Journal of Hydroinformatics

Journal of Hydroinformatics Vol 25 No 5, 1927 doi: 10.2166/hydro.2023.092

Methodology for the study of the traceability of runoff water feeding reservoirs

Lidia M. Ortega [a,*], M. Isabel Ramos [b], Carlos Enríquez Turiño [b] and Juan José Cubillas Mercado [c]

[a] Department of Computer Science, University of Jaén, Paraje Las lagunillas, A3-136, Jaén 23071, Spain
[b] Department of Cartographic, Geodetic and Photogrammetric Engineering, University of Jaén, Paraje Las lagunillas, Jaén 23071, Spain
[c] Department of Information and Communication Technology applied to Education, International University of la Rioja, Av. de la Paz, 137, 26006 Logroño, Spain
*Corresponding author. E-mail: lidia@ujaen.es

LMO, 0000-0002-7320-7382; MIR, 0000-0002-3006-4166; CE, 0000-0002-0070-0133; JJ, 0000-0002-5138-8872

ABSTRACT

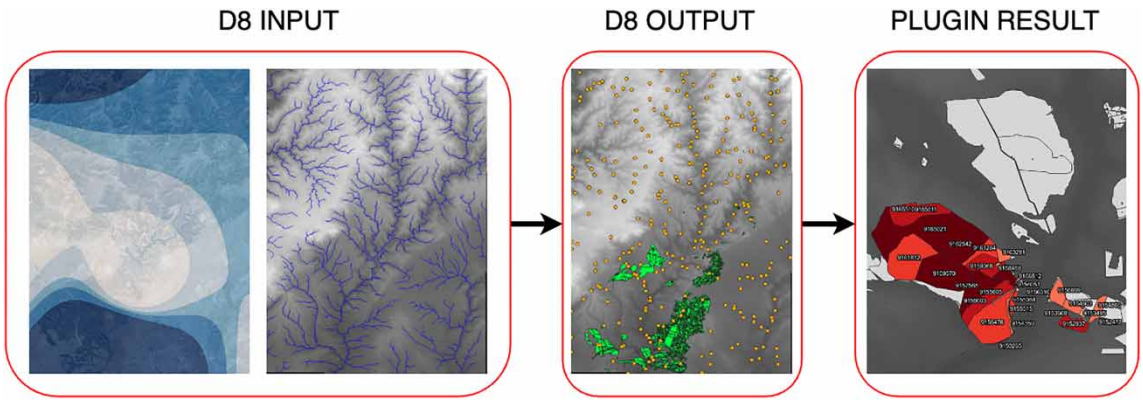
Water reservoirs are essential to ensure water supply to both the population and agriculture, especially in the Mediterranean basin. In some cases, analyses of water intended for human consumption have detected high levels of agrochemicals. Knowing the possible origin of these products is complex because there may be many agricultural plots within the reservoir basin. In this paper, we introduce a methodology to obtain the set of agricultural plots whose rainwater reaches the reservoir and in what proportion they affect the points where chemical analyses are performed. The method implements an extension of the D8 algorithm for the calculation of the drainage network, in which additional information about the land-use type of the area, as well as rainfall maps, are also considered. In order to facilitate the user's analysis of the data, a plugin has been implemented in QGIS. This allows usability and easy interaction with the visual information. The Rumblar reservoir basin, located in Andalusia (Spain) has been studied as a use case, surrounded by olive orchards. The result is a replicable methodology for any other water reservoir and for carrying out an individualized study of agricultural plots.

Key words: D8 algorithm, olive orchards, QGIS plugin, rainfall maps, reservoir basin, water traceability

HIGHLIGHTS

- Water reaching a reservoir is of great environmental interest.
- Farms that contribute the most water to the reservoir can be found out.
- Modification of the D8 algorithm allows adding new functionality to the drainage network creation process.
- The methodology allows for the addition of real information such as land use type and rainfall maps.
- The end-users can use an interactive tool implemented as QGIS plugin.

GRAPHICAL ABSTRACT



1. INTRODUCTION

In the countries of the Mediterranean basin, such as Spain, many municipalities are supplied with drinking water stored in reservoirs. In this area, rainfall is scarce and mostly concentrated in autumn and spring. In general terms, the pluviometry is very changeable, with particularly dry periods that seriously affect the agricultural sector. Thanks to the reservoir infrastructure, the supply to the citizens has been assured so far. However, not only is the quantity of water of concern, but also its quality indexes.

For this work, we have studied a specific case, the Rumblar reservoir, located in the province of Jaén (Spain), which supplies freshwater to about 88,200 inhabitants belonging to the Rumblar Consortium. Company technicians in charge of the reservoir have observed that, after a period of rainfall, chemical analysis reveals a higher concentration of nitrates and other unwanted components. These substances come mostly from olive groves, which are predominant in the entire river basin of the area. This situation is aggravated when it rains after a long period of drought. At the beginning of the year 2023, the reservoir is at only 10% of its capacity. This water scarcity increases the concentration of fertilizers and phytosanitary products from agricultural activities.

This paper is focused on developing a methodology for the traceability of the water (and consequently of entrained chemicals) that reaches a reservoir, in order to determine their possible origin considering the cultivation areas of the basin that supplies water to the reservoir. We have focused our study on the previously described area of the Rumblar, but the methodology can be extended to any other reservoir or rush of water whose watershed is affected by agricultural activity. Preliminary studies carried out using Geographic Information Systems (GIS) tools do not allow a true traceability of the origin of these discharges, as they use only the digital elevation model (DEM) as input data. Nor have we obtained results using simulation tools even though realistic parameters are taken into account, since they do not allow customizing input files to better adapt the simulation to the realities of the area. As a consequence of the above, we have developed our own method based on the calculation of the drainage network. In this way, we have been able to control the entire path of rainwater throughout the basin that drains into the reservoir while considering its origin and the land use type that is carried out. Although the study area is not very large (the reservoir basin), rainfall maps have been considered as input data. These maps show certain differences in rainfall over a given period of time. We have also added parameters related to soil absorption and the orography of the terrain.

The results are promising since it is possible to obtain both the farms from which the water comes and the percentages that affect the sample. In this way, water samples taken in nearby runoffs can know the most probable origin and thus make a follow-up of the type of agricultural activity they carry out. This makes it possible to focus the study on certain sites to determine whether they are performing this activity in accordance with current regulations.

The paper is structured as follows. Section 2 elaborates on this topic and describes similar solutions in the literature. In Section 3, we describe the work area and the input data needed to run the process. The methodology carried out to obtain water traceability is described in Section 4. In terms of application usability, a plugin has been implemented, as introduced in Section 5. Section 6 analyzes the results and usability of the QGIS plugin. Section 7 shows conclusions and future work, and finally, Section 8 provides links to data and source code repositories.

2. STATE OF THE ART

Soil, along with air and water, is one of the environmental resources that most influences the environment and biodiversity. Soil conditions are mainly affected by the use made of it by human beings, especially in agriculture and livestock farming. In recent decades, the use of chemicals in agriculture has become widespread for various reasons, such as increasing soil fertility and eliminating pests and weeds. As a result, some areas have been affected by an excess of fertilizers, pesticides and herbicides (Lizaga *et al.* 2020).

In particular, nitrogen is one of the essential elements for plant growth. Nitrates are the natural result of atmospheric nitrogen fixation and the decomposition of organic matter by microorganisms. Low concentrations of nitrates are found naturally in surface and groundwater; however, the amounts of nitrogen deposited in the soil are due to human activities that contribute to elevated concentrations in the aquatic environment (Min *et al.* 2002; Yue *et al.* 2020; Kim *et al.* 2023; Qiu *et al.* 2023). Fertilizers are one of the main anthropogenic sources of nitrogen, used to improve growth and increase crop production. Due to its importance and the trans-boundary nature of nitrate pollution in Europe, Council Directive 91/676/EEC of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources was adopted as

early as 1991.[1] However, it is noted that progress has been made in water protection, but that certain obligations need to be clarified or additional measures put in place in order to improve this objective.

The Spanish government has recently approved a national regulation (BOE 2022) for the protection of water against diffuse pollution caused by nitrates from agriculture. In this sense, the European Union and the United Nations are carrying out initiatives to promote efficient soil management to achieve the great challenge of increasing food production while minimizing soil degradation through sustainable development plans. According to the European Commission's proposals, sustainable soil management will be a crucial component of several objectives in the coming years, in particular those focusing on landscapes and biodiversity, natural resources and climate action. Due to its starting in 2023, the new common agricultural policy for 2023–2027 (CAP 2021) will include stronger support for healthy soil, in line with the goals of the European Green Deal. As indicated in the regulatory reports, the problem of the extensive use of these chemicals is both in their passage into the food chain and also into aquifers or water sources. This requires a thorough knowledge of the agricultural plots and their impact along the drainage basin.

In short, addressing the challenge of efficient land and water management requires action on several fronts, including the ICT (Information and Communication Technologies) capabilities. The European Commission is determined to make this Europe's 'Digital Decade'. The Global Soil Partnership (GSP) of the Food and Agriculture Organization of the United Nations (FAO) has established as one of its objectives (Pillar 4): *Improving the quantity and quality of soil data and information: data collection (generation), analysis, validation, reporting, monitoring and integration with other disciplines* (FAO-ITPS 2020). In general terms, this involves the continuous processes of data collection from the physical sites, their storage in specific information systems and subsequent analysis to reach conclusions and perform concrete actions.

Depending on the specific problem to be addressed, diverse types of data can be captured, ranging from the chemical and biological properties of the soil, the terrain orography or the land use, whether agricultural or forestry. This heterogeneous information must be processed considering the geospatial nature of these variables that contribute to correlating crop inventory and water management (Mitran *et al.* 2021).

We are interested in monitoring not only the soil but also the water that flows through these territories and eventually reaches the water reservoirs that serve both for agricultural irrigation and human consumption. The scarcity and pollution of freshwater throughout the planet is a recurring concern, especially in the Mediterranean basin. It also should be considered that part of the rainwater that falls infiltrates into the ground, filling pores and fissures which, when saturated, cause the water to flow by gravity towards springs, rivers or seas, giving rise to underground runoff.

Water quality is measured with different indexes and techniques. Normally a chemical analysis is carried out to detect different types of substances in the water. However, this is often time-consuming, costly and gives the information on a specific date and position. In order to provide water analysis over wide areas and over time, multispectral satellite images are being used for mapping water quality considering parameters such as optical clarity or colored dissolved organic matter (Griffin *et al.* 2018; Sufia Sultana & Dewan 2021). The problem with satellite images is that sometimes the frequency or intervals of capture does not correspond with those required. For this reason, some authors choose to use other types of remote capture mechanisms using airplanes and Unmanned Aerial Vehicles (UAVs) (Sibanda *et al.* 2021; Isgró *et al.* 2022).

In any case, this tells us whether or not there is deterioration in the quality of the water and the type of substance that affects it, but not its most likely origin. In order to carry out this type of study, it is necessary to know the topography of the terrain and the land use throughout the basin that discharges water into the reservoir. Considering that chemicals are washed away by water, watersheds, streams and rivers, the drainage network represents an important data structure for modeling the distribution and movement of surface water over the terrain. The most important feature used in this case is the cartographic representation of the surface given by the DEM.

Traditionally, this and similar studies have been carried out using GIS tools. In particular, it is possible to find both free and commercial software performing hydrological analysis. Some examples are ESRI's ArcGIS, QGIS or Saga, which include standard tools to extract the stream segments and basin watersheds from DEMs under the premise that water flow is governed by the terrain topography. The drainage system of a particular drainage basin provides the patterns formed by rivers and streams which tend to discharge water into major rivers or marshes or lakes. Along these paths, the water carries soil and the organic and inorganic substances it contains.

[1] https://www.eea.europa.eu/policy-documents/council-directive-91-676-eec

The hydrological algorithms supported by GIS provide different outputs such as drainage basins and networks (Jasiewicz & Metz 2011). However, the solutions they provide are conditioned by pre-established parameters and variables that in some cases are not adapted to the reality of the characteristics of the work area or the type of analysis to be carried out (Li 2014). Often, the models of certain software are better adapted to some specific conditions of the study area and poorly adapted to others. This is shown by works such as Astagneau *et al.* (2021), which review the advantages and disadvantages of using hydrological models implemented within packages in the R environment and compare and evaluate their applicability and usability. Most of the algorithms implemented in GIS work well with satellite images, such as Landsat, where the spatial resolution is low (Still & Shih 1985). However, when a greater level of detail is required, they need input files including the drainage network but without any control over the calculation process workflow.

A hydrological model is often needed to analyze spatially variable hydrologic behavior. Such a model can be difficult to set up, especially for an ungauged basin, as it demands a massive amount of data. Moreover, there is an additional challenge of selecting a proper grid resolution, which generally leads to predictive uncertainty and also directly determines the amount of work required. In this sense, satellite Earth observations are an accurate and reliable source of data. Their increasing spatial and temporal resolution, as well as their availability, make them attractive for hydrological modeling. Data such as precipitation, evaporation, soil moisture or snow depth obtained by satellite are incorporated into the structure of the hydrological model through a data assimilation scheme (Alfieri *et al.* 2022). This type of model, although very effective, nevertheless requires multidisciplinary and not merely geometric knowledge of the territory, which is not always feasible. A model that integrates all these parameters but that can be inserted in a user-friendly way is one of the challenges to be met in the work presented here.

In order to generate the drainage network, traditional methods use the DEM as the only input data. That is, they are based solely on the geometry of the terrain and not on other types of factors such as the absorption capacity of the soil. Neither do they take into account the possible sources of groundwater. Additionally, the solution is sometimes a set of 2D maps and data tables that need to be interpreted by qualified personnel. Breaking the abstraction barrier through user-friendly interfaces is important when actions are to be carried out by farmers and technicians without GIS training.

On the other hand, fluid simulation software could also provide interesting results to this monitoring problem. However, these are also closed tools in which the process cannot be manipulated at each iteration of the algorithm. In addition, they do not provide real-time response when handling large areas, such as a real river basin. On the contrary, they are able to generate realistic 3D visualization and simulation, which is well received because it is easily understood by authorities and non-technical personnel (Ramos *et al.* 2022).

For all of the above, many researchers have opted for specific implementations to control those aspects of the hydrological process. ICT techniques are present in many different forms. Statistical analysis is increasingly used, but mostly to correlate different hydrochemical parameters (Yue *et al.* 2020). For instance by means of Bayesian models on the related information about chemicals present in water and the land-use types (Kim *et al.* 2023). The source of data is normally soil and water sampling. In any case, this implies chemical analysis in many different locations which does not give direct information about its possible origin. The methodology proposed in this work helps to perform a more efficient sampling, allowing it to go more directly to the source of the substances under study.

Accurate tracing methods are also based on hydrologic analysis, more focused on the study about the distribution of water in the interest area. The study of the drainage network has been used for the study of soil composition in floodplain soils (Kaur *et al.* 2023), or the generation of geochemical maps of alluvial sediments (Lipp *et al.* 2023). The drainage system is normally computed as a mathematical graph or by triangulated irregular networks (TIN) (Rheinwalt *et al.* 2019). One of the most used graph-based methods implements the D8 algorithm for the construction of the drainage network (O'Callaghan & Mark 1984). The original algorithm has been improved and modified in different ways, for instance, to consider also specific geomorphological characteristics (Dávila-Hernández *et al.* 2022). In particular, it has been adapted to solve definite problems where the original versions of the algorithm do not provide valid solutions, such as in flat or steep areas (Gwak & Kim 2013; Ye *et al.* 2022) or when seepage occurs in the subsoil (Sahoo & Sahoo 2019). At other times, the focus is more on particular applications, such as obtaining flood-risk areas (Omran *et al.* 2016; Shen & Zhang 2018). Discharges of different substances in storm-water are of great environmental concern. Drainage systems of specific locations have been studied in relation with the presence of chemicals such as nitrogen or phosphorus (Yue *et al.* 2020; Kim *et al.* 2023; Qiu *et al.* 2023), or microplastics in marine environments (Lutz *et al.* 2021). In these studies, the focus is on the chemical analysis of these substances, rather than on their origin. Thus, to the best of our knowledge, these approaches have not been used for determining the areas from which they initially come from.

In short, the traceability of water reaching a reservoir is an interesting issue, which is not directly solved by GIS or simulation software. We propose a methodology to solve this problem based on a modification of the D8 algorithm. This new extended version allows more parameters to be considered in the calculation of the drainage network. Thus, it allows to load additional input information such as land uses, precipitation maps, and the different absorption rates of the soil, in addition to the terrain orography. This gives the algorithm greater versatility than the other functionalities provided by the conventional D8 method. In addition, a QGIS plugin has been implemented to facilitate the user's task of analyzing the data obtained by the process.

3. STUDY AREA AND DATASET

The study has been carried out in the area around the Rumblar reservoir, built on the Rumblar river (38°09′42″N 3°48′15″W, ETRS89), in the municipality of Baños de la Encina, in the province of Jaén, southern Spain (Figure 1). This region has a semi-continental climate with a mean annual temperature of 25.1 °C, a mean relative humidity of 70% and a mean annual precipitation of 1,341 mm (IDEAndalucia 2023). The capacity of the reservoir is 126 hm^3, although by the beginning of 2023, it was at 10% of its capacity (MITECO.GOB 2023). The reservoir is located in the heart of the Sierra Morena mountain range, surrounded by Mediterranean woodland. However, a few kilometers from the reservoir, the relief and vegetation change considerably, with olive groves alternating with hilly pastures. The swamp is mainly fed by river water coming from the Guadalevín and Pinto rivers, which cross both forested and cultivated areas along their courses. SOMAJASA company[2] is in charge of the quality control of this water, making chemical analyses periodically.

In this work, the hydrographic elements of the study area are obtained from the DEM with a spatial resolution of 2×2 m that forms part of the Spanish Spatial Data Infrastructure (IDEE) platform which follows INSPIRE Regulation (EC) No 1205/2008. This model was downloaded from the Download Center of the Autonomous Section of the National Center for Geographic Information.[3] It is a model obtained by interpolation from the terrain class of the LIDAR flights of the second coverage of the National Aerial Orthophotography Plan (PNOA).[4] The covered area is 56,497 ha.

Another important layer of information to be integrated into the system, and which will allow the identification of areas to monitor the dumping of phytosanitary products, is the land use types in the Rumblar basin (forest, agricultural or urban area). The land-use types established in the information layer have been obtained from the Land Cover Map in Andalusia (SIOSE-Andalucía 2016). This information is compiled by CORINE Land Cover and follows the Guidelines of the High Geographical

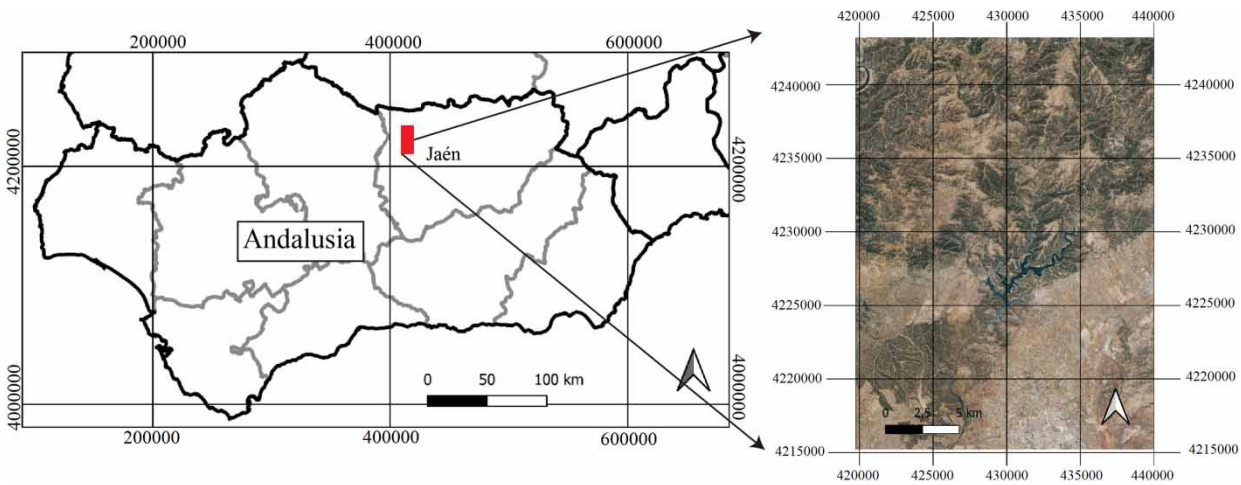


Figure 1 | Study area in which this research was conducted.

[2] https://www.somajasa.es/

[3] https://centrodedescargas.cnig.es/CentroDescargas/index.jsp

[4] https://pnoa.ign.es/web/portal/pnoa-lidar/estado-del-proyecto

Council and INSPIRE directive (INSPIRE 2023), which were obtained by photo-interpretation on satellite images Sentinel 2, Landsat in false color and Spot.

The cartographic data on land occupation are available in WMS (Web Map Service) format. The information shown to the user has different levels of disaggregation depending on the visualization scale chosen. At small scales, a synthetic classification of only four categories is shown, corresponding to the main types of use (forest, agricultural, artificial areas and water surfaces); at medium scales, there is an intermediate level of disaggregation with 16 categories that specify the type of use within each synthetic group; and at large scales only the polygon contours are shown, which are classified into 182 occupancy classes that can be found by interactive consultation, using the GetFeatureInfo property of the WMS. This feature, which allows the detailed occupancy of each parcel to be queried, is available at any scale.

Finally, the algorithm can also be provided with a rainfall map of the area. The original O'Callaghan algorithm to obtain the drainage network (O'Callaghan & Mark 1984) pours a water unit in each cell representing the terrain, that is, a homogeneous rainfall map. We have also changed this behavior by adding as input data a rainfall map with the same resolution as the DEM. This map can be obtained from different data sources and can record rainfall data for a specific period of time, 1 month's accumulation for instance. Our study area is not too large, even so, we have observed areas with higher and lower accumulation. As these maps are not directly accessible for any date, they have been obtained from meteorological stations available in the area and in nearby regions, even outside the basin under study. These stations are managed by the National Meteorological Institute (AEMET),[5] thus guaranteeing the reliability of the data. The accumulated rainfall data for 1 month was taken and a raster map was generated. Since the arrangement of these stations does not cover the study area homogeneously, the interpolation IDW (Inverse Distance Weighted) has been performed with the 14 nearby stations. This method assigns a weight to each of the starting points, and the influence of a point with respect to the rest decreases as the distance between them increases (Chen & Liu 2012). Figure 2 shows an example of a rainfall map of the study area for different periods. Figure 2(a) is the map available from the Web Coverage Service (WCS) of the Andalusian Environmental Portal.[6] It is a

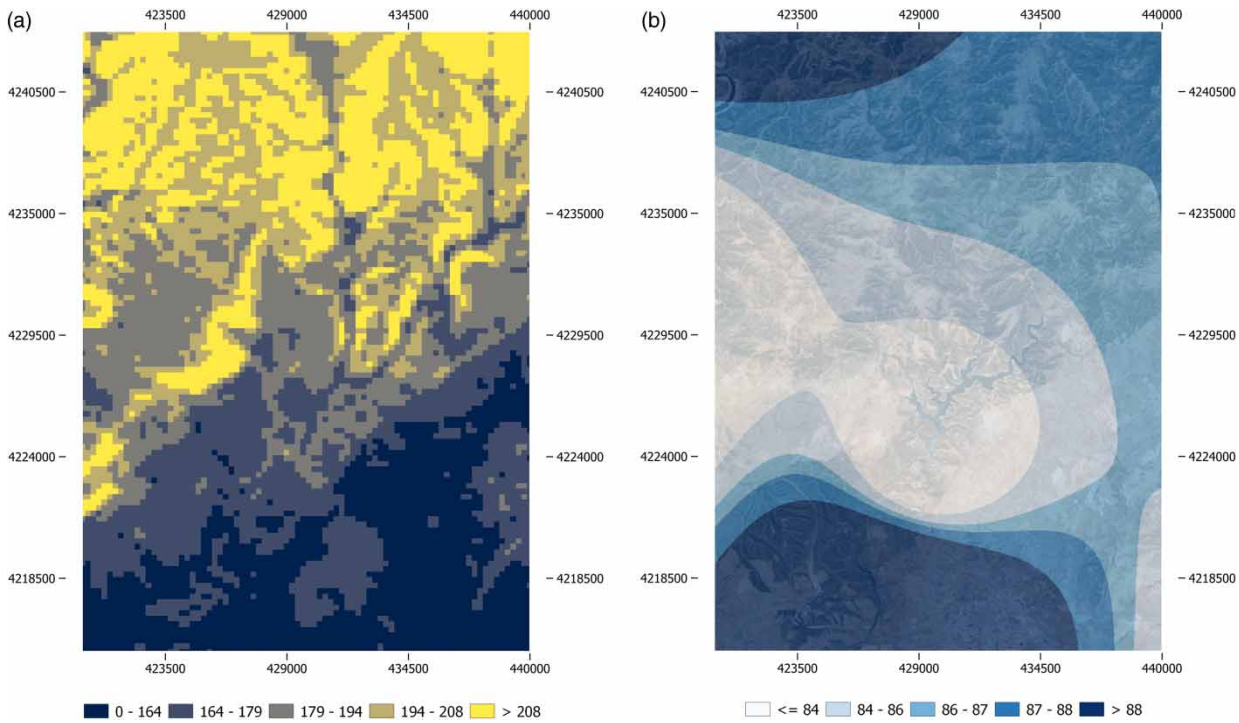


Figure 2 | Rainfall maps (mm). The coordinates (m) are Universal Transverse Mercator (UTM), zone 30, referred to ETRS89: (a) Spring 2016. (b) March 2022.

[5] https://www.aemet.es/es/datos_abiertos/AEMET_OpenData

[6] https://https://www.juntadeandalucia.es/medioambiente/portal/web/guest/home

100 × 100 m spatial resolution raster layer of precipitation climate variables of Andalusia for the year 2016. As this is the most recent map available from this service, we have generated a more recent map (Figure 2(b)). Each pixel represents the water discharged in each cell as input to the D8 algorithm. This last layer of information allows for a more realistic simulation of the amount of water reaching the reservoir from each parcel.

4. METHODOLOGY

This section describes the general process to obtain the traceability of water runoff along its path to the reservoir. The objective is to know the areas where the rainwater comes from and reaches specific points of the river basin, and which finally discharges their water into the reservoir. We also obtain the percentage of water that each of these zones contributes to the points under study. It is important to know whether this origin is agricultural or not in order to make a more detailed study of the areas that contribute to the concentration of discharges or make improper use of fertilizers or phytosanitary products. Figure 3 summarizes the methodological process in two main steps for obtaining the information, and which finally can be easily analyzed by means of a QGIS plugin.

4.1. First phase: obtaining representative points (RPs)

The first phase of the algorithm is described in Figure 3. Step 1 pre-processes the input DEM of the terrain and prepares it to obtain the drainage network, following the D8 method of O'Callaghan & Mark (1984). Initially, each cell $C(i,j)$ of the file is assigned with a specific direction codified as $d_{ij} = \{1|2|4|8|16|32|64|128\}$, representing the eight directions where the cell $C(i,j)$ drains its water accumulation at each iteration of the process. This is codified, respectively, as *{East(→), South-East(↘), South(↓), South-West(↙), West(←), North-West(↖), North(↑) and North-East(↗)}*. This direction is computed considering which of the neighboring cells have the steepest slope regarding $C(i,j)$. These directions will be used for obtaining the drainage network, as described below.

Once directions are fixed, the following process is detecting the so-called sinks, considered as those cells that do not discharge water to their neighbors, that is, they only accumulate water. Depending on the approach followed, these types of points are worked on differently. In our approach, it is assumed that no natural lakes are located in the current orography and therefore they are seen as data errors or artifacts. Thus, the method fills sinks in order to force them to drain into one of the surrounding neighbors as in Jenson & Domingue (1988). This approach does not generate any problems in the methodology and does prevent accumulation areas outside the reservoir.

Once the DEM is pre-processed, as Figure 3 describes, the D8 algorithm obtains a drainage network, similarly to the one implemented in O'Callaghan & Mark (1984). However, this new version of the algorithm also finds a set of representative points (RPs) following Algorithm 1. The cell $C(i,j)$ is considered an RP if a significant accumulation of water has passed through at any time during the drainage calculation process. We are going to focus our study on these points in the second phase of the methodology. Sometimes these RPs are junction points between different torrents or streams. Definitely, they are strategic points in which to focus the study over time.

Algorithm 1 summarizes this process. As input data, the algorithm provides the grid of (nxm) cells, $Grid[n][m]$, each of these cells with this information:

- **h**: **h**eight of the terrain in the area associated with the cell.
- **d**: **d**irection of the cell where water is poured codified as {1|2|4|8|16|32|64|128}.
- **lu**: the code of the **l**and **u**se associated with this cell. As referred to in Section 3, each cultivation plot has a different codification. Along the drainage calculation process, this value is propagated to the RPs and the reservoir.
- **ab**: the percentage of **ab**sorption associated with that particular land use type. This value also records the slope of the terrain in order to consider whether the water advances faster or slower in the cell area (see also Section 3).
- **rp**: is a Boolean value initialized false that indicates if this cell is a **R**epresentative **P**oint or not. This is considered the output of the algorithm.

Algorithm 1. Calculate Representative Points (RP)

Input: *Cell Grid*[*n*][*m*], for each *Cell*(*i*, *j*):{*h*, *d*, *lu*, *ab*, *rp* ← *false*}
Input: *RP_threshold*
Output: Cell *Grid*[*n*][*m*], for each cell *Cell*(*i*, *j*):{*rp*)}
1:**local variables**

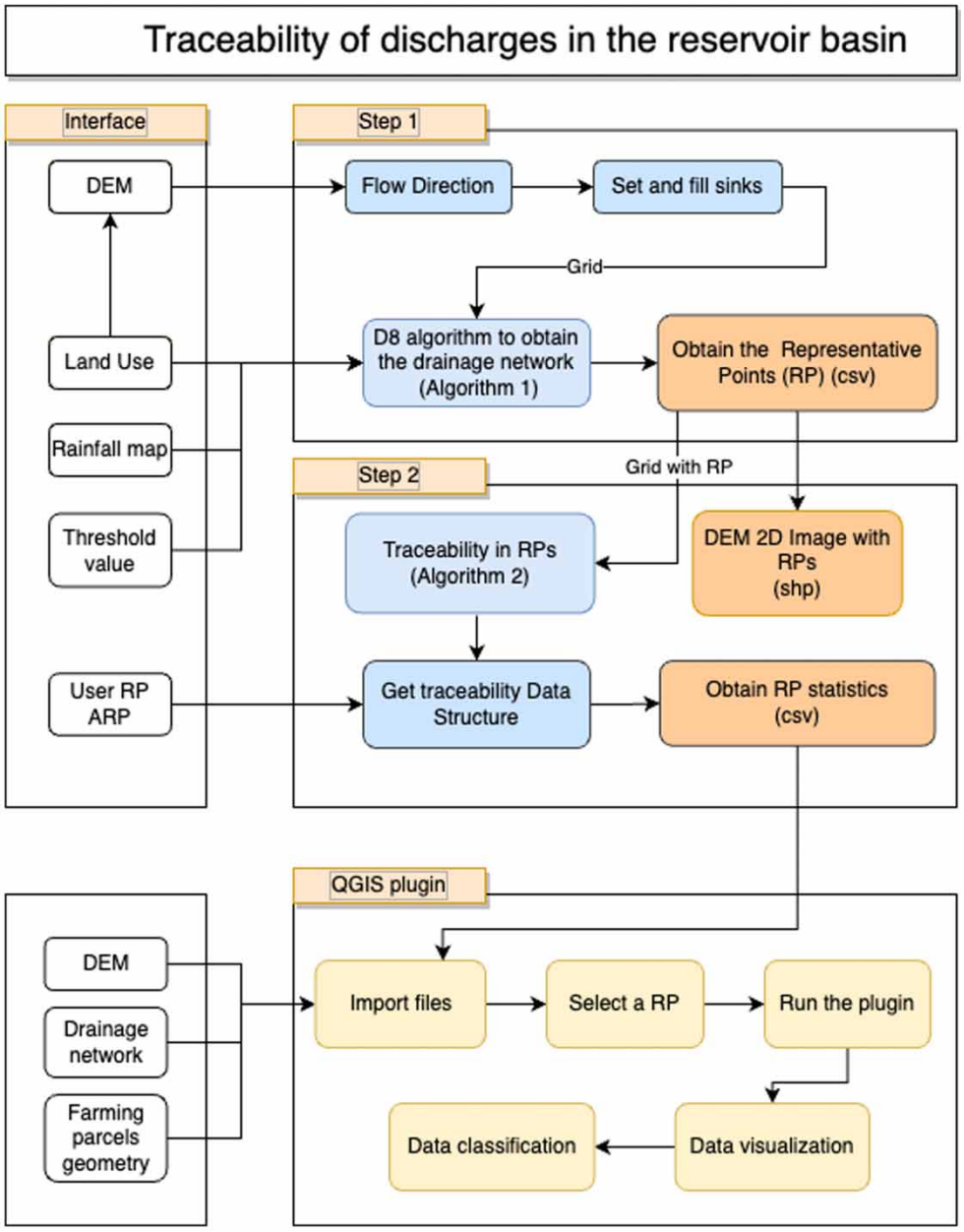


Figure 3 | General scheme of the traceability of discharges methodology.

2: *Integer* $oldFlow[n][m]$, $newFlow[n][m]$, i, j, k, l
3: *Boolean* $contProcess \leftarrow true$
4: *Float* oldf, newf
5:**end local variables**
6:**BEGIN**
7:**for all** $i, j, \quad i \in [0..n-1], \quad j \in [0..m-1]$ **do**
8: $oldFlow[i][j] \leftarrow 1$
9: $newFlow[i][j] \leftarrow 0$
10:**end for**
11:**while** $(contProcess == true)$

```
12:  contProcess ← false
13:  for all i, j,  i ∈ [0..n − 1],  j ∈ [0..m − 1] do
14:     oldf ← oldFlow[i][j]
15:     ifoldf ≠ 0 then
16:        contProcess ← true
17:        k ← i + grid[i][j].d.Y
18:        l ← j + grid[i][j].d.X
19:        newf ← oldf∗(1 − (Cell(i, j).ab/100))
20:        newFlow[k][l]+ = newf
21:    end if
22:  end for
23:  for all i, j,  i ∈ [0..n − 1],  j ∈ [0..m − 1] do
24:     if newFlow[i][j] > RP_thereshold then
25:        grid[i][j].rp ← true
26:     end if
27:     oldFlow[i][j] ← newFlow[i][j]
28:     newFlow[i][j] ← 0
29:  end for
30:end while
31:END
```

The method is an iterative process in which all cells with accumulated water yield in parallel their contents to one of their eight neighbors, according to the previously assigned direction. The algorithm uses two matrices in order to simulate this synchronous water transfer at an instant of time. This allows it to work sequentially to perform a process of a parallel nature. A complete transfer of water over the entire surface is considered a full iteration of the algorithm *(Alg1:while loop)*. In the original algorithm, all cells are given one water unit to start the simulation process in matrix *oldFlow* (Alg1:line 8). We have also used rainfall maps as input data instead, as described above, that are not included in the algorithm for simplicity. In case of being considered, as these maps have the same DEM resolution, the corresponding value is obtained from the corresponding cell of the image. On the other hand, and according to studies of the area, there are no natural springs in the area of interest. If this were to occur in any specific site, the method would add a higher value to the associated cell. The water that accumulates the *oldFlow* matrix is discharged into the *newFlow* matrix at the end of each iteration of the process (while loop) in order to simulate the parallelism. At the end of each of these iterations, *oldFlow* gets the values of *newFlow* to continue with the iterative process (Alg1:line 27), and the *newFlow* matrix is initialized to zero to obtain new values in the next loop.

As cells give their contents to other cells, many cells receive from one or more neighboring cells their accumulated contents. The algorithm finishes when no more water is moved in any cell during the course of an iteration. This condition is reflected with the variable $contProcess == true$, which determines if the process continues (Alg1:line 11). Once into the loop, the variable is set to false (Alg1:line 12) but again reverted to true in case any water movement is detected (Alg1:line 16). This implies that a new iteration of the algorithm is performed until no more water is transferred along the matrix.

The first *for* loop inside the while (Alg1:lines 13–22) represents the synchronous movement of all cells in unison to pour the contents of $Cell(i, j)$ into a neighboring $Cell(k, l)$ according to the previously defined direction **d**. Each cell always pours in one cell (except those on the edge).

Not all content of $Cell(i, j)$ is eventually discharged into the neighboring cell. The original method of the algorithm only provides the drainage network and all water units move through the drainage system as if they were particles that are not lost along the process. In a real river basin, not all rainwater is incorporated into the flow to the reservoir. Several factors are involved, such as the amount of water that falls, the slope of the land, its absorption capacity or evaporation factor. These parameters are considered as follows: (1) The area under study has no natural sources, so all runoff is rainwater from the watershed. (2) Evaporation is not considered in this method. In practice, this implies considering this factor as a constant throughout the study territory. (3) Taking advantage of the fact that we have zoned in plots the entire area of the site, and all of these polygonal areas have an associated land use type, we have associated this type of land with a specific

absorption factor, which can be configured individually. This factor takes into account whether a forested area has a greater capacity to absorb rainfall than an area used for olive cultivation. (4) Rainfall is quite homogeneous in the study area, in any case, rainfall maps have been added to be considered the methodological level. These maps modify the original algorithm by increasing or decreasing the initial poured water in each cell. (5) The slope of the terrain is considered as it is a parameter associated with the topography of the terrain. For each cell $C(i, j)$ and for each iteration of the drainage algorithm, it is considered what percentage of water is yielded to the neighboring cell and what percentage is lost by absorption. This reduces the percentage of water falling in the highlands that reaches the reservoir, but does not nullify it. All these factors are considered in order to take advantage of the knowledge of the terrain, described in Section 3, in order to be more accurate in assessing the origin of the flows.

In summary, the absorption value depends on land use type (agricultural or forestry) and soil slope. It is assumed that the greater the slope of the terrain, the lower the absorption. Then, each cell precomputes the absorption rate, ab according to the formulation:

$$ab = \frac{\left(\frac{\pi}{2} - \arctan(\triangle h / CellSize\right) * abs\_rate}{\frac{\pi}{2}},$$

where $\pi/2$ represents the maximum slope allowed 90°, that is, a vertical slope. $\triangle h$ represents the height difference between the two annexed cells, origin and destination. *CellSize* is a constant which determines the width of any of the cells. Finally, *abs_rate* is the absorption rate associated with the land use type of the cell. Each land use type has a specific value configured in the input data described in Section 3. This factor is applied by deducting a percentage from the amount of water that is discharged into the neighboring cell (Alg1:line 19).

The second *for* loop (Alg1:lines 23–29) is executed once all cells have received water from their neighbors. Then, it is possible to know which of them have received a greater amount of water than others. If they exceed a fixed threshold value, they are considered as RP or representative points. This threshold value is defined in advance as input data to the algorithm and can be changed in order to obtain more or less of these points. A flow rate greater than the threshold may pass through the cell on more than one occasion. The first time this happens, the **rp** value is set to true forever (Alg1:line 25). Finally, the loop interchanges the role of *oldFlow* with the *newFlow* matrices and initializes the latter to zero in order to start another iteration.

In summary, the first phase of the algorithm obtains the drainage network and obtains the RPs as described in Algorithm 1. It is possible to add specific additional representative points (ARP) that are treated as the original RPs. These are points expressly chosen by the technicians to carry out chemical analyses. All this information is the input data of the second phase of the methodology: the study of the traceability of the water reaching these points under study.

4.2. Water traceability study

Once RPs are obtained in Algorithm 1 (Step 1 in Figure 3), the next step is to add the information associated with the sequence of the drainage process at each of these RPs, represented as Step 2 in Figure 3. It is possible to include new RPs customized by the user as an annexed file (previously defined as ARPs). These can be useful if the technicians in charge of monitoring water quality want to follow up on a specific geographical point.

Algorithm 2. Traceability of RPs

Input: Cell $Grid[n][m]$, for each $Cell(i, j)$:$\{h, d, lu, ab, rp, tc \leftarrow \emptyset, TC \leftarrow \emptyset\}$
Input: *MAX_ITER*
Output: Cell $Grid[n][m]$, for each cell $Cell(i, j)$:$\{TC\}$
1:**local variables**
2: *Integer oldFlow*$[n][m]$, *newFlow*$[n][m]$, $i, j, k, l, nIter \leftarrow 0$
3: *Boolean contProcess* $\leftarrow$ *true*
4: *Float* oldf, newf
5:**end local variables**
6:**BEGIN**
7:**for all** $i, j, \quad i \in [0..n-1], \quad j \in [0..m-1]$

```
    oldFlow[i][j] ← 1
    newFlow[i][j] ← 0
end for
while(contProcess == true  AND  nIter < MAX_ITER) do
    nIter + = 1
    contProcess ← false
    for all i, j,   i ∈ [0..n − 1],   j ∈ [0..m − 1] do
        oldf ← oldFlow[i][j]
        if oldf ≠ 0 then
            contProcess ← true
            k ← i + grid[i][j].d.Y
            l ← j + grid[i][j].d.X
            newf ← oldf ∗ (1 − (Cell(i, j).ab/100))
            newFlow[k][l] + = nflow
        end if
    end for
    for all i, j,   i ∈ [0..n − 1],   j ∈ [0..m − 1] do
        grid[i][j].tc  ← addTreeDE({nIter, newFlow[i][j], grid[i][j].lu})
        if newFlow[i][j].rp == true then
            grid[i][j].TC + = grid[i][j].tc
        end if
        grid[i][j].tc ← 0
        oldFlow[i][j] ← newFlow[i][j]
        newFlow[i][j] ← 0
    end for
end while
END
```

Algorithm 2 describes the process. The procedure is similar to that of Algorithm 1, but this time a specific data structure is associated with each RP in each iteration of the process. In both algorithms, the accumulated water from one cell is poured into the cell next to it, but now at each iteration of the process, it also adds information about the original plot from which the water comes. Summarizing, each RP point stores for each iteration:

- The algorithm iteration number in order to know the temporal sequencing.
- The water cumulus that passes through (which coincides with that of Algorithm 1).
- The number of the parcel where the water comes from. This reference number is associated with a type of land use.

Thus, for N RPs and K iterations of the algorithm and with P plots cultivated inside the basin of the reservoir, we can obtain a data structure with a capacity of $O(N{*}K{*}P)$ data. However, this is considered the worst case, since the information is only stored for those iterations where there has been flow movement. Additionally, only those plots pouring water into an RP in a specific iteration are considered.

For maintaining such information on every RP point, we defined a data structure called TC (Tree-based Cumulative data), only available for those cells considered as RP. This structure is a tree-map with N entries, one for each RP cell $C(i, j)$. Associated with all RPs cells, the tree stores an array as much of K size. Each position of this array stores the information about the water traceability at one specific iteration of the algorithm (flow from each plot). Figure 4 represents this data structure. Each element of the array is again a new tree-based data structure with a single entry for each land use type that visits the RP cell. Each time that a specific land use type (**lu**) visits the cell, it adds the accumulation of water value to this entry. In case of a new land use type, a new entry is created in the tree. Once the algorithm finishes, the data structure has the whole information about the water traceability of the reservoir basin.

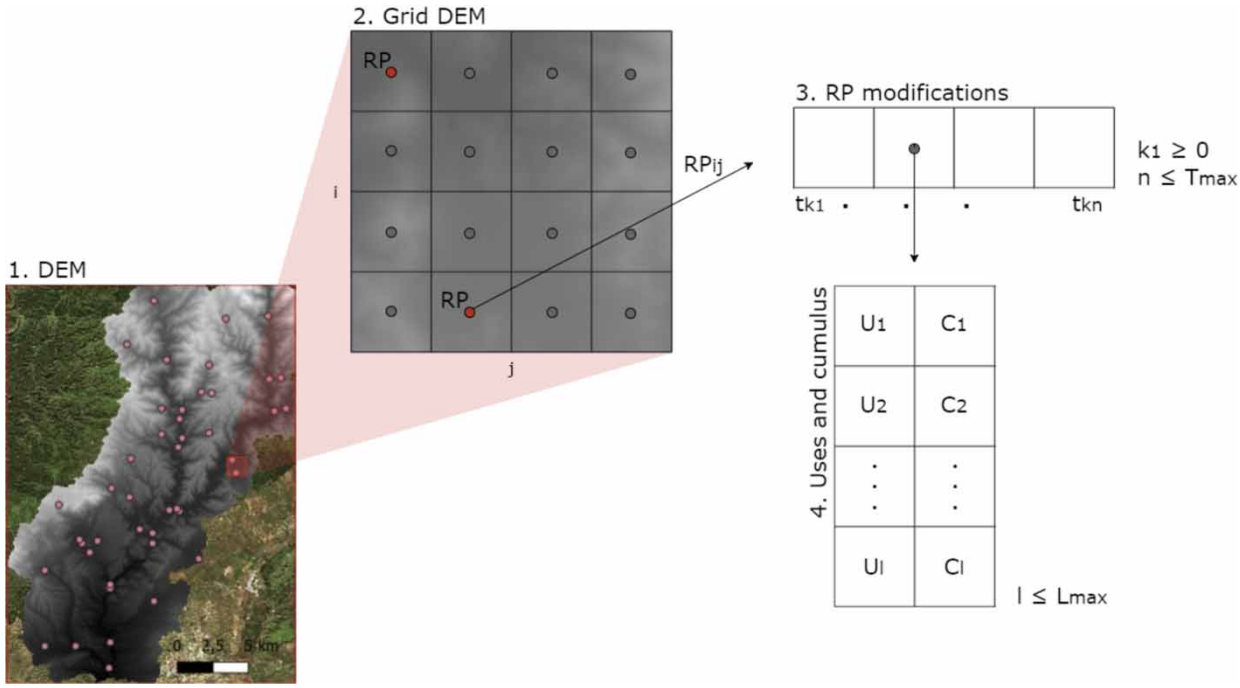


Figure 4 | TC is data structure associated with RPs.

Algorithm 2 works as follows: it initializes the accumulation flow matrices (lines 7–10). Then, it performs the accumulation values during the first for-loop (lines 14–23, process identical to that of Algorithm 1). The second for-loop (lines 24–32) adds to each **tc** data structure, and for each iteration *nIter* the new set of two values: flow accumulation and land use type. If *Cell*(*i*, *j*) is an RP, then the **TC** data structure stores the temporal sequence. TC (in capital letters) is only available for RPs. Otherwise, the *tc* structure is finally reset in order to obtain new values in the next iteration.

The result is a set of RP which stores for each iteration both the cumulus water that has been passing through the cell and its land use origin. This data sequence is later analyzed in the following section of the process with the help of a QGIS plugin (see Figure 3).

5. QGIS PLUGIN FOR WATER TRACEABILITY

QGIS is a versatile open-source Geographic Information System for managing spatial data. Among its features, we highlight the capacity of adding Python plugins with new functionality. Plugins are developed by independent organizations and developers and can be used by the user community.

In order to facilitate the analysis task to technicians about the agricultural plots discharging rainwater on these RPs, we have developed a specific QGIS plugin. This way the data analysis requires only a few notions about this GIS tool.

Before the plugin can be used, the user must load the following files into QGIS (see Figure 3). Section 8 has links to these available data.

- The geo-referenced DEM file in tiff format within the terrain rectangle in which the RPs obtained by Algorithm 1 are included. This region should include the entire reservoir basin.
- The drainage network image was obtained from Algorithm 1 (tiff format). This is not compulsory but interesting to observe the route of the water. Figure 5(a) overlaps the DEM and the drainage network.
- The RPs also calculated by Algorithm 1 are specific points of this drainage network (csv format).
- The geometry of agricultural plots within the basin of the reservoir area (.shp format). The geo-referenced information of this file includes the polygon of each plot, the type of land use (agricultural, forestry, etc.) and the parcel number according to the standard numbering. Figure 5(b) shows RPs and the plot polygons.

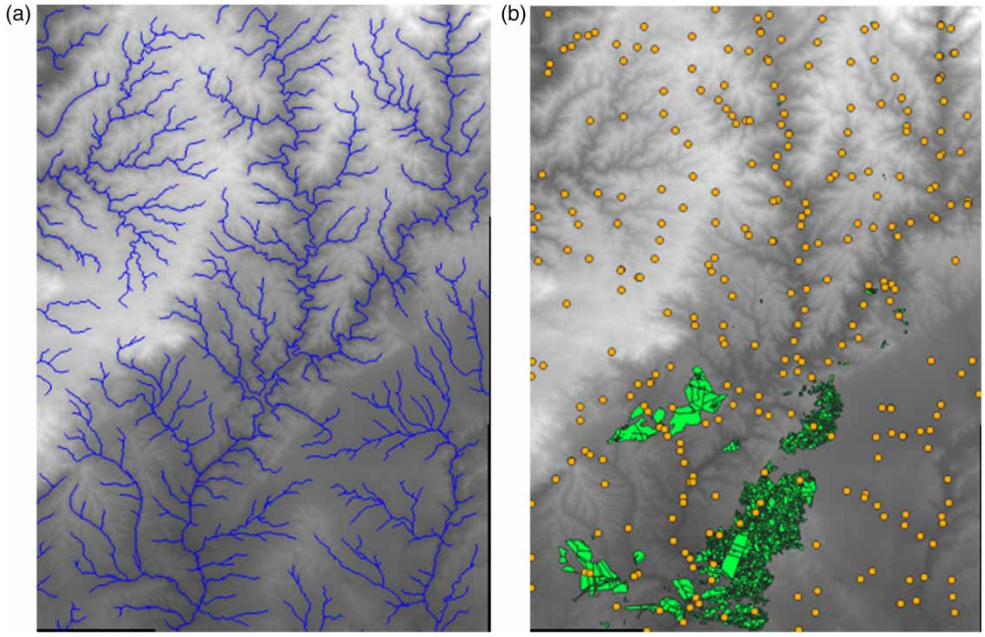


Figure 5 | (a) DEM and drainage network. (b) RPs and plot polygons.

- The output file (csv format) of Algorithm 2 with the information obtained associated with each RP. As mentioned in Section 4, each of these RPs has an entry for all those plots, which sooner or later, pour water into that specific point position.

Once this information has been obtained, the user can open the plugin entitled **Traceability of water origin**, depicted in Figure 6(a). The picture shows the information that must be given to run the plugin:

- The RP layer (.asc) with the information obtained from Algorithm 2 for all RPs, and ARPs if necessary.
- The polygon layer of plots (.shp).
- If the checkbox *Load existing result layer* is unchecked, the process of searching for plots associated with the selected RP is done from the beginning, generating a new layer with the found plots. Otherwise, the layer used has been previously generated in another execution of the plugin, and can be directly selected.

After the plugin execution, the list of plots whose water has passed through the selected RP is presented, as depicted in Figure 6(b). The next step is the analysis of the results.

6. RESULTS AND DISCUSSION

After the user runs the 'Traceability of water origin' plugin, as described above, the data can be easily interpreted. Figure 7 shows the process for managing the plugin. Once an existing or new QGIS project is opened, and the files described above are selected, the layer information can be depicted on the screen by representing plots, the drainage network and RPs. As the information obtained from Algorithm 2 is associated with each representative point, the most natural way for the user to interact with a graphical interface is by 'clicking' on the screen in the desired RP, as Figure 7(a) shows. Then, the specific information of this RP can be easily interpreted by visualizing the data (Figure 7(b))

This section first analyzes the impact of cumulative rainfall maps with respect to the original O'Callaghan algorithm. Secondly, we see the process of analyzing data from the user view point by using the designed plugin.

6.1. Impact of rainfall mapping on methodology

As previously mentioned, the method for calculating the drainage network has been modified in order to simulate reality as much as possible. For this reason, input data such as soil and land use types, or rainfall maps have been added. Specifically,

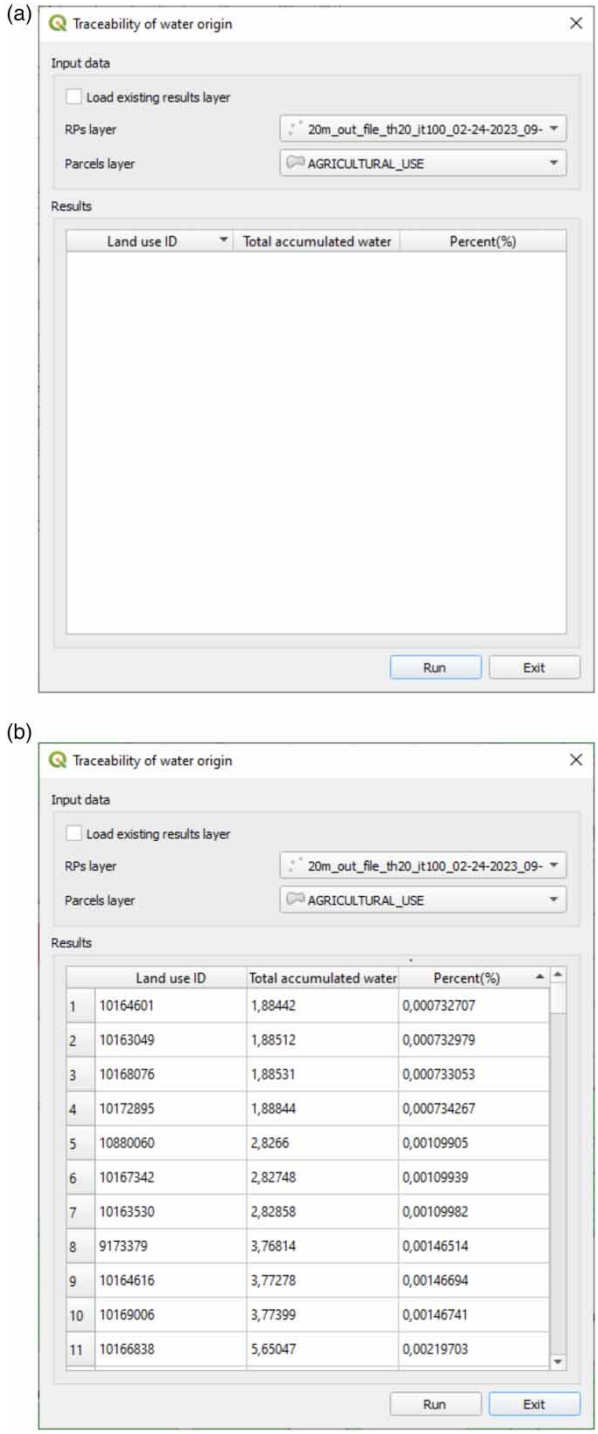


Figure 6 | (a) Home screen of the 'Traceability of water origin'. (b) Information about RPs.

we have tested the impact of the use of rainfall maps, in particular the maps of Figure 2 described in Section 3, associated, respectively, with March 2022 and the whole spring 2016. In particular, we analyze whether the rainfall maps really have an influence on data analysis. The chart of Figure 8(a) shows the study on the same RP but executing the algorithm with the two aforementioned maps and a third without using rainfall maps, that is, as the original D8 algorithm works. In the latter case,

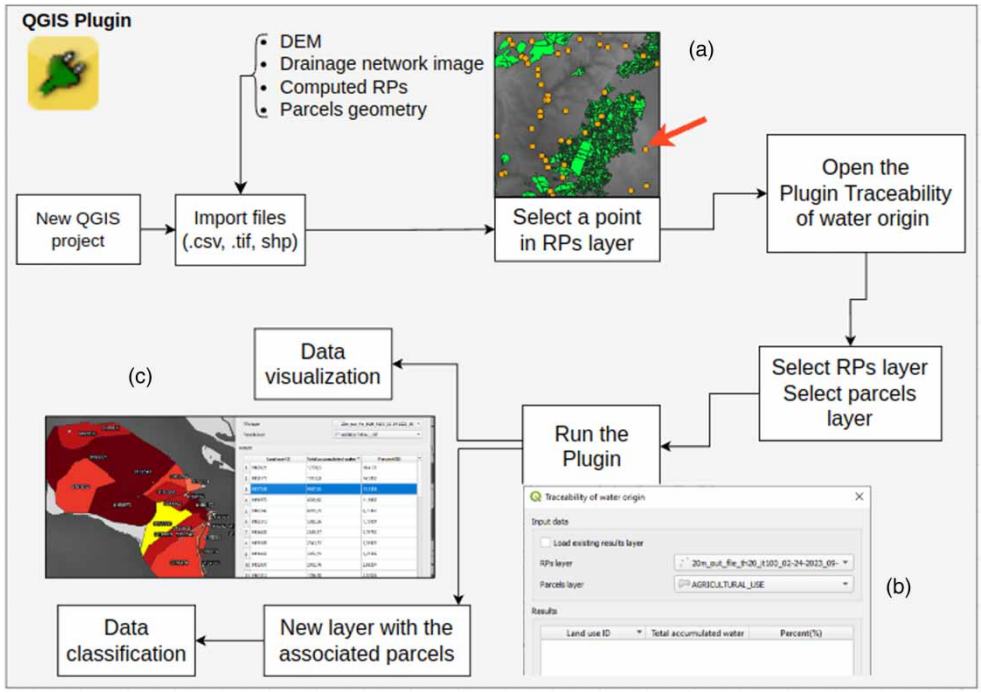


Figure 7 | Use of the QGIS 'Traceability of water origin' plugin.

the method always contemplates a value of 1 as the initial amount of water in each cell. In the rainfall maps, however, the rainfall values have been interpolated in the range [0.8, 1.2].

The horizontal axis represents the plots that have the greatest influence when discharging water in that RP. The vertical axis provides water accumulation values passing through that RP. As observed, there is no appreciable difference between the three lines. The reason is that the variations in water accumulation are not very representative, partly because it is a short period of time and also because the work area is small. In any case, this methodology can be interesting for larger areas with more marked climatic variations, or in order to observe the behavior of rainfall over specific points in certain periods of time.

In the graphs of Figure 9, we observe the amount of accumulated water for two different RPs. In order to make the results more visible, we only show the data for those plots that exceed 1% of the total accumulated water. In each chart, the two rainfall maps are represented (March 2022 and spring 2016), as well as two different numbers of iterations: 100 and 284 (the maximum). In both cases, as the number of iterations increases, the amount of water accumulated in each plot calculated is greater. There are many plots that have a very small difference in accumulated water between the execution of 100 and 284 iterations. On the other hand, in the plots that have a greater amount of accumulated water, this difference is more noticeable. This allows us to observe a certain correlation of the results between the number of iterations (the timing of the simulation) and the rainfall maps (climatic conditions). The higher the extreme conditions, the higher the ratio will be.

6.2. Percentage of water from plots

The process of visual representation is very configurable in QGIS. The user can select a color ramp to represent a range of values, as Figure 10(a) shows. Once the visual criteria are established, this range highlights with intense colors which are the plots that discharge the most in the point under study, as depicted in Figure 10(b). This provides a visual tool for interacting with these results, as described below.

6.3. Study of individual plots in the Rumblar basin

The technicians in charge of water quality control carry out sampling in certain areas to perform chemical analyses on the different water streams that reach the reservoir. Figure 11(a) shows most of the plots in the basin area of the

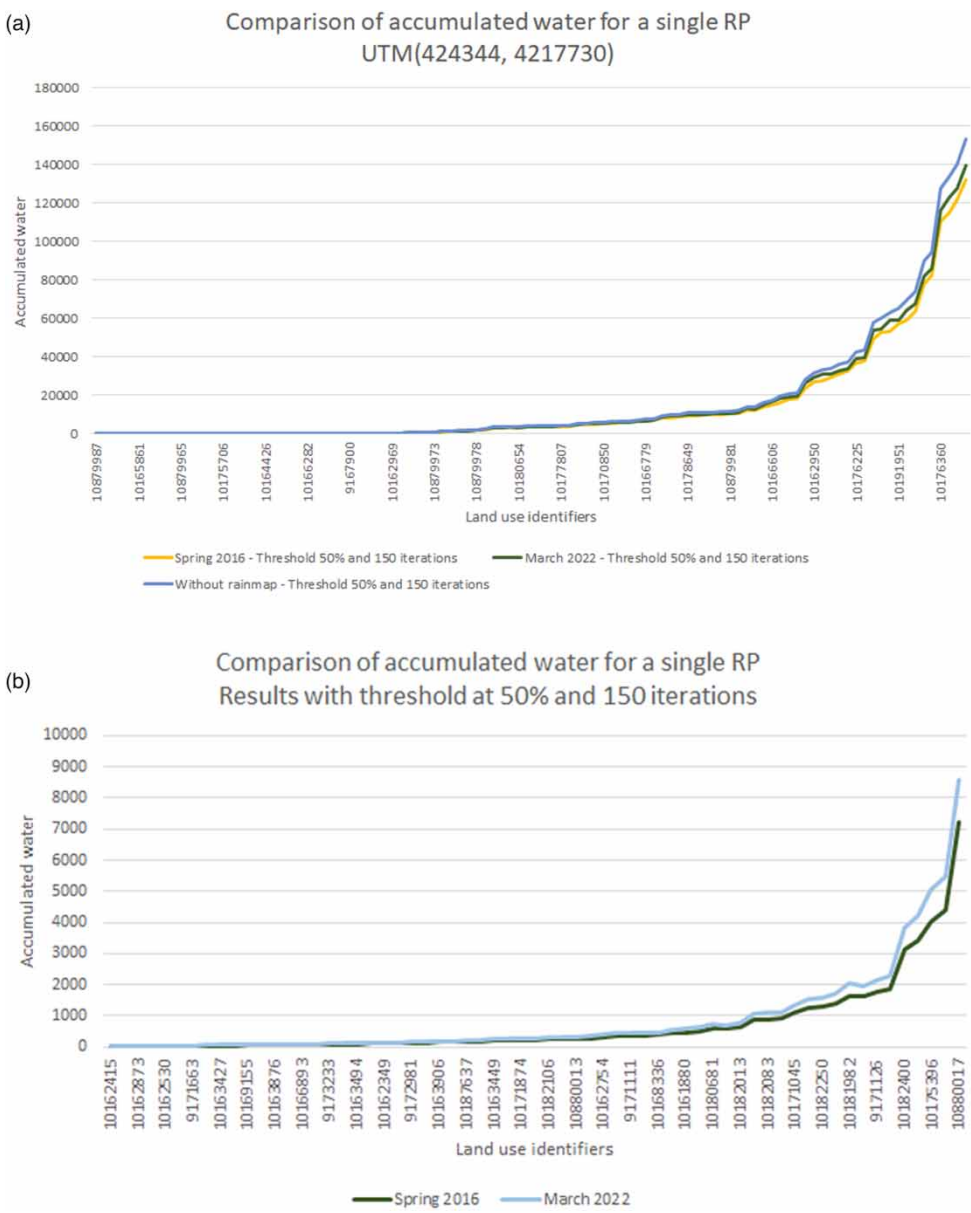


Figure 8 | (a) Influence of rainfall maps in the study. (b) Information about RPs.

reservoir. In white color, we distinguish those that do not drain water into the RP represented in yellow color (bottom of the image). In a range of red color, we see those that do pour at that specific point. Figure 11(b) is centered in this latter area. The table on the right side of the image shows the information of a plot in each line. Just by clicking on one of these lines, the selected plot is displayed differently. In the picture, the plot identified as 9,157,568 has been selected and the left side of the image shows its polygon in yellow color. We also are informed about the amount of water that has been accumulated in the point. Although this value is only approximate, it is valid for comparative purposes. The third column gives us the percentage of water discharged from this specific plot, considering the total number of plots. This allows the technicians to focus the attention in those places that contribute the most, not only in the amount of water but also the chemicals that are also washed away in the drainage process. This way, information analysis can be done easily for technicians with limited use of geographic information tools while establishing control mechanisms over large areas.

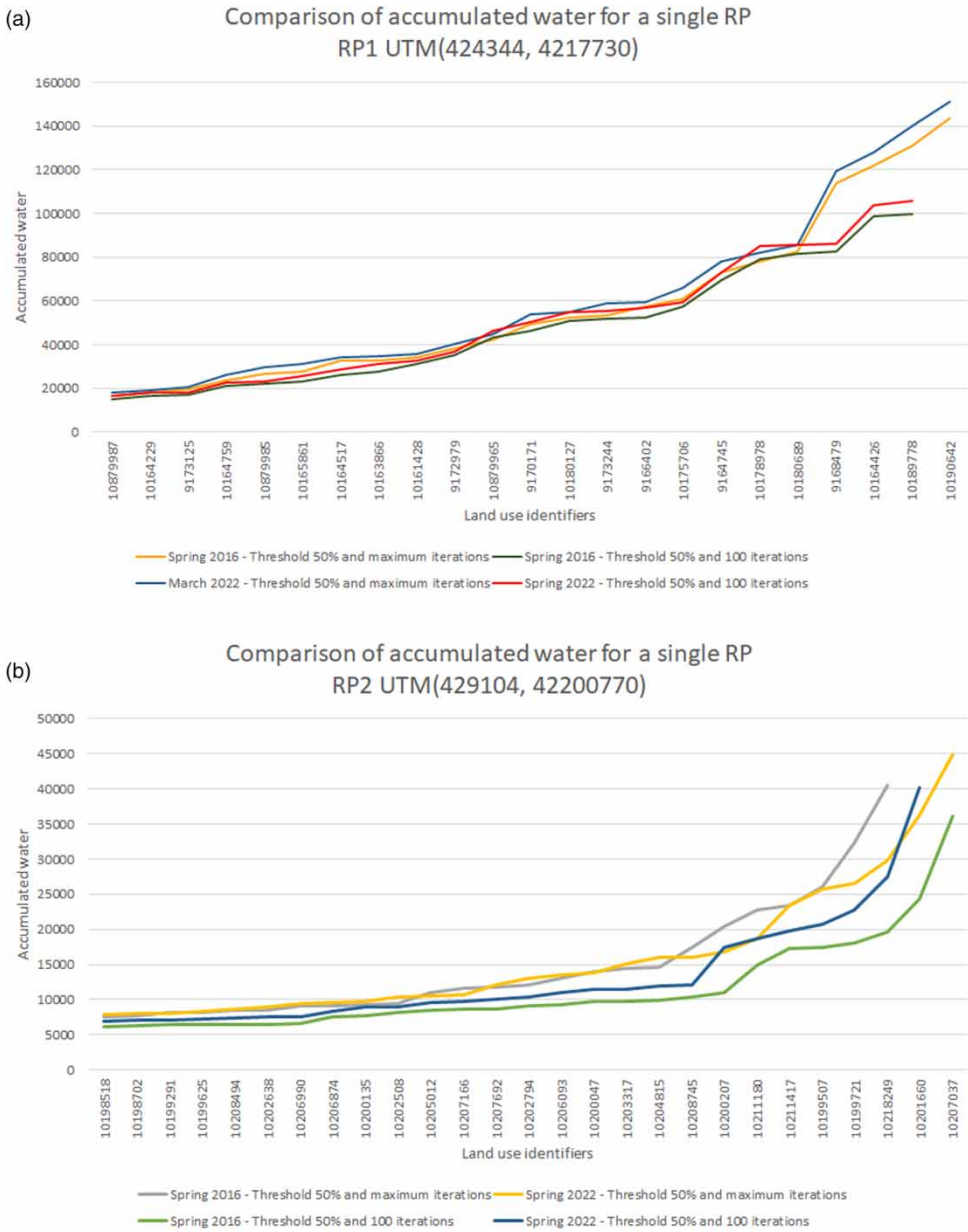


Figure 9 | Distribution of accumulated from each plot in points (a) RP1 and (b) RP2.

6.4. Results in the Cazorla mountain area

In order to test the proposed methodology in another natural area and check its validity, we have run the algorithmic process in another study area. This is the Natural Park of Cazorla, Segura and Las Villas, a Biosphere Reserve located in the northeast of the province of Jaén (38° 07′52 'N 2° 45′52 'W, ETRS89). The reason for selecting an area also belonging to the province of Jaén is that the olive trees are treated with the same type of phytosanitary products, so the problem posed by the arrival of these residues in the drinking water reservoirs is the same. On the other hand, it is an extensive mountainous massif of abrupt relief, the source of the Guadalquivir and Segura rivers, in which karst formations abound. There are differences in the morphology of the terrain, with a topography and steeper slopes, up to 1,800 m. However, forestry and agricultural land use also coexist. An area of smaller land area has been selected, the Rumblar area was 63,186.50 Ha, and this second one is 1,098.71 Ha. A spatial resolution of 20 m has been selected, in order to evaluate the results when processing at the same level of detail.

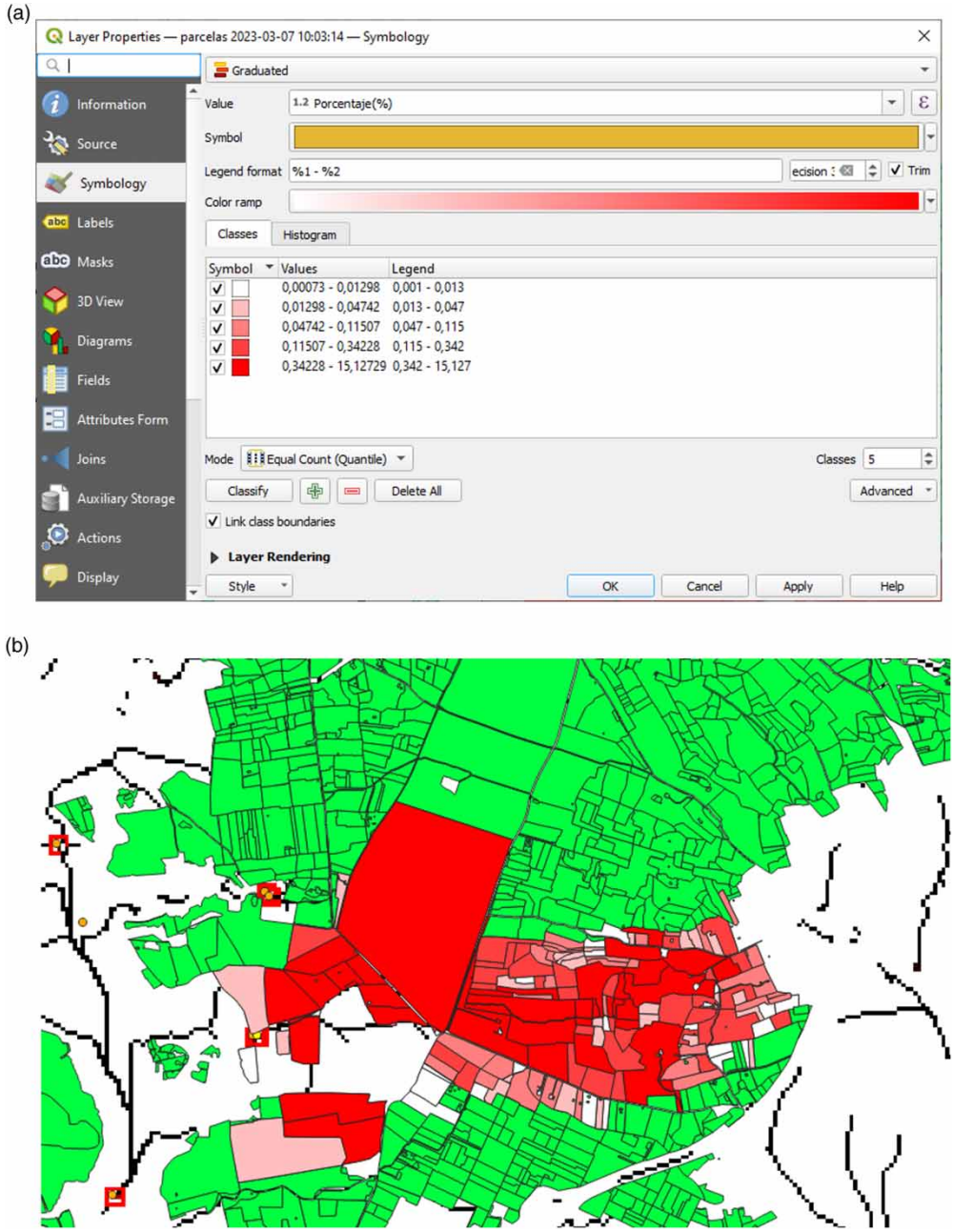


Figure 10 | (a) Classification by intervals. Darker color represents a higher percentage in discharges in the RP. (b) Example of plots discharging water into that RP.

The area under study comprises part of the basin of two rivers, as depicted in Figure 12(a). The drainage network and the RPs are superimposed to the DEM image. The RP under study is highlighted in Figure 12(c). The list of plots (given with its cadastral code) whose water drains in this RP, is shown in Figure 12(b). As can be observed, the point is located among a cultivated area, near the intersection of two tributaries of the river. In fact, Figure 12(b) shows the table listing the plots sorted by the percentage of water contribution to this point.

7. CONCLUSIONS AND FUTURE WORK

In this work, we introduce the methodology for obtaining the information about those agricultural plots that end up discharging their waters into specific areas, such as a stream at the entrance to a reservoir. The process consists of

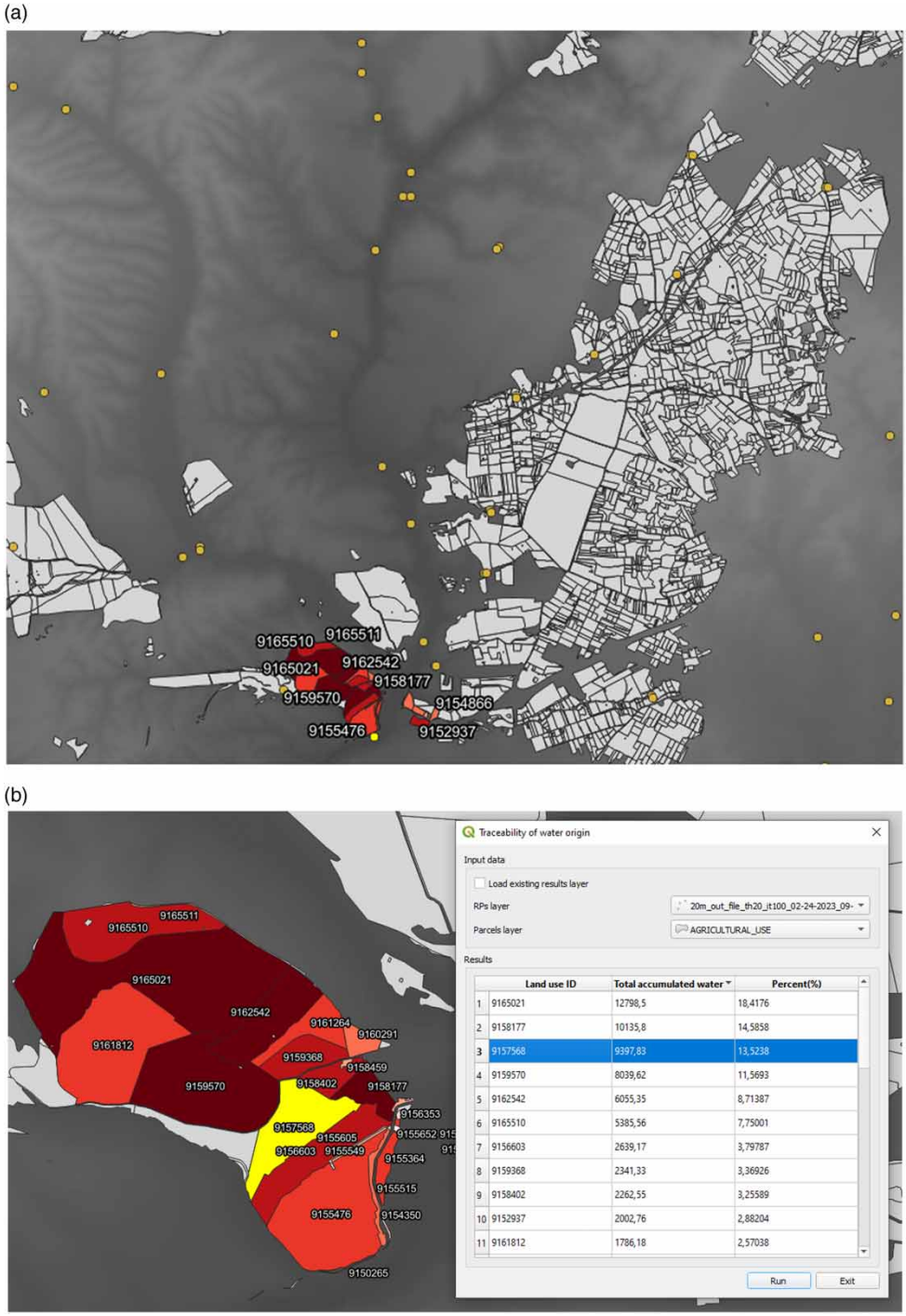


Figure 11 | (a) Set of plots. In white those that do not affect the selected RP (in yellow) and in red those pouring into the RP. (b) Visualization of a specific plot (yellow color). Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/hydro.2023.092.

modifying the classical algorithm to obtain the drainage network, the so-called O'Callaghan's D8. For this purpose, the origin of the water is tracked at all times along its path until it reaches the point of interest (or RP). Each individual agricultural parcel that discharges at that point is taken into account, also considering the proportion in which it discharges.

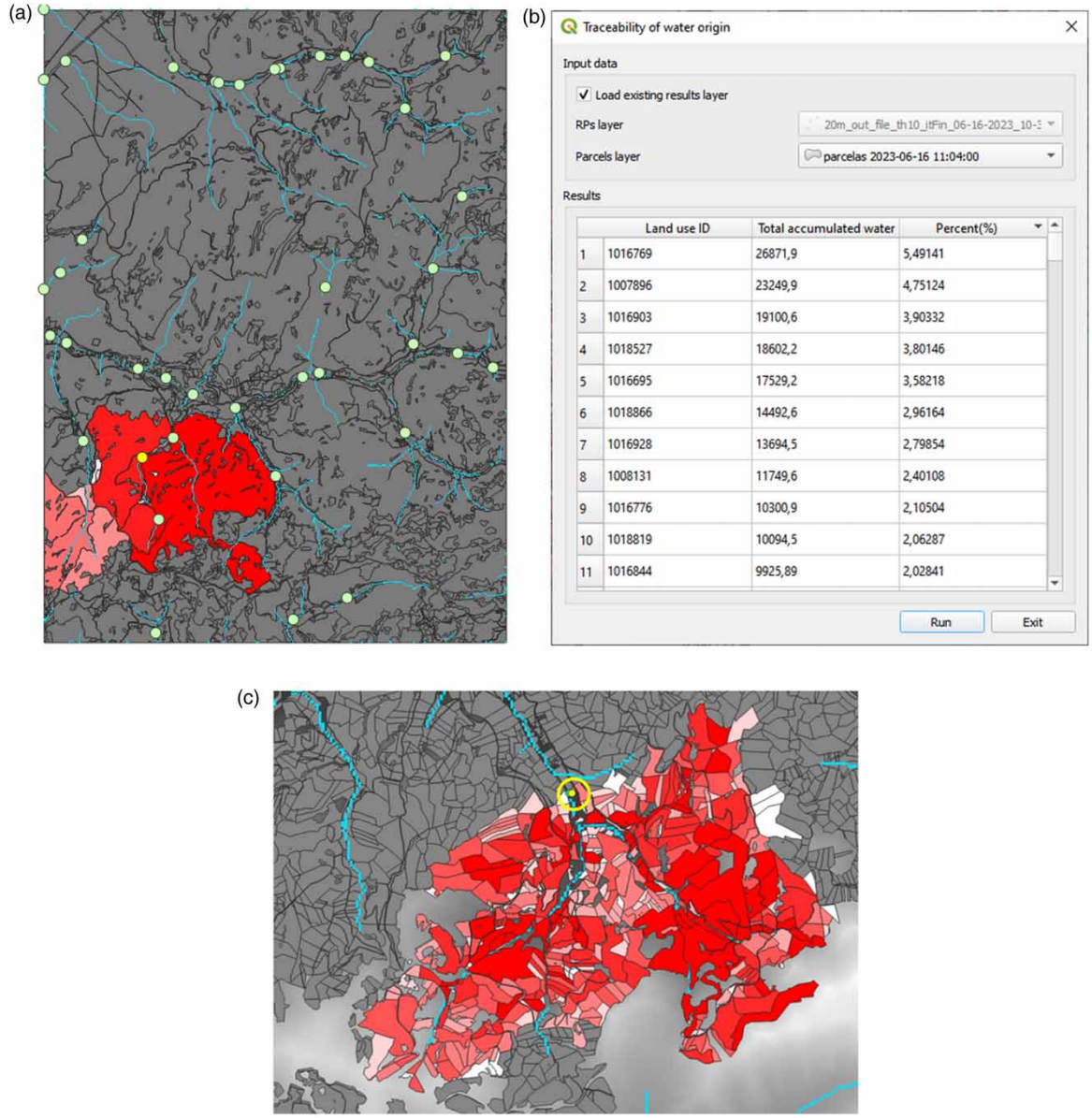


Figure 12 | (a) Drainage network and RPs in Cazorla area. (b) List of plots and percentages of water reaching the RP of marked in (c). (c) RP and the plots that most contribute in darker red. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/hydro.2023.092.

Users interested in this application may be ecologists or environmental technicians, so the usability of the process is important, in case they are not very familiar with GIS tools. To facilitate this task, a plugin has been implemented in QGIS that allows direct interaction with the maps. The user can click on points of interest and plots to obtain the information both graphically and textually. Many of the input data, as well as the parameters, are configurable to perform different types of simulation and also in different time phases. The latter allows the algorithm to be adapted to different study objectives

The initial premise of the work is to be able to control the agrochemicals that are carried by the water and reach the water reserves. The pending work consists of contrasting this methodology with chemical analysis of water at sensitive points. This should also be completed with the chemical analysis of the soil of the agricultural plots considered most representative by this methodology. With the analysis, it is possible to focus the study on the areas that are most likely to be affected.

8. DATA AND CODE AVAILABILITY

In order to make the data and code of this paper findable, accessible, interoperable and reusable according to the FAIR Guiding Principles for scientific data management and stewardship (Wilkinson *et al.* 2016), we have included both the data and the code in public repositories.

Data and QGIS plugin is available through this Drive link: https://drive.google.com/drive/folders/1yOJ_rarpUfYV-ylcdsYAgiOYmTidziNV.

- **Data** available for the Rumblar use case:
  - **Input data**: all data required to run the application.
  - **Output data**: the data generated after running the application in this area.
- **QGIS plugin**: the plugin and installation instructions

The application is implemented in C++. The source code including Algorithms 1 and 2 as well as the end-user interface is available through this GitHub public repository:
https://github.com/lidiort/WaterTraceability_D8.git.

ACKNOWLEDGEMENTS

This research has been partially funded through the research projects PYC20-RE-005-UJA and PID2021-126339OB-I00. This first one is co-financed with the European Union (ERDF funds) and the second is partially supported by the Spanish Ministry of Science and Innovation and the European Union. We are also grateful for the support provided by the Ministry for Ecological Transition and the Demographic Challenge, Spanish Government (AEMET, Agencia Estatal de Meteorología). We would also like to thank SOMAJASA for their collaboration.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Alfieri, L., Avanzi, F., Delogu, F., Gabellani, S., Bruno, G., Campo, L., Libertino, A., Massari, C., Tarpanelli, A., Rains, D., Miralles, D. G., Quast, R., Vreugdenhil, M., Wu, H. & Brocca, L. 2022 High-resolution satellite products improve hydrological modeling in northern Italy. *Hydrology and Earth System Sciences* **26** (14), 3921–3939. Publisher: Copernicus GmbH.

Astagneau, P. C., Thirel, G., Delaigue, O., Guillaume, J. H. A., Parajka, J., Brauer, C. C., Viglione, A., Buytaert, W. & Beven, K. J. 2021 Technical note: hydrology modelling R packages – a unified analysis of models and practicalities from a user perspective. *Hydrology and Earth System Sciences* **25** (7), 3937–3973. Publisher: Copernicus GmbH.

BOE 2022 *Real Decreto 47/2022, de 18 de enero, sobre protección de las aguas contra la contaminación difusa producida por los nitratos procedentes de fuentes agrarias*.

CAP 2021 New CAP: 2023-27.

Chen, F.-W. & Liu, C.-W. 2012 Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy and Water Environment* **10** (3), 209–222.

Dávila-Hernández, S., González-Trinidad, J., Júnez-Ferreira, H. E., Bautista-Capetillo, C. F., Morales de Ávila, H., Cázares Escareño, J., Ortiz-Letechipia, J., Robles Rovelo, C. O. & López-Baltazar, E. A. 2022 Effects of the digital elevation model and hydrological processing algorithms on the geomorphological parameterization. *Water (Switzerland)* **14** (15), 2363.

FAO-ITPS 2020 *Protocol for the Assessment of Sustainable Soil Management*. Technical report, FAO, Rome.

Griffin, C. G., McClelland, J. W., Frey, K. E., Fiske, G. & Holmes, R. M. 2018 Quantifying CDOM and DOC in major Arctic rivers during ice-free conditions using Landsat TM and ETM+ data. *Remote Sensing of Environment* **209**, 395–409.

Gwak, Y. & Kim, S. 2013 Distribution characteristics of hydraulic properties on a mountainous hillslope. *Geosciences Journal* **17** (3), 339–352.

IDEAndalucia. 2023 *REDIAM. WCS Variables Climáticas de Precipitaciones*.

INSPIRE. 2023 *Enhancing Access to European Spatial Data*.

Isgró, M. A., Dolores Basallote, M. & Barbero, L. 2022 Unmanned aerial system-based multispectral water quality monitoring in the Iberian pyrite belt (SW Spain). *Mine Water and the Environment* **41** (1), 30–41.

Jasiewicz, J. & Metz, M. 2011 A new GRASS GIS toolkit for Hortonian analysis of drainage networks. *Computers and Geosciences* **37** (8), 1162–1173.

Jenson, S. K. & Domingue, J. O. 1988 Extracting topographic structure from digital elevation data for geographic information-system analysis. *Photogrammetric Engineering and Remote Sensing* **54** (11), 1593–1600.

Kaur, H., Nelson, K. A., Singh, G., Veum, K. S., Davis, M. P., Udawatta, R. P. & Kaur, G. 2023 Drainage water management impacts soil properties in floodplain soils in the midwestern, USA. *Agricultural Water Management* **279**, 108193.

Kim, S.-H., Lee, D.-H., Kim, M.-S., Rhee, H.-P., Hur, J. & Shin, K.-H. 2023 Systematic tracing of nitrate sources in a complex river catchment: an integrated approach using stable isotopes and hydrological models. *Water Research* **235**, 119755.

Li, Z. 2014 Watershed modeling using arc hydro based on DEMs: a case study in Jackpine watershed. *Environmental Systems Research* **3** (1), 11.

Lipp, A. G., de Caritat, P. & Roberts, G. G. 2023 Geochemical mapping by unmixing alluvial sediments: an example from northern Australia. *Journal of Geochemical Exploration* **248**, 107174.

Lizaga, I., Gaspar, L., Latorre, B. & Navas, A. 2020 Variations in transport of suspended sediment and associated elements induced by rainfall and agricultural cycle in a Mediterranean agroforestry catchment. *Journal of Environmental Management* **272**, 111020.

Lutz, N., Fogarty, J. & Rate, A. 2021 Accumulation and potential for transport of microplastics in stormwater drains into marine environments, Perth region, Western Australia. *Marine Pollution Bulletin* **168**, 112362.

Min, J.-H., Yun, S.-T., Kim, K., Kim, H.-S., Hahn, J. & Lee, K.-S. 2002 Nitrate contamination of alluvial groundwaters in the Nakdong River basin, Korea. *Geosciences Journal* **6** (1), 35–46.

MITECO.GOB. 2023 *Inventario de Presas Y Embalses*.

Mitran, T., Meena, R. S., Chakraborty, A., 2021 Geospatial technologies for crops and soils: an overview. In: *Geospatial Technologies for Crops and Soils, Volume 1 of Geospatial Technologies for Crops and Soils* (Mitran, T., Meena, R. S. & Chakraborty, A. eds.). Springer, Singapore, pp. 1–48.

O'Callaghan, J. F. & Mark, D. M. 1984 The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing* **28** (3), 323–344. Publisher: Academic Press.

Omran, A., Dietrich, S., Abouelmagd, A. & Michael, M. 2016 New ArcGIS tools developed for stream network extraction and basin delineations using Python and Java script. *Computers & Geosciences* **94**, 140–149.

Qiu, H., Niu, J., Baas, D. G. & Phanikumar, M. S. 2023 An integrated watershed-scale framework to model nitrogen transport and transformations. *Science of the Total Environment* **882**, 163348.

Ramos, M. I., Ortega, L., Cruz, J., Jurado, J. M. & Martínez, M. L. 2022 Methodology for the Traceability of Human Consumption Water using Maya 3D Animation Software. In *Proceedings of ACADEMICS WORLD INTERNATIONAL CONFERENCE*, Lisbon (Portugal), pp. 9–14.

Rheinwalt, A., Goswami, B. & Bookhagen, B. 2019 A network-based flow accumulation algorithm for point clouds: facet-flow networks (FFNs). *Journal of Geophysical Research: Earth Surface* **124** (7), 2013–2033. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018JF004827.

Sahoo, S. & Sahoo, B. 2019 Modelling the variability of hillslope drainage using grid-based hillslope width function estimation algorithm. *ISH Journal of Hydraulic Engineering* **25** (1), 71–78.

Shen, J. & Zhang, Q. 2018 Calculation of urban rainstorm waterlogging with drainage area. *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/ Geomatics and Information Science of Wuhan University* **43** (3), 356–363.

Sibanda, M., Mutanga, O., Chimonyo, V. G. P., Clulow, A. D., Shoko, C., Mazvimavi, D., Dube, T. & Mabhaudhi, T. 2021 Application of drone technologies in surface water resources monitoring and assessment: a systematic review of progress, challenges, and opportunities in the global south. *Drones* **5** (3), 84. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

SIOSE-Andalucía. 2016 *Mapa de Ocupación del Suelo en Andalucía*. SIOSE-Andalucía.

Still, D. A. & Shih, S. F. 1985 Using landsat data to classify land use for assessing the basinwide runoff index1. *JAWRA Journal of the American Water Resources Association* **21** (6), 931–940. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1752-1688.1985.tb00188.x.

Sufia Sultana, M. & Dewan, A. 2021 A reflectance-based water quality index and its application to examine degradation of river water quality in a rapidly urbanising megacity. *Environmental Advances* **5**, 100097.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L. B. d. S., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B. 2016 The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* **3** (1), 160018.

Ye, S., Zhang, Q., Yan, F., Ren, B. & Shen, D. 2022 A novel approach for high-quality drainage network extraction in flat terrains by using a priori knowledge of hydrogeomorphic features to extend DEMs: a case study in the Hoh Xil region of the Qinghai-Tibetan Plateau. *Geomorphology* **403**, 108138.

Yue, F.-J., Li, S.-L., Waldron, S., Wang, Z.-J., Oliver, D. M., Chen, X. & Liu, C.-Q. 2020 Rainfall and conduit drainage combine to accelerate nitrate loss from a karst agroecosystem: insights from stable isotope tracing and high-frequency nitrate sensing. *Water Research* **186**, 116388.

First received 13 April 2023; accepted in revised form 10 August 2023. Available online 29 August 2023