva els diferents tipus de microorganismes.

## 1.3 Eliminació de microorganismes en les ETAPs

### 1.3.1 Indicadors microbiològics

Les diferents fonts d'aigua que s'utilitzen en les ETAPs contenen una alta biodiversitat microbiològica. En termes pràctics, ni és possible mostrejar totes les espècies que pugui contenir una aigua, ni és necessari per fer operar les potabilitzadores. No tots són infecciosos ni produeixen les mateixes malalties. És per això que s'utilitzen unes espècies microbianes concretes que serveixen com a indicadors de grans grups microbians (WHO, 2009).

Els microorganismes indicadors s'utilitzen com a subrogats per avaluar la presència de patògens en l'ambient. Aquestes espècies indicadores solen ser els patògens més difícils d'eliminar i/o que tenen més incidència en la població. A més, es prefereix que els indicadors no es reproduueixin en l'aigua, i que siguin detectables a baixes concentracions (Motlagh i Yang, 2019). A la Taula 1 es mostren els patògens indicadors més habituals pels diferents tipus de microorganismes patògens.

**Taula 1.** Patògens indicadors dels grans grups de microorganismes aquàtics (ordenats de major mida a menor). Adaptat de: (Medema et al., 2006; WHO, 2009).

Espècie	
<b>Protozous</b>	<i>Entamoeba histolítica, Cryptosporidium parvum, Giardia intestinalis</i>
<b>Bacteris</b>	<i>Campylobacter sp.</i> , Coliformes fecals ( <i>Escherichia Coli, Enterococcus sp.</i> )
<b>Virus</b>	Enterovirus, Norovirus

Segons quin tipus de microorganisme hi hagi a l'aigua (protozous, bacteris o virus) s'haurà d'aplicar un tractament o un altre per eliminar-los.

### 1.3.2 Eliminació logarítmica

Per comparar l'eficiència d'eliminació de les diferents operacions unitàries d'una ETAP s'utilitza el valor "eliminació logarítmica" (LRV, de l'anglès, Log removal values). Aquest és un valor que relaciona la concentració d'entrada i la concentració de sortida d'un microorganisme en una etapa concreta (Eq.1) (Water Research Australia, 2014).

$$LRV = \log_{10} \frac{\text{Concentració microorganisme influent}}{\text{Concentració microorganisme efluent}} \quad (\text{Eq. 1})$$

El fet que la fórmula sigui logarítmica suposa que un LRV=1 vol dir que s'ha eliminat el 90% dels microorganismes d'entrada, un LRV= 2 és equivalent al 99% d'eliminació, un LRV = 3 al 99,9% d'eliminació, etc.

El valor de l'LRV es calcula igual independentment de la forma com s'han eliminat els microorganismes, ja sigui de forma física (filtració, coagulació – floculació, UV, decantació, etc.) o de forma química (cloració, ozonització...) (Water Research Australia, 2014). A continuació s'exposen les operacions unitàries d'una ETAP més importants on hi ha eliminació de microorganismes.

### 1.3.3 Operacions unitàries implicades en l'eliminació de microorganismes

- **Filtració:** Els filtres de sorra solen ser crucials per eliminar els protozous com *Cryptosporidium sp.* o *Giardia sp.*, arribant fins als 2 logaritmes d'eliminació. Els bacteris també hi queden retinguts, però acaben de ser inactivats amb la cloració (Fujioka et al., 2019), en canvi, els virus són els microorganismes més permeables als filtres degut a la seva mida.

La filtració per membranes, que és una tecnologia més avançada, pot eliminar més de 4 logs (>99,99%) en el cas del gènere *Cryptosporidium* (WHO, 2009).

- **Coagulació - Floculació - Decantació:** Els microorganismes funcionen com partícules: com més gran sigui l'agrupació o el propi microorganisme, més pes tindrà i per tant més ràpid sedimentarà.
- **Desinfecció per clor:** La cloració elimina molt bé els bacteris i virus; però els protozous (i els seus oòcits i cists) en són molt resistents (CDC, 2012; WHO, 1996).
- **Desinfecció per UV:** L'aplicació d'UV un període de temps concret és molt efectiva per eliminar protozous (WHO, 2016).

A mode de resum, la Taula 2 mostra el grau d'eliminació que poden tenir les diferents operacions unitàries pels diferents grups microbiològics.

**Taula 2.** Grau d'eliminació qualitatiu de protozous, bacteris i virus per diferents operacions unitàries. Nota: els protozous són els individus més grans i els virus els més petits. Adaptat de: (CDC, 2012; Hijnen et al., 2006; WHO, 2016).

	<b>Filtració</b>	<b>Coagulació - Floculació - Decantació</b>	<b>Desinfecció física (UV)</b>	<b>Desinfecció química (Cl)</b>
<b>Protozous</b>	Alt	Alt	Alt	Baix
<b>Bacteris</b>	Mig	Mig	Alt	Mig - Alt
<b>Virus</b>	Baix	Baix	Mig	Alt

És important que les ETAPs combinin diferents d'aquestes etapes perquè tal com s'observa en la Taula 2, els diferents microorganismes són més o menys resistents a les diferents operacions unitàries. Així, els virus solen ser el tipus de microorganisme més persistents degut a la seva mida i resistència (Tandukar et al., 2020), sobretot si no s'aplica una desinfecció

química (amb Clor) durant el tren de tractament. En canvi, tractaments on l'eliminació de microorganismes recaigui sobretot en la desinfecció per Clor tindran una persistència més alta de protozoos com per exemple del gènere *Cryptosporidium*.

#### **1.4 Brots de malalties transmeses per l'aigua**

S'ha demostrat que l'aigua pot ser un vehicle per múltiples enteropatògens bacterians, protozoos o vírics. Ingerir aigua amb enteropatògens causa típicament símptomes gastrointestinals en l'hoste com nàusees, vòmits i diarrea, podent ser alguns símptomes concrets més greus (Medema et al., 2006).

En països desenvolupats, la probabilitat de patir una malaltia transmesa per l'aigua és baixa. Per exemple, a Espanya, hi va haver una mitjana anual de 103 casos de Criptosporidiosis (infecció per *Cryptosporidium*, una de les possibles malalties transmeses per l'aigua) entre 1995 i 2003 (un 0,00024% de la població espanyola del 2003) (García et al., 2004). Tot i aquesta baixa probabilitat, com que la distribució d'aigua potable es fa a un grup tan gran de població, encara que es produeixin malalties a una proporció petita de consumidors aquestes poden suposar una amenaça important a la salut pública (Medema et al., 2006).

#### **1.5 Quantitative Microbiological Risk Assessment (QMRA)**

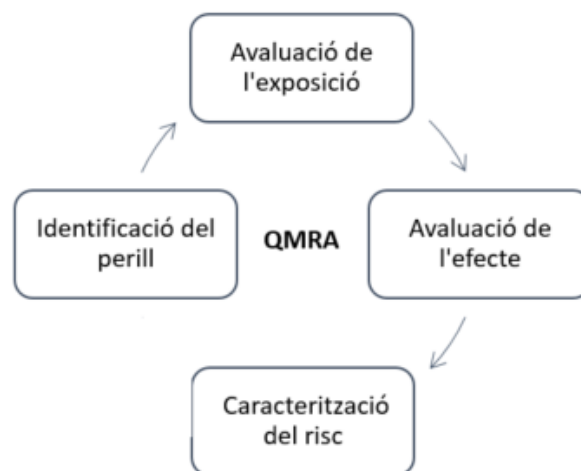
Durant el procés de potabilització i en l'etapa de distribució fins als habitatges poden produir-se incidents que poden provocar que l'aigua de consum presenti contaminació per la presència de microorganismes, la qual s'associa amb un important risc per a la salut pública (Valero, s.d.).

Aquesta contaminació microbiològica és difícil de monitoritzar i quantificar ja que els cultius microbiològics que es fan rutinàriament a les ETAPs solen fer-se només un cop a la setmana i es tenen els resultats a les 24 – 48 h, de manera que és probable que l'aigua potabilitzada ja s'hagi distribuït o inclús consumit abans de saber els resultats. A més, l'aïllament d'alguns patògens pot ser un procés complex i car (Medema et al., 2006).

Per garantir una bona qualitat microbiològica de l'aigua es poden combinar aquests mètodes amb altres aproximacions estadístiques com el "Quantitative Microbiological Risk Assessment" (QMRA). És important la combinació d'aquests dos mètodes, ja que, tot i que les dades obtingues a partir del QMRA poden ser diàries, les dades extretes de cultius microbiològics són més fiables que el risc calculat amb QMRA ja que no són probabilitats (Eregno, 2017).

El QMRA és una metodologia estadística per avaluar el risc microbiològic d'una aigua concreta que passa per un tractament de potabilització concret. És un procés de 4 passos del qual se'n obté un valor numèric final: el risc microbiològic. Els detalls de la metodologia es poden trobar a Medema et al. (2006) i WHO (2016).

En la Figura 2 es mostren les etapes que es realitzen per desenvolupar el QMRA.



**Figura 2.** Etapes de la metodologia QMRA

### 1.5.1 Identificació del perill

La primera fase del QMRA consisteix en establir quines són les qüestions específiques que es volen respondre i quins límits s'estableixen.

D'una banda, s'identifica un o diversos microorganismes que puguin causar malalties als humans. Normalment, s'utilitza un sol microorganisme anomenat "patogen índex", i se'n descriuen les seves característiques i la malaltia que desenvolupen.

D'altra banda, es descriuen les situacions perilloses (per exemple, pluges torrencials, o desajustos en la potabilització i/o contaminacions durant la distribució de l'aigua) on es podrien donar pics de concentració elevada dels patògens.

Al final d'aquesta etapa s'ha estimat la concentració del patògen índex a l'entrada de l'ETAP segons altres paràmetres fisicoquímics de l'aigua d'entrada. Es poden monitoritzar aquests paràmetres físics i químics de l'aigua per tal de ser utilitzats com a mesures subrogades que indiquin la presència de patògens, aprofitant que la qualitat microbiològica de l'aigua varia segons les condicions ambientals.

#### 1.5.2 Avaluació de l'exposició

En aquesta fase es quantifica la probabilitat d'ingerir patògens de l'aigua ja tractada. Per fer-ho, s'avaluen els canvis que pateix la concentració de patògens (estimada en la fase anterior) degut al procés de potabilització. Aquests canvis es quantifiquen gràcies al LRV.

En aquesta fase també s'avalua el volum mitjà d'aigua consumida per part de la població. Aquest valor pot variar segons la persona, per això se solen agafar valors conservatius, és a dir, valors per sobre de la mitjana per tal de tenir en compte el pitjor escenari possible.

Amb la concentració microbiològica estimada de l'aigua potable i el volum d'aigua ingerit, es calcula la dosi (o exposició): el nombre de patògens consumits per unitat de temps (Eq. 2).

#### 1.5.3 Avaluació de l'efecte

Un cop s'ha obtingut el valor de la dosi s'analitza com aquesta afecta a la població. Això es fa amb corbes dosi-resposta, on es relaciona la magnitud de la dosi, la infectivitat del patògen i la probabilitat que un individu causi infecció amb els efectes sobre la salut quantificats a una població exposada al patògen.

Aquesta informació s'obté d'estudis fets amb adults sans voluntaris, i això fa que la informació sigui limitada i poc extrapolable a altres grups poblacionals. Tampoc es tenen en compte efectes sinèrgics entre diferents patògens en els models QMRA actuals.

#### 1.5.4 Caracterització del risc (DALY)

En aquesta etapa es fa una recopilació de tota la informació obtinguda en els altres punts i es fa un càlcul del risc microbiològic de l'aigua. Això s'aconsegueix calculant els Anys de Vida Ajustats per Discapacitat (AVAD) o DALY en les sigles en anglès.

El DALY pondera cada efecte sobre la salut d'un agent concret per la seva severitat en una escala del 0 a l'1, sent l'1 el més greu, que normalment és la mort de la persona. Aquest valor es multiplica per la duració de l'efecte sobre la salut (quan es tracta de la mort, la duració és l'esperança de vida restant que li quedava a la persona) i pel nombre de persones en una població afectada.

El valor final es considera l'estimació de la càrrega de la malaltia atribuïble a un agent concret.

El fet d'utilitzar el DALY i no un altre indicador es fa perquè permet la comparació del seu valor amb altres riscos d'emmalaltir d'altres maneres, com per exemple el risc que suposa tenir una certa concentració de carcinògens en l'aigua.

El càlcul del DALY es pot fer de forma puntual, és a dir, agafant valors concrets de totes les variables i obtenir un sol valor; o fer-lo a través de l'anàlisi de Monte Carlo, on es té en compte la incertesa d'alguna de les variables gràcies a afegir tota una distribució. Per exemple, en comptes d'assumir que el volum d'aigua ingerit és 1 L, s'agafen rangs (ex. de 0,5 a 3 L) i la probabilitat de cada valor de volum. El resultat no és un sol valor del DALY, si no tota una distribució probabilística d'aquest paràmetre. És d'aquesta manera que es quantifica també la incertesa del propi valor final del risc.

### 1.6 Motivació i hipòtesi d'estudi

Davant del context de pandèmia actual i del context de canvi climàtic que s'està vivint, on es preveu que nous episodis pandèmics i que les condicions climàtiques extremes (ex. episodi de pluges torrencials) en els països mediterranis es tornin més freqüents, cal aprofundir i millorar els càlculs del risc microbiològic aplicats a casos reals perquè les ETAPs puguin preveure quan caldrà extremar la precaució per evitar problemes de salut pública.

Arran d'aquest context aquest estudi s'emmarca en el projecte DrinkIA, projecte de col·laboració entre l'Ens d'Abastament d'Aigües Ter-Llobregat (ATL) i la Universitat de Girona, per tal de sistematitzar i codificar la presa de decisions a les seves instal·lacions d'aigua potable.

La hipòtesi d'aquest treball proposa la utilització de paràmetres fisicoquímics subrogats obtinguts diàriament a l'entrada de l'ETAP per tal d'ajudar a millorar la quantificació del risc microbiològic d'una aigua potabilitzada segons un tren de tractament concret. A més, amb l'ús de subrogats s'obtindrà la informació a temps real pel fet de no dependre de cultius microbiològics, de manera que es podrà actuar de forma molt més ràpida davant de situacions d'estrès.

## 2 Objectives

The main objective of this TFG is to quantify the microbiological risk of a treated water from a study case by predicting the microorganism concentration at the DWTP (Drinking Water Treatment Plant) inlet in real time.

To achieve this goal, four sub-objectives have been defined:

- To analyse the distributions of microorganisms at the DWTP inlet from an historical serial.
- To determine which surrogates are better for estimating the microbiological concentration at the DWTP inlet.
- To develop an algorithm capable of determining the qualitative concentration of microorganisms according to the quantitative values of the chosen surrogates.
- To determine the risk associated with different predicted microorganism concentration scenarios.



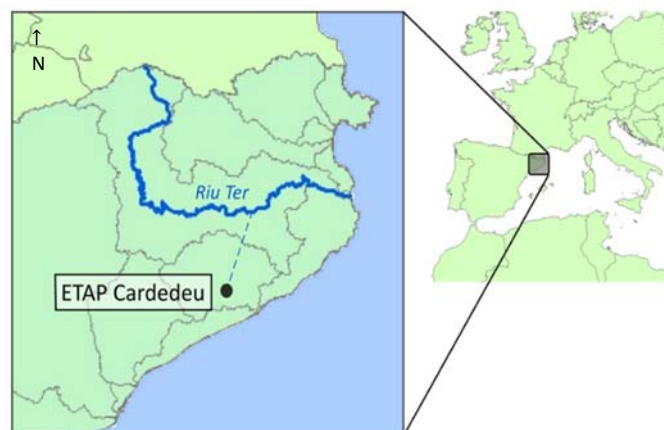
## 3 Metodologia

### 3.1 Cas d'estudi

L'anàlisi QMRA s'ha realitzat en un cas d'estudi concret: l'ETAP del Ter, localitzada a Cardedeu (província de Barcelona, Catalunya) (Figura 3).

L'ETAP del Ter capta l'aigua a la presa del Pasteral, l'últim embassament del complex Sau-Susqueda-Pasteral, i arriba a la planta mitjançant una galeria de 56 km de longitud. Té una capacitat de tractament de 8 m<sup>3</sup>/s i és capaç de subministrar aigua a una bona part de l'Àrea Metropolitana de Barcelona, arribant a un número potencial de consumidors de 4 milions d'habitants (Ens d'Abastament d'Aigua Ter-Llobregat, s.d.).

Les operacions unitàries que s'hi duen a terme són les pròpies d'un tractament convencional: es fa una desinfecció primària, després una coagulació/floculació, sedimentació, filtració amb carbó actiu granular (CAG) i finalment s'aplica un procés de desinfecció (Godo-Pla, 2020).



**Figura 3.** Localització de l'ETAP Cardedeu a Catalunya i el riu Ter. Mapa de conques hidrogràfiques. Elaboració pròpia.

### 3.2 Obtenció de dades

La majoria de dades utilitzades han sigut proporcionades per l'Ens d'Abastament d'Aigua Ter-Llobregat (ATL), que és l'empresa que gestiona l'ETAP Ter. Les variables Cabal de sortida del Pasteral i Volum de Sau i Susqueda s'han extret de l'Agència Catalana de l'Aigua (ACA), i la variable Precipitació (Estació Meteorològica del Pantà de Sau, codi KE) del Meteocat.

Les sèries històriques utilitzades, així com alguns dels paràmetres estadístics més habituals estan resumides en la Taula 3.

Taula 3. Dades proporcionades per ATL i ACA, i alguns valors estadístics d'interès.

Variable		Sèrie temporal	Lapse temporal entre mostres	Nº de dades	Mín	Mitjana ± SD	Màx	Unitats
Microorganismes	Total coliformes	05/03/2001 – 10/12/2019	Setmanal	3505	0	200 ± 600	4106	UFC/100 mL
	<i>Escherichia coli</i>	02/10/2000 – 30/12/2019	Setmanal	3657	0	8 ± 33	1733	UFC/100 mL
	<i>Clostridium perfringens</i>	02/10/2000 – 30/12/2019	Setmanal	3401	0	9 ± 12	275	UFC/100 mL
	<i>Cryptosporidium sp.</i>	05/03/2001 – 10/12/2019	Mensual	112	0	0,004 ± 0,02	0,2	Org/L
	<i>Giardia sp.</i>	05/03/2001 – 10/12/2019	Mensual	110	0	0,03 ± 0,18	1,334	Org/L
Variables fisicoquímiques / ambientals	Precipitació	01/01/2000 – 22/05/2020	Diari	7446	0	2 ± 7	164,7	mm
	Terbolesa	02/10/2000 – 30/12/2019	Setmanal	3985	0,11	1 ± 3	130	NTU
	Temps de residència hidràulic (TRH)	01/01/2000 – 18/01/2019	Diari	6958	57,2	268 ± 99	780,2	dies
	Cabal sortida Pasteral	21/06/2002 – 01/01/2020	Diari	5488	0	8 ± 12	369,5	m³/s
	Volum Sau i Susqueda	01/01/2000 – 01/01/2020	Diari	7306	81,1	270 ± 80	402	hm³

### 3.3 Processament de dades

#### 3.3.1 Variables

Les variables utilitzades en aquest treball es divideixen en dos grups: microorganismes i variables fisicoquímiques/ambientals (veure Taula 3).

- Microorganismes: les mostres es prenen a l'entrada de l'ETAP Ter cada setmana en el cas dels bacteris (Total coliformes, *E. coli*, *Clostridium perfringens*) o cada mes en el cas dels protozous (*Giardia sp.* i *Cryptosporidium sp.*).
- Variables fisicoquímiques/ambientals: mostres que es prenen diàriament a l'entrada de l'ETAP Ter (Terbolesa) o als embassaments del riu Ter (Precipitació, TRH, Cabal sortida Pasteral i Volum Sau i Susqueda). A més, amb aquestes variables també se n'han derivat d'altres com la Precipitació acumulada durant 7 dies, o el Canvi en el volum de Sau i Susqueda durant 7 dies.

Amb aquestes variables ambientals s'ha volgut estimar la concentració microbiològica, però no totes les variables s'han utilitzat en el model final.

#### 3.3.2 Correlacions entre variables

El primer pas per aplicar la metodologia QMRA ha sigut buscar relacions entre les variables ambientals i les concentracions de microorganismes. Això s'ha fet a través d'un Anàlisi de Components Principals (PCA) i un anàlisi de correlacions amb els coeficients de Pearson o de correlació (R). El PCA és un mètode estadístic que permet simplificar la complexitat i variabilitat de les dades en un nombre definit de dimensions anomenades components principals (PC). Cada PC té una variància explicada (eigenvalue) associat a ell, i s'ordenen de PC amb eigenvalue major fins al PC amb eigenvalue menor. Les variables es relacionen amb cada PC i entre sí d'una manera més o menys forta: com més propers es trobin els seus eigenvectors (els vectors en la representació gràfica), més relacionats estaran.

S'han exclòs les variables de protozous (*Cryptosporidium sp.* i *Giardia sp.*), ja que el nombre de mostres és molt inferior al de les altres variables (Taula 3).

Les dades s'han transformat utilitzant logaritmes neperians en les dades de Total Coliformes, *Clostridium perfringens*, *E. coli* i Terbolesa per tal que totes segueixin una distribució normal.

### 3.3.3 Algoritme Naive Bayes Classifier

El Naive Bayes Classifier és un algoritme d'aprenentatge supervisat capaç de classificar en diferents categories (variables discretes) a partir dels valors d'unes variables d'entrada (variables contínues). Aquesta classificació és probabilística. Com a algoritme supervisat, li calen unes dades d'entrenament per desenvolupar-lo (les sèries històriques de la Taula 3).

Està basat en el Teorema de Bayes, que s'utilitza per determinar la probabilitat d'una hipòtesi tenint en compte la seva distribució i les vegades que s'ha donat la hipòtesi en la sèrie històrica. És a dir, l'algoritme classificarà de forma més probable en les categories que més s'han repetit en la sèrie.

L'algoritme es diu Naïve (de l'anglès, ingenu) ja que assumeix que l'aparició d'una determinada característica és independent de l'aparició d'altres característiques; cada funció contribueix individualment a identificar que es tracta d'un dels components classificadors sense dependre les unes de les altres (MathWorks, 2021).

Pel cas d'estudi, s'ha entrat a l'algoritme les diferents variables ambientals disponibles per tal de predir diferents escenaris concrets de concentració microbiana.

### 3.3.4 Càlcul del risc microbiològic

Pel càlcul del risc microbiològic s'han seguit els 4 passos del QMRA exposats en la introducció. El risc s'ha quantificat amb el paràmetre DALY per les concentracions microbiològiques dels 3 escenaris contemplats en l'algoritme Naive Bayes Classifier.

S'ha calculat la dosi a partir de l'equació 2 extreta de l'estudi fet per Godo-Pla (2020).

$$Dosi = C \cdot \frac{1}{R} \cdot f \cdot 10^{-LRV} \cdot V \cdot I \quad (\text{Eq. 2})$$

On C és la concentració de microorganismes a l'entrada de l'ETAP (oòcits/100 mL), R és el percentatge de recuperació del mètode de detecció, f és la fracció de patògens capaços d'infectar, LRV són els logaritmes d'eliminació del microorganisme durant el tractament, V és el volum d'aigua consumit per persona i dia (L) i I és la infectivitat del microorganisme.

Amb l'equació 3 i el valor de la dosi es pot calcular el DALY (Godo-Pla, 2020).

$$DALY = (1 - (1 - (1 - e^{\tau \cdot Dosi}))^{365}) \cdot Ip \cdot DBF \quad (\text{Eq. 3})$$

On  $\tau$  és la probabilitat que un individu causi infecció,  $Ip$  és la probabilitat d'infecció degut a *Cryptosporidium sp.* i DBF és el Disease Burden Factor (factor de càrrega de la malaltia).

No s'ha escollit un valor puntual de l'LRV, si no que s'ha assumit que aquest paràmetre és variable segons com treballi la ETAP, i per tant, s'ha agafat tota la distribució probabilística d'aquesta variable. Així es quantifica també la incertesa associada al tractament.

### 3.4 Programari utilitzat

El tractament de dades s'ha fet combinant Excel versió 16.48, R Studio 3.6.1 i MATLAB R2020a. Les bases de dades es van rebre en format Excel, la PCA i l'anàlisi de correlacions s'ha fet amb R Studio i l'algoritme Naive Bayes Classifier i el càlcul del DALY s'ha fet amb MATLAB.

### 3.5 Planificació metodològica

Aquest estudi va començar el novembre de 2020, quan es va fer un primer tractament de les dades. Més endavant, es va desenvolupar l'algoritme Naive Bayes Classifier, pel qual es va tardar 2 mesos per perfeccionar-lo. Per últim, amb els escenaris possibles predits pel Naive Bayes Classifier, es va calcular el DALY de l'aigua potabilitzada a l'ETAP Ter (Figura 4). Durant 4-5 mesos s'ha fet una reunió setmanal o quinquennal per resoldre dubtes i marcar les tasques a fer.



**Figura 4.** Esquema de la planificació metodològica d'aquest TFG.

## 4 Resultats i discussió

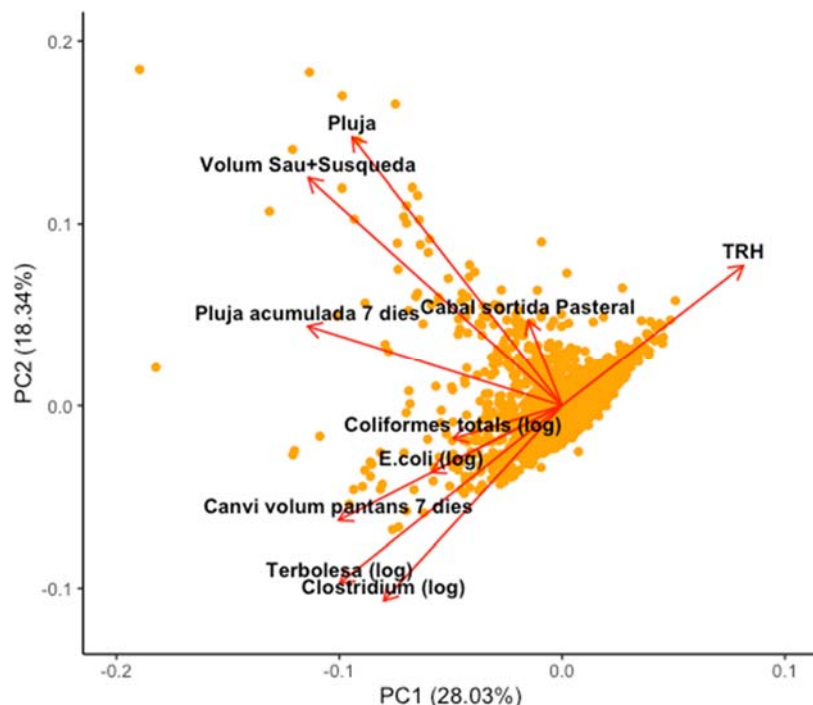
### 4.1 Correlacions entre variables

Per trobar correlacions entre les variables ambientals amb la concentració de microorganismes presents a l'entrada de l'ETAP s'ha fet un PCA i un anàlisi de correlacions.

#### 4.1.1 PCA

En la Figura 5 es mostren els resultats del PCA feta amb totes les variables ambientals i les variables microbiològiques (excepte protozous). Aquest PCA s'ha fet transformant les variables *E. Coli*, *Clostridium perfringens*, Total Coliformes i Terbolesa per tal que totes les variables segueixin una distribució gaussiana. El valor KMO és de 0,58 (valors menors a 0,5 indiquen que les dades no son bones per aplicar un PCA) (Gascón, 2019).

Els eixos del PCA mostrat a la Figura 5 expliquen el 46,3% de tota la variabilitat de les dades històriques: el primer component explica el 28,0% de tota la variabilitat i el segon component, el 18,3%.



**Figura 5.** Anàlisi de components principals (PCA) de les variables estudiades. Les variables Coliformes Totals, *E. Coli*, Terbolesa i *Clostridium perfringens* s'han transformat amb un logaritme neperià per tal que compleixin el supòsit de normalitat.

Els eixos del PCA mostrat a la Figura 5 expliquen el 46,3% de tota la variabilitat de les dades històriques: el primer component explica el 28,0% de tota la variabilitat i el segon component, el 18,3%.

Aquesta variabilitat no és molt elevada, però permet observar algunes relacions entre variables: la Terbolesa, *Clostridium perfringens* i Canvi del volum dels embassaments en 7 dies es troben relacionades de forma directament proporcional (els seus Eigenvectors es troben força propers), i el TRH dels embassaments es relaciona de forma inversament proporcional amb aquestes 3 variables. D'altra banda, les variables Pluja, Pluja acumulada en 7 dies i el Volum de Sau i Susqueda no tenen relació amb aquestes variables, ja que les seves fletxes es troben perpendiculars les altres.

#### 4.1.2 Anàlisi de correlacions

Els resultats de l'anàlisi de correlació es mostren en la Taula 4. Els resultats són similars als del PCA. Les variables més relacionades entre sí ( $R > |0,25|$ ) són: Terbolesa amb *Clostridium perfringens*, Canvi del volum a 7 dies dels embassaments amb *Clostridium perfringens*, TRH amb *Clostridium perfringens* i Cabal de sortida de Pasteral amb Total coliformes.

**Taula 4.** Anàlisi de correlació amb coeficients de Pearson (R) de les variables microbiològiques (columnes) i les variables ambientals (files). En negreta i subratllat, aquelles variables amb un coeficient  $R > |0,25|$ .

	<b>Coliformes totals</b>	<b><i>E. coli</i></b>	<b><i>C. perfringens</i></b>
<b>Pluja</b>	0,047	0,048	0,018
<b>Pluja acumulada en 7 dies</b>	0,146	0,171	0,110
<b>Volum de Sau i Susqueda</b>	0,091	0,104	0,105
<b>Canvi del volum dels pantans en 7 dies</b>	0,110	0,159	<b><u>0,270</u></b>
<b>Terbolesa</b>	0,166	0,221	<b><u>0,523</u></b>
<b>TRH</b>	-0,132	-0,092	<b><u>-0,322</u></b>
<b>Cabal de sortida Pasteral</b>	0,166	0,083	0,178

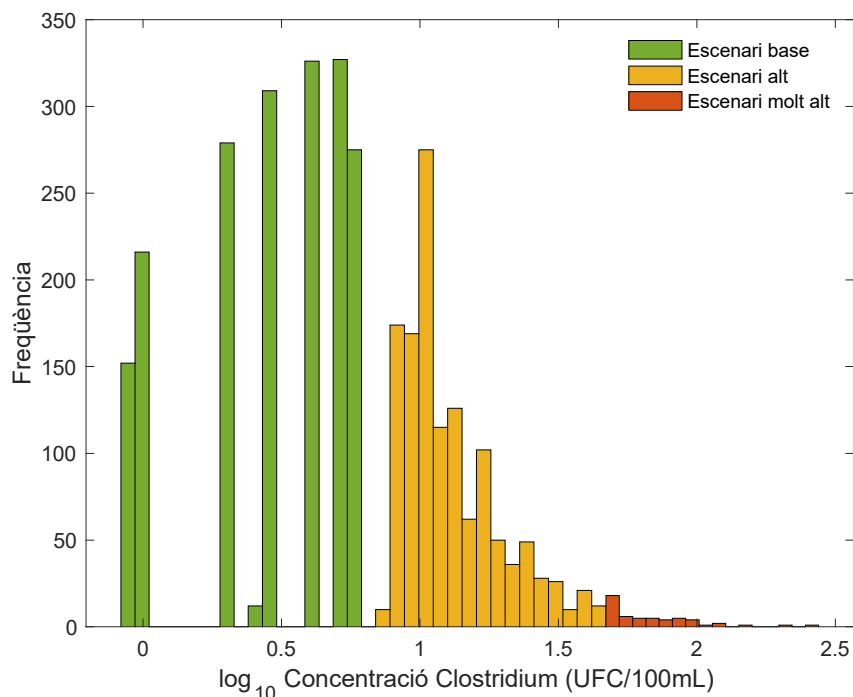
Les correlacions entre variables fisicoquímiques i microbiològiques en general són baixes. No obstant, ens permeten tenir una primera idea de quines seran les millors variables amb les quals treballar.

## 4.2 Algoritme Naive Bayes Classifier

En els resultats 4.1 s'observa que el microorganisme més correlacionat amb les variables ambientals és el *Clostridium perfringens*. És per això que l'algoritme s'ha desenvolupat per classificar les dades ambientals en un dels escenaris possibles d'aquest microorganisme.

### 4.2.1 Definició dels escenaris segons la distribució de *Clostridium perfringens*

L'algoritme Naive Bayes Classifier analitza variables contínues i les classifica en una variable discreta. En aquest cas, la variable discreta és la concentració de *Clostridium perfringens*. Les dades rebudes d'ATL, però, són quantitatives (nº d'UFC/100 mL). És per això que s'ha fet un pas previ de definició de diferents escenaris de *Clostridium perfringens* segons diferents rangs de concentració. Els rangs de concentració s'han definit observant la distribució històrica de *Clostridium perfringens* entre els anys 2000 – 2020 (n= 3391) (Figura 6).



**Figura 6.** Distribució de la concentració de *Clostridium perfringens* (transformat amb logaritme en base 10) entre els anys 2000 i 2020. Nota: les mostres on no hi ha hagut detecció (*Clostridium perfringens*=0) s'han transformat a 0,9 per tal que a l'hora de fer el logaritme aquestes dades no es perdessin (ja que  $\log_{10} 0 = \infty$ ) i quedessin properes a 0).

S'observa com la majoria de les dades es concentren a la banda esquerra de la figura, és a dir, les concentracions són majoritàriament baixes. Així, s'ha establert com a límit un primer escenari on les concentracions van de 0 a 7 UFC/100 mL (-0,08 fins a 0,84 en la Figura 6).



Aquest escenari de concentracions de *Clostridium perfringens* baixes se l'ha anomenat escenari base ja que és el més habitual (n=2083, 60% de les dades). Les dades restants (>7 UFC/100 mL; 0,84 a 1,672 en la Figura 6) s'han considerat concentracions de *Clostridium perfringens* altes. I d'aquestes s'ha volgut diferenciar un escenari molt alt d'aquells dies on la concentració ha sigut superior a 50 UFC/100 mL. Aquest escenari no s'ha triat pel seu volum de dades (n=36), si no per veure com afecten en el risc microbiològic les concentracions "extremes".

En el cas d'estudi, l'escenari de concentració més probable és el baix en la majoria dels dies analitzats degut a què l'origen de l'aigua a potabilitzar és un sistema d'embassaments (Sau i Susqueda) que fan un "pretractament natural" ja que actuen com grans decantadors.

#### 4.2.2 Model escollit

Arrel de l'estudi de correlacions i PCA, s'han escollit les variables ambientals "Canvi de volum dels pantans en 7 dies", "Terbolesa" i "TRH" com les variables més correlacionades i per tant més probables que siguin variables explicatives de la concentració de *Clostridium perfringens*.

A continuació, s'han organitzat les dades ambientals amb la variable categòrica de concentració (segons 4.2.1) en una sola matriu.

A fi de desenvolupar el model Naive Bayes Classifier, s'ha dividit el conjunt de dades de les sèries històriques de la matriu entre les dades de calibració (el 66% de les dades) i les dades de validació (34% de les dades). El conjunt de dades de calibració s'utilitza per entrenar l'algoritme: analitza les distribucions de cada variable ambiental i les freqüències de la variable resposta (concentració de *Clostridium perfringens*) per treure'n patrons. Les dades de validació serveixen per avaluar aquests patrons i veure com de bé prediu els escenaris de concentració (quins percentatges d'encert té l'algoritme).

Per optimitzar l'algoritme, s'han provat diferents combinacions de les variables ambientals (les 3 més correlacionades però també les altres), així com altres hiperparàmetres propis de l'algoritme, com els pesos assignats a l'error de cadascuna de les variables categòriques (el cost de mala classificació de l'algoritme, és a dir, quan aquest no encerta).

#### Combinació variables ambientals

Després de fer totes les combinacions possibles amb les dades disponibles, s'ha escollit finalment com a millor un model amb només les variables Terbolesa i TRH, ja que afegir més variables no ha millorat el percentatge d'encert significativament.

#### Cost de mala classificació

No tots els resultats de mala classificació s'han penalitzat de la mateixa manera: s'ha buscat penalitzar amb més contundència els errors produïts a l'hora de predir la concentració molt alta (ja que és on es produirà el risc microbiològic més alt), per tal que l'algoritme "prioritzi" equivocar-se a l'hora de classificar en concentracions baixes i no en concentracions molt altes. Això s'ha fet perquè una de les característiques del Naive Bayes Classifier és que té en compte el nombre de dades que conté cada escenari; per exemple, si l'escenari baix té més dades històriques, siguin quins siguin els valors de les variables ambientals l'escenari baix tindrà més possibilitats de ser predit. Canviant el cost de mala classificació s'evita aquest problema.

La variació del cost del model es troba representat a la Taula 5, on 0 és cap tipus de penalització ja que l'algoritme ha encertat la classe, i 1 és el màxim cost atorgat per l'error.

**Taula 5.** Costos atorgats a les diferents possibilitats de classificació del model.

Classe real	Molt alt	0	0,75	1
	Alt	0,5	0	0,75
	Baix	0,25	0,5	0
		Molt alt	Alt	Baix
		Classe predita		

En aquest cas (Taula 5), el cost màxim s'atorga quan l'algoritme prediu classe baixa i la real era molt alta ja que es considera l'error més greu aplicat en el cas real. Els altres costos s'han seguit de la forma més simètrica possible, tenint en compte que si l'algoritme prediu que la classe és molt alta i la real és baixa no es considera un error greu en termes pràctics (se li atorga un 0,25; 4 vegades menys de penalització que el cost de predir baix en un cas molt alt).

Variant el cost del model s'ha perdut una mica de % d'encert global, de gairebé el 70%, però s'ha aconseguit augmentar una mica l'encert de l'escenari molt alt.

## Resultat del model

Finalment, en la Figura 7 es mostra l'encert i l'error entre les dades que l'algoritme ha predit i les reals del model escollit.

Classe real	AA Molt Alt	11	20	5
	Alt	32	682	559
	Baix	6	428	1606
		AA Molt Alt	Alt	Baix
		Classe predita		

**Figura 7.** Comparació entre la classe predita (eix X) per l'algoritme i la real (eix Y) pel model que contempla com a variables ambientals la Terbolesa i el TRH. Els nombres són el nº de dades que s'han utilitzat per calibrar el model. En colors vermells-rosats es representen les males classificacions de l'algoritme, i en blau les classificacions correctes. Com més intens és el color més dades s'han encertat/errat. La mostra total és de 3349.

A mode de resum, la Taula 6 mostra el percentatge d'encert de cada categoria pel model Naive Bayes Classifier amb les variables ambientals de Terbolesa i TRH.

**Taula 6.** Percentatge d'encert per cada escenari i en global del algoritme Naive Bayes Classifier desenvolupat.

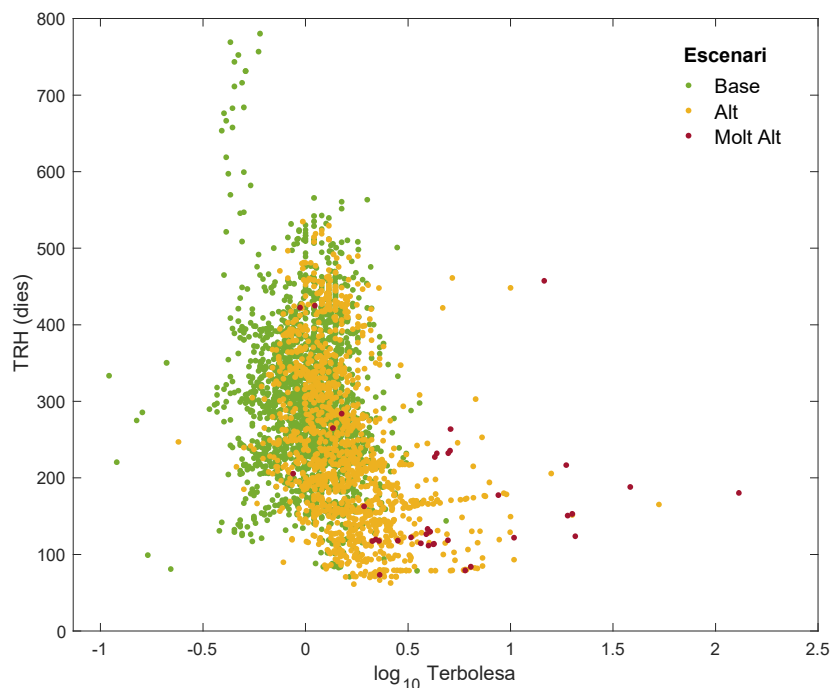
Escenari	Percentatge d'encert (%)
Base	74,01
Alt	60,35
Molt alt	22,45
<b>Total</b>	<b>68,65</b>

El percentatge d'encert global és bo (74,01%), però en el cas de l'escenari molt alt, la falta de dades i el fet de contemplar només dues variables ambientals fa que aquest sigui baix (22,45%).

#### 4.2.3 Aplicació de l'algoritme al cas d'estudi

L'algoritme desenvolupat permet predir l'escenari més probable de concentració de *Clostridium perfringens* a partir dels valors de terbolesa (en logaritme) i TRH del dia concret en què es vulgui mirar el risc microbiològic.

Com que s'ha escollit el model amb 2 variables ambientals, es pot representar en un eix X-Y el funcionament de l'algoritme (Figura 8): analitza la distribució de punts (els escenaris) en funció les variables ambientals, de manera que quan s'afegeix uns valors de les variables ambientals fa una extrapolació de l'escenari més probable.



**Figura 8.** Concentració de *Clostridium perfringens* qualitativa (escenari base, alt i molt alt) en funció de log<sub>10</sub> Terbolesa i TRH.

En la Figura 8 s'observa com molt dels punts dels diferents escenaris es superposen, però es pot observar lleugerament com els punts verds (escenari base) es troben preferentment quan la terbolesa és baixa i quan el TRH és alt. En canvi, com més alta és la terbolesa i com més baix és el TRH, normalment més alta és la concentració de *Clostridium perfringens*. Això s'explica perquè quan la terbolesa és alta aquells microorganismes que queden adherits als sediments passen a estar en suspensió (Medema et al., 2006). I quan el TRH és baix significa que la

retenció de les aigües en els pantans és baixa, és a dir, s'està desembassant. Aquest fet també provoca cabals més grans d'aigua que provoquen que augmentin les partícules en suspensió. Cal dir també, que els punts de l'escenari molt alt, que hipotèticament s'haurien de trobar quan la terbolesa és molt alta i el TRH molt baix, presenten molta dispersió, i és per això que, tot i haver previst un alt cost de mala classificació per l'escenari molt alt, el percentatge d'encert de l'algoritme per aquest escenari és baix, del 22% (veure Taula 6).

### 4.3 Càlcul del risc microbiològic

El risc microbiològic s'ha calculat aplicant la fase 2, 3 i 4 (avaluació de l'exposició, avaluació de l'efecte i caracterització del risc) de la metodologia QMRA amb dades majoritàriament bibliogràfiques.

El patògen índex utilitzat ha sigut *Cryptosporidium sp.* ja que és un dels microorganismes més difícils d'eliminar per desinfecció i dels més utilitzats en altres estudis d'aplicació del QMRA (WHO, 2009). Com que les dades de *Cryptosporidium sp.* són molt limitades (veure Taula 3), s'ha utilitzat el microorganisme *Clostridium perfringens* com a subrogat de *Cryptosporidium sp.* La concentració de *Clostridium perfringens* pot ser utilitzada com a espècie indicadora de la presència de *Cryptosporidium sp.* (Valero, s.d; WHO, 2017). La relació establerta entre aquests dos organismes en el cas d'estudi es considera de 1:1000 basat en les sèries històriques (1 oòcit *Cryptosporidium sp.*/L  $\approx$  0,001 UFC *Clostridium perfringens*/100 mL).

Els valors de concentració mitjana de *Cryptosporidium* per cada escenari definit anteriorment aplicant aquesta relació es troben resumits a la Taula 7.

**Taula 7.** Oòcits/100 mL de *Cryptosporidium sp.* per cada un dels 3 escenaris de concentració.

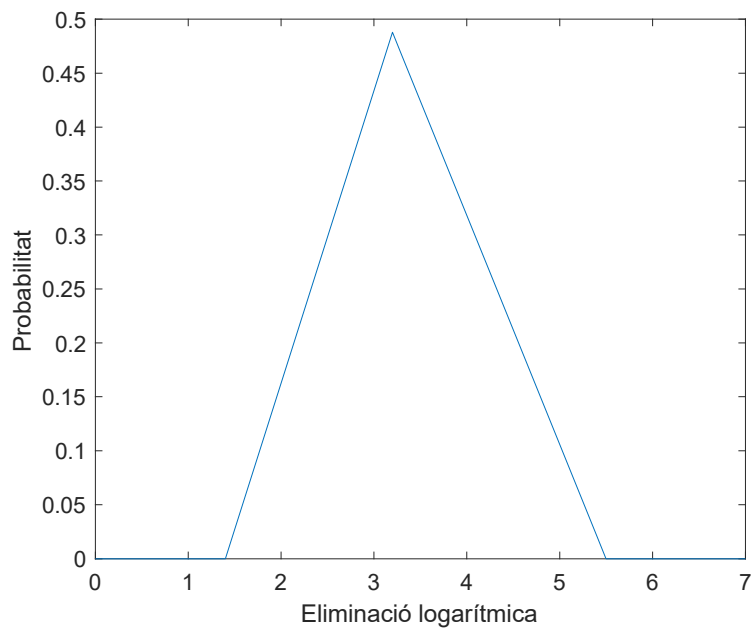
Concentració <i>Cryptosporidium sp.</i> (C)	Valor mitjà	Unitats
Escenari base (baix)	0,00366	
Escenari alt	0,01458	Oòcits/L
Escenari molt alt	0,08713	

#### 4.3.1 Avaluació de l'exposició

Aquest apartat ha consistit en calcular la dosi que pot rebre la població quan consumeix aigua segons els tres escenaris definits.

Per una banda, cal avaluar com disminueix la concentració de *Cryptosporidium sp.* degut al tractament de potabilització. Aquesta disminució, calculada amb logaritmes d'eliminació (LRV), no és sempre constant, ja que depèn de si la planta està funcionant a òptim rendiment o no. És per això que en comptes d'utilitzar un sol valor d'eliminació, s'utilitza tota la distribució probabilística obtinguda a partir de valors trobats en la bibliografia (Medema et al., 2006). D'aquesta manera es delimita la incertesa associada al tractament de potabilització.

Aquesta distribució probabilística és diferent per cada potabilitzadora, però en el cas de l'ETAP de Cardedeu, al tenir un tren de tractament convencional, s'han utilitzat dades bibliogràfiques conservatives de WHO (2009) de LRV de potabilitzadores convencionals (Figura 9).



**Figura 9.** Distribució probabilística triangular de l'eliminació microbiològica logarítmica d'una potabilitzadora convencional. Mín: 1,4; MEC: 3,2; Màx: 5,5.

D'altra banda, per calcular la dosi (Eq. 2), fa falta tenir en compte el volum d'ingesta d'aigua, el percentatge de recuperació del mètode, la infectivitat i la fracció de patògens capaços de produir una infecció.

Les dades que s'han utilitzat es troben recollides a la Taula 8 (WHO, 2009).

**Taula 8.** Dades bibliogràfiques utilitzades per calcular la dosi. Font: WHO (2009).

Variable	Paràmetre	Valor	Unitats
Consum d'aigua per persona i dia	V	2	L
Percentatge de recuperació del mètode de detecció	R	0,136	%
Infectivitat	I	0,9	-
Fracció de patògens capaços de produir una infecció	f	1	-

#### 4.3.2 Avaluació de l'efecte

En aquesta fase s'ha buscat la probabilitat que un individu de *Cryptosporidium sp.* causi infecció. Com s'ha introduït al principi del treball, aquesta probabilitat varia segons el grup de població que s'hagi exposat a l'aigua (segons edat, patologies prèvies, sexe, etc). La variabilitat present no s'ha tingut en compte i s'han agafat valors concrets.

Amb aquests paràmetres (Taula 9) s'ha pogut aplicar l'Eq. 3.

**Taula 9.** Dades bibliogràfiques per calcular el DALY. Font: (Havelaar i Melse, 2003; WHO, 2009)

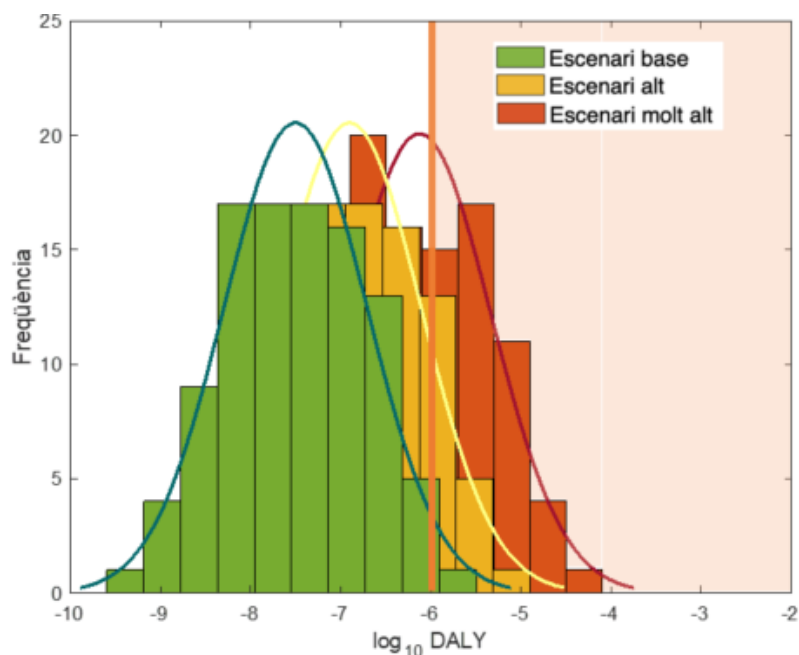
Variable	Paràmetre	Valor
Probabilitat que un individu causi infecció	$\tau$	0,004005
Probabilitat que un individu de <i>Cryptosporidium sp.</i> causi infecció	$I_p$	0,7
Factor de càrrega de la malaltia	Dbf	0,00169

#### 4.3.3 Avaluació d'escenaris de tractament en el risc microbiològic

Aplicant les Equacions 2 i 3 amb les Taules 7, 8 i 9 i els valors probabilístics de l'LRV (Figura 9), s'ha obtingut els valors de risc és més probables per cada escenari (Figura 10). La incertesa associada a aquest rang del risc ve donada únicament per la incertesa de l'eliminació dels microorganismes durant la potabilització (LRV). Així, dins d'un mateix escenari, els valors més alts del DALY es donen quan la potabilització no està al màxim rendiment, i viceversa.

**Taula 10.** Valors mitjans del DALY pels 3 escenaris.

Escenari	DALY mitjà
Base	$5,28 \cdot 10^{-8}$
Alt	$2,105 \cdot 10^{-7}$
Molt alt	$1,25 \cdot 10^{-6}$



**Figura 10.** Rangos més probables del risc per cada escenari. El risc (DALY) es mostra en logaritme en base 10. En taronja, els valors més grans que -6, que és el valor límit on la OMS considera que hi ha risc microbiològic. Distribució elaborada a partir de la generació de valors aleatoris de la distribució LRV.

En la Figura 10 i a la Taula 10 s'observa com quan major és la concentració de l'escenari, més alt és el rang del DALY. Així, quan la concentració de *Cryptosporidium sp.* és baixa (escenari base), gairebé mai es supera el límit de  $10^{-6}$  marcat per la OMS. En canvi, en l'escenari alt s'observen valors per sobre de  $10^{-6}$ , i en l'escenari molt alt, aquests valors arriben gairebé al 50% de les dades.

El fet que hi hagi episodis on la concentració a l'entrada sigui molt elevada no implica de forma directa que hi hagi risc pel consumidor, si no que el que reflecteix és que hi ha un risc associat a l'aigua d'entrada que obligaria a modificar el tractament de l'ETAP per garantir sempre els estàndards de qualitat de l'aigua potable. Tot i això en aquest treball s'han usat valors conservatius per calcular el DALY per tal d'avaluar escenaris d'alt risc i validar si la metodologia emprada seria suficientment sensible per identificar-los. Aquest estudi és la etapa prèvia d'un projecte que té com a objectiu usar la metodologia QMRA com a alerta anticipada per la gestió de les pròpies ETAPs i justificar així la modificació del seu tractament.



## 5 Conclusions

In this work, an algorithm was developed to predict the concentration of a microbiological indicator (*Clostridium perfringens*) at the inlet of a drinking water treatment plant (DWTP). The classification algorithm, based on a Naïve Bayes Classifier, showed good performance at predicting low concentrations (accuracy of 74,01%), but had lesser predictive accuracy at predicting very high concentrations (22,45%). The environmental variables used to train this algorithm had some limitations when used in this study case because of the presence of Sau – Susqueda reservoirs. For instance, rain variable (highly related to turbidity) could not be related to the microorganism concentration, because, even if it rained, water was retained in the reservoirs and didn't reflect an increase in water flow or turbidity in the DWTP inlet. That is why complimentary data would be needed to develop a more robust, efficient algorithm. The outcomes of the classification algorithm were used to calculate the microbiological risk of tap water, assuming normal operational conditions at the case study DWTP. The main conclusion of this work is that in the vast majority of the cases, the microbiological risk of tap water treated by Ter DWTP is below the WHO's limit and therefore, the microbiological safety of drinking water was demonstrated. Nevertheless, a combination of an extreme event (for instance, heavy rains, which are associated to high microbiological concentrations) and a bad management strategy during the water treatment (for instance, a by-pass of an operation unit), could lead to exceed the recommended values of risk, set by the WHO (DALY=10<sup>-6</sup>). These results highlight the importance to predict these extreme events and establish treatment strategies to avoid situations leading to high microbiological risk, because the potabilization process has to be as efficient as possible. Under extreme events, DWTP managers should anticipate and take preventive actions like retain the water in settlers for longer time, to put more filters into service, or increase the disinfectant dose.

## 6 Ètica i sostenibilitat

Amb aquest treball s'explora com varia el risc microbiològic de l'aigua potable en funció de la concentració microbiològica de l'aigua captada. L'aigua ha de ser un recurs públic a l'abast de tothom, i com a tal, se n'ha de garantir la màxima qualitat.

En un context de canvi climàtic, on s'espera que els fenòmens extrems augmentin la seva freqüència, i per tant on els dies de concentració microbiològica molt alta es facin més abundants, és important saber avançar-se a tals situacions per tal que les potabilitzadores treballin a ple rendiment i es minimitzi així un possible risc microbiològic a l'hora de beure aigua de l'aixeta.

A més, aquest treball s'ha desenvolupat de forma 100% online, de manera que no s'han produït desplaçaments ni a camp ni a laboratori.

## 7 Bibliografia

- Aguirre, P. A., García, J. i Mujeriego, R. (2004). Desinfección con cloro y luz UV en un proceso convencional de regeneración de agua. *Ingeniería del agua*, 11(1), 75-89.  
<https://doi.org/10.4995/ia.2004.2524>
- Asami, T., Katayama, H., Torrey, J. R., Visvanathan, C. i Furumai, H. (2016). Evaluation of virus removal efficiency of coagulation-sedimentation and rapid sand filtration processes in a drinking water treatment plant in Bangkok, Thailand. *Water Research*.  
<https://doi.org/10.1016/j.watres.2016.05.012>
- Casero, D. (2008). *Potabilización del agua. Módulo IV: Abastecimientos y Saneamientos Urbanos*. Máster en Ingeniería Medioambiental y Gestión del Agua 2007/2008, EOI.  
<https://static.eoi.es/savia/documents/componente45469.pdf>
- CDC. (2012). *Effect of Chlorination on Inactivating Selected Pathogen*. Recuperat 16 de maig 2021, de <https://www.cdc.gov/safewater/effectiveness-on-pathogens.html>
- Comas, J. (2019). *Tema 6. Potabilització de les aigües* [Apunts acadèmics]. UdGMoodle.
- Crittenden, J., Trussell, R., Hand, D., Howe, K. i Tchobanoglous, G. (2012). *MWH's Water Treatment: Principles and Design* (3rd ed.).  
[https://www.academia.edu/37374890/MWH\\_s\\_Water\\_Treatment\\_Principles\\_and\\_Design\\_Third\\_Edition](https://www.academia.edu/37374890/MWH_s_Water_Treatment_Principles_and_Design_Third_Edition)
- Ens d'Abastament d'Aigua Ter-Llobregat. (s.d.). *ETAP del Ter*. Recuperat 22 d'abril 2021 de [https://www.atl.cat/ca/etap-del-ter\\_3170](https://www.atl.cat/ca/etap-del-ter_3170)
- Eregno, F. E. (2017). *Quantitative Microbial Risk Assessment and Water Quality Modelling - On Drinking Water Sources, Recreational and Recycled Waters*. (Tesi doctoral, Norwegian University of Life Sciences, Noruega).  
<https://doi.org/10.13140/RG.2.2.24317.05608>
- Fujioka, T., Ueyama, T., Mingliang, F. i Leddy, M. (2019). Online assessment of sand filter performance for bacterial removal in a full-scale drinking water treatment plant. *Chemosphere*, 229, 509-514. <https://doi.org/10.1016/j.chemosphere.2019.04.197>
- García, A. M., Fernández, C., López García, C., Martos García, P. i Marín Casanova, P. (2004). Brotes epidémicos de Criptosporidiosis. *Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica (SEIMC)*, 1-10.  
<https://www.seimc.org/contenidos/ccs/revisionestematicas/parasitologia/Brotcripto.pdf>
- Gascón, S. (2019). *Mètodes d'ordenació indirectes* [Apunts acadèmics]. UdGMoodle.
- Godo-Pla, L. (2020). *Design and implementation of an environmental decision support system for the control and management of drinking water treatment plants*. (Tesi doctoral, Universitat de Girona, Catalunya). <http://hdl.handle.net/10256/18811>
- Havelaar, A.H, i Melse, J.M. (2003). Quantifying Public Health Risk in the WHO Guidelines for Drinking Water Quality. *WHO RIVM report*, 1-49.  
<https://www.ircwash.org/resources/quantifying-public-health-risk-who-guidelines->

[drinking-water-quality-burden-disease](#)

- Hijnen, W. A. M., Beerendonk, E. F. i Medema, G. J. (2006). Inactivation credit of UV radiation for viruses, bacteria and protozoan (oo)cysts in water: A review. *Water Research*, 40(1), 3-22. <https://doi.org/10.1016/j.watres.2005.10.030>
- IWA (2021). *Filtration processes*. Recuperat 12 de maig 2021 de <https://www.iwapublishing.com/news/filtration-processes>
- MathWorks, (2021). *Naive Bayes (R2021a)*. Recuperat el 26 de maig 2021 de <https://www.mathworks.com/help/stats/classificationnaivebayes.html>
- Medema, G., Loret, J. F., Stenström, T. A. i Ashbolt, N. (2006). Quantitative microbial risk assessment in the water safety plan. *Final Report on the EU MicroRisk Project*. <https://www.kwrwater.nl/wp-content/uploads/2016/09/MICRORISK-FINAL-REPORT-Quantitative-microbial-risk-assessment-in-the-Water-Safety-Plan.pdf>
- Montaño, B. (2019). El crecimiento de la población y la escasez hídrica. *Congreso Nacional del Agua 2019: innovación y sostenibilidad. Temática: economía del agua*, 509-519. <https://doi.org/10.14198/congreso-nacional-del-agua-orihuela-2019>
- Motlagh, A. M. i Yang, Z. (2019). Detection and occurrence of indicator organisms and pathogens. *Water Environment Research*, 91(10), 1402-1408. <https://doi.org/10.1002/wer.1238>
- Real Decreto 140/2003, de 7 de febrero, por el que se establecen los criterios sanitarios de la calidad del agua de consumo humano. Publicado en: «BOE» núm. 45, de 21/02/2003.
- Tandukar, S., Sherchan, S. P. i Haramoto, E. (2020). Reduction of Pathogenic and Indicator Viruses at a Drinking Water Treatment Plant in Southern Louisiana, USA. *Food and Environmental Virology*. <https://doi.org/10.1007/s12560-020-09436-y>
- Valero, F. (s.d.). El control microbiológico del agua de consumo ante el nuevo escenario legislativo.
- Water Research Australia. (2014). *Log removal values in wastewater treatment*. Recuperat 2 de maig 2021, de <https://www.waterra.com.au/publications/document-search/?download=949>
- WHO. (1996). *Fact Sheets on Environmental Sanitation*. <http://helid.digicollection.org/es/d/Js13461e/2.htm>
- WHO. (2009). *Risk Assessment of Cryptosporidium in Drinking Water*. <https://apps.who.int/iris/handle/10665/70117>
- WHO. (2016). *Quantitative Microbial Risk Assessment: Application for Water Safety Management*. <https://apps.who.int/iris/handle/10665/246195>
- WHO. (2017). *Guidelines for drinking-water quality: fourth edition incorporating the first addendum*. <https://www.who.int/publications/i/item/9789241549950>