Environmental Modelling & Software 82 (2016) 308–320

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A user-driven case-based reasoning tool for infilling missing values in daily mean river flow records

Laura Giustarini*, Olivier Parisot, Mohammad Ghoniem, Renaud Hostache, Ivonne Trebs, Benoît Otjacques

Luxembourg Institute of Science and Technology (LIST), Environmental Research & Innovation Department (ERIN), 5, avenue des Hauts-Fourneaux, L-4362 Esch/Alzette, Luxembourg

ARTICLE INFO

Article history:
Received 19 November 2015
Received in revised form
23 March 2016
Accepted 12 April 2016

Keywords:
Gap filling
Hydrological time series
Case-based reasoning
Knowledge database
Human in the loop

ABSTRACT

Missing data in river flow records represent a loss of information and a serious drawback in water management. In this work, we introduce gapIt, a user-driven case-based reasoning tool for infilling gaps in daily mean river flow records. Given a set of flow time series, gapIt builds a database of artificial gaps for which it computes several flow estimates, to find the best combinations of infilling algorithm and automatically selected donor station(s), according to state-of-the-art performance indicators. We obtained satisfactory results with Nash-Sutcliffe >0.7 for more than half of the ~5000 synthetic gaps of various lengths and positions, randomly created along the available records. gapIt was evaluated on 24 daily river discharge time series recorded in Luxembourg over seven years from 01/01/2007 to 31/12/2013. We also discuss the benefits of coupling this approach with user-expertise for an improved infilling of real data gaps.

Software availability

Name of software: gapIt
Developer: Olivier Parisot (olivier.parisot@list.lu)
Programming language: Java
Required hardware: 4 GB RAM minimum
Supported systems: Windows, Unix, Linux, Mac
Required softwares: Maven (≥3.0.2), JDK (≥1.7)
Availability: https://github.com/ERIN-LIST/gapIt
License: GNU General Public License version 3

1. Introduction

Long uninterrupted hydrological time series are often not available for many of the stream gauges in the world. Rather, time series of hydrological data are often affected by data gaps, which are discontinuities in the record of data. They are an inevitable consequence of factors such as station maintenance, equipment malfunctioning, human errors, changes in instrumentation and data processing issues (Harvey et al., 2010). Missing data in river flow records represent a loss of information and a serious drawback in water management. The existence of gaps results in difficulties in data interpretation and is a large source of uncertainty in data analysis. Specifically, the presence of discontinuities precludes the computation of hydrological statistics and physiographic indices. It also limits the use of such data for hydrological or hydrodynamic model calibration/validation purposes. A consequence of these issues is the need of data infilling methods to reconstruct missing data, when appropriate and before hydrological time series can be used in a number of applications.

From a technical point of view, a wide choice of data analysis tools is nowadays offered to hydrologists. For instance, specific user friendly software tools are already available or can be developed in platforms like R[1] or Matlab[2] to interpolate missing data and/or address hydrological problems. But most of these tools require some data mining and machine learning expertise, as well as fine-tuning in order to meet user needs and be properly exploitable by end-users (Serban et al., 2013). As a result, hydrologists have access to a collection of usable tools, but they still need to deal with several technical issues (like *data wrangling, tuning predictive algorithms*)

* Corresponding author.
E-mail address: laura.giustarini@gmail.com (L. Giustarini).

[1] http://www.r-project.org/.
[2] http://www.mathworks.com/products/matlab/.

http://dx.doi.org/10.1016/j.envsoft.2016.04.013

before solving their initial problem, i.e. infilling missing values.

Data infilling is a challenging task that has been addressed by previous research work.

2. Related work

For infilling gaps in hydrological time series, classical methods of data analysis have long been applied (Salas, 1980) and recent studies have proposed more efficient techniques (Harvey et al., 2010; Mwale et al., 2012). Most of the methods proposed in the literature are based on data transfer from one or more donor stations (gauges) to a target station. Among all possible infilling methods, the choice of the most appropriate one is not a trivial task. The same holds true for the selection of a set of donor stations. Moreover, results greatly depend on the context and, in non automated techniques, also on the user expertise.

Recently, Harvey et al. (2010) tested different infilling techniques, simulating an entire target flow record, for several stations in the UK. For each target station, two donor stations were selected a priori, based on the hydrological knowledge of the region and catchment metadata. Their work focused on the performance analysis of gap infilling techniques. In a follow-up study, Harvey et al. (2012) assessed a wide range of target-donor combinations, trying at the same time to improve data infilling performance by either seasonally grouping flows or excluding known inhomogeneity.

Gyau-Boakye and Schultz (1994) presented a Decision Support System (DSS) for selecting the most appropriate infilling model, as a function of gap length, season, climatic region and data characteristics of the records. The main disadvantage of their approach is that all rules are *hard-coded* and specific to a given region, namely West Africa. The same idea was applied by Johnston (1999) to build a DSS that helps experts select an estimation method for missing rainfall data in the United States. More recently, Griffioen et al. (2006) proposed a Case-Based Reasoning system (CBR) to intercompare water stress among different catchments in Europe. In their work, CBR was presented as a retrieval method to offer large amounts of filtered information to the end-user. In a broader context, Matthies et al. (2007) provided a review on environmental DSS, showing a general tendency towards integration and visualization of temporal and spatial results.

Despite the numerous studies available in the literature, a standardized procedure for gap filling in hydrological time series is still missing. One of the main limitations of many of the currently available approaches is their incomplete level of automation. Generally, donor stations are often determined a priori and tend to be specific to only a given region of interest. The user expertise is fundamental for this type of settings but it also limits the level of automation and the transferability of the approach to different areas.

In this work, we present a first attempt towards standardization, providing an interactive tool that allows performing gap filling in a consistent and traceable manner, bridging the gap between data-driven and user-expertise approaches.

gapIt is an interactive and visual data-driven tool that offers several infilling techniques, coupled with different sets of donor stations. It assesses the performance of all possible configurations, i.e. combinations of infilling method and set of donor station(s), to fill a given gap in a consistent way, eventually providing the best data-driven solution according to performance indicators. The visual interface allows users to select different infilling methods and/or donor station(s) than those automatically proposed by the tool, according to their expertise and specific knowledge of the region of interest. The fact that users can interactively inject their knowledge allows an iterative refinement of the results, while keeping track of all modifications.

In the general practice, infilling techniques require both a strong methodological background and a significant knowledge of the application domain (Maimon and Rokach, 2005; Domingos, 2012). In this paper, we show how gapIt can provide a bridge between a purely data-driven approach and an infilling method based on user expertise only. The automated approach, coupled with a visual inspection system for user-defined refinement, allows for standardized infilling, where subjective expert decisions can easily be incorporated in a traceable manner.

In the remainder of this paper, we will present a case study and the related data sources (section 3). Then we describe the proposed gapIt algorithm in section 4, which is followed by an analysis of the results obtained for both synthetic and real gaps in section 5. Advantages and limitations of the method are summarized at the end.

3. Case study

The dense river network of hydrometric stations in Luxembourg offers an excellent opportunity to test the proposed tool. The gauge network considered here is composed of 24 stations, displayed in Fig. 1, including both very responsive and groundwater-fed rivers. The region has a temperate, semi-oceanic climate. Precipitation is relatively uniform throughout the year, although strong seasonality in low flow exists due to higher evapotranspiration from July to September. High discharge values are recorded in winter (maximum January–February), sometimes leading to inundations, while low flows are observed particularly in September. The influence of snow can be considered negligible.[3]

We use discharge data, originally available as 15-min time series and subsequently aggregated using gapIt itself to daily values, covering the period from 01/01/2007 to 31/12/2013. A total number of 28 gaps are present in the dataset; most of them have been observed in winter.

4. Methods and tools

In this section, algorithm implementation and input data requirements for gapIt will be described. It has to be noted that this approach is based on a single variable, discharge, provided as input to the software. This loosens dependency on other types of variables, for instance catchment rainfall, which may not always be available (Harvey et al., 2010). In the following, we designate as target station (respectively, donor station(s)) the station characterized by a gap to be infilled (respectively, the station(s) whose data is used to derive infilled data for the target). The underlying hypothesis of the presented tool is the availability of a sufficiently dense river network that provides continuous measurements. gapIt infills gaps in discharge time series, providing the final user with the best solution that is possible to obtain, given the available donor stations. The best solution is individuated based on performance measures. As we are dealing with both synthetic and real gaps, two different strategies will be proposed to compute performance measures, depending on the type of gap. The insertion of estimated values in the database in lieu of gaps is subject to the acceptance by the end-user.

All infilled data are consistently flagged, for the sake of traceability of the reconstructed values. Moreover, the configuration used for infilling each gap is stored in the database, for the sake of reproducibility.

[3] http://www.hydroclimato.lu.

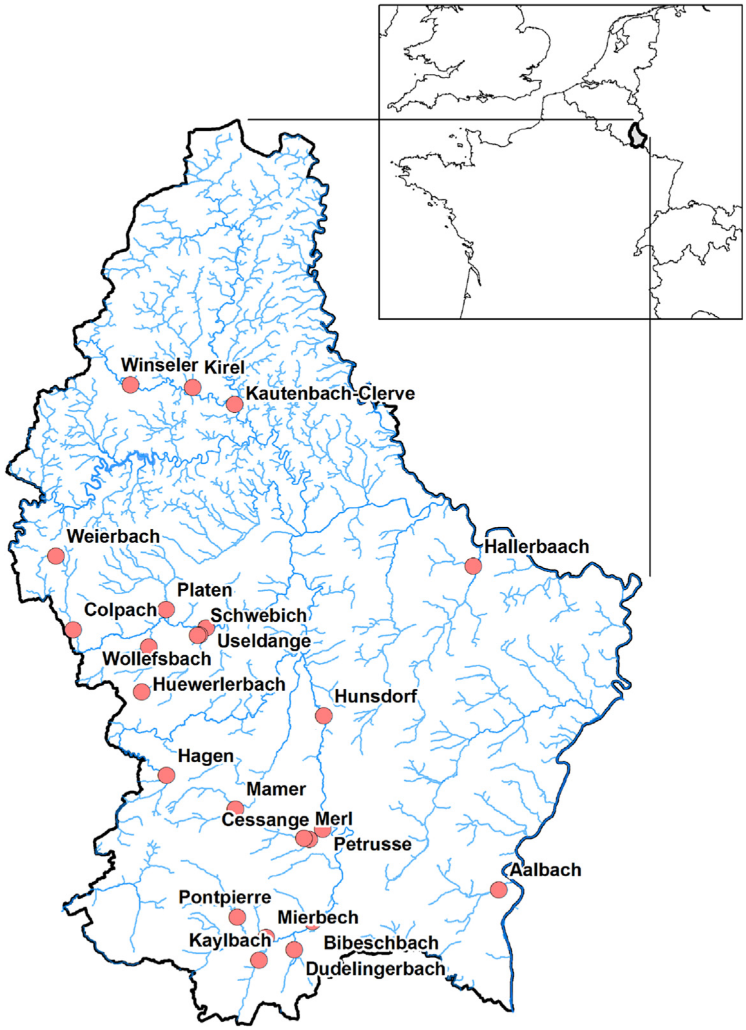

Fig. 1. Gauging stations in the river network of Luxembourg. The upper right panel shows the location of Luxembourg in North West Europe.

4.1. Software architecture and third-party libraries

gapIt is a multiplatform standalone tool developed in Java. It is mainly based on Cadral, a data analysis platform (Didry et al., 2015) developed in-house and leveraging WEKA (Witten et al., 2011) for data mining purposes, and JFreeChart (Gilbert, 2002) for graphical data representation.

4.2. Input data

gapIt is capable of importing/exporting data in CSV format or in WEKA's ARFF format. The data infilling process uses the following input data:

- Time series of hydrological data (discharge), recorded at several gauging stations within the river network described earlier. For each station, the data consists of a sequence of numerical values with timestamps.
- Geographic coordinates of gauging stations (longitudes and latitudes).
- Upstream/downstream relationships (if applicable, depending on the stations); this information is stored in the form of a dependency graph, a simple scheme displaying the different relationships among stations.

In case an upstream/downstream station is not present or available in the same river as the target station, we use the closest station in a tributary as a proxy upstream gauge and, as a proxy downstream gauge, the closest station in the river where the stream with the target station flows.

4.3. Visualization and data preprocessing

A graphical user interface (GUI) allows the visual inspection of data, in terms of time series, maps and relationships. The user can typically visualize the different gaps (Fig. 2) in their spatio-temporal context. The GUI consists of three panels showing:

- The list of gaps present in all time series (Fig. 2, panel 1).
- A map (Fig. 2, panel 2) showing the locations of the gauging stations. Based on the user selection, the gauge of interest is interactively highlighted, to easily put it in its geographic context.
- A line chart (Fig. 2, panel 3) to let end-users inspect temporal trends of the selected time series and put gaps in their temporal context.

Moreover, the tool offers several data preprocessing features. In particular, the user can set the desired temporal resolution of the

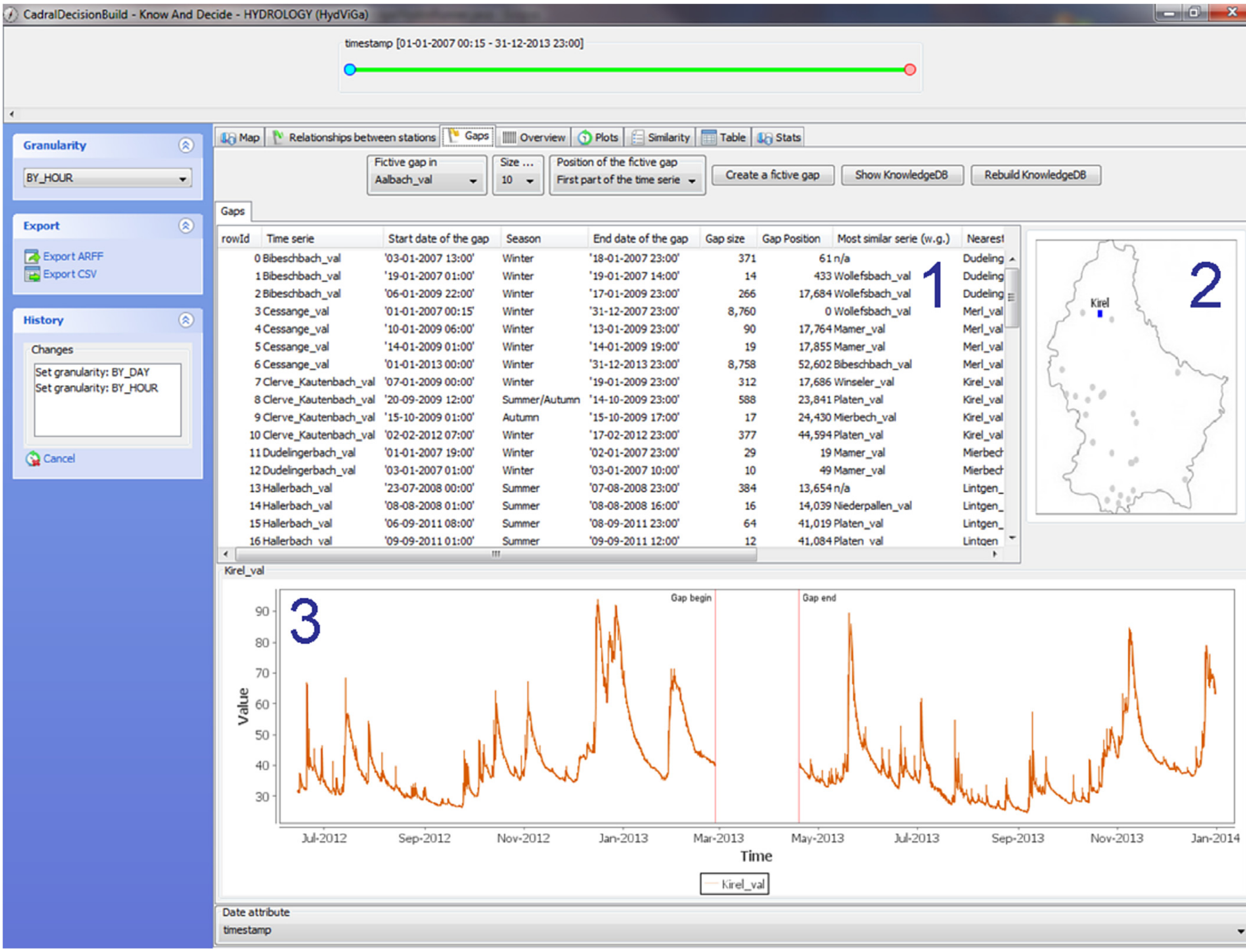

Fig. 2. Graphical user interface of gapIt: (1) gap list, (2) map showing the locations of the available stations, (3) time series visualization.

time series, i.e. creating hourly, daily and monthly aggregate values. In the present study, it was deemed more appropriate to perform gap infilling on daily values, rather than on the noisier original 15-min observations (Harvey et al., 2010, 2012).

4.4. Gap characterization and inspection

The choice of any particular gap-filling technique depends on several factors: gap length, season, high/low flow conditions, etc. (Gyau-Boakye and Schultz, 1994). As a consequence, these characteristics are determined in gapIt and are accessible by the user through the GUI (Fig. 2).

- Flow (low/medium/high): given the minimum and maximum value of a certain time series (i.e. station), three equispaced ranges are defined in order to characterize the type of flow before and after a gap.
- Rising limb: an outlier detection method is used to detect the presence of a rising limb, which may indicate the *probable* presence of a flood event for the considered gap. More precisely, the method is based on the *local outlier factor* which computes a score for each value of the time series by taking into account values before and after it (Breunig et al., 2000).
- Geographical proximity: the closest station is computed by applying a simple Euclidean distance on station coordinates.
- Upstream/downstream relationships: the identification of upstream and downstream station(s), if present, is carried out on the basis of the input dependency graph, using a simple nearest-neighbor search.
- Similarity between time series: this is computed using Dynamic Time Warping (DTW) (Berndt and Clifford, 1994). This method is rather popular in time series analysis as it takes into account time shift and distortion and was already used in hydrology to find patterns in discharge data (Ouyang et al., 2010). In practice, due to the high time complexity of DTW, we use an empirically defined time window of size *N*gapsize* (with *N* fixed by the end-user — a reasonable value for *N* between 3 and 10 helps defining a compact and sufficient time window).

These properties are computed and displayed in the GUI (Fig. 2).

4.5. Gap infilling

To fill a given gap, we propose a two-phase approach described in Algorithm 1: the tool starts by computing the best solution automatically. Then, the end-user can either accept or refine it.

More precisely, different configurations are considered to infill a gap. By configuration we mean a combination of donor station(s) and gap infilling method. For a given gap, gapIt computes all possible configurations and ranks them according to errors and performance measures (Section 4.6). The best configuration is automatically selected by the tool as the one yielding the smallest root mean squared error (RMSE) (Please see Section 4.6 for more details). Subsequently, the user/expert can refine it by adjusting any of its constituents.

Eventually, the configuration approved by the user is applied to reconstruct the missing data which gapIt stores and flags as *infilled* data.

Algorithm 1 gapIt infilling approach.

Require:
a gap (GAP)

Ensure:
/* *Automated selection of the donor stations (DS)* */
$selectedDS \Leftarrow$ empty list
add the geographically closest station into $selectedDS$
add the downstream station into $selectedDS$ (if applicable)
add the upstream station into $selectedDS$ (if applicable)
add the most similar station into $selectedDS$
/* *Automated selection of an infilling method* */
$selectedM \Leftarrow$ identify combination of method and set of $selectedDS$ providing the best infilling performance
/* *User refinement phase* */
while the user is not satisfied **do**
    $selectedDS \Leftarrow$ manual user refinement of the donor stations
    $selectedM \Leftarrow$ manual user selection of an infilling method
    evaluation of infilling performance for GAP by using $selectedDS$ and $selectedALGO$
end while
apply $selectedALGO$ by using DS in order to fill GAP

4.5.1. Selection of the donor stations

Selecting donor stations represents the most critical step of the infilling procedure, as their data is going to be used to estimate the missing values. Based on the gap characterization data (Section 4.4), the software offers several options to automatically select donor station(s) among:

- the geographically closest station;
- the station having the most similar time series (based on DTW as explained in Section 4.4);
- the upstream and/or downstream station.

It is important to highlight that both the geographically closest station and the one with the most similar time series may not belong to the same catchment as the target station. In addition, all different combinations of donor stations are potentially applicable depending on the case at hand. For example, the algorithm and/or the user may use the downstream station and ignore the geographically closest one. The final set of donor stations depends on several factors, such as the context, and user expertise, among others.

4.5.2. Selection of an infilling method

Observations with missing data can of course simply be omitted in the user application: this trivial approach can be sufficient in several cases (Enders, 2010) or when infilling may risk being detrimental (Beven and Westerberg, 2011; Beven and Smith, 2014) leading to periods of disinformative hydrological data. However, it goes without saying that data infilling is an extremely helpful approach to make the best possible use of time series, for example to derive accurate long-term statistics. Dealing with missing values is a well-known topic in data mining and approaches generally used in this field can be easily transferred to hydrology.

Among the various infilling methods proposed in the literature (Pigott, 2001; Marwala and Global, 2009; Van Buuren, 2012), a comprehensive, though not exhaustive, range of available methods was selected and integrated into gapIt:

- Interpolation (INTERP) is an easy and efficient solution if time series do not present steep increases or decreases of measured values (jumps) in data and when the gap length is rather small.
- Mean value (ZeroR) is a simple solution that consists in replacing missing numerical values by the mean value: this approach is still used in many statistical software packages. However, it can highly disrupt the data structure, thus degrading the performance of statistical modelling (Junninen et al., 2004). In this paper, we use Weka's ZeroR classifier.
- The Nearest-neighbors (NN) technique as implemented in Weka was applied as follows: for each incomplete record, similar records are identified (by using the Euclidean distance as a *brute force* search) among the already selected donor stations, and then used to estimate missing values.
- Multiple linear regressions (REG) are rather frequently used, particularly when links are evident among sensors of the network: they can capture relationships between downstream and upstream gauges (Bennis et al., 1997). In gapIt, we integrated the *Ordinary Least Square* method, provided by the Apache Commons Mathematics Library.[4]
- Regression Trees (RT) are suitable too as they are efficient and easy to visualize/interpret (Kotsiantis, 2013): rules are explicitly

[4] http://commons.apache.org/proper/commons-math/index.html.

described by the tree and are more expressive than the classical linear regression formula (Witten et al., 2011, section 3.3). gapIt uses Weka's REPTree implementation.

- Model trees (MT) are included by applying the M5 method (Quinlan, 1992); this technique was recently used to forecast flows in Turkey (Sattari et al., 2013). gapIt uses Weka's M5P implementation (Witten et al., 2011, section 6.6).
- Artificial Neural Networks (ANN) have been recently used to preprocess missing hydrological data (Mwale et al., 2012; Tfwala et al., 2013). Although some work has been done to ease the interpretation of ANN (Féraud and Clérot, 2002), they are still discounted as *black box* models. Yet, they represent an extremely helpful approach to build powerful predictive models, capable of providing satisfactory results. gapIt uses the implementation of the multilayer perceptron with back propagation (Witten et al., 2011, section 6.4).
- Expectation-Maximization (EM) method (Van Hulse and Khoshgoftaar, 2008) was also included in gapIt, through a Weka plugin which uses EM to replace missing values with a multivariate normal model.[5]

The above listed techniques are implemented and available in gapIt. As a complement to all of them, temporal discretization is included as an additional option. In other words, for all the eight methods implemented in gapIt, when applicable, one test is made considering the discretization in time of the governing equations, while the second computation is performed excluding this additional step. The idea behind this is to take the best advantage of time discretization in all cases where it helps the algorithm in capturing temporal patterns (i.e. adding date-derived periodic attributes like *month of the year* or *quarter*).[6]

While the application of ANN is accepted as a form of rainfall-runoff modelling (Gao et al., 2015), it has to be noted that conventional hydrological modelling was not included in this work. According to Harvey et al. (2010), current rainfall-runoff models are still too demanding in terms of computation time, resources and input data, with the need of calibration limiting transferability among catchments.

More sophisticated techniques, like flow-flow models for donor stations, as well as forms of inverse hydrology (Croke, 2006; Kirchner, 2009; Kretzschmar et al., 2014) could be implemented in gapIt. However, for the moment we deliberately intended to limit the involved inputs and the required knowledge and systemic understanding of hydrological processes.

4.6. Evaluation of the infilling accuracy

Eventually, the user can inspect the results in the GUI, together with an evaluation of the infilling performance (Fig. 3). The accuracy of the gap-filling procedure can be assessed using several errors and performance measures: RMSE, mean absolute error (MAE) and Nash-Sutcliffe coefficient (NS) (Nash and Sutcliffe, 1970). In practice, a perfect gap infilling will lead to MAE and RMSE values equal to 0 and an NS value of 1. In the literature, ranges of satisfactory fits are provided for various performance indicators (Moriasi et al., 2007a; Harmal et al., 2014). The NS coefficient is frequently used in hydrology and has the interesting characteristic of being dimensionless, which allows the comparison between different catchments and periods. Its main limitation lies in the fact that the differences between observed and predicted values are squared. Thus, larger values in a time series are overestimated while lower values are neglected (Krause et al., 2005; Legates and McCabe, 1999). This leads to an overestimation of the model performance during peak flows and an underestimation during low flow conditions. One should also note that all performance indicators have their shortcomings. For example, the index of agreement leads to relatively high values even for poor model fits and, like the NS, is not sensitive to systematic model over- or underprediction. Other additional measures are implemented in the tool, like the index of agreement and the percent bias (PBIAS) (Moriasi et al., 2007b), but they are not included in this study for the sake of brevity.

As anticipated we deal with both synthetic and real gaps and, correspondingly, the strategy to compute the evaluation measures slightly differs for the two situations.

4.6.1. Infilling accuracy for synthetic gaps

In the case of a synthetic gap, the estimated series for gap infilling is compared to the actual observations, only in the time window of the gap itself (Evaluation strategy A). Visually, the tool simply presents the result by indicating the errors and by plotting the observed and estimated time series on the same plot (Fig. 4).

4.6.2. Infilling accuracy for real gaps

In contrast, when dealing with a real gap, the software will first individuate a time window that is centered on the gap itself but larger than it. The enlargement is set as a percentage of the gap size itself. The infilling performance (Evaluation strategy B) is computed on values before and after a given real gap, where observed values are available.

Obviously in case of a real gap, the performance measures can only hint to the algorithm performance, as they are computed over a period that does not exactly match the time span of the gap.

4.7. Selection of the best solution

In conclusion, after computing all possible configurations, i.e. donor station(s) and infilling methods, the software will automatically provide the best solution, characterized by an optimal configuration, i.e. the one yielding the smallest RMSE (Evaluation strategy A for synthetic gaps, Evaluation strategy B for real gaps). All other indicators, e.g. NS, index of agreement, are provided to the user to better contextualize the infilling performance and accuracy. The user can always adjust the configuration, to identify the solution that is deemed the most appropriate. The user expertise represents an invaluable resource to interpret the results and improve and/or reject model outputs.

The best solution retained in the end can be different from the one initially found by the algorithm in a purely data-driven approach. Users may simply accept the solution with the best performance measures or use it as a starting point, to be refined based on their expertise, by adjusting the configuration. They can also completely discard the proposed solution, leaving the gap unfilled if none of the configurations is deemed appropriate. This kind of post-processing combines the best of a data-driven approach and a manual infilling based on the human knowledge of the problem/context.

4.8. A case-based reasoning module for traceability and decision support

For a given gap (synthetic/real), gapIt determines the best solution as the one with the best performance (Evaluation strategy A/B), given the available inputs. However, to take the best advantage of the tool, a knowledge database can be built with synthetic gaps appropriately created and infilled. In this case, to

[5] http://weka.sourceforge.net/packageMetaData/EMImputation/index.html.
[6] http://tinyurl.com/padzrt4.

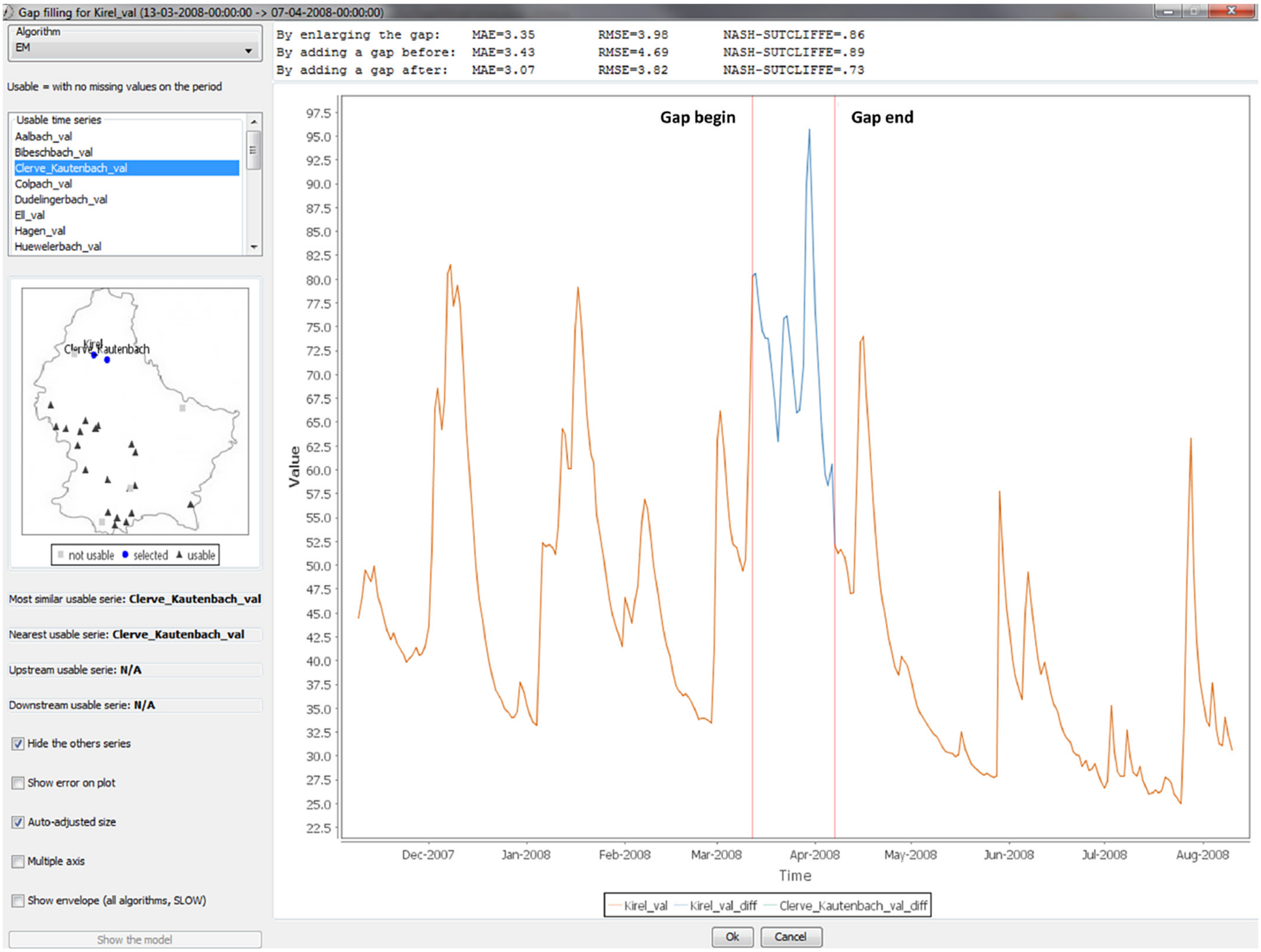

Fig. 3. User interface to visualize and infill a true gap: by default, a solution is selected by the tool: the donor stations and the infilling method are automatically chosen (Algorithm 1). Moreover, users can refine it by selecting different donors and infilling methods according to their knowledge.

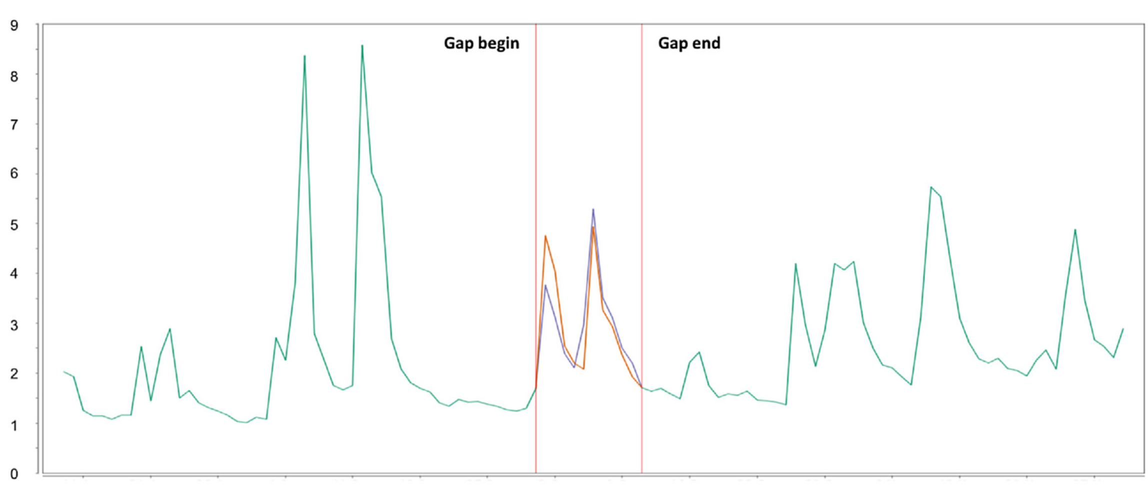

Fig. 4. A synthetic gap in the discharge (m^3/s) series at Hunsdorf station in 2008. Observed values are depicted in orange, while the estimated ones are in purple. In this case, the best infilling is achieved by ANN using Useldange (the station having the most similar time series) and Schwebich (the geographically closest station) as donor stations, yielding the following performance values: MAE = 0.42, RMSE = 0.54 (m^3/s), NS = 0.73. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

build the database itself, the identification of best solutions always relies on the application of Evaluation strategy A. The main goal is to create a DSS, in the form of a CBR system that solves new cases based on the solutions of similar past ones (Aamodt and Plaza, 1994). More precisely, the CBR module is an alternative to the default selection of donor stations and infilling method (Algorithm 1, lines 2–9).

Synthetic gaps are automatically infilled by a batch process, applying on them all possible configurations (method/use of nearest station data/use of most similar series/use of downstream station data/use of upstream station data/use of nearest station data/use periodic attributes/etc.). As a result, gapIt generates a knowledge database that stores all the created synthetic gaps and the best solution (Evaluation strategy A) obtained for each of them:

- For each configuration, the gap filling accuracy is evaluated (MAE/RMSE/NS/etc.), comparing estimated and observed values, and then it is stored into the knowledge database.
- In addition, for each synthetic gap, the configuration yielding the highest accuracy is flagged as the best solution.

In the knowledge database, a case is defined by gap characteristics and properties, configuration (infilling method, donor stations), and performance measures (i.e. RMSE, NS) (Table 1).

For any new gap, the most similar cases in the knowledge database are retrieved by using the k-nearest-neighbors technique. Users can choose to apply one of the best past configurations suggested by gapIt on the new gap (Fig. 5). Otherwise, as discussed earlier, they can use one of the suggested past configurations as a starting point, for further expert-driven refinement.

In conclusion, for a new real gap, gapIt provides more than one data-driven best solutions: the configuration with the smallest RMSE (Evaluation strategy B) and the most similar case(s), based on past cases found in the knowledge database. Any of these can be used as a starting point for user-driven refinement consisting in modifying the infilling method and/or donor station(s), in order to define the best solution.

After the infilled values have been approved by the end-user, gapIt stores them in the database keeping a flag on all reconstructed values, along with the configuration used to infill each data value. For each reconstructed time series, the database contains the previously-computed errors that are related to the missing values estimations, in order to propagate the uncertainty information (Table 1). This traceability is fundamental for further use of the processed data and helps deal with the uncertainty inherent in any data infilling approach.

5. Results

The gapIt software was applied to infill gaps in discharge times series measured at 24 gauging stations of the Luxembourgish gauge network. Before infilling real gaps, a first analysis was performed on synthetic gaps, in order to test the tool's capabilities and, at the same time, to build a knowledge database with synthetic gaps appropriately created and infilled.

5.1. Synthetic gaps

Removing actually observed data, 5108 gaps were randomly created, ranging in length from 2 to 100 days. We made sure that the different gap lengths are equally represented w.r.t. the total number of gaps. The synthetic gaps were uniformly distributed over the entire record duration and across seasons, in order to have a wide range of gaps, characterized by different lengths, distributed across various periods of the year. As low flow conditions are the most common in rivers and streams, the majority of synthetic gaps, 65% of the total, were created in low flow regime, while the remaining share was equally distributed between high and middle flow conditions.

The synthetic gaps thus created are representative of all types of gaps encountered in real time series, in terms of lengths and distribution across seasons and flow regimes.

After computing all possible configurations, gapIt provided the best solution to fill each synthetic gap, i.e.the configuration having the smallest RMSE. Needless to say, we encountered cases where also the best solution, out of all possible configurations, was nevertheless a sub-optimal one, because its RMSE was too high (and its NS too low and even negative) to be subsequently accepted by the end-user. This occurs when very few donor stations are available for the specific period of the gap and/or with little similarity to the target station.

The following analysis focuses on the best solutions, as provided by gapIt, for each of the 5108 synthetic gaps. To allow the comparison between different catchments and periods, results are discussed in terms of NS coefficients, even though gapIt uses RMSE to find the best solution.

The NS values characterizing the 5108 best solutions were classified in a set of intervals with 0.2 resolution, leading to 5 intervals: [<0.2], [0.2–0.4], [0.4–0.6], [0.6–0.8], [0.8–1.0]. For each interval, we counted the number of best solution yielding an NS value included in that interval. Subsequently, the count of best solutions per interval was divided by the total number of best solutions, i.e. 5108, to obtain a percentage. Fig. 6 shows the

Table 1
Subset of the knowledge database for the stations located at Petrusse and Hunsdorf. This subset includes several synthetic gaps and the respective infilling procedure. For each gap, the following characteristics are listed: station, gap characteristics, configuration of the best solution (infilling method and used donor station(s)), infilling performance (NS).

Gap characteristics				Properties		Infilling method	Donor stations				NS
Station	Length	Season	Year	Rising	Flow		Most similar	Closest	Downstream	Upstream	
Petrusse	4	Winter	2009	yes	low	NN	yes	yes	yes	yes	0.95
Petrusse	8	Winter	2009	yes	low	REG	no	yes	no	no	0.91
Petrusse	10	Autumn	2007	yes	low	MT	no	yes	yes	yes	0.70
Petrusse	20	Autumn	2007	yes	mid	NN	yes	yes	no	yes	0.98
Petrusse	50	Spring	2008	yes	low	RT	yes	no	yes	no	0.88
Hunsdorf	2	Summer	2013	no	low	NN	yes	yes	no	yes	0.72
Hunsdorf	3	Summer	2013	no	low	EM	yes	yes	no	yes	0.94
Hunsdorf	4	Summer	2013	no	low	RT	yes	no	no	no	0.92
Hunsdorf	5	Summer	2013	no	mid	NN	yes	no	no	no	0.91
Hunsdorf	6	Summer	2013	no	low	NN	yes	no	no	no	0.85
Hunsdorf	4	Winter	2009	yes	low	NN	yes	yes	yes	no	0.96

Fig. 5. The CBR module supports gap infilling by inspecting and applying similar past configurations. In other words, it provides alternative results to infill a current gap. Firstly, it identifies *similar* gaps that were corrected in the past (by applying the k-nearest-neighbors technique). Secondly, it retrieves the configuration that provided the best results. Finally, it reuses these settings to infill the gap at hand.

distribution of best solutions across different NS ranges.

It is reassuring to observe that 65% of the total gaps were reconstructed with NS > 0.6. Moreover, the percentage of best solutions with a given NS value increases with increasing NS values: for example, only 17% of the total best solutions have a NS between 0.0 and 0.2, whereas 48% of the total gaps were reconstructed with NS > 0.8.

As explained above, there is a small percentage of gaps (12%) that, even with the best solution proposed by gapIt, shows negative NS values. For example, a synthetic gap of 20 days created in summer in the Wollefsbach time series was reconstructed with a best NS of −0.02. Although the donor stations, namely Schwebich and Heuwelerbach, are both close to the target and lie in similar catchment areas, they present a discharge peak higher than what was observed in Wollefsbach. This is most likely caused by very localized rainfall in the region of the donor stations. Localized rainfall events are typical of summer and hinders the reconstruction of missing data with data transfer techniques, based on a single variable, i.e. discharge. Moreover, one must take into account that headwater catchments, like those in this example, are largely controlled by the underlying bedrock geology which may result in different hydrological responses even if the distance between streams is small (Wrede et al., 2015).

Furthermore, the 5108 best solutions were grouped by gap length. For each gap length group, the percentage of best solutions with different NS values was computed, following the same procedure as for Fig. 6. Some examples are plotted in Fig. 7.

For all gap lengths, the distribution of best solutions is left-tailed (more best solutions in high NS bins than in low NS bins). Note that the percentage of gaps filled with negative NS values decreases as

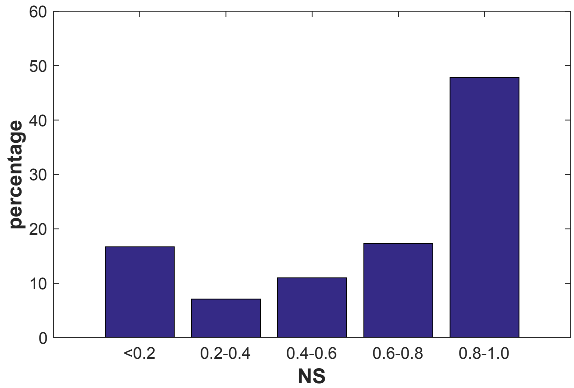

Fig. 6. The distribution of best solutions with different NS values.

gap length increases. This may be explained by the fact that NS tends to overestimate model performance during peak flows, which are more likely to be present in larger time windows, and to underestimate it during low flow conditions, which represent the majority of the shortest gaps.

Grouping the 5108 best solutions by season, the percentage of gaps with NS > 0.6 for spring, summer, autumn and winter are comparable and equal to 56%, 56%, 75%, 74%, respectively. If clustering by flow regime, practically all gaps in high and middle flow conditions are infilled with NS > 0.6, while this figure becomes 63% for low flow regime. As anticipated, this outcome is consistent with the fact that 65% of the gaps occur in low flow condition, the most common situation for river systems.

Additionally, a second test was performed by running gapIt one more time on the 5108 synthetic gaps, discarding the option of using (when available) upstream and/or downstream station(s). In this case, the share of best solutions having NS > 0.5 drops from 71% to 58%, indicating a decrease in performance even if this can still be considered a reasonable result. This indicates a limited influence, in our gauge network, of upstream/downstream gauge availability. In other words, the data-driven approach leads to good results in more than half of the analyzed cases.

A final test was conducted by infilling gaps only using the closest station as donor. This option further decreases to 49% the percentage of gaps filled with NS > 0.5. This result can be partially explained by the fact that, even though it can generally be assumed that nearby stations are affected by similar rainfall patterns, it cannot be expected that they belong to the same catchment (i.e. upstream/downstream, tributary, …) or to catchments showing similar hydrological responses to a given input.

For the following analysis, we revert to letting the algorithm freely select among all available donor stations. The 5108 best solutions were grouped with respect to the infilling method yielding the best solution. Obviously, the number of bins corresponds to the

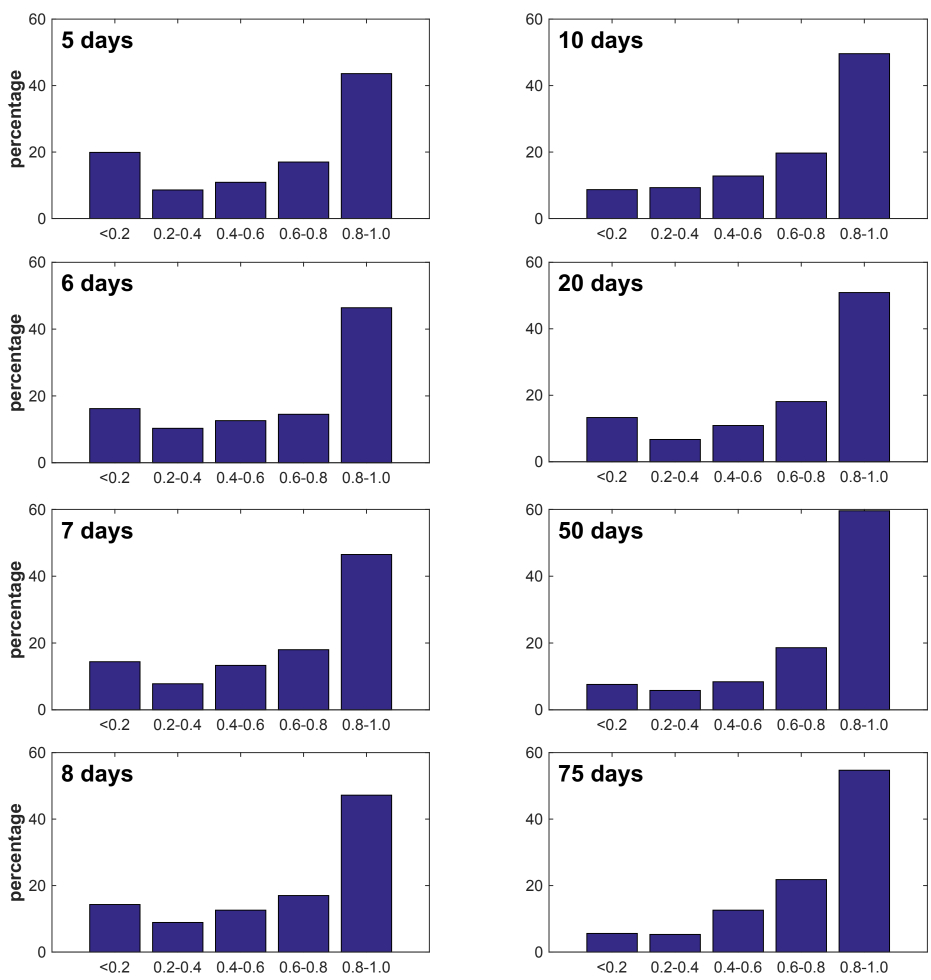

Fig. 7. Distribution of best solutions with different NS values, split by gap lengths.

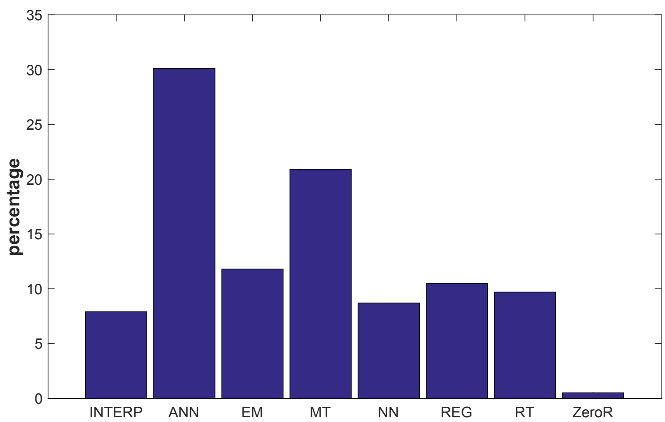

Fig. 8. The distribution of best solutions across different infilling methods. Here all gaps are considered, irrespective of their lengths.

number of infilling methods. For each bin, we divided the number of best solutions by the total number of best solutions, i.e. 5108, to obtain the percentage.

Fig. 8 shows the distribution of best solutions across all infilling methods, regardless of the number/type of donor stations and gap lengths. Under these conditions, ANN and MT were found to be the most accurate methods for infilling the majority of the synthetic gaps, with low MAE and RMSE values and high NS coefficients.

A similar trend to that displayed in Fig. 8 has been observed with seasonal gaps. When clustering by flow regimes, the percentages of selected methods are comparable to what was found for the total number of gaps. Interestingly, however, INTERP was never selected to infill gaps in high flow condition, while it was chosen for infilling 1% and 8% of middle and low flow regime gaps, respectively. This can be explained by the fact that low flow conditions tend to show less important discharge variations, while in case of a flood event, a rapid increase and/or decrease of values is usually observed.

Subsequently, we investigated the influence of gap length on the method that leads to the best solution. The total 5108 synthetic gaps were grouped by their gap length and for each group we applied the procedure used to obtain Fig. 8. Fig. 9 shows the distribution of best solutions, across gap length values, for all infilling methods.

For the groups of gaps characterized by small lengths (<5 days),

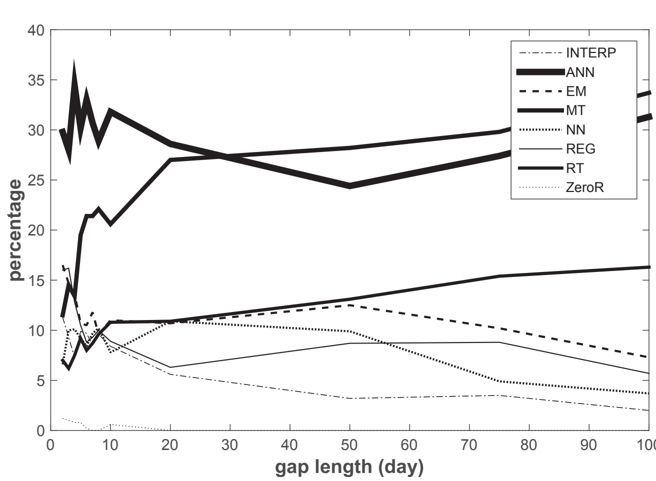

Fig. 9. The distribution of best solutions, grouped by gap lengths, for different infilling methods.

ANN provides the best solution for one case out of three, while other methods share equally the remaining cases. For gap lengths >5 days a clearer trend emerges: ANN maintains a high percentage (25–35%) of best solutions, while MT is more frequently the best method with increasing gap lengths. This confirms the general finding of Fig. 8, for the total number of synthetic gaps. In Fig. 9, also RT appears to be chosen more often with increasing gap lengths, even though its share is smaller than for ANN and MP. On the other hand, INTERP, EM, REG and NN show a decreasing trend with increasing gap lengths.

Cases with a gap length ≥50 days are regarded as an exercise of analysis rather than as suitable for real applications. Although good results were obtained in terms of NS for synthetic gaps, infilling such extended gaps may not be recommended in real cases, due to uncertainty generally increasing with gap length.

So far, the analysis of the best solutions was focused on either performance or infilling method. When considering the influence of donor station(s) that are automatically selected for the best solution, some interesting insights can be revealed. Investigating some of the best solutions, i.e. characterized by a high NS value, we found that donor stations had a catchment area comparable to that of the target station. Moreover, land cover classes, e.g. forest, grassland, agricultural, urban, and geology types, e.g. sandstone, marls, schists, alluvials, were found to be present in similar fractions in both donor station(s) and corresponding target station. Features like catchment area, land cover and geology are considered to be some of the drivers of the catchment response, in terms of rainfall-runoff transformation. At the same time, one must highlight that DTW was implemented in a non dimensionless form: in other words, in the selected time window gapIt compares the actually measured values of discharge, and not their dimensionless trends. Therefore, in case of similar rainfall patterns, it can be hypothesized that this particular implementation of DTW detects similarity among catchments in terms of catchment area, land use and geology, through their response to rainfall, also in cases when donor and target stations do not belong to the same catchment. In other words, gapIt helps identifying an appropriate donor station, even though this does not seem evident at first glance from a hydrological perspective.

When excluding human interaction, gapIt is designed to automatically find, for each gap, the configuration having the smallest RMSE, regardless of the number of donor stations considered for that specific infilling solution. In this standard setup, 71% of the 5108 synthetic gaps were infilled with NS > 0.5. To test the robustness of the proposed tool, we also tested gapIt performances in different setups, re-infilling the 5108 synthetic gaps and obtaining the results listed in Table 2.

From this analysis, we concluded that in its standard setup gapIt provides the highest percentage of gaps infilled with NS > 0.5. The inclusion of the station with the most similar series provides an improvement in performance from 61% to 71%, proving the added value of DTW for selecting donor stations. All the tests performed with different setups show the robustness of the tool, with NS > 0.5 for more than half of the synthetic gaps in all setups, with the exception of the one performed considering only the station with the most similar series (selected through DTW).

5.2. Real gaps

Half of the 28 real gaps are characterized by the presence of a downstream station that can be potentially used for reconstructing the missing time series, whereas only 4 have both upstream and downstream stations available and usable. In terms of best performing infilling method, we did not observe a trend as clear as with the synthetic gaps. This could be expected given the small

Table 2
Comparison of different setups.

Setup	Gaps infilled with NS > 0.5
Standard setup	71%
Standard setup but excluding the station with the most similar series	61%
Using only the station with the most similar series	44%
Using the station with the most similar series and the geographically closest one	58%

number of cases. However, it is interesting to report that MT was the top performer in 8 cases, while EM was never selected.

To assess the quality of infilling, performance measures were computed according to Evaluation strategy B. Out of the 28 real gaps, 19 achieved an NS coefficient >0.8. As mentioned before, this can be only regarded a simple hint and not as a real performance value. For instance, in an extreme case, a peak discharge may have occurred, due to localized rainfall, only in a limited area and it also may have been recorded (potentially) by only a single station, whereas all other stations in the surrounding region are not affected by rainfall and, hence, do not detect any particular discharge variation. In case the first station would be affected by a data gap, exactly during that flood event, none of the proposed infilling methods would be able to correctly reconstruct it. However, the performance indicators would be quite high, indicating good infilling results. This highlights the importance of combining automatic techniques with user expertise and knowledge, as implemented in gapIt, to obtain reliable gap infilling results.

6. Conclusions

In this paper, we presented gapIt, a tool for infilling gaps in hydrological discharge time series.

It was tested in the gauging network of Luxembourg to perform gap infilling on daily values, leading to satisfactory results on synthetic gaps. The tool was used for infilling ~5000 synthetic gaps, of different lengths and positions, randomly created along the entire records of all stations. More than half of the synthetic gaps were reconstructed with NS > 0.7. The software showed stable performance, regardless of gap length and flow regime. Superior performance was obtained by the use of methods such as neural networks and regression trees. Subsequently, gapIt was applied to infill 28 real gaps, ranging in length from 2 to 95 days. The good performance values obtained for more than half of them need to be considered as simple hints, as it is precisely in these situations that the added value of gapIt is revealed, consisting in combining an automated data-driven approach with the interactive user-driven refinements based on domain expertise through the visual inspection of the proposed solutions in their spatio-temporal context.

The proposed gapIt software provides a framework for a more standardized and traceable gap infilling process. The solution reached for infilling any specific gap can be retrieved for future analysis, while the infilled data are appropriately flagged. Ideally, the infilled data should also be characterized with an estimate of uncertainty. If the infilling is then used to calibrate or validate models, such uncertainty can be taken into account: that would comply with the good practice also suggested by Beven (2015). A proper uncertainty analysis should distinguish between any uncertainty deriving, for instance, from measurement instruments themselves, poor rating curves, etc., and uncertainty introduced by the infilling procedure. In the present case study, the uncertainty of the infilled data would derive partially from the uncertain data (discharge values computed through uncertain rating curves) used to reconstruct them and partially from the uncertainty inherent in the infilling method adopted. Future work will focus on uncertainty analysis, disentangling the different components of uncertainty (Beven and Westerberg, 2011), to better deal with the risks of injecting, through infilling, disinformative or inconsistent values.

In its current setup, the main advantage of the software is that it offers the opportunity of taking the best advantage of software automation and human expertise. When the user encounters a real gap, gapIt may help find the best possible solution through CBR. Nevertheless, the user can always modify both the proposed infilling method and/or the set of donor stations. In fact, a database of infilled synthetic gaps is available from the start in gapIt and from there the tool can suggest the best solution. However, any solution obtained for a real gap by an expert working with this tool is added to the database, indirectly including part of the human expertise into the database itself. The more cases are stored in the system, the better results it will achieve.

It is important to note that at present no limit for the distance between target and donor station was set. This is a consequence of the limited area of our case study. However, when dealing with larger regions, it is legitimate to foresee the application of spatial limits in the search for donor station(s). These limits may be set according to the knowledge of the hydraulic and hydrological behavior of the region and should probably be different from station to station. Like most CBR systems, gapIt will always provide a solution, regardless of the number of similar cases available, or the strength of the similarity. For example, if all donor stations except one have missing data in the same time period as the target station, the algorithm will use that single station as donor, as it is the only one available with recorded data, regardless of its location in space and/or its similarity in terms of DTW. This could be an issue in larger regions with several periods of instrument malfunctioning. Coupling the automatic algorithm with the visual inspection tool will allow the user to compensate the unavoidable drawbacks of a purely data-driven approach.

First promising results were obtained infilling 15-min values (i.e. original resolution of the present case study). However, due to the complexity of dealing with high frequency sampling, further testing is needed. Despite its versatility and capability of working with different temporal resolutions, in the present implementation gapIt cannot deal with irregularly sampled data and/or very sparse measurement networks, where the need for infilling is arguably more critical.

Acknowledgements

This study was funded by the Luxembourg Institute of Science and Technology (LIST). We are grateful to the anonymous reviewers for their valuable comments and suggestions to improve the paper.

References

Aamodt, A., Plaza, E., 1994. Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun. 7 (1), 39–59.
Bennis, S., Berrada, F., Kang, N., 1997. Improving single-variable and multivariable techniques for estimating missing hydrological data. J. Hydrol. 191 (1), 87–105.
Berndt, D.J., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. In: KDD Workshop'94, pp. 359–370.
Beven, K., 2015. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrol. Sci. J. http://dx.doi.org/10.1080/02626667.2015.1031761.

Beven, K., Smith, P., 2014. Concepts of information content and likelihood in parameter calibration for hydrological simulation models. J. Hydrol. Eng. 20.
Beven, K., Westerberg, I., 2011. On red herrings and real herrings: disinformation and information in hydrological inference. Hydrol. Process. 25, 1676–1680.
Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J., 2000. Lof: identifying density-based local outliers. In: ACM Sigmod Record, vol. 29. ACM, pp. 93–104.
Croke, B., 2006. A technique for deriving an average event unit hydrograph from streamflow-only data for ephemeral quick-flow-dominant carchments. Adv. Water Resour. 29, 493–502.
Didry, Y., Parisot, O., Tamisier, T., 2015. Engineering data intensive applications with cadral. In: Cooperative Design, Visualization, and Engineering: 12th International Conference, CDVE 2015, Mallorca, Spain, September 20-23, 2015. Proceedings. Springer.
Domingos, P., 2012. A few useful things to know about machine learning. Commun. ACM 55 (10), 78–87.
Enders, C.K., 2010. Applied Missing Data Analysis. Guilford Publications.
Féraud, R., Clérot, F., 2002. A methodology to explain neural network classification. Neural Netw. Off. J. Int. Neural Netw. Soc. 15 (2), 237–246.
Gao, C., Yao, M., Wang, Y., Zhai, J., Buda, S., Fischer, T., Zeng, X., Wang, W., 2015. Hydrological model comparison and assessment: criteria from catchment scales and temporal resolution. Hydrol. Sci. J. http://dx.doi.org/10.1080/02626667.2015.1057141.
Gilbert, D., 2002. The Jfreechart Class Library. Developer Guide. Object Refinery 7.
Griffioen, J., Vermooten, S., Kukuric, N., Vasak, S., Demuth, S., Gustard, A., Planos, E., Seatena, F., Servat, E., et al., 2006. A european case-based reasoning tool to inter-compare hydrological information on water stress among drainage subbasins. Clim. Var. Change Hydrol. Impacts 91–96.
Gyau-Boakye, P., Schultz, G., 1994. Filling gaps in runoff time series in West Africa. Hydrol. Sci. J. 39 (6), 621–636.
Harmal, R., Smith, P., Migliaccio, K., Chaubex, I., Douglas-Mankin, K., Benham, B., Shukla, S.,R.,M.-C., Robson, B., 2014. Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use: a review and recommendations. Environ. Model. Softw. 57, 40–51.
Harvey, C.L., Dixon, H., Hannaford, J., 2010. Developing best practice for infilling daily river flow data. Role Hydrol. Manag. Consequences a Chang. Glob. Environ. 816–823.
Harvey, C.L., Dixon, H., Hannaford, J., 2012. An appraisal of the performance of data-infilling methods for application to daily mean river flow records in the UK. Hydrol. Res. 43 (5), 618–636.
Johnston, C.A., 1999. Development and Evaluation of Infilling Methods for Missing Hydrologic and Chemical Watershed Monitoring Data. Ph.D. thesis. Virginia Polytechnic Institute and State University.
Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., Kolehmainen, M., 2004. Methods for imputation of missing values in air quality data sets. Atmos. Environ. 38 (18), 2895–2907.
Kirchner, J., 2009. Catchments as simple dynamical systems: catchment characterization, rainfall-runoff modeling, and doing hydrology backward. Water Resour. Res. 45.
Kotsiantis, S., 2013. Decision trees: a recent overview. Artif. Intell. Rev. 39 (4), 261–283.
Krause, P., Boyle, D., Bsw, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 78–97.
Kretzschmar, A., Tych, W., Chappell, N., 2014. Reversing hydrology: estimation of sub-hourly rainfall time-series from streamflow. Environ. Model. Softw. 60, 290–301.
Legates, D., McCabe Jr., G., 1999. Evaluating the use of "goodness-of-fit" measures in hyfrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241.
Maimon, O.Z., Rokach, L., 2005. Data Mining and Knowledge Discovery Handbook, vol. 1. Springer.
Marwala, T., Global, I., 2009. Computational Intelligence for Missing Data Imputation, Estimation and Management: Knowledge Optimization Techniques. Information Science Reference Herhsey, USA.
Matthies, M., Giupponi, C., Ostendorf, B., 2007. Environmental decision support systems: current issues, methods and tools. Environ. Model. Softw. 22 (2), 123–127.
Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007a. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.
Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., Veith, T., 2007b. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. Asabe 50 (3), 885–900.
Mwale, F., Adeloye, A., Rustum, R., 2012. Infilling of missing rainfall and streamflow data in the shire river basin, malawi—a self organizing map approach. Phys. Chem. Earth Parts A/B/C 50, 34–43.
Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part i a discussion of principles. J. Hydrol. 10 (3), 282–290.
Ouyang, R., Ren, L., Cheng, W., Zhou, C., 2010. Similarity search and pattern discovery in hydrological time series data mining. Hydrol. Process. 24 (9), 1198–1210.
Pigott, T.D., 2001. A review of methods for missing data. Educ. Res. Eval. 7 (4), 353–383.
Quinlan, J.R., 1992. Learning with continuous classes. In: Proceedings of the 5th Australian Joint Conference on Artificial Intelligence. Vol. 92. Singapore, pp. 343–348.
Salas, J.D., 1980. Applied Modeling of Hydrologic Time Series. Water Resources Publication.
Sattari, M.T., Pal, M., Apaydin, H., Ozturk, F., 2013. M5 model tree application in daily river flow forecasting in sohu stream, Turkey. Water Resour. 40 (3), 233–242.
Serban, F., Vanschoren, J., Kietz, J.-U., Bernstein, A., 2013. A survey of intelligent assistants for data analysis. ACM Comput. Surv. (CSUR) 45 (3), 31.
Tfwala, S.S., Wang, Y.-M., Lin, Y.-C., 2013. Prediction of missing flow records using multilayer perceptron and coactive neurofuzzy inference system. Sci. World J. 2013. Article ID 584516, 7 pages, http://dx.doi.org/10.1155/2013/584516.
Van Buuren, S., 2012. Flexible Imputation of Missing Data. Chapman & Hall/CRC. Interdisciplinary Statistics.
Van Hulse, J., Khoshgoftaar, T.M., 2008. A comprehensive empirical evaluation of missing value imputation in noisy software measurement data. J. Syst. Softw. 81 (5), 691–708.
Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Teechniques. Elsevier.
Wrede, S., Fenicia, F., Martnez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H., Uhlenbrook, S., Kavetski, D., Pfister, L., 2015. Towards more systematic perceptual model development: a case study using 3 luxembourgish catchments. Hydrol. Process. 29 (12), 2731–2750.