DOCTORAL THESIS

Innovative Graph-Theory Solutions for the Future Urban Water Networks

David Martínez

2024

Universitat de Girona

DOCTORAL THESIS

Innovative Graph-Theory Solutions for the Future Urban Water Networks

David Martínez

2024

DOCTORAL PROGRAM IN TECHNOLOGY

Supervisor:
Ph.D. Eusebi Calle

This thesis is presented in fulfillment of the requirement for the conferral of the degree of Doctor of Technology by the University of Girona

El Dr. Eusebi Calle, de la Universitat de Girona,

DECLARO:

Que el treball titulat Innovative Graph-Theory Solutions for the Future Urban Water Networks, que presenta David Martínez Álvarez per a l'obtenció del títol de doctor, ha estat realitzat sota la meva direcció.

I, perquè així consti i tingui els efectes oportuns, signo aquest document.

EUSEBIO CALLE ORTEGA - DNI 77915254V

Firmado digitalmente por EUSEBIO CALLE ORTEGA - DNI 77915254V
Fecha: 2024.04.02 10:22:57 +02'00'

Signatura

Girona, 2 d'abrl de 2024

Dedicated to my partner,
who has supported me tirelessly
throughout this journey.

Acknowledgments

Realitzar una tesi doctoral és tant un viatge personal com professional. No és únicament la feina d'una sola persona, sinó més aviat un esforç col·lectiu que involucra totes aquelles persones que es troben pel camí. Per tant, la gratitud que expresso en aquesta pàgina d'agraïments l'estenc a tots aquells que ens hem trobat de manera casual i aquells que ens hem fet amics durant aquests tres anys. A tots vosaltres, gràcies!

Dedico un agraïment sincer a la meva família i especialment a la meva parella per aguantar, encoratjar i donar-me suport durant aquest llarg i, a vegades, incert viatge. El vostre suport incondicional ha estat increïble, i avui no estaria aquí, ni seria la persona que sóc, sense tots vosaltres. Gràcies!

También quiero dedicar un gran reconocimiento a mis abuelos. Han estado ahí apoyándome desde el primer día, especialmente mi abuela. Ella no tuvo la oportunidad de estudiar, y aun así siempre me ha animado a dar lo mejor de mí. Para ambos, ¡gracias!

Un segon gran agraïment és pel meu director de tesi, Eusebi Calle, a qui li dono les gràcies pel seu suport no només en l'àmbit professional sinó també personal. Gràcies per ajudar-me i guiar-me durant aquests anys i durant tot el meu procés acadèmic, des que vaig entrar a fer pràctiques en el grup de recerca BCDS, a tercer del grau d'Enginyeria Informàtica l'any 2018.

També m'agradaria agrair a tots els membres del grup de recerca BCDS per l'ajut incondicional que he rebut sempre i en especial a en Josep Lluís Marzo, que ha confiat en mi des del principi i m'ha permès continuar encoratjat la meva carrera acadèmica.

Un agraïment especial va dedicat als investigadors d'ICRA que han estat una peça clau per a la meva recerca. A partir d'un primer treball conjunt de col·laboració hem estat capaços d'iniciar una nova línia de recerca ajuntant els coneixements d'enginyeria informàtica del BCDS amb els seus coneixements d'aigua. Lluís, Quim, Josep i Gigi, moltes gràcies!

M'agradaria agrair també a en Marc Comas, professor del departament d'IMAE, que sempre s'ha mostrat disposat a ajudar-me i sempre m'ha resolt tots els dubtes estadístics que he tingut durant tot el camí.

I would like to also thank Professor Rob Kooij and Ph.D. candidate Fenghua Wang from Delft University of Technology (TU Delft) for their invaluable support in providing mathematical and telecommunication formulations and measures that significantly helped and contributed to this thesis. They graciously welcomed me during my visit to their research group facilities and were always willing to help with our research, laying the groundwork for some exciting collaborations.

Un altre gran agraïment el dedico als meus companys de laboratori que també estan realitzant la tesi doctoral i també a tots els becaris que han passat durant aquest llarg viatge. En Sergi i Miquel han estat uns grans companys de viatge i agraeixo molt tota l'ajuda i el suport emocional que he rebut d'ells durant tots aquests anys. No són només companys de doctorat, sinó també grans amics. Vull agrair a cadascun dels companys becaris que han format part del laboratori, on sempre hem mantingut un bon ambient i ens hem ajudat mútuament.

Finally, I would like to express my gratitude to all the entities and projects that provided financial support for this thesis. Special thanks are extended to the VirWASTE project (2020PANDE00044), funded by AGAUR, and the projects CLEaN-TOUR (CTM2017-85385-C2-1-R), INVEST (RTI2018-097471-B-C21), and ReUseMP3 (PID2020-115456RB-I00) funded by the Spanish Ministry of Economy and Competitiveness. Additionally, we acknowledge the ReWaT project (AEI-010500-2023-339) funded by the Spanish Ministry of Industry, Commerce and Tourism.

I would also like to extend a special acknowledgment to ABM Consulting for their support in providing economic data used by the algorithms, as well as to the Municipality of Girona and Cicle de l'Aigua del Ter S.A. for providing real water consumption data. Furthermore, my thanks go to the Municipalities of Girona, Lloret de Mar, and el Prat de Llobregat for their collaboration and support in testing our algorithms in real water network scenarios.

List of Acronyms

WWTP Wastewater Treatment Plant

COVID-19 COronaVIrus Disease of 2019

SARS-CoV-2 Severe acute respiratory syndrome coronavirus 2

WDNs Water Distribution Networks

SDN Software-Defined Networking

BCDS Broadband Communications and Distributed Systems

NRS Network Research Simulator

ANA Average Network Availability

PCR Polymerase Chain Reaction

CPU Central Processing Unit

AMD Advanced Micro Devices, Inc

RAM Random Access Memory

API Application Programming Interface

IGN Instituto Geográfico Nacional

DEM Digital Elevation Model

XML Extensible Markup Language

GraphML Graph Markup Language

QGIS Quantum Geographic Information System

CATSA Cicle de l'Aigua del Ter S.A.

MSE Monitoring Sites Evaluation

RNA RiboNucleic Acid

ANOVA ANalysis Of VAriance

SWMM Storm Water Management Model

ABS Àrea Bàsica de Salut

TPZs Tree Protection Zones

AS Australian Standard

IAs Impact Areas

CCTV Closed-Circuit TeleVision

ER Element Rearrangement

PVC PolyVinyl Chloride

REWATnet REclaimed WATer Network Tool

EPANET Environmental Protection Agency Network Evaluation Tool

WNTR Water Network Tool for Resilience

WA Water Availability

LB Limited Budget

UdG University of Girona

List of Tables

List of Figures

Related Author's Publications

This doctoral thesis has been structured in the form of a compendium of publications, comprising four distinct articles that have been submitted, accepted, and published. All publications are indexed in the Journal Citation Report database by Clarivate Analytics (JCR). The submission process for each article, along with their respective Impact Factors (IF), is outlined below:

(1) Calle, E.; Martínez, D.; Brugués-i-Pujolràs, R.; Farreras, M.; Saló-Grau, J.; Pueyo-Ros, J.; Corominas, L. Optimal selection of monitoring sites in cities for SARS-CoV-2 surveillance in sewage networks. *Environment International* **2021**, *157*, JCR Impact Factor (IF): **11.8, Q1** (Environmental Science, miscellaneous), 106768, DOI: https://doi.org/10.1016/j.envint.2021.106768.

(2) Calle, E.; Martínez, D.; Buttiglieri, G.; Corominas, L.; Farreras, M.; Saló-Grau, J.; Vilà, P.; Pueyo-Ros, J.; Comas, J. Optimal design of water reuse networks in cities through decision support tool development and testing. *NPJ Clean Water* **2023**, *6*, JCR Impact Factor (IF): **11.4, Q1** (Water Science and Technology), 23, DOI: https://doi.org/10.1038/s41545-023-00222-4.

(3) Martínez, D.; Bergillos, S.; Corominas, L.; Comas-Cufí, M.; Calle, E. Seamless integration of sewer system topology and tree location data: An algorithm to diagnose the potential impact of tree roots on pipes and propose rearrangement solutions. *Heliyon* **2024**, *10*, JCR Impact Factor (IF): **4, Q2** (Multidisciplinary Sciences), DOI: https://doi.org/10.1016/j.heliyon.2023.e23382.

(4) Martínez, D.; Bergillos, S.; Corominas, L.; Comas, J.; Wang, F.; Kooij, R.; Calle, E. Enhancing reclaimed water distribution network resilience with cost-effective meshing. *Science of the Total Environment* **2024**, *938*, JCR Impact Factor (IF): **9.8, Q1** (Environmental Science, miscellaneous), DOI: https://doi.org/10.1016/j.scitotenv.2024.173051.

Contents

Abstract

Urban water networks serve as lifelines for densely populated areas, facilitating the essential functions necessary for sustaining urban life. These critical infrastructures encompass water supply, distribution, and wastewater management, ensuring access to clean water for drinking, sanitation, and industrial purposes.

The critical role of urban water networks cannot be overstated, as they form the backbone of urban infrastructure. However, in recent years, water science has faced several urgent challenges that require attention. Two particular research areas include the necessity for wastewater surveillance to detect viruses and the optimization of resilient reclaimed water networks. These efforts are crucial for addressing the significant challenge of water scarcity exacerbated by climate change. Without critical infrastructures, cities would grind to a halt, facing severe repercussions on public health, economic activity, and overall quality of life.

In our rapidly evolving technological era, graph theory provides cutting-edge solutions to emerging challenges, primarily in telecom networks and high-tech fields. This thesis applies the principles of graph theory, which explores the study of graphs that are essential mathematical structures crafted to encapsulate pairwise relationships between objects composed of nodes and edges. Addressing challenges in urban water networks modeled as graphs, it aims to offer innovative strategies with significant societal impact amid COVID-19 and water scarcity. Graph theory's applications transcend disciplines, with urban water distribution networks defined by undirected graphs and wastewater networks by directed graphs.

The methodology in this thesis involves a straightforward yet efficient five-phase approach: bibliographic revision and literature gap identification, data acquisition, data preparation, algorithm development, and data analysis and processing of results. This methodology automated data gathering and processing, simplifying algorithm implementation, resulting in significant efficiencies and revealing notable benefits. Additionally, the cities of Girona and Lloret de Mar serve as primary real case studies

to test the developed algorithms and present their main findings.

First, the sewage monitoring site selection algorithm emerged as a relevant tool for pandemic management, meeting the original goal of optimal balance between coverage, interference, and maintenance hole distribution. This algorithm was developed based on prior experience with optimization and heuristic algorithms commonly used in telecom networks, including a foundation derived from an algorithm used to position backup controllers.

After sewage network monitoring, efforts focused on improving network resilience, starting with analyzing tree root impacts on wastewater networks. Substantial benefits were discovered through a quantitative assessment of pipe failure risk. Additionally, the tree rearrangement algorithm significantly decreased pipe failure risk and repair costs in Girona's network, even offering substantial economic savings despite initial investment requirements.

After applying graph theory in wastewater networks, two proposals for designing cost-effective, resilient, and hydraulically feasible reclaimed Water Distribution Networks (WDNs) emerged. These proposals efficiently compute optimal network graphs, delivering reclaimed water up to three times more efficiently than manual planning. Our algorithms evaluate and prioritize resilience, leading to significant cost savings over a 50-year period compared to conventional networks. Additionally, they conserve a substantial amount of clean water, crucial amid ongoing drought emergencies. These algorithms have been seamlessly integrated into the REWATnet tool and repository.

This doctoral thesis contributes significantly to urban water network management by integrating interdisciplinary research in computer science, graph theory, and water sciences. Addressing challenges from the COVID-19 pandemic, environmental preservation, and water scarcity, our results pave the way for efficient, accessible, and economical application of innovative algorithms, tools, and decision systems. These contributions aim to minimize costs, enhance resilience, and provide sustainability and efficacy in water resource management for future urban water systems.

The algorithms developed have not only been published in prestigious academic journals but have also been integrated into numerous applied research projects, including CLEaN-Tour, ReUseMP3, and ReWaT, receiving a total funding of €517,000. These integrated solutions have made significant contributions to the community and society. Particularly noteworthy is the interest shown by the city councils of Girona, Lloret de Mar, and el Prat de Llobregat in applying our solutions to their systems, as evidenced by their active participation in official meetings where the results of this thesis were

presented.

Resumen

Las redes de agua sirven como arterias vitales para áreas densamente pobladas, facilitando las funciones esenciales necesarias para sostener la vida urbana. Estas infraestructuras críticas abarcan el suministro de agua, distribución y gestión de aguas residuales, garantizando el acceso a agua limpia para beber, saneamiento y fines industriales.

El papel de estas redes no puede pasar desapercibido, ya que constituyen los cimientos de la infraestructura urbana. Sin embargo, en los últimos años, la ciencia del agua se ha enfrentado a varios desafíos urgentes que requieren atención. Dos áreas de investigación particulares incluyen la necesidad de vigilancia de aguas residuales para detectar virus y la optimización de redes de agua regenerada resilientes. Estos esfuerzos son cruciales para abordar el desafío de la escasez de agua agravado por el cambio climático. Sin infraestructuras críticas, las ciudades se paralizarían, provocando graves repercusiones en la salud pública, la actividad económica y la calidad de vida.

En nuestra era tecnológica, la teoría de grafos ofrece soluciones innovadoras a desafíos emergentes, principalmente en redes de telecos y campos de alta tecnología. Esta tesis aplica los principios de la teoría de grafos, que explora el estudio de grafos, estructuras matemáticas esenciales diseñadas para encapsular relaciones por pares entre objetos compuestos de nodos y aristas. Al abordar los desafíos en las redes urbanas de agua modeladas como grafos se pueden ofrecer estrategias innovadoras con un gran impacto en la sociedad en problemas relacionados con la COVID-19 y la escasez de agua. La teoría de grafos trasciende disciplinas, con redes de distribución de agua definidas por grafos no dirigidos y redes de aguas residuales por grafos dirigidos.

La metodología de esta tesis incluye un enfoque de cinco fases sencillo y eficiente: revisión bibliográfica e identificación de brechas en la literatura, adquisición de datos, preparación de datos, desarrollo de algoritmos y análisis de datos y procesamiento de resultados. Esta metodología automatiza la recopilación y procesamiento de datos, simplificando la implementación de algoritmos, resultando en beneficios notables. Además, las ciudades de Girona y Lloret de Mar sirven como casos de estudio reales

para probar los algoritmos desarrollados y presentar sus resultados.

Primero, el algoritmo de selección de sitios de vigilancia de aguas residuales surgió como una herramienta relevante para la gestión de pandemias, cumpliendo el objetivo original de un equilibrio óptimo entre cobertura, interferencia y distribución de pozos de mantenimiento. Este algoritmo se desarrolló basado en experiencias previas con algoritmos de optimización y heurísticos comúnmente utilizados en redes de telecos, a partir de una base derivada de un algoritmo utilizado para posicionar controladores.

Después del monitoreo de la red de alcantarillado, los esfuerzos se enfocaron en mejorar la resiliencia de estas redes, empezando con el análisis del impacto de los árboles en la infraestructura de aguas residuales. Se descubrieron grandes beneficios a través de una evaluación del riesgo de falla de tuberías. Además, el algoritmo de reordenamiento de árboles redujo el riesgo de falla de tuberías y los costos de reparación en la red de Girona, incluso ofreciendo ahorros económicos a pesar de los requisitos de inversión.

Después de aplicar la teoría de grafos en las redes de aguas residuales, surgieron dos propuestas para diseñar redes de distribución de agua regenerada (WDNs) rentables, resilientes e hidráulicamente factibles. Estas propuestas calculan grafos de red óptimos, sirviendo agua hasta tres veces más eficientemente que la planificación manual. Nuestros algoritmos evalúan y priorizan la resiliencia, lo que lleva a ahorros significativos en costos en comparación con las redes convencionales. Además, se conserva una cantidad sustancial de agua, crucial en medio de las emergencias de sequía en curso. Estos algoritmos se han integrado en la herramienta y repositorio REWATnet.

Esta tesis doctoral contribuye notablemente a la gestión de redes urbanas de agua mediante la integración de ciencias de la computación, teoría de grafos y ciencias del agua. Al abordar desafíos como la pandemia de COVID-19 y escasez de agua, nuestros resultados allanan el camino para la aplicación eficiente, accesible y económica de algoritmos innovadores y sistemas de decisión. Estas contribuciones tienen como objetivo minimizar costos, mejorar la resiliencia y proporcionar sostenibilidad y eficacia en la gestión de recursos hídricos para futuros sistemas de agua urbanos.

Los algoritmos no solo se han publicado en revistas académicas, sino que también se han integrado en varios proyectos de investigación, incluidos CLEaN-Tour, ReUseMP3 y ReWaT, recibiendo un financiamiento total de €517,000. Estas soluciones han hecho contribuciones importantes a la comunidad y la sociedad. Es especialmente notable el interés mostrado por los ayuntamientos de Girona, Lloret de Mar y el Prat de Llobregat en aplicar nuestras soluciones a sus sistemas, como lo demuestra su participación activa en reuniones oficiales donde se presentaron los resultados de esta tesis.

Resum

Les xarxes d'aigua serveixen com a artèries vitals per a àrees densament poblades, facilitant les funcions essencials necessàries per sostenir la vida urbana. Aquestes infraestructures crítiques inclouen el subministrament d'aigua, la distribució i la gestió d'aigües residuals, garantint l'accés a aigua per beure, sanejament i usos industrials.

El paper d'aquestes xarxes no pot passar desapercebut, ja que constitueixen els fonaments de la infraestructura urbana. Tanmateix, en els darrers anys, la ciència de l'aigua s'ha enfrontat a diversos reptes urgents que requereixen atenció. Dues àrees de recerca particulars inclouen la necessitat de vigilància d'aigües residuals per detectar virus i l'optimització de xarxes d'aigua regenerada resilients. Aquests esforços són crucials per abordar el repte de l'escassetat d'aigua agreujat pel canvi climàtic. Sense infraestructures crítiques, les ciutats es paralitzarien, provocant greus repercussions en la salut pública, l'activitat econòmica i la qualitat de vida en general.

En la nostra era tecnològica, la teoria de grafs ofereix solucions innovadores a reptes emergents, principalment en xarxes de telecomunicacions i camps d'alta tecnologia. Aquesta tesi aplica els principis de la teoria de grafs, que explora l'estudi de grafs, estructures matemàtiques essencials dissenyades per encapsular relacions per parells entre objectes compostos de nodes i arestes. En abordar els reptes en les xarxes urbanes d'aigua modelades com grafs, es poden oferir estratègies innovadores amb un gran impacte en la societat en problemes relacionats amb la COVID-19 i l'escassetat d'aigua. La teoria de grafs transcendeix disciplines, amb xarxes de distribució d'aigua definides per grafs no dirigits i xarxes d'aigües residuals per grafs dirigits.

La metodologia d'aquesta tesi inclou un enfocament de cinc fases senzill i eficient: revisió bibliogràfica i identificació de mancances en la literatura, adquisició de dades, preparació de dades, desenvolupament d'algoritmes i anàlisi de dades i processament de resultats. Aquesta metodologia automatitza la recopilació i el processament de dades, simplificant la implementació d'algoritmes, amb beneficis notables. A més, les ciutats de Girona i Lloret de Mar serveixen com a principals casos d'estudi reals per provar els

algoritmes desenvolupats i presentar els resultats.

Primer, l'algoritme de selecció de punts de vigilància a les aigües residuals va sorgir com una eina rellevant per a la gestió de pandèmies, complint l'objectiu original d'un equilibri òptim entre cobertura, interferència i distribució dels pous de manteniment. Aquest algoritme es va desenvolupar basat en experiències prèvies amb algoritmes d'optimització i heurístics comunament utilitzats en xarxes de telecomunicacions, a partir d'una base derivada d'un algoritme utilitzat per posicionar controladors.

Després del seguiment de la xarxa de clavegueram, els esforços es van centrar a millorar la resiliència d'aquestes xarxes, començant amb l'anàlisi de l'impacte dels arbres en la infraestructura d'aigües residuals. Es van descobrir grans beneficis a través d'una avaluació del risc de fallida de canonades. A més, l'algoritme de reordenament d'arbres va reduir significativament el risc de fallida de canonades i els costos de reparació en la xarxa de Girona, fins i tot oferint estalvis econòmics malgrat els requisits d'inversió.

Després d'aplicar la teoria de grafs en les xarxes d'aigües residuals, van sorgir dues propostes per dissenyar xarxes de distribució d'aigua regenerada (WDNs) rendibles, resilients i hidràulicament factibles. Aquestes propostes generen grafs de xarxa òptims, servint aigua fins a tres vegades més eficientment que la planificació manual. Els nostres algoritmes avaluen i prioritzen la resiliència, aportant estalvis significatius en costos en comparació amb les xarxes convencionals. A més, es conserva una quantitat substancial d'aigua, crucial enmig de les emergències de sequera en curs. Aquests algoritmes s'han integrat en l'eina i repositori REWATnet.

Aquesta tesi doctoral contribueix notablement a la gestió de xarxes urbanes d'aigua mitjançant la integració de ciències de la computació, teoria de grafs i ciències de l'aigua. En abordar reptes com la pandèmia de COVID-19 i l'escassetat d'aigua, els nostres resultats obren el camí per a l'aplicació eficient, accessible i econòmica d'algoritmes innovadors i sistemes de presa de decisions. Aquestes contribucions tenen com a objectiu minimitzar costos, millorar la resiliència i proporcionar sostenibilitat i eficàcia en la gestió de recursos hídrics per a futurs sistemes d'aigua urbans.

Els algoritmes no només s'han publicat en revistes acadèmiques, sinó que també s'han integrat en diversos projectes de recerca, inclosos CLEaN-Tour, ReUseMP3 i ReWaT, rebent un finançament total de €517,000. Aquestes solucions han fet contribucions importants a la comunitat i la societat. És especialment notable l'interès mostrat pels ajuntaments de Girona, Lloret de Mar i el Prat de Llobregat en aplicar les nostres solucions als seus sistemes, com ho demostra la seva participació activa en reunions oficials on es van presentar els resultats d'aquesta tesi.

Chapter 1

Introduction

1.1 Background and motivation: the COVID-19 pandemic, water scarcity, and the integration of graph theory

Graph theory, a foundational discipline in mathematics and computer science, explores the study of graphs, essential mathematical structures crafted to encapsulate pairwise relationships between objects. These graphs, composed of nodes and connecting edges, serve as maps illustrating connections between various elements. Within this framework, edges may represent undirected paths, highlighting equality between connected nodes, or directed, guiding singular paths from one node to another. This versatility positions graphs as indispensable tools in discrete mathematics, enabling the modeling and analysis of complex systems, particularly evident in the context of transport networks. Indeed, the applicability of graph theory transcends disciplines, finding resonance in diverse domains such as telecommunications and water science. In telecommunications, Software-Defined Networking (SDN) is often represented as graphs, similarly, urban water networks in water science can also be modeled using graph structures. Urban water distribution networks, characterized by bidirectional flows, are typically represented by undirected graphs. In contrast, wastewater networks, where flow is unidirectional towards the Wastewater Treatment Plant (WWTP), are depicted using directed graphs.

Urban water networks serve as lifelines for densely populated areas, facilitating the essential functions necessary for sustaining urban life [1]. These critical infrastructures encompass water supply, distribution, and wastewater management, ensuring access to clean water for drinking, sanitation, and industrial purposes. Amidst the bustling activity

of urban centers, these networks operate quietly and often go unnoticed. In our rapidly evolving technological era, computer science offers advanced solutions, predominantly applied to telecommunication networks and other high-tech sciences, yet often overlooked in other scientific fields.

The critical role of urban water networks cannot be overstated [2]. They form the backbone of urban infrastructure, enabling the functioning of households, businesses, healthcare facilities, and industries. However, in recent years, water science has encountered several pressing issues that demand attention, particularly the need for wastewater surveillance for SARS-CoV-2 detection and the optimization of resilient reclaimed water networks to address the significant challenge of water scarcity, especially in Mediterranean areas. Applying computer science solutions to these challenges can shed light and offer novel strategies for the future of urban water networks, thereby delivering a profoundly positive impact on our society. Without reliable water supplies, cities would grind to a halt, facing severe repercussions on public health, economic activity, and overall quality of life.

Increasing evidence suggests that sewage serves as a valuable, unbiased indicator of virus prevalence within a population. Research groups worldwide have reported successful detection of SARS-CoV-2 in sewage samples [3, 4]. With confirmation that COVID-19 patients shed SARS-CoV-2 in feces, numerous studies have demonstrated a significant correlation between the concentration of the virus in sewage and the prevalence of COVID-19 within the corresponding population [5–8]. While this approach has proven successful at the WWTP level, integrating data from all inhabitants within a municipality, there remains limited experience in applying it at the neighborhood level. Implementing 'upstream' surveillance for Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) may enable more precise spatial detection of the virus in catchments with varying COronaVIrus Disease of 2019 (COVID-19) disease burdens. Moreover, it has the potential to provide valuable information regarding any mitigation actions implemented at the community level, thus offering a proactive approach to managing public health risks.

The transition to circularity and the minimization of potable water consumption necessitate a comprehensive redesign of water infrastructure, encompassing water reuse networks [9]. Treated wastewater presents a valuable resource for various non-potable purposes, such as irrigation, toilet flushing, car washing, cleaning, and industrial uses [10], thereby conserving a significant portion of freshwater resources. Initiatives promoting water reuse, exemplified by municipalities like Sant Cugat del Vallès in Spain [11, 12], have gained traction, particularly in multi-story buildings. However, such

applications remain limited, and comprehensive information on the subject is scarce in existing literature. Therefore, there is an urgent need to enhance planning strategies for water reuse within urban settings by designing efficient and resilient water reuse networks.

Resilience is a crucial consideration for urban water networks in general, especially amidst escalating climate variability and extreme weather events. These networks face diverse challenges, ranging from droughts affecting Water Distribution Networks (WDNs) to tree roots infiltrating wastewater systems, necessitating their ability to maintain uninterrupted service delivery. Enhancing resilience entails not only strengthening physical infrastructure but also developing adaptive strategies and integrating redundancy to enhance preparedness and response capabilities.

The wastewater network is vulnerable to various threats, often resulting in leakages that lead to significant economic losses and pose public health hazards [13]. Pipe failures incur direct economic costs for repairs, along with indirect costs such as infrastructure damage, business disruption, and production losses [14]. The concept of 'urban water security' has emerged to address such vulnerabilities, fostering a new understanding of the intricate dynamics between human and natural systems, and paving the way for expanded risk management strategies [15]. Leveraging available databases containing pertinent information on pipe and risk element geolocation, such as tree roots or street works, offers the opportunity to automate processes using mathematical algorithms. This automation enables the determination of whether it is more prudent to address pipe failures or leakages as they occur or to prevent them by rearranging nearby risk elements.

Ensuring resilience is also a critical issue in Water Distribution Networks (WDNs). The consequences of pipe failures extend far beyond immediate concerns, encompassing profound economic and environmental ramifications. Without proactive measures to provide resilient network designs, the risk of such failures escalates markedly, particularly amid the ongoing challenges of water scarcity driven by climate change in numerous regions worldwide. These repercussions may manifest as compromised public health, disruption of essential services, significant economic burdens, and wastage of substantial water resources [16]. By offering cost-effective and hydraulically validated approaches, automatic resilient design solutions present a viable means of preventing network failures in both drinking and reclaimed WDNs, ultimately conserving large quantities of water resources over the long term.

In this way, the present thesis aims to address the pressing challenges faced by urban water networks, particularly in the context of the COVID-19 pandemic and water scarcity. Through the incorporation of cutting-edge graph-theory methodologies,

this research seeks to enhance surveillance strategies for early detection of pathogens in wastewater, while also designing resilient reclaimed water networks capable of withstanding the impacts of climate change. Through interdisciplinary approaches that leverage computer and water sciences, this thesis endeavors to provide novel insights and practical cost-effective strategies for ensuring the optimal surveillance, design, and resilience of future urban water networks, thus positively impacting the overall quality of life in urban areas.

1.2 Objectives

The **general objective** of this thesis is **to advance the cost-effective management of urban water networks by integrating interdisciplinary research from computer science and water sciences to tackle challenges arising from the COVID-19 pandemic and water scarcity**. The primary focus is on enhancing surveillance capabilities for pathogen detection in wastewater, optimizing water reuse networks, improving resilience against climate variability, automating risk management processes, and designing cost-effective solutions for urban water networks. By leveraging innovative approaches and insights from both fields, this research aims to contribute to the overall enhancement of urban living standards through the development of more efficient, resilient, and sustainable water management strategies.

To reach the ultimate purpose of the research, the following specific objectives have been set:

1. **Development of Advanced Surveillance Techniques:** Implement innovative graph-theory-based methods to enhance surveillance strategies for early detection of pathogens, particularly focusing on SARS-CoV-2, in urban wastewater networks. This objective involves designing algorithms capable of intelligent positioning sensors at the neighborhood level for precise and cost-effective spatial detection of virus prevalence.
2. **Optimization of Water Reuse Network Designs:** Conduct comprehensive planning and optimization strategies for water reuse within urban settings. This objective includes designing efficient water reuse networks to minimize potable water consumption and network construction costs, leveraging treated wastewater for various non-potable purposes while ensuring environmental sustainability and economic benefits.
3. **Automation of Risk Management Processes:** Utilize available databases and

innovative mathematical algorithms to automate the evaluation of pipe failure risks within urban water networks. This objective aims to optimize decision-making regarding pipe maintenance and leakage prevention by analyzing geolocation data of risk elements such as tree roots or street works and other operational factors.

4. **Design of Cost-Effective Resilient Network Solutions:** Develop cost-effective and hydraulically validated approaches for designing resilient WDNs. This objective focuses on preventing network failures and conserving water resources by implementing automatic resilient design solutions capable of withstanding the challenges posed by climate change and water scarcity.

5. **Integration of Solutions with Open Data Platforms and Applied Research Projects** The intention is to disseminate research findings not only through academic articles but also to translate them into practical, modular tools freely accessible to network managers. All data and code implementations generated will be openly available in public code repositories. Additionally, the tools and algorithms will be linked with specific applied research projects to ensure their applicability beyond academia.

1.3 Presentation and justification of the articles included in the thesis

The findings of this thesis are derived from four distinct studies published in prestigious journals, each carrying its own substantial significance. These studies are presented in four results sections (Sections 5.1 – 5.4), with each subsection dedicated to addressing the objectives 1 to 4 outlined in this research.

The articles are interconnected, with some evolving in a "sequential" manner while others are developed concurrently (Table 1.1, Figure 1.1). This dual approach facilitates both a progressive exploration of research topics and simultaneous investigation of related themes. While certain articles build upon the findings of preceding ones, others explore distinct but complementary aspects of the research domain.

Table 1.1: Order and structure of the various results chapters comprising this thesis in relation to the research objectives.

Paper 1.	
Title:	Optimal selection of monitoring sites in cities for SARS-CoV-2 surveillance in sewage networks

Authors:	Calle, E.; **Martínez, D.**; Brugués-i-Pujolràs, R.; Farreras, M.; Saló-Grau, J.; Pueyo-Ros, J.; Corominas, L.
Publication:	Environment International
DOI:	https://doi.org/10.1016/j.envint.2021.106768
Year:	2021
Volume:	157
Pages:	106768
Abstract:	Selecting sampling points to monitor traces of SARS-CoV-2 in sewage at the intra-urban scale is no trivial task given the complexity of the networks and the multiple technical, economic and socio-environmental constraints involved. This paper proposes two algorithms for the automatic selection of sampling locations in sewage networks. The first algorithm, is for the optimal selection of a predefined number of sampling locations ensuring maximum coverage of inhabitants and minimum overlapping amongst selected sites (static approach). The second is for establishing a strategy of iterations of sample&analysis to identify patient zero and hot spots of COVID-19 infected inhabitants in cities (dynamic approach). The algorithms are based on graph-theory and are coupled to a greedy optimization algorithm. The usefulness of the algorithms is illustrated in the case study of Girona (NE Iberian Peninsula, 148,504 inhabitants). The results show that the algorithms are able to automatically propose locations for a given number of stations. In the case of Girona, always covering more than 60% of the manholes and with less than 3% of them overlapping amongst stations. Deploying 5, 6 or 7 stations results in more than 80% coverage in manholes and more than 85% of the inhabitants. For the dynamic sensor placement, we demonstrate that assigning infection probabilities to each manhole as a function of the number of inhabitants connected reduces the number of iterations required to detect the zero patient and the hot spot areas.
Keywords:	COVID-19; Wastewater-based epidemiology; Sewer; Sensor placement; Graph theory
Oobjectives:	1. Development of Advanced Surveillance Techniques 5. Integration of Solutions with Open Data Platforms and Applied Research Projects

Paper 2.

Title:	Optimal design of water reuse networks in cities through decision support tool development and testing

Authors:	Calle, E.; **Martínez, D.**; Buttiglieri, G.; Corominas, L.; Farreras, M.; Saló-Grau, J.; Vilà, P.; Pueyo-Ros, J.; Comas, J.
Publication:	npj Clean Water
DOI:	https://doi.org/10.1038/s41545-023-00222-4
Year:	2023
Volume:	6
Pages:	23
Abstract:	Water scarcity and droughts are an increasing issue in many parts of the world. In the context of urban water systems, the transition to circularity may imply wastewater treatment and reuse. Planning and assessment of water reuse projects require decision-makers evaluating the cost and benefits of alternative scenarios. Manual or semi-automatic approaches are still common practice for planning both drinking and reclaimed water distribution networks. This work illustrates a decision support tool that, based on open data sources and graph theory coupled to greedy optimization algorithms, is able to automatically compute the optimal reclaimed water network for a given scenario. The tool provides not only the maximum amount of served reclaimed water per unit of invested cost, but also the length and diameters of the pipes required, the location and size of storage tanks, the population served, and the construction costs, i.e., everything under the same architecture. The usefulness of the tool is illustrated in two different but complementary cities in terms of size, density, and topography. The construction cost of the optimal water reclaimed network for a city of approximately 100,000 inhabitants is estimated to be in the range of €0.17–0.22/m^3 (for a payback period of 30 years).
Keywords:	wastewater treatment; planning; reclaimed; circular economy; graph theory; clustering; costs
Objectives:	2. Optimization of Water Reuse Network Designs 5. Integration of Solutions with Open Data Platforms and Applied Research Projects

Paper 3.

Title:	Seamless integration of sewer system topology and tree location data: An algorithm to diagnose the potential impact of tree roots on pipes and propose rearrangement solutions
Authors:	**Martínez, D.**; Bergillos, S.; Corominas, L.; Comas-Cufí, M.; Calle, E.
Publication:	Heliyon

DOI: https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e23382
Year: 2024
Volume: 10
Pages: –
Abstract: Wastewater networks are subject to several threats leading to wastewater leakages and public health hazards. External elements such as natural factors and human activities are common causes of wastewater leakages and require more in-depth analysis. Prevention and rehabilitation work is essential to secure wastewater networks and avoid pipe failures. This work presents a new algorithm that allows for the seamless integration of sewer topology and tree location data to diagnose the potential impact of tree roots on pipes. The algorithm also proposes tree rearrangement options that balance the cost of tree rearrangement with the cost of pipe repair. The paper also showcases a real-world case study in the city of Girona to evaluate the performance of the presented algorithms for a specific case focusing on tree roots as a natural factor. Results show that it is possible to optimally rearrange a number of the trees with the greatest impact, significantly minimizing pipe failures and wastewater leakages (82% risk reduction with only rearranging a 12% of the most impactful trees). The rearrangement solution not only protects the environment and prevents public health hazards, but also achieves a positive economic payback during the operational period of the pipes, saving up to 1.33M€ for a tree rearrangement of 7%. The presented methodology is applicable to other natural or human factors.
Keywords: Wastewater leakages; Risk assessment; Pipe failures; Underground networks; Environmental protection; Rearrangement solutions
Objectives: 3. Automation of Risk Management Processes
5. Integration of Solutions with Open Data Platforms and Applied Research Projects

Paper 4.

Title: Enhancing reclaimed water distribution network resilience with cost-effective meshing
Authors: **Martínez, D.**; Bergillos, S.; Corominas, L.; Comas, J.; Wang, F.; Kooij, R.; Calle, E.
Publication: Science of The Total Environment
DOI: https://doi.org/10.1016/j.scitotenv.2024.173051
Year: 2024
Volume: 938

Pages: 173051

Abstract: Water Distribution Networks (WDNs) are critical infrastructures that ensure a continuous supply of safe water to homes. In the face of challenges, like water scarcity, establishing resilient networks is imperative, especially in regions vulnerable to water crises. This study evaluates the resilience of network designs through graph theory, including its hydraulic feasibility using EPANET software, an aspect often overlooked. Novel mathematical algorithms, including Resilience by Design (RbD) and Resilience-strengthening (RS) algorithms, provide cost-effective and resilient network designs, even with budget constraints. A novel metric, Water Availability (WA), is introduced to offer a comprehensive measure of network resilience, thereby addressing ongoing discrepancies in resilience evaluation methods. Practical benefits are illustrated through a case study in which a resilient-by-design reclaimed water network is created, and an existing equivalent non-resilient network is improved.

The resilient-by-design network demonstrates remarkable better results compared to the equivalent non-resilient design, including up to a 36% reduction in the probability of service disruptions and a nearly 65% decrease in the annual average unserved water due to service disruptions. These findings underscore the enormous advantages of a resilience-focused network design approach. When compared to the equivalent non-resilient design, the resilience-by-design network generated effectively safeguards up to a significant 91,700 m^3 of water from the impacts of water disruption events over a 50-year operational period. Over a 50-year operational period, our reclaimed resilient-by-design WDN solution incurs only a modest increase in overall costs compared to consuming tap water from the drinking WDN baseline, highlighting a cost-effective approach that saves a significant amount of water. This paper builds on our previous research by expanding its scope to include resilience considerations, providing algorithms that can be easily adapted from reclaimed to drinking water distribution networks. Ultimately, we contribute to the enhancement of water resource management and infrastructure planning in ever-evolving urban environments.

Keywords: resilience; water availability; planning; graph theory; costs; hydraulic feasibility

Objectives: 3. Automation of Risk Management Processes

4. Design of Cost-Effective Resilient Network Solutions

5. Integration of Solutions with Open Data Platforms and Applied Research Projects

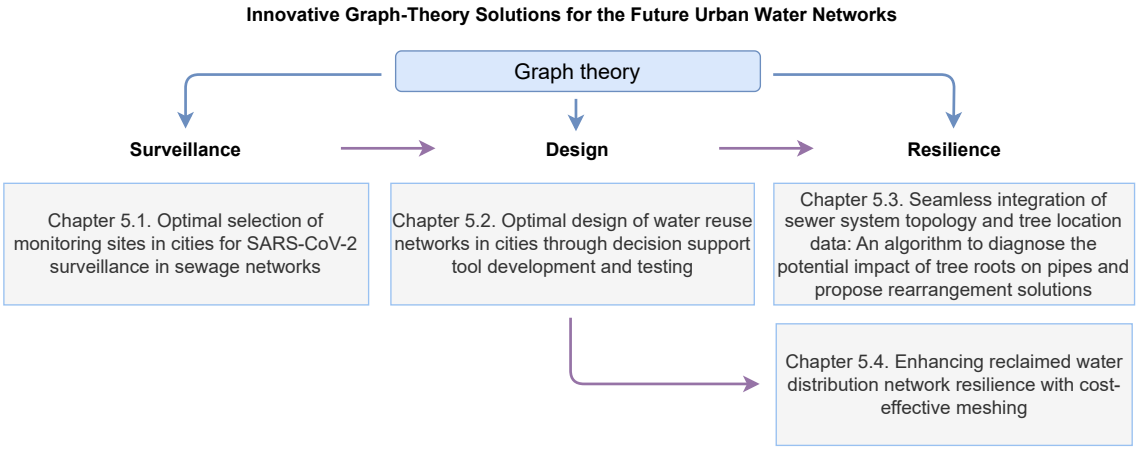


Figure 1.1: Interconnections among the different result chapters comprising this research. **Source:** Original work.

Furthermore, Chapter 6 serves as a general discussion section that intertwines and examines the findings of each article in an integrated and complementary manner. Its purpose is to emphasize the overarching themes derived from the conclusions of the various publications comprising this research.

1.4 Structure of the thesis

This thesis is presented in the format of a compendium of publications, each associated with one of the specific objectives. It is structured into seven chapters, including the present introduction and sections for bibliographic references.

Chapter 1 summarizes the background and motivation behind this research work (Section 1.1), the stated general and specific objectives (Section 1.2), and the publications comprising this thesis (Section 1.3), along with its overall structure (Section 1.4).

Chapter 2 establishes the theoretical foundation upon which the thesis is built, building upon the background presented in the previous chapter. First, water network types and challenges are introduced (Section 2.1). Next, graph theory and its role in telecommunications and applied sciences are presented (Section 2.2).

Chapter 3 illustrates the general methodology followed in this thesis and Chapter 4 presents the justification (Section 4.1) and characterization (Section 4.2) of Girona and Lloret de Mar case-study scenarios.

Chapter 5 presents the four publications encompassed in this thesis, including: (i) the optimal selection of monitoring sites in cities for SARS-CoV-2 surveillance in sewage networks (Section 5.1), (ii) the optimal design of water reuse networks in

cities through decision support tool development and testing (Section 5.2), (iii) the seamless integration of sewer system topology and tree location data: an algorithm to diagnose the potential impact of tree roots on pipes and propose rearrangement solutions (Section 5.3), and (iv) the enhancement of reclaimed water distribution network resilience with cost-effective meshing (Section 5.4). Chapter 6 serves as a general discussion section that intertwines and examines the findings of each article in an integrated and complementary manner.

Finally, Chapter 7 presents comprehensive conclusions and outlines future research directions.

Chapter 2

Theoretical Basis

2.1 Water networks: types and challenges

Managing water networks involves addressing numerous challenges, such as safeguarding public health with clean drinking water, tackling environmental concerns through wastewater management, and minimizing water drought effects through innovative solutions like reclaimed water systems. In this section, two pivotal types of water networks are defined: wastewater networks, and water distribution networks (WDNs). Each network fulfills a specific role, presenting unique technical, regulatory, and societal challenges.

2.1.1 Wastewater networks

A wastewater network is a conglomerated system mainly comprised of underground pipes, maintenance holes, pumping stations, and overflow structures that work together to collect and drain wastewater from households or industrial centers to wastewater treatment plants. This network ensures the efficient transport of wastewater, accommodating variations in flow and preventing overflows during heavy rainfall or increased usage. Additionally, these networks include monitoring systems and control mechanisms to maintain optimal performance and compliance with environmental standards. Once treated, the water is returned to the environment or reused for beneficial purposes such as agriculture, irrigation, or even potable water supplies [17], ever more widely accepted by the general public [18]. Hence, wastewater networks are critical infrastructures and essential assets for the proper functioning of society and the economy [19, 20].

In the context of the COVID-19 global pandemic, disease monitoring emerged as

crucial information, particularly for decision-makers, guiding the implementation of necessary countermeasures to flatten the epidemic curve. Given that sewage has been confirmed to retain concentrations of the SARS-CoV-2 virus from human feces [5–8], monitoring these concentrations at a neighborhood level can provide a precise snapshot of the current state of the epidemic curve, often days or weeks ahead of other methods such as Polymerase Chain Reaction (PCR) or antigen tests conducted in medical centers [21] (Figure 2.1). However, there is limited literature available on the optimal placement of monitoring sites or autonomous wastewater sensors to effectively monitor the epidemic while: (i) covering the maximum population with the fewest sensors possible, and (ii) minimizing data interference between different monitoring sites.

Figure 2.1: Sensor deployment in a wastewater network maintenance hole for flow monitoring purposes. **Source:** Cicle de l'aigua del Ter.

External elements such as tree roots or street works are widely recognized for their role in causing pipe failures and network disruptions (Figure 2.2). These threats are particularly concerning in wastewater networks, as they can result in leakages that not only incur economic losses but also pose risks to public health and contribute to environmental contamination [22–26]. In the city of Girona, wastewater network managers confirm that such external factors, especially tree roots, are responsible for a significant proportion of pipe failures. With the increasing trend of cities publishing data on open portals, including tree inventories and locations, a new line of research emerges to understand the potential impact of trees on wastewater networks. This information becomes crucial in planning potential tree rearrangements aimed at minimizing pipe failures.

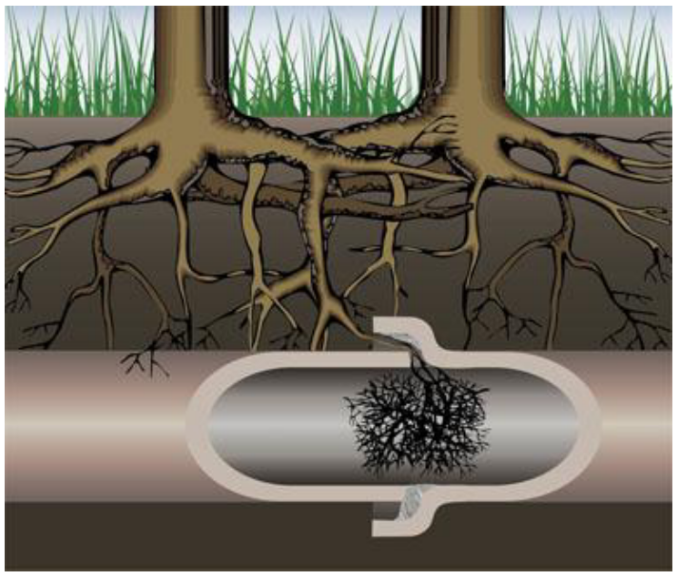


Figure 2.2: Tree root intrusions in wastewater pipes. **Source:** [26].

2.1.2 Drinking and reclaimed water distribution networks

A water distribution network (WDN) ensures a reliable supply of water to households and other urban infrastructures. This network comprises a system of pipes, valves, pressure pumps, and tank facilities designed to transport water from its source to homes, businesses, and public facilities. The network operates on a city-wide scale, spanning across neighborhoods, and undergoing a series of processes to maintain water quality and meet regulatory frameworks. The maintenance and monitoring of these networks is crucial, with emerging challenges such as design costs and water leaks caused by pipe failures [27].

Amid the escalating climate emergency, various regions worldwide are challenged with increasingly severe water shortages, particularly evident in areas like the Mediterranean [28]. In Catalonia, for instance, nearly 80% of the population in 2024, amounting to approximately 6 million people, is currently facing drought emergencies, placing immense pressure on available clean water reservoirs. As a consequence, restrictions on drinking water distribution networks (WDNs) are anticipated. In response to this pressing challenge, the integration of reclaimed WDNs has become imperative [9, 10]. These networks offer a viable solution by significantly reducing dependence on clean water sources. Through advanced treatment processes at wastewater treatment plants (WWTPs), and sometimes even through decentralized methods [11, 12], wastewater is transformed into reclaimed water suitable for various non-potable purposes such as garden irrigation and toilet flushing.

Constructing reclaimed water distribution networks (WDNs) necessitates substantial

economic investments during the design and planning phases, requiring extensive technical expertise and local knowledge. Consequently, the development of these networks presents novel challenges, particularly in optimizing costs to achieve up to 10% savings over the network's lifetime, which is the typical cost for the manual design phase [29]. Moreover, the focus extends beyond cost minimization; it encompasses the creation of resilient networks capable of reducing repair expenses and minimizing service disruptions and water loss caused by unforeseen events.

Efforts to minimize costs are intertwined to optimize network size while maximizing water distribution efficiency. Simultaneously, resilience considerations emphasize the importance of establishing alternative pathways for water delivery to ensure continuous service availability in the event of pipe failures, thus preventing disruptions. In addition to cost-effective designs, evaluating water availability measures is crucial for assessing network resilience. This evaluation enables a comparative analysis between non-resilient and resilient network designs, providing insights into the effectiveness of different strategies in ensuring consistent and reliable water distribution even under adverse conditions.

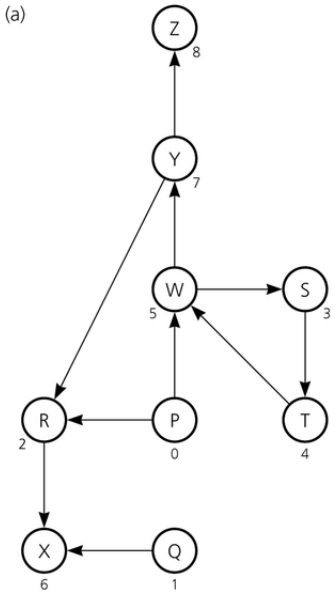
Through a comprehensive approach encompassing cost minimization and resilience enhancement, the development of reclaimed WDNs can not only effectively address water scarcity challenges but also promote sustainable and resilient water infrastructure systems.

2.2 Graph theory

Graph theory, a cornerstone discipline in mathematics and computer science, delves into the study of graphs, fundamental mathematical constructs designed to represent pairwise relationships between objects. These graphs, meticulously crafted from nodes and connecting edges, function as dynamic maps that visually depict the connections and relationships among diverse elements within a system or network. This versatility positions graphs as indispensable tools in modeling a wide range of real-world systems, from social networks to transportation networks, biological networks, water networks, and beyond [30].

Graphs can be represented using various data structures that define their nodes, edges, and associated data. The adjacency matrix, a popular representation form, is a square matrix used to depict finite graphs, with its elements indicating node adjacency (Figure 2.3). However, it lacks the ability to illustrate additional node or edge data. Addressing this limitation, Graph Markup Language (GraphML) emerged as an Extensible

Markup Language (XML)-based file format for graphs, facilitating the exchange of graph structure data within the graph drawing community [31]. Choosing the right graph representation is crucial, depending on the specific problem and computational requirements of algorithms [32]. With the advent of AI techniques, selecting appropriate and efficient graph representations becomes paramount for learning systems [33]. Graphs are mathematically represented as $G = (V, E)$, consisting of a set V of vertices (also called nodes) and a set E of edges. In particular, a directed graph is a specific graph in which the edges may only be traversed in one direction, in contrast with an undirected graph in which the edges may be traversed in both directions.



(b)

		0	1	2	3	4	5	6	7	8
		P	Q	R	S	T	W	X	Y	Z
0	P	0	0	1	0	0	1	0	0	0
1	Q	0	0	0	0	0	0	1	0	0
2	R	0	0	0	0	0	0	1	0	0
3	S	0	0	0	0	1	0	0	0	0
4	T	0	0	0	0	0	1	0	0	0
5	W	0	0	0	1	0	0	0	1	0
6	X	0	0	0	0	0	0	0	0	0
7	Y	0	0	1	0	0	0	0	0	1
8	Z	0	0	0	0	0	0	0	0	0

Figure 2.3: A directed graph represented with its adjacency matrix. **Source:** [34]

Graph theory exemplifies its broad applicability across diverse fields, showcasing its prowess in solving practical problems and fostering innovation in real-world scenarios. Several notable applications of graph theory include:

1. Computer Networking: Graph theory is crucial in routing algorithms, network topology analysis, and network security, as well as for clusterization purposes [35–37]. Fundamental algorithms like Dijkstra's shortest path and the Bellman-Ford algorithm are pivotal for efficiently routing data in networks.

2. Social Network Analysis: Graph theory is essential for understanding relationships and interactions within social networks, enabling the detection of communities and the analysis of information or disease spread among populations [38].

3. Bioinformatics: In the context of bioinformatics, graph theory aids in the analysis of biological data, including protein-protein interaction networks, metabolic

pathways, and genetic regulatory networks [39]. Graph algorithms facilitate tasks such as predicting protein structures and identifying functional modules in biological networks.

4. Operations Research: Graph theory is instrumental in optimizing various processes and systems, such as logistics, supply chain management, scheduling, facility layout planning, and project management [40]. Graph algorithms optimize transportation routes, warehouse locations, inventory management, and scheduling tasks.

2.2.1 Graph theory's role in telecommunication

Telecommunication networks are essential for modern society, facilitating a wide range of activities from social interactions to emergency services. The recent global challenges, such as the COVID-19 pandemic, have underscored the critical nature of these networks, highlighting the need for their reliability and resilience. Graph theory, a mathematical framework for studying relationships between interconnected entities, emerges as the backbone of understanding and optimizing these networks. By representing network components as nodes and their Internet connections as edges, graph theory provides vital insights into the structure and behavior of the network.

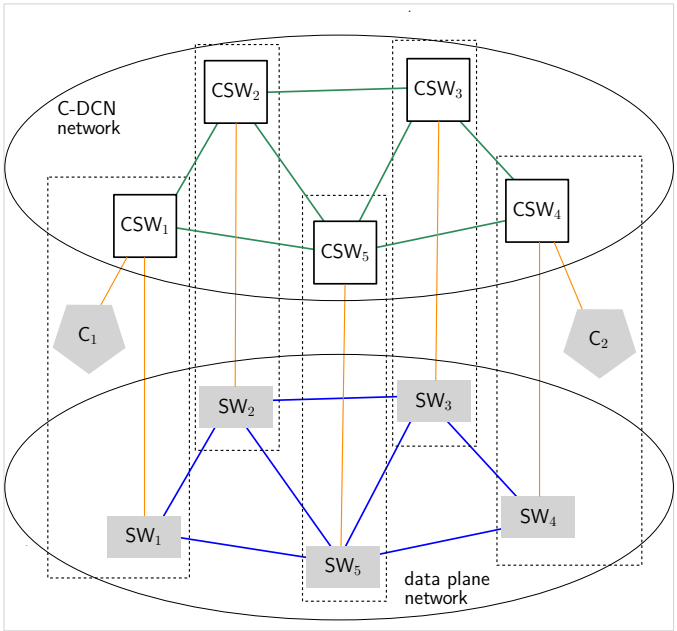


Figure 2.4: Data plane network and control plane data communication network (SDN).
Source: [41]

Another application of graph theory focuses on the network robustness problem. The Broadband Communications and Distributed Systems (BCDS) research group at the

University of Girona (UdG) is well-known for tackling this problem [35, 42], which resulted in the recent implementation of a complete Network Research Simulator (NRS) 2.0 tool [43]. This tool is capable of computing not only robustness but also conducting generic experiments and simulations for any kind of network infrastructure represented through nodes and edges. Additionally, NRS allows for very rich interactive visualization of networks, in 2D, 3D, or even in map mode for networks including geographical data.

In previous research works [41, 44] we presented an optimization approach that can be used by the SDN network operator to properly locate the controllers by taking into account predictable sets of critical targeted attacks on network topology. The proposed approach included an algorithm for predicting the sets of most dangerous attacks. Such sets are then used as input data for controller placement optimization, which is performed by means of mixed-integer and time-efficient heuristic programming methods. In the optimization, the impact of the considered attacks is measured by a novel Average Network Availability (ANA) measure. To minimize the consequences of attacks we considered additional backup controllers. The results show an efficient maximization of the ANA measure through the optimal SDN controller placement.

Therefore, graph theory empowers SDN network defenders to address vulnerabilities and develop robust defense strategies, facilitating the creation of resilient architectures to tackle evolving digital challenges.

2.2.2 Graphing the flow: graph theory's application in water networks

The challenges encountered in water networks can find their solutions through the application of graph theory [45], a well-established methodology commonly utilized in telecommunications networks. By representing water networks in a graph format, innovative algorithms grounded in graph theory principles can be effectively deployed. While telecommunications networks typically depict switches as nodes and their connections as edges within the graph, water networks are represented with junctions serving as nodes and pipes acting as edges (Figure 2.5). Ultimately, both telecommunication and water networks encounter flow optimization challenges. What data represents in telecommunication networks, water symbolizes in water networks. This approach facilitates the adaptation of graph theory techniques to analyze and optimize water network operations, offering a systematic framework to address complex network design issues and enhance overall efficiency and resilience. Therefore, graph theory serves not only as a suitable tool for analysis but also for solution development.

Figure 2.5: Design example of an urban water distribution network in the Girona Domeny neighborhood represented in a graph format. **Source:** ReWaT project.

Monitoring site selection, or sensor placement, has received considerable attention within the context of drinking water networks, yet research in sewage networks remains limited. Only a handful of studies, such as those by [46] and [47], have proposed methodologies for sensor placement in wastewater treatment plants (WWTPs) based on graph theory principles. These methodologies aim to maximize data quality assessment and control while minimizing costs, although without offering solutions at the neighborhood level. By employing graph-theory algorithms to strategically position SARS-CoV-2 sensors at the neighborhood level, it becomes possible to enhance pandemic monitoring capabilities. This decentralized approach not only facilitates the identification of hotspots but also enables early-stage detection of patient zero during a pandemic outbreak.

Previous studies have explored distribution water network modeling, approached through graph theory principles. For instance, [48] proposed a methodology for designing water distribution networks based on the hydraulically balanced loop method, while [49] developed a comprehensive analytical framework to examine the resilience patterns of water distribution systems in relation to their topological characteristics. Despite these efforts, graph theory has yet to be fully utilized for the advanced and automated design and resilience optimization of water networks. While various algorithms exist for designing water reuse networks using graph theory coupled with existing greedy optimization algorithms [50, 51], these approaches typically involve multiple tools and frameworks. Thus, there remains a need for the development of a unified solution that eradicates the need for data exchange, potentially saving both time and effort in the design

process.

Chapter 3

Methodology

3.1 General methodology

The aim of this section is to introduce the general methodology that has been utilized throughout the doctoral thesis. A comprehensive description of this methodology and its particularities is provided within the publications comprising Chapter 5.

Figure 3.1 outlines the synthesis of methods employed in the study, distinguishing between data acquisition techniques, data preparation and curation techniques, algorithm definitions, and, lastly, data analysis techniques. It is crucial to note that all algorithms developed within the scope of this thesis, along with the obtained results, were implemented using an Ubuntu 20.04 LTS server (Central Processing Unit (CPU): Advanced Micro Devices, Inc (AMD) Ryzen 5 5600X, 32GB Random Access Memory (RAM)), executed within a Python 3 notebook (Jupyter Lab, [52]). However, the methodology can be easily adaptable to other systems, such as Microsoft Windows.

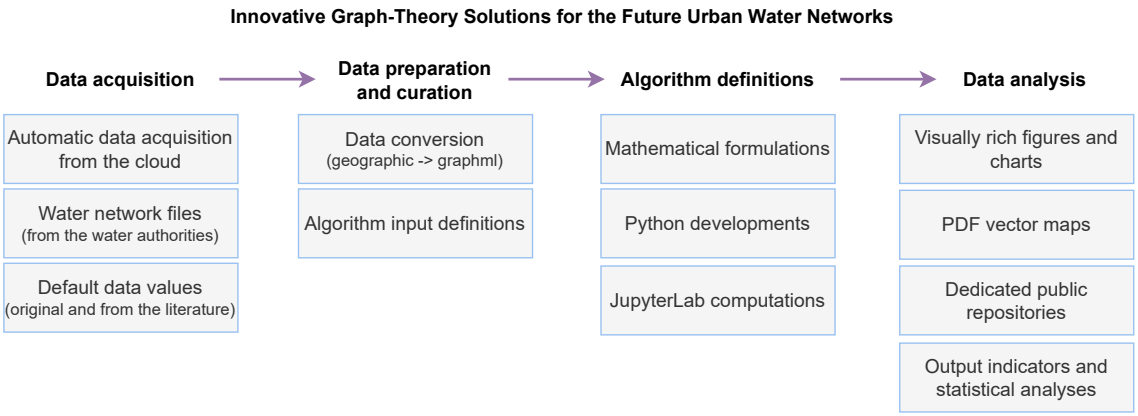


Figure 3.1: General methodology employed in the thesis. **Source:** Original work.

In essence, the general methodology adheres to the following procedures:

0. Bibliographic review and identification of literature gaps: A crucial initial step in any research endeavor involves identifying gaps in existing literature through a bibliographic review. This process entails integrating various resources, such as Google Scholar and ResearchGate, and utilizing keywords relevant to the thesis's focus areas within the context of graph theory. Using this method, a database of relevant bibliography items is analyzed, and those deemed pertinent are thoroughly reviewed before being cited.

1. Data acquisition: Since all publications within this thesis aim to automate surveillance, design, and resilience processes via graph theory, specific methodologies start with the automatic acquisition of data from open sources. If data is publicly accessible in the cloud, all necessary information for algorithm computation is automatically gathered through Python scripts. In cases where more specific public data cannot be retrieved automatically, it is integrated into the developed algorithms as default data values, easily modifiable as required. Several examples of publicly accessible data in the cloud include: OpenStreetMap Application Programming Interface (API) [53, 54], Instituto Geográfico Nacional (IGN) Digital Elevation Model (DEM) [55], Cadastral data [56], Overpass API [57], and Shapely [58].

2. Data preparation and curation: The data gathered from the data acquisition process often comes in formats that are not readily compatible with graph-theory code frameworks. Given graph theory's primary focus on telecommunication networks, popular frameworks like the well-known *networkx* Python library [59] typically require graph data to be formatted in the standard Graph Markup Language (GraphML) format [31]. This differs from geographic or Quantum Geographic Information System (QGIS) data, often handled in water networks [60, 61]. Therefore, this stage of the methodology involves developing techniques to automatically convert the raw data into the appropriate format, ensuring it is ready for the computation of the algorithms.

3. Algorithm definitions: With the data prepared and structured appropriately, algorithms are formulated and executed to obtain results. Each presented algorithm addresses a specific issue aligned with the objectives of the thesis. All newly devised algorithms presented in this thesis are built upon the *networkx* Python library [59], which offers fundamental graph-theory routing and analysis algorithms of immense utility.

4. Data analysis: The results obtained by the algorithms undergo processing to generate detailed and visually rich PDF vector map figures and charts. This ensures their clarity and detail remain intact even upon zooming. Furthermore, the network designs and other geographic-based outputs are transformed into interactive HTML maps, offering network operators a valuable tool for assessing results and making informed decisions [62]. This interactive presentation of data provides a comprehensive understanding of the network's performance, enhancing decision-making capabilities. In contrast with the predominant trend in related research, a significant portion of the algorithm definitions, implementations, and output indicator results for the case study, comprising numerical, on-map, and graphical data visualizations, are accessible via dedicated public repositories [63, 64]. This commitment to transparency and accessibility underscores the dedication to fostering collaboration and facilitating broader application of the research findings in real-world scenarios. All algorithms and data will be readily available to municipalities, empowering them to leverage the outcomes of this thesis as needed.

Finally, an important aspect of our methodology involves ensuring transparent access to the data outputs generated through this thesis. This includes the algorithm definitions, implementations, and the resultant output indicator findings for the case studies. We are committed to making all this information openly accessible through dedicated public repositories. By doing so, municipalities and network managers will have direct access to these algorithms and data, empowering them to leverage the insights and findings of this thesis as per their requirements.

Chapter 4

Case-study scenarios

This chapter explains the reasons behind selecting the cities of Girona and Lloret de Mar as suitable case studies for this thesis's objectives. It also provides a concise description of city characteristics.

4.1 Justification of the case-study scenarios

Girona has been strategically chosen as the primary focus of this thesis, prominently featured across all four papers comprising its content (Sections 5.1 - 5.4). The decision to select this city stems from several key factors. Firstly, Girona benefits from proximity and strong collaborative ties with both the Girona City Hall and Cicle de l'Aigua del Ter S.A. (CATSA), the company responsible for managing water systems in the area. This facilitates access to essential data and expertise crucial for comprehensive research. Additionally, Girona's status as a moderately sized city, with approximately 103,369 inhabitants, renders it an ideal case study. Its scale strikes a delicate balance, offering sufficient complexity to rigorously test algorithmic approaches without introducing unnecessary computational burdens while remaining grounded enough to ensure the practical applicability of findings. Hence, Girona serves as an excellent location, offering a compelling model for exploring innovative graph-theory solutions within the context of urban water networks.

Furthermore, Girona has faced significant challenges beyond the context of the COVID-19 pandemic, notably grappling with severe drought conditions prevalent in the Mediterranean region. Since the onset of 2024, Girona has been formally declared to be in a state of drought emergency, a condition that persists to this day, with water resources in Catalonia's internal basins dwindling to a mere 15% capacity as of March 28, 2024 [65].

Compounded by these circumstances is the absence of an operational reclaimed water distribution network (WDN) in Girona. This notable gap in infrastructure renders the city an even more compelling choice for inclusion as a case study within this thesis.

The selection of Lloret de Mar as an additional case study, particularly within the context of designing cost-effective reclaimed Water Distribution Networks (WDNs) as outlined in Section 5.2, is backed by various strategic considerations. Firstly, Lloret de Mar enjoys geographical proximity to Girona, situated within the Costa Brava region of the Girona province, facilitating collaboration with urban water network authorities. Moreover, the city faces significant water supply challenges, particularly during the summer months when its population surges from 39,089 to approximately 55,394 mean inhabitants due to tourism, employment, and educational activities. Unlike Girona, Lloret de Mar already operates a reclaimed WDN primarily utilized for public garden irrigation purposes, with plans underway for its expansion. This existing infrastructure provides a valuable basis for exploring more ambitious applications of reclaimed water, including irrigation for private gardens and toilet flushing. Hence, Lloret de Mar presents an intriguing alternative case study for devising comprehensive reclaimed WDN solutions.

Additionally, it is worth mentioning that the algorithms developed in this thesis have been applied in el Prat de Llobregat as part of the ReWaT project. El Prat de Llobregat features a reclaimed WDN, and the optimal algorithm's design solutions have been compared with the existing infrastructure, although not within the publications included in this thesis. This underscores the versatility and practicality of the solutions and algorithms outlined in this thesis, which have been validated across three real cities. Such successful applications pave the way for further expansion to include additional real scenarios in future works and projects.

4.2 Characterization of the case-study scenarios

The cities of Girona and Lloret de Mar, situated in Catalonia on the northeastern coast of the Iberian Peninsula, offer complementary urban landscapes distinguished by their size, density, and topography. Girona, home to 103,369 residents residing in 47,446 households (equating to an average of 2.4 citizens per household), exemplifies a typical Western Mediterranean city characterized by its compact layout, diverse land uses, and a clear division between the historic old town and the modern periphery (Figure 4.1). Its urban expanse spans 12.7 km^2, featuring a population density of 8,139 inhabitants per square kilometer, an average slope of 5.1, and an altitude range of 177 meters. In contrast, Lloret

de Mar, nestled along the northeastern Mediterranean coastline of Spain (Figure 4.2), boasts a year-round population of 39,089, with a seasonal influx increasing its equivalent population by 16,305 (resulting in 2.35 citizens per household). Encompassing an urban area of 7.8 km^2, the city exhibits a population density of 5,011 inhabitants per square kilometer, an average slope of 13.3, and an altitude range of 344 meters. Notably, Lloret de Mar's economy heavily relies on tourism, evidenced by its 120 hotels offering a total of 29,147 beds, with an average annual occupancy rate of approximately 65% in 2016 [66]. Moreover, the city attracts over one million visitors in a year alone [67].

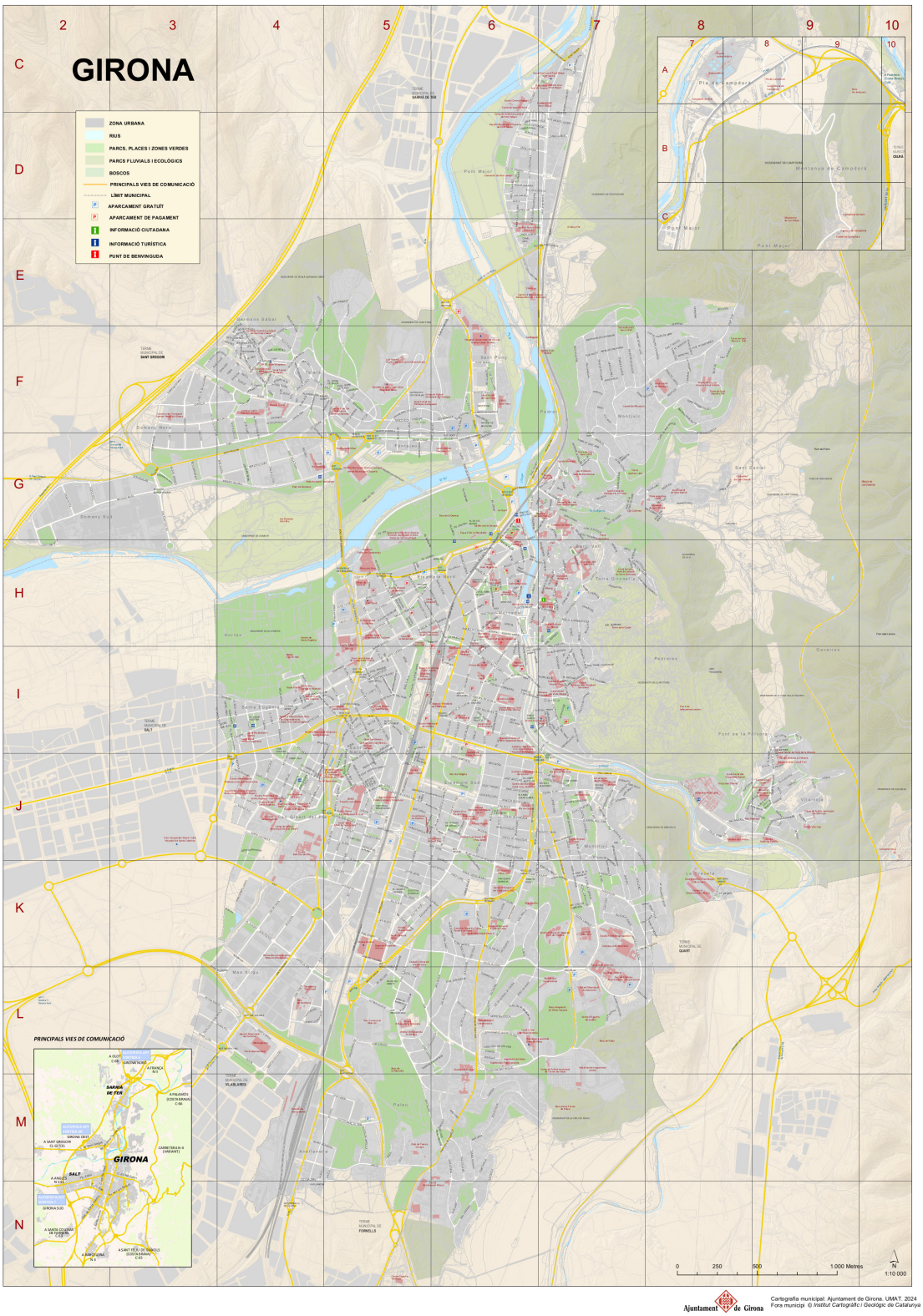


Figure 4.1: Basic information map of Girona 2024. **Source:** Ajuntament de Girona.

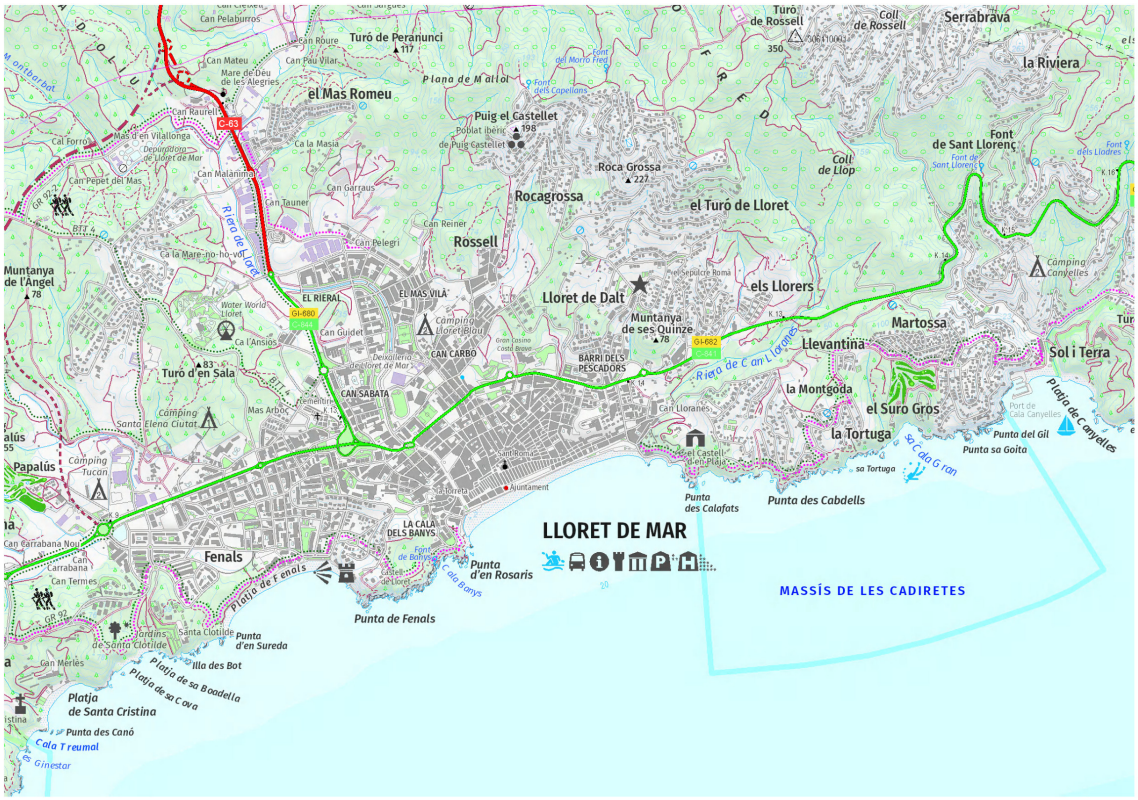


Figure 4.2: Basic information map of Lloret de Mar 2024. **Source:** Institut Cartogràfic i Geològic de Catalunya.

Chapter 5

Results

5.1 Optimal selection of monitoring sites in cities for SARS-CoV-2 surveillance in sewage networks

Environment International 157 (2021) 106768

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Optimal selection of monitoring sites in cities for SARS-CoV-2 surveillance in sewage networks

Eusebi Calle [a], David Martínez [b], Roser Brugués-i-Pujolràs [a], Miquel Farreras [a], Joan Saló-Grau [a], Josep Pueyo-Ros [b], Lluís Corominas [b,*]

[a] Institute of Informatics and Applications, Universitat de Girona, Girona, Spain
[b] Catalan Institute for Water Research, Emili Grahit 101, 17003 Girona, Spain

ARTICLE INFO

Handling Editor: Frederic Coulon

Keywords:
COVID-19
Wastewater-based epidemiology
Sewer
Sensor placement
Graph theory

ABSTRACT

Selecting sampling points to monitor traces of SARS-CoV-2 in sewage at the intra-urban scale is no trivial task given the complexity of the networks and the multiple technical, economic and socio-environmental constraints involved. This paper proposes two algorithms for the automatic selection of sampling locations in sewage networks. The first algorithm, is for the optimal selection of a predefined number of sampling locations ensuring maximum coverage of inhabitants and minimum overlapping amongst selected sites (static approach). The second is for establishing a strategy of iterations of sample&analysis to identify patient zero and hot spots of COVID-19 infected inhabitants in cities (dynamic approach). The algorithms are based on graph-theory and are coupled to a greedy optimization algorithm. The usefulness of the algorithms is illustrated in the case study of Girona (NE Iberian Peninsula, 148,504 inhabitants). The results show that the algorithms are able to automatically propose locations for a given number of stations. In the case of Girona, always covering more than 60% of the manholes and with less than 3% of them overlapping amongst stations. Deploying 5, 6 or 7 stations results in more than 80% coverage in manholes and more than 85% of the inhabitants. For the dynamic sensor placement, we demonstrate that assigning infection probabilities to each manhole as a function of the number of inhabitants connected reduces the number of iterations required to detect the zero patient and the hot spot areas.

1. Introduction

There is increasing evidence that sewage is a good, unbiased indicator of the prevalence of a virus in a population. The ability to detect SARS-CoV-2 in sewage has been reported by research groups worldwide. Upon confirmation that COVID-19 patients shed SARS-CoV-2 in feces, different studies have provided significant correlation between the concentration of SARS-CoV-2 in sewage and the prevalence of COVID-19 in the corresponding population (Lenzen et al., 2020; Mallapaty, 2020; Medema et al., 2020; Schmidt, 2020). So far, the approach has been successful when monitoring at the wastewater treatment plant (WWTP) level (i.e. integrating all inhabitants from a municipality), but there is limited experience when bringing the approach to a neighborhood level. 'Upstream' surveillance for SARS-CoV-2 may facilitate finer spatial detection of the virus in catchments with differing COVID-19 disease burdens, and may help provide information about any mitigation actions implemented at the community level. An example of monitoring at the neighborhood level can be found in Wu et al. (2020) where 11 urban neighborhoods within the wastewater treatment facility's catchment, representing populations ranging from ~4,000 to ~40,000 individuals, were monitored. GIS (geographic information system) data with catchment outlines was used to aggregate the demographic information for the catchment. Yet, while the selection of the sampling points in Wu et al. (2020) serves the purpose of the study, this might not be optimal from the perspective of a municipality.

Monitoring the traces of SARS-CoV-2 in sewage at the intra-urban scale implies establishing a surveillance network inside the sewage network. Sewer systems are long complex networks of pipes. As an example, the total length of the sewage network across the EU has been estimated at around 3 M kilometers (EurEAU, 2017). A city of about 100,000 inhabitants might have around 300 km of small sewer pipes (building sewer pipes and lateral sewers) and 60 km of bigger main pipes (community sewers collecting sewage from the lateral sewers and transporting it to the WWTPs). It is then not evident where to place

* Corresponding author.
E-mail addresses: eusebi.calle@udg.edu (E. Calle), dmartinez@icra.cat (D. Martínez), u1059680@campus.udg.edu (R. Brugués-i-Pujolràs), miquel.farreras@udg.edu (M. Farreras), u1953621@campus.udg.edu (J. Saló-Grau), jpueyo@icra.cat (J. Pueyo-Ros), lcorominas@icra.cat (L. Corominas).

https://doi.org/10.1016/j.envint.2021.106768
Received 14 April 2021; Received in revised form 25 June 2021; Accepted 6 July 2021
Available online 12 July 2021

autosamplers (or sensors if available in the future) to monitor the concentration of the RNA traces of SARS-CoV-2. Selecting sampling/sensor placement locations can follow several criterion. Given a predefined number of sampling locations to be installed, a municipality might be interested in ensuring maximum coverage of manholes (or of inhabitants) and ensuring a balance in the number of inhabitants covered by each of them (static monitoring site selection from now on). Another approach would be to establish a monitoring procedure to detect the area where the virus is more present (dynamic monitoring site selection from now on) or the zero patient if the analytical method for SARS-CoV-2 would be sufficiently sensitive to detect one virus shedder in an entire community. For either of the two approaches, it is relevant to collect demographic and socioeconomic indicators for the community connected to a specific sampling point. Otherwise, it is not possible to correlate the virus concentration to the number of diagnosed cases, for instance, or to socioeconomic indicators. Furthermore, the cost of autosampler/sensor ownership, maintenance efforts in particular, can still be cost-prohibitive and a balance between the number of sampling locations and costs needs to be guaranteed.

Larson et al. (2020) is the only paper about sampling points selection in sewers related to SARS-CoV-2; the authors propose two strategies to detect the zero patient and to identify zones with high levels of infection. Larson et al. (2020) assume that near-real time SARS-CoV-2 concentrations can be measured, for instance by employing fast tests (Mao et al., 2020b) (Mao et al., 2020a) which are under development but not yet available. Existing approaches to analyze SARS-CoV-2 concentrations imply lab analyzes and deliver results in 24–72 h. Wang et al. (2020) propose adaptive sampling site allocation for the sewage surveillance of the pathogen S. Typhi, by which the locations of sampling sites are dynamically updated to increase the probability of detecting a positive signal of the pathogen. It uses a model to simulate pathogen shedding, pathogen transport and fate in the sewage network, sewage sampling, and detection of the pathogen. Wang et al. (2020) propose stratified sampling for the initial selection of sampling sites, by which the geographic area is divided into a certain number of subareas and one sampling unit is randomly selected from each subarea. Within the scope of wastewater-based epidemiology (not related to SARS-CoV-2) the work from Matus et al. (2019) proposed monitoring sites selection using GIS analysis with city-wide demographic and sewage network information. The approach was semi-automatic, based on the definition of constraints, and did not use optimization.

Monitoring site selection (or sensor placement) has been widely studied for drinking water networks, but only a few studies exist on sewage networks. Kang et al. (2013) determined key sensor locations for non-point pollutant sources management in sewage networks by means of clustering analysis and ANOVA on top of SWMM simulated results. A few examples exist on sensor placement for illicit intrusion detection in sewage networks based on single and multi-objective optimization (Yazdi, 2018) or on Bayesian decision networks (Sambito et al., 2020). Vonach et al. (2018) proposed best sampling locations with the objective of calibrating a hydrodynamic model. Finally, Villez et al. (2016) and Villez et al. (2020) proposed methodologies for sensor placement in WWTPs based on graph theory and mass balances for maximizing the ability to assess and control data quality while minimizing the cost of ownership. Given this background, this paper proposes two algorithms for SARS-CoV-2 monitoring site selection in sewage networks. The algorithms are based on graph-theory and are coupled to a greedy optimization algorithm. To the best of our knowledge, this is the first paper which proposes an algorithm for static SARS-CoV-2 monitoring site selection. Furthermore, this paper enhances the approach proposed in Larson et al. (2020) for dynamic monitoring site selection by i) defining infection probabilities as a function of the number of inhabitants connected to each manhole and ii) evaluating the benefit of combining the static sensor placement outcomes to the dynamic placement algorithms. This paper as well contributes to enhance the selection of the initial selection sites from Wang et al. (2020) by proposing a static sensor placement algorithm which uses optimization. The usefulness of the two types of algorithms is illustrated with a case study in the city of Girona.

2. Materials and methods

This section describes the general methodology followed to obtain optimal monitoring sites for SARS-CoV-2 surveillance and includes a description of the algorithms proposed and the description of the case study used to illustrate their usefulness.

2.1. General methodology

The overall approach for obtaining optimal monitoring site selection for SARS-CoV-2 surveillance involves the following steps: i) goal and scope definition, ii) data collection, iii) graph generation, iv) linking demographic and socioeconomic indicators to the graph, v) implementation and execution of the algorithms, and vi) analysis of results. Fig. 1 describes the process flow of this general methodology. Details on the actual application of the general methodology to the specific case-study are provided in Section 2.3.

Goal and scope definition. This first step consists of defining the goal and scope of the study which, in turn, will influence all subsequent steps in terms of data intensity, data quality, algorithm selection, etc. Some examples of "goal and scope" are given: i) monitoring the spread of the COVID-19 disease in different communities with homogenous socio-economic status of a city during a pandemic; ii) identify in a given city a hotspot area with much higher disease prevalence than others. This step also involves the definition of the monitoring site selection needs (e.g. static vs dynamic), criteria and constraints together with the client profile (municipality or local health authority). Criteria are related to demographic, environmental, health or economic indicators; as an example, population density or socio-economic status of inhabitants are criteria which can be used to make the decision on the sampling sites selection. Constraints can be applied to the criteria, but also can be related to physical constraints in given manholes which do not allow to install equipment for wastewater sampling.

Data collection. The following data are collected: i) sewage network topology; ii) Digital Elevation Model (DEM) of the case under study; iii) cadastral data from the city parcels and iv) demographic and socioeconomic indicators associated to each cadastral parcel.

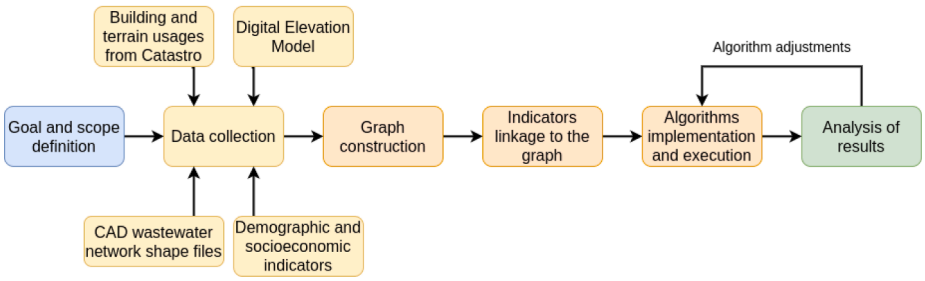


Fig. 1. General methodology flow.

Graph construction. The sewage network topology is transformed into a graph, where the edges represent the pipes and the nodes represent the manholes. The original sewage network topology must include the pipes, manholes, elevations and direction of the wastewater flow. In most cases, wastewater companies keep this data in GIS databases, which means an additional step is required to transform that data into a graph. The key elements are the coordinates of the nodes and the relationship between the nodes and edges. Data verification and reconciliation is essential at this stage to verify that all the pipes that are supposed to be connected through a manhole are actually connected, and to validate the sewer network slopes.

Indicators linkage to the graph. The physical attributes of the graph (e.g. topological such as node elevation) are assigned. The challenge here is to connect inhabitants (virus shedders) to the nearest manhole. This also provides the possibility to link the manhole with some socioeconomic indicators fed by census or other socioeconomic databases. Hydraulic models can be a source of data (e.g. biomarkers travel time) which might be used as a criteria or constraint by the algorithms. Yet, hydraulic models have not been used in this case-study.

Implementation and execution of the algorithms. Static and dynamic monitoring site selection require computationally expensive algorithms and thus need to be implemented in a relevant computing environment to achieve simulation results at reasonable times. This is of special interest in the case of dynamic monitoring site selection.

Results analysis. The results are analyzed, and if the targets set up in the goal and scope phase are satisfied, the project can be accomplished. Otherwise, a feedback loop to previous phases is needed until the goals are reached, as shown in Fig. 1. Iterations are conducted in the feedback loop to enhance the performance of the algorithms in terms of results and computational time. The final algorithms deployed in this study are transferable to other case-studies with no further upgrade; for transferability it is only necessary to construct the graph of each new case-study and run the algorithms developed in this paper.

2.2. Static and dynamic algorithms

The algorithms are based on techniques borrowed from graph theory which allow pipe network configurations to be analyzed (Kesavan and Chandrashekar, 1972). In the case of sewage networks, the pipes correspond to the graph edges and the manholes represent the graph nodes. We provide algorithms for both static and dynamic monitoring site selection. The static challenge implies selecting the locations for permanent monitoring sites which meet the client's (e.g. a municipality or health agency) needs (e.g. maximum coverage at minimal investment). In the case of SARS-CoV-2 sewage surveillance, permanent monitoring sites can be selected where 24 h composite samples are taken and brought to the lab for microbiological analysis that will deliver results in between 24 and 72 h (Ahmed et al., 2020; Rusiñol et al., 2020). The concept of interference is applied in the static monitoring site selection. Interferences appear when two or more monitoring sites overlap in covered manholes. In some cases, interferences might appear as a result of the graph being improperly constructed, and hence a reconciliation exercice is needed before launching the optimization algorithms. The dynamic monitoring site selection challenge implies dynamically and adaptively developing a sequence of manholes to sample and test until it finds the manhole in which the first infected person in a city is connected, or until the group of manholes in a city in which the largest group of infected people are connected is found. For the dynamic approach we build on what was presented in Larson et al. (2020) but also include one enhancement. In Larson et al. (2020), the Bayesian probabilities of infection are equally assigned to all manholes, whereas in our proposal, the Bayesian probabilities are assigned according to the number of inhabitants connected to each manhole; hence we assume that there is a greater chance of finding the origin of infection there. It is assumed in the dynamic monitoring site selection that a portable and fast analytical method (results delivered in few minutes) is available to guarantee several sample-and-test iterations to be executed in a short period of time. The turnaround time for an iteration of sampling, testing and adjustment should ideally be of 24 h; since the concentrations of SARS-CoV-2 can change during the course of a day (and the dynamics are even more pronounced at the community level), it is recommended to analyze 24-h composite samples (Medema et al., 2020), and then the analysis of SARS-CoV-2 concentrations and the launch of the proposed algorithms should take less than an hour. The whole process for detecting the hot spot should ideally be shorter than 1 week.

Table 1 specifies the notation used for the static and dynamic algorithms. In brief, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the sewage network graph, with a V-element set of nodes $\mathcal{V}$ representing manholes, and an E-element set of links $\mathcal{E} \subset \mathcal{V}^{|2|}$ representing pipes. Additionally, $\mathcal{S}$ (where $\mathcal{S} \subseteq \mathcal{V}$) denotes an S-element set of nodes with placed monitoring sites.

2.2.1. Static monitoring site selection algorithm

A novel algorithm called monitoring site selection (MSS, see Algorithm 1) is proposed. The MSS algorithm presents a greedy approach that optimizes the placement of several K SARS-CoV-2 monitoring sites within the sewage network nodes, dividing the network into K monitoring sites areas or subgraphs. We define the set of nodes that are present in the monitoring site k coverage area (i.e., source nodes) as $\mathcal{C}(k)$, $k \in \mathcal{K}$.

Table 1
Notation concerning the static and dynamic algorithms.

K	number of monitoring sites to place; $1 \leqslant K \leqslant V$ (fixed parameter)
$\mathcal{K} = \{1,2,\ldots,K\}$	set of (indices) of the monitoring sites
$\mathcal{N}(v), v \in \mathcal{V}$	set of neighbor nodes of the node v; $\mathcal{N}(v) \subseteq \mathcal{V} \setminus \{v\}$
$\mathcal{C}(k), k \in \mathcal{K}$	set of nodes that are present in the monitoring site k coverage area (i.e., source nodes); $\mathcal{C}(k) \subseteq \mathcal{V}$
$\mathcal{C}$	set of nodes that are present in at least one monitoring site coverage area; $\mathcal{C} := \bigcup_{k \in \mathcal{K}} \mathcal{C}(k)$; The size (number of nodes) of this set is called **coverage** $C = \lvert\mathcal{C}\rvert$
$\mathcal{I}(k), k \in \mathcal{K}$	set of nodes in the monitoring site k coverage area that are present also in at least another monitoring site j coverage area; $\mathcal{I}(k) \subseteq \mathcal{C}(k)$; $\mathcal{I}(k) := \{v \in \mathcal{C}(k) \cap \mathcal{C}(j) : j \in \mathcal{K}, j \neq k\}$
$\mathcal{I}$	set of nodes that are present in at least two monitoring site coverage areas; $\mathcal{I} := \bigcup_{k \in \mathcal{K}} \mathcal{I}(k)$; The size (number of nodes) of this set is called **interference** $I = \lvert\mathcal{I}\rvert$
$\mathcal{U}(k), k \in \mathcal{K}$	set of nodes that are present only and exclusively in the monitoring site k coverage areas; $\mathcal{U}(k) := \{v \in \mathcal{C}(k) \setminus \mathcal{I}(k)\}$
$\mathcal{U}$	set of nodes that are present in one and only one monitoring site coverage area; $\mathcal{U} := \bigcup_{k \in \mathcal{K}} \mathcal{U}(k)$; The size (number of nodes) of this set is called **unique coverage** $U = \lvert\mathcal{U}\rvert$
C_{max}	number of nodes of the largest monitoring site coverage area; $C_{max} := \max(\lvert\mathcal{C}(k)\rvert, k \in \mathcal{K})$
C_{min}	number of nodes of the smallest monitoring site coverage area; $C_{min} := \min(\lvert\mathcal{C}(k)\rvert, k \in \mathcal{K})$
D	number of nodes **difference** between the largest and the smallest monitoring site coverage areas; $D := C_{max} - C_{min}$
$\mathcal{A}$	set of **artificial** source nodes in the area covered by a sensor with $deg^- := 0$ and population associated to it.
$\mathcal{P}$	set of nodes in the area covered by a sensor after **normalising** and **simplifying** $\mathcal{C}$.
$\mathcal{T}$	set of nodes of Hot Spot neighborhood/node of Patient Zero in the area covered by a sensor. For PZ, $\lvert\mathcal{T}\rvert := 1$.

The MSS algorithm starts from a random combination of nodes with placed monitoring sites $\mathcal{S}$, and then iterates to find better combinations by moving each monitoring site $s \in \mathcal{S}$ through its neighbouring nodes. We propose an evaluation function called monitoring site selection evaluation (MSE, see Function 1) which allows an optimization metric after each execution of the MSS algorithm to be estimated. The MSS algorithm halts when it is not possible to find an $\mathcal{S}' \neq \mathcal{S}$ monitoring site set by which for all neighboring nodes of each $s \in \mathcal{S}$ the optimization metric does not improve as compared to the $\mathcal{S}$ set.

Algorithm 1. Monitoring site selection (MSS) algorithm.

Input: K: number of monitoring sites to place.
$\mathcal{G} \leftarrow \{\mathcal{V}, \mathcal{E}\}$: wastewater network with node set $\mathcal{V}$ and link set $\mathcal{E}$.
$\mathcal{C}(k), k \in \mathcal{K}$: set of nodes that are present in each monitoring site area k.
$\mathcal{N}(v), v \in \mathcal{V}$: set of neighbor nodes of each node v.
Output: $\mathcal{S}, \forall_{s \in \mathcal{S}}\ s \in \mathcal{V}, S = K$: set of nodes with collocated monitoring sites ($S = |\mathcal{S}|$).
O: optimization value of the monitoring sites placed on the node set $\mathcal{S}$.
1. Obtain a random sample of nodes with collocated monitoring sites $\mathcal{S}$ with K elements of $\mathcal{V}$.
2. Compute the optimization value O for the sample $\mathcal{S}$.
3. $\forall s \in \mathcal{S}$ picked randomly:
   (a) $\forall n \in \mathcal{N}(v), n \notin \mathcal{S}$ picked randomly:
      i. Obtain a new sample $\mathcal{X}$ removing s and adding n ($\mathcal{X} \leftarrow \mathcal{S} \bigcup \{n\} \setminus \{s\}$).
      ii. Compute the optimization value P for sample $\mathcal{X}$.
      iii. If $P > O$, set $\mathcal{S} \leftarrow \mathcal{X}, O \leftarrow P$, and go to **step 3)**.
4. $\mathcal{S}$ contains the resulting set of nodes with collocated monitoring sites, and O contains the optimization value of the placed monitoring sites.

The MSS searches for large, non-interference, equal-sized coverage areas. Starting from $\mathcal{S}$ monitoring sites, the MSE maximizes the unique coverage U and minimizes the difference D between the maximum C_{max} and minimum C_{min} sizes of the resulting network coverage areas. These measures are normalized taking into account the total number of network nodes V. The U measure is proposed in order to take into account the interference I between the coverage areas of each monitoring site that we want to minimize. In that way, the maximization of the unique coverage U also minimizes the interference I between coverage areas.

Function 1. Monitoring sites evaluation (MSE) function.

Input: $\mathcal{G} \leftarrow \{\mathcal{V}, \mathcal{E}\}$: wastewater network with node set $\mathcal{V}$ and link set $\mathcal{E}$, $V = |\mathcal{V}|$.
$\mathcal{K}$: set of (indices) of the monitoring sites to test, $K = |\mathcal{K}|$.
$\mathcal{C}(k), k \in \mathcal{K}$: set of nodes that are present in the monitoring site k coverage area; $\mathcal{C}(k) \in \mathcal{V}$.
Output: O: optimization value of the provided monitoring sites sample $\mathcal{K}$.
1. Initialize the set of nodes covered by a unique monitoring site $\mathcal{U}$. $\mathcal{U} \leftarrow \varnothing$
2. Initialize the set of interference nodes, covered by multiple monitoring sites $\mathcal{I}$. $\mathcal{I} \leftarrow \varnothing$
3. Obtain the maximum C_{max} and minimum C_{min} values of nodes covered by each single monitoring site $k, \forall k \in \mathcal{K}$. $C_{max} = \max(|\mathcal{C}(k)|, k \in \mathcal{K}), C_{min} = \min(|\mathcal{C}(k)|, k \in \mathcal{K})$
4. $\forall k \in \mathcal{K}$:
   (a) $\forall v \in \mathcal{C}(k)$:
      i. If the node $v \notin \mathcal{I}$ and $v \notin \mathcal{U}$, then add v to $\mathcal{U}$. $\mathcal{U} \leftarrow \mathcal{U} \bigcup \{v\}$
      ii. Otherwise if the node $v \notin \mathcal{I}$ and $v \in \mathcal{U}$, then remove v from $\mathcal{U}$ and add it to $\mathcal{I}$. $\mathcal{U} \leftarrow \mathcal{U} \setminus \{v\}$ $\mathcal{I} \leftarrow \mathcal{I} \bigcup \{v\}$
5. Compute and return the optimization value that is the difference between the unique coverage $U = |\mathcal{U}|$ and the difference between C_{max} and C_{min}. This is also normalized with V. There are also two weight variables w, y (1 by default) that could be modified in order to prioritize one measure over the other. $o \leftarrow \frac{(w \times U - y \times (C_{max} - C_{min}))}{V}$

The MSS needs to be computed several times (i.e., iterations) to find the best K node combination to place the monitoring sites on $\mathcal{S}$ according to the MSE. The optimal number of required iterations may vary depending on sewage network size and topology. It is up to the network administrator to define the number of iterations as an stop criteria upfront, that may be input manually by the user or an automated decision based on the accumulated O value improvement in the MSS algorithm iteration results.

2.2.2. Dynamic monitoring site selection algorithm

After running the MSE Algorithm 1 and obtaining a subgraph of the sewage network with a monitoring site (sensor) s as output (i.e., $\mathcal{G}(s)$), we then use the dynamic monitoring approach proposed in Larson et al. (2020) to home in on either a possible patient zero or the hot spot neighborhood in that subgraph when its sensor s detects SARS-CoV-2 RNA traces $\mathcal{F}$. Larson et al.'s approach consists of assigning Bayesian probabilities of infection to all possible *source nodes* based on professional beliefs and applying a "binary search" using these probabilities. Our implementation assigns the Bayesian probabilities according to inhabitants connected to each node as we believe that the higher the population in the area is, the greater the chances are of finding infected people.

Two algorithms have been implemented: i) the Patient Zero (PZ) algorithm and ii) the Hot Spot (HS) algorithm. The PZ algorithm assumes there exists only one case of COVID-19 in a community (Patient Zero) and tries to find the minimum sequence of manholes to test in order to locate that first source of infection. The HS algorithm, on the other hand, assumes that many individuals are already infected and seeks to find the cluster in the sewage network with the largest SARS-CoV-2 RNA load; in other words, locate the hot spot.

The HS algorithm works as follows. At each iteration it seeks for the manhole whose Bayesian probability of infection is the highest. After testing it, if the viral load $\mathcal{F}'$ is high compared to the previously tested manhole, we know that the infected area is upstream from this point and we can discard all the network nodes downstream. Otherwise, the upstream nodes are discarded. Hence, at each iteration the population associated to the remaining nodes is approximately the same as those associated to the eliminated ones. In order to simplify the simulation, for each iteration we assume that the viral load $\mathcal{F}'$ is boolean. The algorithm halts when the *stopping rule*, which is defined by the user, is reached. The PZ algorithm is a special instance of the HS algorithm, where the *stopping rule* is reached when there is only one source node left.

Both PZ and HS algorithms share a prior three-step process which has been defined below for the sake of clarity.

Given $\mathcal{G}(s) \leftarrow \{\mathcal{V}(s), \mathcal{E}(s)\}, c_v$ is the population associated to the node $v \in \mathcal{V}(s)$:

i **Create artificial nodes:** All *source nodes* must have $deg^-(v) = 0$. To achieve this, for each inner node in the graph (so-called "original node") with the associated population, we create a new node (so-called "artificial node") with the same associated population and connected to that original node. From any artificial node we can easily obtain its original node. Let $\mathcal{A}$ be the set with all artificial nodes from $\mathcal{G}(s)$.
(a) Create empty set $\mathcal{A}$
(b) $\forall v \in \mathcal{V}(s)$ such that $deg^-(v) > 0$ and $c_v > 0$, add new node v' and edge (v', v) to $\mathcal{G}(s)$, make the citizens $c_{v'} \leftarrow c_v$, define $original(v') \leftarrow v$, add v' to $\mathcal{A}$.
(c) Return $\mathcal{A}$

ii **Simplify and normalise:** Dispense with useless nodes for the calculus in order to simplify the graph and assign a Bayesian probability to the *source nodes* based on its associated population.
(a) $\forall v \in \mathcal{V}(s)$, if $deg^-(v) = 0$ and $c_v = 0$, remove it.
(b) $\forall v \in \mathcal{V}(s)$ with $deg^-(v) = 1, deg^+(v) = 1$ and $c_v = 0$, add edge to $\mathcal{G}(s)$ going from predecessor of v to the successor of v. Remove vertex v
(c) Let $\mathcal{P} \subset \mathcal{V}(s)$ be the set of all $v \in \mathcal{V}(s)$ such that $deg^-(v) = 0$.
(d) Normalise $\mathcal{P}$: $\forall p \in \mathcal{P}, probability(p) \leftarrow \frac{c_p}{\sum_{p \in \mathcal{P}} c_p}$
(e) Return $\mathcal{P}$

iii **Propagate probabilities:** Propagate probabilities from *source nodes* to all other nodes, such that each inner node's probability is the sum of probabilities of upstream nodes. Let $\mathcal{P}$ be the set obtained after **simplifying and normalising** $\mathcal{G}(s)$.

(a) $\forall v \in (\mathcal{V}(s) \setminus \mathcal{P})$, and $\mathcal{W}$ is all the predecessors of v : $probability(v) \leftarrow \sum_{w \in \mathcal{W}} \frac{probability(w)}{deg^{+}(w)}$

Algorithm 2. Hot Spot detection algorithm.

Input: $\mathcal{G}(s) \leftarrow \{\mathcal{V}(s), \mathcal{E}(s)\}$: Directed graph such that it only has one node $s \in \mathcal{S} | deg^{+}(s) = 0$ corresponding to *sensor node*.
$c_v \mid \forall v \in \mathcal{S}, c_v$ are the citizens of v.
$\mathcal{F}$: Viral load detected in *sensor node* s.
Output: $\mathcal{T} | \mathcal{T} \subset \mathcal{V}(s)$: Nodes of Hot Spot neighborhood.
1. $\mathcal{A} \leftarrow$ Create artificial nodes
2. $\mathcal{P} \leftarrow$ Simplify and normalise graph
3. Define *stopping rule*.
4. While not *stopping rule*:
   (a) Propagate probabilities
   (b) Find node $t \in (\mathcal{V}(s) \setminus \mathcal{A})$ such that $t \leftarrow \min_{v \in (\mathcal{V}(s) \setminus \mathcal{A})} \left| probability(v) - \mathcal{F}/2 \right|$. In case of a tie, choose the node with the larger *probability*.
   (c) Let $\mathcal{G}'$ be a subgraph of $\mathcal{G}(s)$ having t and all the ancestors of t.
   (d) Test node t.
      i. $\mathcal{F}' \leftarrow$ *detected viral load*
      ii. If $\mathcal{F}' \geqslant \mathcal{F}/2$ then $\mathcal{G}(s) \leftarrow \mathcal{G}', F \leftarrow F'$
      iii. Else, $\mathcal{G}(s) \leftarrow \mathcal{G}(s) \setminus \mathcal{G}', \mathcal{F} \leftarrow \mathcal{F} - \mathcal{F}'$
   (e) $\mathcal{P} \leftarrow$ Simplify and normalise graph
5. $\mathcal{T} \leftarrow \mathcal{V}(s) \setminus \mathcal{A}$
6. Return $\mathcal{T}$

2.3. Case study

The usefulness of the algorithms was illustrated with the sewage network of Girona (Girona, northern Catalonia, Spain). Girona is a city of 101,852 inhabitants, with a metropolitan area shaped by seven municipalities that together have a total of 148,504 inhabitants (Source: 2019 electoral roll). Girona is a typical compacted western Mediterranean city, with mixed uses and clearly divided between the old town and the modern peripheral. It extends 39.1 km^2 at the confluence of the Ter, Onyar, Galligants, and Güell rivers and has a population density of 2,605 inhab./km^2. The Girona sewage network consists of 9,718 manholes with a total number of 148,504 inhabitants connected on them, resulting in a large network of 13.79 km in diameter and with a total of 338 kms of pipes. The basic topological characteristics of the network are: 9,718 nodes (V); 10,185 edges (E); an average node degree of 2.1 ($\overline{D}$); a diameter of 9,718 ($\varnothing$); and an average shortest path length of 9,580 ($\overline{d}$).

The topological data from the community sewer network was provided by the municipality of Girona through GIS that included feature geometry, attributes, etc. First, these files were combined to generate a GraphML file format which is compatible with the Network Robustness Simulator (NRS) (BCDS, 2021) used for graph analysis. The output format is a unique file in GraphML format which contains both nodes and edges, including their attributes. GraphML is an XML based format (GraphML, 2001). Next, a data verification and reconciliation approach was followed. The obtained graph was then checked for inconsistencies in disconnected nodes and/or edges. It was also important to check the additional data for outliers and discuss possible errors with the water company to ensure greater precision.

The citizens living in a household are estimated to be 2.7 citizens per household. This assumption is taken from the ratio of inhabitants in 2019 in Girona (101,852 citizens) and the surrounding villages connected to the sewer system, including Salt (31,362 citizens), Vilablareix (2,897 citizens), Sarrià de Ter (5,170 citizens), Aiguaviva (756 citizens), Fornells de la Selva (2,650 citizens) and Sant Gregori (3,817 citizens), which gives a final total of 148,504 citizens. These data were obtained from the Catalan Statistics Institute (idescat, 2019) and are divided by the total number of households (53,466 households in 2019).

The cadastral parcels from the city were associated to each manhole. The cadastral database for Girona was downloaded from the official Spanish Spatial Data Infrastructure, which is based on the European INSPIRE Directive (2007/2/EC), and transformed into a geojson file that contained the geometries as well as the alphanumeric information linked to the cadaster parcels. Next, each cadastral parcel was linked to the nearest manhole following a negative slope, adhering to the assumption that water is transported by gravity. The official DEM from the Catalan Cartographic and Geological Institute at a 2×2 resolution (ICGC, 2020) was used to estimate the z coordinate of the centroid of each cadastral parcel and manhole; hence, each cadastral parcel was connected to the nearest sewer origin with an equal or lower elevation. However, to overcome EDM and other inaccuracies, when the distance between the parcel and the manhole exceeded a defined threshold (100 m in this case), the algorithm searched for the closest higher manholes in a progressive way (1 m added in each new search) until the distance was lower than the threshold or the maximum z tolerance was reached (3 m in this case). To run these calculations, the scripts were developed on a PyQGIS console on QGIS v. 3.10 (QGIS, 2008). The number of inhabitants connected to each manhole was used as an input to the dynamic sensor placement algorithm to assign the Bayesian probabilities of infection to the source nodes. It was not used as an input to the static sensor placement algorithm.

3. Results and discussion

Below, we discuss the numerical results that illustrate the considerations of this paper. For that purpose, we used the Girona sewage network instance, i.e., *girona-wastewater*.

3.1. Static monitoring site selection

The results obtained for the case study of Girona confirm that the MSS algorithm performed well. The solution obtained for each of the eight tests (each of them fixing the number of monitoring sites from one to eight, $K = \{1,2,\ldots,8\}$), result in a coverage (in manholes) larger than 60%, an interference smaller than 3% and a maximum difference of 25% amongst monitoring sites coverage (Table 2).

The solution obtained when fixing one monitoring site is arbitrary, as it corresponds to the selection of the sampling point at the end of the sewage network (the entrance of the WWTP) with 100% coverage and 0% interference. For the remaining tests, an optimal solution was found which balances coverage, interference and manholes' coverage equity amongst sites. The results show that when fixing two and three monitoring sites, the coverage reduces down to 65%; after fixing four or more monitoring sites the total coverage is always larger than 75% (Fig. 2a).

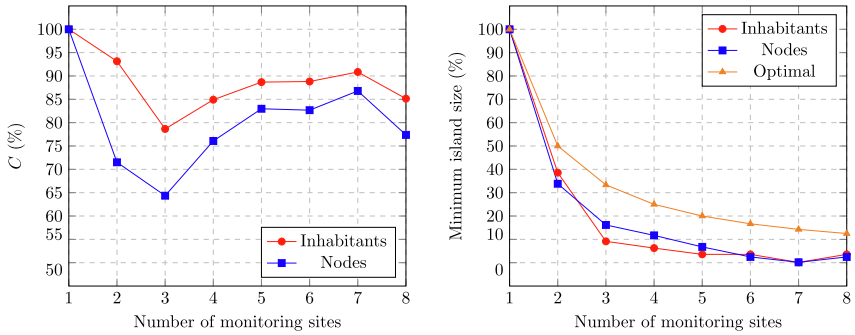
The algorithm uses the number of manholes as an input (not the inhabitants connected to each manhole). As the MSS algorithm minimizes the difference of the number of nodes between the largest and the smallest monitoring site coverage areas (D), our results show adequate minimum island sizes from one to five monitoring site placements (Fig. 2b) (up to a maximum of five monitoring site placements is recommended in the *girona-wastewater* network as a larger number of placements result on small-sized islands, which should be avoided). The minimum island sizes for nodes and inhabitants are compared with the theoretical optimal solution, which considers that all of the obtained monitoring site coverage areas are equally sized. The solutions provided show a good correlation between the number of manholes and inhabitants covered for each monitoring site coverage area for all tests, this is the particular case in the Girona catchment with a population density range of 17.7–759.9 inh/km^2 amongst neighborhoods.

The results show that the larger the number of monitoring sites the smaller the distance from the furthest node to the respective sampling points (ID values in the table). Given the potential attenuation of RNA signal along the sewage network transport Hart and Halden (2020) a constraint might be added in the selection of monitoring sites related to the maximum distance between the points of discharge of SARS-CoV-2 RNA traces and each monitoring site.

Table 2
MSS results data for *girona-wastewater* network, from one to eight monitoring sites.

K	*C* (%)	*I* (%)	*D* (%)	IN	WD (m)	ID (m)	IH	HC (%)
1	100	0	0	9718	0	13785	151248	100
2	71.51	2.45	6.4	3909, 3287	4561, 4457	6513, 9328	58331, 82561	93.15
3	64.35	0.01	8.3	2378, 2306, 1571	6247, 4452, 4618	7538, 6618, 4232	61007, 44102, 13845	78.65
4	76.07	0.01	12.76	2378, 2306, 1571, 1138	6247, 4452, 4618, 4182	7538, 6618, 4232, 3655	61007, 44102, 13845, 9468	84.91
5	82.97	0.01	17.83	658, 1571, 2306, 2391, 1138	812, 4618, 4452, 6021, 4182	3825, 4232, 6618, 7764, 3655	5404, 13845, 44102, 61313, 9468	88.68
6	82.67	0.1	21.24	1138, 240, 2143, 658, 1571, 2294	4182, 6255, 6253, 812, 4618, 4756	3655, 1465, 7531, 3825, 4232, 6319	9468, 7920, 53589, 5404, 13845, 44102	88.81
7	86.79	0.44	24.31	2306, 658, 408, 2379, 17, 1571, 1138	4452, 812, 4447, 6239, 5085, 4618, 4182	6618, 3825, 2778, 7546, 214, 4232, 3655	44102, 5404, 3401, 61007, 186, 13845, 9468	90.85
8	77.38	0.99	15.15	240, 1712, 1512, 1130, 596, 658, 1138, 630	6255, 6371, 6263, 4736, 5317, 812, 4182, 6301	1465, 4699, 4607, 4034, 2376, 3825, 3655, 7484	7920, 25075, 47133, 11067, 16254, 5404, 9468, 6456

K – number of placed monitoring sites, *C* (%) – normalized network nodes coverage (*C*, in %), *I* (%) – normalized network nodes interference (*I*, in %), D (%) – normalized difference between C_{max} and C_{min} (*D*, in %), IN – islands number of nodes, WD – distance between the WWTP and each node where the monitoring sites are placed (in meters), ID – islands diameter (i.e., monitoring sites coverage areas diameter, in meters), IH – island number of inhabitants (inhabitants size), HC – inhabitants coverage (in %).



(a) Coverage. (b) Minimum island sizes.

Fig. 2. Coverage and minimum island sizes (one to eight monitoring sites, *girona-wastewater*).

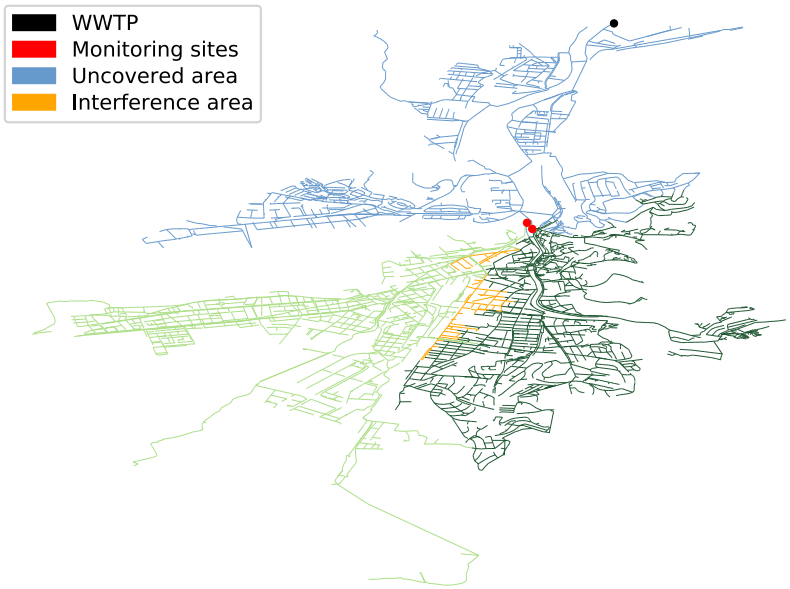
Fig. 3 shows the results for the tests that fixed two and five monitoring sites. When fixing two monitoring sites, the two covered areas (in green) represent 71% of the manholes. The uncovered areas (in blue) are the ones located further downstream in the sewage network (closest to the WWTP). When fixing five monitoring sites, the coverage increases (up to 83%) and smaller residential areas (as compared to the ones covered by the fixing two sites test) are included. Again, the manholes in the areas close to the WWTP cannot be captured in the final solutions because of their small number of manholes and their potential to generate interference as they are located downstream.

This is the first time that a static monitoring site selection algorithm has been proposed for SARS-CoV-2 monitoring at fine spatial resolution. The static monitoring site selection algorithm can effectively be applied to municipalities wishing to monitor SARS-CoV-2 RNA traces at a finer scale than an entire municipality. The presence of RNA traces is normally analyzed in tandem with health (number of infected cases, (Medema et al., 2020)), demographic and socioeconomic indicators (Wu et al., 2020). The module developed in this study which links each household to a manhole is essential to be able to aggregate these indicators (from individuals to inhabitants connected to a specific monitoring site). In the particular case of Catalonia, the public COVID-19 prevalence data is only aggregated at the municipality level and at the ABS (primary health area, which are areas defined around primary care health centers in the city) level. As an illustration exercise, we fixed the placement of sensors at the end of each ABS (the downstream manhole of each ABS subcatchment) and estimated the interference. Appendix A shows that locating sensors as a function of ABS results in large interference amongst monitoring sites and hence would not be the preferred option. Therefore, municipalities should make a request to the health authorities to aggregate prevalence data according to the areas covered by the sampling points. The main limitation to bringing the approach into practice is the data quality on the topology of the sewage network; data reconciliation is the most time-consuming step. Launching the optimization for each of the tests took between one and three minutes on a modern laptop (CPU Ryzen 5 4800U, 16 GB RAM). Finally, the approach is equally valid for the placement of monitoring sites for purposes other than tracking the spread of COVID-19, such as estimating the consumption of pharmaceuticals (Escolà Casas et al., 2021) and illicit drugs (González-Mariño et al., 2020) at fine spatial resolution, or detecting illicit discharges of pollutants from industries (Sambito et al., 2020).

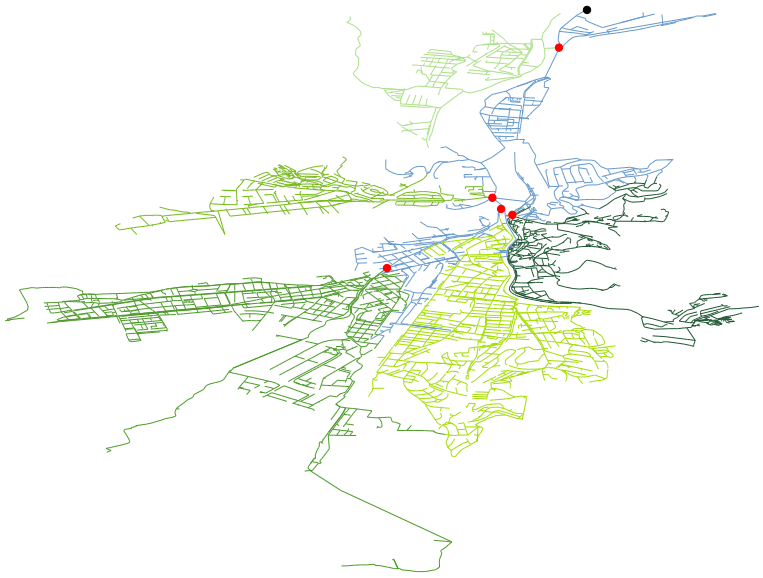
3.2. Dynamic monitoring site selection

The dynamic monitoring site selection algorithm is executed over both the entire sewage network (cases 1 and 2) and a reduced network, which includes 1,099 manholes (equivalent to 44,102 citizens), found as an outcome of the static monitoring algorithm (cases 3 and 4). For cases 1 and 3, the Bayesian probabilities of source nodes are assigned according to the connected inhabitants, while for cases 2 and 4 the probabilities are assigned using random numbers from a unit probability distribution. Cases 2 and 4 would be comparable to the methodological proposal from Larson et al. (2020).

Patient Zero. When applying the PZ algorithm to the entire network and with probabilities as a function of population (case 1), the sampling iterations required to identify the first individual discharging SARS-CoV-2 RNA traces in the sewage system range from 8 to 15. When the probabilities are randomly assigned (case 2) the range is smaller, between 10 and 14 iterations. Looking at the median of the distributions one less iteration would be needed when assigning the probabilities as a



(a) Two monitoring sites (71.51% cov.).



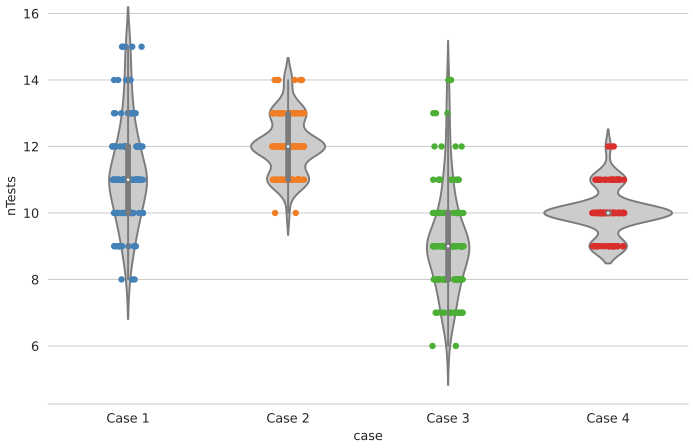
(b) Five monitoring sites (82.97% cov.).

Fig. 3. Coverage areas for two and five monitoring sites placement (*girona-wastewater*).

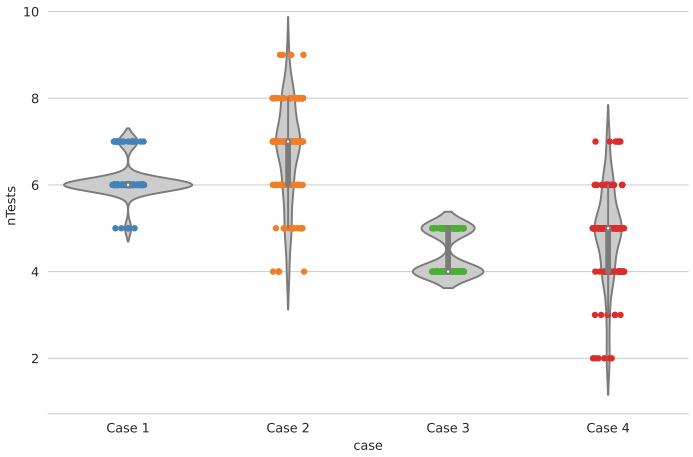
function of the population. This gain might increase in cities with larger variability in population density amongst neighborhoods (in the case of Girona, the population density across neighborhoods varies between 17.66 and 759.94 inh/km^2).

The application of the PZ algorithm to the reduced network results in a decrease of two iterations to identify the first individual discharging SARS-CoV-2 RNA traces (cases 3 and 4 as compared to 1 and 2). Overall, the strategy resulting in a smaller number of iterations is case 3, which involves departing from a reduced network resulting from the static monitoring site selection algorithm and with the assignment of probabilities as a function of connected inhabitants. The frequency distribution of the number of required samples for the PZ algorithm is shown in Fig. 4a.

Hot Spot. The HS algorithm allows hot spot areas with a large number of potential infected people to be identified (we set 3000 infected inhabitants as the *stopping rule* in this exercise). As compared to the PZ algorithm, the number of iterations reduces by five no matter the case. As shown in Fig. 4b, when assigning the probabilities as a function of the connected inhabitants, one less iteration is needed (looking at the median of the distributions); yet, the spread of the distribution of the resulting number of iterations is much smaller as compared to the assignment of uniform probabilities; this is the opposite of what is shown in Fig. 4a. The most favourable (and conservative) option would be case 3 with large certainty that with four or five iterations the hot spot area would be identified. As compared to case 1, one or two iterations would be saved when departing from a sewage network obtained from the



(a) Patient Zero algorithm.



(b) Hot Spot algorithm.

case 1 – entire wastewater network and Bayesian probabilities as a function of population.
case 2 – entire wastewater network and Bayesian probabilities randomly assigned.
case 3 – island from the MSS algorithm and Bayesian probabilities as case 1.
case 4 – island from the MSS algorithm and Bayesian probabilities as case 2.

Fig. 4. Distribution of sampling points required.

static approach.

The dynamic algorithms are comparable to the ones discussed in Larson et al. (2020), where the Bayesian probabilities are assigned randomly in the same way as we did in both cases 2 and 4. The enhancements in this paper relate i) to the use of the outcome from the static sensor placement algorithm as a starting point which allows the overall number of samples needed to be reduced in both PZ and HS algorithms, and ii) to the addition of information about the number of inhabitants to each manhole which also allows the number of samples on average to be reduced; the latter is relevant in cities like Girona, which show high spatial variability in population density (and hence in the number of inhabitants connected to a manhole). As stated in Larson et al. (2020), the usefulness of the dynamic approach to detect a hot spot within a city is constrained by the availability of devices which offer a fast response in the detection and quantification of SARS-CoV-2 RNA traces. Current methods imply the transport of the samples to a lab where the RNA traces are concentrated (e.g. (Forés et al., 2021)) and then the qPCR is executed, overall with a result being available in between 24 and 48 h. In case 3, iterations are needed to locate the Hot Spot, which means that between three and six days would be needed in total, which is probably too slow to make a decision on an effective mitigation action.

The applicability of the dynamic monitoring site selection algorithms is also constrained by the detection limit of the SARS-CoV-2 analysis in sewage. The lowest incidence resulting in quantifiable SARS-CoV-2 concentration in wastewater differed between community sizes; Rusiñol et al. (2021) found the lowest quantifiable incidence to be 0.11 and 0.82 cases per 1,000 inhabitants for the large and small sized communities respectively and Hata et al. (2021) reported 0.05–0.10 detectable cases per 1,000 inhabitants. Hart and Halden (2020) reported that under a best-case scenario of no in-sewer RNA signal loss, wastewater generation (50–500 L/person/d) and virus shedding (56.6 million-113.2 billion viromes/d) are important variables determining the detectability in community wastewater of a single infected person among one hundred to two million healthy individuals, assuming homogeneous distribution of cases. In the case of SARS-CoV-2 it is really challenging to

detect the zero patient; the patient zero might be moving around the city, but also the limit of detection of the analysis of SARS-CoV-2 in wastewater might not allow to detect 1 infected amongst the surveilled community. Results of patient zero are provided in this paper to compare the performance of the algorithm against Larson et al. (2020). Yet, the patient zero algorithm can be used for other purposes than SARS-CoV-2 surveillance, such as the detection of an illegal industrial discharges in sewer systems. Furthermore, there are substantial uncertainties in estimating SARS-CoV-2 loads (Li et al., 2021) which propagate to the calculation of increase or decrease of virus load between two collected samples; high-frequency flow-proportional sampling would reduce uncertainties (yet this is challenging at the intra-city scale) as well as using surrogate viruses as internal or external standards during the analysis, and further improvement on analytical approaches.

Future work will be conducted to connect the algorithm to a mechanistic model that includes SARS-CoV-2 concentration as a variable, using a hydraulic model to estimate the dilution capacity in the sewer network, a SARS-CoV-2 load generation pattern and implementing the in-sewer degradation of SARS-CoV-2. The concentration can then be used by the algorithm as a criteria or a constraint to define best sampling sites. Some attempts in that sense have been published for SARS-CoV-2 (Hart and Halden, 2020) and for other pathogens (Ranta et al., 2001; Wang et al., 2020).

4. Conclusions

This paper demonstrates that it is possible to optimally select sampling points for SARS-CoV-2 sewage surveillance in cities. An algorithm is proposed for the placement of a predefined number of monitoring sites which result in maximum coverage of manholes and minimum interference amongst them (static sensor placement). Two other algorithms are proposed to dynamically sample and analyze to identify patient zero and hot spots in cities (dynamic sensor placement). For the case study of Girona, a static sensor placement of five monitoring sites (or more) results in a coverage greater than 80% of both manholes and inhabitants. The best option for detecting a patient zero and a hotspot area implies assigning probabilities as a function of the number of inhabitants connected to each manhole. Results have demonstrated that when using these probabilities our proposed algorithms enhanced previous proposals in all presented scenarios. As a conclusion for the city of Girona, 11 iterations would be needed to detect the patient zero, and six iterations for identifying a hotspot of about 3,000 infected inhabitants. In the case of combining both algorithms, the number of iterations can be reduced to nine and four, respectively.

CRediT authorship contribution statement

Eusebi Calle: Conceptualization, Funding acquisition, Writing – review & editing, Supervision, Project administration. **David Martínez:** Methodology, Software, Formal analysis, Visualization, Writing – review & editing, Data curation. **Roser Brugués-i-Pujolràs:** Software, Visualization. **Miquel Farreras:** Software, Visualization. **Joan Saló-Grau:** Software, Visualization. **Josep Pueyo-Ros:** Methodology, Data curation, Software, Writing – review & editing. **Lluís Corominas:** Conceptualization, Funding acquisition, Investigation, Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

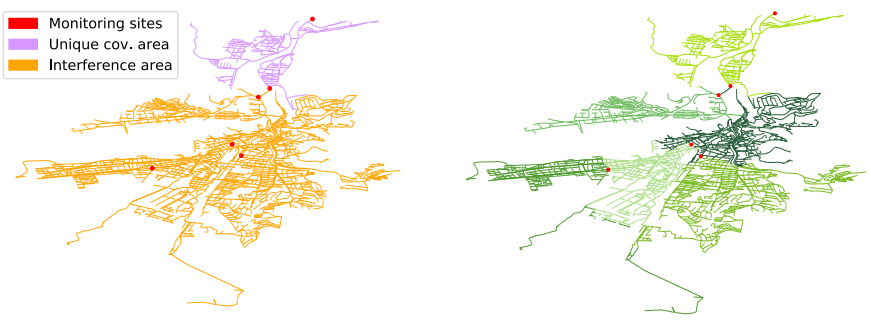
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Lluís Corominas acknowledges the Ministry of Economy and competitiveness for the Ramon and Cajal grant (RYC-2013-14595) and its corresponding I3 consolidation. The authors acknowledge the project VirWASTE (2020PANDE00044) funded by AGAUR. The authors acknowledge the CLEaN-TOUR (CTM2017-85385-C2-1-R) and INVEST (RTI2018-097471-B-C21) projects from the Spanish Ministry of Economy and Competitiveness and thank Generalitat de Catalu-nya through Consolidated Research Group 2017 SGR 1318. ICRA researchers thank funding from the CERCA program. UdG researchers thank funding from Red temática Go2Edge (Ref.: RED2018-102585-T) and Ajut PontUdG2020/23.

Appendix A. ABS monitoring site selection

ABS (Área Básica de Salut) or Basic Health Area, is the clustering method used by the Spanish government to divide a city into different areas. The main criteria is that all of them include at least one primary Health-care facility. A priori it would be interesting to establish the monitoring points considering these areas. However, if we locate the monitoring points considering only the coverage of these areas the 'interference' between them can report a huge error in the expected results. Fig. 5 shows the *girona-wastewater* network placing $k = 6$ monitoring sites, each one monitoring one of the six ABS areas on the network. Each ABS monitoring site is placed in the closest node from an ABS area to the sewage treatment plant. The large level of interference provided by this approach cannot be considered as a feasible solution (88.79% of interference depicted on the orange area (Fig. 5a)). This is produced, as expected, because part of the monitoring points are located closed to the WWTP, covering by themselves the major part of the city.



(a) ABS interference area. (b) ABS areas.

Fig. 5. ABS monitoring site selection ($k = 6$, 88.79% of interference).

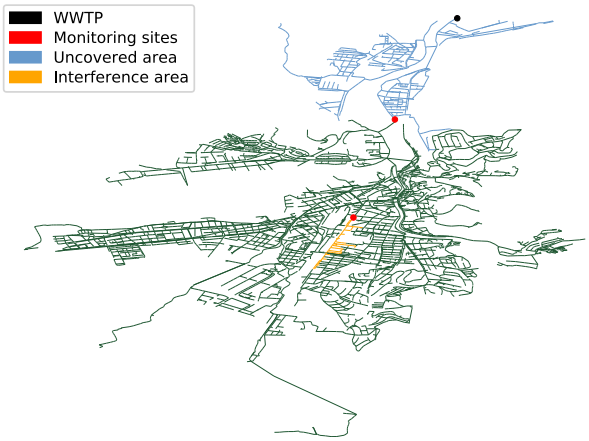


Fig. 6. Interference between two ABS areas.

Consequently, the rest of the monitoring points, located downstream from the WWTP, overlap the covered area. Moreover, in the case of reducing the number of ABS monitored areas, the interference between areas would still persist. This is described in Fig. 6 where the six ABS areas (in green) are also shown. In this case, the interfence area reports a percentage of 1.22. Consequently, it is shown that using ABSs as a clustering method to monitor the city is not possible if the results have to be relevant (without interferences between areas).

References

Ahmed, W., Bertsch, P.M., Bivins, A., Bibby, K., Farkas, K., Gathercole, A., Haramoto, E., Gyawali, P., Korajkic, A., McMinn, B.R., Mueller, J.F., Simpson, S.L., Smith, W.J., Symonds, E.M., Thomas, K.V., Verhagen, R., Kitajima, M., 2020. Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. Sci. Total Environ. 739 (June), 139960.

BCDS, 2021. Network Robustness Simulator. Universitat de Girona, BCDS Research Group. http://nrs.udg.edu [Online; accessed 16-mar-2021].

Escolà Casas, M., Schröter, N.S., Zammit, I., Castaño-Trias, M., Rodriguez-Mozaz, S., Gago-Ferrero, P., Corominas, L., 2021. Showcasing the potential of wastewater-based epidemiology to track pharmaceuticals consumption in cities: Comparison against prescription data collected at fine spatial resolution. Environ. Int. 150, 106404.

EurEAU, 2017. Europe's water in figures. An overview of the European drinking water and waste water sectors. Technical report. The European Federation of National Associations of Water Services.

Forés, E., Bofill-Mas, S., Itarte, M., Martínez-Puchol, S., Hundesa, A., Calvo, M., Borrego, C.M., Corominas, L.L., Girones, R., Rusiñol, M., 2021. Evaluation of two rapid ultrafiltration-based methods for SARS-CoV-2 concentration from wastewater. Sci. Total Environ. 768, 144786.

González-Mariño, I., Baz-Lomba, J.A., Alygizakis, N.A., Andrés-Costa, M.J., Bade, R., Bannwarth, A., Barron, L.P., Been, F., Benaglia, L., Berset, J.D., Bijlsma, L., Bodík, I., Brenner, A., Brock, A.L., Burgard, D.A., Castrignanò, E., Celma, A., Christophoridis, C.E., Covaci, A., Delémont, O., Devoogt, P., Devault, D.A., Dias, M. J., Emke, E., Esseiva, P., Fatta-Kassinos, D., Fedorova, G., Fytianos, K., Gerber, C., Grabic, R., Gracia-Lor, E., Grüner, S., Gunnar, T., Hapeshi, E., Heath, E., Helm, B., Hernández, F., Kankaanpaa, A., Karolak, S., Kasprzyk-Hordern, B., Krizman-Matasic, I., Lai, F.Y., Lechowicz, W., Lopes, A., de Alda, M.L., López-García, E., Löve, A.S., Mastroianni, N., McEneff, G.L., Montes, R., Munro, K., Nefau, T., Oberacher, H., O'brien, J.W., Oertel, R., Olafsdottir, K., Picó, Y., Plósz, B.G., Polesel, F., Postigo, C., Quintana, J.B., Ramin, P., Reid, M.J., Rice, J., Rodil, R., Salgueiro-Gonzàlez, N., Schubert, S., Senta, I., Simões, S.M., Sremacki, M.M., Styszko, K., Terzic, S., Thomaidis, N.S., Thomas, K.V., Tscharke, B.J., Udrisard, R., van Nuijs, A.L., Yargeau, V., Zuccato, E., Castiglioni, S., Ort, C., 2020. Spatio-temporal assessment of illicit drug use at large scale: evidence from 7 years of international wastewater monitoring. Addiction 115 (1), 109–120.

GraphML, 2001. The GraphML File Format. http://graphml.graphdrawing.org/ [Online; accessed 21-dec-2020].

Hart, O.E., Halden, R.U., 2020. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. Sci. Total Environ. 730, 138875.

Hata, A., Hara-Yamamura, H., Meuchi, Y., Imai, S., Honda, R., 2021. Detection of SARS-CoV-2 in wastewater in Japan during a COVID-19 outbreak. Sci. Total Environ. 758, 143578.

ICGC, 2020. Digital Elevation Model 5x5m. https://www.icgc.cat/Descarregues/Elevacions/Model-d-elevacions-del-terreny-de-5x5-m [Online; accessed 21-dec-2020].

idescat, 2019. El municipi en xifres. https://www.idescat.cat/emex/?id=170792 [Online; accessed 21-dec-2020].

Kang, O.Y., Lee, S.C., Wasewar, K., Kim, M.J., Liu, H., Oh, T.S., Janghorban, E., Yoo, C.K., 2013. Determination of key sensor locations for non-point pollutant sources management in sewer network. Korean J. Chem. Eng. 30 (1), 20–26.

Kesavan, H.K., Chandrashekar, M., 1972. Graph-Theoretic Models for Pipe Network Analysis. J. Hydraulics Div. 98.

Larson, R.C., Berman, O., Nourinejad, M., 2020. Sampling manholes to home in on SARS-CoV-2 infections. PLoS ONE 15.

Lenzen, M., Li, M., Malik, A., Pomponi, F., Sun, Y.Y., Wiedmann, T., Faturay, F., Fry, J., Gallego, B., Geschke, A., Gómez-Paredes, J., Kanemoto, K., Kenway, S., Nansai, K., Prokopenko, M., Wakiyama, T., Wang, Y., Yousefzadeh, M., 2020. Global socio-economic losses and environmental gains from the coronavirus pandemic. PLoS ONE 15 (7 July), 1–13.

Li, X., Zhang, S., Shi, J., Luby, S.P., Jiang, G., 2021. Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. Chem. Eng. J. 415, 129039.

Mallapaty, S., 2020. How sewage could reveal true scale of coronavirus outbreak. Nature 580 (9), 176–177.

Mao, K., Zhang, H., Yang, Z., 2020a. An integrated biosensor system with mobile health and wastewater-based epidemiology (iBMW) for COVID-19 pandemic. Biosens. Bioelectron. 169 (January).

Mao, K., Zhang, H., Yang, Z., 2020b. Can a Paper-Based Device Trace COVID-19 Sources with Wastewater-Based Epidemiology? Environ. Sci. Technol. 54 (7), 3733–3735.

Matus, M., Duvallet, C., Soule, M.K., Sean M. Kearney, S., Endo, N., Ghaeli, N., Brito, I., Ratti, C., Kujawinski, E.B., Alm, E.J., 2019. 24-hour multi-omics analysis of residential sewage reflects human activity and informs public health. bioRxiv preprint.

Medema, G., Been, F., Heijnen, L., Petterson, S., 2020. Implementation of environmental surveillance for SARS-CoV-2 virus to support public health decisions: Opportunities and challenges. Curr. Opin. Environ. Sci. Health 17, 49–71.

QGIS, 2008. A Free and Open Source Geographic Information System. https://www.qgis.org/en/site/ [Online; accessed 21-dec-2020].

Ranta, J., Hovi, T., Arjas, E., 2001. Poliovirus surveillance by examining sewage water specimens: Studies on detection probability using simulation models. Risk Anal. 21 (6), 1087–1096.

Rusiñol, M., Martínez-Puchol, S., Forés, E., Itarte, M., Girones, R., Bofill-Mas, S., 2020. Concentration methods for the quantification of coronavirus and other potentially pandemic enveloped virus from wastewater. Curr. Opin. Environ. Sci. Health 17, 21–28.

Rusiñol, M., Zammit, I., Itarte, M., Forés, E., Martínez-Puchol, S., Girones, R., Borrego, C., Corominas, L., Bofill-Mas, S., 2021. Monitoring waves of the COVID-19 pandemic: Inferences from WWTPs of different sizes. Sci. Total Environ. 787, 147463.

Sambito, M., Di Cristo, C., Freni, G., Leopardi, A., 2020. Optimal water quality sensor positioning in urban drainage systems for illicit intrusion identification. J. Hydroinformat. 22 (1), 46–60.

Schmidt, C., 2020. Watcher in the wastewater. Nat. Biotechnol. 38 (8), 917–920.

Villez, K., Vanrolleghem, P.A., Corominas, L., 2016. Optimal flow sensor placement on wastewater treatment plants. Water Res. 101 (1), 75–83.

Villez, K., Vanrolleghem, P.A., Corominas, L., 2020. A general-purpose method for Pareto optimal placement of flow rate and concentration sensors in networked systems – With application to wastewater treatment plants. Comput. Chem. Eng. 139, 106880.

Vonach, T., Tscheikner-Gratl, F., Rauch, W., Kleidorfer, M., 2018. A heuristic method for measurement site selection in sewer systems. Water (Switzerland) 10 (2), 1–16.
Wang, Y., Moe, C.L., Dutta, S., Wadhwa, A., Kanungo, S., Mairinger, W., Zhao, Y., Jiang, Y., Teunis, P.F., 2020. Designing a typhoid environmental surveillance study: A simulation model for optimum sampling site allocation. Epidemics 31, 100391.
Wu, F., Xiao, A., Zhang, J., Moniz, K., Endo, N., Armas, F., Bonneau, R., Brown, M.A., Bushman, M., Chai, P.R., Duvallet, C., Erickson, T.B., Foppe, K., Ghaeli, N., Gu, X., Hanage, W.P., Huang, K.H., Lee, W.L., Matus, M., McElroy, K.A., Nagler, J., Rhode, S.F., Santillana, M., Tucker, J.A., Wuertz, S., Zhao, S., Thompson, J., Alm, E.J., 2020. SARS-CoV-2 titers in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. medRxiv.
Yazdi, J., 2018. Water quality monitoring network design for urban drainage systems, an entropy method. Urban Water J. 15 (3), 227–233.

5.2 Optimal design of water reuse networks in cities through decision support tool development and testing

npj | clean water

www.nature.com/npjcleanwater

ARTICLE OPEN

Check for updates

Optimal design of water reuse networks in cities through decision support tool development and testing

Eusebi Calle[1], David Martínez[1,2], Gianluigi Buttiglieri[2,3], Lluís Corominas[2,3], Miquel Farreras[1], Joan Saló-Grau[1,2], Pere Vilà[1], Josep Pueyo-Ros[2,3] and Joaquim Comas[2,4]✉

Water scarcity and droughts are an increasing issue in many parts of the world. In the context of urban water systems, the transition to circularity may imply wastewater treatment and reuse. Planning and assessment of water reuse projects require decision-makers evaluating the cost and benefits of alternative scenarios. Manual or semi-automatic approaches are still common practice for planning both drinking and reclaimed water distribution networks. This work illustrates a decision support tool that, based on open data sources and graph theory coupled to greedy optimization algorithms, is able to automatically compute the optimal reclaimed water network for a given scenario. The tool provides not only the maximum amount of served reclaimed water per unit of invested cost, but also the length and diameters of the pipes required, the location and size of storage tanks, the population served, and the construction costs, i.e., everything under the same architecture. The usefulness of the tool is illustrated in two different but complementary cities in terms of size, density, and topography. The construction cost of the optimal water reclaimed network for a city of approximately 100,000 inhabitants is estimated to be in the range of €0.17–0.22/m^3 (for a payback period of 30 years).

npj Clean Water (2023)6:23 ; https://doi.org/10.1038/s41545-023-00222-4

INTRODUCTION

Water resources are limited and unequally distributed in space and time. Water scarcity and droughts are an increasing problem in many areas, at least seasonally, in terms of intensity and frequency[1]. Tourism has been recognized as one of the most significant water-consuming sectors on local, regional, and global scales[2,3], as its viability and sustainability depends on adequate water supply quantity and quality[4].

In the context of urban water systems, the transition to circularity and minimizing potable water consumption, requires the redesign of the water infrastructure, including (waste)water treatment and water reuse[5]. Treated wastewater can be used for non-potable purposes, including irrigation, toilet flushing, car washing, cleaning purposes, and industrial uses[6], where appropriate technologies should be carefully selected. EU legislation (EU 2020/741) sets minimum requirements, especially for agricultural water reuse purposes. It does, however, not specifically regulate water reuse in tourist facilities or water reuse for general urban uses, such as toilet flushing. Spain is one of the only five European countries, besides Cyprus, Greece, France, and Italy, which have implemented a national legally binding water reuse regulation (RD 1620/2007). The Spanish water reuse regulation, in fact, is currently the regulation in the EU with the highest number of well-defined water reuse applications, including toilet flushing and garden irrigation. Beyond Europe, other countries worldwide, such as USA, Australia, Singapore, and South Africa, also allow using reclaimed water in cities and specifically for domestic uses. Besides, it is expected that more and more countries will soon consider water reuse as a reliable alternative resource. Several municipalities in Spain (e.g., Sant Cugat del Vallès) have been promoting water reuse in multi-story buildings[7,8]. Nonetheless, applications are still very limited, and related information is largely lacking in the literature.

Efficient and sustainable water reuse requires feasible water reuse projects (i.e., water reclamation treatment plants and distribution to potential uses). Planning and assessing water reuse projects require decision-makers answering a number of questions concerning issues such as: (i) the best tertiary/advanced treatment to be implemented, (ii) the number of uses/users in the city (i.e., how much wastewater needs to be reclaimed), and (iii) how to select the optimal water distribution network. The answer to these questions requires considering different challenges (environmental, economic) and technologies for water reclamation and potential uses of reclaimed water, while evaluating the cost and benefit of all the scenarios by means of different criteria.

In a seminal work[9], a life cycle assessment study was carried out to evaluate the impact of water reuse in the city of Lloret de Mar (Catalonia, NE of Spain), a mass tourism destination on the Mediterranean coast with a high density of high-rise hotels. This study considered four distinct scenarios: non optimized (only potable water consumption), decentralized, hybrid, and centralized. All the water distribution networks were designed manually, assuming the shortest path with lowest terrain elevation. In fact, the multiple factors that need to be considered (terrain elevation, street graph, pipe diameters, terrain usages, etc.) mean that manual or semi-automatic design is still the common practice in planning distribution networks, both for drinking and reclaimed water.

Multiple data, knowledge from different disciplines, and computational capabilities need to be integrated. This is a complex problem where model-based and decision support tools can help by offering a variety of solutions. Previous research on decision support tools for wastewater management has been mainly focused on wastewater treatment[10,11], with a few recent examples dealing with the selection of the most adequate advanced treatment technologies for water reclamation[12,13]. The design of reclaimed water distribution networks has attracted little attention.

[1]Institute of Informatics and Applications, University of Girona, Girona, Spain. [2]Catalan Institute for Water Research (ICRA-CERCA), Emili Grahit 101, 17003 Girona, Spain. [3]University of Girona, Girona, Spain. [4]LEQUIA, Institute of Environment, University of Girona, E-17071 Girona, Spain. ✉email: joaquim.comas@udg.edu

The problem of planning and identifying the most suitable economic schemes for centralized wastewater infrastructure has been partially solved[14]. This solution is based on the potential of a geographical information system to design and locate the water collection pipe network. However, it only considers on-site water reuse, i.e. no water reclamation networks are required. In addition, an improved data-reduced method for wastewater management using globally available data has been proposed[15] (i.e., GIS and statistical data), enabling the approach to be applied worldwide. On the other hand[16], del Teso et al. (2019) aims at energy optimisation in drinking water distribution networks, considering not only operational losses but also structural (or topographic) ones. However, none take into account the initial (or brand-new) design of water reclamation networks. Moreover, more than one tool is sometimes used in a sequential manner, with the corresponding conversion of variables and parameters between the tools, thus becoming highly time-consuming work, especially when aiming at optimal designed networks[17].

In[18], the authors evaluate the life cycle costs and benefits of decentralized greywater reuse planning based on two scales of decentralization: satellite and onsite. However, these two decentralization scales require separating raw greywater from wastewater at the source, which is often not possible in many cities. A centralized water reclamation plant and the corresponding water distribution network is not considered. There is also literature on urban stormwater management. The work of Khurelbaatar et al. (2021)[17] shows an approach that uses the software package MIKE URBAN from DHI (MIKE URBAN, Hørsholm, Danmark) for estimating the potential for managing urban stormwater in already existing urban environments to mitigate the impact of urban stormwater runoff. However, few of their proposed scenarios allow for stormwater reuse.

Distribution water and wastewater network modeling can be approached through graph theory[19], such as[20], for designing water distribution networks based on loops hydraulically balanced method, and wastewater sensor placement approaches for SARS-CoV-2 detection[21], but has not been used yet for the advanced and automated design optimization of water networks.

Given this background, the aim of this paper is to describe a decision support tool for planning water reuse networks in cities. Our approach integrates several algorithms for designing water reuse networks based on graph theory coupled to existent greedy optimization algorithms[22,23]. Our proposal is made up of one single tool, in contrast with the literature, avoiding the need for data exchange and thus resulting in potential savings in time and effort. This tool combines city characteristics (i.e., terrain characteristics, including plot and building usages, elevation, and slope) and water consumption rates to automatically propose an optimal network for water reuse. This paper proposes advanced algorithms to design and optimize large-scale water reuse networks. The usefulness of our solution is also illustrated when testing in real cities. And in this line, two cities of different scales and significantly diverse water uses and requirements have been compared. Construction costs and benefits in terms of water savings are estimated for each scenario. Finally, the optimal water network can be provided when only a limited budget is available.

RESULTS

Overview

The section first describes the innovative decision support tool, called REWATnet, for planning water reuse projects, and highlights the steps involved and the outputs produced. Next, the application of the tool to generate and analyze potential water reuse projects for the cities of Girona and Lloret de Mar is presented. The location of water reclamation treatment plants for reclaimed water production in each city is considered the same as the existing wastewater treatment plants in centralized scenarios, assuming they include the necessary water treatments for the desired reclaimed water quality. The reclaimed water networks obtained by using different algorithms and different scenarios are compared, and the usefulness of the REWATnet tool for network optimisation, when a limited budget is available, is also illustrated and compared with a semi-manual approach.

Description of the REWATnet decision support tool

The new REWATnet decision support tool for planning urban water reuse projects involves the following steps: (i) defining the scenarios; (ii) generating the initial graph; (iii) generating the reclaimed water network; and (iv) estimating key output indicators. The tool provides an innovative mechanism for designing reclaimed water networks in Spain, which can be easily adapted to any country. Figure 1 illustrates a simplified scheme of the water reuse planning tool.

Defining the scenarios. Defining the scenarios to be simulated consists of defining the user inputs, gathering open source data and, only when necessary, customizing default values. The only required user inputs are the target city identifier, the city's cadastral data files, and the location of the wastewater treatment plant. On the one hand, the city identifier is used to obtain the city's street graph through the OpenStreetMap API (see "Methods" section), and must be specified by name, postal code or OpenStreetMap identifier. On the other hand, the same city identifier is also used to gather the city's topography Digital Elevation Model (DEM), which is essential to obtain the city street graph nodes elevation (see "Methods" section). The elevation data can be commonly obtained from the country's official geographic data provider: In Spain this is the Instituto Geográfico Nacional[24]. The DEM is obtained from this source with a 5x5 meter precision.

The city's cadastral data files are used to gather land plot and building data. In Spain, land plot and building data can be obtained from the official cadastre online database. This data is divided into four files that can be freely downloaded: two ".cat" extension files containing the urban and rural plots and buildings, and two ".shp" extension files containing the public gardens in both urban and rural zones. As these files are presented in a hard-to-process format, and for faster and simpler data treatment, said files are converted and combined to a unique standard JSON (JavaScript Object Notation[25]) file using a Python[26] custom implemented script. For other countries, it would be necessary to process the corresponding authority land plot data and generate the standard JSON file. In the case of several municipalities connected to the same reclamation water treatment plant, the files from every municipality must be downloaded, converted to JSON with the script, and merged to form a unique file. Additionally, the OverPass API[27] and Shapely[28] library allow for the location and surface of public gardens to be obtained and appended since they are not included in cadastre files.

Optional user inputs are the definition of a subset of the considered uses for reclaimed water, a threshold for minimal water consumption and specific city areas particularly suitable for water reuse. The subset of considered uses for reclaimed water will define the network destinations. The threshold for water consumption may specify whether a minimum amount of reclaimed water must be served in each network destination. The selection of specific city areas especially suitable for water reuse may be relevant to target new city settlements (i.e., neighborhoods or sectors) and for statistical and future work purposes. The OverPass API is used to extract specific city areas (downloading the geographic polygons that define the divisions) together with the Shapely library (checking to which polygon a certain node belongs). By default, all water use categories, a threshold for water consumption of 0 m^3/d for all network

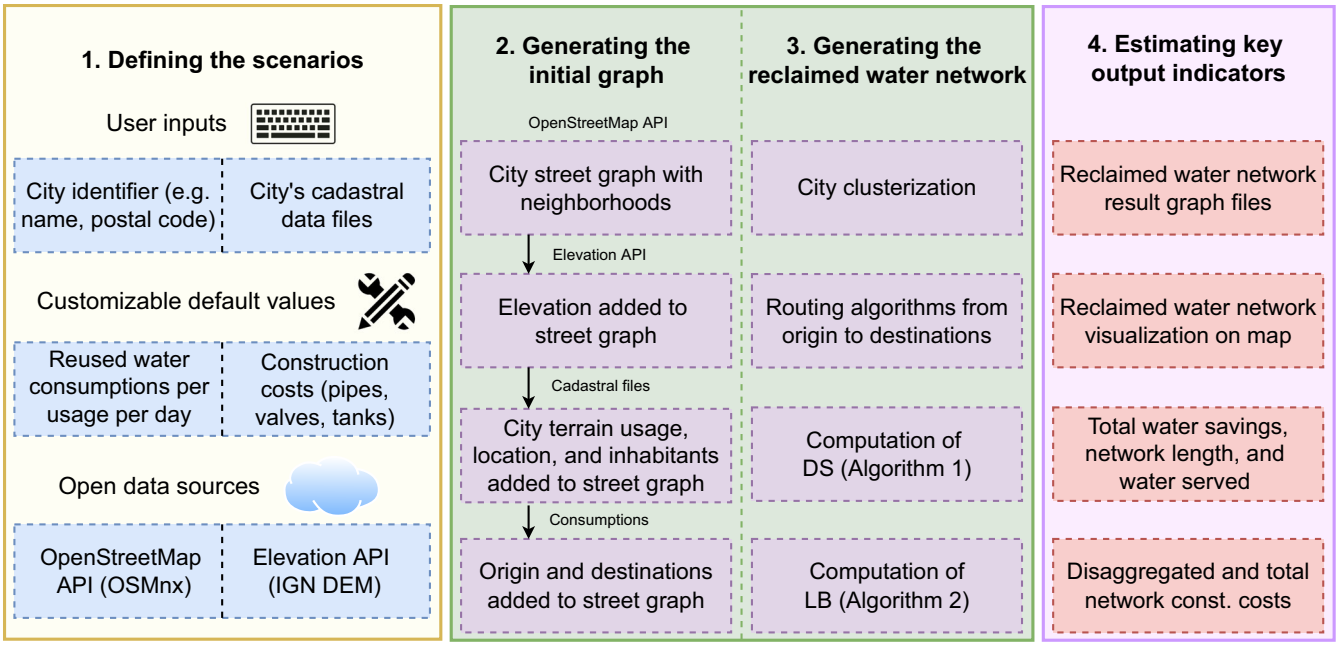


Fig. 1 REWATnet decision support tool. REWATnet simplified scheme. The data flow is as follows: (i) defining the scenarios, including user inputs, customizable values, and open data sources; (ii) generating the initial graph, where an initial city street graph is generated from the OpenStreetMap API, the elevations are added to each node from the Elevation API, and the cadastral file is processed where consumptions are added to nodes; (iii) generating the reclaimed water network, which includes the application of the algorithms such as city clusterization, routing algorithms, pipe diameter selection (DS) and limited budget (LB), to the initial graph; and (iv) estimating key output indicators, where the result graph files with the reclaimed water network, visualizations on map, water saving, network length, water served, and disaggregated and total network construction costs are obtained.

destinations (see Methods section) and the whole city area are considered for water reuse projects.

Generating the initial graph and the reclaimed water network. Once the scenario is defined and all data has been gathered, the REWATnet decision support tool is able to generate the reclaimed water network. First, an initial graph is generated and then the different optimization algorithms are applied (see Methods section).

The initial graph is generated by adding the data to the city street graph, including the node elevation, street slopes, city terrain usage, location, and inhabitants, as well as the origin node and the destination nodes. Then, the routing algorithms are computed from the initial graph to obtain the optimal reclaimed water network based on two approaches: (i) serving water to the destinations defined in the scenarios whatever the budget might be, or (ii) maximizing the water served with a limited budget. In the first approach, city clustering algorithms may be first applied to the initial graph depending on the targeted city size. Then, the routing algorithms are computed for the main network (i.e., from the initial node (water tank next to the water treatment plant) to the cluster water tanks distributed along the city) and the branched network areas (i.e., from each cluster water tank to all its destinations). The routing algorithms generate the base for the reclaimed water network, although the network construction costs remain unknown as the pipe diameters have not yet been computed. Then, the computation of the pipe diameter selection (DS) algorithm (Algorithm 1) generates all the pipe diameters of the reclaimed water network and enables the construction costs to be calculated (see Supplementary Table 1, Supplementary Table 2, and Supplementary Table 3).

In the second approach, the Limited Budget (LB) availability algorithm (Algorithm 2) is computed considering a single branched network area. The LB algorithm uses the routing algorithms and the DS algorithm to build a reclaimed water network that maximizes the water volume served for a specific budget (see Methods section).

Estimating key output indicators. The REWATnet decision support tool estimates the following key output indicators:

- Graph files of the optimal reclaimed water network: the optimal reclaimed water network is provided in a file with a standard graphml format for further analysis.
- Visualization on a map of the optimal reclaimed water network: a clear visual representation of the reclaimed water network drawn on a plane and over a map, provided in PDF vector image files.
- Network length, pipe diameters, population served, and total water savings: this relevant data related to each reclaimed water network is provided in a simple text file. The total water savings is assumed as the total reclaimed water consumption, as this amount of water will be subtracted from the drinking water distribution network.
- Disaggregated and total network construction costs: the construction costs of the reclaimed water network are provided disaggregated by main network costs, branched network costs, and water tank costs expressed in thousands of euros (€K), provided in another simple text file.

Before the execution and analysis of the different scenarios, a preliminary validation of the REWATnet tool was carried out for the case study of Girona. The reclaimed water consumption of potential users estimated by the tool was compared to actual water consumption data provided by the water authorities (see Table 1). The estimated model consumption was calculated based on the methods and references of Supplementary Table 4 in Supplementary Materials, together with the information extracted from the land plots (see "Methods" section). The validation shows a remarkably accurate model consumption with a minor overall error of 6.4%. According to the water use categories, the error

Table 1. Water use consumption validation (Girona).

Water use categories	Model consumption (m^3/year)	Actual consumption (m^3/year)	Error
Economic operations (Domestic with operations + Large industrial consumption)	530,564	525,588	2.8%
Public uses (Urban equipment + Public garden irrigation)	251,405	203,064	23.8%
Domestic (Housing + Private garden irrigation)	1,690,877	1,594,735	6%
All validated uses* (Economic operations + Public uses + Domestic)	2,472,846	2,323,387	6.4%

*All validated uses do not include the vegetable garden irrigation model consumption (estimated as 269,005 m^3/year) as there is no actual consumption data to be compared.

between actual and estimated consumption is especially low for both economic operations and domestic use categories, as shown in Table 1. In the case of the public uses, which suppose the lowest water consumption, the model shows an error of 23.8%. This error barely affects the overall error of 6.4% as it represents only 10% of the total consumption. The model estimates a higher public uses consumption as the toilet flushing and irrigation consumption data in sports centers is complicated to estimate, and the land plot data may be incomplete or outdated. Finally, it is worth noting that the consumption for vegetable garden irrigation has not been included since there is no real data to be compared with. It should also be noted that when validation results are not satisfactory (which is not the case of the present work), the tool allows for the fine-tuning of some parameter values of the estimated consumption rates (Table 1).

Comparison of different routing algorithms

Among the Kou, Takahashi, and Mehlhorn routing algorithms, the Mehlhorn algorithm was found to be the most suitable in the literature due to its lower computational complexity (see "Methods" section). Nonetheless, all routing algorithms were tested with our case study cities to evaluate their accuracy, where lower reclaimed network length provides better accuracy. Thus, the planning tool has been first tested for designing the optimal reclaimed water network for the whole urban area of Girona and Lloret de Mar cities applying the Kou, Takahashi, and Mehlhorn algorithms. In both cities, reclaimed water is produced in a centralized water reclamation plant and stored in an initial water tank placed alongside it. The whole city of Lloret de Mar is considered as a unique cluster (i.e., one branched network area), while city clustering techniques are applied in the case of Girona to determine the optimal placement of water tanks. In fact, due to the city's size, intermediate water tanks are needed along the reclaimed water network for the case of Girona. In both case studies, only public water uses are considered. Besides, we contemplate water reuse destinations within 300 m between any land plot centroid (i.e., the geographical center of the physical entity that uses reclaimed water) and the nearest node of the city's initial street graph. This consideration is necessary to ignore water destinations obtained from land plots' official data that are too far from the street graph (e.g., a farm outside the city), as the reclaimed water network is built based on the streets. Given this scenario, the initial graph of Girona considers 328 destinations with total water consumption of 2129 m^3/d, while the initial graph of Lloret de Mar considers 144 destinations with total water consumption of 1182 m^3/d.

Table 2 shows the comparison between the REWATnet output indicators for the different routing algorithms and the two cities. The first thing to notice is that, for both Girona and Lloret de Mar, the Kou and Mehlhorn algorithms present exactly the same accuracy (the same reclaimed network pipe length), while the Mehlhorn algorithm provides a significantly lower computation time (about 100 times faster). Although the Takahashi algorithm present the best accuracy (lower reclaimed network pipe length),

Table 2. Comparison of the decision support tool output indicators for the different routing algorithms.

City	Algorithm	Network length (m)	Execution time (s)
Girona	Takahashi	67,149	28,123 (≈7.8 h)
	Kou	68,068	38.06
	Mehlhorn	68,068	0.24
Lloret de Mar	Takahashi	49,498	1442 (≈24 m)
	Kou	50,413	8.39
	Mehlhorn	50,413	0.08

compared to the Kou and Mehlhorn algorithms, its execution time becomes intractable on cities with a large set of destinations (about 18,025 and 117,179 times slower than the Mehlhorn algorithm).

Comparison of reclaimed water networks

For both cities and using the best routing algorithm, the reclaimed water network was computed considering two scenarios serving water to all the destinations: (i) only for public water uses and (ii) for both public and private water uses. This section presents some of the most relevant key output indicators.

Regarding Lloret de Mar, Fig. 2 illustrates the graphs representing the reclaimed water networks generated for the two scenarios. The first scenario results in a 44 km network, a total construction cost of €3628K, a pipe diameter average of 64 mm, and total consumption of 283 m^3/d to serve 4.1% of the total demanded water (Water served/total demand × 100). The second scenario results in a 104 km network, a total construction cost of €9429K, a pipe diameter average of 82 mm, and total consumption of 6844 m^3/d to serve the entire demand for water. It is worth noting that the first scenario presents a construction cost per cubic meter of €12.82K/m^3/d compared with the €1.38K/m^3/d of the second scenario, which is more than nine times bigger. This difference might be related to the fact that private water use needs connections that are already (partially) included in the first scenario.

Regarding Girona, Table 3 contains the disaggregated and total water reclaimed network construction costs obtained for different clustering solutions (from one to eight clusters) for the scenario (ii) of public and private water uses. The observed total water consumption is 7142 m^3/d, leading to an average cost increase for each cluster of 4.2%. It is worth noting the low increase in the costs of the pipe network, as the branched network costs from two to eight clusters are analogous. Hence, the most significant elements that increase the cost among the different clustering solutions is the main network and the number of water tanks. Note that, the whole infrastructure payback period considered is 30 years, and the accumulative water savings should also be considered (up to 78,204,900 m^3 of total consumption). The resulting reclaimed water network of five clusters is shown in Fig. 3, with 155 km of pipes and a total construction cost of €15,996K. Interestingly, in this case, the

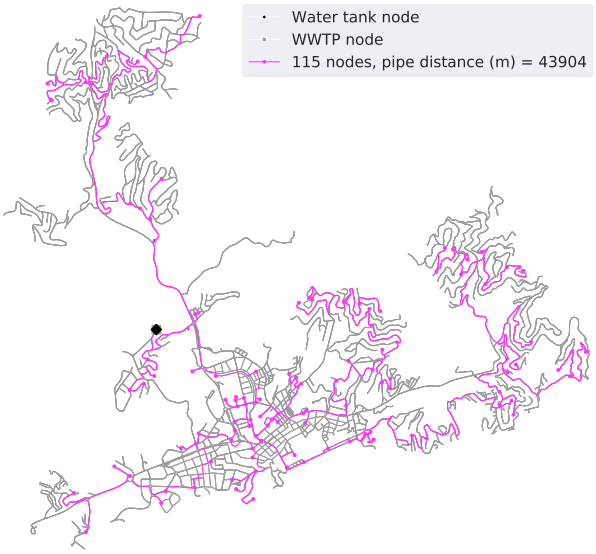


Fig. 2 Reclaimed water network visualization (Lloret de Mar). The diagram on the left shows the resulting reclaimed water network for only public cases, while the one on the right shows the resulting reclaimed water network for both public and private uses. The city street graph is represented in gray paths and the reclaimed water network in magenta paths.

Table 3. Reclaimed water network construction costs for different clustering solutions, both public and private uses (Girona).

Number of clusters	Main network costs (€K)	Branched network costs (€K)	Water tank costs (€K)	Total costs (€K)	Cost (€K/m^3/d)
1	0	12,361	760	13,121	1.84
2	1177	11,278	1200	13,655	1.91
3	1421	11,213	1460	14,094	1.97
4	1715	11,300	1610	14,625	2.05
5	2533	11,502	1960	15,996	2.24
6	2812	11,198	2310	16,320	2.29
7	3249	11,254	2550	17,053	2.39
8	3305	11,291	2870	17,466	2.45

placement positions for the water tanks computed by the tool match well with the ones that are actually in the Girona drinking water distribution network[29]. In Supplementary Materials, the clusterization results of Girona with three (Supplementary Fig. 1) and seven (Supplementary Fig. 2) clusters are also illustrated.

Comparison of optimal network with current practice

The optimal network with limited budget availability is illustrated in the case study of Girona. The Limited Budget (LB) availability algorithm (Algorithm 2) results in the maximum amount of reclaimed water served given a maximum budget *B* (see Methods section). The LB algorithm is executed considering eight different budgets from €500K to €2000K, with intervals of €250K over a randomly selected branched network area, obtained from the previously generated five-cluster reclaimed water network (see Fig. 3). Table 4 shows these results for the case study of Girona where, considering the blue branched network area water tank as the origin (from the five cluster solution, see Fig. 3), each limited budget *B*, pipe network length, reclaimed water served, percentage of water served over total demand, and execution time output indicators are presented. The results show a linear evolution of the percentage of water served over total demand *C* as a function of budget *B*. Supplementary Figs. 3, 4, 5, 6, 7, 8, and 9 illustrate the networks generated for each budget (computation performed) in Table 4.

The LB algorithm optimization is compared with the so-named "current practice", which considers a manual approach based on actual reclaimed water network planning experience. In the current practice, unlike the LB algorithm, the best profit *P* (water served per cost ratio) is considered based purely on the lowest distance without considering water served (i.e., for each iteration, the closest destination is added to the current reclaimed water network until budget *B* is reached). The current practice is applied with the same budgets considered in the testing of the LB algorithm, showing a linear evolution with a considerably lower slope of the percentage of water served over total demand *C* as a function of the budget. Both linear functions are illustrated in Fig. 4, where the LB algorithm results in a function $C = 17.40B - 5.07$ and the current practice in a function $C = 6.16B - 0.92$. As can be seen in Fig. 4, the benefits of the optimal network approach are more evident as the budget increases, since the slope of the LB algorithm is almost three times (2.82) that of the current practice.

DISCUSSION

In the context of the global climate change, where water scarcity regions are increasing and water reuse is becoming paramount, an optimal return (e.g. served water per unit cost invested) from each water reclamation project is sought. Moreover, tourism is recognized as a major water consuming sector, and the growth in tourism establishments has been matched by a growth in water demand[3]. As such, the new decision support tool presented has been applied to two case studies in Spain, but it can be easily adapted and applied to any region world wide. This would only require configuring the proper open online services and open data sources (e.g. cadaster) and a potential customization of default values for water consumption and costs. According to our knowledge, there are no similar tools in the literature able to plan and economically asses an optimal water reclamation network for a city with little computing effort. Besides, the tool may be used for usual water distribution networks, although it should be validated and revised.

Urban planners from municipal or regional authorities, consulting companies and/or water utilities can use the tool for planning urban water reuse projects, i.e., to identify a city's most critical water consumption hotspots, compare different solutions by using technical and economic criteria and then select the optimal alternative. Another application for this innovative tool could be to assess which water reuse scenario - centralized, semi-decentralized, or decentralized - offers a better cost-efficient water reuse scheme, optimizing the number and location of decentralized treatment plants. Another question worth examining with the tool would be determining the minimum number of inhabitants to be decentralized for a water reuse solution to be sustainable for a specific city/neighborhood. REWATnet can also be used as a dissemination and training tool for planning water reuse schemes.

This decentralized water treatment and reuse can be extremely relevant for touristic cities, easing the pressure on scarce water resources and/or significantly reducing wastewater generation. For example, in Lloret de Mar, with 40,000 inhabitants in winter and up to 200,000 in summer, the portion of generated wastewater coming from tourist facilities was estimated as more than 10,000 m^3/d and at least half of this amount was gray water, which can be more easily reclaimed than wastewater[30].

A validation of the reclaimed network design (pipe length and diameters) can be carried out by coupling the graph files generated by our tool and EPANET[31]. Besides, the current

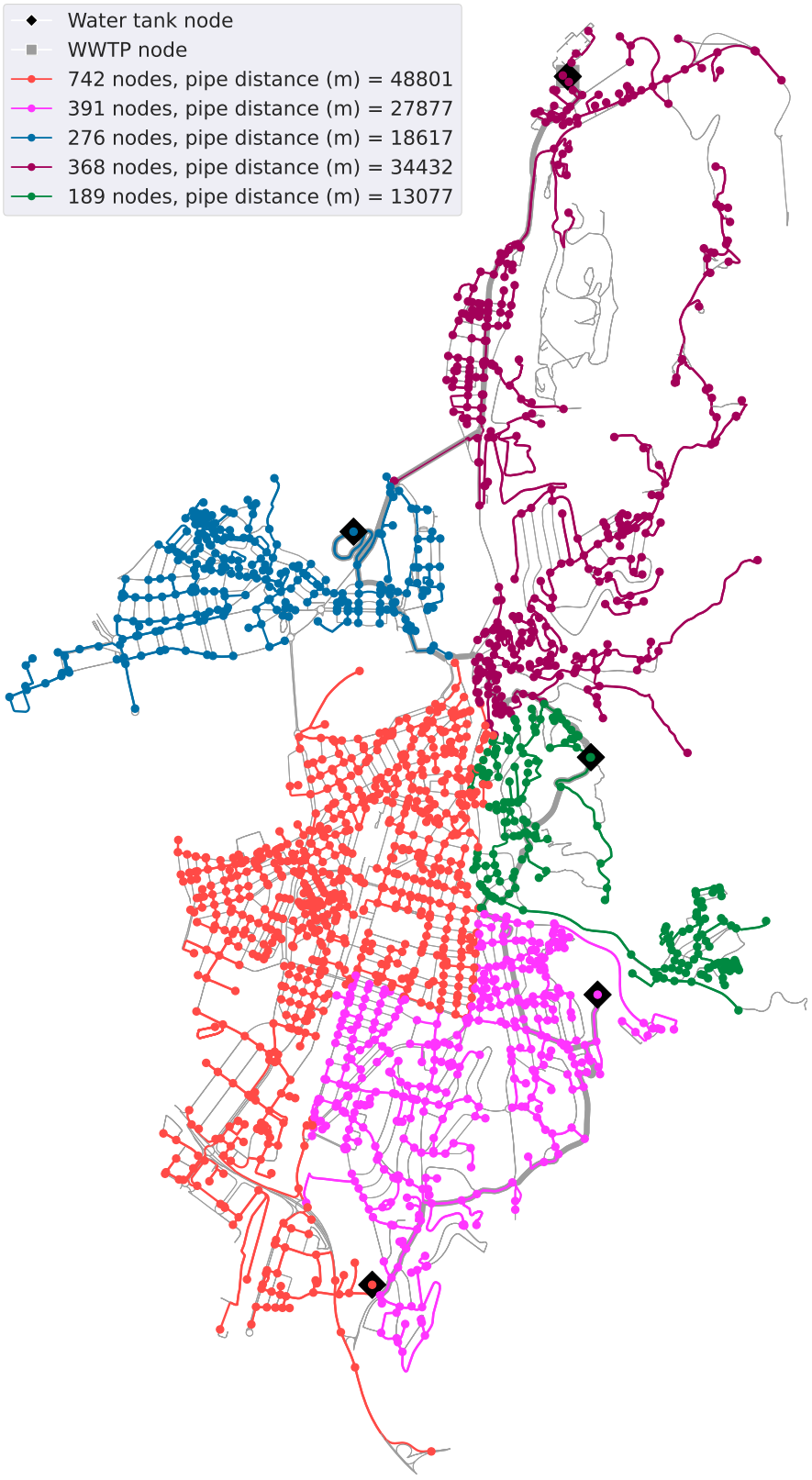


Fig. 3 Reclaimed water network visualization (Girona). The reclaimed water network for the case study of Girona is illustrated with five clusters and for both public and private uses. The city street graph is represented in gray paths and the generated branched networks in different colors (for each cluster). Black rhombuses indicate the location of water tanks (origin point of a branched network). The main network is represented in wider gray paths.

clustering algorithm provides a simplified approach based on cost analysis. Further developments of the tool will include meshed network designs, which are common designs to better deal with potential pipe failures, increasing the scope of our tool. Besides, construction costs for the advanced treatments for water reclamation, as well as the operational and maintenance costs for both the water reclamation treatment plant and network, will also be estimated. Moreover, the use of reclaimed water is not accepted in many countries or only accepted for specific purposes according to their regulations. Therefore, an adaptation of our tool based on the country may be considered.

The innovative REWATnet decision support tool to help plan optimal networks for reclaimed water reuse in cities has been developed and tested. With little input data from the users themselves, and by using open data, the tool is able to compute the maximum amount of reclaimed water served per unit of cost invested, including the length and pipe diameters of the network, the location of storage tanks as well as the population served and construction costs. In other words, everything under the same architecture. A comparison of the estimated water consumption for industry/commerce, public and private uses to actual water consumption data gives an overall error of 6.4%. Gaining private user trust in reclaimed water is a key factor for sustainable water reuse networks since these users have the highest consumption rates, 60–70% share of the total in the tested cities (with the differences being due to intense tourism activities). The optimal network graph is computed by using the Mehlhorn routing and clustering algorithms and, when needed, the Limited Budget (LB) availability algorithm. The construction cost of an optimal water reclaimed network for a city of approximately 100,000 inhabitants is estimated to be in the range of €0.17-0.22/m^3 (for a payback period of 30 years), thus demonstrating a reasonable cost compared to actual costs of drinking water networks. For the same city, the automatic tool computes (in less than 10 minutes) an optimal network able to serve reclaimed water up to three times more than the water served using the current (manual) planning practice. Finally, the tool also provides a map of a user-friendly visualization of the optimal reclaimed water network, including the main and branched networks and colored city clusters, when needed.

Table 4. LB algorithm results over a randomly selected branched network area (Girona five clusters, blue cluster water tank in Fig. 3, public and private uses).

Budget (€K)	Pipe network length (m)	Transported water (m^3)	Water served / total demand (%)	Execution time (s)
500	3420	337	4.7	57
750	5191	529	7.4	59
1000	8464	777	10.9	159
1250	11,412	1228	17.2	273
1500	14,680	1477	20.7	384
1750	17,658	1906	26.7	392
2000	20,426	2086	29.2	493

METHODS

Overview

This section first presents the data collection, distinguishing the different data sources; on the one hand, the open data sources to automatically obtain city characteristics and, on the other hand, the consumption and costs-related databases. The combination of multiple data sources is indeed one of the key features of our proposal. Second, the definition of the potential water reuse scenarios is introduced in order to target the desired water destinations in cities based on their water usage. Third, the routing algorithms, based on graph theory and optimization, that support the generation of the reclaimed water network, as well as the algorithms for city clustering (for water tank allocation), pipe diameter selection, and limited budget availability are presented. Finally, the case studies (i.e., cities) used for the testing of the decision support tool are presented.

City characteristics: open data sources

The city characteristics are collected automatically from open data sources. In particular, we obtain and link together: (i) the city street graph; (ii) the city land plots and building data; and (iii) the city topography.

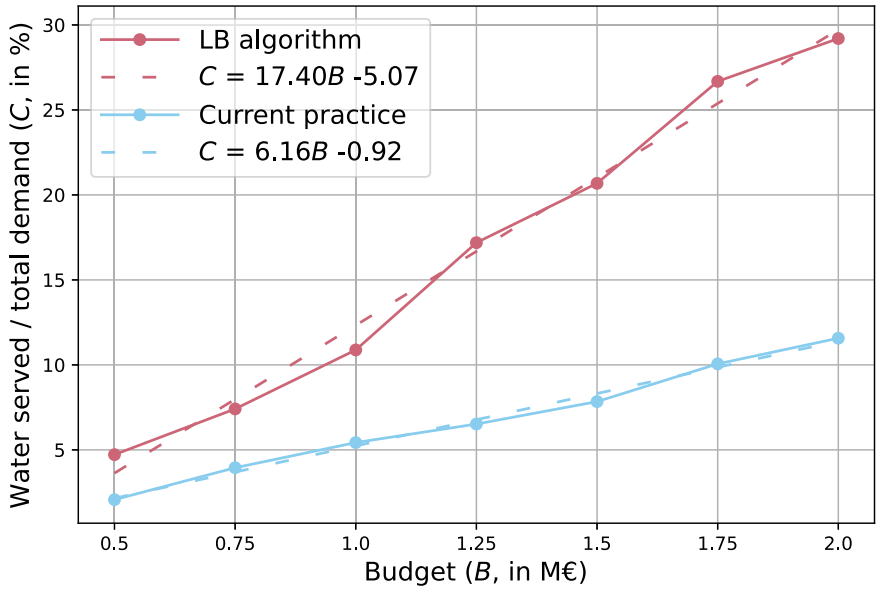


Fig. 4 LB algorithm water served per budget results, compared with current practice. Percentage of water served over total demand comparison between the LB algorithm and current practice (manual approach). The linear regression functions for both the LB algorithm and current practice are defined in the figure legend and represented as dotted lines.

City street graph. The entire city street graph is obtained from the OpenStreetMap[32] API (a software intermediary that allows two applications to communicate), using the OSMnx library[33]. This API provides a street graph containing the city streets (edges) and intersection points or direction changes on street turns (nodes). It is assumed that the pipes for the new reclaimed water network will be installed following the streets of the city. This information layer is the basis for the data aggregation related to the city (i.e., land plots and building data, water uses, inhabitants, and consumption), and for later running the algorithms.

Land plots and building data. The land plots and building data of a city are required to extract all the possible reclaimed water destinations and consumption demands. Each land plot geographical location is linked and clusterized to the nearest points of the city street graph, where the pipes of the reclaimed water network will be placed. The land plot data provides the surface and terrain use (e.g., household, hotels, gardens, or sports facilities), while the building data indicates whether the plot is occupied by a single household or a building of several floors and apartments. The building data is needed to estimate the number of inhabitants per plot, which is required to estimate daily water consumption. The water consumers living in a household are estimated as the ratio between the number of inhabitants in a given city and the number of households of that city, which can usually be obtained from the national statistics institutes.

City topography. The city topography is required to know the elevation of the plots, which is essential to calculate the optimal route for the reclaimed water network and place water tanks properly, and assure adequate pipe diameter and minimal operation and maintenance costs due to pumping.

Consumption and construction costs database

A relational database including the data required to estimate water consumption for different water reuse purposes and network construction investment costs was developed. Default values are included in the database of the decision support tool but all parameters, based on the appearance of new or more precise information available, can be customized to user needs. The database for the estimation of the reclaimed water consumption for different uses is based on bibliographic and practitioner information (see Supplementary Table 4)[34–42], and can also be extended with additional water uses providing water consumption and destination locations. The database for the reclaimed water network construction costs has been obtained from a tool for life cycle analysis of sewer systems construction[43], which is based on a standard database frequently used by practitioners[44] (see Supplementary Table 1, Supplementary Table 2, and Supplementary Table 3).

Scenario definition

The definition of potential water reuse scenarios involves: (i) the selection of the reclaimed water origin and destinations (among all potential water reuse purposes, see Supplementary Table 4); and (ii) the identification of the city area that will be considered, i.e., the whole city area or only some parts of the city with special interest for water reuse (e.g., new developments) or complying with optional constraints such as a minimum reclaimed water flow rate or population served. The origin of reclaimed water is the centralized wastewater treatment plant, which incorporates a tertiary or advanced treatment for enhancing effluent water quality. The possible uses for reclaimed water would normally be defined by the end user, whereas the areas of application can be either automatically identified by the algorithms or by the user.

Reclaimed water network generation

In brief, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the reclaimed water network graph, with a *V*-element set of nodes $\mathcal{V}$ representing the set of destination (water consumption) nodes, the water source node, and junction points, and an *E*-element set of links $\mathcal{E} \subset \mathcal{V}^2$ representing pipes. Additionally, r (where $r \in \mathcal{V}$) denotes the source node (e.g., the reclaimed water treatment plant or an initial water tank), and $\mathcal{C}$ (where $\mathcal{C} \subseteq \mathcal{V}$) denotes a *C*-element set of consumption nodes. First, we introduce the routing algorithms, and then the optimization algorithms for city clustering (branched network areas definition and water tank allocation), pipe diameter selection, and limited budget availability are presented.

Routing algorithms. The routing algorithms to generate and analyze the reclaimed water networks are based on techniques borrowed from graph theory[19]. In the case of water distribution networks, the pipes correspond to the graph edges and the junctions represent the graph nodes from the city street graph. Hence, these networks follow the existing street paths. Using that representation, we generate a network, solving the problem of covering the paths from the centralised water reclamation plant to all the required destinations with minimum costs, using graph routing techniques such as Steiner Tree algorithm variations.

The Steiner tree problem in graphs is well known to be computationally intractable since it is an NP-hard problem[45]. A preliminary performance and complexity study has been carried out on improved Steiner tree optimization and greedy algorithms to select the proper routing algorithm to use. In particular, Table 5 shows the complexity of the Kou[46], Takahashi[47], and Mehlhorn[23] algorithms, and where it can be seen that the Mehlhorn algorithm provides the better complexity.

The execution of the routing algorithms over a city street graph results in a reclaimed water network graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E}$ contains the edges of the city street graph that are the most suitable to build the water distribution network (i.e., the pipe route that minimizes the network length). However, in this stage, the pipes defined by $\mathcal{E}$ do not yet contain their diameters.

Since some land plots potentially using reclaimed water may be too far from the nearest node or connection point on the street graph, the routing algorithms omit all plots that are farther away than a certain distance from the land plot centroid, which is given by a threshold in meters customizable by the user.

City clustering algorithms. In the case of small cities, a water tank placed alongside the water reclamation plant can be enough to supply all the destination nodes on the distribution network, acting as a unique branched network. Because of large distances, this is not possible in medium to large cities, where a clusterization of the city street network is necessary for scalability reasons. The allocation of additional water tanks along large water reuse networks is needed for practical reasons such as topography issues (i.e., issues that disable gravity distribution), pressure losses, and leak localization[48,49]. Thus, a clustering approach to build the reclaimed water network in medium to large cities is presented.

Table 5. Technical comparison of Steiner tree algorithms.

Algorithm	Complexity
Kou	$O(\lvert\mathcal{V}\rvert \times \lvert\mathcal{E}\rvert \times \log \lvert\mathcal{V}\rvert)$
Takahashi	$O\left(\lvert\mathcal{S}\rvert \times \lvert\mathcal{V}\rvert^2\right)$
Mehlhorn	$O(\lvert\mathcal{V}\rvert \times \log \lvert\mathcal{V}\rvert + \lvert\mathcal{E}\rvert)$

Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, search the Steiner tree with terminals $\mathcal{S}$.

Although several graph clustering optimization algorithms exist, a medium to large city scenario requires an efficient algorithm to provide a feasible solution in a reasonable amount of time. In Blondel et al. (2008)[22], the authors propose the so-called Louvain algorithm, a heuristic method based on modularity optimization that outperforms all the other known clustering methods in terms of computation time. Their results show a significant reduction of the network computation time compared to the well-known algorithms of Clauset, Newman and Moore[50], of Pons and Latapy[51], and of Wakita and Tsurumi[52].

Thus, our proposal is to apply the Louvain heuristic algorithm to generate city clusters based on node-pairs proximity, each city cluster representing a branched network area. First, it is necessary to place an initial water tank alongside the distribution network source node (i.e., the water reclamation plant). Next, for each cluster, a simple algorithm optimizes the placement of a water tank. This algorithm selects as candidates the subset of the cluster nodes equal to or higher (in elevation) than the highest destination node. From these candidates, the algorithm finds the node that minimizes the cluster minimum Steiner tree (i.e., that minimizes the branched network area). With this method, we assume that, for each cluster, the water will reach all the destination nodes by gravity. Exceptionally, in order to save costs, the initial water tank also behaves as its cluster water tank. Once the clusters are defined and the water tanks allocated, the main network is constructed based on the minimum Steiner tree between the initial water tank and the other cluster water tanks.

Pipe diameter selection algorithm. Once the network graph $\mathcal{G}$ is generated by applying a routing algorithm, it is necessary to select the appropriate construction pipe diameters for each edge from a limited set of available pipe diameters based on the reclaimed water demand of the destination nodes. The diameter selection (DS) algorithm (Algorithm 1) selects the proper pipe diameter for each edge of the reuse water network $\mathcal{G}$. First, the algorithm obtains the expected daily reclaimed water flow volume w (in m^3/s) of each edge $e \in \mathcal{E}$ based on the consumption of the destination nodes $c \in \mathcal{C}$ where the edge e is present in the route $\mathcal{E}(r,c)$, $r, r \in \mathcal{V}$ being the water distribution source node. Then, the minimum required diameter $d(e)$ is computed from the edges expected reclaimed water flow $w(e)$ and the desired flow speed s using the Eq. (1). The flow speed s is set to 1 m/s by default, extracted from Simpson and Elhay (2008)[53], who proposed pipe velocity range of 0.5 to 1.5 m/s. Finally, based on the user-specified set of available pipe diameters $\mathcal{D}$, the algorithm selects for each edge e the next greater value $d'(e)$ from the computed minimum required diameter $d(e)$. Table 6 specifies the full notation used for the diameter selection (DS) algorithm.

$$d(e) := \sqrt{\frac{w(e)}{s \times \pi}} \times 2 \tag{1}$$

Algorithm 1. Diameter selection (DS) algorithm.
Step 1: Initialize the node r and sets $\mathcal{C}$; $\mathcal{D}$; m; $\mathcal{E}(r,c), c \in C$; $\mathcal{X} := \varnothing$; $\mathcal{Y} := \mathcal{E}$.
Step 2: Choose at random an edge with unassigned water flow, i.e., an edge $e \in \mathcal{Y}$, set $w(e) := 0$, and update sets $\mathcal{X}; \mathcal{Y}$.
Step 3: For each water distribution consumption node $c \in C$:
(a) if $e \in \mathcal{E}(r,c)$, then set $w(e) := w(e) + w(c)$.
Step 4: If $w(e) > 0$, then:
(a) compute $d(e) := \sqrt{\frac{w(e)}{s \times \pi}} \times 2$, and set $d'(e) := \max(\mathcal{D})$.
(b) for each available pipe diameter $p \in \mathcal{D}$:
(i) if $p >= d(e)$ and $p < d'(e)$, then set $d'(e) := p$.
Step 5: If $\mathcal{Y} \neq \varnothing$, then go to Step 2.
Step 6: If $\mathcal{Y} = \varnothing$, then stop ($d'(e)$ contains the assigned pipe diameter $\forall e \in \mathcal{E}$).

Table 6. Full notation concerning the algorithms.

$r, r \in \mathcal{V}$	reclaimed water source node
$\mathcal{C}$	set of water distribution consumption nodes; $\mathcal{C} \subseteq \mathcal{V}$
$\mathcal{D}$	set of available pipe diameters (each one in mm)
s	float constant indicating the desired water flow speed (in m/s, 1 by default)
$m(c), c \in \mathcal{C}$	integer indicating the consumption of destination node c (volume, in m^3)
$\mathcal{E}(a,c), c \in \mathcal{C}$	set of edges forming the shortest path from the source node a to the destination node c; $\mathcal{E}(a,c) \subseteq \mathcal{E}$
$l(e), e \in \mathcal{E}$	float indicating the length the edge e (in m)
$w(e), e \in \mathcal{E}$	float indicating the water flow of the edge e (in m^3/s)
$\mathcal{X}$	set of edges with assigned water flows; $\mathcal{X} := \{e : e \in f\}$
$\mathcal{Y}$	set of edges with unassigned water flows; $\mathcal{Y} := \mathcal{E} \setminus \mathcal{X}$
$d(e), e \in \mathcal{E}$	integer indicating the minimum required diameter of the edge e (in mm)
$d'(e), e \in \mathcal{E}$	integer indicating the assigned diameter of the edge e; $d'(e) \in \mathcal{D}$ (in mm)

Limited budget availability algorithm. The limited budget availability (LB) algorithm (Algorithm 2) uses the routing algorithms and the DS algorithm (Algorithm 1) to build a reclaimed water network that maximizes the water volume served for a specific budget B. The LB algorithm follows a greedy approach that is an adaption of the algorithm provided in[54], which presents fast heuristics for the Steiner tree problem with revenues, budget, and hop constraints. The main idea of the algorithm is to iteratively build a reclaimed water network while its cost does not exceed the provided budget. It starts from an initial graph $\mathcal{T}$ with only the reclaimed water source node r. For each iteration, and while the construction costs are below the budget, the algorithm adds to $\mathcal{T}$ the destination node c ($c \notin \mathcal{T}$) that provides the best profit P (water served per cost ratio). The profit P is obtained by dividing the node c cubed daily water consumption by the extra construction cost of adding c to the graph $\mathcal{T}$.

Algorithm 2. Limited Budget availability (LB) algorithm.
Step 1: Initialize the node r, the budget B, and sets $\mathcal{C}$; $\mathcal{D}$; m.
Step 2: Let $\mathcal{T}$ be the initial graph with $\mathcal{V}' := \{r\}$ and $\mathcal{E}' := \varnothing$.
Step 3: Set the profit $P := 0$, the iteration candidate node $n := \varnothing$, and its current network's closest node $o := \varnothing$.
Step 4: For each reclaimed water consumption node $c : c \in \mathcal{C}, c \notin \mathcal{V}'$:
(a) Get the node $a \in \mathcal{V}'$ that minimizes the path to join $\mathcal{T}$ with c, such that:
$\sum l(e), e \in \mathcal{E}(a,c) := \min((\sum l(e), e \in \mathcal{E}(v',c)), v' \in \mathcal{V}')$
(b) Copy the graph $\mathcal{T}$ to $\mathcal{U}$, such that $(\mathcal{V}'', \mathcal{E}'') := (\mathcal{V}', \mathcal{E}')$.
(c) Add the (a,c) path to graph $\mathcal{U}$, such that $\mathcal{V}'' := \mathcal{V}'' \bigcup \{a\}$, and $\mathcal{E}'' := \mathcal{E}'' \bigcup \mathcal{E}(a,c)$.
(d) Compute Algorithm 1 (DS) with $\mathcal{U}$ and $\mathcal{D}$, to obtain the pipe diameters $d'(e), e \in \mathcal{E}''$.
(e) Calculate the pipe network construction cost Z of $\mathcal{U}$ (including the initial water tank) from $d'(e)$ and $l(e)$, $e \in \mathcal{E}''$ (see Supplementary Table 1 and Supplementary Table 3).
(f) If $Z <= B$, then:
(i) Compute the profit P' of adding a to $\mathcal{T}$, such that $P' := \frac{m(a)^3}{L}$, where $L := \sum l(e), e \in \mathcal{E}(a,c)$.
(ii) If $P' > P$, then set $P := P'$, $n := a$, and $o := c$.
Step 5: If $P > 0$, then:
(a) Add the (n,o) path to graph $\mathcal{T}$, such that $\mathcal{V}' := \mathcal{V}' \bigcup \{o\}$, and $\mathcal{E}' := \mathcal{E}' \bigcup \mathcal{E}(n,o)$.
(b) Go to Step 3.
Step 6: $\mathcal{T}$ represents the final reclaimed network graph $\mathcal{G}$.

Case studies

The usefulness of the decision support tool presented here is illustrated in the cities of Girona and Lloret de Mar, both in Catalonia (North-East of the Iberian Peninsula), two different but complementary cities in terms of size, density and topography. Girona with its 103,369 inhabitants and 47,446 households (2.4 citizen per household), is a typical Western Mediterranean city; compact, with mixed uses and clearly divided between the old town and the modern peripheral. Its urban area extends 12.7 km^2 on a rivers' crossing, has a population density of 8139 hab/km^2, an average slope of 5.1 and an altitude range (difference between minimum and maximum altitudes) of 177 m. Lloret de Mar is a city located on the northeastern Mediterranean coast of Spain. The city has a year-round population of 39,089 and a seasonal population equivalent (non-residents who either reside, work, study or spend holidays in Lloret de Mar multiplied by a weighting factor based on the total number of days in a year the person stays in Lloret de Mar) of 16,305 (leading to 2.35 citizen per household). Its urban area extends 7.8 km^2, it has a population density of 5011 hab/km^2, an average slope of 13.3 and an altitude range of 344 m. Much of the city's economy is dependent on tourism. In fact, the city has about 120 hotels, which translates into 29,147 hotel beds with a year-round average occupancy rate of about 65% in 2016[34]. In addition, the number of visits to the city in 2014 surpassed one million (Lloret Turisme Press Office). Real water consumption data from 2019 was provided by the water public utility of Girona for the validation of the decision support tool[42].

The scenarios illustrated in the results section of this paper include: (i) a comparison of the reclaimed water networks generated by different routing algorithms for public water uses in the cities of Girona and Lloret de Mar; (ii) with the best routing algorithm, a comparison of the reclaimed water networks generated for scenarios with only public water use and with both public and private water uses; and (iii) the optimal reclaimed water network with limited budget availability for the case of Girona compared with current practice (i.e., semi-manual approach). The results have been obtained using an Ubuntu 20.04 LTS server (CPU AMD Ryzen 5600X, 32GB RAM), although the tool can be used on other systems. All the computations have been spawned in a Python notebook (Jupyter Hub).

DATA AVAILABILITY

Datasets and algorithms related to this study will be made available upon request to the corresponding author.

CODE AVAILABILITY

Code implementations related to this study will be made available upon request to the corresponding author.

Received: 18 March 2022; Accepted: 23 January 2023;
Published online: 17 March 2023

REFERENCES

1. Agency, E.E. European waters assessment of status and pressures 2018. *EEA Report*, **7** (2018).
2. Gössling, S. New performance indicators for water management in tourism. *Tour Manag.* **46**, 233–244 (2015).
3. Mendoza, E., Ferrero, G., March Slokar, Y., Amores, X., Azzellino, A. & Buttiglieri, G. Water management practices in euro-mediterranean hotels and resorts. *Int. J. Water Resour. Dev.* **38**, 1–22 (2022).
4. Deyà Tortella, B. & Tirado, D. Hotel water consumption at a seasonal mass tourist destination. The case of the island of Mallorca. *J. Environ. Manag.* **92**, 2568–2579 (2011).
5. Masi, F., Langergraber, G., Santoni, M., Istenic, D., Atanasova, N. & Buttiglieri, G. Possibilities of nature-based and hybrid decentralized solutions for reclaimed water reuse. Wastewater Treatment and Reuse - Present and Future Perspectives in Technological Developments and Management Issues. *Adv. Chem. Pollut. Environ. Manag. Prot.* **5**, 145–187 (2020).
6. Chrispim, M. C. & Nolasco, M. A. Greywater treatment using a moving bed biofilm reactor at a university campus in Brazil. *J. Clean. Prod.* **142**, 290–296 (2017).
7. Domènech, L. & Saurí, D. Socio-technical transitions in water scarcity contexts: Public acceptance of greywater reuse technologies in the Metropolitan Area of Barcelona. *Resour. Conserv. Recycl.* **55**, 53–62 (2010).
8. Vallès-Casas, M., March, H. & Saurí, D. Decentralized and user-led approaches to rainwater harvesting and greywater recycling: the case of Sant Cugat del Vallès, Barcelona, Spain. *Built Environ.* **42**, 243–257 (2016).
9. Santana, M. V. E., Cornejo, P. K., Rodríguez-Roda, I., Buttiglieri, G. & Corominas, L. Holistic life cycle assessment of water reuse in a tourist-based community. *J. Clean. Prod.* **233**, 743–752 (2019).
10. Castillo, A. et al. Validation of a decision support tool for wastewater treatment selection. *J. Environ. Manag.* **184**, 409–418 (2016).
11. Poch, M., Comas, J., Rodríguez-Roda, I., Sànchez-Marrè, M. & Cortés, U. Designing and building real environmental decision support systems. *Environ. Model Softw.* **19**, 857–873 (2004).
12. Chhipi-Shrestha, G., Hewage, K. & Sadiq, R. Fit-for-purpose wastewater treatment: Conceptualization to development of decision support tool (i). *Sci. Total Environ.* **607**, 600–612 (2017).
13. Sadr, S. M. et al. A multi expert decision support tool for the evaluation of advanced wastewater treatment trains: A novel approach to improve urban sustainability. *Environ. Sci Policy* **90**, 1–10 (2018).
14. Van Afferden, M., Cardona, J. A., Müller, R. A., Lee, M. Y. & Subah, A. A new approach to implementing decentralized wastewater treatment concepts. *Water Sci. Technol.* **72**, 1923–1930 (2015).
15. Khurelbaatar, G., Al Marzuqi, B., Van Afferden, M., Müller, R. & Friesen, J. Data reduced method for cost comparison of wastewater management scenarios–case study for two settlements in Jordan and Oman. *Front. Environ. Sci.* **9**, 626634 (2021).
16. del Teso, R., Gómez, E., Estruch-Juan, E. & Cabrera, E. Topographic energy management in water distribution systems. *Water Resour. Manag.* **33**, 4385–4400 (2019).
17. Khurelbaatar, G. et al. Management of urban stormwater at block-level (MUST-B): a new approach for potential analysis of decentralized stormwater management systems. *Water* **13**, 378 (2021).
18. Yerri, S. & Piratla, K. R. Decentralized water reuse planning: Evaluation of life cycle costs and benefits. *Resour Conserv Recycl* **141**, 339–346 (2019).
19. Kesavan, H. K. & Chandrashekar, M. Graph-theoretic models for pipe network analysis. *J. Hydraulics Div.* **98**, 345–364 (1972).
20. Ahmadullah, R. & Dongshik, K. Designing of hydraulically balanced water distribution network based on GIS and EPANET. *Int. J. Adv. Comput. Sci. Appl.* **7**, 118–125 (2016).
21. Calle, E. et al. Optimal selection of monitoring sites in cities for sars-cov-2 surveillance in sewage networks. *Environ. Int.* **157**, 106768 (2021).
22. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, 10008 (2008).
23. Mehlhorn, K. A faster approximation algorithm for the steiner problem in graphs. *Inf. Process Lett.* **27**, 125–128 (1988).
24. Instituto Geográfico Nacional: Instituto Geográfico Nacional, Digital Elevation Model provider for Spain. https://www.ign.es (2019).
25. Pezoa, F., Reutter, J. L., Suarez, F., Ugarte, M. & Vrgoč, D. Foundations of json schema. Proceedings of the 25th International Conference on World Wide Web 263–273 (2016).
26. Van Rossum, G. & Drake Jr, F.L. Python Reference Manual. Department of CS, CWI, R 9525 (1995).
27. Olbricht, R. et al. Overpass API. https://dev.overpass-api.de/index.html (2011).
28. Gillies, S. et al. Shapely: manipulation and analysis of geometric objects. https://github.com/Toblerity/Shapely (2007).
29. Mind engineers: Informe tècnic d'abastament d'aigua potable gestionat per aigües de Girona, Salt i Sarrià de Ter, S.A. Technical Report 1, Aigües de Girona, Salt i Sarrià de Ter. (2014). https://seu.girona.cat/portal/dades/web/doc/2014_auditoria_AGISSA_InformeTecnic.pdf.
30. Atanasova, N., Dalmau, M., Comas, J., Poch, M., Rodrıguez-Roda, I. & Buttiglieri, G. Optimized MBR for greywater reuse systems in hotel facilities. *J. Environ. Manag.* **193**, 503–511 (2017).
31. Rossman, L. A. et al. EPANET 2: users manual.US Environmental Protection Agency. (2000). https://www.epa.gov/water-research/epanet.
32. OpenStreetMap contributors: City data retrieved from https://planet.osm.org. (2017). https://www.openstreetmap.org.

33. Boeing, G. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban Syst.* **65**, 126–139 (2017).
34. Catalan Statistics Institute: Girona population. [Online; accessed 21-dec-2020] (2019). https://www.idescat.cat/emex/?id=170792.
35. Gabarda-Mallorquí, A., Garcia, X. & Ribas, A. Mass tourism and water efficiency in the hotel industry: A case study. *Int. J. Hosp. Manag.* **61**, 82–93 (2017).
36. Garcia Acosta, X. Nous processos d'urbanització i consum d'aigua per a usos domèstics una exploració de relacions a l'àmbit gironí. (2012). http://www.tdx.cat/handle/10803/109220.
37. Gascon, L., Arregui, F., Cobacho, R. & Cabrera, E. Urban Water Demand in Spanish Cities By Measuring End Uses. 2004 Water Sources Conference 11, (2004).
38. Salas, J. J. Cuantificación y caracterización de mis aguas residuales. [Online; accessed 27-jul-2021] (2020). https://www.iagua.es/blogs/juan-jose-salas/cuantificacion-y-caracterizacion-mis-aguas-residuales-i.
39. March, J. G., Gual, M. & Orozco, F. Experiences on greywater re-use for toilet flushing in a hotel (mallorca island, spain). *Desalination* **164**, 241–247 (2004).
40. Ministerio de Fomento Código Técnico de la Edificiación. (2019). https://www.codigotecnico.org/.
41. Boneta Herrero, A., Rufí-Salís, M., Ercilla Montserrat, M., Gabarrell Durany, X. & Rieradevall, J. Agronomic and environmental assessment of a polyculture rooftop soilless urban home garden in a mediterranean city. *Front. Plant Sci.* **10**, 341 (2019).
42. Girona municipality Quantification and location of water consumption in the city of Girona. Annual Technical report. (2019). https://terra.girona.cat/apps/observatori/media/observatori/estudis/152/fitxers/basics_consumsaigua19_tot.pdf.
43. Morera, S., Remy, C., Comas, J. & Corominas, L. Life cycle assessment of construction and renovation of sewer systems using a detailed inventory tool. *Int J Life Cycle Assess* **21**, 1121–1133 (2016).
44. Banc BEDEC: Structured data bank of building elements. Funded by the Spanish Ministry of Economy and Competitivenes (2013). https://itec.cat/nouBedec.c/bedec.aspx.
45. Garey, M. R., Graham, R. L. & Johnson, D. S. The complexity of computing steiner minimal trees. *SIAM J. Appl. Math* **32**, 835–859 (1977).
46. Kou, L., Markowsky, G. & Berman, L. A fast algorithm for steiner trees. *Acta Inform* **15**, 141–145 (1981).
47. Takahashi, H. & Matsuyama, A. An approximate solution for the steiner problem in graphs. *Math. Japonic* **24**, 573–577 (1980).
48. Escribá Bonafé, D. Hidráulica para ingenieros. Librería Editorial Bellisco. Madrid (1988).
49. Aqualia: NORMAS TÉCNICAS DE ABASTECIMIENTO DE AGUA. Aqualia, (2013).
50. Clauset, A., Newman, M. E. & Moore, C. Finding community structure in very large networks. *Phys. Rev. E* **70**, 066111 (2004).
51. Pons, P. & Latapy, M. Computing communities in large networks using random walks. *J. Graph Algorithms Appl* **10**, 191–218 (2006).
52. Wakita, K. & Tsurumi, T. Finding community structure in mega-scale social networks. Proceedings of the 16th International Conference on World Wide Web 1275-1276 (2007).
53. Simpson, A. R. & Elhay, S. Formulating the water distribution system equations in terms of head and velocity. Water Distribution Systems Analysis 2008. *Trans. Am. Soc. Civ. Eng.* 1–13 (2008).
54. Costa, A. M., Cordeau, J.-F. & Laporte, G. Fast heuristics for the steiner tree problem with revenues, budget and hop constraints. *Eur. J. Oper. Res.* **190**, 68–78 (2008).

ACKNOWLEDGEMENTS

G. B., E. C., J. C., and D. M. acknowledge the research funded by the Spanish State Research Agency of the Spanish Ministry of Science and Innovation with the project code: PID2020-115456RB-I00/MCIN/AEI/10.13039/501100011033; ReUseMP3. J. P.-R. acknowledges the H2020 EU project MULTISOURCE (GA101003527). We thank Generalitat de Catalunya through Consolidated Research Groups 2021-SGR-01125 and 2021-SGR-01283. ICRA researchers thank the CERCA program/Generalitat de Catalunya for their funding. Miquel Farreras thanks the Generalitat de Catalunya and European Social Fund for his FI fellowship (2020 FISDU00590). D. M. thanks the Universitat de Girona for his FI fellowship (IFUdG 46 2022). G. B. acknowledges Spanish State Research Agency of the Spanish Ministry of Science, Innovation and Universities for the Grant to the Creation of a permanent position Ramon y Cajal 2014 (RYC-2014-16754). We would like to thank the company ABM Consulting for providing support on the economic data and the Municipality of Girona and Aigües de Girona, Salt i Sarrià de Ter for providing water consumption real data.

AUTHOR CONTRIBUTIONS

E. C.: Conceptualization; Formal analysis; Investigation; Methodology; Supervision; Validation; Visualization; Writing, review & editing. D. M.: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing, review & editing. G. B.: Conceptualization; Supervision; Visualization; Writing, review & editing. L. C.: Conceptualization; Supervision; Visualization; Writing, review & editing. M. F.: Data curation; Investigation; Methodology; Writing. J. S.-G.: Data curation; Formal analysis; Investigation; Methodology; Writing. P. V.: Conceptualization; Supervision; Visualization; Writing, review & editing. J. P.-R.: Conceptualization; Data curation; Formal analysis; Investigation; Writing, review & editing. J. C.: Conceptualization; Formal analysis; Methodology, Supervision, Validation, Writing, review & editing.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41545-023-00222-4.

Correspondence and requests for materials should be addressed to Joaquim Comas.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

5.3 Seamless integration of sewer system topology and tree location data: An algorithm to diagnose the potential impact of tree roots on pipes and propose rearrangement solutions

Heliyon 10 (2024) e23382

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Seamless integration of sewer system topology and tree location data: An algorithm to diagnose the potential impact of tree roots on pipes and propose rearrangement solutions

David Martínez [a,*], Sergi Bergillos [a], Lluís Corominas [b], Marc Comas-Cufí [a], Eusebi Calle [a]

[a] Institute of Informatics and Applications, University of Girona, Maria Aurèlia Capmany 61 (Edifici PIV), 17003, Girona, Spain
[b] Catalan Institute for Water Research, Emily Grahit 101, 17003, Girona, Spain

ARTICLE INFO

Keywords:
Wastewater leakages
Risk assessment
Pipe failures
Underground networks
Environmental protection
Rearrangement solutions

ABSTRACT

Wastewater networks are subject to several threats leading to wastewater leakages and public health hazards. External elements such as natural factors and human activities are common causes of wastewater leakages and require more in-depth analysis. Prevention and rehabilitation work is essential to secure wastewater networks and avoid pipe failures. This work presents a new algorithm that allows for the seamless integration of sewer topology and tree location data to diagnose the potential impact of tree roots on pipes. The algorithm also proposes tree rearrangement options that balance the cost of tree rearrangement with the cost of pipe repair. The paper also showcases a real-world case study in the city of Girona to evaluate the performance of the presented algorithms for a specific case focusing on tree roots as a natural factor. Results show that it is possible to optimally rearrange a number of the trees with the greatest impact, significantly minimizing pipe failures and wastewater leakages (82% risk reduction with only rearranging a 12% of the most impactful trees). The rearrangement solution not only protects the environment and prevents public health hazards, but also achieves a positive economic payback during the operational period of the pipes, saving up to 1.33M€ for a tree rearrangement of 7%. The presented methodology is applicable to other natural or human factors.

1. Introduction

A wastewater network is formed by a conglomeration of underground pipes and maintenance holes that work together in order to collect and drain wastewater from households or industrial centers to wastewater treatment plants. Once treated, the water is returned to the environment or reused for beneficial purposes such as agriculture, irrigation, or even potable water supplies [1], ever more widely accepted by the general public [2]. Hence, wastewater networks are critical infrastructures and essential assets for the proper functioning of society and the economy [3,4].

The wastewater network is subject to several threats that lead to leakages and, thus, significant economic losses and public health hazards [5]. Pipe failures cause not only direct economic costs through repairs, but also indirect costs such as damage to

* Corresponding author.
E-mail addresses: david.martineza@udg.edu (D. Martínez), sergi.bergillos@udg.edu (S. Bergillos), lcorominas@icra.cat (L. Corominas), marc.comas@udg.edu (M. Comas-Cufí), eusebi.calle@udg.edu (E. Calle).

https://doi.org/10.1016/j.heliyon.2023.e23382
Received 14 July 2023; Received in revised form 1 December 2023; Accepted 1 December 2023

Available online 7 December 2023

infrastructures, disruption to business, and production losses [6]. The concept of "urban water security" emerged to address such vulnerabilities leading to a new understanding of the complex dynamics between human and natural systems and can pave the way to extend the scope of risk management [7].

Christodoulou et al. [8] and Obradović [9] agree that natural factors such as tree roots affect pipe failure hazard rates. In particular, they recommend checking pipes in the proximity of trees and evaluating the possibility of tree rearrangement, although a more in-depth analysis is needed. According to their work, the risk of pipe failure incidents concerning tree roots increases over time and can be exacerbated on old or corroded pipes [10]. Sydney Water [11] agrees with this statement, justifying that the roots of trees planted in the wrong place can find their way into wastewater pipes, causing about 80% of all dry weather sewage overflows and seriously affecting public health and the environment. However, there is an active discussion on whether tree roots are able to crack pipes. According to Hartley [12], tree roots may affect pipes in other ways such as joint intrusion or tensile forces. For example, no matter the individual root size, the total volume of the tree roots in a joint could develop a surface big enough to break the pipe collar.

The concept of Tree Protection Zones (TPZs) is widely known in the world of arboriculture, and is defined as the calculated area above and below ground at a given distance from the tree trunk to provide for the protection of the tree's roots and canopy during construction works [13]. Although the TPZ is tied to tree protection, it can be used the other way around. In other words, although TPZ defines an area where construction works can affect tree roots, this area can also be considered the impact zone where tree roots can break into infrastructures such as wastewater pipes. The Australian Standard [14] is the most widely-accepted method for calculating TPZ, although it has led to discussion [15].

Tree species also influence the hazard rate depending on their capacity to build and extend root systems. Ward and Clatterbuck [16] provided a list of slow-growing tolerant trees considered "sewer-safe", and Sydney Water [11] a list of fast-growing sensitive species. Hence, planting the right species of tree also reduces pipe failure hazard rates [9]. Moreover, Östberg et al. [17] identified which tree species are most likely to crack pipes through fieldwork inspections of wastewater networks. Specifically, it is advisable to conduct CCTV sewer surveys at increased intervals, particularly when previous surveys have revealed a history of high-risk pipe failures [18].

Municipalities often tend to maintain an inventory of trees and their respective species, as well as details about the wastewater network (e.g., pipe age, material, length). Given the extensive datasets available to municipalities, there is potential to apply AI techniques to predict failure risks. Dawood et al. [19] achieve this by performing a literature review on AI algorithms to predict drinking water pipes' risk of failure and highlighted that the main parameters related to this are physical factors such as age, length, diameter, and material. Other factors, including environmental (e.g., tree roots) and operational ones, present a lower effect on pipe failure risk, although they should also be considered.

Apart from the AI approaches, Amiri-Ardakani and Najafzadeh [20] presented a probabilistic framework, considering both natural and human factors, for pipe break rate estimation. Other works require Monte Carlo simulations to predict risk probabilities [21]. Moreover, Vishwakarma and Sinha [22] introduced a prediction method using a fuzzy inference system to assess the comprehensive failure impacts of water pipes based on economic, social, and environmental impacts, operational characteristics, and renewal complexity.

Human factors such as street works or building constructions have been proven as other of the main factors leading to accidents and also contribute to wastewater leakages and pipe failures [23]. Managing human activities is essential to prevent wastewater leakages. Hence, the concept of TPZ can be applied to other human factors, such as urban construction works, and used as an Impact Area (IA) metric for physical elements that can make their way into wastewater pipes (Fig. 1).

Intelligent tree positioning in cities has recently been applied in other fields of research, such as the strategic planning of trees in a city area to improve the walkability of the outdoor space [24]. However, tree rearrangement has not been studied yet for minimizing the impact of tree roots on pipes and, thus, going towards securing wastewater networks.

Drawing from this background and the availability of relevant databases for pipe and risk element geolocalization, it is possible to automate processes through mathematical algorithms to determine whether it is better to wait and repair pipes when failures or leakages occur or avoid them by rearranging nearby elements. The novelty of this approach is to automatically cross these databases for fast environmental protection and city planning, providing cost-effective impact and risk assessments, as it requires already available data in municipalities and without the need for fieldwork in sewers. The contributions are the following:

1. A new concept of Impact Areas (IAs), which defines the impact zone where physical elements can break into underground network infrastructures, evaluates the impact of the elements and detects the worst threats based on the idea of Tree Protection Zones (TPZs).
2. Mathematical algorithms and methods are introduced to analyze the impact of risk elements on the pipes and evaluate the pipe failure hazard risk on wastewater networks.
3. A novel algorithm provides Element Rearrangement (ER) solutions that minimize pipe failure hazard risk and mitigate wastewater leakages.
4. For a specific case study considering tree roots as natural elements, a comparison between the results obtained and an "all sewer-safe tree city", which is an ideal city where all trees would be slow-growing or tolerant, quantifies the benefit and wastewater leakage risk reduction of planting "sewer-safe" trees at convenient locations.
5. A method to estimate the probabilities of pipe failures is also introduced and used to extract the expected cost of repair. The original and ER scenarios are then compared and analyzed.

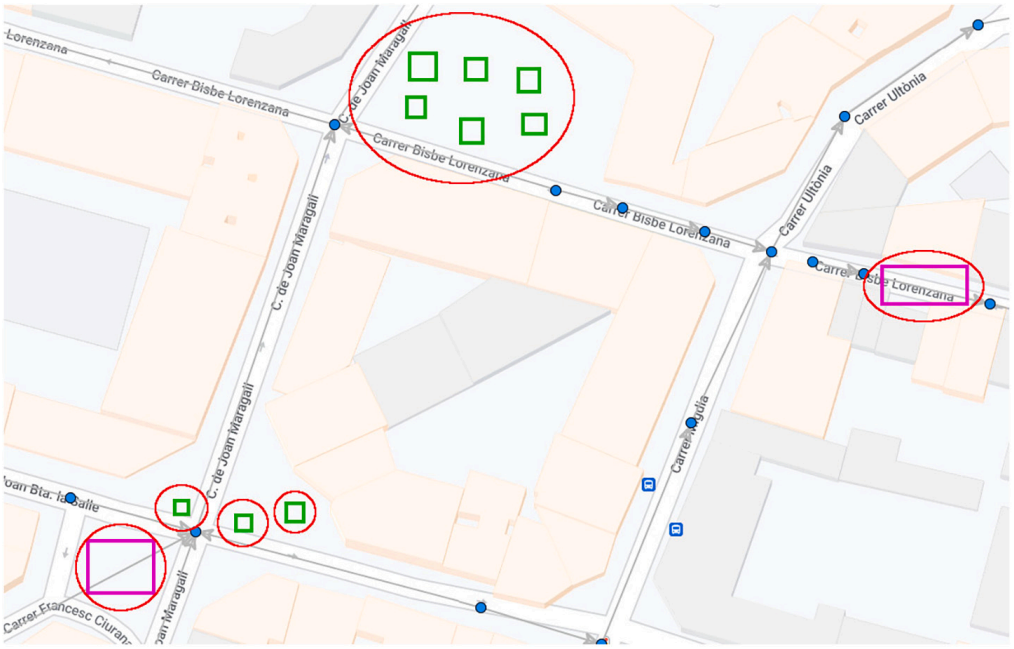


Fig. 1. Examples of potential Impact Areas (IAs) drawn in red ellipses considering natural (green rectangles representing urban trees) and human (purple rectangles representing street works) factors in a portion of the wastewater network of Girona (pipes in gray lines, junctions in blue-filled circles).

6. The economic savings (i.e., payback) considering the expected cost of repairs and the cost of the ER works are analyzed for both the original and ER scenarios.

The remainder of the paper is organized as follows. Section 2 presents the methodology followed; including an explanation of the general methodology (Section 2.1), the definition of the algorithms and methods (Section 2.2), and the introduction of the case study (Section 2.3). Section 3 illustrate the results and effectiveness and considers the approach described in the paper together with the discussion and future work, including the results of the tree impact and pipe failure risk analyses (Section 3.1), the most damaging trees detection (Section 3.2), the pipe risk reduction from the Element Rearrangement (ER) algorithm (Section 3.3), the pipe failure probabilities and expected cost of repairs (Section 3.4), and the economic savings of the ER algorithm (Section 3.5). Finally, Section 4 summarizes the results and contributions of the paper.

2. Materials and methods

The following methods are based on graph theory to build and manage the layer of wastewater pipe networks [25]. Several previous studies have used graph theory in water distribution and wastewater networks: Ahmadullah and Dongshik [26] for designing drinking water networks; Calle et al. [27] for wastewater sensor placement approaches concerning SARS-CoV-2 detection; and Meng et al. [28] for proposing a comprehensive analytical framework for examining the resilience pattern of water distribution systems against topological characteristics (i.e., the correlations between resilience and topological features).

2.1. General methodology

The methods and mathematical algorithms presented in this paper involve the following steps: (i) defining the scenario; (ii) preparing the scenario; and (iii) estimating key output indicators (Fig. 2).

2.1.1. Defining the scenario

The required user inputs are (i) a city's water network graph and (ii) the risk elements data. The first parameter (i) must be in a graph format (e.g., Graphml [29]), usually converted from a geographical information format (e.g., GIS) provided by water companies. The second parameter (ii) must include at least the location and, in the case of trees as natural risk elements, an interval (or exact value) of trunk perimeter or diameter and, optionally, the species of the trees. If provided, tree species allow for a more finely-tuned calculation of the Impact Area (IA) depending on the Tree Protection Zone (TPZ) computation method.

A safe radius proportional to the risk element sizes is used to calculate the IAs, which varies in function of the method being considered. In the specific case of trees, Table 1 compares the three most used TPZ methods: Day et al., the Australian Standard, and considering all trees as "sewer-safe". The last method is used to analyze the gain in pipe failure risks (i.e., failure minimization) in a hypothetical scenario where all the city trees would be "sewer-safe" (i.e., an ideal city where all trees would be slow-growing or tolerant).

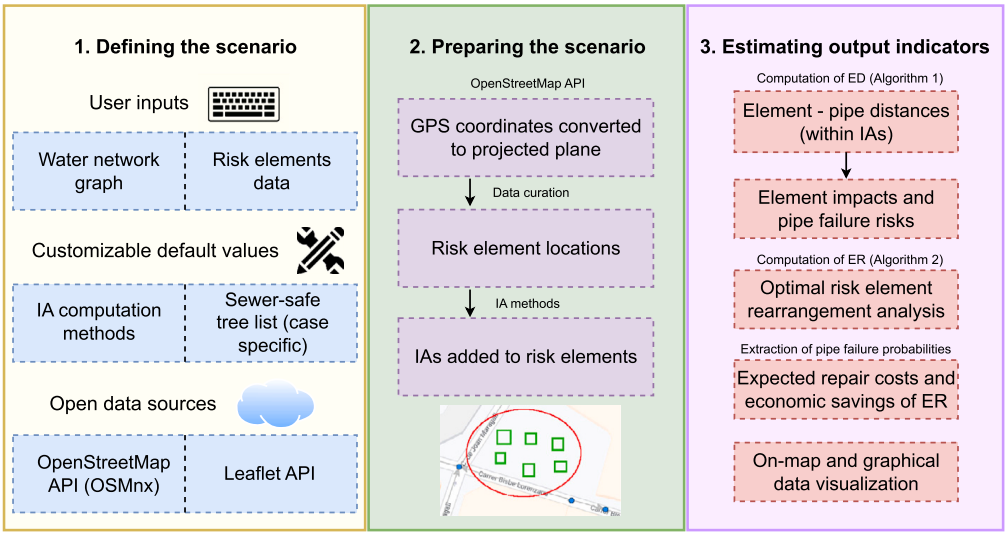


Fig. 2. Methodology scheme for the presented methods and algorithms.

Table 1
Comparison between the three most used Tree Protection Zone (TPZ) methods used to calculate the Impact Area (IA) in the case of trees as risk elements.

IA	TPZ method	References
Day et al.	6:1 ratio (radius of TPZ:trunk diameter) for tolerant or "sewer-safe" trees and 18:1 for sensitive fast-growing species	[30,31]
Australian Standard	12:1 ratio (radius of TPZ:trunk diameter) for all trees	[14]
All "sewer-safe"	6:1 ratio (radius of TPZ:trunk diameter) for all trees	[30,31,16]

2.1.2. *Preparing the scenario*

The preparation of the scenario starts with converting the original network coordinates onto a projected EPSG 3857 plane, a Spherical Mercator projection coordinate system popularized by Google Maps and later OpenStreetMap. Next, risk element locations are processed together with their IAs, which depend on the risk element size and type.

In the case of trees as a risk element type, tree diameter is estimated based on trunk perimeter or diameter intervals depending on the original dataset. In the case of diameter intervals, this is considered the upper value (i.e., worst-case scenario). Then, "sewer-safe" trees that are identified through the computation of the Tree Protection Zone (TPZ) may differ based on the method. Finally, the TPZ is computed for each tree based on the selected TPZ method, obtaining the Impact Areas (IAs) and the scenario is ready for the computation of the algorithms.

2.1.3. *Estimating output indicators*

The impact of a risk element is the value representing the effect of this element on the nearby pipes inside its Impact Area (IA) based on the distances between the element and the pipes. The pipe failure risks are calculated from the aggregation of the affected impacts of each element for each pipe. The calculated IAs are used together with the distances obtained from the execution of the ED algorithm (Algorithm 1) to calculate the element impacts and the pipe failure risks. Furthermore, impacts are used for the optimal analysis of rearrangement to minimize pipe failure risk through the execution of the Element Rearrangement (ER) algorithm (Algorithm 2).

The probabilities of pipe failures are also calculated considering the failure risks, thus enabling the extraction of the expected repair costs. The expected repair costs can be calculated for both the original scenario (i.e., original element locations) and considering rearranged elements (ER algorithm). Furthermore, rich on-map and graphical data visualization of the results are generated to help visualize the numerical results through the open data sources (e.g., data histograms, geographical pipe risk and most-impacting-element maps, line charts, etc.).

2.2. *Algorithms*

In brief, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the wastewater network graph, with a V-element set of nodes $\mathcal{V}$ representing the set of origin (wastewater entries) nodes, the wastewater treatment plant, and junction points, and an E-element set of links $\mathcal{E} \subset \mathcal{V}^{|2|}$ representing pipes. Additionally, $\mathcal{T}$ denotes a T-element set of risk elements. Table 2 specifies the notation used for the algorithms.

First, the element-pipe distances (ED) algorithm (Algorithm 1) makes use of the *nearest_edges* function, a key component of the OSMnx library, as detailed in Boeing's work [32]. The *nearest_edges* function serves a straightforward purpose: it identifies the closest water network pipe e to a specific geographic point, representing the location of a risk element t within the context of the study.

Table 2
Full notation concerning the algorithms.

$\mathcal{T}$	set of risk elements
$\mathcal{E}$	set of wastewater network pipes
$z(t), t \in \mathcal{T}$	Impact Area (IA) of risk element t
$d(t,e), t \in \mathcal{T}; e \in \mathcal{E}$	distance between risk element t and pipe e
$i(t,e), t \in \mathcal{T}; e \in \mathcal{E}$	impact of risk element t on pipe e
$i(t), t \in \mathcal{T}$	impact of risk element t on the wastewater network
$r(e), e \in \mathcal{E}$	failure risk of pipe e based on risk element impact aggregation
$p(e), e \in \mathcal{E}$	probability of failure of pipe e (caused by risk elements during its operational period)
$l(e), e \in \mathcal{E}$	length of pipe e
$c(e), e \in \mathcal{E}$	expected repair cost of pipe e based on $p(e)$

Algorithm 1: Element - pipe distances (ED) algorithm.

Step 1: Initialize the wastewater network graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and risk element set $\mathcal{T}$.
Step 2: For each risk element $t : t \in \mathcal{T}$, obtain its nearest pipe $e : e \in \mathcal{E}$ and distance $d(t,e)$ using the OSMnx *nearest_edges* function.
Step 3: For each risk element $t : t \in \mathcal{T}$ and its nearest edge $e : e \in \mathcal{E}$:
(a) Let $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ be a copy of the graph $\mathcal{G}$.
(b) While $d(t,e) \leq z(t)$:
(i) Remove pipe e from graph $\mathcal{G}'$, such that $E' := E' \setminus \{e\}$.
(ii) Obtain the nearest pipe $e' : e' \in \mathcal{E}'$ of risk element t and the distance $d(t,e')$ using the OSMnx *nearest_edges* function.
(iii) Set the current nearest pipe as $e := e'$.

Step 4: $d(t,e), t \in \mathcal{T}; e \in \mathcal{E}$ represents a data structure with element-pipe distances for all risk elements $t \in \mathcal{T}$ and pipes $e \in \mathcal{E}$ in which $d(t,e) \leq z(t)$.

The ED algorithm plays a pivotal role in determining the set $d(t,e)$, which characterizes the distances between each identified risk element t and all the water network pipes e that are situated within the risk element t Impact Area (IA). For each risk element t and its nearest pipe e, acquired through the *nearest_edges* function, the ED algorithm systematically checks if the pipe e falls within the IA of the risk element t. If it does, the algorithm removes this pipe e from the original wastewater network $\mathcal{G}$ and executes the *nearest_edges* function once more to identify the next nearest pipe e'. This process continues until the algorithm identifies pipes and element-pipe distances that fall outside the IA for all the risk elements. In essence, the ED algorithm harnesses the capabilities of spatial analysis and geographic data to precisely compute these $d(t,e)$ distances.

Then, Equation (1) describes the risk element impacts for each pipe $i(t,e), t \in \mathcal{T}, e \in \mathcal{E}$. In other words, the impact of element t on a pipe e is defined as a normalized value between 0 and 1. The maximum value represents the pipe passing through the center of the IA, and the minimum represents the pipe passing just at the edge of the IA (Fig. 3). If the pipe is outside the IA, the value is considered 0.

$$i(t,e) := \max\{1 - \frac{d(t,e)}{z(t)}, 0\} \tag{1}$$

The impact aggregation of risk element $i(t)$ on the whole wastewater network is simply the summation of risk element t impact over the network pipes $e \in \mathcal{E}$ (Equation (2)). Note that the impact aggregation $i(t)$ values can be greater than one, and taking the example of Fig. 3, a value of $i(t) = 1.1$ represents the impact aggregation of the pipe e_1 ($i(t,e_1) = 0.4$) and the pipe e_2 ($i(t,e_2) = 0.7$), which are those inside the IA of risk element t.

$$i(t) := \sum i(t,e), e \in \mathcal{E} \tag{2}$$

The same procedure is followed in Equation (3) to obtain the pipe failure risk of each pipe $r(e)$ considering all risk elements, i.e., the summation of each risk element $t \in \mathcal{T}$ impact over the network pipe e.

$$r(e) := \sum i(t,e), t \in \mathcal{T} \tag{3}$$

Although the pipe failure risk $r(e)$ is a quantitative measure, Table 3 shows a proposal of qualitative assessment, which assists in interpreting the results of the measure presented later in the results section. Note that the assessment value intervals can be adjusted with more research or using other requirements.

The Element Rearrangement (ER) algorithm (Algorithm 2) minimizes the pipe risk by rearranging a portion of the most impactful risk elements. The algorithm contemplates two options: (i) risk element removal or (ii) element replacement in the case of trees as a natural risk element type, as trees can be replaced with a smaller "sewer-safe" tree alternative). Moreover, the algorithm also considers an exception element list $\mathcal{U}$ for cases where it is not possible to remove specific elements (e.g., in the case of trees for cultural or historical significance or technical challenges; or in the case of street building constructions for the original location being the only available option).

In the context of tree roots as a natural factor, the ER algorithm operates as follows: it begins by selecting the top p percentage of the most impactful trees from the entire set of risk elements $\mathcal{T}$, creating a new set denoted as $\mathcal{T}'$. Subsequently, each tree t in $\mathcal{T}'$ is

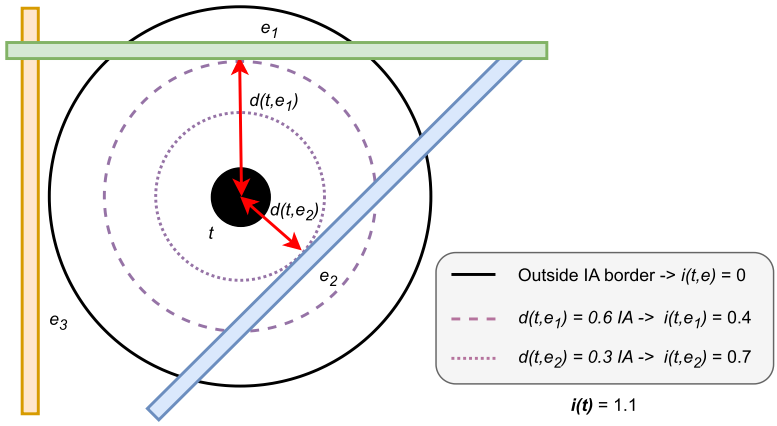


Fig. 3. Example of risk element impacts $i(t,e)$ computation within an IA (t as a tree).

Table 3
Proposal of pipe qualitative risk assessment.

Pipe risk $r(e)$ value	Qualitative assessment
$r(e) = 0$	The risk e is "non-existent". The pipe does not pass through any Impact Area (IA).
$0 < r(e) \leq 1$	The risk e is "low". At most, the risk element impact aggregation on the pipe reaches one unit.
$1 < r(e) \leq 3$	The risk e is "moderate". The risk element impact aggregation on the pipe outreaches at most three units and surpasses one at least.
$3 < r(e)$	The risk e is "high". The risk element impact aggregation on the pipe surpasses three units.

Algorithm 2: Element Rearrangement (ER) algorithm.

Step 1: Initialize the wastewater network graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, risk element set $\mathcal{T}$, exception risk element set $\mathcal{U}$, set $i(t)$, percentage p (from 0 to 100, natural-factor specific), replace r (True, False), replace tree perimeter r_p (in centimeters, natural-factor specific), and rearranged risk element set $\mathcal{V}$.
Step 2: Let $\mathcal{T}'$ be the set of p percentage most impactful risk elements based on highest $i(t)$ for all $t \in \mathcal{T} \notin \mathcal{U}$.
Step 3: For each risk element $t : t \in \mathcal{T}'$:
(a) Remove risk element t from set $\mathcal{T}$, such that $\mathcal{T} := \mathcal{T} \setminus \{t\}$.
(b) Add risk element t to set $\mathcal{V}$, such that $\mathcal{V} := \mathcal{V} \bigcup \{t\}$.
(c) If r = True, then:
(i) Place a "sewer-safe" tree risk element s on the same location of risk element t, such that $d(t,e) = d(s,e), \forall e \in \mathcal{E}$.
(ii) Compute $z(s)$ considering a tree perimeter of r_p.
(iii) Add "sewer-safe" tree risk element s to set $\mathcal{T}$, such that $\mathcal{T} := \mathcal{T} \bigcup \{s\}$.

Step 4: $\mathcal{T}$ contains the new risk element set and $\mathcal{V}$ contains the list of rearranged elements. Recompute $i(t,e)$, $i(t)$, and $r(e)$ to obtain the new elements's impact and pipe risk analysis.

removed from $\mathcal{T}$ and integrated into $\mathcal{V}$, signifying the exclusion of the specific tree t from the original set. When the "replace" option r is activated, a new "sewer-safe" tree s with a perimeter of r_p is introduced at the same location as the original tree t. The Impact Area (IA) of the new tree s, denoted as $z(s)$, is then calculated, and this newly introduced tree s is incorporated into the set of risk elements $\mathcal{T}$.

Consequently, the algorithm proceeds to recompute the impact of the trees and the risks associated with the pipes (i.e., $i(t,e)$, $i(t)$, and $r(e)$). As a result, the $\mathcal{T}$ set now contains the updated risk element set, while $\mathcal{V}$ maintains a record of the relocated elements.

To extract the probabilities of pipe failures caused by risk elements $p(e), e \in \mathcal{E}$, a new customizable threshold R_{fail} sets the value of risk $r(e)$ in which there is a 100% probability of failure of pipe e during its operating time period, often considered 30 years (i.e., if $r(e) = R_{fail}$, then $p(e) = 1$) [33]. Depending on the desired R_{fail} value, some probabilities may be greater than one due to data outliers (i.e., if $R_{fail} < r(e), e \in \mathcal{E}$), which may indicate the probability of more than one failure during the operation period. From the R_{fail} threshold, the failure probability of each pipe $p(e)$ can be calculated through its risk value $r(e)$ normalized with R_{fail} (Equation (4)).

$$p(e) := \frac{r(e)}{R_{fail}} \tag{4}$$

According to most sources consulted, including ABM Consulting, entire pipe sections affected by physical elements failures such as tree roots must be replaced. The repair costs depend on multiple factors, such as pipe diameter, material, and terrain. The total cost of pipe repair per meter R has to include the material, placement, earthmoving works, and eventually affected services (e.g., economic losses from a temporary road closure), which will depend on the country of the case study.

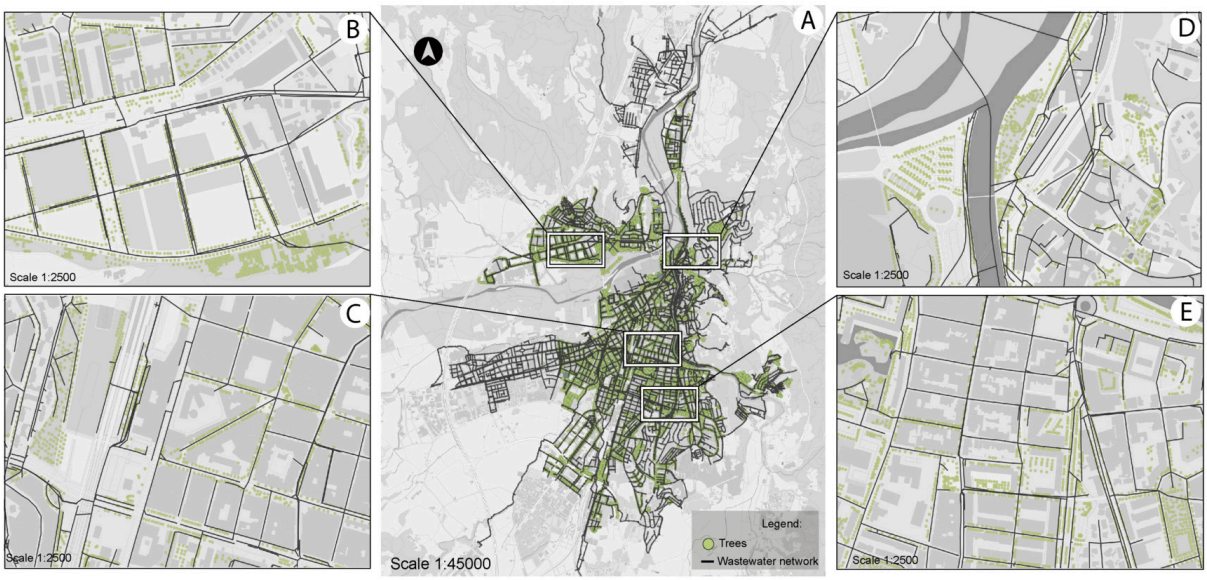


Fig. 4. The case study area of Girona, showing tree locations labeled as green points and wastewater network pipes as black lines (QGIS generated). A - Entire city area; Scale 1:45000. B - Zoomed area of Domeny neighborhood; Scale 1:2500. C - Zoomed area of Eixample neighborhood; Scale 1:2500. D - Zoomed area of Barri Vell neighborhood; Scale 1:2500. E - Zoomed area of Montilivi neighborhood; Scale 1:2500.

The expected cost of a failure repair for each pipe $c(e), e \in \mathcal{E}$ can be estimated through failure probability $p(e)$ and the concept of expected value (i.e., multiplying the total cost of repair per meter R by its length and the likelihood pipe failure will occur $p(e)$), such that (Equation (5)):

$$c(e) := R \times l(e) \times p(e) \tag{5}$$

Finally, the total expected repair cost for the whole wastewater network C is simply the summation of the expected repair cost of each pipe, such that (Equation (6)):

$$C := \sum c(e), e \in \mathcal{E} \tag{6}$$

2.3. Case study

The usefulness of the methods and algorithms presented in this paper is illustrated in the city of Girona, Catalonia (northeast of the Iberian Peninsula, see Figs. 4, 4A, 4B, 4C, 4D, 4E), considering the entire dataset of the city trees as natural factor risk elements. Girona, with its 102,666 inhabitants and 47,446 households (2.2 citizens per household), is a typical compact Western Mediterranean city [34]. Its urban area extends 12.7 km^2 on a rivers' crossing, has a population density of 8,139 hab/km^2, an average slope of 5.1, and an altitude range (difference between the minimum and maximum altitudes) of 177 m. Its sewage network consists of more than 6,000 maintenance holes, resulting in a large network totaling 265 km of pipes. The basic topological characteristics of the network layer are 7,946 nodes (V); 8,303 edges (E); an average nodal degree of 2.1 ($\overline{D}$); a diameter of 11,071 meters (∅); and an average shortest path length of 47 meters ($\overline{d}$).

On the one hand, the company Cicle de l'Aigua del Ter, which manages the wastewater network of Girona, provided the topological data from the city sewer network in geographic information system (GIS) format files, including feature geometry and attributes. First, the GIS files needed to be converted to a GraphML file format (i.e., to an XML-based format [29]). The GraphML file format is compatible with the Network Robustness Simulator (NRS) [35], used for graph analysis and the execution of the algorithms. The final wastewater network graph is in a unique file in GraphML format containing both nodes and edges and their attributes. Next, data verification and reconciliation processes are performed based on previous research [27].

On the other hand, the city tree data was obtained from the Girona Open Data portal [36], which includes an extended dataset (in CSV format) of 32,881 trees updated in January 2023 with the following attributes: (i) scientific name; (ii) common name in Catalan; (iii) trunk size, as perimeter intervals in centimeters (20-50 / 50-80 / 80-120 / +120); (iv) tree pit frame, as size of the larger side intervals in centimeters (-40 / 40-90 / +90); (v) x coordinate, in UTM ETRS89 format; and (vi) y coordinate, in UTM ETRS89 format. In order to apply the algorithms to this case study, it is necessary to prepare the data of the tree dataset according to our methodology, converting the coordinates and estimating tree diameters.

3. Results and discussion

The results are grouped and presented alongside the discussion, and future work is also introduced at the end of each subsection. The results presented in this section are the following: (i) tree impact and pipe failure risk analysis based on the ED algorithm

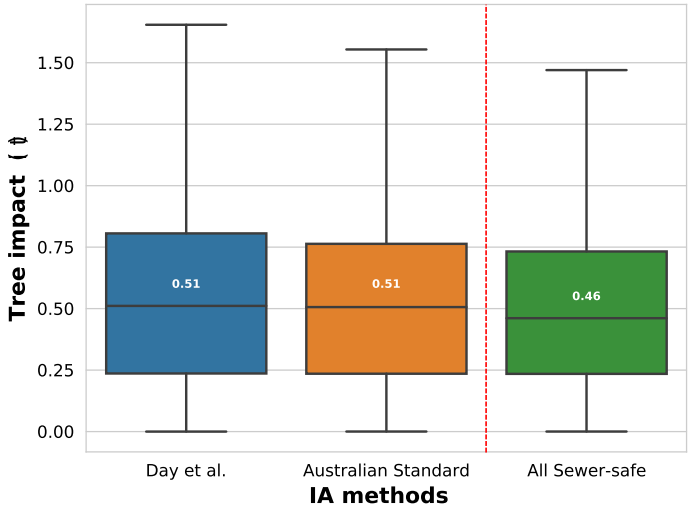


Fig. 5. Histogram of tree impacts $i(t), t \in \mathcal{T}$: Day et al. comparison with Australian Standard (AS) IA methods, with the All Sewer-safe method to contrast (without outliers).

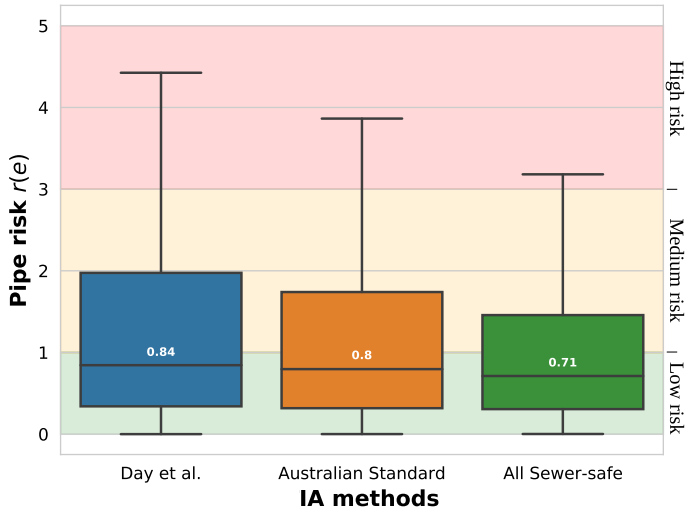


Fig. 6. Histogram of pipe risk $r(e), e \in \mathcal{E}$ grouped by the three Impact Area (IA) computation methods (without outliers).

(Algorithm 1) computation; (ii) most impactful tree detection; (iii) pipe risk minimization through the ER algorithm (Algorithm 2); (iv) extraction of pipe failure probabilities caused by tree roots and the expected cost of repairs; and (v) the actual economic savings (i.e., payback) of the ER algorithm. The results have been obtained using an Ubuntu 22.04 LTS server (CPU AMD Ryzen 5 5600X, 32 GB RAM). All the computations have been spawned in a Python [37] notebook (Jupyter Hub).

3.1. Tree impact and pipe failure risk analysis

Tree impacts $i(t)$ have been calculated based on the element-pipe distances obtained from the ED algorithm (Algorithm 1). Fig. 5 illustrates the histogram comparison of tree impacts $i(t) > 0$ for Day et al. and Australian Standard (AS) TIA methods. It is worth noting that the tree impact data distribution is almost equal for both Day et al. and AS methods (i.e., a median of 0.51), which demonstrates that the AS is an excellent approximation, without the need for tree specie data, for the IA computation. In contrast, Fig. 5 also shows the tree impact $i(t)$ results in a hypothetical situation where all city trees were "sewer-safe" with the same trunk size and location as the actual ones (i.e., identical tree sizes above the surface with much less tree root areas below). The comparison between the Day et al. and AS methods and the All Sewer-safe suggests a significant minimization of pipe failure risk, proving that the median of tree impacts would be reduced by 10% if all the city trees were "sewer-safe". However, more research and other case studies are needed to check this tendency.

The histogram of pipe failure risks concerning tree roots $r(e) > 0$ is shown in Fig. 6, where the background colors represent the qualitative failure risk assessment values (see Table 3). The data distribution follows a similar pattern to tree impacts, with the Day et al. method presenting a slightly higher 0.84 median compared to the 0.8 of the AS. As the AS method keeps showing an excellent approximation, it has been selected as the IA method for the rest of the results. The All Sewer-safe method reveals a more significant reduction in pipe risks data distribution, as low as 0.71, with very little presence of high-risk pipes.

Table 4
Top 10 of the most impactful trees in Girona (IA method: Australian Standard).

No.	Scientific name	Trunk per. (cm)	Trunk diam. max. (cm)	IA A.S. (m)	Tree impact $i(t)$
1	Celtis australis	80-120	38.20	4.60	4.98
2	Tilia platyphyllos	80-120	38.20	4.59	4.59
3	Magnolia grandiflora	50-80	25.46	3.06	4.35
4	Melia azedarach	80-120	38.20	4.58	4.19
5	Platanus x hispanica	>120	50.93	6.11	3.77
6	Celtis australis	80-120	38.20	4.58	3.67
7	Pinus pinea	20-50	15.92	1.91	2.86
8	Celtis australis	>120	50.93	6.11	2.86
9	Tilia platyphyllos	>120	50.93	6.11	2.78
10	Tilia platyphyllos	>120	50.93	6.11	2.77

The potential impact analysis of tree roots on pipes in wastewater networks throughout multi-layer crossing was a gap in the current literature. It is worth noting that the pipe failure risk $r(e)$ is a relevant quantitative measure that will be very useful for planning and prioritizing preventive actions on wastewater network pipes.

Future work will be conducted to consider pipe material in the computation of failure risks which, in practice, affects the failure hazard rate [19]. In addition, it would be worthwhile to analyze the effect pipe depth has on risk element impacts. Finally, it would also be interesting to verify our approach with the actual city failure records, as this is, to the best of our knowledge, the first study that quantifies pipe failure risk concerning tree roots at a theoretical level. Our approach does not require sewer fieldwork and is based only on the topology of the network and the tree inventory. The obtained data can also be used to improve the existing Artificial Intelligence (AI) algorithm predictions.

3.2. Most impactful tree detection

Table 4 shows the top 10 most impactful trees considering the Australian Standard IA method, illustrating that a small percentage of trees cause the most impact on pipes due to the large number of upper-bound outliers on the data (as shown in Fig. 5) and the significant 44% decrease in the values between the first and the tenth-placed tree. The most impactful trees are expected to be large or in a critical location where many pipes are present (e.g., street crossroads), or both. The majority of the most impactful trees have large IAs, except for number seven, the *Pinus pinea*. A manual check on the location of this tree revealed that it is placed in a critical spot, on a roundabout with a union of seven pipes. It is also worth noting that the most present species in the top 10 are the Celtis australis and the Tilia platyphyllos, which are not considered “sewer-safe” and are well-known for their relatively extensive root systems.

Fig. 7 visualizes the central part of the wastewater network of Girona represented with plane coordinates, with the color of each pipe representing the risk category and the black dots representing each location of the top 10 most impactful trees. The algorithms generate high-quality PDF maps as well as interactive HTML maps (available on the dedicated public repository [38]), both generated automatically, clearly helping to identify critical pipe risk areas and tree locations. As can be seen in the figure, the location of the most impactful trees matches, in most of the cases, where pipe failure risks are high or moderate, showing in a visual way that these trees are significantly affecting the pipe failure risks.

The most impactful trees and the pipe failure risk map are extremely useful reports for city councils and wastewater network managers, providing an excellent first image of the current scenario. They can be generated easily and without the necessity for additional resources, lots of data, or fieldwork in sewers, in contrast with the existing literature [17,18]. After this preliminary assessment, decision-makers may require the application of additional methods to minimize the risks, such as localized CCTV inspections in the most critical areas.

In line with the advances in smart cities, future research may develop intelligent tree-planting approaches to indicate in which city zones trees can be planted without being a risk to the wastewater network. In any case, a prioritized list of the most impactful trees concerning wastewater pipes is highly useful information for cities to plan future proceedings.

3.3. Element Rearrangement (ER): pipe risk reduction algorithm

The Element Rearrangement (ER) algorithm is expected to decrease the pipe risk significantly by rearranging a portion of the most impactful trees obtained from the previous analysis. Fig. 8 proves this statement showing the Element Rearrangement (ER) algorithm risk reduction considering the Australian Standard TPZ method with tree replacement enabled from 0 to 13% of the dataset trees (i.e., within the percentages that present a clear improvement). The replacement approach considers planting a “sewer-safe” alternative tree in the same spot as the original one, with an assumed trunk perimeter of 50 cm (i.e., the upper value of the smallest trunk perimeter interval of the case-study dataset). With a rearrangement of only 4% of the dataset trees, the number of high-risk pipes is reduced drastically by 75%, and medium-risk ones by 30%. Moreover, the number of medium-risk pipes is reduced significantly by 77% with a ER of 8%. Finally, the number of low-risk pipes is also reduced sharply by 79% with a ER of 13%. With a small percentage of rearranged trees, the pipe failure risk can be lowered substantially, especially for high-risk pipes.

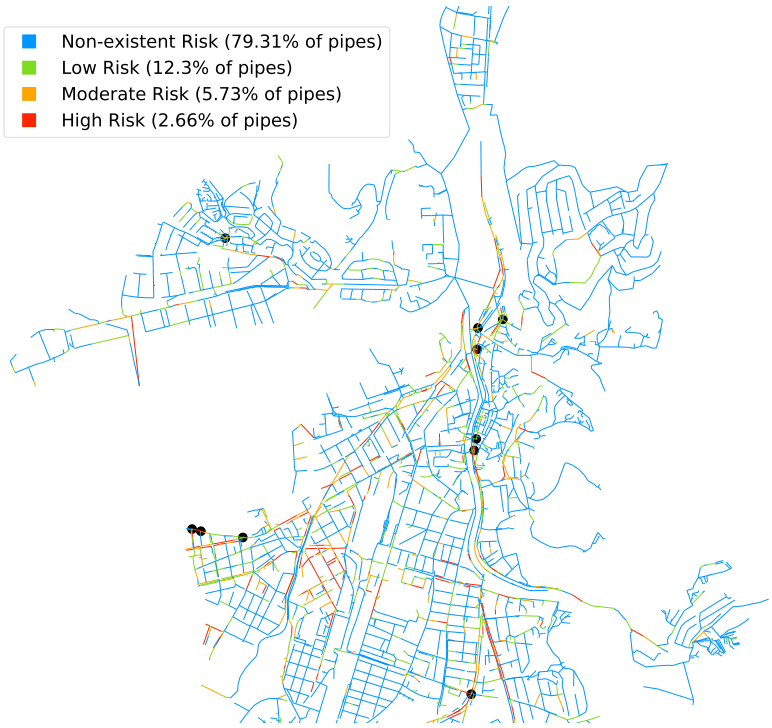


Fig. 7. Pipe risk and top 10 most impactful trees (black dots) in part of the wastewater network of Girona.

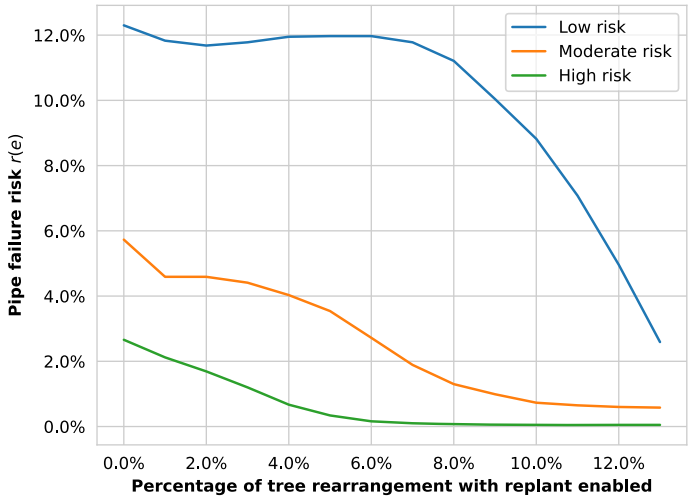


Fig. 8. Element Rearrangement (ER) algorithm pipe failure risk minimization grouped by risk categories (Girona tree dataset as a risk element, Australian Standard method, with replant enabled).

Fig. 9 shows the aggregated pipe failure risk $r(e)$ of the previously categorized risks through the application of the ER algorithm (Australian Standard method, with replant enabled). The risk median is reduced steadily by 33% (4% ER), 62% (8% ER), and 82% (12% ER) from the original tree dataset. For higher ER percentages, the reduction of pipe failure risks tends to stabilize.

The ER algorithm results show a clear benefit of the rearrangement of the city's most impactful trees in terms of pipe failure risks, thus preventing environmental and public health hazards by avoiding wastewater leakages caused by tree roots.

For future research, it would be interesting to consider an automatic feature in the algorithms that would detect if a replanted "sewer-safe" tree does not reduce the risk significantly compared to the high-risk tree it replaced. In this case, a warning should appear along with a recommendation not to plant a tree in that location.

3.4. Pipe failure probabilities and expected cost of repairs

The probabilities of pipe failures caused by tree roots have been calculated for the case study for both the original tree dataset and the rearranged tree scenarios. Based on the proposal of the pipe risk qualitative assessment shown in Table 3, a value of $R_{fail} = 5$ has been proposed and introduced in the case study. This decision was taken considering that the risk value of $r(e) > 3$ is high and

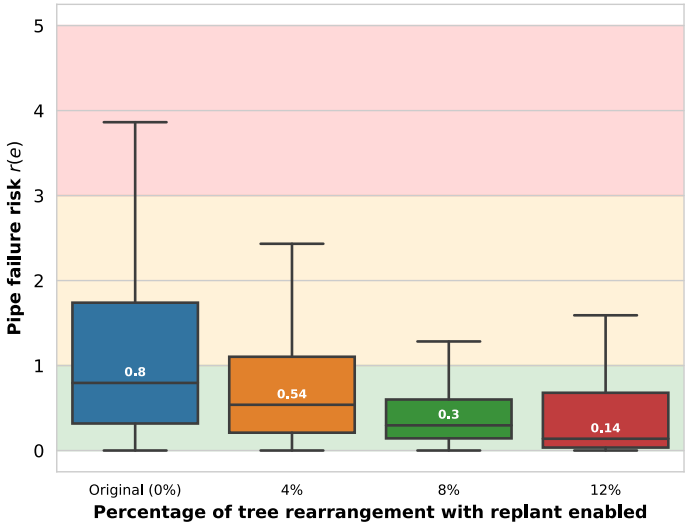


Fig. 9. Aggregated pipe failure risk minimization (Element Rearrangement (ER) algorithm, Australian Standard method, with replant enabled, without outliers).

Table 5
Expected cost of pipe repairs, costs of the Element Rearrangement (ER), and payback from 1 to 13 % of ER.

ER (%)	ECoR (€)	CoER (€)	P (€)	ER (%)	ECoR (€)	CoER (€)	P (€)
0	5.14 m	– –	– –	7	1.46 m	2.35 m	1.33 m
1	4.46 m	336k	346k	8	1.15 m	2.68 m	1.31 m
2	4.03 m	671k	439k	9	878k	3.02 m	1.24 m
3	3.39 m	1.01 m	742k	10	680k	3.35 m	1.10 m
4	2.78 m	1.34 m	1.01 m	11	537k	3.69 m	910k
5	2.29 m	1.68 m	1.17 m	12	465k	4.02 m	646k
6	1.84 m	2.01 m	1.29 m	13	446k	4.36 m	330k

ER (%) – percentage of Element Rearrangement (ER), ECoR (€) – expected cost of repairs, CoER (€) – cost of ER, P (€) – payback (i.e., economic savings of ER).

the interval of the medium risk qualitative category is two units (i.e., from 1 to 3). Therefore, a risk of $r(e) >= 5$ may be considered extreme and is only present in a few data outliers in this case study. An R_{fail} of the maximum value of $r(e), e \in \mathcal{E}$ is a poor approach as a few extreme data outliers $r(e) > 10$ would significantly affect the whole sample and consider extremely low probabilities on low- and medium-risk pipes.

The expected cost of the pipe failure repairs caused by tree roots has been calculated based on the original scenario, and the cost of pipe repairs of €230 per meter ($R = 230$). According to ABM Consulting, this estimation is valid in Spanish case studies when considering the material to be new 300 mm diameter PVC pipes. The expected cost resulted in about €5.14 m, which is reasonable as the repair works are considered during the entire operating lifetime of the pipes (approx. 30 years), as mentioned in the methods section.

3.5. Economic savings of the Element Rearrangement (ER) algorithm

The actual economic savings of the Element Rearrangement (ER) approach with tree replant enabled are calculated based on the expected reduction in repair costs from the original scenario combined with an estimation of the ER costs (i.e., payback based on the savings from avoiding pipe failures caused by tree roots during the operational period of the pipes). An ER cost of €1,020 is estimated for each tree in Spain based on the following quotes from several local companies: (i) €350 for big tree removals; (ii) €200 to deposit 5 tonnes wood; (iii) €110 for two-hour rental of a dump truck with loading crane; (iv) €340 for removing the tree stump; and (v) €20 for the cost of a new tree. Table 5 summarizes the expected cost of pipe repairs, costs of the ER, and payback from 1 to 13% of ER. The payback increases within the first 7% of ER up to €1.33 m, although it starts decreasing from 8 to 13% of ER as shown in Fig. 10. After this point, the ER costs start to cause economic losses.

The line chart in Fig. 10 visually illustrates the economic savings presented in Table 5. The dashed black line represents the expected costs without rearrangement (i.e., 0% ER), contrasted with the red line representing the expected costs of pipe repairs and ER. The expected cost of pipe repairs and ER red line forms a quadratic function shaping a parabola with the vertex being the maximum payback of €1.33 m for 7% of ER. Despite the initial investment required, the ER algorithm not only prevents environmental and public health hazards by avoiding wastewater leakages caused by tree roots, but also demonstrates a significant economic payback during the pipes' operation lifetime. In the cases where it is prioritized the minimization of pipe risks instead of the maximization of the economic payback, an ER of 14% should be considered.

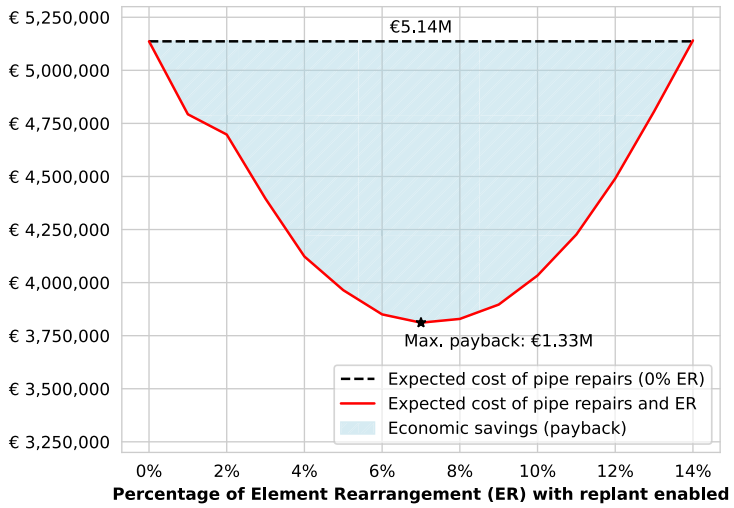


Fig. 10. Economic savings of the Element Rearrangement (ER) approach.

It would be interesting to extend these results with more case studies once more cities provide open data with detailed information about tree localization and trunk sizes. Fortunately, cities trend to issue increasingly more open data [39] as novel cost-effective methods are emerging to establish city-wide tree inventories [40].

3.6. Final thoughts

It is worth noting that the approach in this paper provides innovative methods that can be applied to any city in the world, as the only information needed is wastewater network and risk element data, including their location. The methodology and algorithms can be extrapolated to other water and underground (e.g., telecommunication or electricity) networks. Furthermore, the Impact Area (IA) concept can be applied to not only trees but also any other risk elements, such as street works or building constructions, in order to quantify their impacts and help create secure wastewater and other underground networks. All algorithm definitions, implementations, and output indicator results for the case study, including numerical, on-map, and graphical data visualizations, are available from a dedicated public repository [38].

The algorithms and methods described in this paper provide a simple and cost-effective approach to diagnose the impact of external elements (e.g., natural factors and human activities) on wastewater networks. Our approach uses the Impact Areas, but not actual impact volumes. We believe, that using Impact Areas makes it simple to understand by decision makers. Adding impact volumes implies adding sources of uncertainty, as limited knowledge exists on the volumes of tree roots of different ages and species. Future work will be dedicated to enhance the algorithm by including impact volumes.

4. Conclusions

This paper demonstrates that it is possible to perform an automatic diagnosis of potential impacts of tree roots on wastewater pipes, and to propose cost-effective rearrangement options. The Element Rearrangement (ER) algorithm not only prevents environmental and public health hazards, but also obtains a positive economic payback during the operational period of the pipes within the optimal rearrangement percentages. The proposed novel algorithms could also be applied to other natural and human factors. Furthermore, pipe failure probabilities are calculated and used to estimate the expected cost of pipe repairs during their operational period.

For the case study of Girona, the Australian Standard Tree Protection Zone (TPZ) method is the most practical approach to calculate the Impact Areas (IA) of the city trees, showing a tree impact median of 0.51 and pipe risk median of 0.84. The top 10 most impactful trees cause the majority of pipe damage, given the significant difference (44%) within the values for the first and tenth-placed trees. The Element Rearrangement (ER) algorithm reduced the pipe failure risk median considerably (from 0.8 to 0.14) with a small percentage of ER (from 1 to 12% of the trees). Based on the computed pipe failure probabilities, the expected cost of the pipe repairs caused by tree roots is about €5.14 m during the operational period, which is reduced to almost half (€2.78 m) with only a 4% ER. Finally, the economic savings of the ER algorithm show a payback of up to €1.33 m for a 7% rearrangement despite the required initial inversion.

This study illustrates a cost-effective approach for evaluating the influence of external factors on wastewater networks and pipe failure risks, all without the need for fieldwork in sewers. Despite some limitations, the method's utility lies in its global applicability using existing data in municipalities. It serves as a valuable preliminary study for prioritizing preventive measures and providing a detailed initial assessment, making it particularly useful for city councils and wastewater network managers.

Ethics statement

Review and/or approval by an ethics committee and informed consent were not needed for this study because it solely relies on publicly available data from open and unrestricted sources. The data used in this research contains no personally identifiable information and does not involve human participants or patients. Therefore, all aspects of this study adhere to ethical standards pertaining to the use of publicly accessible data without requiring formal ethical review or informed consent.

CRediT authorship contribution statement

David Martínez: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sergi Bergillos:** Writing – review & editing, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Lluís Corominas:** Writing – review & editing, Visualization, Resources. **Marc Comas-Cufí:** Supervision, Conceptualization. **Eusebi Calle:** Writing – review & editing, Visualization, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data, algorithms, and code implementations that support the findings of this study are openly available in "Mitigating Wastewater Leakages for Enhanced Network Protection: Risk Assessment and Practical Solutions." at https://doi.org/10.5281/zenodo.7704666, reference [38].

Acknowledgements

Universitat de Girona researchers thank the Generalitat de Catalunya for their support through Consolidated Research Group (2021 SGR 01125). ICRA researchers acknowledge the support from the Economy and Knowledge Department of the Catalan Government through a Consolidated Research Group (ICRA-TECH - 2021 SGR 01283). David Martínez thanks the Universitat de Girona for his FI fellowship (IFUdG 46 2022). Sergi Bergillos thanks the Departament de Recerca i Universitats de la Generalitat de Catalunya and the European Social Fund for his FI fellowship (2023 FI-1 00751). Marc Comas-Cufí holds the Serra Húnter Fellow. The authors would like to thank ABM Consulting for providing support regarding the pipe repair economic data, Cicle de l'Aigua del Ter for providing the wastewater network, and the Girona city council for providing the detailed tree dataset on its Open Data portal. We acknowledge Lide Jaurrieta for her contribution to the QGIS maps.

References

[1] E. Calle, D. Martínez, G. Buttiglieri, L. Corominas, M. Farreras, J. Saló-Grau, P. Vilà, J. Pueyo-Ros, J. Comas, Optimal design of water reuse networks in cities through decision support tool development and testing, npj Clean Water 6 (1) (2023) 23.
[2] J. Rice, A. Wutich, D.D. White, P. Westerhoff, Comparing actual de facto wastewater reuse and its public acceptability: a three city case study, Sustain. Cities Soc. 27 (2016) 467–474.
[3] E. Luiijf, M. Klaver, Analysis and lessons identified on critical infrastructures and dependencies from an empirical data set, Int. J. CIP 35 (2021) 100471.
[4] L.J. Van Leuven, Water/wastewater infrastructure security: threats and vulnerabilities, in: Handbook of Water and Wastewater Systems Protection, 2011, pp. 27–46.
[5] B. Salman, O. Salem, Risk assessment of wastewater collection lines using failure models and criticality ratings, J. Pipeline Syst. Eng. Pract. 3 (3) (2012) 68–76.
[6] T. Laakso, S. Ahopelto, T. Lampola, T. Kokkonen, R. Vahala, Estimating water and wastewater pipe failure consequences and the most detrimental failure modes, Water Sci. Technol.: Water Supply 18 (3) (2018) 901–909.
[7] A. Nazemi, K. Madani, Urban water security: emerging discussion and remaining challenges, Sustain. Cities Soc. 41 (2018) 925–928.
[8] S. Christodoulou, A. Agathokleous, B. Charalambous, A. Adamou, Proactive risk-based integrity assessment of water distribution networks, Water Resour. Manag. 24 (13) (2010) 3715–3730.
[9] D. Obradović, The impact of tree root systems on wastewater pipes, in: Zajednički Temelji'17: Zbornik Radova, pp. 65–71, 2017.
[10] S. Kumar, R. Singh, N.S. Maurya, Modelling of corrosion rate in the drinking water distribution network using design expert 13 software, Environ. Sci. Pollut. Res. 30 (15) (2023) 45428–45444.
[11] Sydney Water, How you can help stop blockages: tree planting and the wastewater system, Sydney Water (2017).
[12] M. Hartley, Tree root damage to pipes, Arborist Netw. (2012).
[13] Macedon Ranges Shire Council, Tree protection guidelines for developments, https://www.mrsc.vic.gov.au/files/assets/public/live-amp-work/environment/tree-protection-guidelines.pdf, 2022.
[14] AS 4970-2009, Protection of Trees on Development Sites, Standard, Standards Australia, 2009.
[15] G. Moore, Tree protection, Australian standards and the law: getting it right, in: 19th National Treenet Symposium, 2018.
[16] B. Ward, W.K. Clatterbuck, Choosing "Sewer Safer" Trees?, UT Extension, 2005.
[17] J. Östberg, M. Martinsson, Ö. Stål, A.-M. Fransson, Risk of root intrusion by tree and shrub species into sewer pipes in Swedish urban areas, Urban For. Urban Greening 11 (1) (2012) 65–71.
[18] E. Kuliczkowska, A. Parka, Management of risk of tree and shrub root intrusion into sewers, Urban For. Urban Greening 21 (2017) 1–10.

[19] T. Dawood, E. Elwakil, H.M. Novoa, J.F. Gárate Delgado, Water pipe failure prediction and risk models: state-of-the-art review, Can. J. Civ. Eng. 47 (10) (2020) 1117–1127.
[20] Y. Amiri-Ardakani, M. Najafzadeh, Pipe break rate assessment while considering physical and operational factors: a methodology based on global positioning system and data-driven techniques, Water Resour. Manag. 35 (11) (2021) 3703–3720.
[21] Y. Li, Z. Ye, Y. Yu, Y. Li, J. Jiang, L. Wang, G. Wang, H. Zhang, N. Li, X. Xie, et al., A combined method for human health risk area identification of heavy metals in urban environments, J. Hazard. Mater. 449 (2023) 131067.
[22] A. Vishwakarma, S. Sinha, Consequence of failure modeling for water pipeline infrastructure using a hierarchical ensemble fuzzy inference system, J. Infrastruct. Syst. 29 (1) (2023) 04022040.
[23] X. Xie, D. Guo, Human factors risk assessment and management: process safety in engineering, Process Saf. Environ. Prot. 113 (2018) 467–482.
[24] I. Estacio, R. Hadfi, A. Blanco, T. Ito, J. Babaan, Optimization of tree positioning to maximize walking in urban outdoor spaces: a modeling and simulation framework, Sustain. Cities Soc. 86 (2022) 104105.
[25] H. Kesavan, M. Chandrashekar, Graph-theoretic models for pipe network analysis, J. Hydraul. Div. 98 (2) (1972) 345–364.
[26] R. Ahmadullah, K. Dongshik, Designing of hydraulically balanced water distribution network based on gis and epanet, Int. J. Adv. Comput. Sci. Appl. 7 (2) (2016).
[27] E. Calle, D. Martínez, R. Brugués-i Pujolràs, M. Farreras, J. Saló-Grau, J. Pueyo-Ros, L. Corominas, Optimal selection of monitoring sites in cities for Sars-Cov-2 surveillance in sewage networks, Environ. Int. 157 (2021) 106768.
[28] F. Meng, G. Fu, R. Farmani, C. Sweetapple, D. Butler, Topological attributes of network resilience: a study in water distribution systems, Water Res. 143 (2018) 376–386.
[29] GraphML, The GraphML file format, http://graphml.graphdrawing.org/, 2001. (Accessed 21 December 2020), Online.
[30] S.D. Day, P.E. Wiseman, S.B. Dickinson, J.R. Harris, Contemporary concepts of root system architecture of urban trees, Arboric. Urban For. 36 (4) (2010) 149–159.
[31] R. Harris, J. Clark, N. Matheny, Arboriculture: Integrated management of landscape trees, shrubs, and Vines 4th ed prentice Hall upper saddle river, 2004.
[32] G. Boeing, Osmnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks, Comput. Environ. Urban Syst. 65 (2017) 126–139.
[33] F. Du, G.J. Woods, D. Kang, K.E. Lansey, R.G. Arnold, Life cycle analysis for water and wastewater pipe materials, J. Environ. Eng. 139 (5) (2013) 703–711.
[34] Statistical Institute of Catalonia, The municipality in figures: Girona (gironès), 2022. (Accessed 11 January 2023).
[35] J.L. Marzo, D. Martinez, S. Bergillos, E. Calle, Network research simulator. An abstract model formulation, in: 2022 18th International Conference on the Design of Reliable Communication Networks (DRCN), IEEE, 2022, pp. 1–4.
[36] Girona Open Data, Arbrat 2022: Localització i característiques de l'arbrat públic de la ciutat, 2023. (Accessed 17 January 2020). Published on 12.01.2023.
[37] G. vanRossum, Python Reference Manual, Department of Computer Science [CS], 1995.
[38] D. Martínez, Extended version of the algorithms and results of the research work "Mitigating wastewater leakages for enhanced network protection: risk assessment and practical solutions", https://doi.org/10.5281/zenodo.7704666, 2023.
[39] F.T. Neves, M. de Castro Neto, M. Aparicio, The impacts of open data initiatives on smart cities: a framework for evaluation and monitoring, Cities 106 (2020) 102860.
[40] D. Liu, Y. Jiang, R. Wang, Y. Lu, Establishing a citywide street tree inventory with street view images and computer vision techniques, Computers, Environment and Urban Systems 100 (2023) 101924.

5.4 Enhancing reclaimed water distribution network resilience with cost-effective meshing

Science of the Total Environment 938 (2024) 173051

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Enhancing reclaimed water distribution network resilience with cost-effective meshing

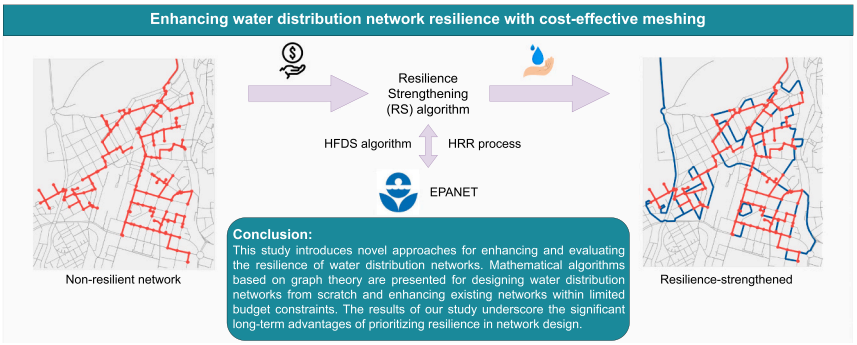
David Martínez [a,*], Sergi Bergillos [a], Lluís Corominas [b], Joaquim Comas [b,c], Fenghua Wang [d], Robert Kooij [d,e], Eusebi Calle [a]

[a] Institute of Informatics and Applications, University of Girona, 17003 Girona, Spain
[b] Catalan Institute for Water Research, Emili Grahit 101, 17003 Girona, Spain
[c] LEQUIA Institute of Environment, University of Girona, E-17071 Girona, Spain
[d] Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, the Netherlands
[e] Unit ICT, Strategy and Policy, Netherlands Organisation for Applied Scientific Research (TNO), 2595 DA Den Haag, the Netherlands

HIGHLIGHTS

- Novel algorithms for enhancing the resilience of water distribution networks
- Design of resilient and cost-effective networks in limited budget scenarios
- The Water Availability (WA) provides a comprehensive measure to evaluate resilience.
- Prioritizing resilience in network design demonstrated substantial long-term benefits.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Resilience
Water Availability
Planning
Graph theory
Costs
Hydraulic feasibility

ABSTRACT

Water Distribution Networks (WDNs) are critical infrastructures that ensure a continuous supply of safe water to homes. In the face of challenges, like water scarcity, establishing resilient networks is imperative, especially in regions vulnerable to water crises. This study evaluates the resilience of network designs through graph theory, including its hydraulic feasibility using EPANET software, an aspect often overlooked. Novel mathematical algorithms, including Resilience by Design (RbD) and Resilience-strengthening (RS) algorithms, provide cost-effective and resilient network designs, even with budget constraints. A novel metric, Water Availability (WA), is introduced to offer a comprehensive measure of network resilience, thereby addressing ongoing discrepancies in resilience evaluation methods. Practical benefits are illustrated through a case study in which a resilient-by-design reclaimed water network is created, and an existing equivalent non-resilient network is improved. The resilient-by-design network demonstrates remarkably better results compared to the equivalent non-resilient design, including up to a 36 % reduction in the probability of service disruptions and a nearly 65 % decrease in the annual average unserved water due to service disruptions. These findings underscore the

* Corresponding author.
E-mail addresses: david.martineza@udg.edu (D. Martínez), sergi.bergillos@udg.edu (S. Bergillos), lcorominas@icra.cat (L. Corominas), jcomas@icra.cat (J. Comas), f.wang-8@tudelft.nl (F. Wang), r.e.kooij@tudelft.nl (R. Kooij).

https://doi.org/10.1016/j.scitotenv.2024.173051
Received 5 October 2023; Received in revised form 29 March 2024; Accepted 6 May 2024
Available online 11 May 2024

enormous advantages of a resilience-focused network design approach. When compared to the equivalent non-resilient design, the resilient-by-design network generated effectively safeguards up to a significant 91,700 m^3 of water from the impacts of water disruption events over a 50-year operational period. In addition, the resilient-by-design WDN solution incurs a subtle decrease in overall costs compared to consuming tap water from the drinking WDN baseline over a 50-year operational period. These findings highlight the cost-effectiveness of the approach, even offering financial benefits. This paper builds on our previous research by expanding its scope to include resilience considerations, providing algorithms that can be easily adapted from reclaimed to drinking WDNs. Ultimately, we contribute to the enhancement of water resource management and infrastructure planning in ever-evolving urban environments.

1. Introduction

Water Distribution Networks (WDNs) are critical infrastructures of modern urban life, ensuring the seamless flow of safe water from reservoirs to homes. It is paramount for households and essential industries and public services to have a consistent water supply, particularly within the context of drinking water networks that safeguard public health and support societal functions (Liu and Song, 2020). However, in light of growing challenges such as water scarcity and climate change, it has become imperative to also plan reclaimed water networks for non-drinking purposes (Kristensen et al., 2018). Regions around the world, especially those most susceptible to water scarcity crises, must develop and prioritize their reclaimed water networks as critical infrastructures (Domènech and Saurí, 2010; Vallès-Casas et al., 2016). Only by effectively ensuring the resilience of these WDNs (both drinking and reclaimed water) can we ensure the proper long-term functioning of these vital systems that play such a pivotal role in sustaining urban life.

The repercussions of WDN failures have broader implications involving both economic and environmental consequences. In the absence of proactive measures aimed at enhancing resilient network designs, the likelihood of such failures increases significantly, amplifying the potential for devastating outcomes. These may include compromising public health, disrupting essential services, and incurring substantial economic costs (Ahmad et al., 2023).

The debate over the definition of resilience is a recurring aspect in this field (Soldi et al., 2015; Piratla et al., 2016; Herrera et al., 2016). In this paper, we adhere to the definition of resilience as 'the capacity of a system to withstand stress and recover from failures' (Soldi et al., 2015; Piratla et al., 2016; Herrera et al., 2016), thus forming the foundational concept with which to address this challenge. Resilience defies a single performance indicator, instead it comprises multiple dimensions, including structural robustness and adaptive recovery considerations influenced by factors such as network topology, failure rates, recovery rates, and severity (Meng et al., 2018; Lindhe et al., 2009). This interdisciplinary issue offers a range of approaches that consider different hazard categories, methodologies, and enhanced measures.

In the literature, diverse metrics and approaches have been explored to evaluate the resilience of WDNs (Christodoulou et al., 2017; Cimellaro et al., 2016; Zhao et al., 2015). While some works offer insights into minimizing network disruptions during large-scale cascade failures such as natural disasters (McAllister, 2015; Chang and Shinozuka, 2004), they focus predominantly on evaluating resilience rather than providing preventive design solutions. In particular, Cimellaro et al. (2016) introduced a Resilience Index (R) for WDNs which focuses on parameters like the number of users temporarily without water, water tank levels, and water quality. Similarly, other studies have explored resilience evaluation measures like the Resilience Index (RI) and Network Resilience Index (NRI) (Baños et al., 2011; Todini, 2000; Tumula and Park, 2004). It is important to note that while both R and RI are named equally, they represent distinct measures. RI, for instance, aims to assess network resilience by ensuring demand satisfaction. However, these metrics frequently rely on operational data which is rarely available during the initial design stages. Furthermore, while RI and NRI excel at over-demand resilience analysis, they do not consider other crucial causes of such as failures and fail to provide specific design improvement measures to enhance network resilience.

This discrepancy highlights the need for a more comprehensive approach. Recent research by Taiwo et al. (2023) systematically analyzed the causes leading to network failures, categorizing them into three main categories: pipe-related, environment-related, and operation-related (Zhang et al., 2009), identifying a total of 33 distinct causes, and also providing a detailed table of relative weights of the causes of water pipe failure. These causes encompass a wide spectrum of elements, such as pipe age or diameter, some of which have not received attention in previous metrics.

While assessing the resilience of a network is crucial, it is equally important that the proposed designs conform to hydraulic feasibility constraints, an aspect that has received less attention in prior works (Yazdani et al., 2011; Herrera et al., 2016). In this regard, the EPANET software plays a pivotal role (Rossman et al., 2000), offering hydraulic simulation capabilities with which to assess the validity and performance of Water Distribution Networks (WDNs). Recent advances in the field have introduced the Water Network Tool for Resilience (WNTR), an EPANET-compatible Python package designed to simulate and analyze the resilience of WDNs (Klise et al., 2017b, 2018). While WNTR provides valuable insights into resilience analysis, it primarily focuses on evaluating resilience rather than offering concrete design improvement measures (Klise et al., 2017a).

The sole work offering resilience improvements through design and hydraulically validated for cost-effectiveness is, to the best of our knowledge, presented in Todini (2000). However, it has several limitations that warrant consideration. First, their study focuses on designing networks from scratch, i.e., neglecting the potential for enhancing existing systems. Moreover, its data and algorithms are outdated and inaccessible to the public. Additionally, the approach lacks automation, relying on an initial set of fixed diameters determined by the designer's experience. The study itself acknowledges the need for further investigation and development to create efficient and easy-to use tools.

Although there is an extensive body of literature on this topic, it is evident that this specific field of study lacks comprehensive improvement measures or strategies that would ensure resilience by design. Given this context, automating the process of water distribution network design through mathematical algorithms becomes feasible to provide both resilient and cost-effective networks. The contributions of this paper are the following:

1. Water Availability (WA): A novel metric is introduced, the Water Availability (WA), which serves as a comprehensive measure for assessing the resilience of water distribution networks. This metric is based on the concept of Network Availability (NA), which is commonly employed in the placement of controllers within telecommunication networks (Lu et al., 2019; Hu et al., 2014; Gaur et al., 2021; Rosenthal, 1977). NA is defined as the probability that all nodes can reach at least one controller with an operational probability of each link, which can be estimated using Monte Carlo simulations or computed precisely via a brute force algorithm. In cases where a single controller is present in the network, NA aligns with all-terminal reliability which is the probability that the network is

connected. To compute the exact value of all-terminal reliability, the path decomposition algorithm is recommended, particularly for medium-sized networks, due to its lower computational complexity compared to the brute force approach (Carlier and Lucet, 1996).
2. Mathematical Algorithms for Resilient Networks Design and Improvement: Not only cutting-edge mathematical algorithms are introduced for designing water distribution networks from scratch, but also several strategies are offered to enhance existing networks within limited budget constraints, which will provide resilient and cost-effective networks.
3. Hydraulic Feasibility with EPANET: An innovative and interactive process that integrates the developed algorithms with EPANET software (U.S. EPA, 2000) ensures that the generated network designs not only prioritize cost-effective resilience, but also adhere to hydraulic feasibility constraints. Something that is hardly found in the reviewed literature.

The algorithms presented in this paper are based on applying graph theory (Kesavan and Chandrashekar, 1972) coupled with hydraulic validations to design the water distribution networks. Several previous works have used graph theory in water networks: Ahmadullah and Dongshik (2016) for designing drinking water networks; Calle et al. (2021) for wastewater sensor placement approaches concerning SARS-CoV-2 detection; and Meng et al. (2018) for proposing a comprehensive analytical framework for examining the resilience pattern of water distribution systems against topological characteristics (i.e., the correlations between resilience and topological features). The use of EPANET for hydraulic validations is also present in several previous studies (Soldi et al., 2015; Klise et al., 2017b; Todini, 2000).

2. Materials and methods

2.1. Generation of the initial graph

The REWATnet tool automates the generation of the initial graph, illustrating the paths of city streets for potential reclaimed water network designs. It achieves this by collecting data from diverse open sources. Subsequently, additional algorithms are then employed to develop resilient network designs based on this initial representation (Calle et al., 2023). The process of generating the initial graph involves 5 steps (see Table 1):

1. City Street Graph Acquisition: REWATnet utilizes OpenStreetMap to obtain the city street graph based on the city's name and source point (i.e., initial tank) coordinates. City's topography Digital Elevation Model (DEM) is gathered for node elevation required for hydraulic validations.
2. Land Plot and Building Data Retrieval: Cadastral data files are used to gather land plot and building data for the identification of water consumption destinations. Overpass API and Shapely library locate and determine the surface area of public gardens.
3. Construction Cost Retrieval: REWATnet utilizes its open database to provide available PE100 (HDPE) pipe diameters and associated costs for materials, labor, valves, and water tanks.
4. Water Use Consideration: A list of water uses is essential for estimating water demand. The full list of water uses considered for reclaimed water networks is available in Calle et al. (2023).
5. Graph Generation: REWATnet automatically generates the initial graph in standardized graphml format by processing land plot and building data and estimating their water demands. The estimated water demand for each land plot is then linked to the nearest node in the city street graph.

Table 1
Initial graph generation process.

Step	Data source	Description	References
1	OpenStreetMap API, IGN Digital Elevation Model (DEM)	Obtain city street graph and node elevations.	Bennett (2010); Boeing (2017); Instituto Geográfico Nacional (2023)
2	Cadastral data, Overpass API, Shapely	Gather land plot, building, and public garden data.	Pezoa et al. (2016); Olbricht et al. (2011); Gillies et al. (2007)
3	REWATnet database	Retrieve available pipe diameters and construction costs.	Christodoulou and Agathokleous (2012)
4	Water use list	Consider water uses for estimating water demand.	Calle et al. (2023)
5	REWATnet tool	Generate initial graph by processing land plot and building data and estimating their water demands.	Calle et al. (2023); GraphML (2001)

2.2. Resilience metrics and optimization algorithms

In brief, $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ represents the solution of a resilient water network graph resulting from the computation of the algorithms, with a V-element set of nodes $\mathscr{V}$ representing the set of destination (water consumption) nodes, the water source node, and junction points, and an E-element set of links $\mathscr{E} \subset \mathscr{V}^{|2|}$ representing pipes. Additionally, r (where $r \in \mathscr{V}$) denotes the source node (i.e., the initial water tank), and $\mathscr{C}$ (where $\mathscr{C} \subseteq \mathscr{V}$) denotes a C-element set of consumption nodes. Table 2 specifies the full notation used for the algorithms.

2.2.1. Water Availability (WA)

While the concept of Network Availability (NA) has traditionally found its application in telecommunication networks, it seamlessly extends to serve as an ideal approach for assessing resilience in water distribution networks. In the context of water distribution networks, we introduce Water Availability "WA($\mathscr{G}, \mathscr{P}, r$)" as the probability that the initial water tank r can effectively supply water to all the destination nodes $\mathscr{V}$ within the network $\mathscr{G}$ while considering variable pipe failure probabilities for each pipe in the network $p(e) \in \mathscr{P}, \forall e \in \mathscr{E}$.

As a reference point for the pipe failure probabilities $p(e)$, we utilize a well-established criterion of a maximum of 0.4 failures per kilometer per year (MIMAM, 2000). This criterion, coupled with the average duration of a failure, known as the Mean Time To Repair (MTTR) measured in hours, is used in deriving the unavailability per pipe kilometer denoted as q. This q value represents the probability of pipe failures per kilometer

Table 2
Full notation concerning the algorithms.

$r, r \in \mathscr{V}$	Reclaimed water source node
$\mathscr{C}$	Set of water distribution consumption nodes; $\mathscr{C} \subseteq \mathscr{V}$
$\mathscr{D}$	Set of available pipe diameters, in ascending order (each one in mm)
s_{max}	Float constant indicating the maximum desired water flow speed (in m/s, 1 by default)
s_{min}	Float constant indicating the minimum desired water flow speed (in m/s, 0.6 by default)
$pr(c), c \in \mathscr{C}$	Float indicating the water pressure of destination node c (in m)
$t(c), c \in \mathscr{C}$	Float indicating the water travel time from the origin r to the destination node c (in minutes)
$m(c), c \in \mathscr{C}$	Integer indicating the consumption of destination node c (volume, in m^3)
$\mathscr{E}(a,c), c \in \mathscr{C}$	Set of edges forming the shortest path from the source node a to the destination node c; $\mathscr{E}(a,c) \subseteq \mathscr{E}$
$l(e), e \in \mathscr{E}$	Float indicating the length of the pipe link e (in m)
$w(e), e \in \mathscr{E}$	Float indicating the water flow of the pipe link e (in m^3/s)
$s(e), e \in \mathscr{E}$	Float indicating the water speed of the pipe link e (in m/s)
$v(e), e \in \mathscr{E}$	Integer indicating the valve diameter for pipe link e (in mm); 0 by default (i.e., no valve installed on link e)
$d(e), e \in \mathscr{E}$	Integer indicating the assigned diameter of the pipe link e; $d(e) \in \mathscr{D}$ (in mm)

within the network, as delineated in Eq. (1). The choice of MTTR is crucial and may vary depending on the scenario and the criticality of the network. For drinking water networks, a recommended MTTR value is 19 h (Darvini et al., 2020). Conversely, for reclaimed water networks primarily used for non-critical purposes such as public garden irrigation, a higher MTTR may be considered. Considering the criticality of reclaimed water networks for toilet flushing, we recommend a default MTTR of 24 h, easily modifiable as an input in the algorithms to accommodate different case-study scenarios. Assuming an MTTR of 24 h, which means the average down time per kilometer of pipe is 9.6 h per year, yields an approximate q value of 0.0011.

$$q = \frac{0.4 \times \mathrm{MTTR}}{(24 \times 365)} \tag{1}$$

Then, the estimation for each pipe unavailability can be written as Eq. (2), where $u(e)$ denotes the unavailability of pipe e with its length $l(e)$ transformed to kilometers. This formula is common practice in telecommunication (Mezhoudi and Chu, 2006), with values between 0 and 1, increasing as a function of the pipe length. The expression $u(e)$ goes to 1 if $l(e)$ goes to infinity, and for a $l(e) = 1000$, we retrieve $u(e) = q$.

$$u(e) = 1 - (1 - q)^{l(e)/1000} \tag{2}$$

While $u(e)$ serves as a valuable reference point for evaluating the overall likelihood of each pipe failure, assuming an average repair time of 24 h (i.e., the average duration of the failure), it does not fully capture the individual variability of each pipe. This is because each pipe has its own unique set of related features contributing to its failure. In water distribution networks, pipe failures can be attributed to three primary causes, each with its own distinct probability of incidence (Taiwo et al., 2023): (i) pipe-related (probability w_{pr}: 0.396), (ii) environmental-related (probability w_{er}: 0.413), and (iii) operational-related (probability w_{or}: 0.191). As we are approaching resilience from a network design perspective, only pipe-related attributes can be directly quantified, including pipe diameter (d) with a probability of w_d at 0.122; age (a) with a probability of w_a at 0.105; material composition (m) with a probability of w_m at 0.076; pipe length (l) with a probability of w_l at 0.066; and wall thickness (t) with a probability of w_t at 0.027. The sum of all the pipe-related attribute probabilities is equal to $w_{pr} = 0.396$.

Thus, it becomes feasible to adapt and normalize the probability of pipe-related failure causes w_{pr} for each individual pipe $w'_{pr}(e)$ through a table of normalized attributes (see Table 3). This table maps potential attribute values (within specified ranges) to their corresponding normalized values, falling from 0 to 1. In this normalized range, a value of 0 indicates the lowest severity, while a value of 1 represents the highest severity. Subsequently, the established probability of pipe-related failure causes w_{pr} is adapted for each pipe individually in $w'_{pr}(e)$, which is the weighted sum of these normalized attributes, as defined in Eq. (3). As the pipe unavailability $u(e)$ is calculated based on an established maximum number of failures, the adapted $w'_{pr}(e)$ results in a value between 0 and 0.396 (i.e., from lowest to highest-vulnerable pipe-related attributes), where a value of $w'_{pr}(e) = 0.396$ would represent a pipe e with the most vulnerable pipe-related features according to the table of normalized attributes (i.e., all the normalized values of pipe-related attributes are 1, therefore $w'_{pr}(e) = w_{pr}$). Importantly, it should be noted that the pipe unavailability $u(e)$ already incorporates the length of each individual pipe; therefore, the length is set as $l = 1$ in the equation.

$$w'_{pr} = (d \times w_d) + (a \times w_a) + (m \times w_m) + (l \times w_l) + (t \times w_t) \tag{3}$$

Given this context, the failure probabilities $p(e)$ for each pipe $e \in \mathscr{E}$ are calculated by multiplying the reference point of the overall likelihood of each pipe failure, $u(e)$, by the sum of the failure weights $w'_{pr}(e)$, w_{er}, and w_{or} (see Eq. (4)).

Table 3
Proposed table of normalized attributes for pipe-related failure causes.

Attribute	Value intervals	Normalized value
Pipe diameter (d)	0 to 90 mm	0
	90 mm to 250 mm	0.33
	250 mm to 560 mm	0.67
	More than 560 mm	1
Reference: Wilson et al. (2017)		
Age (a)	0 to 33 years	0
	33 to 67 years	0.33
	67 to 100 years	0.67
	More than 100 years	1
Reference: Zangenehmadar et al. (2020)		
Material (m)	HDPE (PE 100)	0
	MDPE (black)	0.33
	MDPE (blue), GI, LDPE (black), AC	0.67
	UPVC, DI	1
Reference: Christodoulou and Agathokleous (2012)		
Wall thickness (t)	More than 33.2 mm	0
	14.8 to 33.2 mm	0.33
	3.8 to 14.8 mm	0.67
	0 to 3.8 mm	1
Reference: Wilson et al. (2017)		

$$p(e) = u(e) \times \left(w'_{pr}(e) + w_{er} + w_{or}\right) \tag{4}$$

The annual volume of water affected by the pipe failures is also a crucial metric for assessing resilience. Initially, the volume of Average Unserved Water per service Disruption event (AUW/D) is computed by examining Water Availability (WA) during the Monte Carlo realizations. AUW/D represents the average volume of unserved water in each Monte Carlo realization that results in a water service disruption. Next, utilizing the computed WA value, a Mean Time To Repair (MTTR), and the Mean Time Between Disruptions (MTBD), the number of Average Disruptions per Year (AD/Y) is calculated, as demonstrated in Eq. (5), where the Mean Time To Repair (MTTR) is 0.002739726 (1 day converted to years). Finally, the volume of Average Unserved Water per Year (AUW/Y) is determined by multiplying AD/Y by AUW/D, as illustrated in Eq. (6). This comprehensive measure provides valuable insights into the impact of disruptions on the annual water supply.

$$\mathrm{AD/Y} = \frac{1}{\mathrm{MTBD} + \mathrm{MTTR}}; \text{where } \mathrm{MTBD} = \frac{-\mathrm{WA} \times \mathrm{MTTR}}{\mathrm{WA} - 1} \tag{5}$$

$$\mathrm{AUW/Y} = \mathrm{AD/Y} \times \mathrm{AUW/D} \tag{6}$$

It is important to highlight that the formulation presented in Eq. (5) draws inspiration from a standard telecommunication Availability (A) formula, which is typically expressed as a function of the Mean Time to Repair (MTTR) and the Mean Time Between Failures (MTBF), as detailed in Calle (2004). In the context of this paper, this formula has been adapted to suit the specific needs of water networks. Here, the MTBF corresponds to what it is referred to as the Mean Time Between Disruptions (MTBD), as the calculations focus on service disruptions when computing Water Availability (WA).

2.2.2. Hydraulic-feasible diameter selection

To create resilient water networks, an algorithm is needed to efficiently select the optimal diameter for each pipe while simultaneously ensuring the hydraulic feasibility of the design. The Hydraulic-Feasible Diameter Selection (HFDS) algorithm (Algorithm 1) aims to initially predict water flows within pipes, considering destination demands and the pipe network design. Subsequently, it determines the optimal pipe diameters from a predefined set of available options (Calle et al., 2023), ensuring a suitable water speed according to the predicted flows.

The first phase of predicting water flows $w(e), e \in \mathscr{E}$ starts with the assumption that all water pipes in the network have similar speeds. As a result, the water flows from the initial tank to the farthest node while traversing the network and passing through nodes based on their respective distances. In accordance with this premise, the Breadth-First Search (BFS) exploration graph theory algorithm generates an ordered list of water destinations reflecting the sequence of the water's journey and the directional flow within the pipelines, with minor adjustments to prioritize same-level neighbor water destinations based on increasing pipe distances. Afterwards, the HFDS algorithm traverses this list in reverse order. For each evaluated water destination, it predicts the flow within its connected pipes based on the accumulated demand and the various sources from which water comes.

The second phase in determining pipe diameters involves iterating through each water pipe $e, e \in \mathscr{E}$ from the list of available pipe diameters $d \in \mathscr{D}$, arranged in ascending order, until an appropriate flow speed $s(e)$ is achieved based on the predicted pipe flow $w(e)$. The algorithm defines an acceptable speed range, denoted as $s_{min} \leq s \leq s_{max}$, typically falling between 0.6 and 1 m/s (Simpson and Elhay, 2008; MIMAM, 2000). Speeds lower than 0.6 m/s may lead to pipe sediment accumulation issues, while speeds higher than 1 m/s can cause vibrations. Importantly, this range remains consistent across various water qualities, making it applicable to both reclaimed and drinking water networks. When the algorithm identifies an evaluated diameter d resulting in an excessively low speed $s < s_{min}$, it selects the previous, smaller diameter, provided the prior velocity s_p is below the maximum threshold $s_p \leq s_{max}$. Otherwise, although the current diameter results in a low speed, it is chosen. It is crucial to avoid selecting diameters that result in speeds exceeding the maximum limit $s > s_{max}$, as this can lead to pipe failures due to factors such as vibration (Rezaei et al., 2015).

Algorithm 1. Hydraulic-feasible diameter selection (HFDS).

Step 1: Given a water network design $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, initialize the node r and sets $\mathcal{C}$; $\mathcal{D}$; m; $\mathcal{E}$. Initialize s_{min} and s_{max} to the desirable values. Create a new empty set of visited nodes $\mathcal{W} := \varnothing$.

Step 2: Create a new variable $f(v), v \in \mathcal{V}$, which is a float indicating the accumulated consumption of each node in the graph, and initialize to 0 $f(v) := 0, \forall v \in \mathcal{V}$.

Step 3: Order the nodes $c \in C$ by performing a reverse Breadth-First Search (BFS) traversal of graph G starting from node r while considering the weights of node neighbors based on pipe distances.

Step 4: For each consumption node $c \in C$:

(a) set a new variable that indicates the number of sources of node c, such that $o := 0$.

(b) add node c to visited nodes, such that $\mathcal{W} := \mathcal{W} \bigcup \{c\}$.

(c) for each neighbor n of node c:

(i) if $n \notin \mathcal{W}$, then set $o := o + 1$.

(d) for each neighbor n of node c if $n \notin \mathcal{W}$:

(i) set $f(n) := f(n) + ((m(c) + f(c))/o)$

(ii) set current diameter as the smallest diameter in the list $d(\mathcal{E}(c, n)) := \mathcal{D}[0]$.

(iii) set a counter $i := 0$ and previous speed $s_p := Inf$

(iv) for every diameter $d \in \mathcal{D}$:

- obtain the speed s of $\mathcal{E}(c, n)$ considering flow $((m(c) + f(c))/o)$ and diameter d.
- if $s \leq s_{min}$ and $s_p \leq s_{max}$, then $d(\mathcal{E}(c, n)) := \mathcal{D}[i - 1]$
- otherwise if $s \leq s_{min}$, then $d(\mathcal{E}(c, n)) := \mathcal{D}[i]$
- set $s_p := s$ and $i := i + 1$

Step 5: $d(e)$ contains the assigned pipe diameter $\forall e \in \mathcal{E}$.

The HFDS algorithm is integrated into an interactive process with the EPANET software to perform hydraulic validation on the generated network designs, called Hydraulic Refinement Relay (HRR). For each design requiring hydraulic validation, the HFDS algorithm is initially executed to predict water flows and determine appropriate pipe diameters. Subsequently, the network design is automatically converted into an EPANET-compatible INP file and processed. EPANET then generates a result file, from which key output indicators are extracted. These indicators include whether the water supply meets the demands of all nodes, if node pressures $pr(c), c \in \mathscr{C}$ fall within the acceptable range (i.e., between 15 and 60 m (Desta et al., 2022; MoWR, 2006)), and if all water travel times $t(c), c \in \mathscr{C}$ adhere to a maximum residence time. A maximum residence time value of 72 h was determined based on a thorough survey encompassing more than 800 utilities across the USA. This duration is widely acknowledged as the accepted maximum, though actual residence times within a system may diverge considerably due to variations in design and usage patterns (World Health Organization, 2014). While no other references specifically address water reuse systems, a value of 72 h is recommended as a precautionary principle. However, it can be easily adapted as an input value depending on the specific case-study scenario. If any of these indicators fail, highlighting a potential issue with the network design, the HRR process takes specific corrective actions. Depending on which indicator has failed, adjustments are made to the acceptable range of flow speeds, denoted as an interval $[s_{min}, s_{max}]$. The following adjustments are made automatically:

1. If any speed within the network $s(e), e \in \mathscr{E}$ exceeds s_{max}, the HRR slightly reduces s_{max}.
2. If the water supply fails to meet the demands of all consumption nodes $m(c), c \in \mathscr{C}$, the HRR slightly reduces s_{min} while maintaining s_{max}.
3. If node pressures $pr(c), c \in \mathscr{C}$ fall outside the acceptable range, adjustments are made by slightly increasing s_{max} or reducing s_{min}.
4. If any water travel time $t(c), c \in \mathscr{C}$ exceed the required criteria, the HRR adjusts both s_{min} and s_{max} as necessary.

The modified design is then subjected to another round of validation through the EPANET software. The HRR process continues until a feasible hydraulic solution is obtained. However, if such a solution cannot be achieved due to the speeds $s(e), e \in \mathscr{E}$ not falling within absolute interval constraints (i.e., where the minimum speed s_{min} should not be lower than 0.4 m/s, and the maximum speed s_{max} should not exceed 1.2 m/s), the process is terminated because the hydraulic feasibility is unachievable for this design. This outcome may be attributed to factors such as an excessively large network or an insufficient initial tank elevation.

2.2.3. Resilience by design

Designing a resilient water distribution network from its inception can be an arduous challenge, particularly when faced with tight budget constraints. The Resilience by Design (RbD) algorithm offers an effective and automatic solution. It operates as a greedy algorithm, crafting network designs that enhance resilience within the confines of budget limitations. This approach takes advantage of both the HFDS algorithm and the HRR process to guarantee hydraulic feasibility throughout the design process.

The Resilience by Design (RbD) algorithm (Algorithm 2) starts with an initial setup that includes a starting point capable of gravity-based water distribution (i.e., an elevated initial water tank r without requiring water pumps) and a set of locations requiring water ($\mathscr{C}$). The operation of the RbD algorithm is illustrated as:

1. Building the Resilient Network: The RbD algorithm goes through a step-by-step process to construct a resilient network based on the initial graph (see Section 2.1). It evaluates each destination where water is needed, one at a time, until the budget is exhausted.
2. Ensuring Resilience: One of the key goals is to ensure resilience in the network. This means that every consumption point should have at least two independent paths to receive water (i.e., a meshed pipe network design). To achieve this, whenever a path to a consumption point is added to the network, the algorithm also looks for an alternative path to the same point.
3. Optimizing for Cost-Effectiveness: In each iteration, the algorithm selects the candidate destination that provides the most cost-effective solution. This is determined by looking at the water

consumption at each node along the path and considering the length of the path itself.

Algorithm 2. Resilience by Design (RbD) algorithm.

Step 1: Initialize the initial graph $\mathcal{I}(\mathcal{V},\mathcal{E})$, node r, the budget B, sets $\mathcal{C}$; $\mathcal{D}$; m, and current expenses $Z := 0$.
Step 2: Let $\mathcal{T}(\mathcal{V}',\mathcal{E}')$ be the resilient network design graph initialized with $\mathcal{V}' := \{r\}$ and $\mathcal{E}' := \varnothing$.
Step 3: Create an empty prioritized list $\mathcal{C}' := \varnothing$ of consumption node candidates $c \in \mathcal{C}$ to be added to $\mathcal{T}$.
Step 4: For each consumption node $c : c \in \mathcal{C}, c \notin \mathcal{V}'$:
(a) get the node $a \in \mathcal{V}'$ that minimizes the path to join $\mathcal{T}$ with c, such that:
$$\sum l(e), e \in \mathcal{E}(a,c) := \min((\sum l(e), e \in \mathcal{E}(v',c)), v' \in \mathcal{V}')$$
(b) if Z+min_cost$(\sum l(e), e \in \mathcal{E}(a,c)) \leq B$, then:
(i) compute profit $P := \dfrac{\sum m(v), (v,\_) \in \mathcal{E}(a,c)}{\sum l(e) \times 2, e \in \mathcal{E}(a,c)}$.
(ii) add node c to $\mathcal{C}'$ such that $\mathcal{C}' := \mathcal{C}' \bigcup \{c\}$, prioritized by profit P. Also store its a and $\mathcal{E}(a,c)$
Step 5: For each consumption node candidate $c : c \in \mathcal{C}'$ and its related node a and path $\mathcal{E}(a,c)$ ordered by descending profit:
(a) remove every $e \in \mathcal{E}(a,c)$ if it does not disconnect the initial graph $\mathcal{I}$.
(b) find an alternative shortest path $\mathcal{S} := \mathcal{E}(a,c)$ from c to a following the equation of Step 5(a).
(c) reinstate removed $e \in \mathcal{E}(a,c)$ edges to graph $\mathcal{I}$ from Step 6(a).
(d) add the two paths to the resilient network graph $\mathcal{T}$, such that $\mathcal{E}' := \mathcal{E}' \bigcup \mathcal{E}(a,c) \bigcup \mathcal{S}$.
(e) compute HFDS algorithm and HRR process to get pipe diameters $d(e), e \in \mathcal{E}'$ of network $\mathcal{T}$.
(f) if Z+cost$(\mathcal{T}) \leq P$, then:
(i) for each node $(v,\_)$ in $\mathcal{E}(a,c) \bigcup \mathcal{S}$, if $v \in \mathcal{C}$ then $\mathcal{C} := \mathcal{C} \setminus \{v\}$
(ii) set $Z := Z$+cost$(\mathcal{T})$.
(iii) go to Step 3.
Step 6: $\mathcal{T}$ represents the final water distribution network design $\mathcal{G}$.

In order to benefit from the resilient meshed design, the RbD algorithm places valves $v(e)$ on each pipe $e \in \mathscr{E}$ connected to nodes with more than one downstream branch. Valve placement is crucial since a failure on an individual non-valve pipe e can result in the service disruption of an entire network section until its flow is successfully isolated both upstream and downstream.

2.2.4. Resilience strengthening

In many cases, existing water distribution networks (WDNs) have already been established and are currently in operation but may lack the necessary level of resilience. If not addressed promptly, inadequately resilient networks can lead to substantial economic losses and service disruptions. In such scenarios, it becomes crucial to initiate projects aimed at enhancing the network's resilience. The Resilience Strengthening (RS) algorithm offers an effective solution to this challenge. RS is a greedy algorithm, and its primary purpose is to enhance the resilience of pre-existing network designs while working within predefined budget constraints. Similar to the Resilience by Design (RbD) algorithm (see Section 2.2.3), the RS takes advantage of both the HFDS algorithm and the HRR process to guarantee hydraulic feasibility. In this scenario, the HFDS algorithm is applied to pre-existing water distribution networks, where the original pipe diameters are fixed and cannot be altered. Adapted to the context of RS, the HFDS algorithm focuses on predicting flows and determining diameters only for the newly added pipes, while leaving the existing network layout unchanged.

The RS (Algorithm 3) starts with a pre-existing network design and its operation can be outlined as follows:

1. Enhancing resilience: The RS algorithm introduces additional pipes and valves to enhance resilience in the network. This means, in a similar way to the RbD algorithm, that every consumption point should have at least two distinct paths to receive water. To achieve this, the algorithm searches for alternative routes to the same destination based on the initial graph (see Section 2.1).
2. Optimizing for Cost-Effectiveness: In each iteration, the algorithm selects the candidate that provides the most cost-effective solution. This selection process considers the volume of water delivered to each node via at least two distinct paths, factoring in the additional length of the newly introduced pipes.

Algorithm 3. Resilience Strengthening (RS) algorithm.

Step 1: Initialize the initial graph $\mathcal{I}(\mathcal{V},\mathcal{E})$, node r, the budget B, sets $\mathcal{C}$; $\mathcal{D}$; m, and current expenses $Z := 0$.
Step 2: Let $\mathcal{T}(\mathcal{V}',\mathcal{E}')$ be the pre-existing network design graph.
Step 3: Create an empty prioritized list $\mathcal{C}' := \varnothing$ of consumption node candidates $c \in \mathcal{C}, \forall c \in \mathcal{V}'$ to be evaluated.
Step 4: For each consumption node $c : c \in \mathcal{C}$:
(a) remove every $e \in \mathcal{E}(r,c)$ if it does not disconnect the initial graph $\mathcal{I}$.
(b) find an alternative shortest path $\mathcal{S} := \mathcal{E}(r,c)$ from r to c, such that:
$$\sum l(e), e \in \mathcal{E}(r,c) := \min((\sum l(e), e \in \mathcal{E}(v',c)), v' \in \mathcal{V}')$$
(c) reinstate removed $e \in \mathcal{E}(r,c)$ edges to graph $\mathcal{I}$ from Step 4(a).
(d) if Z+min_cost$(\sum l(e), e \in \mathcal{S} \notin \mathcal{E}(r,c)) \leq B$, then:
(i) compute profit $P := \dfrac{\sum m(v), (v,\_) \in \mathcal{E}(r,c) \bigcup \mathcal{S}}{\sum l(e), e \in \mathcal{S} \notin \mathcal{E}(r,c)}$.
(ii) add node c to $\mathcal{C}'$ such that $\mathcal{C}' := \mathcal{C}' \bigcup \{c\}$, prioritized by profit P. Also store its $\mathcal{S}$.
Step 5: For each consumption node candidate $c : c \in \mathcal{C}'$ and its alternative path $\mathcal{S}$ ordered by descending profit:
(a) add every $e \in \mathcal{S} \notin \mathcal{E}(r,c)$ to $\mathcal{T}$.
(b) if Z+cost$(\mathcal{T}) \leq P$, then:
(i) for each node $(v,\_)$ in $\mathcal{E}(r,c) \bigcup \mathcal{S}$, if $v \in \mathcal{C}$ then $\mathcal{C} := \mathcal{C} \setminus \{v\}$
(ii) set $Z := Z$+cost$(\mathcal{T})$.
(iii) go to Step 3.
(c) otherwise remove every $e \in \mathcal{S} \notin \mathcal{E}(r,c)$ from $\mathcal{T}$.
Step 6: $\mathcal{T}$ represents the final water distribution network design $\mathcal{G}$.

3. Results and discussion

3.1. Case study

The usefulness of the methods and algorithms presented in this paper is illustrated in the city of Girona, Catalonia (northeast of the Iberian Peninsula). Girona, with its 102,666 inhabitants and 47,446 households (2.2 citizens per household), is a typical compact Western Mediterranean city (Statistical Institute of Catalonia, 2022). Its urban area extends 12.7 km^2 on a rivers' crossing, has a population density of $8,139$ hab/km^2, an average slope of 5.1, and an altitude range (difference between the minimum and maximum altitudes) of 177 m. Within the results section of this paper, we present two scenarios:

(i) Achieving a cost-effective design of a new resilient reclaimed water network within a limited budget, ensuring hydraulic feasibility through the Resilience by Design (RbD) algorithm (Algorithm 2 and Section 3.2).
(ii) Enhancing resilience based on previous research regarding a non-resilient but cost-effective reclaimed water network design (Calle et al., 2023), using the Resilience-Strengthening (RS) algorithm (Algorithm 3 and Section 3.3).

Both scenarios share the same initial water tank location in the Fontajau neighborhood, situated at 111 m above sea level. The initial

graph (see Section 2.1) encompasses Girona's entire urban area, including estimated reclaimed water demands for various public and private purposes (see Fig. 1), as detailed in previous works (Calle et al., 2023).

The results were obtained using an Ubuntu 20.04 LTS server (CPU: AMD Ryzen 5 5600X, 32GB RAM), executed within a Python notebook (Jupyter Hub, vanRossum (1995)). However, the tool is adaptable to other systems.

3.2. Resilient-by-design network

In scenario (i), a cost-effective and resilient network, named 'resilient-by-design', has been designed to ensure hydraulic feasibility within a limited budget of €1,500,000. The Resilience by Design (RbD) algorithm (Algorithm 2) was employed for this purpose. Fig. 2 illustrates the network graph created for this scenario, resulting in a 16-kilometer network. The 'resilient-by-design' network accommodates a total consumption of 1527 cubic meters per day, effectively serving 21.4 % of the city's total reclaimed water demand (Water served / total demand × 100). Remarkably, the algorithm demonstrated a rapid execution, completing its task within seconds.

Contrasting with prior research works (Yazdani et al., 2011; Herrera et al., 2016), the 'resilient-by-design' network not only conforms to a cost-effective, resilient layout consistent with the principles of the RbD algorithm, ensuring that each destination point is supplied by at least two distinct paths, but it also undergoes thorough hydraulic feasibility validation. This hydraulic feasibility assessment acts as a critical link between theoretical design and practical implementation.

The RbD algorithm integrates the Hydraulic-Feasible Diameter Selection (HFDS) algorithm (Algorithm 1) and the Hydraulic Refinement Relay (HRR) process. It utilizes the EPANET software to validate hydraulic feasibility at every stage of the reclaimed water network design process. The hydraulic simulation results for the 'resilient-by-design' are detailed in Table 4, presenting essential hydraulic output indicators categorized into three main groups: Water Service, Node Pressure, and Water Quality.

In the Water Service category, two vital aspects are assessed. First, a verification is conducted to check whether water can be effectively delivered to all destination points. Second, it is confirmed that water speeds remain below the established maximum limit of 1.2 m/s, thereby ensuring the correct functioning of the network.

Within the Node Pressure category, the focus lies on the distribution of node pressures to ensure that all node pressures fall within the acceptable range (i.e., between 15 and 60 m (Desta et al., 2022; MoWR, 2006)).

Addressing Water Quality concerns, especially in areas characterized by lower water speeds, we calculate the required travel time for water to reach each destination from the initial tank. This calculation is essential to ensure that water quality remains within acceptable limits. In this case study, the recommended maximum water age value of 72 h (World Health Organization, 2014) has been utilized (see Section 2.2.2). This analysis provides assurance that, despite lower speeds in some specific parts of the network (e.g., the smallest 32 mm pipes and distant endpoints with low demands), water quality is compliant and within acceptable limits.

In future works, we will explore incorporating pressure pumps within the water distribution network designs. This proactive approach ensures an uninterrupted water supply, even in scenarios where initial tank elevation alone may not be sufficient to serve water to all the destinations.

The resilience assessment of the 'resilient-by-design' network utilized the Water Availability (WA) metric, as outlined in Section 2.2.1. This assessment involved an extensive series of 1,000,000 Monte Carlo simulations. Remarkably, these simulations resulted in a Water Availability value of WA = 0.9949, indicating a substantial 99.49 % probability that the initial water tank can consistently fulfill the water demands of all destination nodes within the network. This translates to a mere 0.51 % chance of service disruption.

The Monte Carlo simulations conducted to evaluate WA also facilitated the computation of the Average Unserved Water per service Disruption event (AUW/D), which quantifies the average volume of water that cannot be served in the event of a service disruption, yielding a value of 512 m^3. Moreover, the Average Disruptions per Year (AD/Y) was computed using the Mean Time To Repair (MTTR) of 24 h and the Mean Time Between Disruptions (MTBD) derived from the Water Availability (WA) metric, as detailed in Section 2.2.1, yielding an average of 1.86 disruptions per year. Consequently, the Average Unsupplied Water per Year (AUW/Y) was determined to be a total volume of 952 m^3/year.

3.3. Resilience-strengthening of a current non-resilient network

In this section, scenario (ii), involving the resilience enhancement of a current 'non-resilient' network through the application of the Resilience Strengthening (RS) algorithm (Algorithm 3) is introduced. The 'non-resilient' network illustrates a cost-effective design for a reclaimed water network within the same city of Girona. This network's design is based in the same topology obtained by the algorithms introduced in Calle et al. (2023).

The original 'non-resilient' network design is not hydraulically feasible, as flows were predicted without the EPANET validation. Thus, before the application of the Resilience Strengthening (RS) algorithm (Algorithm 2), the network's pipe diameters have been adapted to ensure hydraulic feasibility, thus facilitating a fair comparison of resilience and cost output indicators. Consequently, the adapted 'non-resilient' network, illustrated in Fig. 3, represents a tree-based topology characterized by hydraulic feasibility, 10 kilometer pipe length, and the same 1527 m^3/day volume of transported water as the 'resilient-by-design' network in scenario (i), effectively serving 21.4 % of the city's total reclaimed water demand. The adapted pipe diameters resulted on a total construction cost for the network of €1,135,000.

The application of the RS algorithm over the current 'non-resilient' network resulted in the 'resilience-strengthened' design, as illustrated in Fig. 4. The 'resilience-strengthened' design evolved to a meshed topology characterized by hydraulic feasibility, 20 kilometer pipe length, and maintaining the same 1527 m^3/day volume of transported water. The added pipes resulted in an extra 10 km and a construction cost of €742,000, which results in a total cost of €1,880,000. In the figure, the original pipes of the 'non-resilient' network are illustrated in red, while the new pipes extension are in blue.

3.4. Comparative analysis of key output indicators

This section provides a comprehensive comparative analysis of key output indicators among the following networks: the current 'non-resilient' network, the 'resilience-strengthened' network enhanced through the application of the Resilience Strengthening (RS) algorithm (Algorithm 2) to the 'non-resilient' network, and the 'resilient-by-design' network created in scenario (i), as presented in Table 5. In the context of this study, it becomes evident that while the daily reclaimed water volume remains constant, notable variations emerge in the output indicators of the three network designs.

Beginning with the 'non-resilient' design, it is apparent that its construction cost is the lowest at €1,135,000, primarily attributed to its limited pipe length of just 10 km. As expected, this design exhibits the least favorable Water Availability (WA) at 99.20 %, resulting in a 0.80 % probability of service disruption. Furthermore, it registers the highest values for Average Unserved Water per Disruption (AUW/D), Average Disruptions per Year (AD/Y), and Average Unserved Water per Year (AUW/Y), culminating in a substantial volume of 2786 m^3/year affected by service disruptions. To ensure the sustained and reliable operation of these critical networks over the long term, it is imperative to minimize

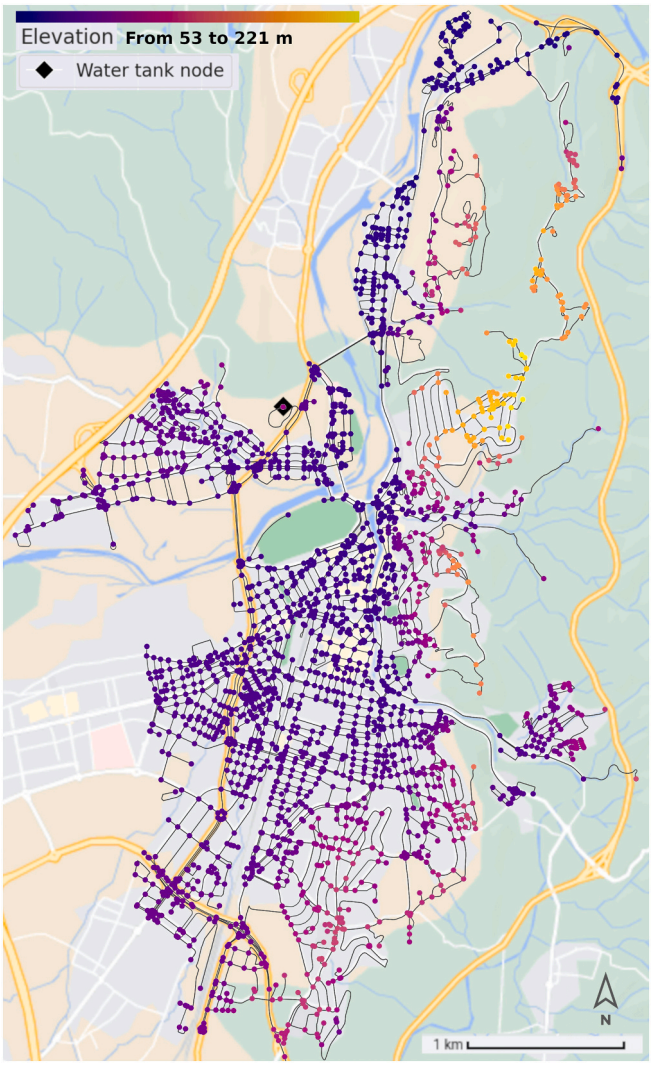


Fig. 1. Visualization of the Girona case study area, highlighting nodes color-coded by elevation, and indicating the placement of the initial water tank.

both the probability of service disruptions (SD) and the impact on water service in the event of such disruptions as much as possible (AUW/D). Concerning additional costs, the AUW/Y cost amounts to €1530 and the cost of D/KY rises to €9980, marking them as the highest among the three designs, reflecting the low resilience of the network.

The enhanced 'resilience-strengthened' design presents a network expansion of 20 km, which is almost twice the length of the pipe network compared to the original 'non-resilient' design. The total construction cost of the network is the sum of the cost of the original 'non-resilient' design (i.e., €1,135,000) and the cost of the expansion resulting from the RS algorithm (i.e., €742,000), which results in €1,877,000. Thus, the required investment to enhance resilience represents 65 % of the initial construction cost of the 'non-resilient' network. While this initial financial commitment is substantial, it yields significant improvements in resilience. The 'resilience-strengthened' design achieves a Water Availability (WA) of 99.30 %, translating to a noteworthy 12.5 % reduction in service disruptions (SD). Furthermore, it reduces the Average Unserved Water per Year (AUW/Y) to 1656 m^3, constituting a nearly 65 % reduction in the impact of water service disruptions. Notably, despite the 'resilience-strengthened' design featuring nearly twice the pipe length of the original 'non-resilient' network, the Average Disruptions per Year (AD/Y) also decrease by 11 % (from 2.92 to 2.56). This reduction underscores the effectiveness of the RS algorithm's strategy, which enhances network resilience by meshing the

Fig. 2. Visualization of the 'resilient-by-design' €1,500,000 reclaimed water network in Girona (highlighted in red) overlaying the Girona street graph (in grey).

infrastructure and ensuring multiple distinct water paths from the initial tank to all destinations, alongside efficient valve placement. Regarding additional costs, the AUW/Y cost of this design amounts to €910, which marks a 40 % decrease compared with the 'non-resilient' design, reflecting its improved resilience. Although the cost of D/KY decreases to €9530, it only represents a 5 % reduction, underscoring the trade-off between network resilience and maintenance expenses.

Notably, it is worth emphasizing that, to the best of our knowledge, there are currently no other cost-effective solutions available in existing research that offer comparable enhancements to the resilience of pre-existing network designs. By employing RS, organizations can address the critical issue of network resilience within the limitations of available resources, thereby reducing the potential for economic losses and service interruptions while maximizing the resilience and functionality of their water distribution systems.

In the context of planning the new design of water distribution

Table 4
Summary of hydraulic feasibility indicators provided by EPANET ('resilient-by-design', scenario (i)).

Indicator	Result
Water service	
- All destination points are fully supplied	Yes
- Water speed remains within the maximum (<12 m/s)	Yes
Node pressure	
- All nodes present adequate pressures (in meters)	Yes
- Quartiles [Q1, Q2-Median, Q3]	[33.3, 36.5, 40.5]
- Minimum and maximum values	[29.9, 44.5]
Water quality	
- Water age is acceptable (in minutes)	Yes
- Quartiles [Q1, Q2-Median, Q3]	[60.5, 75.5, 90.6]
- Minimum and maximum values	[6.7, 363.3]

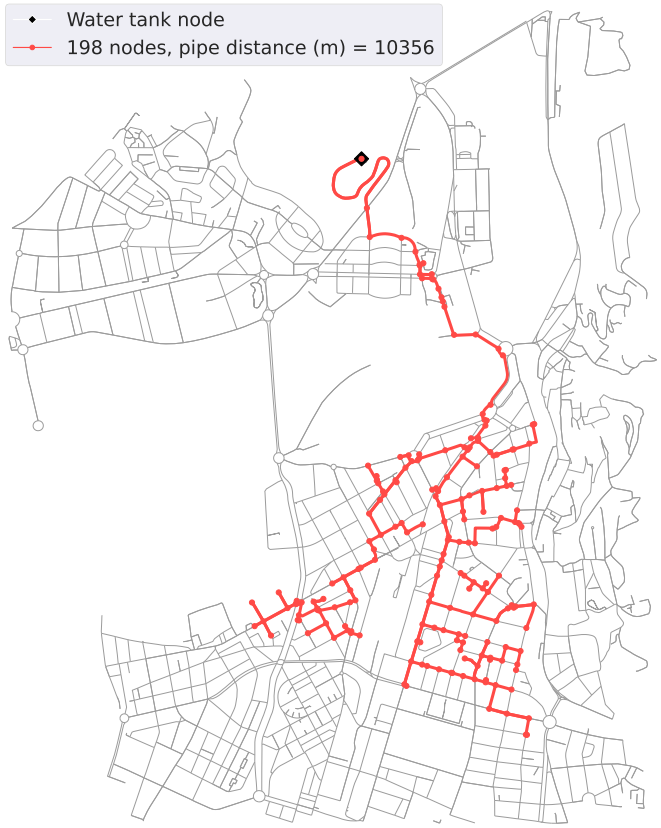


Fig. 3. Visualization of the 'non-resilient' current reclaimed water network in Girona (highlighted in red) overlaying the Girona street graph (in grey).

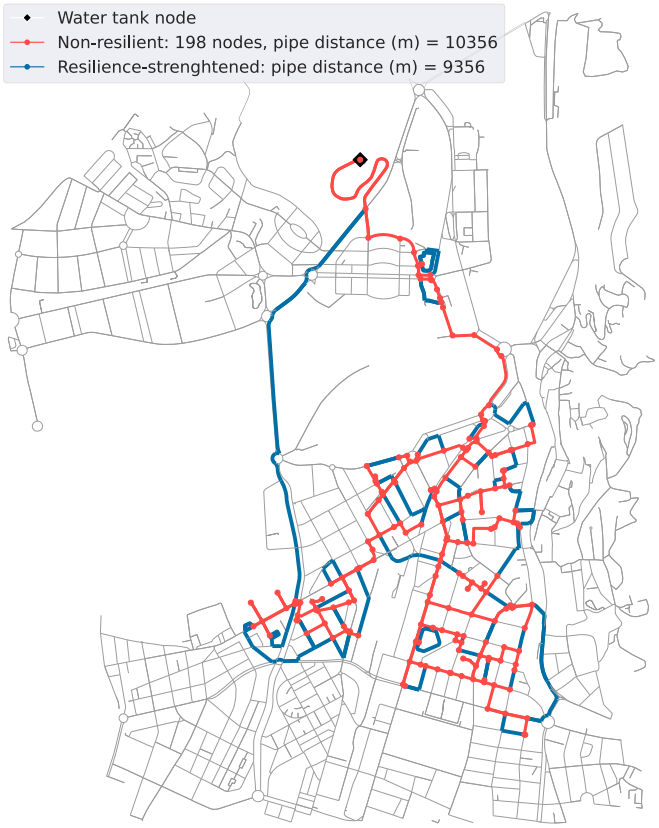


Fig. 4. Visualization of the 'resilience-strengthened' network extension in Girona (highlighted in blue) over the 'non-resilient' network existing pipes (highlighted in red).

networks (WDNs), particularly in newly developed neighborhoods, it becomes essential to compare 'non-resilient' and 'resilient-by-design' networks. This comparison aids in determining the overall prioritization strategy, which can either emphasize solely the cost-effectiveness of the design or take into account both resilience and cost-effectiveness indicators.

Notably, the 'resilient-by-design' network incurs an initial cost that is 32 % higher, primarily due to the requirement for a more interconnected and meshed topology. However, it is worth noting that, despite the higher initial cost, the 'resilient-by-design' approach still offers a cost reduction of 20 % compared to making an already existing network resilient. This upfront investment yields significant benefits, including a 36 % reduction in the probability of service disruption and a nearly 50 % decrease in the average unserved water volume per disruption. Consequently, the annual average unserved water drops from 2786 to 952 m^3, representing a substantial 65 % reduction. In terms of additional costs, the 'resilient-by-design' approach demonstrates significant advantages. In particular, it shows the lowest AUW/Y cost of €520, indicative of its maximal resilience. Furthermore, the cost of D/KY amounts to €6770, marking a notable 32 % reduction from the cost observed in the 'non-resilient' design at €9980. This decrease in disruption repair costs per kilometer per year, despite a considerable increase in the network's pipe length, underscores the effectiveness of the 'resilient-by-design' network. It indicates that despite the infrastructure's significant expansion, the associated disruption repair expenses have notably decreased, demonstrating the cost-effectiveness of resilience-focused strategies. This reaffirms the efficacy of the resilient-by-design approach, highlighting its capacity to mitigate disruptions while efficiently managing maintenance expenditures.

Fig. 5 illustrates the progression of Average Unserved Water per Year (AUW/Y) over a 50-year operational period. When extending our perspective to this extended timeframe, opting for a 'resilient-by-design' network would effectively shield up to a substantial 91,700 m^3 of water from the impact of water disruption events, in stark contrast to the 'non-resilient' network. Even when factoring in the subsequent resilience enhancement of the RS algorithm for the 'non-resilient' network, resulting in the 'resilience-strengthened' design with notable improvements, the associated costs would still be 25 % higher, and the annual average unserved water volume would remain 76 % higher. This

Table 5
Comparative analysis of key output indicators across 'non-resilient', 'resilience-strengthened', and 'resilient-by-design' network designs.

	Non-resilient	Resilience-strengthened	Resilient-by-design
Length (m)	10,356	19,712	16,113
Cost (k€)	1135	1877	1500
WA (%)	99.20	99.30	99.49
SD (%)	0.80	0.70	0.51
AUW/D (m^3)	954	647	512
AD/Y	2.92	2.56	1.86
AUW/Y (m^3)	2786	1656	952
Cost AUW/Y (k€)	1.53	0.91	0.52
Cost D/KY (k€/km)	9.98	9.53	6.77

WA (%) – percentage of Water Availability; SD (%) – percentage of Service Disruption (SD = 1 - WA); AUW/D (m^3) – volume of Average Unserved Water per Disruption; AD/Y – number of Average Disruptions per Year; AUW/Y (m^3) – volume of Average Unserved Water per Year; Cost AUW/Y (k€) – cost of the AUW/Y, derived from the mean price per cubic meter of €0.550502; Cost D/KY (k€/km) – cost of disruption repairs per kilometer per year, derived from the mean price of disruption repairs of the WA Monte Carlo realizations, the AD/Y, and the length of the network (based on the cost of entire pipe replacements in the events of disruptions).

underscores the long-term benefits of integrating resilience into the initial design considerations (i.e., resilience by design).

3.5. Cost analysis: reclaimed network vs. tap water baseline

In addition to evaluating the effectiveness of the 'resilient-by-design' network solution, we conducted a comparative cost analysis against a baseline scenario using tap water. In this baseline scenario, the supply of all water relies solely on the drinking water distribution network as no reclaimed water network exists. Upon implementing the 'resilient-by-design' network, 1527 cubic meters per day of reclaimed water are now supplied, reducing the reliance on the drinking water network for this portion of water demand.

To assess the benefits of this transition, we compared the cost savings of water originally provided by the baseline drinking water network, calculated using a price of €1.09 per cubic meter. This baseline scenario's cost is calculated from the mean price per cubic meter of drinking water in Spain in 2022 (AEAS-AGA, 2022).

On one hand, the reclaimed WDN includes the costs extracted from the algorithms and methods (i.e., design phase), which include construction costs, the cost of disruption repairs per year, and Water Availability (WA) costs; WA costs represent the expenses associated with the Average Unserved Water per Year (AUW/Y), which during disruption events would otherwise be supplied by the original baseline drinking water network.

On the other hand, the reclaimed WDN costs also encompass the adequate wastewater treatment to obtain the appropriate reclaimed water quality for specified water usages, the construction of the supplying WDN from the wastewater treatment plant (WWTP) to the reclaimed WDN initial tank, and operation costs of the pressure pump needed for the supplying WDN. These costs have been manually calculated for our case study. Construction costs of the supplying WDN include expenses for pipes, valves, and an additional water tank with the same capacity as the reclaimed WDN initial tank placed at the WWTP site. These costs have been meticulously computed utilizing the same REWATnet database as the construction costs of reclaimed WDNs. Additionally, the operational costs of the pressure pump essential for the supplying WDN have been determined based on an average water pumping energy consumption of 0.475 kWh/m^3 (Yerri and Piratla, 2019), and the final average energy price in Spain during 2023, standing at 0.0996 €/kWh (Statista, 2024).

The costs associated with wastewater treatment at the WWTP have been derived from the Suggereix tool (Catalan Water Agency, 2021), specifically designed to streamline access to information and resources concerning reclaimed water in Catalonia. This tool aids in decision-making processes regarding the costs of water regeneration and reuse procedures. It has been used in conjunction with the Spanish guide for the application of R.D. 1620/2007 (Spanish Ministry of Environment and Rural and Marine Affairs, 2010), which proposes the following wastewater treatment train: conventional activated sludge system with biological nutrient removal, including a secondary clarifier + coagulation/flocculation process + sand filter + microfiltration or ultrafiltration membrane + final disinfection (using Cl_2). This treatment train is recommended for achieving the best water quality, fully eliminating *Escherichia coli* concentrations, which is necessary for the water uses selected in our case study, including public and private garden irrigation and toilet flushing. The expenses linked to this wastewater treatment train, encompassing both construction and operation and maintenance (O&M) costs, for the water volume of 1527 m^3/day for 50 years have been calculated as €1100/day.

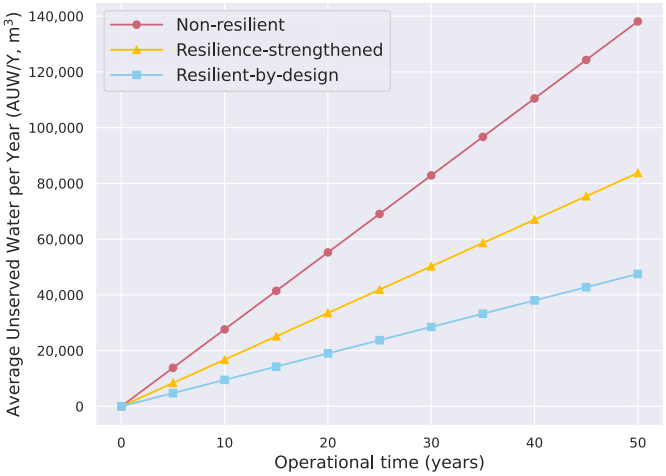


Fig. 5. Evolution of Average Unserved Water per Year (AUW/Y) over a 50-year operational period.

Table 6
Cost analysis of the 'resilient-by-design' reclaimed WDN, including the supplying WDN and wastewater treatment costs, considering a 50-year operational period.

Metric	Cost (k€)	Cost (%)
Construction	1500	5.12
Disruption repairs	5453	18.62
Tap water served during disruption events	26	0.09
Supplying WDN construction	906	3.10
Supplying WDN operational costs	1318	4.50
Wastewater treatment at WWTP	20,075	68.57
Total	29,278	100

Table 6 outlines the cost analysis of the 'resilient-by-design' reclaimed WDN over a 50-year operational period, providing a comprehensive assessment of its cost-effectiveness. The total cost of the reclaimed WDN for this period amounts to €29,278,000 (€1.05 per cubic meter), compared to €30,375,848 (€1.09 per cubic meter) for the tap water baseline, showing that they are both in the same level of magnitude. In addition to the slightly 3.67 % cost decrease, the reclaimed WDN saves up to a remarkable 27,820,150 cubic meters of water from the drinking water network. Significantly, the complex wastewater treatment train emerges as the primary expense, comprising up to 68.57 % of the total cost of the reclaimed WDN. This underscores the substantial investment required to treat wastewater to a quality suitable for reuse in our case-study scenario.

It is noteworthy that Spain's pricing model for drinking water does not fully account for all associated costs, including operational, infrastructure maintenance, renovation, and quality assurance measures (AEAS-AGA, 2022). In contrast, these associated costs, which are integral to ensuring water quality and service reliability, are fully integrated into the cost analysis of our reclaimed WDN solution. Therefore, taking these factors into account, the potential cost savings offered by our reclaimed WDN solution could be even more significant compared to the conventional tap water baseline.

While the 'resilient-by-design' reclaimed WDN demonstrates cost-effectiveness over a 50-year operational period, an initial investment is required to build the infrastructure (i.e., construction costs). Given the current drought emergency in our case study region, it becomes imperative to seek public aid initiatives aimed at promoting water reuse. Such subsidies are crucial in incentivizing water reuse practices, especially considering the severe strain on water resources. These initiatives not only yield significant ecological benefits but also promise substantial long-term cost savings.

At present, the methodology outlined in this paper does not include pressure pumps in the algorithms generating reclaimed WDN designs, thus omitting direct consideration of operation and maintenance (O&M) costs from the initial elevated tank to the consumption destinations. In our specific case study, the deliberate absence of pressure pumps aligns with the geographical reality of Girona, where an elevated location with pre-treated reclaimed water readily available for designated purposes

eliminates the need for additional pumping. Future research will incorporate pressure pumps and O&M costs directly into the reclaimed WDN design phase, allowing for the evaluation of additional case studies that necessitate their use. This advancement will build upon insights from relevant literature, such as studies by Khurelbaatar et al. (2021) and Friesen et al. (2023), which provide valuable frameworks for integrating such considerations.

3.6. Final thoughts

While there exists a substantial body of literature on water network resilience and its evaluation, it becomes evident that this specific field of study has been lacking comprehensive measures or strategies for ensuring resilience through design. This paper addresses this gap in the literature by presenting algorithms and metrics that are well-suited for integration with existing approaches and tools, thus offering practical solutions for enhancing reclaimed WDN resilience. Notably, our work serves as an ideal supplement to the Water Network Tool for Resilience (WNTR) (Klise et al., 2017a, 2018). In addition to its primary function of resilience assessment, our algorithms offer cost-effective solutions for enhancing resilience in two key scenarios: the initial design of a network (RbD algorithm), and the optimization of resilience in existing operational networks (RS algorithm).

The figures showcased in the results section of this paper are automatically generated and exported as high-quality PDF vector maps. This ensures that their clarity and detail are preserved even when zooming in. Additionally, the network designs and hydraulic feasibility outputs are parsed into interactive HTML maps, a valuable resource for network operators when evaluating results and making informed decisions (Crickard III, 2014). Specifically, our Hydraulic Refinement Relay (HRR) process extracts key data from the EPANET output indicators, including node pressures, water supply, pipe velocities, and flow rates. This interactive presentation of information offers a comprehensive understanding of the network's performance, enhancing decision-making capabilities.

In contrast with the majority of the evaluated literature, all algorithm definitions, implementations, and output indicator results for the case study, including numerical, on-map, and graphical data visualizations, are available on a dedicated public repository (Martínez, 2023). This transparency and accessibility underscore our commitment to fostering collaboration and enabling wider application of our research findings in practice. All these algorithms and data will be available to municipalities, empowering them to utilize our work as needed. Given that the design and planning phase of water distribution networks is known to cost up to 10 % of the initial investment (Khurelbaatar et al., 2021), our work not only facilitates decision-making but also potentially saves significant financial resources.

4. Conclusions

This study fills a critical gap in the existing literature by introducing novel approaches for enhancing and evaluating the resilience of reclaimed water distribution networks. We have presented mathematical algorithms for both designing resilient water distribution networks from scratch and enhancing existing networks within limited budget constraints. These algorithms not only prioritize resilience but also optimize cost-effectiveness. Remarkably, the Resilience by Design (RbD) algorithm (Algorithm 2) guarantees both cost-effective and resilient water network designs, even with limited budgets. Additionally, the Resilience-strengthening (RS) algorithm (Algorithm 3) demonstrated excellent performance by significantly enhancing resilience output indicators within an existing non-resilient network design.

Furthermore, we introduced a novel metric, Water Availability (WA), providing a comprehensive measure to evaluate network resilience. Our innovative and interactive process seamlessly integrates these algorithms with the EPANET software validations, ensuring cost-effective resilience while adhering to hydraulic feasibility constraints in network designs. This holistic approach reshapes the way we evaluate and design resilient water distribution systems.

In practical terms, our study demonstrated substantial benefits in prioritizing resilience in network design. Comparing a resilient-by-design network to a non-resilient counterpart, we observed a 36 % reduction in the probability of water service disruption and a significant 65 % decrease in the annual average unserved water due to service disruptions (AUW/Y). Notably, not only is the AUW/Y significantly reduced, but the cost of disruption repairs per kilometer per year (D/KY) also decreases by a significant 32 % compared to the non-resilient design, highlighting its cost-effectiveness. These findings underscore the long-term advantages of resilience-focused network design, with our case study effectively safeguarding up to a significant 91,700 cubic meters of water from the impact of water disruption events compared to an equivalent non-resilient design.

Our cost analysis also revealed a 3.67 % decrease in costs over a 50-year operational period for our 'resilient-by-design' reclaimed water distribution network (WDN) solution compared to the baseline of using the conventional drinking WDN, highlighting its cost-effectiveness. Additionally, a remarkable amount of 27,820,150 cubic meters of clean water can be saved by our reclaimed WDN solution, especially meaningful in light of the ongoing drought emergency present in our case study region.

Overall, our research not only addresses a crucial gap in the literature by focusing on reclaimed water distribution networks but also illustrates its applicability to other WDN systems with minimal input data adaptation. Our work provides valuable insights and practical tools for the resilient design of water distribution networks, significantly contributing to the advancement of water resource management and infrastructure planning.

CRediT authorship contribution statement

David Martínez: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Sergi Bergillos:** Conceptualization, Methodology, Visualization, Writing – review & editing. **Lluís Corominas:** Conceptualization, Resources, Visualization, Writing – review & editing. **Joaquim Comas:** Conceptualization, Resources, Visualization, Writing – review & editing. **Fenghua Wang:** Conceptualization, Methodology, Resources, Writing – review & editing. **Robert Kooij:** Conceptualization, Resources, Writing – review & editing. **Eusebi Calle:** Conceptualization, Resources, Supervision, Visualization, Writing – review & editing.

Declaration of competing interest

There is no conflict of interest.

Data availability

The data, algorithms, and code implementations that support the findings of this study are openly available in "Enhancing resilience of water distribution networks with cost-effective meshing" at https://doi.org/10.5281/zenodo.8398703, reference Martínez (2023).

Acknowledgements

University of Girona researchers thank the Generalitat de Catalunya for their support through a Consolidated Research Group (LEQUIA, 2021 SGR 01125). ICRA researchers acknowledge the support from the Economy and Knowledge Department of the Catalan Government through a Consolidated Research Group (ICRA-TECH - 2021 SGR 01283). David Martínez thanks the University of Girona for his FI fellowship (IFUdG 46 2022). Sergi Bergillos thanks the Departament de Recerca i Universitats de la Generalitat de Catalunya and the European

Social Fund for his FI fellowship (2023 FI-1 00751). F. Wang is supported by the China Scholarship Council (No. 201906040194). The support from the project ReUseMP3 from the Spanish State Research Agency of the Spanish Ministry of Science and Innovation [PID2020-115456RBI00/MCIN/AEI/10.13039/501100011033] is also acknowledged. The University of Girona and ICRA researchers would like to thank the Spanish Ministry of Industry, Commerce and Tourism for their support through the ReWat project [AEI-010500-2023-339]. Finally, the authors would like to thank ABM Consulting for providing support regarding the pipe repair economic data and Cicle de l'Aigua del Ter S.A. for providing access to Girona data. Open Access funding provided thanks to the CRUE-CSIC agreement with Elsevier.

References

AEAS-AGA, 2022. Price of supply and sanitation services in Spain. Retrieved from: https://www.aeas.es/component/content/article/53-estudios/estudios-tarifas/265-estudio-de-tarifas-2022?Itemid=101.

Ahmad, T., Shaban, I., Zayed, T., 2023. Watermain breaks in Hong Kong: causes and consequences. In: 2nd International Conference on Civil Infrastructure and Construction (CIC 2023). Qatar University Press.

Ahmadullah, R., Dongshik, K., 2016. Designing of hydraulically balanced water distribution network based on gis and epanet. Int. J. Adv. Comput. Sci. Appl. 7 (2).

Baños, R., Reca, J., Martínez, J., Gil, C., Márquez, A.L., 2011. Resilience indexes for water distribution network design: a performance analysis under demand uncertainty. Water Resour. Manag. 25, 2351–2366.

Bennett, J., 2010. OpenStreetMap. Packt Publishing Ltd.

Boeing, G., 2017. Osmnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. Comput. Environ. Urban. Syst. 65, 126–139.

Calle, E., 2004. Enhanced Fault Recovery Methods for Protected Traffic Services in Gmpls Networks. Universitat de Girona, Girona.

Calle, E., Martínez, D., Brugués-i Pujolràs, R., Farreras, M., Saló-Grau, J., Pueyo-Ros, J., Corominas, L., 2021. Optimal selection of monitoring sites in cities for sars-cov-2 surveillance in sewage networks. Environ. Int. 157, 106768.

Calle, E., Martínez, D., Buttiglieri, G., Corominas, L., Farreras, M., Saló-Grau, J., Vilà, P., Pueyo-Ros, J., Comas, J., 2023. Optimal design of water reuse networks in cities through decision support tool development and testing. npj Clean Water 6 (1), 23.

Carlier, J., Lucet, C., 1996. A decomposition algorithm for network reliability evaluation. Discret. Appl. Math. 65 (1–3), 141–156.

Catalan Water Agency, 2021. Decision support system for implementation and management of reuse (v1.6.1). Retrieved from: https://suggereix.icradev.cat/.

Chang, S.E., Shinozuka, M., 2004. Measuring improvements in the disaster resilience of communities. Earthq. Spectra 20 (3), 739–755.

Christodoulou, S., Agathokleous, A., 2012. A study on the effects of intermittent water supply on the vulnerability of urban water distribution networks. Water Sci. Technol. Water Supply 12 (4), 523–530.

Christodoulou, S., Fragiadakis, M., Agathokleous, A., Xanthos, S., 2017. Urban Water Distribution Networks: Assessing Systems Vulnerabilities, Failures, and Risks. Butterworth-Heinemann.

Cimellaro, G., Tinebra, A., Renschler, C., Fragiadakis, M., 2016. New resilience index for urban water distribution networks. J. Struct. Eng. 142 (8), C4015014.

Crickard III, P., 2014. Leaflet. js Essentials. Packt Publishing Ltd.

Darvini, G., Ruzza, V., Salandin, P., 2020. Performance assessment of water distribution systems subject to leakage and temporal variability of water demand. J. Water Resour. Plan. Manag. 146 (1), 04019069.

Desta, W.M., Feyessa, F.F., Debela, S.K., 2022. Modeling and optimization of pressure and water age for evaluation of urban water distribution systems performance. Heliyon 8 (11).

Domènech, L., Saurí, D., 2010. Socio-technical transitions in water scarcity contexts: public acceptance of greywater reuse technologies in the metropolitan area of Barcelona. Resour. Conserv. Recycl. 55 (1), 53–62.

Friesen, J., Sanne, M., Khurelbaatar, G., van Afferden, M., 2023. "Octopus" principle reduces wastewater management costs through network optimization and clustering. One Earth 6 (9), 1227–1234.

Gaur, V., Yadav, O.P., Soni, G., Rathore, A.P.S., 2021. A literature review on network reliability analysis and its engineering applications. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 235 (2), 167–181.

Gillies, S., et al., 2007. Shapely: Manipulation and Analysis of Geometric Objects.

GraphML, 2001. The GraphML file format. http://graphml.graphdrawing.org/ (Online; accessed 21-dec-2020).

Herrera, M., Abraham, E., Stoianov, I., 2016. A graph-theoretic framework for assessing the resilience of sectorised water distribution networks. Water Resour. Manag. 30, 1685–1699.

Hu, Y., Wang, W., Gong, X., Que, X., Cheng, S., 2014. On reliability-optimized controller placement for software-defined networks. China Communications 11 (2), 38–54.

Instituto Geográfico Nacional, 2023. Digital elevation model provider for Spain. https://www.ign.es/ (Accessed: 2023-30-05).

Kesavan, H., Chandrashekar, M., 1972. Graph-theoretic models for pipe network analysis. J. Hydraul. Div. 98 (2), 345–364.

Khurelbaatar, G., Al Marzuqi, B., Van Afferden, M., Müller, R.A., Friesen, J., 2021. Data reduced method for cost comparison of wastewater management scenarios–case study for two settlements in Jordan and Oman. Front. Environ. Sci. 9, 626634.

Klise, K.A., Bynum, M., Moriarty, D., Murray, R., 2017a. A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. Environ. Model. Software 95, 420–431.

Klise, K.A., Hart, D., Moriarty, D.M., Bynum, M.L., Murray, R., Burkhardt, J., Haxton, T., 2017b. Water network tool for resilience (WNTR) user manual. In: Technical Report. Sandia National Lab. (SNL-NM), Albuquerque, NM (United States).

Klise, K.A., Murray, R., Haxton, T., 2018. An overview of the Water Network Tool for Resilience (WNTR). In: 1st International WDSA/CCWI 2018 Joint Conference. Sandia National Lab. (SNL-NM), Albuquerque, NM (United States).

Kristensen, P., Whalley, C., Zal, F.N.N., Christiansen, T., et al., 2018. European waters assessment of status and pressures 2018. In: EEA Report, 7/2018.

Lindhe, A., Rosén, L., Norberg, T., Bergstedt, O., 2009. Fault tree analysis for integrated and probabilistic risk analysis of drinking water systems. Water Res. 43 (6), 1641–1653.

Liu, W., Song, Z., 2020. Review of studies on the resilience of urban critical infrastructure networks. Reliability Engineering & System Safety 193, 106617.

Lu, J., Zhang, Z., Hu, T., Yi, P., Lan, J., 2019. A survey of controller placement problem in software-defined networking. IEEE Access 7, 24290–24307.

Martínez, D., 2023. Extended Version of the Algorithms and Results of the Research Work "Enhancing Resilience of Water Distribution Networks With Cost-Effective Meshing" https://doi.org/10.5281/zenodo.8398703.

McAllister, T.P., 2015. Community resilience planning guide for buildings and infrastructure systems, volume I. In: Special Publication (NIST SP).

Meng, F., Fu, G., Farmani, R., Sweetapple, C., Butler, D., 2018. Topological attributes of network resilience: a study in water distribution systems. Water Res. 143, 376–386.

Mezhoudi, M., Chu, C.-H.K., 2006. Integrating optical transport quality, availability, and cost through reliability-based optical network design. Bell Labs Technical Journal 11 (3), 91–104.

MIMAM, E, 2000. Libro blanco del agua en españa. In: Dirección General de Obras Hidráulicas y Calidad de las Aguas. Secretaría de Estado de Aguas y Costas. MIMAM, Madrid.

MoWR, 2006. Urban Water Supply Design Criteria. Water Resources Administration Urban Water Supply and Sanitation Department, pp. 1–60.

Olbricht, R., et al., 2011. Overpass API.

Pezoa, F., Reutter, J.L., Suarez, F., Ugarte, M., Vrgoč, D., 2016. Foundations of json schema. In: Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee.

Piratla, K.R., Matthews, J.C., Farahmandfar, Z., 2016. The role of resilience in the rehabilitation planning of water pipeline systems. In: Pipelines 2016. American Society of Civil Engineers, pp. 1856–1864.

Rezaei, H., Ryan, B., Stoianov, I., 2015. Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks. Procedia Engineering 119, 253–262.

Rosenthal, A., 1977. Computing the reliability of complex networks. SIAM J. Appl. Math. 32 (2), 384–393.

Rossman, L.A., et al., 2000. Epanet 2: Users Manual. US Environmental Protection Agency. Office of Research and Development.

Simpson, A.R., Elhay, S., 2008. Formulating the water distribution system equations in terms of head and velocity. In: Water Distribution Systems Analysis 2008. American Society of Civil Engineers, Kruger National Park, South Africa, pp. 1–13.

Soldi, D., Candelieri, A., Archetti, F., 2015. Resilience and vulnerability in urban water distribution networks through network theory and hydraulic simulation. Procedia Engineering 119, 1259–1268.

Spanish Ministry of Environment and Rural and Marine Affairs, 2010. Guide for the application of Spanish R.D. 1620/2007 establishing the legal framework for the reuse of treated water. Retrieved from: https://www.miteco.gob.es/content/dam/miteco/es/agua/publicaciones/GUIA%20RD%201620_2007_tcm30-213764.pdf.

Statista, 2024. Annual final average price of electricity in Spain from 2010 to 2023. Retrieved from: https://es.statista.com/estadisticas/993787/precio-medio-final-de-la-electricidad-en-espana/.

Statistical Institute of Catalonia, 2022. The Municipality in Figures: Girona (Gironès) (Accessed on 11.01.2023).

Taiwo, R., Shaban, I.A., Zayed, T., 2023. Development of sustainable water infrastructure: a proper understanding of water pipe failure. J. Clean. Prod. 398, 136653.

Todini, E., 2000. Looped water distribution networks design using a resilience index based heuristic approach. Urban Water 2 (2), 115–122.

Tumula, P., Park, N., 2004. Multi-objective genetic algorithms for the design of pipe networks. Journal of Water Resources Planning and Management 130, 73–82.

U.S. EPA, 2000. EPANET 2.0: application for modeling drinking water distribution systems. https://www.epa.gov/water-research/epanet (Accessed: 2022-02-08).

Vallès-Casas, M., March, H., Saurí, D., 2016. Decentralized and user-led approaches to rainwater harvesting and greywater recycling: the case of Sant Cugat del vallès, Barcelona, Spain. Built Environ. 42 (2), 243–257.

vanRossum, G., 1995. Python Reference Manual. Department of Computer Science [CS].

Wilson, D., Filion, Y., Moore, I., 2017. State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. Urban Water J. 14 (2), 173–184.

World Health Organization, 2014. Water Safety in Distribution Systems. World Health Organization.

Yazdani, A., Otoo, R.A., Jeffrey, P., 2011. Resilience enhancing expansion strategies for water distribution systems: a network theory approach. Environ. Model. Software 26 (12), 1574–1582.

Yerri, S., Piratla, K.R., 2019. Decentralized water reuse planning: evaluation of life cycle costs and benefits. Resources, Conservation and Recycling 141, 339–346.

Zangenehmadar, Z., Moselhi, O., Golnaraghi, S., 2020. Optimized planning of repair works for pipelines in water distribution networks using genetic algorithm. Eng. Rep. 2 (6), e12179.
Zhang, Z., Feng, X., Qian, F., 2009. Studies on resilience of water networks. Chem. Eng. J. 147 (2–3), 117–121.
Zhao, X., Chen, Z., Gong, H., 2015. Effects comparison of different resilience enhancing strategies for municipal water distribution network: a multidimensional approach. Math. Probl. Eng. 2015.

Chapter 6

Discussion

The results obtained in this thesis demonstrate how the integration of innovative graph theory solutions addresses critical challenges in urban water network management. These solutions contribute positively to public health, particularly in managing pandemic situations (Section 6.1), while also mitigating risks related to external elements such as tree roots or street works in wastewater networks (Section 6.2). Additionally, they facilitate the development of cost-effective and resilient designs for reclaimed Water Distribution Networks (WDNs), thereby assisting in their implementation and enhancing overall water resource management efficiency (Section 6.3).

The methodology outlined in this thesis (Section 3) facilitated the automation of data gathering and processing, streamlining the implementation of algorithms. This approach not only led to significant efficiencies but also enabled the achievement of objectives outlined in the introduction (Section 1.2), with results revealing notable benefits.

If in the previous chapters the results were presented as individual and independent compartments, this chapter focuses on the multiple relationships that can be established among them. The purpose is, therefore, to analyze the results obtained by interweaving them according to common guidelines that constitute the main thread of the discussion. To achieve this, the chapter is structured into four sections that analyze the results based on their contributions to the overarching objectives of the study.

6.1 Distributing wastewater sampling: a vital tool in pandemic management

The findings from the case study in Girona validate the efficacy of the Monitoring Sites Evaluation (MSE) algorithm, establishing it as a vital tool in pandemic management. The algorithm is designed to optimize the coverage of sewage monitoring sites while minimizing interference and ensuring equitable coverage areas. The results demonstrate coverage exceeding 60% within maintenance holes, interference levels below 3%, and a maximum difference of 25% in coverage among monitoring sites. Thus, the algorithm achieves an optimal balance between coverage, interference, and maintenance hole distribution.

It is noteworthy that a higher number of monitoring sites leads to shorter distances from the furthest node to respective sampling points. Consequently, maintenance holes near the Wastewater Treatment Plant (WWTP) are excluded from the final solutions due to their limited number and potential interference, particularly as they are downstream. Considering the potential attenuation of RiboNucleic Acid (RNA) signals along the sewage network transport [68], future iterations of the algorithm may incorporate constraints on the maximum distance between SARS-CoV-2 RNA discharge points and monitoring sites.

Monitoring site selection (or sensor placement) has been extensively studied for drinking water networks, but relatively few studies exist on sewage networks. [69] determined key sensor locations for managing non-point pollutant sources in sewage networks using clustering analysis and ANalysis Of VAriance (ANOVA) on top of Storm Water Management Model (SWMM) simulated results, thus requiring significant computational power. Similarly, there are examples of sensor placement for illicit intrusion detection in sewage networks based on single and multi-objective optimization [70] or Bayesian decision networks [71], both of which present high complexity. [72] proposed optimal sampling monitoring locations primarily for calibrating a hydrodynamic model, while [46] and [47] developed methodologies for sensor placement in WWTPs based on graph theory and mass balances, prioritizing data quality assessment and control while minimizing ownership costs.

Therefore, our study pioneers the application of a fast, heuristic static monitoring site selection algorithm for fine-resolution SARS-CoV-2 monitoring. This algorithm holds promise for municipalities seeking to monitor RNA traces at a granular scale. Integrating RNA trace analysis with health, demographic, and socioeconomic indicators [7, 73] necessitates a module, developed in this study, linking households to maintenance

holes, enabling data aggregation from individuals to inhabitants associated with specific monitoring sites. In Catalonia, where public COVID-19 prevalence data is aggregated only at the municipality and primary health area (Àrea Bàsica de Salut (ABS)) levels, this finer scale of monitoring fills a crucial gap. Moreover, the applicability of this approach extends beyond COVID-19 tracking. It can be adapted for various purposes such as estimating pharmaceutical consumption [74], detecting illicit drug use [75], and identifying illicit pollutant discharges from industries [71] at fine spatial resolutions.

Overall, this novel distribution of wastewater sampling provides a clear improvement in pandemic data monitoring by strategically placing monitoring sites in sewage systems, becoming a vital tool in future pandemic management.

6.2 Preserving environmental integrity: mitigating risks in wastewater networks

Tree roots and street works are well-known for their constant impact on triggering pipe failures and wastewater leakages, thus mitigating their risks is essential to preserve environmental integrity. Natural factors such as tree roots affect pipe failure hazard rates, as proved by [22, 23]. In particular, they recommend checking pipes in the proximity of trees and evaluating the possibility of tree rearrangement, although a more in-depth analysis is needed. According to their work, the risk of pipe failure incidents concerning tree roots increases over time and can be exacerbated on old or corroded pipes [24]. Sydney Water [25] agrees with this statement, justifying that the roots of trees planted in the wrong place can find their way into wastewater pipes, causing about 80% of all dry weather sewage overflows and seriously affecting public health and the environment. However, there is an active discussion on whether tree roots are able to crack pipes. According to [26], tree roots may affect pipes in other ways such as joint intrusion or tensile forces. For example, no matter the individual root size, the total volume of the tree roots in a joint could develop a surface big enough to break the pipe collar.

The concept of Tree Protection Zones (TPZs) is widely known in the world of arboriculture. It is defined as the calculated area above and below ground at a given distance from the tree trunk that aims to protect the tree's roots and canopy during construction works [76]. While the TPZ primarily serves for tree protection, it can also be used inversely. The Australian Standard (AS) [77] is the most widely accepted method for calculating TPZ, although it has led to discussion [78]. Our tree root impact results for Girona, calculated based on the element-pipe distances obtained through the defined

algorithms, show that the tree impact data distribution is almost equal for both [79] and [77] methods. This demonstrates that the AS is an excellent approximation, without needing tree species data to compute the tree Impact Areas (IAs)). In a hypothetical situation where all city trees were "sewer-safe" with the same trunk size and location as the actual ones (i.e., identical tree sizes above the surface with much smaller tree root areas below), the median of tree impacts would be slightly reduced. However, more research and other case studies are needed to check this tendency.

The top 10 most impactful tree detection results illustrate that a small percentage of trees cause the most impact on pipes. This is due to the large number of upper-bound outliers in the data and the significant 44% decrease in values between the first and the tenth-placed tree. The most impactful trees are expected to be large or in a critical location where many pipes are present (e.g., street crossroads), or both. The majority of the most impactful trees have large Impact Areas (IAs). However, some exceptions were found during manual checks, revealing that they are placed in critical spots, such as roundabouts with multiple pipes. It is also worth noting that the most present species in the top 10 are the Celtis australis and the Tilia platyphyllos, which are not considered "sewer-safe" and are well-known for their relatively extensive root systems, highlighting the importance of planting "sewer-safe" trees in cities. After this preliminary assessment, decision-makers may require the application of additional methods to minimize the risks, such as localized Closed-Circuit TeleVision (CCTV) inspections in the most critical areas.

Moreover, the Element Rearrangement (ER) algorithm results show a clear benefit of rearrangement of a portion of Girona's most impactful trees in terms of pipe failure risks, thus preventing environmental and public health hazards by avoiding wastewater leakages caused by tree roots. In particular, our case study results prove this statement considering the Australian Standard TPZ method with tree replacement enabled (i.e., to plant a sewer-safe tree in the same location) from 0 to 13% of the dataset trees, showing that the number of high-risk pipes is reduced drastically by 75%, and medium-risk ones by 30%.

The anticipated expenses for repairing pipe failures due to tree root intrusion in Girona were assessed at approximately €5.14 million over the pipes' operational lifespan of approximately 30 years. This estimation assumed a repair cost of €230 per meter of new 300 mm diameter PolyVinyl Chloride (PVC) pipes, based on findings from ABM Consulting's Spanish case studies. These repair costs were then used to calculate the actual economic savings of the Element Rearrangement (ER) approach analyzed in our case study. Despite the initial investment required, the ER algorithm not only prevents environmental and public health hazards by avoiding wastewater leakages caused by tree

roots, but also demonstrates a significant economic payback during the pipes' operational lifetime by rearranging a specific percentage of the city trees.

Overall, the current literature lacks a comprehensive analysis of the potential impact of tree roots on pipes in wastewater networks, especially in multi-layer crossing scenarios. It is worth noting that evaluating pipe failure risk and tree impacts quantitatively can be highly beneficial for city councils and wastewater network managers in planning and prioritizing preventive actions on wastewater network pipes. The results provide an excellent initial overview of the current scenario and can be generated easily and without the need for additional resources, extensive data, or fieldwork in sewers, in contrast with the existing literature [80, 81].

6.3 Crafting resilient solutions: designing cost-effective reclaimed water distribution networks to address drought challenges

Water resources, inherently limited and unevenly distributed across both space and time [28], represent a critical global concern. As anticipated in our first related paper (Section 5.2), an increasing number of nations are recognizing the viability of water reuse as a dependable alternative resource. Particularly noteworthy is the expanded focus on developing reclaimed water networks, now a primary agenda item for local and regional governments in Catalonia. This shift is precipitated by the current drought emergency, with water reserves in Catalonia's internal basins plummeting to a mere 15% capacity as of March 28, 2024 [65].

Efficient and sustainable water reuse depends on the realization of feasible water reuse projects, encompassing the establishment of water reclamation treatment plants and the subsequent distribution to potential users. The planning and assessment of such projects necessitate (i) decision-makers to address a multitude of queries, including determining the optimal tertiary/advanced treatment methods; (ii) estimating the volume of wastewater to be reclaimed based on the city's usage patterns; and (iii) devising the most effective water distribution network. While solutions for issues (i) and (ii) already exist, the realm of the issue (iii) remains largely unexplored, with little literature addressing this critical aspect. In response, we have developed an innovative methodology alongside cost-efficient algorithms based on graph theory. This approach automates the design process and conducts comprehensive cost analyses of reclaimed Water Distribution Networks (WDNs). Our solution marks a significant advancement

from existing literature, which predominantly focuses on decision support tools for wastewater systems alone [29, 82–86].

The first iteration of the REclaimed WATer Network Tool (REWATnet) decision support tool, designed to streamline the planning of optimal reclaimed water reuse networks within urban settings, has been meticulously developed and rigorously tested. Designed to operate seamlessly with minimal user input, the tool benefits from the open data methodology (Chapter 3) to compute the maximum volume of reclaimed water served per unit of investment cost. This encompasses the determination of network parameters such as pipe lengths and diameters, storage tank placements, population coverage, and construction expenses, all unified within a cohesive architecture, which presents an evolution from existing paradigms [87].

The crux of the tool lies in its ability to compute the optimal network graph from scratch, leveraging fast and efficient routing and clustering algorithms [50, 51], diverging from prevalent approaches [88]. Moreover, it employs the Limited Budget (LB) availability algorithm as necessary. Notably, for a city of approximately 100,000 inhabitants, the estimated construction cost of an optimal water reclaimed network falls within the range of €0.17-0.22/m^3 over a 30-year payback period. Furthermore, the tool computes, within minutes, an optimal network capable of delivering reclaimed water up to three times more efficiently than current manual planning practices. In addition to its computational efficiency, the tool offers an intuitive user interface, providing a visually appealing map showcasing the optimal reclaimed water network. This includes detailed representations of main and branched networks, complemented by color-coded city clusters for enhanced clarity and comprehension.

Nevertheless, the initial release of the REWATnet tool presented two prominent limitations. Firstly, while the design solutions effectively minimized costs, they often resulted in non-resilient tree-like structures. In such configurations, the failure of a single pipe could disrupt service throughout the entire downstream network, highlighting a significant vulnerability. Secondly, the designs provided by the tool lacked hydraulic verification through dedicated Water Distribution Network (WDN) flow simulation tools, such as Environmental Protection Agency Network Evaluation Tool (EPANET) [89]. Consequently, there remained uncertainty regarding the practical feasibility of these designs—a critical aspect that has been overlooked in much of the existing literature [90, 91]. Despite a substantial body of research on water network resilience and its evaluation, it is apparent that this specific facet lacks comprehensive measures or strategies ensuring resilience through design.

The second iteration of the REWATnet tool, detailed in Section 5.4, represents a

natural progression from its predecessor aimed at addressing its primary limitations. This advancement fills a notable void in the literature by introducing algorithms and metrics specifically tailored to integrate with existing methodologies and tools, thereby providing practical solutions for fortifying the resilience of reclaimed Water Distribution Networks (WDNs). Moreover, the tool now offers designs that have undergone rigorous hydraulic validation through EPANET simulations [92].

Significantly, the REWATnet tool complements the well-established Water Network Tool for Resilience (WNTR) [93, 94], filling a crucial gap that the latter primarily overlooks. While WNTR excels in evaluating resilience, it lacks specific measures for concrete design improvements. In contrast, our algorithms not only assess resilience but also provide cost-effective solutions for enhancing it in two pivotal scenarios: firstly, during the initial network design phase through the Resilience by Design algorithm, and secondly, in optimizing resilience within existing operational networks via the Resilience Strengthening algorithm.

The enhanced REWATnet tool introduces a novel approach to evaluating resilience through the Water Availability (WA) measure. This measure offers a comprehensive assessment while requiring minimal data, an advancement from prevailing practices in the literature [95–100] that often rely on operational data, which is typically scarce during the initial design phases. Additionally, our framework incorporates a broader spectrum of potential failure causes, encompassing pipe-related, environment-related, and operation-related failures [101, 102], in addition to the traditionally considered natural disasters [103, 104]. Furthermore, our approach goes beyond mere assessment, offering specific design improvement measures tailored to enhance network resilience. This aspect distinguishes our work from similar studies such as [99], which lacks accessibility to data or algorithms and automation, relying solely on a designer's experiential judgment and an initial set of fixed pipe diameters.

6.4 Limitations of the study

Although the proposed methodology has been rigorously applied, it is essential to be aware of the potential limitations that may arise from the research, so that they can be considered in future investigations. Below, the main limitations associated with this study are detailed, primarily relating to methodological aspects that invariably influence, to a greater or lesser extent, the final outcome.

The primary challenge in implementing the wastewater sampling approach lies in the quality of data concerning the topology of the sewage network. Often, the

database maintained by network managers is incomplete or inconsistent, necessitating a reconciliation of sewage data, which is one of the most time-consuming aspects. Additionally, logistical constraints arise on the maintenance holes within the network, where installing monitoring sites or sensors may be physically impractical. Furthermore, maintenance holes located in the middle of roads pose significant challenges, as sample collection or sensor installation in such locations could result in substantial transit disruptions. Another minor limitation concerns the concentration of the virus or contaminant requiring analysis, which can be mitigated by employing a hydraulic model to estimate the dilution capacity within the sewer network. Previous efforts in this regard have been documented for SARS-CoV-2 [68] and other pathogens [105, 106].

While analyzing the impact and risks of tree roots and street works on wastewater networks, we have identified a minor limitation in our methodology. Although our approach is effective in diagnosing the influence of external elements such as natural factors and human activities, it focuses on determining exact circumference impact areas [107, 108], which may not fully represent the volume of tree roots. However, we believe that focusing on impact areas simplifies comprehension for decision-makers by providing a clear spatial representation. Additionally, estimating volumes would introduce uncertainty, primarily due to the limited understanding of the volumes of tree roots of different ages and species.

The initial major challenges associated with designing cost-effective reclaimed Water Distribution Networks (WDNs) (Section 5.2) have been effectively addressed in subsequent research, accompanied by the development of innovative algorithms to enhance resilience and ensure hydraulic feasibility (Section 5.4). However, an additional limitation remains to be addressed. While the resilient reclaimed WDN networks now incorporate comprehensive cost analyses spanning a 50-year operational period, they still operate under the assumption that reclaimed water can reach its intended destinations solely via gravity flow from an elevated initial water tank. This assumption necessitates specific altitude differentials in case-study scenarios to ensure water delivery. Notably, while the costs associated with pressure pumping for conveying treated water from the Wastewater Treatment Plant (WWTP) to the initial elevated tank are factored into the supplying WDN, they are not directly integrated into the design considerations of the network itself. Addressing this limitation is crucial for ensuring a more holistic and accurate assessment of the costs and operational requirements associated with implementing reclaimed WDN networks [29, 109]. Lastly, it is worth noting a minor limitation: the widespread non-acceptance of reclaimed water in numerous countries, or its limited acceptance for specific purposes as dictated by regulations. However, it is important to highlight that in the case of Spain, reclaimed water is permitted for all the

purposes discussed in our papers. Nevertheless, the urgency of climate change compels governments to increasingly recognize and authorize the utilization of reclaimed water as a crucial tool in combating its effects.

Chapter 7

Conclusions and future work

This chapter encapsulates the primary findings of this investigation concerning the predefined objectives (Section 7.1). Additionally, it outlines potential avenues for future research endeavors (Section 7.2).

7.1 Conclusions

In this doctoral thesis, we have made significant contributions to the field of urban water network management. Through the integration of interdisciplinary research encompassing computer science, graph theory, and water sciences, we have addressed pressing challenges stemming from the COVID-19 pandemic, environmental preservation efforts, and water scarcity concerns. The results of this thesis pave the way for a more efficient, accessible, and economical application of innovative algorithms, tools, and decision systems. Ultimately, the contributions aim to minimize costs and enhance the resilience of future urban water systems, thus providing sustainability and efficacy in water resource management.

Objective 1. *Development of Advanced Surveillance Techniques*

- Our contributions demonstrate the feasibility of optimally selecting sampling points for SARS-CoV-2 sewage surveillance in cities. This thesis proposes an algorithm for placing a predefined number of monitoring sites to achieve maximum coverage of maintenance holes with minimal interference among them (static sensor placement). Additionally, two other algorithms are proposed for dynamically sampling and analyzing sewage to identify patient zero and hotspots in cities (dynamic sensor placement). These proposed algorithms are based on graph theory methodologies, which draw upon our previous works in the field of

telecommunications.

- In the Girona case study, deploying five static monitoring sites has shown to be the optimal surveillance solution, achieving coverage values of 82% for maintenance holes and 88% for inhabitants without any interference. These results highlight that the proposed algorithms outperform previous proposals across all presented scenarios. Specifically, they demonstrate enhanced efficiency in monitoring SARS-CoV-2 RNA traces at a finer scale.
- This work introduced the first static monitoring site selection algorithm for SARS-CoV-2 monitoring at a fine spatial resolution. The static monitoring site selection algorithm can be effectively applied in municipalities seeking to monitor SARS-CoV-2 RNA traces at a finer scale than the entire municipality. Additionally, a module has been developed to link each household to a maintenance hole, which is essential for aggregating individuals connected to a specific monitoring site.
- The proposed algorithms are equally valid for the placement of monitoring sites for purposes other than tracking the spread of COVID-19, such as estimating the consumption of pharmaceuticals and illicit drugs at fine spatial resolution or detecting illicit discharges of pollutants from industries.

Objective 2. *Optimization of Water Reuse Network Designs*

- The optimization of water reuse networks has enabled the automatic generation of cost-effective network designs, facilitating the development of crucial infrastructure, particularly amidst the current drought emergency in the case study area.
- The newly introduced decision support tool, REclaimed WATer Network Tool (REWATnet), has been successfully applied to two real scenarios in Spain and provides a framework that can be readily adapted and deployed in other regions worldwide. Its implementation only requires the configuration of appropriate open online services and data sources (e.g., cadaster), alongside potential customization of default values for water consumption and costs. No similar tools in the literature offer the capability to plan and economically assess optimal water reclamation networks for cities with minimal computational effort. Additionally, while the tool may be utilized for conventional water distribution networks, it necessitates validation and refinement.
- Urban planners, municipal or regional authorities, and consulting firms can operate the tool to plan urban water reuse projects, pinpoint critical water consumption

areas within a city, compare alternative solutions based on technical and economic criteria, and ultimately select the optimal approach. REWATnet holds significant potential as a valuable dissemination and training tool for water reuse scheme planning.

- With minimal input data from users and leveraging open data sources, the tool calculates the maximum reclaimed water served per unit of investment cost, encompassing network length, pipe diameters, storage tank locations, population served, and construction expenses. In essence, all parameters are seamlessly integrated within a unified architecture as a result of the ReWaT project.

Objective 3. *Automation of Risk Management Processes*

- This thesis results corroborate the initial statement from the Girona municipality that tree roots are a significant cause of pipe failures in both wastewater and drinking water networks. Although the analysis focused solely on the wastewater network due to the unavailability of data for the Drinking Water Distribution Network (WDN), risks from tree roots and potentially other external factors, such as street works, were successfully identified. The Australian Standard (AS) Tree Protection Zone (TPZ) method emerged as the most practical approach for calculating the Impact Areas (IA) of city trees. The findings revealed that although nearly 80% of Girona's wastewater pipes pose no risk, 12% exhibit low risks, 6% moderate risks, and 3% high risks.

- To mitigate pipe risks, the Element Rearrangement (ER) algorithm was introduced, which involves replanting high-impact trees with alternative "sewer-safe" tree species that produce less hazardous root systems. The algorithm proved effective, particularly as the top 10 most impactful trees in Girona contribute the most to pipe risks, with a notable 44% difference in impact values between the first and tenth-placed trees. Implementing the ER algorithm resulted in a considerable reduction in the median pipe failure risk (from 0.8 to 0.14) with a minimal percentage of tree rearrangement (from 1% to 12%). Based on computed pipe failure probabilities, the anticipated cost of pipe repairs due to tree roots over the network's operational period of 30 years is approximately €5.14 million, which can be halved (€2.78 million) with just a 4% implementation of the ER algorithm. Furthermore, the economic benefits of the ER algorithm demonstrate a payback of up to €1.33 million for a 7% rearrangement, despite the required initial investment.

- This approach demonstrates the feasibility of automatically diagnosing the potential impacts of tree roots on wastewater pipes and proposing cost-effective

rearrangement solutions. The Element Rearrangement (ER) algorithm not only mitigates environmental and public health hazards but also yields positive economic returns throughout the operational period of the pipes within optimal rearrangement percentages. These novel algorithms are versatile and can be applied to address other natural and human factors such as street works or building constructions. Additionally, pipe failure probabilities are calculated and utilized to estimate the expected cost of pipe repairs over their operational lifespan.

- The achievement of this objective presents a cost-effective solution for both analyzing and assessing the impact of external factors on wastewater networks and predicting pipe failure risks, all without the necessity of fieldwork in sewers. Despite certain limitations, the method's versatility lies in its global applicability, utilizing existing data available in municipalities. It serves as a valuable preliminary study for prioritizing preventive measures and conducting detailed initial assessments, rendering it particularly beneficial for city councils and wastewater network managers.

Objective 4. *Design of Cost-Effective Resilient Network Solutions*

- This objective represents an evolution from Objective 2, aiming to enhance the optimization of water reuse network designs by ensuring hydraulic feasibility and resilience. In defining resilience, a network is characterized by its ability to minimize service disruptions resulting from pipe failure events, thereby maximizing the number of unaffected destinations.
- Given the challenge of assessing the resilience of one design compared to another, a novel metric termed Water Availability (WA) has been introduced, which offers a comprehensive means to evaluate network resilience. Within water distribution networks, pipe failures typically stem from three main causes: (i) pipe-related factors, (ii) environmental factors, and (iii) operational factors. Since one of the focuses of this thesis is resilience from a network design perspective, only pipe-related attributes can be directly quantified and calculated. These attributes include pipe diameter, age, material composition, length, and wall thickness. As a result, the WA metric provides, at the design phase, a robust approximation of network resilience without requiring environmental or operational-related data. To achieve this, this thesis presents a novel measure of unavailability, incorporating random components and statistical data to account for environmental and operational-related failure causes.
- The innovative and interactive network design process seamlessly integrates

these algorithms with the EPANET software validations, ensuring cost-effective resilience while adhering to hydraulic feasibility constraints in network designs. This holistic approach reshapes the way resilient water distribution systems are evaluated and designed.

- In practical terms, these contributions demonstrate substantial benefits in prioritizing resilience in network design, as evidenced by our case study of Girona. When comparing a resilient-by-design network to a non-resilient counterpart, a 36% reduction was observed in the probability of water service disruption. Notably, this reduction not only significantly decreases the average volume of unserved water per year due to disruption events but also reduces the cost of disruption repairs per kilometer per year by a significant 32% compared to the non-resilient design, highlighting its cost-effectiveness. These findings underscore the long-term advantages of resilience-focused network design. In the case study of Girona, the implementation of resilience measures effectively safeguards up to a substantial 91,700 cubic meters of water per year from the impact of water disruption events compared to an equivalent non-resilient design.
- The cost analysis revealed a 3.67% decrease in costs over a 50-year operational period for the resilient-by-design reclaimed water distribution network (WDN) solution compared to the baseline of using the conventional drinking water distribution network (WDN), emphasizing both its cost-effectiveness and long-term financial benefits. Additionally, the reclaimed WDN solution has the potential to save a remarkable amount of 27,820,150 cubic meters (556,403 cubic meters per year) of clean water, a particularly significant figure given the ongoing drought emergency in our case study region.
- Overall, this thesis not only fills a crucial gap in the literature by focusing on reclaimed water distribution networks but also demonstrates its adaptability to other water distribution network (WDN) systems with minimal input data adaptation. Notably, although resilience is more crucial in drinking WDNs, it is also increasingly crucial in reclaimed WDNs, particularly in the face of water drought emergencies exacerbated by climate change. By providing valuable insights and practical tools for resilient network design, this work significantly advances water resource management and infrastructure planning.

Objective 5. *Integration of Solutions with Open Data Platforms and Applied Research Projects*

- The outcomes of this thesis have been successfully integrated into the applied

research projects CLEaN-Tour and ReUseMP3, which collectively received funding totaling €325,500. Additionally, the achievements of this thesis played a pivotal role in establishing the groundwork for another project, the Reclaimed WaTer networks (ReWaT) Spanish national project, with the code AEI-010500-2023-339. As this project approaches its conclusion and has recently secured funding of 191,437 euros for a one-year duration, bringing together several small and medium-sized companies, plans are underway for a second phase. The invaluable findings from this endeavor highlight the importance of ongoing research efforts. These funding initiatives primarily aim to encourage innovative business groups, thereby enhancing the competitiveness of small and medium-sized companies. In particular, the ReWaT project involves a range of companies, including those focused on sensor technology, auditing, and consulting services.

- A significant portion of the algorithm definitions, implementations, and output indicator results for the case study, including numerical, on-map, and graphical data visualizations, are accessible via dedicated public repositories [63, 64]. All algorithms and data will be readily available to municipalities and network managers, enabling them to leverage the outcomes of this thesis as needed.
- Overall, the integrated solutions presented in this thesis have made significant contributions to the community and society, encompassing various research projects and the promotion of open data initiatives. Furthermore, the methods and algorithm results presented, applied in real-case scenarios, are currently being monitored by the city councils of Girona, Lloret de Mar, and El Prat del Llobregat.

7.2 Future lines of research

This section outlines future research requirements aimed at enhancing the methodological approaches and building upon the contributions made thus far in this thesis:

- In Catalonia, COVID-19 prevalence data is currently aggregated only at the municipality level and at the level of primary health areas (Àrea Bàsica de Salut (ABS)), which are defined around primary care health centers. To illustrate, sensors were positioned in the wastewater network at the end of each ABS (the downstream maintenance hole of each ABS subcatchment) and assessed interference. The resulting findings indicate significant interference among monitoring sites when sensors are placed according to ABS, making it an unfavorable option. Therefore, municipalities should request health authorities to aggregate prevalence data based on other justified areas, such as those identified in this thesis, to enable validation

of optimal approaches with real data.

- It is also noteworthy that future endeavors are currently in their initial stages, intending to utilize the algorithms outlined in this thesis to implement drought emergency and wastewater surveillance measures. The first measure includes establishing controlled water-isolate clusters to save clean water and suspending water service for a specific number of customers during a predetermined timeframe without disrupting the remainder of the network. Additionally, the other ongoing project aims to employ the monitoring site placement algorithm to position sensors for detecting antibiotics in sewage. This initiative seeks to differentiate antibiotic concentrations between more affluent and poorer city sectors. This project is conducted within the framework of the EXPOWASTE project, which integrates human biomonitoring and wastewater-based epidemiology to evaluate exposure to harmful chemicals and biological agents. Notably, this latter project has already successfully identified optimal sensor locations and has positioned sensors in the Girona wastewater network.
- While analyzing the impact and risks of tree roots and street works on wastewater networks, integrating impact volumes instead of solely focusing on circumference impact areas would substantially enhance the precision of the algorithms. In practice, tree roots occupy an irregular but somewhat predictable volume, providing better insights into the potential risks posed to wastewater networks. Moreover, future work will be conducted to consider pipe material in the computation of failure risks which, in practice, affects the failure hazard rate.
- It would also be interesting to verify the tree root risk analysis approach with the actual city wastewater pipe failure records, as this is the first study that quantifies pipe failure risk concerning tree roots at a theoretical level. The obtained data can also be used to improve the existing Artificial Intelligence (AI) algorithm predictions.
- Consideration may be given to adapting the REWATnet tool based on the country, enabling automatic extraction of permitted water uses in a given country to ensure compliance with all regulatory frameworks. This issue may be implemented in the future phase 2 of the ReWaT project.
- Future research will incorporate pressure pumps and O&M costs directly into the reclaimed WDN design phase, enabling the evaluation of additional case studies that require their inclusion in the REWATnet tool. This advancement will further leverage insights from relevant literature, as discussed in Section 6.4, providing

valuable frameworks for integrating these considerations. Additionally, the design solutions could also be compared with decentralized approaches.

- This thesis lays the foundation for tackling more complex problems that have not been explored yet in the literature. One such challenge is the optimal and cost-effective placement of necessary sensors to guarantee water quality in reclaimed WDNs. While the proliferation of the Internet of Things and smart cities has made complex and automatic sensorization a reality, there remains a pressing need for strategies that balance effective monitoring and resilient design while minimizing costs. Therefore, it is proposed that addressing this unresolved issue can be achieved through the application of principles of graph theory, a method whose efficacy in solving future urban water management-related issues has been demonstrated in this thesis.

Bibliography

[1] Bergkamp, G.; Diphoorn, B.; Trommsdorf, C. Water and development in the urban setting. *2015). Water for Development–Charting a Water Wise Path. Report* **2015**.

[2] Marlow, D. R.; Moglia, M.; Cook, S.; Beale, D. J. Towards sustainable urban water management: A critical reassessment. *Water research* **2013**, *47*, 7150–7161.

[3] Crits-Christoph, A.; Kantor, R. S.; Olm, M. R.; Whitney, O. N.; Al-Shayeb, B.; Lou, Y. C.; Flamholz, A.; Kennedy, L. C.; Greenwald, H.; Hinkle, A., et al. Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. *MBio* **2021**, *12*, 10–1128.

[4] Sharma, D. K.; Nalavade, U. P.; Kalgutkar, K.; Gupta, N.; Deshpande, J. M. SARS-CoV-2 detection in sewage samples: Standardization of method & preliminary observations. *Indian Journal of Medical Research* **2021**, *153*, 159–165.

[5] Lenzen, M. et al. Global socio-economic losses and environmental gains from the coronavirus pandemic. *PLoS ONE* **2020**, *15*, 1–13, DOI: 10.1371/journal.pone.0235654.

[6] Mallapaty, S. How sewage could reveal true scale of coronavirus outbreak. *Nature* **2020**, *580*, 176–177.

[7] Medema, G.; Been, F.; Heijnen, L.; Petterson, S. Implementation of environmental surveillance for SARS-CoV-2 virus to support public health decisions: Opportunities and challenges. *Current Opinion in Environmental Science and Health* **2020**, *17*, 49–71, DOI: 10.1016/j.coesh.2020.09.006.

[8] Schmidt, C. Watcher in the wastewater. *Nature biotechnology* **2020**, *38*, 917–920, DOI: 10.1038/s41587-020-0620-2.

[9] Masi, F.; Langergraber, G.; Santoni, M.; Istenič, D.; Atanasova, N.; Buttiglieri, G. In *Wastewater treatment and Reuse – Present and future perspectives in technological developments and management issues*, Verlicchi, P., Ed.; Advances in Chemical Pollution, Environmental Management and Protection, Vol. 5;

Elsevier: 2020, pp 145–187, DOI: https://doi.org/10.1016/bs.apmp.2020.07.004.

[10] Chrispim, M. C.; Nolasco, M. A. Greywater treatment using a moving bed biofilm reactor at a university campus in Brazil. *Journal of Cleaner Production* **2017**, *142*, 290–296, DOI: 10.1016/j.jclepro.2016.07.162.

[11] Domnech, L.; Saurí, D. Socio-technical transitions in water scarcity contexts: Public acceptance of greywater reuse technologies in the Metropolitan Area of Barcelona. *Resources, Conservation and Recycling* **2010**, *55*, 53–62, DOI: 10.1016/j.resconrec.2010.07.001.

[12] Vallès-Casas, M.; March, H.; Saurí, D. Decentralized and User-Led Approaches to Rainwater Harvesting and Greywater Recycling: The Case of Sant Cugat del Vallès, Barcelona, Spain. *Built Environment* **2016**, *42*, 243–257.

[13] Salman, B.; Salem, O. Risk assessment of wastewater collection lines using failure models and criticality ratings. *Journal of pipeline systems engineering and practice* **2012**, *3*, 68–76.

[14] Laakso, T.; Ahopelto, S.; Lampola, T.; Kokkonen, T.; Vahala, R. Estimating water and wastewater pipe failure consequences and the most detrimental failure modes. *Water Science and Technology: Water Supply* **2018**, *18*, 901–909.

[15] Nazemi, A.; Madani, K. Urban water security: Emerging discussion and remaining challenges. *Sustainable Cities and Society* **2018**, *41*, 925–928.

[16] Ahmad, T.; Shaban, I.; Zayed, T. In *2nd International Conference on Civil Infrastructure and Construction (CIC 2023)*, 2023.

[17] Calle, E.; Martínez, D.; Buttiglieri, G.; Corominas, L.; Farreras, M.; Saló-Grau, J.; Vilà, P.; Pueyo-Ros, J.; Comas, J. Optimal design of water reuse networks in cities through decision support tool development and testing. *NPJ Clean Water* **2023**, *6*, JCR Impact Factor (IF): **11.4, Q1** (Water Science and Technology), 23, DOI: https://doi.org/10.1038/s41545-023-00222-4.

[18] Rice, J.; Wutich, A.; White, D. D.; Westerhoff, P. Comparing actual de facto wastewater reuse and its public acceptability: A three city case study. *Sustainable Cities and Society* **2016**, *27*, 467–474.

[19] Luiijf, E.; Klaver, M. Analysis and lessons identified on critical infrastructures and dependencies from an empirical data set. *International Journal of Critical Infrastructure Protection* **2021**, *35*, 100471.

[20] Van Leuven, L. J. Water/wastewater infrastructure security: Threats and vulnerabilities. *Handbook of water and wastewater systems protection* **2011**, 27–46.

[21] Monitoring waves of the COVID-19 pandemic: Inferences from WWTPs of different sizes. *Science of the Total Environment* **2021**, *787*, 147463, DOI: 10.1016/j.scitotenv.2021.147463.

[22] Christodoulou, S.; Agathokleous, A.; Charalambous, B.; Adamou, A. Proactive risk-based integrity assessment of water distribution networks. *Water resources management* **2010**, *24*, 3715–3730.

[23] Obradović, D. The impact of tree root systems on wastewater pipes. *Zajednički temelji'17: zbornik radova* **2017**, 65–71.

[24] Kumar, S.; Singh, R.; Maurya, N. S. Modelling of corrosion rate in the drinking water distribution network using Design Expert 13 software. *Environmental Science and Pollution Research* **2023**, *30*, 45428–45444.

[25] Sydney Water How you can help stop blockages: Tree planting and the wastewater system. *Sydney Water* **2017**.

[26] Hartley, M. Tree root damage to pipes. *The Arborist Network* **2012**.

[27] Ramos, H. M.; Carravetta, A.; Nabola, A. M. New challenges in water systems, 2020.

[28] Agency, E. E. European waters - Assessment of status and pressures 2018. *European waters - Assessment of status and pressures* **2018**, *7*, DOI: 10.2800/303664.

[29] Khurelbaatar, G.; Al Marzuqi, B.; Van Afferden, M.; Müller, R.; Friesen, J. Data Reduced Method for Cost Comparison of Wastewater Management Scenarios–Case Study for Two Settlements in Jordan and Oman. Front. *Environ. Sci* **2021**, *9*, 626634.

[30] Gross, J. L.; Yellen, J.; Anderson, M., *Graph theory and its applications*; Chapman and Hall/CRC: 2018.

[31] GraphML The GraphML File Format, http://graphml.graphdrawing.org/, [Online; accessed 21-dec-2020], 2001.

[32] Balakrishnan, R.; Ranganathan, K., *A textbook of graph theory*; Springer Science & Business Media: 2012.

[33] Hamilton, W. L., *Graph representation learning*; Morgan & Claypool Publishers: 2020.

[34] Carrano, F. M.; Prichard, J. J., *Data abstraction and problem solving with Java: walls and mirrors*; Pearson/Addison Wesley: 2004.

[35] Rueda, D. F.; Calle, E.; Marzo, J. L. Robustness comparison of 15 real telecommunication networks: Structural and centrality measurements. *Journal of Network and Systems Management* **2017**, *25*, 269–289.

[36] Zhang, D.-g.; Tang, Y.-m.; Cui, Y.-y.; Gao, J.-x.; Liu, X.-h.; Zhang, T. Novel reliable routing method for engineering of internet of vehicles based on graph theory. *Engineering Computations* **2018**, *36*, 226–247.

[37] Arat, F.; Akleylek, S. Modified graph-based algorithm to analyze security threats in IoT. *PeerJ Computer Science* **2023**, *9*, e1743.

[38] Liu, W.; Sidhu, A.; Beacom, A. M.; Valente, T. W. Social network theory. *The international encyclopedia of media effects* **2017**, 1–12.

[39] Koutrouli, M.; Karatzas, E.; Paez-Espino, D.; Pavlopoulos, G. A. A guide to conquer the biological network era using graph theory. *Frontiers in bioengineering and biotechnology* **2020**, *8*, 34.

[40] Agarwal, N.; Seth, N.; Agarwal, A. Evaluation of supply chain resilience index: a graph theory based approach. *Benchmarking: An International Journal* **2022**, *29*, 735–766.

[41] Calle, E.; Cosgaya, S. G.; Martínez, D.; Pióro, M. In *11th International Workshop on Resilient Networks Design and Modeling (RNDM)*, 2019, pp 1–7.

[42] Manzano, M.; Marzo, J.; Calle, E.; Manolovay, A. In *17th European Conference on Networks and Optical Communications (EuCNC)*, 2012, pp 1–6.

[43] Marzo, J. L.; Martinez, D.; Bergillos, S.; Calle, E. In *2022 18th International Conference on the Design of Reliable Communication Networks (DRCN)*, 2022, pp 1–4.

[44] Calle, E.; Martínez, D.; Mycek, M.; Pióro, M. Resilient backup controller placement in distributed SDN under critical targeted attacks. *International Journal of Critical Infrastructure Protection* **2021**, *33*, 100422.

[45] Kesavan, H.; Chandrashekar, M. Graph-theoretic models for pipe network analysis. *Journal of the Hydraulics Division* **1972**, *98*, 345–364.

[46] Villez, K.; Vanrolleghem, P. A.; Corominas, L. Optimal flow sensor placement on wastewater treatment plants. *Water Research* **2016**, *101*, 75–83, DOI: 10.1016/j.watres.2016.05.068.

[47] Villez, K.; Vanrolleghem, P. A.; Corominas, L. A general-purpose method for Pareto optimal placement of flow rate and concentration sensors in networked systems – With application to wastewater treatment plants. *Computers and Chemical Engineering* **2020**, *139*, 106880, DOI: 10.1016/j.compchemeng.2020.106880.

[48] Ahmadullah, R.; Dongshik, K. Designing of Hydraulically Balanced Water Distribution Network Based on GIS and EPANET. *International Journal of Advanced Computer Science and Applications* **2016**, *7*, 118–125, DOI: 10.14569/ijacsa.2016.070216.

[49] Meng, F.; Fu, G.; Farmani, R.; Sweetapple, C.; Butler, D. Topological attributes of network resilience: A study in water distribution systems. *Water research* **2018**, *143*, 376–386.

[50] Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008**, *2008*, P10008.

[51] Mehlhorn, K. A faster approximation algorithm for the Steiner problem in graphs. *Information Processing Letters* **1988**, *27*, 125–128.

[52] vanRossum, G. Python reference manual. *Department of Computer Science [CS]* **1995**.

[53] Bennett, J., *OpenStreetMap*; Packt Publishing Ltd: 2010.

[54] Boeing, G. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* **2017**, *65*, 126–139.

[55] Instituto Geográfico Nacional Digital Elevation Model provider for Spain. `https://www.ign.es/`, Accessed: 2023-30-05, 2023.

[56] Pezoa, F.; Reutter, J. L.; Suarez, F.; Ugarte, M.; Vrgoč, D. In *Proceedings of the 25th International Conference on World Wide Web*, 2016, pp 263–273.

[57] Olbricht, R. et al. Overpass API, openstreetmap.org, 2011.

[58] Gillies, S. et al. Shapely: manipulation and analysis of geometric objects, toblerity.org, 2007.

[59] Hagberg, A. A.; Swart, P. J.; S Chult, D. A. NetworkX: High productivity software for complex networks, `https://networkx.org/`, Accessed: March 26, 2024, 2008.

[60] Butler, S. GeoJSON, `https://github.com/frewsxcv/python-geojson`, Accessed: March 26, 2024, 2008.

[61] QGIS A Free and Open Source Geographic Information System, `https://www.qgis.org/en/site/`, [Online; accessed 21-dec-2020], 2008.

[62] Crickard III, P., *Leaflet. js essentials*; Packt Publishing Ltd: 2014.

[63] Martínez, D. Extended version of the algorithms and results of the research work "Mitigating Wastewater Leakages for Enhanced Network Protection: Risk Assessment and Practical Solutions." `https://doi.org/10.5281/zenodo.7704666`, version 1.0.0, 2023, DOI: `10.5281/zenodo.7704666`.

[64] Martínez, D. Extended version of the algorithms and results of the research work "Enhancing resilience of water distribution networks with cost-effective

meshing." https://doi.org/10.5281/zenodo.8398703, version 1.0.0, 2023, DOI: 10.5281/zenodo.8398703.

[65] Catalan Water Agency Status of water reserves in reservoirs, https://github.com/frewsxcv/python-geojson, Accessed: March 28, 2024, 2024.

[66] Catalan Statistics Institute Girona population, https://www.idescat.cat/emex/?id=170792, [Online; accessed 21-dec-2021], 2019.

[67] Lloret Tourism Press Office Status of water reserves in reservoirs, https://professionals.lloretdemar.org/estadistiques/lloret-en-xifres/, Accessed: March 28, 2024, 2024.

[68] Hart, O. E.; Halden, R. U. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. *Science of the Total Environment* **2020**, *730*, 138875, DOI: 10.1016/j.scitotenv.2020.138875.

[69] Kang, O. Y.; Lee, S. C.; Wasewar, K.; Kim, M. J.; Liu, H.; Oh, T. S.; Janghorban, E.; Yoo, C. K. Determination of key sensor locations for non-point pollutant sources management in sewer network. *Korean Journal of Chemical Engineering* **2013**, *30*, 20–26, DOI: 10.1007/s11814-012-0108-y.

[70] Yazdi, J. Water quality monitoring network design for urban drainage systems, an entropy method. *Urban Water Journal* **2018**, *15*, 227–233, DOI: 10.1080/1573062X.2018.1424215.

[71] Sambito, M.; Di Cristo, C.; Freni, G.; Leopardi, A. Optimal water quality sensor positioning in urban drainage systems for illicit intrusion identification. *Journal of Hydroinformatics* **2020**, *22*, 46–60, DOI: 10.2166/hydro.2019.036.

[72] Vonach, T.; Tscheikner-Gratl, F.; Rauch, W.; Kleidorfer, M. A heuristic method for measurement site selection in sewer systems. *Water (Switzerland)* **2018**, *10*, 1–16, DOI: 10.3390/w10020122.

[73] SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases. *mSystems* **2020**, *5*, DOI: 10.1128/msystems.00614-20.

[74] Escolà Casas, M.; Schröter, N. S.; Zammit, I.; Castaño-Trias, M.; Rodriguez-Mozaz, S.; Gago-Ferrero, P.; Corominas, L. Showcasing the potential of wastewater-based epidemiology to track pharmaceuticals consumption in cities: Comparison against prescription data collected at fine spatial resolution. *Environment International* **2021**, *150*, 106404, DOI: 10.1016/j.envint.2021.106404.

[75] Spatio-temporal assessment of illicit drug use at large scale: evidence from 7 years of international wastewater monitoring. *Addiction* **2020**, *115*, 109–120, DOI: 10.1111/add.14767.

[76] Macedon Ranges Shire Council Tree protection guidelines for developments, https://www.mrsc.vic.gov.au/files/assets/public/live-amp-work/environment/tree-protection-guidelines.pdf, 2022.

[77] AS 4970-2009 *Protection of trees on development sites*; Standard; Standards Australia, 2009.

[78] Moore, G. In *19th National Treenet Symposium*, 2018.

[79] Day, S. D.; Wiseman, P. E.; Dickinson, S. B.; Harris, J. R. Contemporary concepts of root system architecture of urban trees. *Arboriculture & Urban Forestry* **2010**, *36*, 149–159.

[80] Östberg, J.; Martinsson, M.; Stål, Ö.; Fransson, A.-M. Risk of root intrusion by tree and shrub species into sewer pipes in Swedish urban areas. *Urban Forestry & Urban Greening* **2012**, *11*, 65–71.

[81] Kuliczkowska, E.; Parka, A. Management of risk of tree and shrub root intrusion into sewers. *Urban forestry & urban greening* **2017**, *21*, 1–10.

[82] Castillo, A.; Porro, J.; Garrido-Baserba, M.; Rosso, D.; Renzi, D.; Fatone, F.; Gómez, V.; Comas, J.; Poch, M. Validation of a decision support tool for wastewater treatment selection. *Journal of Environmental Management* **2016**, *184*, 409–418.

[83] Poch, M.; Comas, J.; Rodríguez-Roda, I.; Sànchez-Marrè, M.; Cortés, U. Designing and building real environmental decision support systems. *Environmental Modelling & Software* **2004**, *19*, Environmental Sciences and Artificial Intelligence, 857–873, DOI: https://doi.org/10.1016/j.envsoft.2003.03.007.

[84] Chhipi-Shrestha, G.; Hewage, K.; Sadiq, R. Fit-for-purpose wastewater treatment: Conceptualization to development of decision support tool (I). *Science of the Total Environment* **2017**, *607*, 600–612.

[85] Sadr, S. M.; Saroj, D. P.; Mierzwa, J. C.; McGrane, S. J.; Skouteris, G.; Farmani, R.; Kazos, X.; Aumeier, B.; Kouchaki, S.; Ouki, S. K. A multi expert decision support tool for the evaluation of advanced wastewater treatment trains: A novel approach to improve urban sustainability. *Environmental science & policy* **2018**, *90*, 1–10.

[86] Van Afferden, M.; Cardona, J. A.; Müller, R. A.; Lee, M. Y.; Subah, A. A new approach to implementing decentralized wastewater treatment concepts. *Water*

Science and Technology **2015**, *72*, 1923–1930, DOI: 10.2166/wst.2015.393.

[87] Khurelbaatar, G.; van Afferden, M.; Ueberham, M.; Stefan, M.; Geyler, S.; Müller, R. Management of Urban Stormwater at Block-Level (MUST-B): A New Approach for Potential Analysis of Decentralized Stormwater Management Systems. Water 2021, 13, 378, 2021.

[88] Del Teso, R.; Gómez, E.; Estruch-Juan, E.; Cabrera, E. Topographic Energy Management in Water Distribution Systems. *Water Resources Management* **2019**, *33*, 4385–4400, DOI: 10.1007/s11269-019-02375-9.

[89] U.S. EPA EPANET 2.0: Application for Modeling Drinking Water Distribution Systems, https://www.epa.gov/water-research/epanet, Accessed: 2022-02-08, 2000.

[90] Yazdani, A.; Otoo, R. A.; Jeffrey, P. Resilience enhancing expansion strategies for water distribution systems: A network theory approach. *Environmental Modelling & Software* **2011**, *26*, 1574–1582.

[91] Herrera, M.; Abraham, E.; Stoianov, I. A graph-theoretic framework for assessing the resilience of sectorised water distribution networks. *Water Resources Management* **2016**, *30*, 1685–1699.

[92] Rossman, L. A. et al. EPANET 2: users manual, US Environmental Protection Agency. Office of Research and Development, 2000.

[93] Klise, K. A.; Murray, R.; Haxton, T. In *1st International WDSA / CCWI 2018 Joint Conference*, 2018.

[94] Klise, K. A.; Bynum, M.; Moriarty, D.; Murray, R. A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. *Environmental modelling & software* **2017**, *95*, 420–431.

[95] Christodoulou, S.; Fragiadakis, M.; Agathokleous, A.; Xanthos, S., *Urban water distribution networks: assessing systems vulnerabilities, failures, and risks*; Butterworth-Heinemann: 2017.

[96] Cimellaro, G.; Tinebra, A.; Renschler, C.; Fragiadakis, M. New resilience index for urban water distribution networks. *Journal of Structural Engineering* **2016**, *142*, C4015014.

[97] Zhao, X.; Chen, Z.; Gong, H. Effects comparison of different resilience enhancing strategies for municipal water distribution network: A multidimensional approach. *Mathematical problems in engineering* **2015**, *2015*.

[98] Baños, R.; Reca, J.; Martínez, J.; Gil, C.; Márquez, A. L. Resilience indexes for water distribution network design: a performance analysis under demand uncertainty. *Water resources management* **2011**, *25*, 2351–2366.

[99] Todini, E. Looped water distribution networks design using a resilience index based heuristic approach. *Urban water* **2000**, *2*, 115–122.

[100] Tumula, P.; Park, N. Multi-Objective genetic algorithms for the design of pipe networks. *Journal of Water Resources Planning and Management* **2004**, *130*, 73–82.

[101] Taiwo, R.; Shaban, I. A.; Zayed, T. Development of sustainable water infrastructure: A proper understanding of water pipe failure. *Journal of Cleaner Production* **2023**, 136653.

[102] Zhang, Z.; Feng, X.; Qian, F. Studies on resilience of water networks. *Chemical Engineering Journal* **2009**, *147*, 117–121.

[103] McAllister, T. P. Community resilience planning guide for buildings and infrastructure systems, volume I. *Special Publication (NIST SP)* **2015**.

[104] Chang, S. E.; Shinozuka, M. Measuring improvements in the disaster resilience of communities. *Earthquake spectra* **2004**, *20*, 739–755.

[105] Ranta, J.; Hovi, T.; Arjas, E. Poliovirus surveillance by examining sewage water specimens: Studies on detection probability using simulation models. *Risk Analysis* **2001**, *21*, 1087–1096, DOI: 10.1111/0272-4332.t01-1-216174.

[106] Wang, Y.; Moe, C. L.; Dutta, S.; Wadhwa, A.; Kanungo, S.; Mairinger, W.; Zhao, Y.; Jiang, Y.; Teunis, P. F. Designing a typhoid environmental surveillance study: A simulation model for optimum sampling site allocation. *Epidemics* **2020**, *31*, 100391, DOI: https://doi.org/10.1016/j.epidem.2020.100391.

[107] Smith, A.; Astrup, R.; Raumonen, P.; Liski, J.; Krooks, A.; Kaasalainen, S.; Åkerblom, M.; Kaasalainen, M. Tree root system characterization and volume estimation by terrestrial laser scanning and quantitative structure modeling. *Forests* **2014**, *5*, 3274–3294.

[108] Lantini, L.; Tosti, F.; Giannakis, I.; Zou, L.; Benedetto, A.; Alani, A. M. An enhanced data processing framework for mapping tree root systems using ground penetrating radar. *Remote Sensing* **2020**, *12*, 3417.

[109] Friesen, J.; Sanne, M.; Khurelbaatar, G.; van Afferden, M. "OCTOPUS" principle reduces wastewater management costs through network optimization and clustering. *One Earth* **2023**, *6*, 1227–1234.