Water Research 236 (2023) 119958

Contents lists available at ScienceDirect

Water Research

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

Practice makes the model: A critical review of stormwater green infrastructure modelling practice

Vincent Pons [a,b,*], Elhadi Mohsen Hassan Abdalla [a], Franz Tscheikner-Gratl [a], Knut Alfredsen [a], Edvard Sivertsen [c], Jean-Luc Bertrand-Krajewski [b], Tone Merete Muthanna [a]

[a] *Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim 7031 Norway*
[b] *Univ Lyon, INSA Lyon, DEEP, EA7429, 11 rue de la Physique, F-69621, Villeurbanne cedex, France*
[c] *SINTEF AS, S.P. Andersens veg 3 7465 Trondheim, Norway*

ABSTRACT

Green infrastructures (GIs) have in recent decades emerged as sustainable technologies for urban stormwater management, and numerous studies have been conducted to develop and improve hydrological models for GIs. This review aims to assess current practice in GI hydrological modelling, encompassing the selection of model structure, equations, model parametrization and testing, uncertainty analysis, sensitivity analysis, the selection of objective functions for model calibration, and the interpretation of modelling results. During a quantitative and qualitative analysis, based on a paper analysis methodology applied across a sample of 270 published studies, we found that the authors of GI modelling studies generally fail to justify their modelling choices and their alignments between modelling objectives and methods. Some practices, such as uncertainty analysis, were also found to be limited, despite their necessity being widely acknowledged by the scientific community and their application in other fields. In order to improve current GI modelling practice, the authors suggest the following: i) a framework, called STAMP, designed to promote the standardisation of the documentation of GI modelling studies, and ii) improvements in modelling tools for facilitating good practices, iii) the sharing of data for better model testing, iv) the evaluation of the suitability of hydrological equations for GI application, v) the publication of clear statements regarding model limitations and negative results.

1. Introduction

During recent decades, green infrastructures (GIs), such as green roofs and bio-retention cells, have emerged as sustainable alternatives for the improvement of urban stormwater management. They attempt to mimic the natural water cycle in urbanized catchments by increasing both vegetation cover and surface permeability, thus enhancing catchment evapotranspiration (ET), interception and infiltration. GIs thus serve to delay, attenuate and reduce stormwater runoff (Hamouz et al., 2018; Johannessen et al., 2018; Stovin, 2010; Støvring et al., 2018), and can also be used to improve runoff quality (Huang et al., 2016; Hunt et al., 2006), enhance a city's visual amenity (Jungels et al., 2013), promote citizen health and wellbeing (Tzoulas et al., 2007), as well as improve the urban micro-climate and bioclimate (Back et al., 2021), and enhance biodiversity (Wooster et al., 2022).

Quantification of the hydrological performance of a GI based on its physical properties and climate conditions is crucial. An appropriate modelling tool is required to aid the design and selection of GI geometries and configurations, and the materials needed to achieve desirable hydrological performance while supporting future technology development. Numerous studies have been conducted in order to develop and improve existing modelling tools with the aim of supporting these tasks and acquiring a deeper knowledge of the processes underlying GI (Li and Babcock, 2016; Soulis et al., 2017; Vesuviano et al., 2013).

In response to numerous published studies in this field, a number of review articles have appeared attempting to summarise the current status of GI modelling. The first review of note was completed by Elliott & Trowsdale (2007), in which the authors compared ten existing modelling tools used to simulate the quantity and quality of GI runoff. Kaykhosravi et al. (2018) have since expanded this approach by comparing a further eleven tools developed after the publication of Elliott & Trowsdale (2007). Both papers adhered to a similar methodology, which made it easy to compare their findings and at the same time apply their conclusions as an aid in the selection of fit-for-purpose modelling tools. Lerer et al. (2015) have reviewed a number of tools that can be used to assist in GI design and planning. They categorized these into several groups based on their modelling domains and the types of issues they addressed. Similarly, Kuller et al. (2017) reviewed

---

* Corresponding author at: Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, S.P. Andersens veg 5, Trondheim 7031 Norway.
E-mail address: vincent.pons@ntnu.no (V. Pons).

https://doi.org/10.1016/j.watres.2023.119958
Received 12 October 2022; Received in revised form 7 April 2023; Accepted 8 April 2023

tools, including models, in support of a best practice for GI planning, while Voskamp et al. (2021) compared several decision-making tools for GI planning and implementation by identifying the capabilities and challenges they offered to end-users. Other reviews have presented a more specific focus. For example, Li & Babcock (2013) focused on the hydrological modelling of green roofs, while Jayasooriya & Ng (2014) presented an economic analysis, comparing various financial issues linked to ten selected GI modelling tools, but with a de-emphasis on the hydrological modelling aspect.

The aforementioned reviews have all focused on the advantages and drawbacks of tools such as the Storm Water Management Model (SWMM) (Rossman, 2015) and Hydrus software (Šimůnek et al., 2016). None, however, has specifically assessed how the practice of hydrological modelling has been dealt with in the GI literature. A modelling practice defines the steps required to achieve the aims of a given modelling task. Such steps include the selection of a model structure and relevant equations, model parametrization and testing, sensitivity and uncertainty analyses, the selection of an objective function for model calibration and testing, as well as interpretation of the results (Jakeman et al., 2006). A 'good modelling practice' serves to promote a better understanding of the processes being modelled, leading to an improvement in model accuracy (Crout et al., 2008) and result reproducibility. This in turn enhances the quality of model-based decision-making, because high-quality models and modelling lead to better informed decisions (Rokstad and Ugarelli, 2015).

Guidelines for good modelling practice have been published in several scientific fields in which numerical models serve as essential tools. For example, Jakeman et al. (2006) identified ten iterative steps that ensure good practice in environmental modelling. Crout et al. (2008) published guidelines for environmental modelling with an emphasis on model development, evaluation and application. In the field of hydrological modelling, specific guidelines have been published as aids to model semantics (Beven and Young, 2013), parameterization (Malone et al., 2015) and evaluation (Moriasi et al., 2007).

However, in spite of an abundance of guidelines, 'good modelling practice' is frequently not adhered to. Schmolke et al. (2010) reported having found modelling practice in ecological studies to be unsatisfactory. Similarly, Pappenberger & Beven (2006) proposed several reasons why proper uncertainty analysis is not common practice in hydrological modelling. In acknowledgement of these studies, the present work aims to review current practice in GI hydrological modelling and to suggest improvements. The review comprises two main tasks:

1. A qualitative and quantitative analysis of the practice of GI modelling and its evolution.
2. A proposal of a best practice framework for GI modelling based on findings from current practices.

This paper deals with the hydrological aspects of stormwater GI modelling at 'building to block' scales. A number of other aspects are also of interest in this field, but are regarded as beyond the scope of this paper. These include the planning of stormwater GIs at entire city or catchment scale, water quality processes (pollutants transfer and removal), economic and financial issues, integrated urban water management, and scenarios that couple grey and green infrastructures.

2. Methodology

2.1. Selection of papers

Table 1
Keywords used to search for the relevant articles.

Keyword level 1	AND Keyword level 2	AND Keyword level 3
"SUDS" OR "LID" OR "BMPs" OR "WSUD" OR "GI" OR "IUWM" OR "LIUDD" OR "best management practices" OR "green infrastructure" OR "low impact development" OR "low impact urban design and development" OR "sustainable urban drainage systems" OR "urban stormwater management" OR "water sensitive urban design" OR "Green Roof" OR "Rain garden" OR "Bioretention" OR "permeable pavement“ OR “Bio-swales"	"hydrolog*" OR "Detention" OR "Retention" OR "Runoff“ OR “Hydraulic” OR “Infiltration” OR “Exfiltration” OR “snow” OR “Evapotranspiration” OR “Evaporation”	“Model*” OR “Simulat*” OR “Equation*”

Our review was initiated by first conducting a systematic search for relevant articles. The first step was to generate a list of relevant keywords, as shown in Table 1. Keywords were grouped into one of three levels, where Level 1 contains the different terminologies often used to describe GI, as summarised in the review of Fletcher et al. (2015). Level 2 represents the various hydrological processes within a GI, while Level 3 was included to ensure that the paper encompassed terms related to modelling tasks. The search employed the tools Web of Science (www.webofknowledge.com) and Scopus (www.scopus.com) and produced a total of 945 relevant articles after title screening. Our sample consists of papers that were published before April 2020, which was the start date of the review process. We excluded studies dealing with GI modelling at catchment scale, focusing exclusively on those that were developed based on and/or used a model on single GI units, such as individual buildings. Of the initial 945 papers, 270 were subsequently selected for review once the titles had been read and the abstracts screened. Figure A1 illustrates the steps followed to identify the relevant papers used in the review.

We note here that other terminologies associated with GI, such as 'Sponge City' and 'Blue-Green Infrastructure' have emerged since publication of the review by Fletcher et al. (2015). Such terms were not included as part of the keyword search performed for this review. However, most papers that use these new terminologies also use the keywords selected for our search. Thus, since full exhaustiveness was not our objective (and is in any case almost impossible to achieve due to the extremely high and fast-growing number of papers in this field), we decided that the number of papers selected for review, and the time period chosen, were sufficient to represent the scope of current publications dealing with dissemination of the recent evolution of GI modelling tasks and modelling practice. Dissemination in this context refers to communication of the details of modelling tasks by means of peer-reviewed publications.

2.2. Bias and limitations

We selected relevant papers by employing an approach similar to that described by Moher et al. (2009). This involved a keyword search in two databases, duplicate removal, title screening, abstract screening, and finally, full-text screening. Based on the papers ultimately selected, examples of good practice were highlighted and discussed.

This review aims to be systematic and in so doing restrict bias (Moher et al., 2009). We managed bias in two ways. The first was by applying a systematic reading approach to each paper, and the second by making the reader aware of existing biases.

2.2.1. Selection of papers and selection bias

The papers selected consist of 270 articles published up until April 2020. This in itself constitutes a possible limitation because numerous papers have been published in the period between April 2020 and mid-2022, when the present article was first drafted. However, the reading and detailed analysis of the 270 papers was a lengthy task. We believe that by further increasing the number of papers to include those published after April 2020, this may have introduced an additional bias due to the fact that the authors would not have been able to maintain their reading analysis of the original 270 papers. Of course, reading analysis evolves with time, depending on experience and results linked to the initial phases of the process. One of the aims of this review is to evaluate if and how modelling practice has evolved in recent years. Trends are identified by assessing statistics across yearly subsamples. We believe that extending the period from which the papers were selected would not have significantly changed the results, but only extrapolated them into more recent times.

2.2.2. Paper analysis bias

We reviewed each of the 270 papers in our sample by employing a 'paper analysis methodology', which is set out in Section 3. The methodology consisted in separating the different aspects of modelling in several categories. In each categories items were defined to describe the existing practice. The methodology involved systematically assigning items to different categories as a means of assessing the modelling practices described in each of the papers. Details of these items and categories are given in Section 3. Items were defined iteratively for each category. A first draft was tested on a test subset and the results and limitations analysed by all co-authors as a means of revising the items.

In order to maintain analytical consistency, we conducted the review category by category. On the basis of expertise, one of the co-authors was assigned the task of analysing all the papers in the sample against each defined category. Time constraints made it impossible for this work to be replicated by another co-author. Two actions were implemented with the aim of limiting biases inherent in individual readings. The first was to conduct an automatic word search in each of the papers in order to achieve a prior categorization. Expert knowledge was then employed to analyse each paper and decide on a posterior categorization. In situations where an aspect of modelling practice in a given paper was found to lie between two items of the category considered, a discussion was held amongst the co-authors to decide the matter. Table ***A1*** in the Supplementary Materials provides some examples of where several items could have been chosen for as many as three selected studies.

For each categories, we labelled some papers as 'not relevant' when the context of their study did not make the evaluation of practice relevant in the category considered. These papers are indicated by the suffix 'NR' in Table 2. For the categories related to the choice of hydrological processes and equations, only a few papers were labelled as 'NR'. One of these papers dealt primarily with experimentation, but was retained during the selection process because it referred to modelling practices linked to empirical models. In the case of objective functions and objective function values, some papers were labelled as not relevant if they were based on a previously calibrated model. Some studies did consider neither model calibration nor model testing, and these were also labelled 'NR'. The proportions of papers labelled as 'NR' in the matters of model testing and of objective function selection differ: . some studies were labelled 'NR' for objective function selection while presenting a qualitative testing of the model. In the case of parameter selection, 'NR' was assigned to papers identified as lying in a 'grey zone' during the selection of papers. The terme 'gey zone' refers to papers that were only considered relevant for a few categories of practice. We labelled some papers 'NR' for parameters selection on a case-by-case basis if they were based on non-parametric models, or potential evapotranspiration models relying mostly on measurement.

2.2.3. Data availability and reproducibility

The paper analysis methodology adopted in this review is presented in detail Section 3, and examples of items assignment for three selected papers are presented in the Supplementary Materials (Table A3). A list of the selected 270 papers, together with the items assigned for each category, is also available in the Supplementary Materials. However, in terms of reproducibility, the items assigned should be considered with a number of caveats. As noted in the previous section, some papers fall within the 'grey zone'. Their evaluation may contain biases and should thus be regarded as at least partly subjective. Item assignment was performed with the aim of identifying general trends of practice in the field and not the individual assessment of each papers. Therefore, the assessment of a paper taken out of this context should be considered carefully. Since the analysis was conducted category by category, some residual inconsistencies may exist between categories. Some papers may also describe several practices related to each category, but our approach was such that only one item could be assigned per paper and category. In the main body of the text, we have elected to focus on highlighting good examples of modelling practices rather than sharing examples of papers describing limited practices.

3. Paper analysis methodology

The articles studied in this review were assessed on the basis of ten different categories allocated to five sections based on published guidelines for good modelling practice (Crout et al., 2008; Jakeman et al., 2006). These in turn were adapted to the context of GI modelling and the investigation of alignments between objectives and methods (Table 2). The five sections are as follows: i) general study frame, ii) model assumptions and selection, iii) the use of objective functions, iv) uncertainty and sensitivity and v) parameter selection and model testing. Our assessment criteria are applied to an evaluation of how each of the reviewed studies addressed the ten categories that employ categorical or ordinal variables, as shown in Table 2.

It should be noted that the steps involved in 'good modelling practice', as shown in Table 2, are interconnected and as such should be assessed by means of an iterative process (Jakeman et al., 2006). However, the order in which we present the categories here is not arbitrary. The section 'uncertainty and sensitivity' was deliberately not placed at the end of the list in order to highlight that these aspects are indissociable from modelling tasks, i.e., not separate tasks. Items in the 'parameter selection and model testing' section are directly linked to monitored data, and were thus placed at the end of the list. This is because monitored data is normally considered as a source on which models are developed. Moreover, monitored data must also be used at a later stage for model testing and uncertainty propagation.

3.1. General study framework: how did the authors of the selected papers frame their work and communicate research objectives and limitations?

Good modelling practice requires that the objectives and limitations of a modelling study are clearly stated. In this review, we assessed the quality of stated objectives at three different levels (Table 2a). A clear formulation of objectives serves to aid other researchers in selecting relevant studies for their own research, and this is very important given the increased number of publications currently emerging in this field (Jakeman et al., 2006).

Limitations are defined in Table 2b, and vary from none, to those that are related to methods (model assumptions), those related to the study framework (direction for further research), and those related to both practices. The way in which a study's limitations are formulated defines the boundaries of a study and how well it is embedded in the field of GI. Limitations can help to clarify additional knowledge and the point from which further research can take place. They should consist of clearly formulated assumptions and should make reference to any processes neglected in the modelling methodology (Crout et al., 2008; Jakeman et al., 2006). The use of an existing tool or software also involves some assumptions and limitations, and these should also be clarified as an aid to direct future research.

Table 2

Definitions of the items identified for each section and category of practice. The items in underlined characters ("ObjUncl" and items ending with "None") represent the lower boundary of observed practices. Specifically, this means the absence of mention, explanation, or clarity depending on the category. The items in light grey (ending with "NR") were assigned to papers considered not relevant for the category of practice considered in the assessment.

i) General study frame

a) Are the objectives of the study clear?		**b) Are the modelling task limitations stated?**	
ObjUncl	Some objectives are defined but without clearly stated research questions.	**LimNone**	The limitations of the study and the methodology are not clearly stated
ObjMent	The study presents both a research question and research objectives.	**LimFurt**	Limitations or suggestions for further work are stated (related to objectives)
ObjClear	The study presents a research question, objective for the research and detail how they intend to achieve the objectives.	**LimMeth**	Limitations or suggestions to improve methodology are stated
		LimBoth	Both limitations and suggestions to improve the methodology and for further work

ii) Model assumptions and selection

c) Are the choices of the selected hydrological processes justified?		**d) Are the choices of the selected/major equations justified?**	
HyNone	The hydrological processes are not mentioned	**EqNone**	None of the equations are mentioned
HyMent	The hydrological processes selected are mentioned	**EqPart**	Some of the selected equations are mentioned
HyJust	The selection of hydrological processes is justified (e.g., from literature or conceptualisation of the system)*	**EqMent**	All the selected equations are mentioned
		EqJust	The choice of the equations is justified (e.g., from literature or equation properties)*
HyNR	It is not relevant to mention the hydrological processes	**EqNR**	It is not relevant to mention the selected equations

iii) Use of objective functions

e) How does the objective function choice align with model and study objective?		**f) How are the objective functions used for model evaluation?**	
OFNone	No objective functions used or mentioned	**OFVNone**	No threshold or no explanation
OFPop	Objective functions used because of their wide use and not because of their properties (i.e. not justified)	**OFVPop**	Threshold selected because of its wide use in the literature (i.e., not justified)
OFRed	Redundant objective functions set used without justification	**OFVLit**	Justified from literature or through author argumentation
OFJust	The selection of objective function is justified*	**OFVObj**	Justified based on the objectives of the study*
OFNR	It is not relevant to select objective functions	**OFVNR**	It is not relevant to use objective functions

iv) Uncertainty and Sensitivity analysis

g&h) ** **How is Uncertainty/Sensitivity analysis included in the study?**	
UncNone/SenNone	The concept of uncertainty/sensitivity is not mentioned
UncMent/SenMent	The concept of uncertainty/sensitivity is mentioned in the results
UncUse/SenUse	A method of uncertainty/sensitivity analysis is used but not specified
UncSpec/SenSpec	A method of uncertainty/sensitivity analysis is specified but not justified
UncJust/SenJust	The method (or absence of method) is justified in the context of the study (hypothesis)*
UncDed/SenDed	The objective of the paper directly involves applying uncertainty/sensitivity analysis

v) Parameters selection and model testing

i) How are the model parameter selected?		**j) How is the model tested?**	
ParNone	The parameters selection process is not mentioned or not aligned with the objectives	**TesNone**	The model is not tested, and it is not justified why it is not tested
ParMent	The parameters are selected from literature or measured	**TesLim**	The data for testing are limited (i.e., it is not explained why the dataset is sufficient) and it is not justified
ParLitJust	The parameters are selected from literature or measured; it is explained why their choice is relevant to the study	**TesLJust**	The data for testing are limited, but the limitations are stated, or the dataset is presented
ParMan	A manual calibration is used to select parameters	**TesSJust**	The dataset is sufficient (i.e., it is justified why the data are sufficient through the convergence of performance indicators), and the dataset is presented*

(*continued on next page*)

Table 2 (*continued*)

v) Parameters selection and model testing			
ParJust	The parameters are selected through automatic calibration, or their choice is justified by the authors [*]		
ParNR	Selecting parameters is not relevant in the study	**TesNR**	It is not relevant to test the model.

[*] The term “justification” or “justified” refers to justification how a choice is aligned with the study objectives.

[**] Uncertainty and Sensitivity analysis are two separate categories (g and h) that are merged here because their items are very similar.

3.2. Model assumptions and selection: how did the authors of the selected papers justify the hydrological processes described and their selection of equation(s)?

In good modelling practice, researchers develop a conceptualization of the system they are modelling and demonstrate how inputs to the modelled system are linked to its outputs (Jakeman et al., 2006). In the context of GI modelling, this will include the selection of the hydrological processes that will be included in, or excluded from, the model and a description of how these processes are interlinked (Table 2c). Following conceptualization, choices concerning formulations and assumptions related to the model should be clearly stated and justified (Crout et al., 2008). These choices will include the selection of model equations to simulate the hydrological processes, taking into consideration the theoretical limitations of these equations, as well as their data requirements and suitability for the study objectives (Table 2d). This step is necessary even when using existing software. Suitability requirements must always be checked and the chosen options justified.

3.3. Use of objective functions: how did the authors of the selected papers evaluate the models they used?

In order to ensure good modelling practice, researchers generally aim to employ an unbiased and transparent framework to evaluate their modelling results. Firstly, this framework should involve the use of objective functions (Bennett et al., 2013; Kouchi et al., 2017; Krause et al., 2005) that are aligned to the modelling task (Table 2e). In this regard, our focus was directed at evaluating how researchers in the field of GI modelling justified their selection of objective functions. Indeed, as highlighted by Bennett et al. (2013), objective functions may have different properties that make them suitable for the assessment of different parameter sets, depending on specific modelling objectives. For example, the Nash-Sutcliffe Efficiency (NSE) is best suited to estimations of peak outflow from a GI, but is less suitable for assessing retention performance. Secondly, an evaluation framework should define how the values of objective functions should be interpreted, not least in terms of the level of accuracy required for the modelling task (Table 2f).

3.4. Sensitivity and uncertainty analyses: how do the authors integrate these key concepts into their studies?

3.4.1. Definition and concepts

The increasing use of green infrastructure and its implications for urban water management make it essential to assess uncertainties linked to models, parameters and performance. An evaluation of uncertainty serves to inform both design and decision-making processes. Uncertainty is inherent to all aspects of modelling and can be defined along three dimensions (Tscheikner-Gratl et al., 2017; Walker et al., 2003):

i) the *source* of uncertainty (model input, calibration and model structure)
ii) the *type* of uncertainty (along a continuous spectrum from deep uncertainty to the theoretical horizon of determinism)
iii) the *nature* of uncertainty (differs according to whether it can be reduced with additional knowledge (epistemic), or not (random)).

Modelling uncertainty stems from inputs, assumptions, and simplifications in the model design, and will thus propagate to the outputs of the model (Deletic et al., 2012). Uncertainty analysis (UA) aims to quantify the level of confidence in a modelling result, and two classes of methods have been identified (Renard et al., 2011):

- *Forward* uncertainty quantification (propagating uncertainty from input to outputs)
- *Inverse* uncertainty quantification (parameter estimation and model calibration)

In a modelling context, a sensitivity analysis (SA) refers to how variations in inputs and model assumptions contribute to the variability of the output. In other words, which aspect of the total uncertainty can be allocated to given assumptions and inputs (Saltelli et al., 2007). In this sense, it differs from uncertainty quantification, which simply addresses the magnitude of uncertainty.

According to Saltelli et al. (2007), a sensitivity analysis can be allocated to one of two main categories:

- *Local (LSA)* or *global (GSA),* depending on whether the effect is studied at a local scale by means of derivatives, or for the entire input space;
- *One-At-a-Time (OAT)* or *All-At-a-Time (AAT)*, depending on whether variations in model parameters are simultaneous, or not.

They serve three primary purposes:

- *Screening*, i.e. qualification of inputs that are influential and those that are not.
- *Ranking*, i.e. quantification of the influence of the inputs considered.
- *Mapping*, i.e. identification of regions in the input space that produce a specific output.

3.4.2. Assessment of uncertainty and sensitivity analysis practice

This study has reviewed the common practices involved in uncertainty analysis (UA) and sensitivity analysis (SA) in the field of GI modelling. We checked each reviewed article to see if UA and SA were mentioned or applied, and if application was accompanied by transparent justification and alignment to the study objectives (Tables 2g,h). Due to their very nature, even the most advanced physically-based models are subject to uncertainties. Their use also requires an understanding of their sensitivity. A knowledge of both uncertainty and sensitivity is crucial to our understanding the modelling phenomenon, as well as for model improvement and as a basis for robust decision-making (Pappenberger and Beven, 2006). Uncertainty analysis methods are sometimes used as part of pre- or post-processing approaches by which uncertainty is either measured prior to a modelling task or propagated through a calibrated model, respectively. However, such methods may lead to a misunderstanding of the nature of the models. Deletic et al. (2012) suggest that by describing uncertainty methods as *intrinsic* to a modelling task we may improve practice. This means that the selection of a given model should be carried out *in parallel with* the selection of the of SA and UA methods in a way that is *in accordance with* the objective of the study and the resources, such as computational power, available.

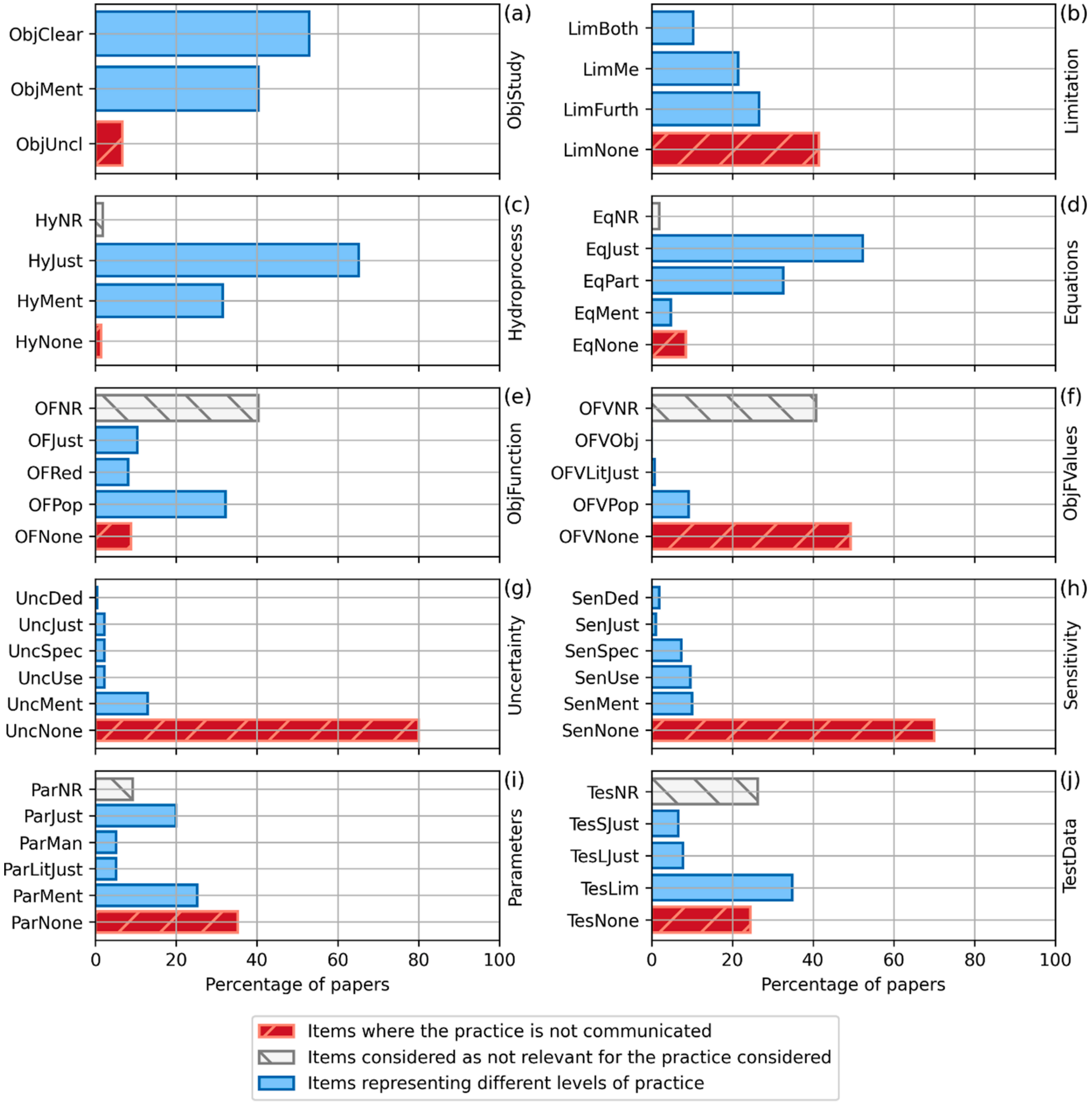

Fig. 1. Summary of practice for each items in each of the practice categories. The items and categories are defined in Table 2. The items coloured dark red are assigned to papers in which the lowest level of practice is identified. With the exception of 'ObjUncl', this applies to papers where the practice is considered to be absent (e.g. where the concept of 'uncertainty' is not mentioned). Items coloured light grey are assigned to papers that were regarded as not relevant to the practice category.

3.5. Parameter selection and model testing: how did the authors select their parameters and test their model?

This review has also evaluated the practice of selecting model parameters and testing datasets (Table 2h,i). In environmental study contexts, models require relevant parameters and inputs that are suited to their objectives. For this to be achieved, a framework is required as an aid to parameter selection. Parameters may be chosen from the literature or may be based on expert knowledge. They can be set by manual calibration, or assigned by automatic calibration. Furthermore, an evaluation of whether a calibrated model is suited to its study objectives will be influenced by the selection of its testing dataset (Beven and Young, 2013; Sargent, 2004; Shen et al., 2022). In this context, Silberstein (2006) emphasized the importance of collaboration between 'experimentalists' and 'modellers'. The difficulties inherent in environmental models (we refer here to 'model equifinality': (Beven, 2006)) mean that achieving a calibrated model suited to all conditions, and which encompasses all processes, is unlikely (Beven and Young, 2013). The selection of testing datasets must therefore be performed with care in order to be able to justify that a model is fit for purpose. The same applies to the selection of calibration datasets, which is outside the scope of this review. A large uncertainty should be assigned to conclusions drawn from models utilized outside their calibration range. Silberstein (2006) has illustrated this point using an example in which the selection of a testing dataset concealed a seasonal pattern, leading to a significant change in model performance. Thus, good practice demands an investigation not only of how parameters are selected, but also of the extent to which the dataset used for model testing is aligned with the objectives of model utilisation.

4. Current hydrological modelling practices as applied to GIs

4.1. Clarity of modelling studies and their limitations

The practice of communicating modelling objectives was standard in most of the reviewed papers. Only very few failed to clearly state the authors' objectives and research questions, as is shown in Fig. 1a. However, only half of the papers selected presented a clear and detailed description of their objectives and the means by which they intended to answer their research questions. We believe that the approach to communicating study objectives should be improved and perhaps standardized. For example, Stovin et al. (2013) first present the aim of their study, followed by an explanation of their modelling approach, before then providing further details in the subsequent sections of the manuscript. Whist, it could be debated which level of detail should be given in the introduction, it should be clear enough to make the information rapidly available for the reader. Peng et al. (2019) first present the aim of their paper, followed by an explanation of how this aim is addressed by means of a number of sub-objectives, which themselves serve effectively to communicate the structure of the paper. We regard this as an example of good practice.

Our study of limitations revealed an opposing trend. There is no systematic statement of the limitations of the studies (Fig. 1b), or more specifically of the modelling approach. This may be linked to the authors experiencing a limited incentive to publish negative results or a failure of their approach. Suggestions for further studies are presented most often in the conclusion. Some papers, such as Locatelli et al. (2015), offer a section in their paper, just prior to the conclusion, in which they describe the limitations of their approach in the light of their modelling assumptions. Others, such as Versini et al. (2015), discuss limitations very clearly in the discussion section. We regard this as good practice because by adopting this approach, the authors succeed in relating the limitations of their model directly to the context of the study and its results. Good practice in communicating limitations should not necessarily be restricted to the model assumptions, but should also be adopted when discussing data and methods. For example, Stovin et al. (2012) acknowledged the limitations of their data and the consequences these have on their results. They further acknowledged the limitations of event definition.

4.2. Selection of hydrological processes and model equations

Our findings show that almost all of the papers reviewed (>95%) clearly indicated all the hydrological processes that were selected for modelling, as shown in Fig. 1c. Moreover, about 64% of the studies included attempts to justify the selection of these processes, either by means of a conceptual description of the GI hydrological cycle (Yanling Li and Babcock, 2016; Stovin et al., 2013; Vesuviano et al., 2013), a description of a laboratory GI model (Carbone et al., 2014; Jahanfar et al., 2018; Martin et al., 2020; Yio et al., 2013) or a sensitivity analysis using a numerical model (Hakimdavar et al., 2014). We found that a few studies also stated which hydrological processes were excluded from their modelling, a feature that we consider to be good practice. For instance, Hakimdavar et al. (2014) justified why the parameter 'ET' was considered negligible in their event-based simulations. They had assessed the sensitivity of the parameter during single events using HYDRUS and found it to be insignificant.

We found that most of the studies (>80%) provided an explicit presentation of all the model equations selected. However, we identified a gap in practice in that it was only in only about half of the studies that the authors presented any justification for the selection of their equations. As we have noted previously, we consider that 'good modelling practice' requires equation selection based on suitability to the modelling objectives, and a clear justification of why the selected equations are fit for purpose. However, it appears that the selection of most GI equations is dictated not by suitability, but by the availability of modelling tools such as SWMM and Hydrus. For example, as is shown in the Supplementary Materials (Figure A2b), we fount that the infiltration equations most commonly selected were 'Green Ampt' (Green and Ampt, 1911) and 'Richards' (Richards, 1931), which simply reflects the popularity of the SWMM and Hydrus tools, respectively. Equation selection may also be the result of a hypothetical consensus regarding the use of established equations for which, in many cases, authors see no reason to justify. However, we identified no study that either indicated or justified such a consensus.

4.3. Objective functions and the interpretation of model results

As is demonstrated in the Supplementary Materials (Figure A2c), the parameters Nash Sutcliffe Efficiency (NSE), percentage bias (PBIAS) and determination coefficient (R^2) were the most popular objective functions applied in the reviewed studies, in spite of the fact that the latter has been shown to exhibit clear shortcomings (Kvalseth, 1985). The majority of studies applied more than one objective function to assess model accuracy. For example, Haowen et al. (2020) applied six objective functions for model calibration and validation. The Kling Gupta Efficiency (Gupta et al., 2009) which, according to (Knoben et al., 2019)), is one of the most common objective functions used in the hydrological modelling of catchments, was not used by the authors of any of the 270 papers reviewed for this study. Only a very few recently published studies in the field were found to use the KGE parameter (Abdalla et al., 2022; Iffland et al., 2021).

We have found indications that objective functions are commonly selected on the basis of popularity rather than their suitability to the modelling task. Some objective functions are known to be more suitable for certain tasks (low flows, high flows, etc.) than for others (Wöhling et al., 2013). However, there is a general lack amongst authors of GI modelling studies both to present a clear justification for the selection of objective functions and to discuss the effects of selection on model calibration and evaluation. We believe that this issue should be the subject of further investigation.

The values of objective functions are commonly interpreted to offer an evaluation of the accuracy of modelling results. We found that most of the reviewed papers used subjective terms such as ‘accurate’, ‘good’, ‘satisfactory’, ‘poor’, etc., without providing definitions of what these terms meant in terms of modelling results. Twenty-eight of the reviewed papers included definitions of threshold values taken from previous literature as a basis for describing modelling results as either ‘good’ or ‘poor’. For example, we found that many authors used a threshold value of NSE >0.5 to indicate ‘satisfactory’ modelling results based on an earlier study by Rosa et al. (2015), who justified their selection by citing the work of Dongquan et al. (2009). The latter authors in turn based their threshold value on a modelling protocol suggested by Engel et al. (2007). Another example is the study of Moriasi et al. (2007), which provided a protocol for hydrological modelling. This paper is commonly cited in GI modelling studies because it offers threshold values as a basis for describing modelling results as either ‘good’ or ‘unsatisfactory’. However, the limits provided by Moriasi et al. (2007) were based on the authors’ experience with catchment modelling, which may not be comparable with GI applications in terms of scale. Moreover, the authors suggested that the limits they stated were valid only for continuous, long-term simulations using monthly time steps, and strongly recommended that these values be adjusted in other situations.

4.4. Sensitivity and uncertainty

As is shown in Fig. 2, uncertainty and sensitivity analyses (UA and SA) are not applied systematically in GI modelling studies. In fact, these topics are not even mentioned in connection with GI modelling before 2008. More than half of the papers reviewed in our study do not mention uncertainty or sensitivity, and UA or SA methods are either named or used in less than 25% of the studies. This low percentage can be explained by the complexity and computational cost of such methods, perhaps combined with a lack of knowledge of their use amongst some authors despite the efforts made to communicate the importance of UA and SA and the establishment of frameworks to facilitate their dissemination (Deletic et al., 2012; Dotto et al., 2010; Pappenberger and Beven, 2006; Tscheikner-Gratl et al., 2019). Fig. 2 shows that the application of best practice in UA and SA methods, as defined in this study by the terms ‘SenJust’ and ‘UncJust’ in Table 2(h,i), has not increased in recent years. This is a major issue in terms of the development of the field. We found that it is rare to apply a method that propagates uncertainties to model outputs, and this places limits on the reliability of the published results. UA and SA methods may have been ignored by researchers in green infrastructure modelling because the most commonly used software applications (SWMM and Hydrus, used by 27% and 10% of the reviewed papers, respectively) do not include

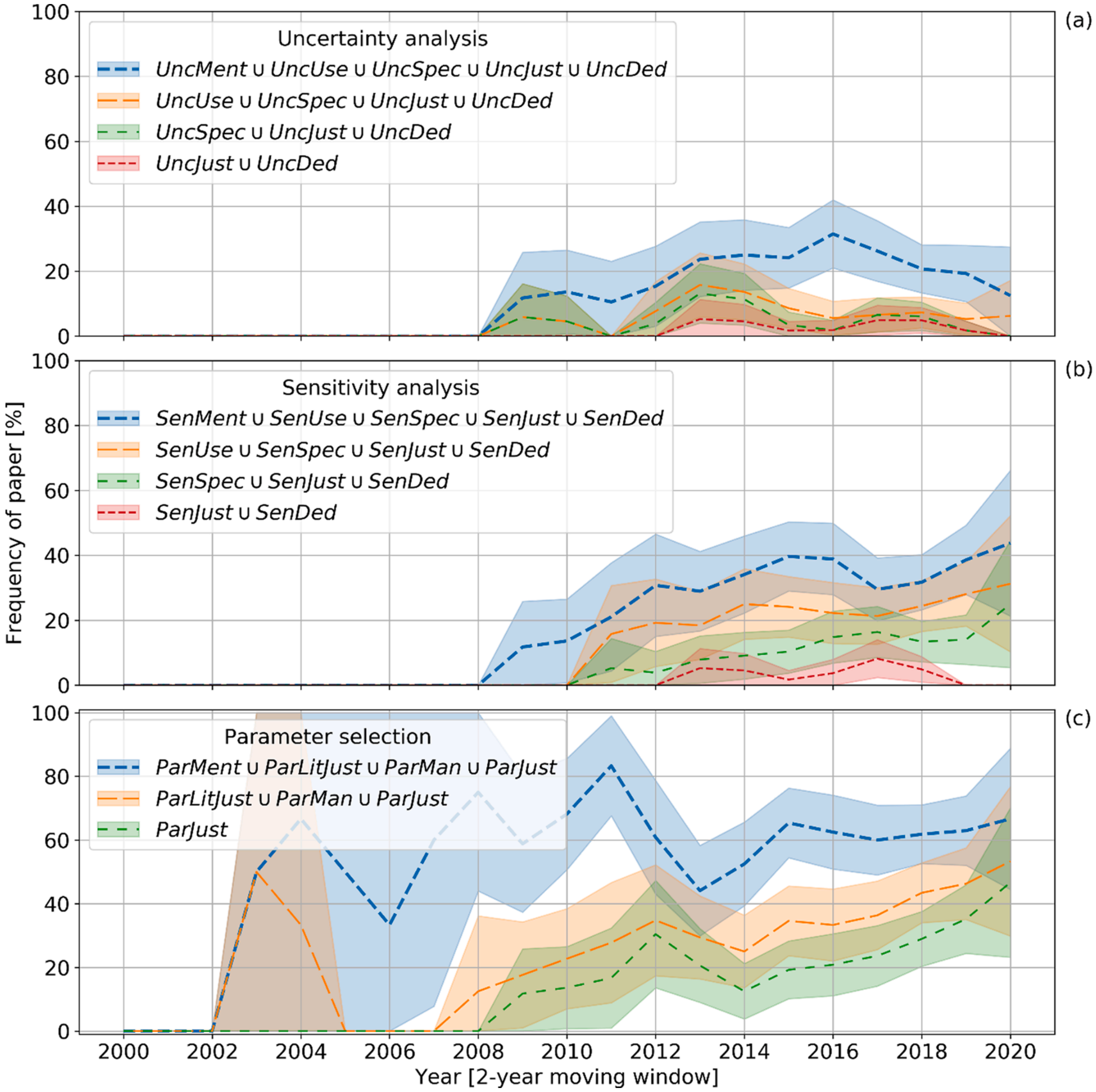

Fig. 2. Evolution of the practice of applying uncertainty and sensitivity analyses in green infrastructure modelling. The figure shows a 95% confidence interval with a 2-year moving window. This accounts for the size of the sample used to estimate usual practice. The categories used are defined in Table 2.

built-in functions for this purpose.

Some of the studies we reviewed in fact exercised ‘good modelling practice’ when it came to UA and SA methods. For example, Šimůnek Brunetti et al. (2018) compared different UA methods based on their performance and computational cost. They recommended the use of formal Bayesian methods such as Markov-Chain Monte-Carlo (MCMC) in preference to the Generalized Likelihood Uncertainty Estimation (GLUE) method applied by (Beven and Binley, 1992). GLUE was also used by authors such as (Feng et al., 2018; Krebs et al., 2016). Fuzzy set theory has recently been used by Lu & Qin (2019) to implement fuzzy parameters in SWMM after sensitivity analysis with the aim of permitting vagueness. In general, UA frameworks remain underused when compared with SA methods.

As mentioned in Section 1, GI placement studies at city or catchment scale are beyond the scope of the present review. However, we make a brief mention of these in the following. As is the case with Lu & Qin (2019), such studies commonly involve a high level of uncertainty (Walker et al., 2013), leading to the use of a different set of approaches, such as ensemble modelling and scenario-based uncertainty. Such methods are often used in climate modelling and climate change adaptation studies (Lee et al., 2021). For example, ensemble models are used to qualify a range of uncertainty when formal statistical methods cannot be applied, and scenario-based approaches may be applied to split a range of uncertainty into subregions in order to facilitate its management.

The Elementary Effect Test is an SA method, also known as the Morris method (Morris, 1991), which has been widely used by the authors of the papers reviewed in our study (Baek et al., 2020; Šimůnek Brunetti et al., 2018; García-Serrana et al., 2018; Jiake Li et al., 2020). It should be noted that although this method is sometimes cited, it is rarely justified or aligned with modelling objectives. Indeed, since the method suffers from the limitations of being a One-At-a-Time (OAT) approach, we agree with Saltelli et al. (2007) that it should only be used in restricted contexts, such as for screening purposes. Other methods such as e-FAST, Monte-Carlo filtering, subset simulation or PAWN (Brunetti et al., 2016; Šimůnek 2018; García-Serrana et al., 2018) may also be mentioned here. However, in most cases, even the authors of the studies reviewed here apply only a local perturbation (one of the LSA methods) to the parameters one-at-a-time. These methods can be applied without programming skills, since the number of simulations can be handled manually. However, the information gained from this practice is limited and can be misleading if not analysed carefully (Saltelli et al., 2007).

4.5. Parameter estimation and model testing

Almost half of the reviewed papers (approx. 48%) selected their modelling parameters not on the basis of calibration but on values taken from the literature or laboratory measurements. These include papers in which parameters were obtained through model calibration by the same authors in a previous study. For example, Palla & Gnecco (2020) used a calibrated hydrological model to quantify the hydrological impact of green roofs after having previously calibrated the same model as part of a formal study (Palla and Gnecco, 2015). As shown in Fig. 2, the practice of model calibration has increased in popularity in the field of GI modelling in recent years, and we believe that this is the result of the recent increase in availability of measured GI-related data.

Although the practice of calibration is gaining popularity, we have found that clear documentation on the calibration methods applied is lacking. Sixty per cent of the studies that performed model calibrations (82 papers) did not clearly state the algorithm or the tools used for the calibration. Moreover, global calibration algorithms are still rarely available GI modelling tools. Only 23% of the studies (33 papers) performed model calibration using automatic algorithms. As is demonstrated in Figure A2 in the Supplementary Materials, there is no consensus on the calibration algorithm used in GI modelling, and manual calibration, employing a trial-and-error method, was stated as being applied in as many as 23 papers.

In some of the more recently published papers, we found evaluations of the transferability of calibrated parameters between similar GIs located in different regions (Abdalla et al., 2022; Johannessen et al., 2019). The results of these studies suggested that even calibrated parameters could yield poor simulations if applied in climatic conditions that differed from those used to calibrate the model. However, in terms of the evaluation of climate change impacts, most studies applied parameters from the published literature or used calibrated models. In both cases (different region, and different period), the climatic conditions might differ significantly leading to a possible model transferability issue. In such cases, the use of ensemble modelling (Le Floch et al., 2022), multi-objective calibration (Fowler et al., 2016), and multi-data (Abdalla et al., 2022) may offer solutions to account for the uncertainty. We recommend the use of such approaches in future studies.

As Fig. 1j, current model testing practice (as defined in Table 2) relies on limited data volumes, and this has characterised 60% of the papers reviewed. In a limited number of studies, we have judged that their data volumes have been sufficient for model testing, as is the case for the multimodel comparison published by (Zhang and Guo, 2015). In general, however, we consider that continuous GI time-series data such as soil moisture, regulated discharge, overflow, infiltration, etc., have not been proven to belong enough to ensure sufficient data representativeness to achieve the modelling objectives. A ‘long enough’ dataset is defined as being of sufficient extent in time to permit convergence of the estimated quantity of interest. More specifically, there may be a difference between an estimator of the performance of a GI in a specific location based on available data, and that based on theoretically sufficient data. The difference between the ‘true’ performance, and an estimator based on limited data, has to be sufficiently small in accordance with the objectives of the study. In addition to proving convergence, good practice in situations where limited data are available must encompass; i) a presentation of the characteristics of the testing dataset that includes the number and characteristics of rainfall events, as is the case in Carson et al. (2013), who classify these events based on their depth, and ii) the consequences and uncertainty linked to a limited dataset, as is the case in Maniquiz et al. (2010), who state that due to limited data, calibration and validation were unable to provide a reliable parameterization of their model. An increase in levels of justification by authors of their testing datasets should help to increase the confidence in developed models. Likewise, the sharing of data may help to push back limitations linked to limited data volumes. Another approach to the management of limited data sets, especially in the case of rare and extreme precipitation events, consists of creating a database based on an experimental setup. Experiments can readily be designed to provide relevant data with the aim of aligning modelling objectives with calibration or testing (e.g., Hamouz et al., 2020; Vesuviano et al., 2013).

5. Recommendations for better practice

5.1. Standardized documentation of modelling studies

We found that a clear lack of effective communication by authors, in particular in relation to their justification of GI modelling practices, was found to be a common issue in the papers reviewed for this study. We assume that the existence of generic frameworks for model development (Crout et al., 2008; Jakeman et al., 2006) are not reaching all GI model developers and users. Indeed, while the paper analysis methodology that we apply in this article is itself generic and could be applied to other fields, we found that practice is less than satisfactory. Saltelli et al. (2007) succeeded in developing a generic framework for good practice in the field of sensitivity analysis (SA) by unifying practices across a number of different fields. However, we find that in the field of GI, SA is only applied to a limited extent. The issues of sensitivity analysis and uncertainty management have recently been introduced to the field of urban drainage (Deletic et al., 2012; Tscheikner-Gratl et al., 2019) but do not appear to have had any impact in the sub-field of GI modelling. It may therefore help to introduce a supplementary step involving implementing these aspects in practice by means of a practice tracking framework. The use of such frameworks in modelling studies could be subject to peer review on the submission of manuscripts for publication. Alternatively, a framework could be filled by journal editors in agreement with reviewers, or verified by a qualified reader as part of an open forum discussion. This may improve the practice in the field if requested by journal editors and reviewers. In this regard, we have attempted here to develop such a framework, called the 'STAMP' framework (STormwater infrastructure Alignment of Modelling Practices), which we believe offers a transparent and efficient means of sharing meta-information related to GI modelling tasks (**Error! Reference source not found.**). Our aim here is not to impose the use of any specific method, but to promote transparency in the communication of modelling tasks.

Method selection options and the directions of current research are manifold, so it may seem presumptuous of us to propose the introduction of a rigid framework for good practice, since such frameworks are likely to evolve in any event. However, we believe that it is important to make the effort to facilitate access to the content of papers in the field of GI modelling, and to promote transparency in the modelling process itself. It should be noted that the meta-information displayed in STAMP includes a revision of some of the items described in Sections 2 and 3 of this paper. This review methodology is aimed at mapping current practice, so it seemed appropriate to adjust some items of good practice in order to achieve better alignment with existing practice. The STAMP framework, on the other hand, aims at enhancing good practice. As such it is aligned with our present recommendation for good practices (Table A2) and not on current practices. The modifications introduced emphasize the distinction between 'mentioning' (by which authors mention an item without explaining it), 'analysing' (by which an item is presented and explained), and justification (by which an item is clearly aligned with the objectives of the study).

5.2. Improving modelling tools that facilitate good practice

Researchers in the field of GI hydrological modelling often confuse the terms 'model' and 'modelling tool'. In the case of the latter, they may be referring to a computer program or software. A model is a simplified representation of a physical world (Wheater et al., 2007), which may be in a mathematical or scaled-down physical form (Jajarmizadeh et al., 2012). A modelling tool is used to facilitate the modelling process and often includes several models and model choices. In the case of mathematical models, such tools may consist of software packages with a user interface, such as SWMM and Hydrus.

Our review indicates that the most commonly used modelling tools exert a major influence on many of the steps of GI modelling practice. For instance, the current version of SWMM, which was applied by 27% of the authors reviewed as part of this study, does not include built-in functions for automatic model calibration, sensitivity analysis or uncertainty analysis. We suggest that extending these open tools to include such functionality will represent a positive step towards promoting a good modelling practice by making the process more user-friendly for non-programmer users.

There have been several attempts made recently to develop software tools and programming packages that will enable SWMM to perform model calibration and other functions. Examples include the '*swmmr* library for R' (Leutnant et al., 2019) and the '*pyswmm* package for Python' (McDonnell et al., 2020). Despite some disparities in their capabilities and documentation, both of these libraries enable the user to modify parameters, execute modelling and extract the results of models developed in the SWMM tool. Such functionality includes modelling tasks such as calibration, sensitivity, and uncertainty analysis. The use of these libraries requires a certain level of programming skills, which may be challenging for some GI researchers and stormwater practitioners, such as engineers and urban planners. Alternatively, a small number of external software tools with user interfaces have been developed with the aim of calibrating models developed by SWMM. These include OSTRICH-SWMM (Shahed et al., 2020) and SWMM2PEST (Lin et al., 2019). The former requires the pre-processing of optimizing files using a Python package.

We believe that the development of user-friendly tools is an important step in the promotion of good modelling practice. Nevertheless, we also argue that it this may not be sufficient to increase the adoption of the steps involved in 'good modelling practice'. For example, the Green Infrastructure Flexible model (GIFMod/https://gifmod.com/) enables users to build GI models at various levels of complexity, and also includes built-in functions for automatic model calibration, sensitivity analysis, and uncertainty analysis (Massoudieh and Aflaki, 2017). To date, according to the Web of Science, GIFMod has only been applied in four published studies (Alikhani et al., 2020; Almadani et al., 2023; Chen and Chui, 2022; Yang and Chui, 2019). In comparison, the SWMM software has been applied in more than 700 published studies since the release of GIFMod. It should be emphasised that our intention in this paper is not to promote a specific tool, but to highlight the fact that alternative, high-capability software tools to SWMM do exist, even though they are still rarely applied.

5.3. Sharing data for improved model evaluation

Collecting GI-related hydrological data is a relatively new and growing field of research. As a result, there are only limited amounts of GI data available on which to apply good modelling practice. This review has also shown that the practice by authors of presenting datasets for model testing is limited. We believe that the sharing of datasets that have been applied and analysed will offer great opportunities for the improvement of GI models and modelling practice. The study by Versini et al. (2020), in which they provided hydrological measurements of a large-area green roof in Paris, represents one of the early attempts to share GI data. In the field of hydrological modelling at catchment scale, many large-scale measurement datasets have been published that have led to the development and verification of a number of modelling techniques. Examples include the CAMEL dataset applied in the US (Newman et al., 2015) and the CAMELS-GB in Great Britain (Coxon et al., 2020).Such practice should be integrated into tthe field of GI and GI modelling, taking the opportunity of the international drive toward data-sharing. Similarly, we encourage the sharing of both knowledge and tools as a means of standardising the management of monitoring data (Bertrand-Krajewski et al., 2021).

5.4. Evaluating the suitability of hydrological equations for GI application

A rational selection of model equations requires comparisons of the outputs of the equations with measurements made under different conditions. We require more studies that compare the choice of the modelling equations under varying climatic conditions because this will help researchers in their attempts to select the relevant equation for their specific cases. We found that many GI studies conducted investigations of the suitability of evapotranspiration (ET) equations for GI modelling in different climatic regions. Examples include (Marasco and Mcgillis, 2015; Hess et al., 2019; Lazzarin et al., 2005; Poë et al., 2015; Skala et al., 2020; Wadzuk et al., 2015), amongst others. Indeed, these papers provided valuable resources as a basis for the selection of suitable ET equations and were cited by the authors of many subsequent GI studies in the justification of selection of their ET equations. In contrast, we have found that there is a lack of studies that compare the suitability of equations used in other GI processes such as infiltration, runoff routing and snowmelt.

We found that very few studies were useful in terms of the selection of suitable infiltration models. A recent study by Parnas et al. (2021) has compared the performance of the three equations ‘Green-Ampt’, ‘Horton’ (Horton, 1939), and ‘Holtan’ (Holtan, 1961), which were designed to simulate infiltration in urban catchments. In addition, (Liu and Fassman-Beck, 2017,2018; Šimůnek 2018; Zhangjie Peng et al., 2019) have provided valuable insights into the estimation of soil properties such as the water retention curve (WRC), the relationship between volumetric water volume and the suction head, and unsaturated hydraulic conductivity, which influences the output of the Richards equation.

The modelling of GI-related runoff routing and snowmelt has received less attention in the literature. We identified only two studies (Hamouz and Muthanna, 2019b, 2019a) that modelled snow in connection with GI, using the conceptual equation for snowmelt available in SWMM. The authors reported unsatisfactory modelling results during snow periods, indicating the unsuitability of the snowmelt equation for their particular GI application.

5.5. *Statements of model limitations and encouragement of the publication of negative results*

We believe that the sharing of negative results and case studies in which models have performed poorly is very important for the progress of science. Beven & Young (2013) have previously stated that a model structure or parameter set is more than likely to fail if it is evaluated continuously under different conditions. However, we believe that the publication of such failures will lead to an improvement in model structures, data collection, and parameterization. Unfortunately, our experience is that there is a tendency amongst authors only to publish positive modelling results, and to omit limitations in the field of GI hydrological modelling – a practice similar to that observed in other scientific fields (Smaldino and McElreath, 2016). In this regard, the recently issued Journal of Trial and Error (https://journal.trialanderror.org/) may offer an incentive to researchers, and in doing so help to improve practice.

Only a few of the GI modelling studies reviewed in this study published negative results with discussions of the implications for current GI models. Peng & Stovin (2017) discussed the unsatisfactory performance of an uncalibrated SWMM-LID module in connection with the modelling of runoff from large-area green roofs. They also discussed issues related to evapotranspiration modelling in their SWMM-LID module. Randall et al. (2020) also evaluated the SWMM-LID module in the case of a single-event modelling of a permeable pavement. They pointed out some of the limitations of the module when it came to simulating events with multiple peaks.

6. Conclusion

The application of a paper analysis methodology to a large sample of 270 papers published in the field of GI modelling has revealed important limitations in the justification by authors of their modelling choices.

- Communication of the limitations of studies and the methods applied can be improved.
- Justification for the choice of modelling equations is lacking and seems to be driven by selection of the modelling tool.
- The choice of objective functions and their interpretation for model evaluation is directed primarily by their use in the literature rather than being based on an alignment between their properties and the modelling objectives.
- In spite of a few examples of good practice, the use of uncertainty and sensitivity analyses was found to be limited and not justified when applied.
- Despite improvements in recent years, authors continue to fail to justify the selection of model parameters and provide clear documentation of the methods they use. The representativeness of data used to test models is only rarely discussed, revealing a clear flaw in model testing practice.

In order to improve the alignment between modelling methods and objectives, the authors propose the adoption of a standardized modelling framework (named STAMP). This could help researchers and reviewers to document and justify their modelling steps. The authors also suggest that modelling tools should be enhanced in order to make the implementation of ‘good practice steps’ more user friendly. We also recommend better sharing of datasets, and more thorough analysis with the aim of enhancing model testing and transferability. Finally, we recommend that the publication of negative results and model limitations could lead to an improvement of current GI models, and that this should be actively pursued and encouraged.

Declaration of Competing Interest

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

Data availability

The data are available in supplementary material. Code can be shared upon request.

Acknowledgment

This study was supported by the *Klima2050* Centre for Research-based Innovation (SFI) and funded by the Research Council of Norway and its consortium partners under grant no. 237859/030. The research was performed in part within the framework of the OTHU (Field Observatory for Urban Water Management – http://othu.org, last access: 26 May 2022) and realized within the Graduate School H2O’Lyon (ANR-17-EURE-0018) and *Université de Lyon* (UdL) as part of the program ‘*Investissement d’Avenir*’ and involving the *Agence Nationale de la Recherche* (ANR).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2023.119958.

References

Abdalla, E.M.H., Alfredsen, K., Merete Muthanna, T., 2022. Towards improving the calibration practice of conceptual hydrological models of extensive green roofs. Journal of Hydrology 607, 127548. https://doi.org/10.1016/j.jhydrol.2022.127548.

Alikhani, J., Nietch, C., Jacobs, S., Shuster, B., Massoudieh, Arash, 2020. Modeling and Design Scenario Analysis of Long-Term Monitored Bioretention System for Rainfall-Runoff Reduction to Combined Sewer in Cincinnati, OH. Journal of Sustainable Water in the Built Environment 6 (2), 04019016. https://doi.org/10.1061/JSWBAY.0000903.

Almadani, M., Nietch, C., Massoudieh, Arash, 2023. Effectiveness of Design and Implementation Alternatives for Stormwater Control Measures Modeled at the Watershed Scale. Journal of Sustainable Water in the Built Environment 9 (1), 04022021. https://doi.org/10.1061/JSWBAY.SWENG-460.

Back, Y., Bach, P.M., Jasper-Tönnies, A., Rauch, W., Kleidorfer, M., 2021. A rapid fine-scale approach to modelling urban bioclimatic conditions. Science of The Total Environment 756, 143732. https://doi.org/10.1016/j.scitotenv.2020.143732.

Baek, S., Ligaray, M., Pachepsky, Y., Chun, J.A., Yoon, K.-S., Park, Y., Cho, K.H., 2020. Assessment of a green roof practice using the coupled SWMM and HYDRUS models. Journal of Environmental Management 261, 109920. https://doi.org/10.1016/j.jenvman.2019.109920.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environmental Modelling and Software 40, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011.

Bertrand-Krajewski, J.-L., Clemens-Meyer, F., Lepot, M., 2021. Metrology in Urban Drainage and Stormwater Management: Plug and Pray. IWA Publishing. https://doi.org/10.2166/9781789060119.

Beven, K., 2006. A manifesto for the equifinality thesis. Journal of Hydrology 320 (1), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007.

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes 6 (3), 279–298. https://doi.org/10.1002/hyp.3360060305.

Beven, K., Young, P., 2013. A guide to good practice in modeling semantics for authors and referees. Water Resources Research 49 (8), 5092–5098. https://doi.org/10.1002/wrcr.20393.

Brunetti, G., Šimůnek, J., Piro, P., 2016. A comprehensive numerical analysis of the hydraulic behavior of a permeable pavement. Journal of Hydrology 540, 1146–1161. https://doi.org/10.1016/j.jhydrol.2016.07.030.

Brunetti, G., Šimůnek, J., Turco, M., Piro, P., 2018. On the use of global sensitivity analysis for the numerical analysis of permeable pavements. Urban Water Journal 15 (3), 269–275. https://doi.org/10.1080/1573062X.2018.1439975.

Carbone, M., Garofalo, G., Nigro, G., Piro, P., 2014. A Conceptual Model for Predicting Hydraulic Behaviour of a Green Roof. Procedia Engineering 70, 266–274. https://doi.org/10.1016/j.proeng.2014.02.030.

Carson, T.B., Marasco, D.E., Culligan, P.J., McGillis, W.R., 2013. Hydrological performance of extensive green roofs in New York City: observations and multi-year modeling of three full-scale systems. Environmental Research Letters 8 (2), 024036. https://doi.org/10.1088/1748-9326/8/2/024036.

Chen, B., Chui, T.F.M., 2022. Optimal design of stepped bioretention cells for slopes. Journal of Hydrology 615, 128697. https://doi.org/10.1016/j.jhydrol.2022.128697.

Coxon, 2020. Catchment attributes and hydro-meteorological timeseries for 671 catchments across Great Britain (CAMELS-GB). NERC Environmental Information Data Centre. https://doi.org/10.5285/8344e4f3-d2ea-44f5-8afa-86d2987543a9.

Crout, N., Kokkonen, T., Jakeman, A.J., Norton, J.P., Newham, L.T.H., Anderson, R., Assaf, H., Croke, B.F.W., Gaber, N., Gibbons, J., Holzworth, D., Mysiak, J., Reichl, J., Seppelt, R., Wagener, T., Whitfield, P., 2008. Chapter Two Good Modelling Practice. In: Jakeman, A.J., Voinov, A.A., Rizzoli, A.E., Chen, S.H. (Eds.), Environmental Modelling, Software and Decision Support. Elsevier, pp. 15–31. https://doi.org/10.1016/S1574-101X(08)00602-9. Vol. 3.

Deletic, A., Dotto, C.B.S., McCarthy, D.T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T.D., Rauch, W., Bertrand-Krajewski, J.L., Tait, S., 2012. Assessing uncertainties in urban drainage models. Physics and Chemistry of the Earth, Parts A/B/C 42–44, 3–10. https://doi.org/10.1016/j.pce.2011.04.007.

Dongquan, Z., Jining, C., Haozheng, W., Qingyuan, T., Shangbing, C., Zheng, S., 2009. GIS-based urban rainfall-runoff modeling using an automatic catchment-discretization approach: a case study in Macau. Environmental Earth Sciences 59 (2), 465. https://doi.org/10.1007/s12665-009-0045-1.

Dotto, C., Kleidorfer, M., McCarthy, D., Deletic, A., Rauch, W., Fletcher, T., 2010. Towards global assessment of modelling errors. In: 6th International Conference on Sewer Processes and Networks (SPN). Gold Coast, Australia.

Elliott, A.H., Trowsdale, S.A., 2007. A review of models for low impact urban stormwater drainage. Environmental Modelling & Software 22 (3), 394–405. https://doi.org/10.1016/j.envsoft.2005.12.005.

Engel, B., Storm, D., White, M., Arnold, J., Arabi, M., 2007. A Hydrologic/Water Quality Model Applicati11. JAWRA Journal of the American Water Resources Association 43 (5), 1223–1236. https://doi.org/10.1111/j.1752-1688.2007.00105.x.

Feng, Y., Burian, S.J., Pardyjak, E.R., 2018. Observation and Estimation of Evapotranspiration from an Irrigated Green Roof in a Rain-Scarce Environment. Water (3), 10. https://doi.org/10.3390/w10030262.

Fletcher, T.D., Shuster, W., Hunt, W.F., Ashley, R., Butler, D., Arthur, S., Trowsdale, S., Barraud, S., Semadeni-Davies, A., Bertrand-Krajewski, J.-L., Mikkelsen, P.S., Rivard, G., Uhl, M., Dagenais, D., Viklander, M., 2015. SUDS, LID, BMPs, WSUD and more – The evolution and application of terminology surrounding urban drainage. Urban Water Journal 12 (7), 525–542. https://doi.org/10.1080/1573062X.2014.916314.

Fowler, K.J.A., Peel, M.C., Western, A.W., Zhang, L., Peterson, T.J., 2016. Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfallrunoff models. Water Resources Research 52 (3), 1820–1846. https://doi.org/10.1002/2015WR018068.

García-Serrana, M., Gulliver, J.S., Nieber, John L., 2018. Calculator to Estimate Annual Infiltration Performance of Roadside Swales. Journal of Hydrologic Engineering 23 (6), 04018017. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001650.

Heber Green, W., Ampt, G.A., 1911. Studies on Soil Phyics. The Journal of Agricultural Science 4 (1), 1–24. https://doi.org/10.1017/S0021859600001441.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377 (1), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hakimdavar, R., Culligan, P.J., Finazzi, M., Barontini, S., Ranzi, R., 2014. Scale dynamics of extensive green roofs: Quantifying the effect of drainage area and rainfall characteristics on observed and modeled green roof hydrologic performance. Ecological Engineering 73, 494–508. https://doi.org/10.1016/j.ecoleng.2014.09.080.

Hamouz, Vladimír, Lohne, J., Wood, J.R., Muthanna, T.M, 2018. Hydrological Performance of LECA-Based Roofs in Cold Climates. Water (3), 10. https://doi.org/10.3390/w10030263.

Hamouz, Vladimír, Muthanna, T.M., 2019a. Hydrological modelling of green and grey roofs in cold climate with the SWMM model. Journal of Environmental Management 249, 109350. https://doi.org/10.1016/j.jenvman.2019.109350.

Hamouz, Vladimír, Muthanna, T.M., 2019b. Hydrological modelling of green and grey roofs in cold climate with the SWMM model. Journal of Environmental Management 249, 109350. https://doi.org/10.1016/j.jenvman.2019.109350.

Hamouz, Vladimír, Pons, V., Sivertsen, E., Raspati, G.S., Bertrand-Krajewski, J.-L., Muthanna, T.M, 2020. Detention-based green roofs for stormwater management under extreme precipitation due to climate change. Blue-Green Systems 2 (1), 250–266. https://doi.org/10.2166/bgs.2020.101.

Haowen, X., Yawen, W., Luping, W., Weilin, L., Wenqi, Z., Hong, Z., Yichen, Y., Jun, L., 2019. Comparing simulations of green roof hydrological processes by SWMM and HYDRUS-1D. Water Supply 20 (1), 130–139. https://doi.org/10.2166/ws.2019.140.

Hess, A., Wadzuk, B., Welker, Andrea, 2019. Predictive Evapotranspiration Equations in Rain Gardens. Journal of Irrigation and Drainage Engineering 145 (7), 04019010. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001389.

Holtan, H. N., & others. (1961). *Concept for infiltration estimates in watershed engineering.*

Horton, R.E., 1939. Analysis of runoff-plat experiments with varying infiltration-capacity. Eos, Transactions American Geophysical Union 20 (4), 693–711. https://doi.org/10.1029/TR020i004p00693.

Huang, J., Valeo, C., He, J., Chu, Angus, 2016. Three Types of Permeable Pavements in Cold Climates: Hydraulic and Environmental Performance. Journal of Environmental Engineering 142 (6), 04016025. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001085.

Hunt, W.F., Jarrett, A.R., Smith, J.T., Sharkey, L.J., 2006. Evaluating Bioretention Hydrology and Nutrient Removal at Three Field Sites in North Carolina. Journal of Irrigation and Drainage Engineering 132 (6), 600–608. https://doi.org/10.1061/(ASCE)0733-9437(2006)132:6(600).

Iffland, R., Förster, K., Westerholt, D., Pesci, M.H., Lösken, G., 2021. Robust Vegetation Parameterization for Green Roofs in the EPA Stormwater Management Model (SWMM). Hydrology 8 (1). https://doi.org/10.3390/hydrology8010012.

Jahanfar, A., Drake, J., Sleep, B., Gharabaghi, B., 2018. A modified FAO evapotranspiration model for refined water budget analysis for Green Roof systems. Ecological Engineering 119, 45–53. https://doi.org/10.1016/j.ecoleng.2018.04.021.

Jajarmizadeh, M., Harun, S., Salarpour, M., 2012. A Review on Theoretical Consideration and Types of Models in Hydrology. Journal of Environmental Science and Technology 5, 249–261. https://doi.org/10.3923/jest.2012.249.261.

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software 21 (5), 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004.

Jayasooriya, V.M., Ng, A.W.M., 2014. Tools for Modeling of Stormwater Management and Economics of Green Infrastructure Practices: a Review. Water, Air, & Soil Pollution 225 (8), 2055. https://doi.org/10.1007/s11270-014-2055-1.

Johannessen, B.G., Hamouz, V., Gragne, A.S., Muthanna, T.M., 2019. The transferability of SWMM model parameters between green roofs with similar build-up. Journal of Hydrology 569, 816–828. https://doi.org/10.1016/j.jhydrol.2019.01.004.

Johannessen, B.G., Muthanna, T.M., Braskerud, B.C., 2018. Detention and Retention Behavior of Four Extensive Green Roofs in Three Nordic Climate Zones. Water (6), 10. https://doi.org/10.3390/w10060671.

Jungels, J., Rakow, D.A., Allred, S.B., Skelly, S.M., 2013. Attitudes and aesthetic reactions toward green roofs in the Northeastern United States. Landscape and Urban Planning 117, 13–21. https://doi.org/10.1016/j.landurbplan.2013.04.013.

Kaykhosravi, S., Khan, U.T., Jadidi, A., 2018. A Comprehensive Review of Low Impact Development Models for Research, Conceptual, Preliminary and Detailed Design Applications. Water (11), 10. https://doi.org/10.3390/w10111541.

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences 23 (10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019.

Kouchi, D.H., Esmaili, K., Faridhosseini, A., Sanaeinejad, S.H., Khalili, D., Abbaspour, K. C., 2017. Sensitivity of Calibrated Parameters and Water Resource Estimates on Different Objective Functions and Optimization Algorithms. Water (6), 9. https://doi.org/10.3390/w9060384.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences 5, 89–97. https://doi.org/10.5194/adgeo-5-89-2005.

Krebs, G., Kuoppamäki, K., Kokkonen, T., Koivusalo, H., 2016. Simulation of green roof test bed runoff. Hydrological Processes 30 (2), 250–262. https://doi.org/10.1002/hyp.10605.

Kuller, M., Bach, P.M., Ramirez-Lovering, D., Deletic, A., 2017. Framing water sensitive urban design as part of the urban form: A critical review of tools for best planning practice. Environmental Modelling & Software 96, 265–282. https://doi.org/10.1016/j.envsoft.2017.07.003.

Kvalseth, T.O., 1985. Cautionary Note about R2. The American Statistician 39 (4), 279–285. https://doi.org/10.2307/2683704.

Lazzarin, R.M., Castellotti, F., Busato, F., 2005. Experimental measurements and numerical modelling of a green roof. Energy and Buildings 37 (12), 1260–1267. https://doi.org/10.1016/j.enbuild.2005.02.001.

Le Floch, N., Pons, V., Abdalla, E.M.H., Alfredsen, K., 2022. Catchment scale effects of low impact development implementation scenarios at different urbanization densities. Journal of Hydrology 612, 128178. https://doi.org/10.1016/j.jhydrol.2022.128178.

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J.P., Engelbrecht, F., Fischer, E., Fyfe, J.C., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S., Zhou, T., 2021. Future Global Climate: Scenario-Based Projections and Near-Term Information [Book Section]. In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021: The Physical Science Basis.

Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 553–672. https://doi.org/10.1017/9781009157896.006.

Lerer, S.M., Arnbjerg-Nielsen, K., Mikkelsen, P.S., 2015. A Mapping of Tools for Informing Water Sensitive Urban Design Planning Decisions—Questions, Aspects and Context Sensitivity. Water 7 (3), 993–1012. https://doi.org/10.3390/w7030993.

Leutnant, D., Döring, A., Uhl, M., 2019. swmmr - an R package to interface SWMM. Urban Water Journal 16 (1), 68–76. https://doi.org/10.1080/1573062X.2019.1611889.

Li, J., Liu, F., Li, Y., 2020. Simulation and design optimization of rain gardens via DRAINMOD and response surface methodology. Journal of Hydrology 585, 124788. https://doi.org/10.1016/j.jhydrol.2020.124788.

Li, Y., Babcock, J., Roger, W., 2013. Green roof hydrologic performance and modeling: a review. Water Science and Technology 69 (4), 727–738. https://doi.org/10.2166/wst.2013.770.

Li, Y., Babcock, R.W., 2016. A Simplified Model for Modular Green Roof Hydrologic Analyses and Design. Water 8 (8). https://doi.org/10.3390/w8080343.

Lin, X., Simon, M., Niu, N., 2019. Releasing Scientific Software in GitHub: A Case Study on SWMM2PEST. 2019 IEEE/ACM 14th International Workshop on Software Engineering for Science (SE4Science) 47–50. https://doi.org/10.1109/SE4Science.2019.00014.

Liu, R., Fassman-Beck, Elizabeth, 2018. Pore Structure and Unsaturated Hydraulic Conductivity of Engineered Media for Living Roofs and Bioretention Based on Water Retention Data. Journal of Hydrologic Engineering 23 (3), 04017065. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001621.

Liu, R., Fassman-Beck, E., 2017. Hydrologic experiments and modeling of two laboratory bioretention systems under different boundary conditions. Frontiers of Environmental Science & Engineering 11 (4), 10. https://doi.org/10.1007/s11783-017-0951-5.

Locatelli, L., Mark, O., Mikkelsen, P.S., Arnbjerg-Nielsen, K., Wong, T., Binning, P.J., 2015. Determining the extent of groundwater interference on the performance of infiltration trenches. Journal of Hydrology 529, 1360–1372. https://doi.org/10.1016/j.jhydrol.2015.08.047.

Lu, W., Qin, X., 2019. An Integrated Fuzzy Simulation-Optimization Model for Supporting Low Impact Development Design under Uncertainty. Water Resources Management 33 (12), 4351–4365. https://doi.org/10.1007/s11269-019-02377-7.

R.J. Malone, W., Yagow, G., Baffaut, C.W., Gitau, M., Qi, Z., Amatya, M., Parajuli, D.B., Bonta, P.V., Green, R., 2015. Parameterization Guidelines and Considerations for Hydrologic Models Transactions of the ASABE 58 (6), 1681–1703 https://elibrary.asabe.org/abstract.asp?aid=46543&t=3.

Maniquiz, M.C., Lee, S.-Y., Kim, L.-H., 2010. Long-Term Monitoring of Infiltration Trench for Nonpoint Source Pollution Control. Water, Air, & Soil Pollution 212 (1), 13–26. https://doi.org/10.1007/s11270-009-0318-z.

Marasco, D.E., Culligan, P.J., McGillis, W.R., 2015. Evaluation of common evapotranspiration models based on measurements from two extensive green roofs in New York City. Ecological Engineering 84, 451–462. https://doi.org/10.1016/j.ecoleng.2015.09.001.

Martin III, W.D., Kaye, N.B., Mohammadi, S, 2020. A physics-based routing model for modular green roof systems. Proceedings of the Institution of Civil Engineers - Water Management 173 (3), 142–151. https://doi.org/10.1680/jwama.18.00094.

Massoudieh, A., Maghrebi, M., Kamrani, B., Nietch, C., Tryby, M., Aflaki, S., Panguluri, S., 2017. A flexible modeling framework for hydraulic and water quality performance assessment of stormwater green infrastructure. Environmental Modelling & Software 92, 57–73. https://doi.org/10.1016/j.envsoft.2017.02.013.

McDonnell, B.E., Ratliff, K., Tryby, M.E., Wu, J.J.X., Mullapudi, A., 2020. PySWMM: The Python Interface to Stormwater Management Model (SWMM). Journal of Open Source Software 5 (52), 2292. https://doi.org/10.21105/joss.02292.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339. https://doi.org/10.1136/bmj.b2535.

R. Moriasi, N., Arnold, D.G., Van Liew, J.W., Bingner, M.L., Harmel, R.D., Veith, L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations Transactions of the ASABE 50 (3), 885–900 https://elibrary.asabe.org/abstract.asp?aid=23153&t=3.

Morris, Max D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics 33 (2), 161–174. https://doi.org/10.1080/00401706.1991.10484804.

Newman, A.J., Clark, M.P., Sampson, K., Wood, A., Hay, L.E., Bock, A., Viger, R.J., Blodgett, D., Brekke, L., Arnold, J.R., Hopson, T., Duan, Q., 2015. Development of a large-sample watershedscale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. Hydrology and Earth System Sciences 19 (1), 209–223. https://doi.org/10.5194/hess-19-209-2015.

Palla, A., Gnecco, I., 2015. Hydrologic modeling of Low Impact Development systems at the urban catchment scale. Journal of Hydrology 528, 361–368. https://doi.org/10.1016/j.jhydrol.2015.06.050.

Palla, A., Gnecco, I., 2020. A continuous simulation approach to quantify the climate condition effect on the hydrologic performance of green roofs. Urban Water Journal 17 (7), 609–618. https://doi.org/10.1080/1573062X.2019.1700287.

Pappenberger, F., Beven, K.J., 2006. Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resources Research (5), 42. https://doi.org/10.1029/2005WR004820.

Parnas, F.E.Å., Abdalla, E.M.H., Muthanna, T.M., 2021. Evaluating three commonly used infiltration methods for permeable surfaces in urban areas using the SWMM and STORM. Hydrology Research 52 (1), 160–175. https://doi.org/10.2166/nh.2021.048.

Peng, Z., Smith, C., Stovin, V., 2019. Internal fluctuations in green roof substrate moisture content during storm events: Monitored data and model simulations. Journal of Hydrology 573, 872–884. https://doi.org/10.1016/j.jhydrol.2019.04.008.

Peng, Z., Stovin, Virginia, 2017. Independent Validation of the SWMM Green Roof Module. Journal of Hydrologic Engineering 22 (9), 04017037. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001558.

Poë, S., Stovin, V., Berretta, C., 2015. Parameters influencing the regeneration of a green roof's retention capacity via evapotranspiration. Journal of Hydrology 523, 356–367. https://doi.org/10.1016/j.jhydrol.2015.02.002.

Randall, M., Støvring, J., Henrichs, M., Jensen, M.B., 2020. Comparison of SWMM evaporation and discharge to in-field observations from lined permeable pavements. Urban Water Journal 17 (6), 491–502. https://doi.org/10.1080/1573062X.2020.1776737.

Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., Franks, S.W., 2011. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. Water Resources Research (11), 47. https://doi.org/10.1029/2011WR010643.

Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. Physics 1 (5), 318–333. https://doi.org/10.1063/1.1745010.

Rokstad, M.M., Ugarelli, R.M., 2015. Evaluating the role of deterioration models for condition assessment of sewers. Journal of Hydroinformatics 17 (5), 789–804. https://doi.org/10.2166/hydro.2015.122.

Rosa, D.J., Clausen, J.C., Dietz, M.E., 2015. Calibration and Verification of SWMM for Low Impact Development. JAWRA Journal of the American Water Resources Association 51 (3), 746–757. https://doi.org/10.1111/jawr.12272.

Rossman, L., 2015. Storm Water Management Model User's Manual Version 5.1. U.S. EPA Office of Research. http://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100N3J6.TXT.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2007. Global sensitivity analysis: the primer. John Wiley. https://doi.org/10.1002/9780470725184.

Sargent, R.G., 2004. Validation and Verification of Simulation Models. In: Proceedings of the 36th Conference on Winter Simulation, pp. 17–28.

Schmolke, A., Thorbek, P., DeAngelis, D.L., Grimm, V., 2010. Ecological models supporting environmental decision making: a strategy for the future. Trends in Ecology & Evolution 25 (8), 479–486. https://doi.org/10.1016/j.tree.2010.05.001.

Shahed, B., Zhu, Z., Matott, L.S., Rabideau, A.J, 2020. A new tool for automatic calibration of the Storm Water Management Model (SWMM). Journal of Hydrology 581, 124436. https://doi.org/10.1016/j.jhydrol.2019.124436.

Shen, H., Tolson, B.A., Mai, J., 2022. Time to Update the Split-Sample Approach in Hydrological Model Calibration. Water Resources Research 58 (3), e2021WR031523. https://doi.org/10.1029/2021WR031523.

Silberstein, R.P., 2006. Hydrological models are so good, do we still need data? Environmental Modelling & Software 21 (9), 1340–1352. https://doi.org/10.1016/j.envsoft.2005.04.019.

Šimůnek, J., van Genuchten, M.Th., Šejna, M, 2016. Recent Developments and Applications of the HYDRUS Computer Software Packages. Vadose Zone Journal (7), 15. https://doi.org/10.2136/vzj2016.04.0033.

Skala, V., Dohnal, M., Votrubova, J., Vogel, T., Dusek, J., Sacha, J., Jelinkova, V., 2020. Hydrological and thermal regime of a thin green roof system evaluated by physically-based model. Urban Forestry & Urban Greening 48, 126582. https://doi.org/10.1016/j.ufug.2020.126582.

Smaldino, P.E., McElreath, R., 2016. The natural selection of bad science. Royal Society Open Science 3 (9), 160384. https://doi.org/10.1098/rsos.160384.

Soulis, K.X., Valiantzas, J.D., Ntoulas, N., Kargas, G., Nektarios, P.A., 2017. Simulation of green roof runoff under different substrate depths and vegetation covers by coupling a simple conceptual and a physically based hydrological model. Journal of Environmental Management 200, 434–445. https://doi.org/10.1016/j.jenvman.2017.06.012.

Stovin, V., 2010. The potential of green roofs to manage Urban Stormwater. Water and Environment Journal 24 (3), 192–199. https://doi.org/10.1111/j.1747-6593.2009.00174.x.

Stovin, V., Poë, S., Berretta, C., 2013. A modelling study of long term green roof retention performance. Journal of Environmental Management 131, 206–215. https://doi.org/10.1016/j.jenvman.2013.09.026.

Stovin, V., Vesuviano, G., Kasmin, H., 2012. The hydrological performance of a green roof test bed under UK climatic conditions. Journal of Hydrology, 414–415 148–161. https://doi.org/10.1016/j.jhydrol.2011.10.022.

Støvring, J., Dam, T., Jensen, M.B., 2018. Hydraulic Performance of Lined Permeable Pavement Systems in the Built Environment. Water 10 (5). https://doi.org/10.3390/w10050587.

Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F., Benedetti, L., Rico-Ramirez, M.A., de Carvalho, R.F., Breuer, L., Shucksmith, J., Heuvelink, G.B.M., Tait, S., 2019. Recent insights on uncertainties present in integrated catchment water quality modelling. Water Research 150, 368–379. https://doi.org/10.1016/j.watres.2018.11.079.

Tscheikner-Gratl, F., Lepot, M., Moreno-Rodenas, A., Schellart, A., 2017. QUICS D.6.7 - A Framework for the application of uncertainty analysis. Zenodo. https://doi.org/10.5281/zenodo.1240926.

Tzoulas, K., Korpela, K., Venn, S., Yli-Pelkonen, V., Kaźmierczak, A., Niemela, J., James, P., 2007. Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review. Landscape and Urban Planning 81 (3), 167–178. https://doi.org/10.1016/j.landurbplan.2007.02.001.

Versini, P.-A., Ramier, D., Berthier, E., de Gouvello, B., 2015. Assessment of the hydrological impacts of green roof: From building scale to basin scale. Journal of Hydrology 524, 562–575. https://doi.org/10.1016/j.jhydrol.2015.03.020.

Versini, P.-A., Stanic, F., Gires, A., Schertzer, D., Tchiguirinskaia, I., 2020. Measurements of the water balance components of a large green roof in the greater Paris area. Earth System Science Data 12 (2), 1025–1035. https://doi.org/10.5194/essd-12-1025-2020.

Vesuviano, G., Sonnenwald, F., Stovin, V., 2013. A two-stage storage routing model for green roof runoff detention. Water Science and Technology 69 (6), 1191–1197. https://doi.org/10.2166/wst.2013.808.

Voskamp, I.M., de Luca, C., Polo-Ballinas, M.B., Hulsman, H., Brolsma, R., 2021. Nature-Based Solutions Tools for Planning Urban Climate Adaptation: State of the Art. Sustainability (11), 13. https://doi.org/10.3390/su13116381.

Wadzuk, B.M., Hickman, J.M., Traver, Robert G., 2015. Understanding the Role of Evapotranspiration in Bioretention: Mesocosm Study. Journal of Sustainable Water in the Built Environment 1 (2), 04014002. https://doi.org/10.1061/JSWBAY.0000794.

Walker, Warren E., Haasnoot, M., Kwakkel, J.H., 2013. Adapt or Perish: A Review of Planning Approaches for Adaptation under Deep Uncertainty. Sustainability 5 (3), 955–979. https://doi.org/10.3390/su5030955.

Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B.A., Janssen, P., von Krauss, M.P.K., 2003. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment 4 (1), 5–17. https://doi.org/10.1076/iaij.4.1.5.16466.

Wheater, H., Sorooshian, S., Sharma, K.D., 2007. Hydrological Modelling in Arid and Semi-Arid Areas. Cambridge University Press. https://doi.org/10.1017/CBO9780511535734.

Wöhling, T., Samaniego, L., Kumar, R., 2013. Evaluating multiple performance criteria to calibrate the distributed hydrological model of the upper Neckar catchment. Environmental Earth Sciences 69 (2), 453–468. https://doi.org/10.1007/s12665-013-2306-2.

Wooster, E.I.F., Fleck, R., Torpy, F., Ramp, D., Irga, P.J., 2022. Urban green roofs promote metropolitan biodiversity: A comparative case study. Building and Environment 207, 108458. https://doi.org/10.1016/j.buildenv.2021.108458.

Yang, Y., Chui, T.F.M., 2019. Developing a Flexible Simulation-Optimization Framework to Facilitate Sustainable Urban Drainage Systems Designs Through Software Reuse. In: Peng, X., Ampatzoglou, A., Bhowmik, T. (Eds.), Reuse in the Big Data Era. Springer International Publishing, pp. 94–99. https://doi.org/10.1007/978-3-030-22888-0_7.

Yio, M.H.N., Stovin, V., Werdin, J., Vesuviano, G., 2013. Experimental analysis of green roof substrate detention characteristics. Water Science and Technology 68 (7), 1477–1486. https://doi.org/10.2166/wst.2013.381.

Zhang, S., Guo, Yiping, 2015. Analytical Equation for Estimating the Stormwater Capture Efficiency of Permeable Pavement Systems. Journal of Irrigation and Drainage Engineering 141 (4), 06014004. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000810.