A fully automated and generic spatial discretization procedure for cultivated landscapes with human-made landscape elements

Philippe Lagacherie, Cécile Dagès, Ekatarina Zadonina, Jean-Christophe Fabre, Jérôme Molenat, Hervé Squividant, Bruce Thomas

corrected Proof

Journal of
Hydroinformatics

Journal of Hydroinformatics Vol 00 No 0, 1 doi: 10.2166/hydro.2022.048

A fully automated and generic spatial discretization procedure for cultivated landscapes with human-made landscape elements

Philippe Lagacherie [a,*], Cécile Dagès[a], Ekatarina Zadonina[a], Jean Christophe Fabre[a], Jérôme Molénat[a], Hervé Squividant[b] and Bruce Thomas[a]

[a] LISAH, Univ Montpellier, INRAE, IRD, Institut Agro, Montpellier, France
[b] SAS, INRAE, Institut Agro, Rennes, France
*Corresponding author. E-mail: philippe.lagacherie@inrae.fr

PL, 0000-0002-2948-1966

ABSTRACT

Landscape discretization is an essential hydrological modeling pre-processing step that comprises the numerical representation of different geographical objects considered in the modeling process and the connections between these objects within a graph structure. GROOV'Scape is a new landscape discretization procedure that (i) produces an oriented tree of a wide range of human-made landscape elements, such as plots, hedges, benches, grass stripes, and ditches, (ii) is fully automated to avoid local drainage anomalies that require user corrections, and (iii) is fast enough to enable sensitivity analysis and the interactive production of mediated modeling landscape management scenarios. GROOV'Scape was tested on a small agricultural catchment in south-west France (Doazit catchment, 9.05 km^2). The results show good agreement between the connections of areal units computed by GROOV'Scape and those observed in the field. The system exhibited substantial sensitivity to the user's choice of minimum area unit size and of digital elevation model (DEM) resolution, suggesting trial-and-error approaches are needed to reach the best landscape discretization for any given modeling purpose. Finally, we demonstrated the feasibility of using GROOV'Scape to find an optimum configuration of an infrastructure network with respect to a given ecosystem function.

Key words: agricultural catchment, connectivity, geomatics, hydrology, sensitivity analysis, spatial modeling

HIGHLIGHTS

- GROOV'Scape produces an oriented tree of human-made landscape elements.
- Fully automated and fast for sensitivity analysis and exploration of scenarios.
- Good agreement of GROOV'Scape-oriented trees with field observations.
- Trial-and-error approaches are needed to reach the best landscape discretization.
- GROOV'Scape helps to find an optimum configuration of an infrastructure network.

INTRODUCTION

Hydrological models, including catchment-scale models of water quantity and water quality, provide effective simulation tools for studying hydrological processes and predicting the effects of change on catchment hydrological response. Spatially distributed models are far from confined to the scientific field of hydrology or to the research of catchments. They have been progressively used to guide policy and water management decisions in the sustainable water management of large areas. In that regard, one expectation of water managers with regard to hydrological distributed models is to quantify the efficiency of scenarios in matters of the spatial organization of land use and landscape infrastructures (Viaud *et al.* 2005; Levavasseur *et al.* 2012; McDowell *et al.* 2014; Lebon *et al.* 2022). Hydrological models are also used in mediated modeling. This structured process is based on dynamic system thinking in which all stakeholders can work together to reach broad and deep consensus (van den Belt 2004).

Regardless of the distributed hydrological model, simulations are based on numerical spatial discretizations of a landscape. Spatial discretization consists in the identification and delineation of geographical objects that are different in nature (areal, linear, or punctual) and the connection between these objects within a graph structure (Heckmann *et al.* 2014). Given the above-cited expected uses of hydrological models, the landscape discretization procedure must meet the following criteria: it must be generic so that it can be applied to a wide range of landscapes; it must be fast and fully automated for use with interactive tools to quickly produce a large number of scenarios with attractive visualizing capabilities. It should be noted that this second feature is mandatory for undertaking hydrological model sensitivity analysis.

Although there have been some recent implementations of hydrological models that follow industrial standards and that fulfill some of the above-cited criteria (e.g., https://www.aquaveo.com), a generic, fast, and fully automated spatial discretization of the landscape remains difficult. The literature provides a great variety of landscape spatial discretization methods in relation to different hydrological models (Lagacherie *et al.* 2010; Pilz *et al.* 2017). Discretization into irregular-shaped polygons associated with semi-distributed models is by far the most common as it provides an acceptable compromise between precision of representation of within-catchment spatial variability and the requirements for achieving acceptable computing times (Pilz *et al.* 2017). Pilz *et al.* (2017) recently proposed an R package that fulfills most operational hydrological modeling criteria. It is free and open source, easily adaptable to a variety of hydrological models, and workflow can be fully automatized. However, this landscape discretization procedure is restricted to hillslope-based hydrological models, assuming that water transfer is only driven by the relief, from upslope areas to stream network, which may limit its application to the cultivated landscapes.

Indeed, cultivated landscapes are characterized by different landscape elements such as plots with different land uses, hedgerows, roads, benches, ditches, or artificial ponds that act as hydrological discontinuities. Many studies have shown the impact on water transfer within a catchment of such landscape elements as hedgerows (Caubel *et al.* 2003), ditches (Dagès *et al.* 2009), terraces (Preti *et al.* 2018), and small water reservoirs (Bouteffeha *et al.* 2014). Some procedures have been developed to take the specific impact of these elements into account when modeling water transfers. Lagacherie *et al.* (2010), for example, observed that a hydrological model accounting for the dense network of ditches of the modeled catchment simulated floods with slower propagation and lower intensity than those simulated by a hillslope-based hydrological model. Similarly, Viaud *et al.* (2005) demonstrated how a hedgerow network modified soil water content and evapotranspiration fluxes. Consequently, these human-made landscape elements need to be represented in mediated modeling hydrological models since they are possible water management levers, the impacts of which should be explicitly addressed and communicated to stakeholders.

Up to now, two landscape discretization procedures have been developed to deal with the human-made landscape elements of cultivated landscapes. These are MNTSurf (Aurousseau *et al.* 2009; Gascuel-Odoux *et al.* 2011) and GeoMHYDAS (Lagacherie *et al.* 2010). Both procedures represent the catchment as a set of plot outlet trees reaching the stream, with a given plot outlet tree representing the pattern of surface flow relationships between individual plots (both MNTsurf and GeoMHYDAS), between individual ditches, or between individual plots and individual ditches (GeoMHYDAS only). Some landscape elements are explicitly represented in each procedure – plots, hedges, and roads in MNTSurf, and plots and ditches in GeoMHYDAS. However, both these procedures are tailored for specific cultivated landscapes – European livestock breeding on bocage landscapes in the case of MNTsurf and Mediterranean vineyard landscapes in the case of GeoMHYDAS – and cannot, therefore, be considered fully generic, particularly with regard to the range of represented landscape elements having different impacts on the waterflow that cannot be handled by the procedures. Furthermore, none of the procedure is fully automated since the fusion between digital elevation model (DEM) and the vector representation of linear features generates drainage anomalies that have to be manually corrected.

This paper presents GROOV'Scape, a new landscape discretization procedure that has the following targeted characteristics: (i) it produces an oriented tree of landscape features, including a wide range of human-made landscape elements, such as plots, hedges, benches, grass stripes, and ditches, (ii) it is fully automated to avoid steps that generate drainage anomalies, (iii) it is adaptable to a wide range of flow representations across the catchment by producing a neutral representation of the connections between landscape elements, and (iv) it is fast enough to enable sensitivity analysis and the interactive production of mediated modeling of landscape management scenarios.

The following sections present the algorithm, the Doazit catchment on which further GROOV'Scape tests and applications will be conducted, and different tests designed to evaluate (i) the quality of representation of connections between plots, (ii) the impact of user-fixed minimum size parameters, and (iii) the impact of the spatial resolution of the DEM used as input. Finally, we provide an example of GROOV'Scape applied to a catchment to analyze the impact of linear infrastructure (LI) that considers both cumulated length and location of the plot connections to the catchment outlet.

METHOD

The GROOV'Scape algorithm

Like the former cultivated landscape procedures (MNTsurf and GeoMHYDAS), GROOV'Scape provides a representation of the areal and linear landscape elements that impact the water flow and documents the connections between these elements that will be considered in the further distributed hydrological modeling. GROOV'Scape outputs consist in three-vector coverages of sub-catchments (spatial unit (SU)), linear elements (LI), and Reachs (RS) with attribute tables documenting the connections between landscape elements (see full description below). To produce these outputs, GROOV'Scape uses as inputs a DEM, a soil map (optional), and a vector coverage for each landscape element.

To obtain the necessary speed required by mediated modeling or sensitivity analysis to which GROOV'Scape is particularly dedicated, a two-fold architecture of the algorithm was designed. The two procedures are the following:

- Building an oriented tree of elementary landscape features connected to each other. This step includes procedures that require substantial computing time that only has to be run once for a given catchment, assuming that changes in relief and plot boundaries are not considered in further landscape management scenarios.
- Building catchment management scenarios with different spatial distributions of land use and linear features within the catchment. This step is fast enough to be repeated a lot of times for exploring in a reasonable time a large number of management scenarios. These two procedures are detailed in the following.

Building an oriented tree of areal and linear features

This step is based on the definition of two types of 'potential' landscape objects from which the real landscape features can be geographically defined. These are potential areal objects (PAOs) and potential linear objects (PLOs). PAOs are within-parcel sub-catchments that can be optionally subdivided into homogeneous soil units. PLOs are portions of PAO boundaries to which the content of each previously defined PAO flows.

Figure 1 shows the different steps to obtain an oriented tree of areal and linear features. These steps are described in the following subsections.

- Step 1: Pre-processing

The inputs for this procedure are a vector coverage of plots and a DEM. As an option, a soil map can be overlaid onto the vector coverage of plots. The initial vector coverage of plots is then replaced by a vector coverage with polygons having homogeneous parcel and soil units. In both cases, all unit attributes selected by the user – at least the identifier of each entity – are propagated through the procedure for further use. Pre-processing consists of (i) building a raster of flow directions from DEM input using the AT least-cost search algorithm (Ehlschlaeger 1989) and (ii) converting the parcel vector coverage into a raster with spatial resolution and pixel locations that match those of the DEM.

Users should select the spatial resolution that enables the most accurate representation of parcel geometry while not dramatically increasing the calculation time. At this early step, a user may decide to simplify the parcel geometry by merging the smallest parcels that are considered to not have a strong impact on water flow.

- Step 2: Defining plot 'outlets' and 'receivers'

Parcel outlets and receivers are specific locations at which connections between parcels occur. The parcel outlet of a given plot (Aurousseau *et al.* 2009) corresponds to a cell where the surface water leaves the parcel. Parcel receivers are cells to which the parcel outlet flows. Parcel outlets and parcel receivers are necessarily located at the parcel boundaries. These rules were applied to flow direction rasters and parcels to label outlet and receiver pixels (Figure 1). Each plot outlet flows to a unique receiver, whereas a receiver may receive flow from more than one outlet.

- Step 3: Identifying sub-parcel catchments

Sub-parcel catchments are defined as the within-plot contributing areas of each outlet plot defined in Step 2. They are calculated using a classical catchment delineation algorithm. In this step, a large number of sub-catchments are delineated, as each plot outlet has its own catchment (Figure 1).

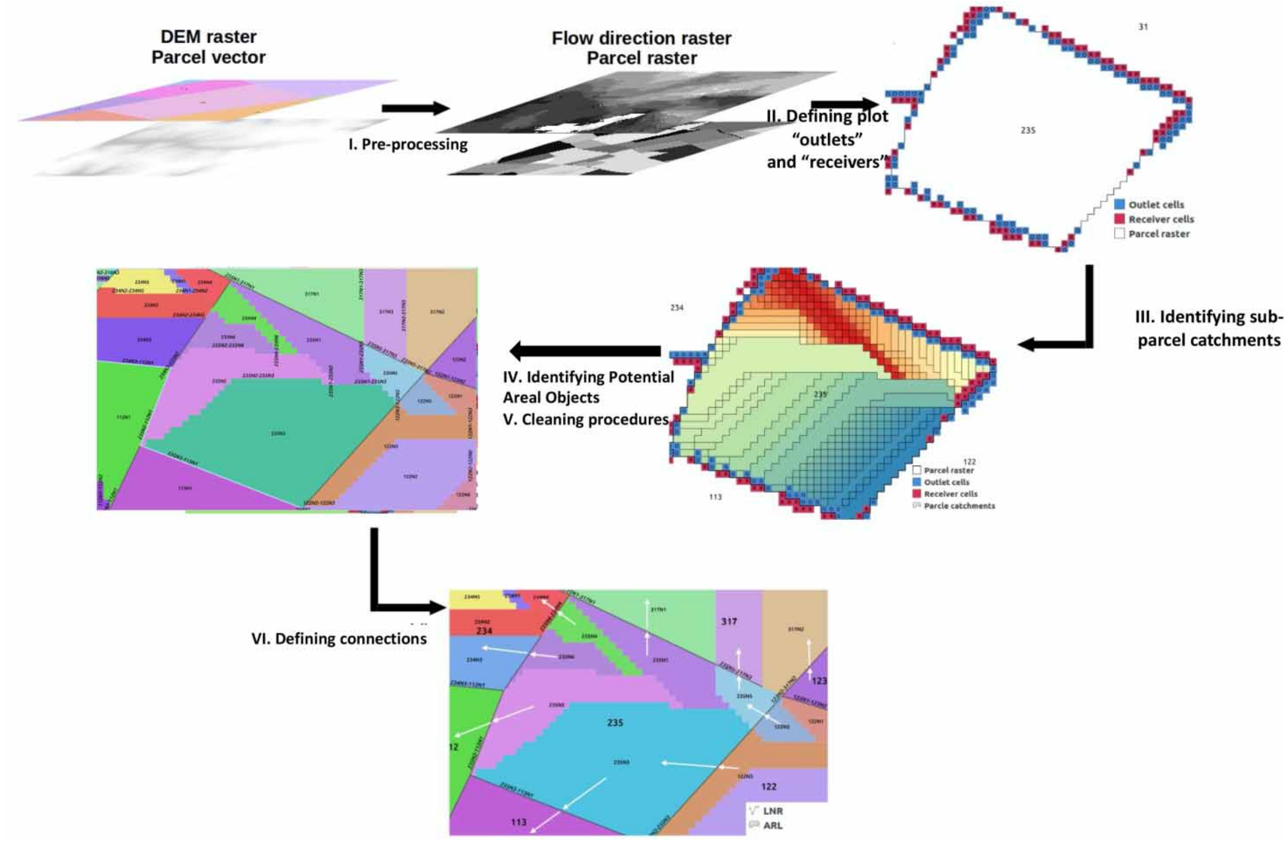

Figure 1 | Steps for building an oriented tree of areal and linear features.

- Step 4: Merging sub-plot-catchment into PAO

PAOs are obtained by the selective grouping of sub-parcel catchments so that all the locations of a given PAO flows to the same downhill PAO (Figure 1). The grouping is realized using an iterative procedure that progresses from sub-parcel catchments that flow outside the study area (downhill sub-parcel catchment) to those located in the highest sub-areas of the catchment. At each step, unique labels are allocated to sets of parcel sub-catchments that share the same parcel ID attributes and the same downhill PAO ID. Once all parcel sub-catchments have been labeled, a vectorization is performed, which produces the PAO polygon coverage. Similarly, those parcel outlets that belong to the same PAO receive a unique label, allowing PLO line coverage to be defined by vectorization.

- Step 5: Cleaning procedures

This step includes several geomatic operations to produce outputs that are better suited to the further allocation of scenario elements (land use and linear features). It consists of (i) restoring the initial geometry of parcel boundaries altered by the rasterization procedure and (ii) eliminating the small squared polygons at the parcel boundaries produced by the previous steps following a user-fixed minimum size of PAO units. This involves applying a selective overlay (Lagacherie *et al.* 2010) that intersects the resulting PAO delineations with the original parcels and allocating the small polygons that are below the minimum area size to a neighboring PAO based on their attributes and topological relationships between objects.

- Step 6: Defining connections

Connections between PAOs and PLOs are defined to represent the flow of water and chemical elements across the study area. In any given location, a PAO is connected to a PLO, which in turn is connected to a downhill PAO (white arrows on bottom map of Figure 1). This is indicated by specific items in the PAO and PLO attribute tables. In Step 4, these items were documented when parcel sub-catchments were merged into PAO and parcel outlets were merged into PLO. However, to restore the lost connections, a specific procedure is necessary due to the deletion of small polygons during the cleaning procedure (Step 5). An alternate connection is defined so that water flowing from the PAO can reach the stream network following the shortest detour. The principle is to connect to a PAO selected among the neighboring PAOs as having a downhill tree that does not include the PAO to be connected (so as to avoid loops). The connection is made through the PLO that represents the boundary between the two PAO to be newly connected.

Building scenarios of catchment management

Building a scenario involves deriving spatial unit coverage and LI coverage from the previously created PAO and PLO coverages. SU and LI coverages inherit the geometries and the connections of PAO and PLO, respectively, while having additional attributes that describe characteristics of the parcels (for SU) or the presence of anthropic linear features (for LI). A scenario can also include an optional third coverage that represents the reach network of the catchment while matching the geometry of the PLO coverage (RS coverage). Since this building scenario procedure does not include any geometry modifications that would require costly re-connections, it is highly time-efficient and can be activated multiple times, both in mediated modeling and in modeling as a water management decision tool.

The input data that together define the scenario to be tested are constituted by three optional vector coverages:

- A coverage of parcels with all the required characteristics to define a scenario (e.g., land use, agricultural practices, owners' details, etc.). This coverage should have the same geometry and the same labeling as the parcel coverage used in the previous procedure (Step 1).
- A coverage that includes anthropic linear features located at the parcel boundaries that may intercept flow, such as, for example, hedges or grass stripes.
- A coverage that includes the different linear features that together constitute the drainage network toward the catchment outlet. These linear features can be streams, ditches, or circulation features, such as roads and pathways.

In the following, we detail how to build the three coverages that define the scenario.

Building a SU coverage

SU coverage is duplicated from the PAO coverage produced in the previous step. The connections between PAOs and neighboring downstream linear features (LI) are also duplicated from those between PAO and PLO in the PAO attribute table. To satisfy some model requirements, there is the option to calculate connections with the neighboring downstream SU using transitivity (SU_A is connected to SU_B if SU_A is connected to LI_C and LI_C is connected to SU_B). The parcel characteristics that define a scenario are documented from the input coverage of parcels using a joint table procedure that considers the parcel label. The same procedure is used for the soil attribute if the initial input is a soil-parcel coverage (see Step 1). Table 1 gives an example of SU attribute documentation.

LI coverage

LI coverage is first duplicated from the PLO coverage produced in the previous step. The connections between LIs and their neighboring downstream areal features (SU) are also duplicated from those between PLOs and PAOs in the PLO attribute table. The attributes of LI segments are optionally documented by the coverage of anthropic linear features by overlaying these linear coverages on the LI coverage. In order to achieve this, a snap procedure is applied to pair the segments of the two coverages. This pairing allows the documentation of the LI coverage with new attributes giving, for each LI segment, the length and ratio of the linear anthropic feature. Table 2 provides an example of the documentation of LI attributes.

Table 1 | An example of an SU coverage attribute table

ID	ID_TO	Xc	Yc	Area	Slope	LC	soil_id	Tillage	Spraying
1	LI#189	8,075	5,111	9,667	0.67	grassland	7	no	no
2	LI#195	8,114	5,558	14,023	3.5	maize	7	yes	yes
3	LI#199	6,267	3,487	9,621	6.6	grassland	4	no	no
4	LI#202	7,615	5,933	21,235	2.5	maize	7	no	no
5	LI#224	7,435	4,761	6,060	5.6	orchard	4	no	yes

Note that the first six attributes are defined or calculated by GROOV'Scape, whereas the last three are derived from initial plot coverage.
ID, SU unique identifier; ID_TO, identifier of down unit (here LI units); Xc and Yc, coordinates of the SU centroïd; LC, Land Cover; soil_id, an identifier of soil type; tillage, occurrence of tillage practice on the plot; spraying, occurrence of pesticide spraying practices on the plot.

Table 2 | An example of an LI coverage attribute table, with hedges as linear infrastructure

ID	ID_TO	Xc	Yc	Length	Flowdist	Hedges
1	SU#26	6,520	5,111	22.4	45.8	0.7
2	RS#22	6,101	4,152	70.3	83.4	0
3	NULL	6,750	5,622	32.9	NULL	1
4	SU#344	7,722	5,198	143	76.1	0
5	SU#60	8,041	5,007	117.7	17.6	0.2

Note that the first six attributes are defined or calculated by GROOV'Scape, while the last, named with the type of infrastructure and corresponding to the coverage rate of the infrastructure on the LI, can be computed by GROOV'Scape or can be derived from the coverage of the linear infrastructure, in this case the hedge.
ID, LI unique identifier; ID_TO, identifier of down unit (here SU or RS units); Xc and Yc, coordinates of the center of the LI segment; flowdist: the distance along the flow direction; type, type of anthropic infrastructure (hedge, ditch, etc.); CoverRate, the ratio of the PLO covered by the linear anthropic infrastructure.

RS coverage

RS coverage is built from the LI coverage and drainage network coverage provided by the user. A snap procedure is applied to identify the PLO segments that match the segments of the drainage network coverage. These PLO segments are then copied into the new RS coverage and the cover rate of the segments is calculated. A connection between each RS segment and its neighboring upstream SU is defined by duplicating the connections between the corresponding PLO and PAO. This connection is documented by an attribute added to the SU attribute table. Connections with the neighboring downstream RS are computed toward the outlet by an iterative process that forces the orientation of each RS segment from the source to the outlet and registers the label of the downstream unit (Lagacherie *et al.* 2010, m.toporeach procedure). Connections between RSs allow for the calculation of drainage area and width for each RS segment. Table 3 gives an example of SU attributes documentation.

Tools and the software development approach

GROOV'Scape software is available as a standalone command line application or as a simulator for the OpenFLUID software environment for spatial modeling (Fabre *et al.* 2010).

The command line application allows either or both of the preparation and scenario procedures of the GROOV'Scape algorithm to be launched. It can be run in a command line terminal with the following usage:

```
groovscape <procedure> <datapath >
```

where the <procedure> command can be discretization or scenario depending on the procedure to be executed, or complete to chain the execution of both procedures. The <datapath> command is the path where the dataset to be processed is located on a disk.

The dataset must contain the needed input vector and raster GIS files, and also an usersetparams.json file that contains the run configuration for the GROOV'Scape algorithm.

GROOV'Scape software mainly consists of a library containing the algorithm parts. It has been developed in C++ language using the C++11 standard. The C++ language was chosen for its performance and its ability to handle complex software design with an object-oriented approach.

Table 3 | An example of an RS coverage attribute table

ID	ID_TO	Xc	Yc	Length	Width	Height	soil_id
1	RS#40	5,686	4,982	64.3	2.7	3.1	2
2	RS#46	6,118	5,504	141.6	2.7	3	2
3	NULL	5,587	6,092	10.5	3.5	3.2	1
4	RS#35	5,629	4,810	7.9	2.7	3	2
5	RS#1	5,670	4,947	14.2	2.7	2.9	2

Note that the first five attributes are defined or calculated by GROOV'Scape, while the last three are derived from the linear infrastructure coverage.
ID, RS unique identifier; ID_TO, identifier of down unit (here only RS units); Xc and Yc, coordinates of the center of the RSsegment; soil_id, an identifier of the soil type.

The GROOV'Scape library relies on reference libraries and tools used in the field of spatial processing: GDAL for handling spatial information, and GEOS and GRASS (as an external command) for spatial processing. It also relies on RapidJSON and Qt libraries for utility functions, and on the OpenFLUID platform for framework facilities and for spatial coupling with other models.

The GROOV'Scape source code is composed of approximately 12,000 lines of C++ and is available as open-source software under the terms of the GPLv3 license. At present, it has only been built and tested on the Linux system using the Ubuntu distribution.

RESULTS

The case study: the Doazit catchment

In order to test and illustrate the GROOV'Scape algorithm, it was applied to the Doazit catchment. In this section, we present the catchment and the data used for running GROOV'Scape.

Description of the Doazit catchment

The Doazit catchment is located in the Landes department of south-west France. It is part of the catchment of the Adour River. The Doazit catchment is 9.05 km^2. The cultivated landscape is characterized by medium sized plots (mean area of 2 ha) mainly devoted to maize and meadows; a dense network of ditches (58 km) connect the plots to the river. The Doazit catchment is in a hilly region with clayey-siliceous soils that are sensitive to crusting.

Doazit catchment management problem

The sensitivity of soils to crusting combined with intense rainfall during summer storms leads to large Horton overland flow within fields. This is associated with both erosion and pesticide contamination. Indeed, the use of a pre-emergent herbicide for the maize crop in spring induces contamination of drinking water abstractions. One out-field management option is to intercept overland flow at the field outlet and to enable its infiltration. Linear agro-ecological infrastructures such as filter strips, hedges, and ditches located around the edges of plots can be used in mitigation strategies (e.g., Lacas *et al.* 2005; Dollinger *et al.* 2015). It is envisaged that such infrastructures will be developed in the Doazit catchment, thus requiring and understanding of the optimal amount and locations of filter strips.

Basic data

As described above, GROOV'Scape uses a DEM and a vector coverage of parcels as input. A 5 m DEM resolution covering the Doazit perimeter was obtained from the RGE ALTI® database of the French National Institute of Geography (IGN 2018). Parcel coverage was elaborated from a field survey performed in 2014. The survey also identified the hydraulic connections between parcels and the parcel boundaries that corresponded to elements of the catchment drainage network. The former was used to evaluate the connections estimated by GROOV'Scape. The latter was used to derive a drainage network which was used as an example of GROOV'Scape application.

GROOV'Scape testing

Ground validation of connections

An oriented PAO and PLO tree was produced by applying it to the Doazit catchment according to the first part of the algorithm described above. Connections between PAO and PLO were first translated into connections between PAO (PAOx flows into PAOy if PAOx flows into PLOz and PLOz flows into PAOy) to enable further comparisons with connections between field plots obtained from the field survey. Since the spatial resolution of the landscape discretization produced by GROOV'Scape (PAO corresponding to within-plot areal units) was finer than that considered by the field surveyor (for entire plots), the comparisons between field and GROOV'Scape connections were not straightforward. Therefore, three comparison modalities were defined:

- Full agreement: The connection between the PAOs according to the algorithm is the same as the connection given by the surveyor for plots containing the PAO.
- Partial agreement: The connection between the PAOs according to the algorithm is not the same as the connection given by the surveyor for plots containing the PAO. However, some PAOs contained within the same plot have a connection that is the same as that of the field surveyor.

- Disagreement: none of the two previous criteria are verified.

Five hundred selected connections representing 19% of all the connections produced by GROOV'Scape were randomly selected and successively examined to determine these modalities.

The comparisons of connections between areal units according to GROOV'Scape and those according to the field survey showed 42, 50, and 8% of full agreement, partial agreement, and disagreement, respectively. Partial agreement was interpreted to be the result of the difference in spatial resolution between compared products. An examination of the 8% of disagreed locations, two main causes were identified:

- Complex slope orientations. A complex slope orientation occurred when the direction of the maximum slope did not clearly follow plot boundaries. In that case, a field surveyor must take account of multiple possible downhill plots, making it difficult to select the downhill path with the greatest slope gradient. The decision did not agree with that of GROOV'Scape, which is based on slopes calculated from the DEM.
- Additional drivers of flow directions. A field surveyor may consider drivers of flow direction other than relief, such as dead furrows and other features related to tillage directions (Ludwig *et al.* 1995).

Sensitivity of discretization to minimum size parameter

The user must define the minimum size of PAO units considered in the discretization algorithm (Step 5). To study the impact of the user's decision, GROOV'Scape considered ten different minimum sizes when run on the Doazit catchment, from the smallest parcel to median parcel sizes. The ten GROOV'Scape outputs were compared (Figure 2, bottom) by calculating the weighted mean number of splits per parcel (WMNSP) as follows:

$$WMNSP = \frac{\sum_{i=1}^{i=p} a_i \times NPAO_i}{\sum_{i=1}^{i=p} a_i} \quad (1)$$

with $NPAO_i$ being the number of PAOs in parcel i and a_i being the parcel area. For some tested minimum sizes, the distributions of PAO sizes were also considered and compared to the initial size of the parcels (Figure 4, top).

As expected, the average number of parcel splits decreased as the fixed minimum size of the split increased. However, there was less of a decrease as soon as the minimum size went beyond 6,928 m^2.

A large minimum size had the positive effect of reducing the distribution queue of large parcels observed on initial distribution (Figure 2, top right). However, some large parcels remained, corresponding to parcels with non-complex relief that did not need to be split to represent the direction of water flux.

A small minimum size (Figure 2, top left) generally increased the splits of small or medium sized parcels, provoking very unbalanced distributions while not significantly reducing the distribution queue of large parcels. This is because small parcels were located in areas of complex relief that needed to be split more to finely represent the direction of water flux.

Another sensitivity analysis on spatial resolution of DEM used as input of GROOVS'scape was also performed. It is presented in the Supplementary Material.

An example of GROOV'Scape application

This section aims to illustrate the possibility of using GROOV'Scape for sensitivity and optimization analysis of spatial distribution scenarios of anthropic features within a catchment. Linear infrastructures, such as ditches, vegetated filter strips, hedges, and fascine, modify water and matter overland flow within hillslopes depending both on their properties (e.g., Schultz *et al.* 1995; Dollinger *et al.* 2015, 2016; Yu *et al.* 2019) and on their location within the catchment, mainly due to their capacity to intercept and reroute flows (e.g., Viaud *et al.* 2005; Levavasseur *et al.* 2012). For example, vegetated filter strips are known to have large infiltration rates and, therefore, to reduce surface overland flow and concentration of MES, phosphate, and pesticides (e.g., Lacas *et al.* 2005). Higher sorption capacities of vegetated filter strip soils compared to plot soils also contribute to enhancing the interception of pesticides by vegetative strip filters (e.g., Benoit *et al.* 1999).

We used the example of a land planning program aimed at granting a given quantity of non-channelized infrastructure (expressed in total length) to mitigate the risks associated with overland flow (flood or water pollution). GROOV'Scape outputs can be used as distributed model inputs. However, to remove the intrinsic efficiency of the infrastructures and the choice

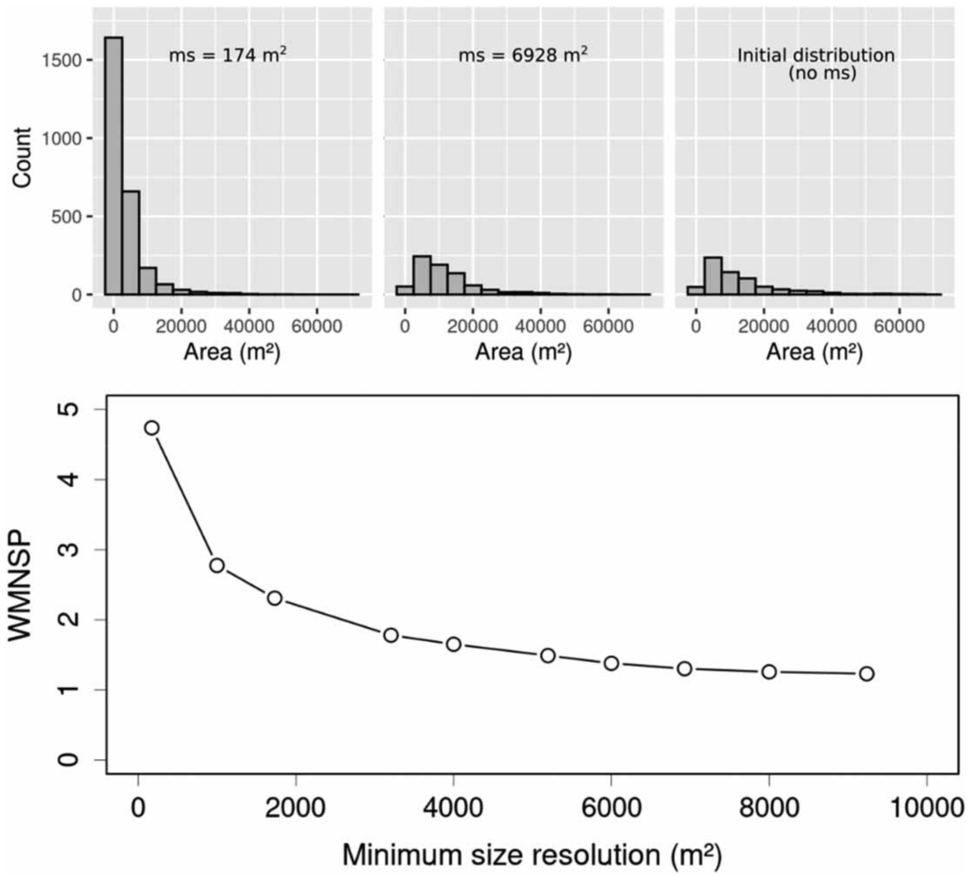

Figure 2 | The impact of minimum size parameter on the parcel splits: (bottom) weighted average of splits per parcels with the fixed minimum size (m^2), (top) from left to right: distribution of PAO area with ms = 174 m^2 (the minimum size of parcels), distribution of PAO area with ms = 6,928 m^2 (the median size of parcels minus 25%), and initial distribution of parcel areas.

of indicators or models that represent it (which is not the intention of this paper to discuss), the different scenarios were evaluated by counting surfaces with runoff that could not be intercepted by any infrastructure. For simplicity, for the remainder of this paper, these surfaces are called 'areas directly connected to the outlet.'

The topography of the Doazit catchment is suitable for examining overland flow issues and was used to support the derivation of different scenarios. These scenarios were envisaged with increasing investments, enabling increasing total infrastructure lengths (11 total lengths from 500 m to 65 km). For each total length, 100 repetitions with random selections of the PLO partially equipped with infrastructure were considered to study the impact of its spatial distribution within the Doazit catchment. The coverage of a PLO by the infrastructure randomly varied from 60 to 100%, with the exception of the last two pieces of infrastructure, for which coverage was adjusted to meet both cumulative target length and percentage coverage requirement criteria.

Having run the first part of GROOV'Scape that built a stable PAO and PLO oriented tree once, the second part of GROOV'Scape was run 100 times for each total length (i.e., 1,100 runs). The first part of the process took 60 minutes on a computer with dual Intel Xeon E5-2690 2.90 GHz CPU (used as a single core architecture) and 128GB of RAM, while each run of the second part took one minute. Although GROOV'Scape outputs can be used as distributed model inputs, we used them in this example to calculate a very simple indicator that is the proportion of area where overland flow cannot be intercepted by the infrastructure. This indicator was obtained by summing the specific contributive areas of PLO that were not or were only partially equipped with infrastructure. For the latter, we assumed that the interception was proportional to infrastructure coverage.

As the landscape structure is represented by an oriented tree, the spatial calculation of these indicators is based on recursive path algorithms. These algorithms are particularly well adapted to this type of graph and are fully documented in graph theory literature (Cormen *et al.* 2001). They provide powerful methods for depth or width traversal of the graph, visiting the graph nodes (which represent spatial objects) by following the edges of the graph (which represent the connections between spatial objects). In this application, we implemented recursive algorithms to perform the traversal of the graph, starting from the outlet and using the opposite direction of the edges. Depending on the characteristics of the computed indicator, the recursive algorithms use preorder or postorder depth-first methods to process the spatial objects.

As expected, the proportion of areas directly connected to the outlet decreased with the total length of infrastructure, i.e., with the amount of investment (Figure 3). A clear drop in the minimal values of the proportions of area directly connected to the outlet (lower intervals of the distribution) was observed at 2,000 m, leading to the conclusion that this length is a good basis for finding compromises that may best optimize investment. In addition, it was interesting to note that the variability of proportions of areas directly connected to the outlet for a given total length, i.e., the impact of the spatial distribution of infrastructure within the Doazit catchment, was only noticeable for small to intermediate total lengths (between 1,000 and 10,000 m). As Colin *et al.* (2012) have shown, in situations where the cumulative length of the infrastructure is large, the degree of freedom of assignment of the infrastructure to the PLO network is small. Therefore, variability of the spatial configuration of the infrastructure is limited, implying low variability of the connected areas.

To establish the best investment scenario, a focus was performed on the impact of the spatial distribution of infrastructures within the Doazit catchment. For doing that, the total length of pieces of infrastructure was fixed at 2,000 m, this length having been observed before (Figure 3) as that containing interesting compromises. Instead of selecting the spatial distribution of infrastructure at random, the selection was constrained by the relative position from the hydrographic network of the catchment. Each candidate PLO for hosting infrastructure was first characterized by its membership to one of the distance quartiles to the hydrographic network. Infrastructure locations were then selected by applying unbalanced stratified PLO sampling using membership to quantile as strata (Table 4). For random selection, the PLO coverage percentage by the infrastructure was randomly set between 60 and 100%. Each selection was repeated 100 times to obtain robust statistics. An R script was built to perform these selections.

Figure 4 shows that the unbalanced stratified sampling of the infrastructures based on the relative positions of the PLO generated more variability than pure random sampling (Figure 4, first white box on left) with regard to water disconnection. The closer to the hydrographic network the infrastructure was clustered (different boxplot colors in Figure 4), the greater the

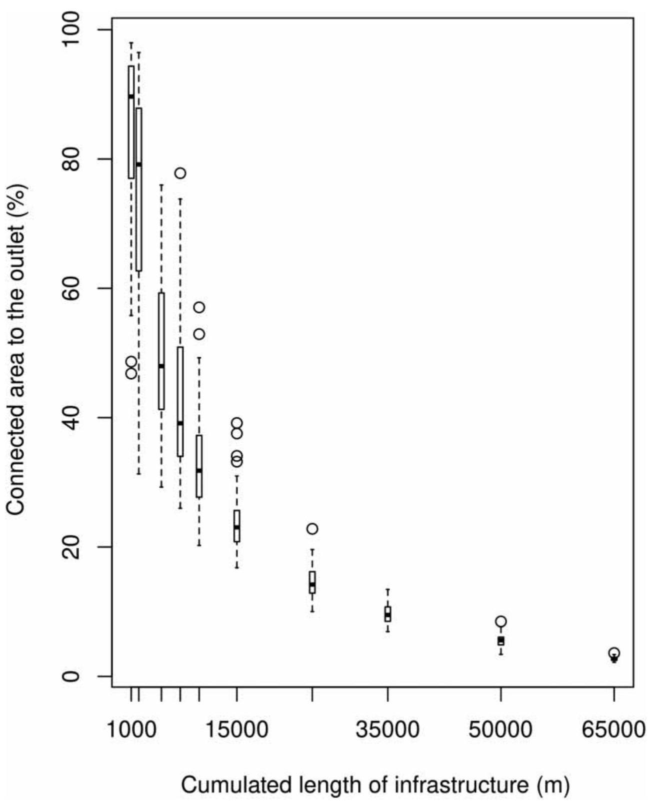

Figure 3 | Evolution of the percentage of the catchment directly connected to the outlet with the total lengths of pieces of infrastructure.

Table 4 | Unbalanced stratified sampling of infrastructures within the Doazit catchment: required proportions of infrastructure located on each quantile of distances to the outlet

Sampling	First quantile (%)	Second quantile (%)	Third quantile (%)	Fourth quantile (%)
100.1	100	0	0	0
100.2	0	100	0	0
100.3	0	0	100	0
100.4	0	0	0	100
85.1	85	5	5	5
85.2	5	85	5	5
85.3	5	5	85	5
85.4	5	5	5	85
70.1	70	10	10	10
70.2	10	70	10	10
70.3	10	10	70	10
70.4	10	10	10	70
55.1	55	15	15	15
55.2	15	55	15	15
55.3	15	15	55	15
55.4	15	15	15	55
40.1	40	20	20	20
40.2	20	40	20	20
40.3	20	20	40	20
40.4	20	20	20	40

chance of finding an effective infrastructure configuration that limited direct flow connection. This is an expected effect since the clustering of infrastructure in the upper part of the hillslope prevents the interception of runoff in lower parts of the hillslope while using infrastructure that is too long to intercept the flow of areas that are too small. This effect was attenuated as infrastructure clustering was reduced (groups of boxplots from left to right in Figure 4) but solutions to intercept a greater part of the flow could still be found with moderate clustering that could be more easily applied in the catchment (see lower bars of the boxplots in Figure 4). This is illustrated in Figure 5 (left) for a spatial pattern with 55% of infrastructure near the hydrographic network (55.1 scenario in Table 4) leading to the direct connection of only 35% of the Doazit catchment. Figure 5 (right) illustrates, for the same scenario, an example of an unsatisfactory situation in which the same scenario (55.1.) leads to the direct connection of 85% of the Doazit catchment.

DISCUSSION

GROOV'Scape is an open-source software that implements a new spatial discretization of cultivated landscapes based on two main procedures. An analysis of a first application to a catchment in France showed that the resulting spatial discretization from the connection between plots is relevant. Its relevance requires further analysis through the application of GROOV'Scape to other case studies, covering a wider range of parcel configurations, soil variability, topography, and infrastructure patterns. In the following, we emphasize the three key characteristics of the new form of processing implemented using GROOV'Scape: (i) processing genericity, (ii) automation and speed, and (iii) discretization sensitivity to users' choice. Finally, we discuss steps toward moving to a fully operational tool.

Key characteristics of GROOV'Scape

Genericity

A procedure dedicated to agricultural landscape discretization would qualify as generic if it could be applied to a large variety of landscape types. In the case of GROOV'Scape, genericity comes first from spatial discretization based on real landscape

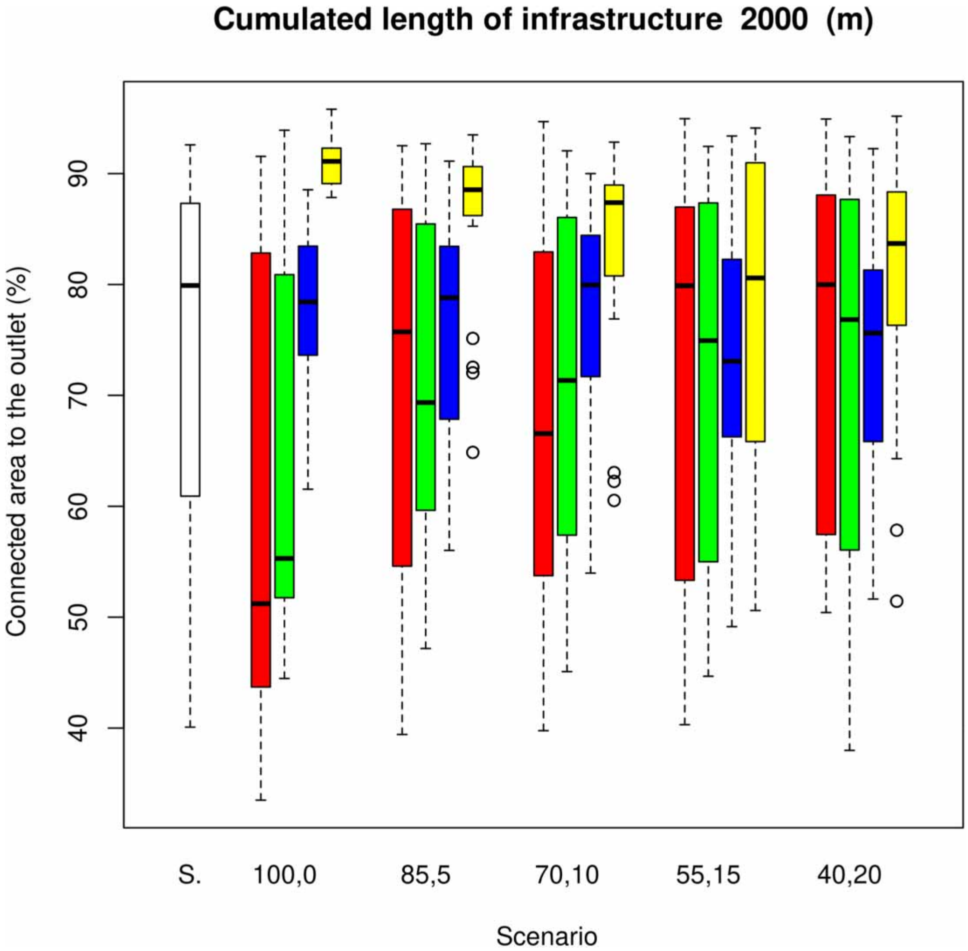

Figure 4 | Proportion of areas directly connected to the outlet for the different spatial distribution of infrastructure scenarios (total length 2,000 m, 100 replications), with S as the stochastic distributions; X and Y as the unbalanced scenarios with X being the percentage in one quartile of distances to the hydrological network and Y being the percentage in each of the other three quartiles. The colors mark the quartile in which the X% is found: the first (red), second (green), third (blue), and fourth (yellow). Stochastic distributions are represented by the white bar. Please refer to the online version of this paper to see this figure in color: https://dx.doi.org/10.2166/hydro.2022.048.

objects (agricultural parcels and, optionally, soil units) to be considered irrespective of the nature of the agricultural landscape. The diversity of anthropic infrastructure considered in the building of landscape scenarios, such as grass strips, ditches, and hedges, also contribute to GROOV'Scape's genericity. Such infrastructure creates water and matter flux discontinuities in the landscape, which must be taken into account in hydrological modeling. Accounting for these discontinuities gives GROOV'Scape an advantage over other processing software (e.g., Pilz *et al.* 2017).

However, in the present version, only LI located at parcel borders are considered, which may be a limitation in agricultural landscapes that also include anthropic LI located within parcels. The most straightforward solution for overcoming this limitation would be the introduction, in the pre-processing step (Step 1, Figure 1), of an operation to split every parcel according to the pattern of real or planned inner LI. A counterpart to such a solution would be a likely increase of computation time induced by this increase in the number of parcels.

A second limitation to GROOV'Scape's genericity is that some common forms of infrastructure within the agricultural landscape have not yet been implemented into the current version. The next step in the development of GROOV'Scape should be the integration of small reservoir networks in discretization processing. Such small reservoirs, used for irrigation, are found in many parts of the world (Lebon *et al.* 2022). Therefore, it will be necessary to consider the different connection configuration between reservoirs and other landscape elements, as small reservoirs located within or between parcels, or across streams.

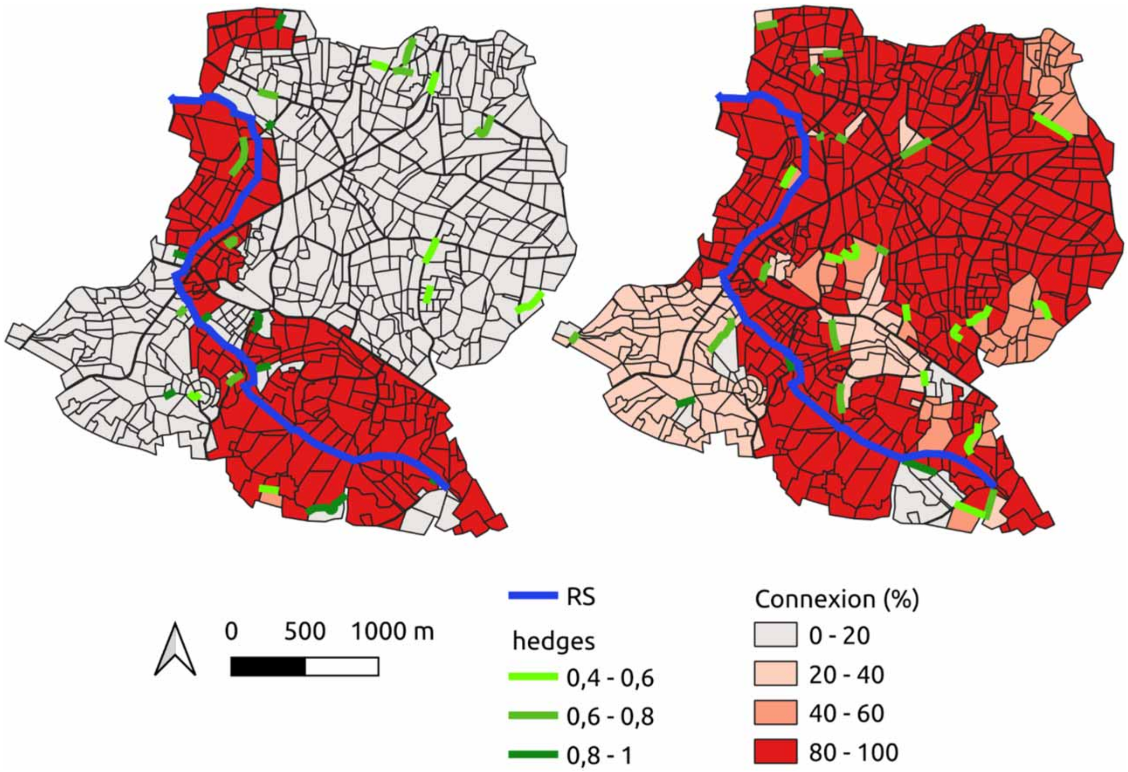

Figure 5 | Examples of spatial distribution of parcel area directly connected to the outlet (% of connection). Hedges are represented in green depending on the coverage percentage of the PLO. RS (in blue) corresponds to the reach segments composing the stream network. Please refer to the online version of this paper to see this figure in color: https://dx.doi.org/10.2166/hydro.2022.048.

Automation and speed

For both scientific and operational water management purposes, a central issue is the search for the optimum configuration of an infrastructure network with respect to a given ecosystem function. Given the complexity of the problem, this cannot be carried out using mathematical optimization approaches. It requires the exploration of the spatial locations of these infrastructures and the effect of location on the geometric or functional properties of the catchment, as illustrated both by the application to the Doazit catchment and in other studies (Levavasseur *et al.* 2012; Gumiere *et al.* 2014). Such an approach requires an exploration of the widest possible field of possibilities and, thus, from a modeling point of view, the generation of a large number of spatial discretizations relative to all spatial configurations. Depending on the scientific or operational objective, the exploration may seek to determine the optimum density of infrastructure (Figure 3) or, for a given infrastructure density, the optimal spatial location of that infrastructure (Figures 4 and 5). In this respect, automation and speed are major assets of GROOV'Scape.

GROOV'Scape's fully automated processing relies on the user's choice of minimum PAO size area. Automation allows the production of a large quantity of discretization (Figures 4 and 5). Non-automated *ad hoc* processing and discretization, involving successive manual and GIS-software-based operations, remain a widespread practice among hydrological modeling scientists and engineers. A tool that automates the processing considerably reduces subjectivity and increases the traceability of the processing conducted. In addition, it increases the determinism of the discretization operations and products. The discretization produced from a given data source (parcel vector, DEM, etc.) is unique for every parameter set.

The division of the procedure into two steps separates the time-consuming procedures that need to be launched once from the ones generating management scenario that are speed enough for being launched a great number of times. This significantly speeds up the exploration of a large set of scenario. A comparison of the computation time of other tools cannot be made as accounts of other spatial discretization processes based on parcel cover and linear infrastructures are not available in the literature. However, the application to the Doazit catchment reveals a time-saving on orders of magnitude as the first procedure does not have to be performed for each new discretization. Performing the second procedure to produce 100 discretizations took 60 times less time than the first step of building the PAO- and PLO-oriented tree connection.

Sensitivity to user's choice

The search for a compromise between the number of objects in the spatial discretization and the quality of the simulation targeted by the hydrological model is a crucial issue for the modeler. A representation at too fine a spatial resolution, generating a large number of objects, leads to long computation times, and potentially limits the application of the model to large areas or, for a given catchment, to a large number of configurations. Conversely, a representation at too coarse a spatial resolution can lead to a degraded representation of hydrological processes, yielding unrealistic simulations. The spatial discretization and the number of objects should depend on the user-fixed minimum size area of the spatial units, namely PAO in GROOV'Scape, and on the spatial resolution of the DEM. Following the discretization produced by GROOV'Scape, the number of PAOs produced was, as expected, dependent on the minimum size, with an optimum size of 6,928 m^2 in the applied case, above which the average split of parcels decreases slightly with size (Figure 3). Besides, sensitivity analysis of the spatial resolution of the DEM (Supplementary Material) showed that it was difficult to draw general recommendations for optimal resolution Depending on whether the simulated variable of interest (runoff, dispersion of contaminants, water table, etc.), the optimum minimum size value can vary for any given catchment. In this respect, the automation and speed of GROOV'Scape processing is an advantage since it allows the user to produce different representations according to the chosen parameter values and to analyze the sensitivity of the hydrological simulations.

CONCLUSION

GROOV'Scape is a software that implements a new landscape discretization procedure that can account for a wide range of human-made landscape elements. It is fully automated and fast enough to enable sensitivity analysis and the interactive production of mediated modeling landscape management scenarios.

The first application to a real catchment (Doazit catchment, France) showed that the spatial discretization analyzed from the connection between plots is relevant. This relevance needs to be further analyzed by applying GROOV'Scpae to other case studies, covering a wider range of parcel configuration, soil variability, topography, and infrastructure patterns.

This first application also analyzed the sensitivity of GROOV'Scape to input DEM resolution and the minimum size parameter needed to guide users toward appropriate modeling practices.

Finally, a test of GROOV'Scape in a simple example that aimed to identify an optimized scenario of spatial distribution of anthropic elements within the Doazit catchment showed its potential utility for sensibility analysis and mediated modeling involving spatially distributed hydrological models.

In the future, several improvements for increasing the genericity of GROOVS'cape will be necessary, e.g., accounting for landscape elements located inside the plots such as ponds (Lebon *et al.* 2022). GROOV'Scape open also new perspectives for massive numerical experiments with hydrological models. In this regard, one of the key issues in hydrology is to better understand and quantify the relationships between landscape infrastructures and catchment functioning. The impacts on stream flow or water use efficiency of spatial organization of crops and related cultural practices, or of the geometry of LI network, such in the Doazit catchment example, constitute questions that GROOVS'cape will help to answer.

ACKNOWLEDGEMENTS

This project was granted by OFB (French Office of Biodiversity) projects 'BV Service' and 'Retenues'. We thank Armel Thöni for his contribution to the software development of GROOV'Scape. We thank Martin André for his contribution to the test of GROOV'Scape on the Doazit Catchment. We thank Adrien Chassan (Coopérative MaïsAdour) for providing the geographical data needed for testing GROOV'Scape on the Doazit Catchment.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Aurousseau, P., Gascuel-Odoux, C., Squividant, H., Trepos, R., Tortrat, F. & Cordier, M. O. 2009 A plot drainage network as a conceptual tool for the spatial representation of surface flow pathways in agricultural catchments. *Computer & Geosciences* **35**, 276–288.

Benoit, P., Barriuso, E., Vidon, P. & Réal, B. 1999 Isoproturon sorption and degradation in a soil from grassed buffer strip. *Journal of Environmental Quality* **28**, 121–129.

Bouteffeha, M., Dagès, C., Bouhlila, R. & Molénat, J. 2014 A water balance approach for quantifying subsurface exchange fluxes and associated errors in hill reservoirs in semiarid regions. *Hydrological Processes* **29** (7), 1861–1872.

Caubel, V., Grimaldi, C., Merot, P. & Grimaldi, M. 2003 Influence of a hedge surrounding bottomland on seasonal soil-water movement. *Hydrological Processes* **17**, 1811–1821.

Colin, F., Moussa, R. & Louchart, X. 2012 Impact of the spatial arrangement of land management practices on surface runoff for small catchments. *Hydrological Processes* **271**, 255–271.

Cormen, T., Leiserson, C. & Rivest, R. 2001 *Introduction to Algorithms*, 1st edn. MIT Press, Cambridge, Massachusetts.

Dagès, C., Voltz, M., Bsaibes, A., Prévot, L., Huttel, O., Louchart, X., Garnier, F. & Negro, S. 2009 Estimating the role of a ditch network in groundwater recharge in a Mediterranean catchment using a water balance approach. *Journal of Hydrology* **375**, 498–512.

Dollinger, J., Dagès, C., Bailly, J. S., Lagacherie, P. & Voltz, M. 2015 Managing ditches for agroecological engineering of landscape: a review. *Agronomy for Sustainable Development* **35**, 999–1020.

Dollinger, J., Dagès, C., Negro, S., Bailly, J.-S. & Voltz, M. 2016 Variability of glyphosate and diuron sorption capacities of ditch beds determined using new indicator-based methods. *Science of the Total Environment* **573**, 716–726.

Ehlschlaeger, C. R. 1989 Using the AT search algorithm to develop hydrologic models from digital elevation data. In: *Proceeding of the International Geographic Information System (IGIS) Symposium*, Baltimore, MD, pp. 275–281.

Fabre, J. C., Louchart, X., Moussa, R., Dagès, C., Colin, F., Rabotin, M., Raclot, D., Lagacherie, P. & Voltz, M. 2010 OpenFLUID: a software environment for modelling fluxes in landscapes. In: *LANDMOD2010*, 3–5 February 2010, Montpellier.

Gascuel-Odoux, C., Aurousseau, P., Doray, T., Squividant, H., Macary, F., Uny, D. & Grimaldi, C. 2011 Incorporating landscape features to obtain an object-oriented landscape drainage network representing the connectivity of surface flow pathways over rural catchments. *Hydrological Processes* **25** (23), 3625–3636.

Gumiere, S., Bailly, J. S., Cheviron, B., Raclot, D., Le Bissonnais, Y. & Rousseau, A. 2014 Evaluating the impact of the spatial distribution of land management practices on water erosion: case study of a Mediterranean catchment. *Journal of Hydrologic Engineering* **20** (6), 1–10.

Heckmann, T., Schwanghart, W. & Phillips, J. D. 2014 Graph theory-Recent developments of its application in geomorphology. *Geomorphology* **243**, 130–146.

IGN 2018 *RGE ALTI® Version 2.0 – Descriptif de contenu – Septembre 2018*.

Lacas, J. G., Voltz, M., Gouy, V., Carluer, N. & Gril, J. J. 2005 Using grassed strips to limit pesticide transfer to surface water: a review. *Agronomy for Sustainable Development* **25**, 253–266.

Lagacherie, P., Rabotin, M., Colin, F., Moussa, R. & Voltz, M. 2010 Geo-MHYDAS: A landscape discretization tool for distributed hydrological modeling of cultivated areas. *Computer & Geosciences* **36** (8), 1021–1032.

Lebon, N., Dagès, C., Burger-Leenhardt, D. & Molénat, J. 2022 A new agro-hydrological catchment model to assess the cumulative impact of small reservoirs. *Environmental Modelling and Software* **153**, 105409.

Levavasseur, F., Bailly, J.-S., Lagacherie, P., Colin, F. & Rabotin, M. 2012 Simulating the effects of spatial configurations of agricultural ditch drainage networks on surface runoff from agricultural catchments. *Hydrological Processes* **26** (22), 3393–3404.

Ludwig, B., Boiffin, J., Chadeuf, J. & Auzet, A. V. 1995 Hydrological structure and erosion damage caused by concentrated flow in cultivated catchments. *Catena* **25** (1), 227–252.

McDowell, R. W., Moreau, P., Salmon-Monviola, J., Durand, P., Leterme, P. & Merot, P. 2014 Contrasting the spatial management of nitrogen and phosphorus for improved water quality: modelling studies in New Zealand and France European. *Journal of Agronomy* **57**, 52–61.

Pilz, T., Francke, T. & Bronstert, A. 2017 Lumpr 2.0.0: an R package facilitating landscape discretisation for hillslope-based hydrological models. *Geoscientific Model Development* **10**, 3001–3023.

Preti, F., Guastini, E., Penna, D., Dani, A., Cassiani, G., Boaga, J., Deiana, R., Romano, N., Nasta, P., Palladino, M., Errico, A., Giambastiani, Y., Trucchi, P. & Tarolli, P. 2018 Conceptualization of water flow pathways in agricultural terraced landscapes. *Land Degradation & Development* **29**, 651–662.

Schultz, R. C., Collettil, J. P., Isenhart, T. M., Simpkins, W. W., Mize, C. W. & Thompson, M. L. 1995 Design and placement of a multi-species riparian buffer strip system. *Agroforestry Systems* **29** (3), 201–226.

Van Den Belt, M. 2004 *Mediated Modeling: A System Dynamics Approach to Environmental Consensus Building*. Island Press, Washington, Covelo, London, p. 296.

Viaud, V., Durand, P., Merot, P., Sauboua, E. & Saadi, Z. 2005 Modeling the impact of the spatial structure of a hedge network on the hydrology of a small catchment in a temperate climate. *Agricultural Water Management* **74** (2), 135–163.

Yu, C., Duan, P., Yu, Z. & Gao, B. 2019 Experimental and model investigations of vegetative filter strips for contaminant removal: a review. *Ecological Engineering* **126**, 25–36.

First received 2 April 2022; accepted in revised form 24 June 2022. Available online 11 July 2022