High-quality observation of surface imperviousness for urban runoff modelling using UAV imagery

Piotr Tokarczyk[1], Joao Paulo Leitao[2], Jörg Rieckermann[2], Konrad Schindler[1], and Frank Blumensaat[2,3]

[1]Institute of Geodesy and Photogrammetry, ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zürich
[2]Swiss Federal Institute of Aquatic Science and Technology, Eawag, Überlandstrasse 133, 8600 Dübendorf
[3]Institute of Environmental Engineering, Chair of Urban Water Systems, ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zürich

Correspondence to: Piotr Tokarczyk (piotr.tokarczyk@geod.baug.ethz.ch)

Abstract. Modelling rainfall-runoff in urban areas is increasingly applied to support flood risk assessment particularly against the background of a changing climate and an increasing urbanization. These models typically rely on high-quality data for rainfall and surface characteristics of the catchment area.

5   While recent research in urban drainage has been focusing on providing spatially detailed rainfall data, the technological advances in remote sensing that ease the acquisition of detailed land-use information are less prominently discussed within the community. The relevance of such methods increase as in many parts of the globe, accurate land-use information is generally lacking, because detailed image data is often unavailable. Modern unmanned aerial vehicles (UAVs) allow acquiring

10   high-resolution images on a local level at comparably lower cost, performing on-demand repetitive measurements, and obtaining a degree of detail tailored for the purpose of the study.

In this study, we investigate for the first time the possibility to derive high-resolution imperviousness maps for urban areas from UAV imagery and to use this information as input for urban drainage models. To do so, an automatic processing pipeline with a modern classification method

15   is proposed and evaluated in a state-of-the-art urban drainage modelling exercise. In a real-life case study (Lucerne, Switzerland) we compare imperviousness maps generated using a fixed-wing consumer micro-UAV and standard large-format aerial images acquired by the Swiss national mapping agency (*swisstopo*). After assessing their overall accuracy, we perform an end-to-end comparison, in which they are used as an input for an urban drainage model. Then, we evaluate the influence which

20   different image data sources and their processing methods have on hydrological and hydraulic model performance. We analyze the surface runoff of the 307 individual subcatchments regarding relevant

1

attributes, such as peak runoff and volume. Finally, we evaluate the model's channel flow prediction performance through a cross-comparison with reference flow measured at the catchment outlet.

We show that imperviousness maps generated from UAV images processed with modern classification methods achieve an accuracy comparable to standard, off-the-shelf aerial imagery. In the examined case study, we find that the different imperviousness maps only have a limited influence on predicted surface runoff and pipe flows, when traditional workflows are used. We expect that they have a substantial influence, when more detailed modelling approaches are employed to characterize land-use and to predict surface runoff. We conclude that UAV imagery represents a valuable alternative data source for urban drainage model applications due to the possibility to flexibly acquire up-to-date aerial images at a quality compared with off-the-shelf image products, and a competitive price at the same time. We believe that in the future, urban drainage models representing a higher degree of spatial detail will fully benefit from the strengths of UAV imagery.

1 Introduction

In the last century we have witnessed increased migration of people from rural areas to cities. Today, a majority of the human population live in cities and this number is estimated to grow constantly, and reach a level of 60% (UN, 2013). The process of a rapid urbanization called on developing an infrastructure capable to cope with a constantly increasing number of its users. Accordingly, ensuring water supply for the people is important, but due to the increased hydrological extremes induced by climate change (Hirabayashi et al., 2013; Hall et al., 2014; Rojas et al., 2013), being able to safely direct stormwater away from populated areas, in order to avoid flooding, is a not least challenging task. It requires predicting the hydraulic behavior of the given drainage infrastructure using reliable hydrological models (Arrighi et al., 2013). Those models, apart from detailed rainfall information, call for surface characteristics such as imperviousness.

Impervious surfaces reduce the infiltration of water into the soil. They can be directly related to a level of urbanization (Stankowski, 1972), because in urban environments, impervious surfaces dominate (*e.g.* rooftops or roads). Monitoring of imperviousness level is substantial for it as it directly impacts many environmental processes. An increasing percentage of impervious surfaces increases surface runoff volume and peak discharge, and decreases soil moisture compensation and groundwater recharge. Moreover, increased peak runoff volumes together with an inefficient drainage network can not only cause urban floods, but also lead to an increased hydraulic stress and increasing the risk of loading waterbeds with sediments, and its associated constituents (*e.g.* nutrients, contaminants and micro-pollutants).

An important step towards automation of the processes applied to map impervious areas was made as a consequence of remote sensing sensors and classification techniques development (for a detailed review of remote sensing methods used to map imperviousness, please refer to the Sup-

plement). In general, most of the studies on extraction of impervious surfaces from remote sensing data focused on satellite images. During the last decade, a rapid improvement of imaging sensors gave the end-user an access to very high spatial resolution (VHR) imagery[1]. Satellite sensors like

60 Ikonos (Chormanski et al., 2008) and QuickBird (Zhou and Wang, 2008) or VHR aerial images (Fankhauser, 1999; Nielsen et al., 2011) were quickly adopted for impervious surfaces mapping. Some studies suggest using highly accurate methods to quantify landscape changes (land-use and land-cover) using multi-sensor approaches (Forzieri et al., 2012b, a). In the context of urban hydrology Ravagnani et al. (2009) attempted to use impervious surfaces extracted from VHR satellite and

65 aerial imagery as an input to urban drainage model, but they did not analyze pipe flow predictions, focusing only on surface runoff component. However, modern urban drainage modelling methods call for up-to-date and detailed input data, which could also be acquired in an efficient way. Even though VHR satellite images able to acquire fine-grained image information (WorldView-3 satellite can achieve up to 0.31m GSD in panchromatic channel) and have short revisit periods, are still

70 expensive and vulnerable to cloud cover. VHR aerial imagery on the other hand, although being able to acquire very detailed imagery, is usually being updated at most once a year, but usually every third year (swisstopo, 2010). Recently, imaging platforms based on UAVs became very popular, finding their application in the fields of photogrammetry, archeology or agriculture (Sauerbier and Eisenbeiß, 2010; Eisenbeiß, 2009; Zhang and Kovacs, 2012). More recently, Leitão et al. (2015) in-

75 vestigated the quality of digital elevation models (DEMs) generated using UAV imagery from urban drainage modellieng applications. In the study the authors show that the quality of UAV DEMs is comparable to that of conventional, off-the-shelf height datasets. However, to our best knowledge no studies exist, that used UAV-based imagery to extract imperviousness information, and to use it in the field of urban drainage modelling. In comparison to a standard, off-the-shelf satellite or aerial re-

80 mote sensing imagery, UAVs demonstrate greater flexibility and are more efficient in terms of money and time. Yet, the classification of UAV VHR imagery, particularly in urban areas, is challenging, because in this level of detail, many small objects appear, and fine-grained texture details of larger objects emerge. Thus, describing an object class using only single raw pixel values is insufficient. Accurate classification needs additional image features, which would characterize the contextual in-

85 formation by describing object's local neighborhood. The value of such approach in classification of surface imperviousness has already been acknowledged (Moser et al., 2013). However, what is highly relevant, but currently unclear, is how to best exploit the rich information, i.e. the unprecedented level of detail and flexibility to acquire problem-specific images. And, whether it is feasible to use imagery acquired with UAVs for urban drainage modelling.

90 Specifically, we present three key aspects:

1. we evaluate whether land-use data based on UAV imagery can be used to assess the performance of urban drainage systems,

---

[1]We refer to a VHR image when sensor's ground sampling distance (GSD) is lower than 1m

2. we propose a unique workflow based on a randomized quasi-exhaustive (RQE) feature bank and a boosting classifier[2]. The RQE feature bank consists of a multitude of multi-scale textural features describing both, spectral and height information (Tokarczyk et al., 2015). The boosting classifier lends itself to the task to only choose the optimal features during training (for details see below), and

3. we perform end-to-end comparison of land-use against high-quality sewer pipe flow data. Although important to correctly interpret the results, this is not routinely done in remote sensing literature.

The key idea of our study was not to solely base the assessment of the usefulness of UAV images for urban drainage applications on the performance of the classifiers. Thus, we demonstrate the usefulness of our approach by means of a case study in a small urban area in Lucerne, Switzerland in two steps (see also Figure 1): first, we compare the UAV data with standard airborne imagery using a maximum likelihood (ML) classifier and the RQE method on both image sources (1). Second, we use a hydrodynamic model to show the consequences of different land-use information on urban drainage performance indicators, here surface runoff (2) and in-sewer pipe flow (3).
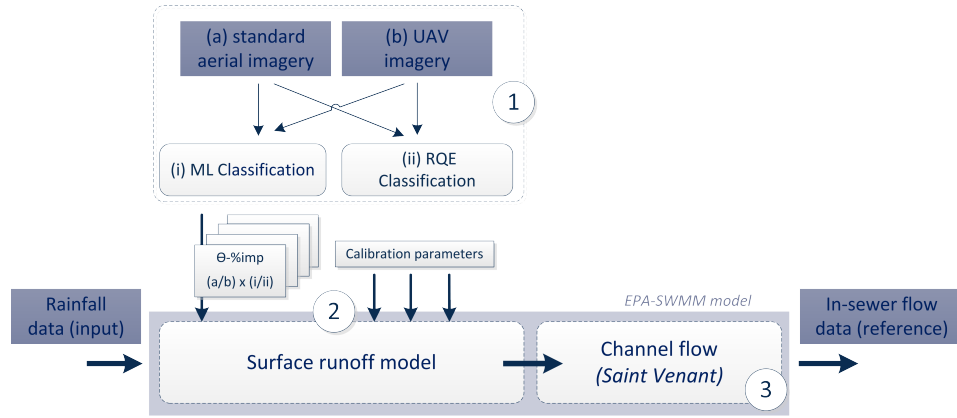

Figure 1. Overall analysis approach (⊖-%imp: model parameter "degree of imperviousness"; ML: Maximum Likelihood; RQE: boosting with randomized quasi-exhaustive feature bank).

The remainder of the paper is structured as follows: first we present general approach and the case study catchment with related material, such as the hydrodynamic rainfall-runoff model, rainfall and runoff observations, and remote sensing data. Then we describe the applied methods, land-use classification, surface runoff and in-sewer flow modelling, as well as the suggested performance criteria. Finally we present results and discuss the potential and limitations of using UAV images in urban hydrology.

---

[2]Boosting classifier used with conjunction with RQE features will be referred to as "RQE method" in this paper

2   Materials and methods

2.1   Case study and datasets

2.1.1   Case study

For our case study we used a residential area, called Wartegg catchment, in the city of Lucerne, Switzerland (see Figure 2). The catchment covers about 77 ha and is home for 6900 residents. It is typical for many suburban areas in Switzerland: high- to moderate-density population, scattered single- to two-story housing embedded in a hilly landscape, including the typical public infrastructure, such as shopping centers and sports grounds.

Storm- and wastewater is drained by separate and combined sewers (see Figure 2) with a total length of 11.2 km. An overflow structure connected to a small storage basin is installed to avoid hydraulic overload in case of heavy rainfall. Excess combined sewage is directly discharged to the lake, the carry-on flow travels by gravity to the wastewater treatment works. Three small creeks, to some extent culverted, cross the catchment and are interlinked with the storm water network.
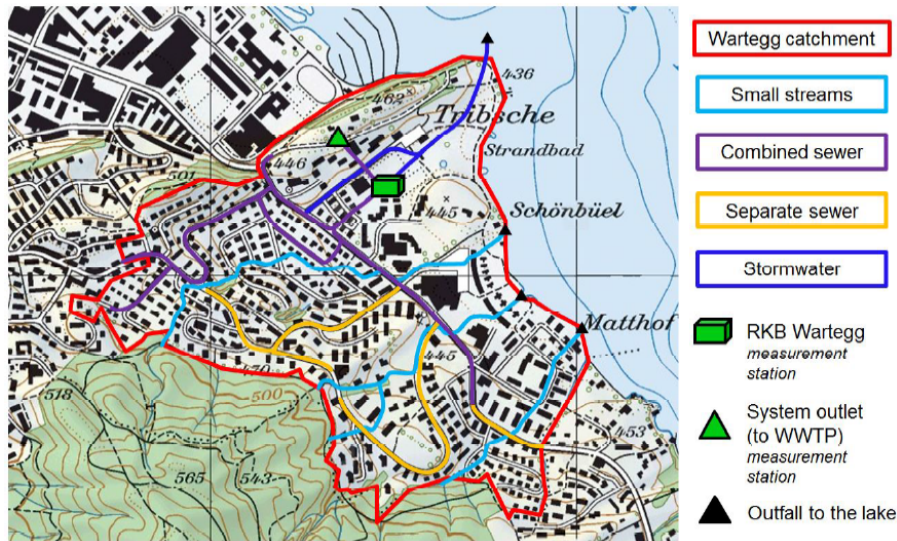

Figure 2. Case study catchment area situated in Lucerne.

2.1.2   Remote sensing datasets

Image data

In this study we used two image datasets. The first image data was acquired by *swisstopo*[3] in June 2013. It is a part of an aerial orthophoto mosaic (RGB channels) with a GSD of 0.0625 m, and consists of images acquired during leaves-on conditions. Although this dataset was acquired on-demand

---

[3]In this paper "ortho" and "orthophoto" terms will be used interchangeably with *swisstopo* imagery.

(standard *swisstopo* orthophotos have a GSD of 0.25 m), images acquired by *swisstopo* are publicly available, and this data source is, to our best knowledge, the standard for hydrological applications in Switzerland. Because *swisstopo* offers off-the-shelf image products, which are already orthorectified

135 and georeferenced, one can avoid costly and time consuming pre-processing of raw image data. On the other hand, image acquisitions are made at most once a year, usually every third year, and try to alternate between leaves-on and leaves-off periods (swisstopo, 2010). Thus, it might happen that one is not able to obtain up-to-date results.

The second dataset was acquired with a Canon IXUS 127 HS digital consumer camera with 16

140 Mpix sensor, mounted on a fixed-wing micro-UAV platform (Sensefly eBee, see Sect. A2 in the Supplement for details). The flight was performed during leaves-off conditions in March 2014. The custom processing software, which is shipped together with the UAV (*cf.* http://www.senseFly.com, based on the Pix4D technology, *cf.* http://pix4d.com/products/) was used to process the images. It is designed for use by non-experts and is highly automated, user interaction is limited to selecting

145 input images, entering flight parameters (camera details and GPS/INS data) and measuring ground control points (GCPs). Orthophotos (RGB channels) generated from the acquired images have a GSD of 0.10 m. In the case of a small catchment, as in our study, a main advantage of UAVs, when compared to manned aircraft with large-format mapping cameras, lies in their flexibility in terms of deployment, and in their low cost. Conducting a standard photogrammetric flight campaign typically

150 requires days of preparation and is more dependent on to weather conditions. Note though, micro-UAVs are at present not suitable for large-area mapping, because of their low speed and limited battery capacity.

Prior to the classification, both datasets were downsampled to a GSD of 0.25 m in order to make the evaluation comparable to standard *swisstopo* imagery available on the market. Furthermore, this

155 step reduces the time needed for training the classifier.

Height model

In this study we used two different height models: *(i)* a DTM provided by swisstopo (swisstopo, 2014) was used to classify the swisstopo dataset and to derive the catchment slope for the urban

160 drainage model. This model features a grid size of 2 m, for the land-use classification it was upsampled to the resolution of corresponding image dataset; *(ii)* a nDSM[4], created by subtracting a DSM extracted using dense image matching from a DTM provided by swisstopo, was used to classify the UAV dataset.

---

[4]A digital terrain model (DTM) represents the bare ground surface; a digital surface model (DSM) represents the surface visible from the top, including buildings, trees etc; the normalized digital surface model (nDSM) is obtained by subtracting the DTM from the DSM and shows the relative height of non-ground objects over the ground.

2.1.3 Rainfall

165   Precipitation data was collected from a meteorological station located $2\ km$ away from the Wartegg catchment area, operated by the Swiss Meteorological Institute (MeteoSwiss). Recordings were taken in a 10 min interval using a tipping bucket rain gauge with a precision of $0.1\ mm$. Hourly precipitation was checked following the quality assurance criteria of MeteoSwiss. Additional quality checks were carried out to ensure that the 10 min data are reliable. Spatial rainfall variability

170   was not considered in the study due to the short distance between the meteorological station and the study area.

2.1.4 Sewer flow reference data

Two flow datasets were obtained from in-sewer flow monitoring located at the outlet of the sub-catchment (see Figure 2). Over a period of four months (17 July 2014 to 18 November 2014) the

175   in-sewer flow was continuously monitored with two different sensors, *(i)* an acoustic Doppler flow sensor (Sigma submerged AV sensor, HACH) – 1 minute monitoring frequency and *(ii)* a digital Doppler radar velocity sensor, along with ultrasonic level-sensing (FLO-DAR, Marsh Mc Birney) - 15 minute monitoring frequency, to provide redundant flow rate information. Correlation analysis between the two reference signals show a high agreement and confirm the solid quality of data.

180 ### 2.1.5 Urban drainage model

Urban drainage models are tools to analyze the hydraulic behaviour of urban drainage systems, and to support risk analysis of urban flooding and receiving water pollution. Typically, these models include two main computing modules: the surface runoff (hydrological) and the in-sewer flow (hydraulic) model. The hydrological model estimates the time and space distribution of the direct runoff

185   under consideration of initial precipitation losses (evaporation, wetting losses) and soil infiltration for pervious areas. The resulting runoff is then used as input for the hydraulic model to simulate the pipe flow in the sewer network.

    In the present study we use the freely available Stormwater Management Model released and constantly developed by the U.S. Environmental Protection Agency (SWMM, Release 5.1.006; (US-

190   EPA, 2010)). SWMM is a widely used and well-accepted state-of-the-art 1-D dynamic rainfall-runoff model. We deliberately chose SWMM despite its limitations (lumped surface runoff model concept) as it represents a widely used state-of-the-art application in urban drainage modelling, and we wanted to keep the modelling use case as simple as possible.

    The description of the surface runoff is based on the MANNING approach, a simplifying, con-

195   ceptual formulation of transport phenomena in the catchment assuming that the surface runoff starts after the rainfall volume has exceeded a representative value of the initial losses in the catchment. Rainfall losses are adjusted throughout the rainfall event according to the changes occurring in the in-

filtration process (pervious part of catchment surface) which is a function of the soil water saturation level. Impervious surfaces are those where no infiltration occurs; the catchment's imperviousness degree and its spatial distribution are then expected to have a great impact on surface runoff and urban drainage system modelling results. Flow routing through a system of sewer pipes, storage basins and regulating devices is accomplished by solving the Saint Venant flow equations, whereas here we applied a type of diffusive wave approximation which neglects inertial terms from the momentum equation when flow becomes supercritical.

2.2 Methodology

2.2.1 Classification

Generally, supervised classification consists of three main steps: *(i)* extraction of the features from raw input image, *(ii)* training the classifier using a small, manually annotated training set (not necessarily from the same image), and *(iii)* classification of all pixels in the area of interest, using the classifier trained in the previous step. In the following we describe two different types of supervised classifiers: *(i)* Gaussian maximum likelihood, and *(ii)* boosting.

Maximum likelihood

The maximum likelihood (ML) classifier is a popular classification method in the field of urban hydrology. It is a simple generative model which assumes that the image features within each target class follow a normal distribution. Under this assumption, each of the target classes can be described by its mean vector and covariance matrix. Given this information one can directly compute the statistical probability of particular pixel belonging to one of the target classes. A serious limitation of ML is that it is not well suited for high-dimensional data. Due to the "curse of dimensionality" (Hughes, 1968) its performance degrades typically beyond a dozen or so feature dimensions. For imagery with a medium spatial resolution imagery, where objects are usually spectrally consistent[5], it might be enough to construct image features consisting only of single raw pixel values. However, the variability of the pixel values within an object class grows with the spatial resolution of the image. For example when roof consists of many pixels and substructures such as planted areas or roof gardens become visible. Therefore one should no longer rely on single pixel values, but has to consider contextual information and, for example, construct features that exploit the neighborhood of a pixel (*e.g.* textural features). Such features expand the dimensionality of data, making generative classifiers inefficient. Here we classified two image datasets using a maximum likelihood classifier implemented in ArcGIS software (ESRI, 2013). As often done in conjunction with the ML method, we use only the single raw pixel values as features.

---

[5]Meaning that they consist of pixels of similar values

Boosting

As an alternative to ML we propose a multiclass extension (Benbouzid et al., 2012) of adaptive boosting (AdaBoost, Freund and Schapire (1995)). Unlike ML, boosting methods (and related discriminative classifiers) are better suited for very high-dimensional feature spaces, as they do not attempt to model the input data distribution. Boosting greedily learns an additive combination of many simple classifiers (in our case shallow decision trees). A useful property of the method is that it performs explicit feature selection as part of the classifier training. Thanks to this, we sidestep manual feature selection. Moreover, at test time only the selected features need to be computed, which significantly reduces the computational effort. Here, we classified the images using randomized quasi-exhaustive (RQE) feature bank (Tokarczyk et al., 2015), which are able to capture multiscale texture properties in a pixel's neighborhood.

Performance assessment of classification

To assess the performance of the two classifiers used in this study, we have manually labeled a subset (5 ha) of each of image datasets (see Figure 3). Hence, we were able to report the classification accuracy for all pixels in an extended area, which in our view is a lot more reliable than sparse, point-wise accuracy assessment. We selected either three (rooftops/streets/vegetation) or two (impervious/pervious) target classes, where in the two-categories case, "impervious" class is an aggregation of "rooftops" and "streets" classes. For the subsequent hydrological analysis, only the two-class maps were used.

Both classifiers were trained using randomly selected subsets of pixels (1%, 2% or 5%, which correspond roughly to 7000, 14000 and 36000 pixels). Thereby we can evaluate how the size of the training data has an influence on the overall classification accuracy. If satisfactory results can be obtained, then a lower number of training samples is preferable, since it reduces the training time and saves annotation effort. Similarly to experiments carried out in Tokarczyk et al. (2015), we trained the boosting classifier using decision trees with eight leaf nodes, and set the number of boosting rounds to 500. As performance metric for the classification we used the overall accuracy (OA), *i.e.* the fraction of correctly classified pixels.

2.2.2 Assessing the importance of input data on surface runoff

To assess the influence of input data accuracy on the surface runoff, we predicted the surface runoff for a rain event of moderate intensity (total volume of 29.7 mm; peak rainfall intensity of 2.9 Ls^{-1}). Then, we analysed the runoff of the 307 individual subcatchments regarding the following attributes: *(i)* peak flow (Q_{peak}) and *(ii)* Volume of the peak (V_{peak}). As it is very challenging to directly measure surface runoff that can be compared with model predictions, we first performed an exploratory analysis of the major influence factors. Second, we investigated interactions between the data source and processing method by means of a regression analysis (see Sect. A3 in the Supplement for de-
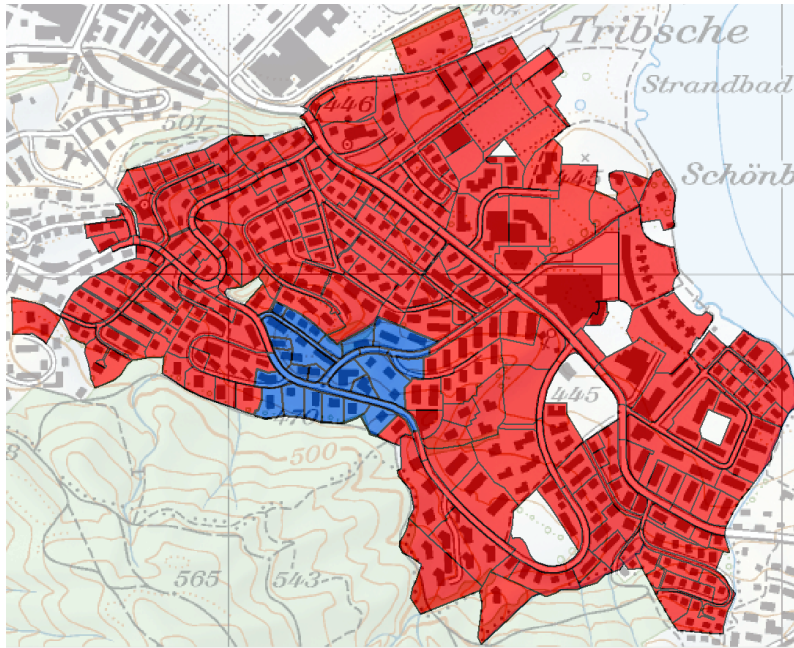
9

Figure 3. Wartegg area containing 307 subcatchments (red polygons including blue polygons) overlaid on a topographic map. The performance of classifiers was assessed on a subset depicted in blue.

tails).

270 **Performance assessment**

Exploratory data analysis of surface and surface runoff characteristics

To summarize the important characteristics of the surface runoff, we visualized important aspects using boxplots and scatterplots (see Figure 6). Main research questions were:

- Which differences in imperviousness (Δ_{Imp}) result for each subcatchment: *(i)* for the two data
275   sources and *(ii)* for the two classification methods?

- Does the image source have a substantial influence on the predictions of surface runoff from a subcatchment? How does this depend on the processing method?

Regression analysis of surface runoff characteristics

To answer the second question, we constructed four regression models with indicator variables (Mont-
280 gomery et al., 2012). This makes it possible to consider the individual effects of the data and the processing method. In addition, a model with an interaction term, unlike an additive model, could add a further adjustment for the "interaction" between the data source and the classification method. Specifically, we would like to explore whether the relationship between the image source and the imperviousness in the subcatchments, and their surface runoff characteristics is different for each

10

285 classifier. The model for a dependent variable y is:

$$y_i = \beta_0 + \beta_1 I_i^{Data} + \beta_2 I_i^{Method} + \beta_3 I_i^{Data*Method} + \epsilon_i \tag{1}$$

where y_i is the i-th observation of the dependent variable, I_i^{Data} an indicator variable which is 1 if y_i was computed from UAV images (UAV) and 0 from orthophotos, I_i^{Method} is an indicator variable which is 1 if y_i was computed with the RQE method and 0 for the ML classifier (ML). $\beta_0...\beta_3$ are

290 the parameters to be estimated and ϵ_i is a random error term. If ϵ_i is normally and independently distributed, $i.e.$, $\epsilon_i \sim N(0, \sigma^2)$, this model is equivalent to a classical least square regression or to a three-way analysis of variance model with treatment contrasts (Montgomery and Runger, 2007).

The imperviousness is bounded between 0 and 1, whereas a linear model could easily predict values beyond this range, which is not admissible. To have a more plausible model, we therefore

295 used a logit-transformation on the imperviousness ($\%imp$):

$$z = 2 * arctanh(2 * Imp - 1) \tag{2}$$

In addition, we analyze the results of this regression analysis on a qualitative basis only. With more correct and more complex models, which better represent the underlying process that generated the data, p-values (see Tables 5, 6 and 7 in supplementary material) would be tend to be larger. Here,

300 however, we are not interested in the magnitude or statistical significance of the individual effect, but we just would like to see whether they are very different or not.

2.2.3 Prediction of pipe flows

To assess the model's capability to predict the resulting in-sewer flow, we predicted stormwater flows at the catchment outlet for 36 independent rain events of different intensity and duration (see below)

305 and compared them with flow data measurements (see Section 3.3). In particular, we compared measured and predicted volume of the total runoff as well as peak flows. Main driving questions for the analysis were:

- How do differences in imperviousness affect pipe flow predictions?

- To what extent may differences regarding input data, i.e. degree of imperviousness of sub-
310 catchment areas, be compensated by the model calibration procedure?

Model calibration

To address the latter question, we compared the results of the different model implementations prior and after calibration. For the calibration/validation procedure we split the reference dataset in a calibration (July to September 2014) and a validation period (September to November 2014). In

315 total, for both periods, 36 independent rain events of different intensity and duration were observed, which we consider sufficient to cover the inherent variability of rain events.

11

To analyse the effect of different input data and how this would be addressed by model calibration, we applied a genetically adaptive multi-objective calibration algorithm (AMALGAM, Vrugt and Robinson (2007)) to adjust the calibration parameters of the four implementations. The model input (two image data sources × two different classifiers) is used to derive the "$\%imp$"-parameter. In the optimization, four different calibration parameters were adjusted to match three objective functions:*(i)* Simulation Bias (SB) and Nash-Sutcliffe-Efficiency (NSE, Nash and Sutcliffe (1970)), *(ii)* total flow balance, and *(iii)* peak flow deviation - all with respect to the flow at the catchment outlet. The input parameter imperviousness "$\%imp$" is derived from orthophotos and not subject to calibration. The calibration parameters are:

- catchment width $[m]$,

- HORTON maximum infiltration rate $[mm\ d^{-1}]$,

- Decay constant for the HORTON curve $[d^{-1}]$, and

- Size of a virtual subcatchment $[ha]$, mimicking groundwater infiltration into the sewer pipe network.

Performance assessment: flow balance and flow dynamics

In a first step, we evaluated the match between modelled hydrographs and reference flow data using the SB and NSE. Both goodness-of-fit measures are well established in urban hydrology to cover deviations regarding the flow balance (bias) and flow dynamics (NSE). The Simulation Bias *B* is defined as follows:

$$B = \left(\overline{E} - \overline{M}\right)^2 \tag{3}$$

whereas $\overline{M}$ is the mean of measured (observed) values and $\overline{E}$ is the mean of estimated (simulated) values. The bias ranges from $-\infty$ until $+\infty$ with an optimum at 0. The Nash-Sutcliffe-Efficiency NSE is defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{N} |M_i - E_i|^2}{\sum_{i=1}^{N} |M_i - \overline{M}|^2} \tag{4}$$

whereas M_i is the measured (observed) and E_i is the simulated value at the time i, $\overline{M}$ is the mean of measured (observed) values, E is the mean of estimated (simulated) values, and N the number of paired data. NSE reaches 0 when the square of the differences between measured and estimated values is as large as the variability in the measured data. In case of negative NSE values the measured mean is a better predictor than the model.

To cover one of the key figures relevant for engineering urban drainage systems, we included an event-specific evaluation of peak flows in a second evaluation step. To this end, we extracted peak flows from observed and modelled hydrographs using an event filter that identifies independent rainfall-runoff events preceding dry weather period at least 6 hours.

3 Results

3.1 Classification

Table 1 presents per-pixel overall classification accuracy achieved using *(i)* two different datasets, *(ii)* two classification methods, and *(iii)* either two or three target classes. Figure 4 and 5 present visual classification results for a subset of each of the two datasets, together with a respective ground truth. We did not perform any pre- or post-processing of the data. Image pre-processing adds no information and typically does not help, except for physically meaningful reflectance calibration, which in our setting, was not feasible. Post-processing of the imperviousness map might improve overall accuracy, but carries the danger of introducing unwanted biases.

	UAV			Orthophoto		
Class. method / % of train data	1%	2%	5%	1%	2%	5%
Three classes						
ML	78.9	72.8	78.4	84.2	84.4	80.8
RQE	93.7	94.3	95.2	95.6	95.8	96.3
Two classes						
ML	87.7	81.6	84.3	90.9	90.8	88.4
RQE	95.5	95.6	96.2	96.6	97.0	97.4

Table 1. RQE vs. ML method: Overall classification accuracies (in %). Boosting with RQE features after 500 iterations. Maximum likelihood classifier was trained with features consisting of single raw pixel intensities (all spectral channels).
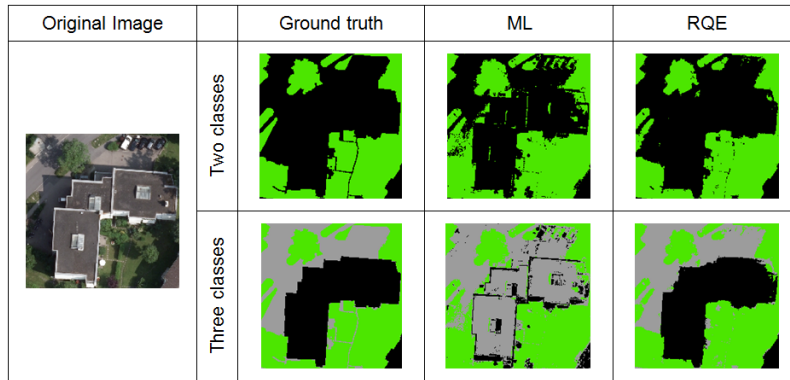

Figure 4. Cutouts of the swisstopo image: original image, manually labeled ground truth, and classification results using ML and RQE (two and three classes). In a case of two classes impervious surfaces are black and pervious are green. In a case of three classes rooftops are black, streets/sidewalks are grey and vegetation is green.
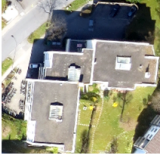
13
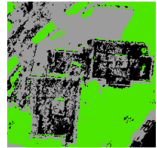
Figure 5. Cutouts of the UAV image: original image, manually labeled ground truth, and classification results using ML and RQE (two and three classes). In a case of two classes impervious surfaces are black and pervious are green. In a case of three classes rooftops are black, streets/sidewalks are grey and vegetation is green.

3.2 Prediction of surface runoff

360 **Exploratory analysis**

We used boxplots and scatterplots to investigate the effect of combining two data sources and two processing methods on *(i)* the imperviousness and the surface runoff characteristics, *(ii)* peak flows, and *(iii)* runoff volumes (see Figure 6).

- *Imperviousness (Imp)*: The boxplot shows that the overall distributions of imperviousness for
365   307 subcatchments do not differ much across the different image sources and classification methods. In general, the UAV images seem to produce slightly lower values of imperviousness than the orthophoto, although this effect might also be dominated by the set of UAV image which was processed by the ML classifier. Regarding the classification methods, the boosting classification method delivers slightly larger imperviousness values for both data sources than
370   the ML method.

- *Peak runoff (Peak)*: Like as for the imperviousness, the different image sources lead to very similar peak runoff values. In general, boosting leads to slightly higher peak flows, which also have a larger variance and slightly higher extreme values for a couple of subcatchments. Regarding the suitability of UAV images in rainfall-runoff modelling, there are no relevant
375   differences between the image sources.

- *Runoff volumes (Volume)*: The exploratory analysis effectively suggest the same patterns for the runoff volume as for the peak flows: boosting leads to larger runoff volumes and the resulting variability of the rainfall runoff from the 307 subcatchments is slightly larger than for the ML classifier. Also, the UAV data seem to be associated with smaller runoff volumes. This
380   is consistent, as this relates to the lower degree of imperviousness in the subcatchments.

14

In general, the relative differences between the different alternatives are very small, with average values of a few percent (see Figure 6). For the imperviousness, there are only a few subcatchments which show rather large differences. These are even less relevant for the peak runoff and runoff volumes.

385   Furthermore, the scatterplots of the different explanatory and dependent variables suggest that there is not a substantial difference between the image sources or classification approaches for the modelled surface runoff in the different subcatchments (see Figure 10 in supplementary material). For the boosting classifier, we observe a weak positive correlation with the degree of imperviousness (see Figure 11 in supplementary material), which means that catchments which are rather impervi-

390   ous (or pervious) based on the ML classifier tend to be even more impervious (or pervious) for the boosting classifier. However, this is difficult to identify by means of visual analysis and is better explored by an analysis of variance or regression analysis.
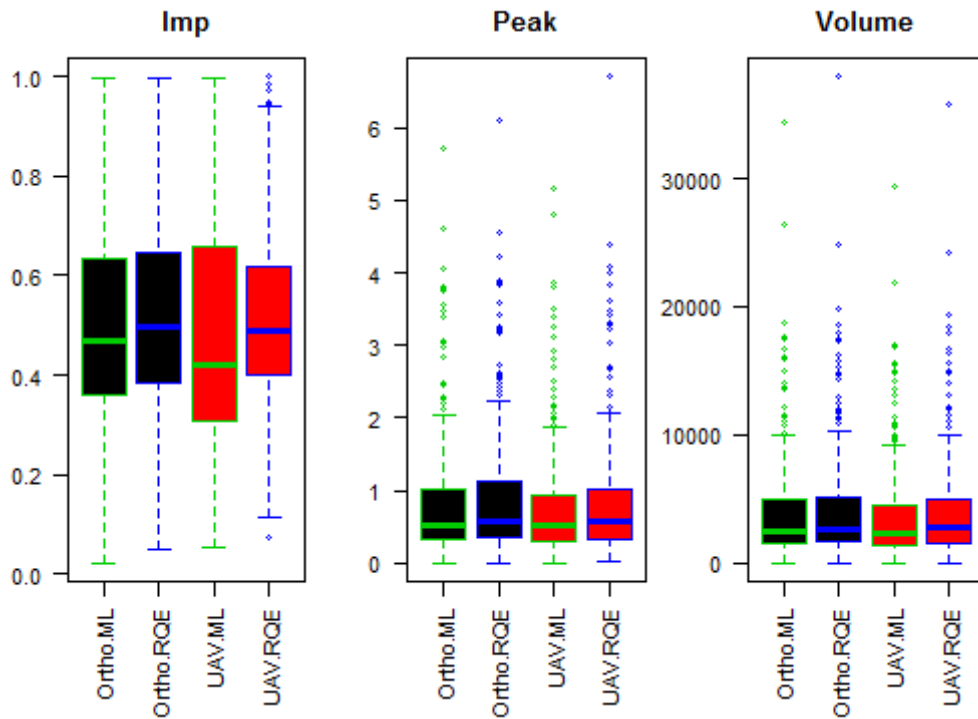

Figure 6. Boxplots of the imperviousness and surface runoff characteristics (Imp $[-]$, Peak $[Ls^{-1}]$, Volume $[m^3]$) for the 307 subcatchments for the four combinations of data sources and processing methods. Black= Ortho, Red= UAV, Green= ML, Blue = RQE.

Regression analysis

The results from the regression analysis are mainly the maximum likelihood estimates of the model parameters and an indicator of their importance (see Tables 5 and 6 in supplementary material).

For the *imperviousness*, as expected neither the image source nor the classifier is strongly correlated. The negative sign of the estimated slope parameter for the image source (β_1 = -0.16) suggests that UAV images generally go together with a lower imperviousness. In addition, the influence of the image source seems to be larger than that of the classification method (β_2 = 0.003), although the large p-values for all parameters suggest that it is not very likely that the observed values of imperviousness were to have occurred under the given statistical model. Therefore, there is virtually no evidence that there are interactions between the image source and the classifiers.

For the *peak runoff*, neither the image source nor the classifier are strongly correlated. The negative sign of the estimated slope parameter for the image source (β_1 = -0.6) suggests that UAV images correlate with a smaller peaks. Here, the influence of the image source seems to be equally important as the classification method (β_2 = -0.6), just with a different sign. Nevertheless the high p-values for all parameters again suggest that it is not very likely that the observed values of imperviousness were to have occurred under the given statistical model. Also, the interaction between the image sources and classifiers is not important.

For the runoff volume, the UAV data generally seem to be correlated with slightly lower runoff volumes (β_1 = -302), whereas the RQE method shows a positive correlation (β_2 = 298). Again, neither the two effects nor their interaction seem to be important.

In summary, the analysis suggests that surface runoffs predicted with SWMM are similar for the different data sources or classifiers. In addition, neither the imperviousness nor peaks nor volumes of the runoff are influenced by interactions between the image sources and the classification methods. As the data source and classifier alone do not represent the data generating process, the underlying statistical assumptions are not met and the numerical results should not be over-interpreted.

3.3 Prediction of in-sewer flow

The evaluation regarding sewer pipe flow is split into two parts: (1) model performance of uncalibrated implementations, and (2) calibrated implementations compared to reference data, *i.e.* flow measured at the outlet of the catchment.

(1) Focusing on the results prior to calibration, it becomes clear that uncalibrated models, among each other, differ particularly regarding the peak flow performance (see boxplot in Figure 7). This clearly corresponds to the findings of the surface runoff analysis (see Section 3.2) in which, for instance the implementation "UAV ML" with the lowest mean degree of imperviousness produces the lowest runoff peaks. The comparison with reference data through hydrological goodness-of-fit measures (see Table 2) underlines the moderate performance regarding flow dynamics (NSE), whereas already good agreement is achieved for the total flow balance (bias). The slightly improved

16

performance of the implementation of which the imperviousness is derived from UAV data classified with the ML method (UAV ML) probably occurs by chance.

	Prior to calibration SB [-] / NSE [-]	After calibration SB [-] / NSE [-]
Ortho ML	2.0 / 0.54	3.16E-5 / 0.72
Ortho RQE	2.0 / 0.52	0.007 / 0.71
UAV ML	0.3 / 0.62	0.1 / 0.75
UAV RQE	2.0 / 0.53	1.38 / 0.73

Table 2. Goodness-of-fit measures prior and after calibration (both quantified for the validation period).
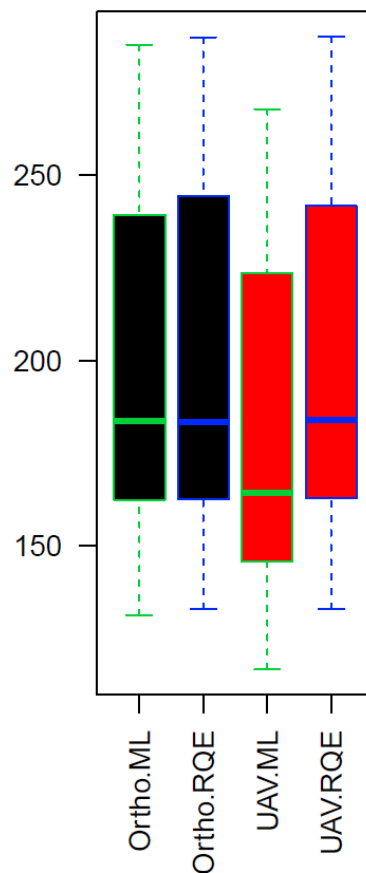

Figure 7. Evaluation of peak flows $[Ls^{-1}]$ for the 13 most intense rain events (prior calibration).

(2) Results from calibrated models (see Figure 8 and Table 2, right) show that conducting a detailed calibration, as expected, leads to an improved model performance (NSE increase, bias reduction) and interestingly smooths out the land-use differences among the four implementations. This is visible in Figure 8, where the hydrographs are practically the same. Even though the results from the

17

UAV ML implementation after calibration still shows slightly different results (see Figure 8, right), the differences in peak flow for the 13 most intense rain events are very similar (see Figure 9).
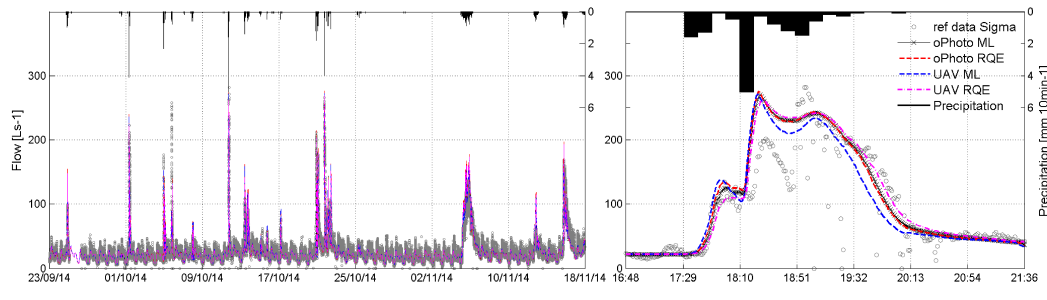

Figure 8. Observed reference and simulations (prior calibration) for the full validation period September to November 2014 (left) and a selected event on 11 October 2014 (right).
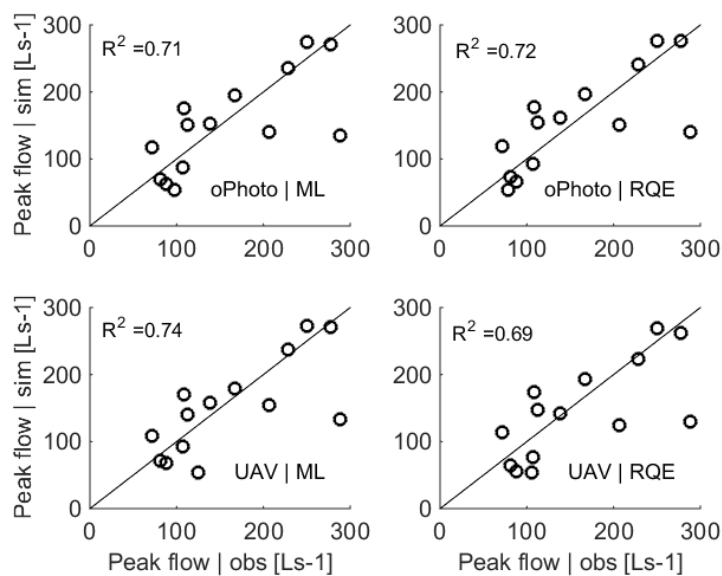

Figure 9. Evaluation of the peak flows for the 13 most intense rain events in the validation period (after calibration).

Interestingly, the very similar performance is achieved with very different parameter estimates (see Figure 15 in supplementary material). Particularly the parameter "width", "maximum infiltration rate" and "Decay K" (influencing the peak flow) vary substantially. Results show that the calibrated runoff model should be fairly robust against variations of the perviousness map, since these can be compensated by changing other, more uncertain parameters, *e.g.* by different parameter defining the infiltration into pervious surfaces.

18

4 Discussion

4.1 Classification

The choice of the classifier has a substantial impact on the overall classification accuracy. While boosting achieves accuracies between 93.7% and 96.2% for the UAV dataset and 95.6% to 97.4% for the *swisstopo* dataset, maximum likelihood yields results which are up to 20% worse. Further, it can be seen that the number of target classes strongly influences the results of the ML method. Classification with three target classes is up to 9% less accurate than with two. Moreover, the amount of data used to train the ML classifier gives inconclusive results. By increasing the number of training samples, overall accuracy should increase. However, in our case the training appears to be unstable, and the expected increase only materializes in a single case (see Table 1, orthophoto dataset, three classes). A possible explanation is that the class distribution is not unimodal, and thus not appropriately captured by the Gaussian model.

In contrast to the ML method, the boosting classifier behaves in a stable manner. Differences in overall accuracy do not exceed 2.5% per dataset. The changes in boosting performance with varying amounts of training data are negligible: 1% (7000 pixels) already yield satisfactory results, *i.e.* the effort for annotation as well as the training time remains low. The efficiency and robustness of boosting used together with features appropriate for VHR aerial imagery, makes this approach a good choice for the task. Also overall classification accuracy achieved using a boosting classifier together with UAV-based imagery shows that in terms of classification accuracy of impervious surfaces, this new imaging platform gives comparable results to the off-the-shelf aerial image products.

Moreover, our experiments show that at the level of surface runoff prediction, the differences between different imaging platforms and between different processing methods are small. Even though the classification accuracy between datasets and methods differs up to 20%, their influence on surface runoff characteristics lies within only few percent on average. We believe that one of the reasons is the spatial size of our subcatchments. Each of them consists hundreds of image pixels, but hydrological model disregards the spatial information and only uses aggregated values, i.e. the sum of all impervious pixels belonging to one subcatchment. A further observation is that the differences in classification accuracy are much larger for the three-class case. This is in line with conventional machine learning wisdom ("only predict what you need to know"), however we have not yet constructed an end-to-end study with the three-class result as an input.

4.2 Prediction of surface runoff

Exploratory analysis of surface runoff
While there are substantial differences when the images are compared pixel-by-pixel (see Figure 4 and 5), these are largely lost for the predicted surface runoff. In our view, this is again explained by the SWMM surface runoff model. It is a lumped model, which aggregates the pixels and thus

480 is smoothing out the differences, already on this small scale. This tendency will be even more pronounced for a higher degree of spatial aggregation, *e.g.* when modelling larger urban areas, where the subcatchments equipped with flow measurements will also be larger. Future experiments that investigate the continuous spatial downsampling of images may reveal when differences fully disappear.

485

Model structure as a bottleneck?

Obvious differences in the input data may be smoothed out due to the simplified, conceptual representation of the surface runoff in SWMM. We do expect different results for more detailed representation of land-use, e.g. with a separate "roof" land-use or modern pixel-based modelling approaches
490 for surface runoff. In the future, this might be even more important considering the increasing popularity of coupled 2D-overland/1D-channel flow models including more detailed overland-flow modelling using raster/pixel-based approaches (*cf.* Leandro et al. (2009)). Traditional models - as currently used in day-to-day engineering practice - will probably never be able to fully make use the amount of detail (pixel basis) provided by such aerial images.

495

High-resolution images provide added value in urban drainage

The effect on surface runoff and pipe hydraulics using spatially aggregating models (two land-use classes) may not be as immense. However, in future investigations, models that allow differentiating between three or more land-use classes should be further investigated. This may be particularly rele-
500 vant for pollutant load modelling, for which detail, accuracy and actuality of land-use characteristics are highly influential. Relevance of input data accuracy may even further increase due to the fact that obtaining adequate pollution load reference data is considered to be very difficult (*cf.* Dotto et al. (2014)).

   Also, other urban drainage tasks would greatly benefit from detailed land-use maps, e.g. precise
505 and justified stormwater fees due to exactly delineated types of impervious areas (*cf.* Figure 4 and 5). An improved feature (gully pots, sewer inlets, curbstone structures) identification is expected to further provide valuable input data for network generation approaches (e.g. as outlined in Blumensaat et al. (2012)) and the coupled 2-D surface runoff/1-D pipe flow model applications. For this, the RQE method seems to be most promising, although for the runoff analysis, a simpler method still
510 seems to produce robust results.

4.3 Pipe flow predictions

The results from the model calibration show that input data deviations are nearly fully compensated by the calibration procedure, involving an adaption of four different calibration parameter sets. The analysis of the final calibration parameter values however reveals that the best fit for each of the
515 implementations is achieved by differing parameter sets (see Figure 15 in supplementary material).

On the one hand side, this may indicate that, even though the full range of *a priori* defined parameter ranges is used during the auto-calibration procedure, for each implementation a different (local) optimum in the Pareto front is identified. On the other hand, it may underline that the given model structure is flexible enough to address different model inputs through different parameter settings. Here, it becomes clear that the compensation is achieved by adjusting parameters in a way that involves the risk that some parameters loose its physically based origin and turn into "conceptual handles". The discussion on this particular question is certainly interesting and would need further analyses, but it cannot be accomplished in this paper contribution as it would blur the main focus of the paper.

5 Conclusions

In this study we investigated the possibility to automatically generate high-resolution imperviousness maps for urban areas from imagery acquired with UAVs, and for the first time assessed the potential of UAVs for high-resolution hydrological applications compared with a standard large-format aerial orthophotos. We proposed an automatic processing pipeline with modern classification methods to extract accurate imperviousness maps from high resolution aerial images, and presented an end-to-end comparison, in which the maps obtained from different sources and processed with different classification methods were used as input for urban drainage models.

The first part of our analysis indicates that using a boosting classifier in conjunction with RQE features we were able to classify UAV imagery with an accuracy comparable to standard aerial orthophotos. The proposed classification method yields more stable results, when compared with those produced using the maximum likelihood method. This improvement is even more apparent when classifying three instead of two classes of land-use.

In the second part of our analysis we have demonstrated how model input data variations propagate in the course of the urban drainage modelling exercise, and how this is reflected in the surface runoff and sewer flow predictions. Results from uncalibrated model implementations actually show deviations in the predictions, which can be explained by input data variations. But still predictions are inaccurate. Conversely, after calibration the performance analysis shows that the calibration process attenuates variations in the input data, suggesting that model predictions are insensitive to these variations. However, the analysis of the resulting model parameter settings also reveals that apparent robustness is achieved by tweaking the parameter in a way which involves the risk of leaving valid parameter ranges.

Because model development and calibration in everyday practice is often based on less accurate information than used in our case study, it is important to underline reliable input data to reduce overall uncertainty in model predictions.

21

550   We note that the conclusions of the study are limited regarding *(i)* the small size of the case study catchment, *(ii)* the degree of detail in which the catchment has been described (more detail may show a more pronounced input error propagation, a more lumped description may absorb input deviations from the start), and *(iii)* the type of hydrological modelling concept used. Therefore we suggest conducting further research to evaluate the impact of the spatial scale, *i.e.* the degree of 555   spatial aggregation linked to the hydrological model approach (ensemble modelling). In the case study presented here we chose a traditional and widely used urban drainage model (EPA SWMM) to deliberately demonstrate the effect of new image sources and processing methods for standard engineering practice.

Still, we suggest using imperviousness maps consisting of three land-use classes as more differ-560   entiated input for a more detailed hydrological model, *i.e.* a pollution load model, which makes a better use of urban land-use differentiation. Because the proposed boosting classifier showed the largest accuracy gain for a three-class case, we strongly believe that introducing this additional information more clearly shows the potential of UAV datasets and advanced classification methods for more accurate urban drainage and pollution load modelling.

565 References

Arrighi, C., Brugioni, M., Castelli, F., Franceschini, S., and Mazzanti, B.: Urban micro-scale flood risk estimation with parsimonious hydraulic modelling and census data, Natural Hazards and Earth System Science, 13, 1375–1391, 2013.

Bauer, M. E., Loffelholz, B. C., and Wilson, B.: Estimating and mapping impervious surface area by regression
570　analysis of Landsat imagery, pp. 3–19, CRC Press, Taylor & Francis Group: Boca Raton, 2008.

Benbouzid, D., Busa-Fekete, R., Casagrande, N., Collin, F.-D., and Kégl, B.: MultiBoost: a multi-purpose boosting package, Journal of Machine Learning Research, 13, 549–553, 2012.

Blumensaat, F., Wolfram, M., and Krebs, P.: Sewer model development under minimum data requirements, Environmental Earth Sciences, 65, 1427–1437, 2012.

575　Boegh, E., Poulsen, R., Butts, M., Abrahamsen, P., Dellwik, E., Hansen, S., Hasager, C., Ibrom, A., Loerup, J.-K., Pilegaard, K., and Soegaard, H.: Remote sensing based evapotranspiration and runoff modeling of agricultural, forest and urban flux sites in Denmark: From field to macro-scale, Journal of Hydrology, 377, 300 – 316, 2009.

Cablk, M. E. and Minor, T. B.: Detecting and discriminating impervious cover with high-resolution IKONOS
580　data using principal component analysis and morphological operators, International Journal of Remote Sensing, 24, 4627–4645, 2003.

Carlson, T. N.: Analysis and Prediction of Surface Runoff in an Urbanizing Watershed Using Satellite Imagery, Journal of the American Water Resources Association, 40, 1087–1098, 2004.

Carlson, T. N. and Arthur, S. T.: The impact of land use - land cover changes due to urbanization on surface
585　microclimate and hydrology: a satellite perspective, Global and Planetary Change, 25, 49 – 65, 2000.

Chabaeva, A., Civco, D., and Hurd, J.: Assessment of Impervious Surface Estimation Techniques, Journal of Hydrologic Engineering, 14, 377–387, 2009.

Chormanski, J., Van de Voorde, T., De Roeck, T., Batelaan, O., and Canters, F.: Improving Distributed Runoff Prediction in Urbanized Catchments with Remote Sensing based Estimates of Impervious Surface Cover,
590　Sensors, 8, 910–932, 2008.

Civco, D. L., Hurd, J. D., Wilson, E. H., Arnold, C. L., and Prisloe, M. P.: Quantifying and describing urbanizing landscapes in the Northeast United States, Photogrammetric Engineering and Remote Sensing, 68, 1083–1090, 2002.

DeBusk, K., Hunt, B., and Sydorovych, O.: Watershed Retrofit and Management Evaluation for Urban
595　Stormwater Management Systems in North Carolina, pp. 3911–3920, 2010.

Dotto, C., Kleidorfer, M., Deletic, A., Rauch, W., and McCarthy, D.: Impacts of measured data uncertainty on urban stormwater models, Journal of Hydrology, 508, 28 – 42, 2014.

Dougherty, M., Dymond, E. L., Goetz, S. J., and Jantz, C. A.: Evaluation of impervious surface estimates in a rapidly urbanizing watershed, Photogrammetric Engineering and Remote Sensing, 2004.

600　Eisenbeiß, H.: UAV photogrammetry, Institut für Geodäsie und Photogrammetrie. Eidgenössische Technische Hochschule Zürich, 2009.

Elvidge, C. D., Tuttle, B. T., Sutton, P. C., Baugh, K. E., Howard, A. T., Milesi, C., Bhaduri, B., and Nemani, R.: Global Distribution and Density of Constructed Impervious Surfaces, Sensors, 7, 1962–1979, 2007.

ESRI: ArcMap 10.2.1. Environmental Systems Resource Institute, Redlands, California., 2013.

23

605 Fankhauser, R.: Automatic determination of imperviousness in urban areas from digital orthophotos, Water
     Science and Technology, 39, 81 – 86, 1999.

Forzieri, G., Battistini, A., and Catani, F.: ES4LUCC: A GIS-tool for Remotely Monitoring Landscape Dynam-
     ics, Comput. Geosci., 49, 72–80, 2012a.

Forzieri, G., Moser, G., and Catani, F.: Assessment of hyperspectral MIVIS sensor capability for heterogeneous
610 landscape classification, ISPRS Journal of Photogrammetry and Remote Sensing, 74, 175–184, 2012b.

Freund, Y. and Schapire, R.: A desicion-theoretic generalization of on-line learning and an application to boost-
     ing, in: Computational Learning Theory, vol. 904 of *Lecture Notes in Computer Science*, pp. 23–37, Springer
     Verlag, Heidelberg, 1995.

Hall, J., et al., Molnar, P., and Schär, C.: Understanding flood regime changes in Europe: A state-of-the-art
615 assessment, Hydrology and earth system sciences, 18, 2735–2772, 2014.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae,
     S.: Global flood risk under climate change, Nature Climate Change, 3, 816–821, 2013.

Hodgson, M. E., Jensen, J. R., Tullis, J. A., Riordan, K. D., and Archer, C. M.: Synergistic use of lidar and color
     aerial photography for mapping urban parcel imperviousness, Photogrammetric Engineering and Remote
620 Sensing, 69, 973–980, 2003.

Hu, X. and Weng, Q.: Estimating impervious surfaces from medium spatial resolution imagery using the self-
     organizing map and multi-layer perceptron neural networks, Remote Sensing of Environment, 113, 2089–
     2102, 2009.

Hughes, G.: On the mean accuracy of statistical pattern recognizers, IEEE Transactions on Information Theory,
625 14, 55–63, 1968.

Leandro, J., Chen, A. S., Djordjević, S., and Savić, D. A.: Comparison of 1D/1D and 1D/2D coupled
     (sewer/surface) hydraulic models for urban flood simulation, Journal of hydraulic engineering, 135, 495–
     504, 2009.

Lee, J. and Heaney, J.: Estimation of Urban Imperviousness and its Impacts on Storm Water Systems, Journal
630 of Water Resources Planning and Management, 129, 419–426, 2003.

Lee, S. and Lathrop, R.: Subpixel analysis of Landsat ETM+ using self-organizing map (SOM) neural networks
     for urban land cover characterization, IEEE Transactions on Geoscience and Remote Sensing, 44, 1642–
     1654, 2006.

Leitão, J. P., Moy de Vitry, M., Scheidegger, A., and Rieckermann, J.: Assessing the quality of Digital Ele-
635 vation Models obtained from mini-Unmanned Aerial Vehicles for overland flow modelling in urban areas,
     Hydrology and Earth System Sciences Discussions, 12, 5629–5670, 2015.

Li, W., Ouyang, Z., Zhou, W., and Chen, Q.: Effects of spatial resolution of remotely sensed data on estimating
     urban impervious surfaces, Journal of Environmental Sciences, 23, 1375 – 1383, 2011.

Lu, D. and Weng, Q.: Use of impervious surface in urban land-use classification, Remote Sensing of Environ-
640 ment, 102, 146–160, 2006.

Lu, D. and Weng, Q.: Extraction of urban impervious surfaces from an IKONOS image, International Journal
     of Remote Sensing, 30, 1297–1311, 2009.

Lu, D., Tian, H., Zhou, G., and Ge, H.: Regional mapping of human settlements in southeastern China with
     multisensor remotely sensed data, Remote Sensing of Environment, 112, 3668 – 3679, 2008.

24

645   Mathieu, R., Aryal, J., and Chong, A. k.: Object-Based Classification of Ikonos Imagery for Mapping Large-Scale Vegetation Communities in Urban Areas, Sensors, 7, 2860–2880, 2007.

Melesse, A. M. . and Wang, X.: Impervious Surface Area Dynamics and Storm Runoff Response, pp. 369–384, CRC Press, Taylor & Francis Group: Boca Raton, 2008.

Mohapatra, R. P., Wu, C., and Weng, Q.: Subpixel imperviousness estimation with IKONOS imagery: an arti-

650   ficial neural network approach, pp. 21–35, CRC Press, Taylor & Francis Group: Boca Raton, 2008.

Montgomery, D. C. and Runger, G. C.: Applied statistics and probability for engineers, Hoboken, NJ: Wiley, 2007.

Montgomery, D. C., Peck, E. A., and Vining, G. G.: Introduction to linear regression analysis, vol. 821, Hoboken, NJ: Wiley, 2012.

655   Moser, G., Serpico, S., and Benediktsson, J.: Land-Cover Mapping by Markov Modeling of Spatial-Contextual Information in Very-High-Resolution Remote Sensing Images, Proceedings of the IEEE, 101, 631–651, 2013.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I: A discussion of principles, Journal of Hydrology, 10, 282 – 290, 1970.

660   Nielsen, N. H., Joergensen, A., and Larsen, A.: Use of spectral analysis in urban drainage modelling, in: International Conference on Urban Drainage, Porto Alegre, Brazil, 11-16 September, 2011, 2011.

Parece, T. E. and Campbell, J. B.: Comparing Urban Impervious Surface Identification Using Landsat and High Resolution Aerial Photography, Remote Sensing, 5, 4942–4960, 2013.

Powell, R. L., Roberts, D. A., Dennison, P. E., and Hess, L. L.: Sub-pixel mapping of urban land cover using

665   multiple endmember spectral mixture analysis: Manaus, Brazil, Remote Sensing of Environment, 106, 253–267, 2007.

Ravagnani, F., Pellegrinelli, A., and Franchini, M.: Estimation of Urban Impervious Fraction from Satellite Images and Its Impact on Peak Discharge Entering a Storm Sewer System, Water Resources Management, 23, 1893–1915, 2009.

670   Rojas, R., Feyen, L., and Watkiss, P.: Climate change and river floods in the European Union: Socio-economic consequences and the costs and benefits of adaptation, Global Environmental Change, 23, 1737 – 1751, 2013.

Sauerbier, M. and Eisenbeiß, H.: UAVs for the documentation of archaeological excavations, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 38, 526–531, 2010.

Small, C.: High spatial resolution spectral mixture analysis of urban reflectance, Remote Sensing of Environ-

675   ment, 88, 170–186, 2003.

Stankowski, S. J.: Population density as an indirect indicator of urban and suburban land-surface modifications, US Geological Survey Professional Paper, Washington D.C., USA, 800, 219–224, 1972.

swisstopo: SWISSIMAGE: Das digitale Farborthophotomosaik der Schweiz. Eidgenössisches Departament für Verteitigung, Bevölkerungsschutz und Sport VBS., http://www.swisstopo.admin.ch/internet/swisstopo/

680   de/home/products/images/ortho/swissimage.parsysrelated1.76752.downloadList.50684.DownloadFile.tmp/ infosi201003deu.pdf, 2010.

swisstopo: swissALTI3D - Ausgabebericht. Eidgenössisches Departament für Verteitigung, Bevölkerungsschutz und Sport VBS., http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/height/ swissALTI3D.html, 2014.

685   Tan, Q., Liu, Z., and Li, X.: Mapping urban surface imperviousness using SPOT multispectral satellite images, in: IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2009), vol. 3, pp. III–346–III–348, 2009.

Tokarczyk, P., Wegner, J., Walk, S., and Schindler, K.: Features, Color Spaces, and Boosting: New Insights on Semantic Classification of Remote Sensing Images, IEEE Transactions on Geoscience and Remote Sensing,

690   53, 280–295, 2015.

UN: World Population Prospects 1950-2050. The 2012 Revision. Key Findings and Advance Tables. United Nations Population Division., http://esa.un.org/wpp/Documentation/pdf/WPP2012_%20KEY%20FINDINGS.pdf, 2013.

US-EPA: Storm Water Management Model (SWMM), Version 5.0.022. United States Environmental Protection

695   Agency., http://www2.epa.gov/water-research/storm-water-management-model-swmm, 2010.

Van de Voorde, T., De Roeck, T., and Canters, F.: A comparison of two spectral mixture modelling approaches for impervious surface mapping in urban areas, International Journal of Remote Sensing, 30, 4785–4806, 2009.

Vrugt, J. A. and Robinson, B. A.: Improved evolutionary optimization from genetically adaptive multimethod

700   search, Proceedings of the National Academy of Sciences, 104, 708–711, 2007.

Weng, Q. and Hu, X.: Medium Spatial Resolution Satellite Imagery for Estimating and Mapping Urban Impervious Surfaces Using LSMA and ANN, IEEE Transactions on Geoscience and Remote Sensing, 46, 2397–2406, 2008.

Weng, Q., Hu, X., and Liu, H.: Estimating impervious surfaces using linear spectral mixture analysis with

705   multitemporal ASTER images, International Journal of Remote Sensing, 30, 4807–4830, 2009.

Wu, C. and Murray, A. T.: Estimating impervious surface distribution by spectral mixture analysis, Remote Sensing of Environment, 84, 493–505, 2003.

Yang, L., Huang, C., Homer, C. G., Wylie, B. K., and Coan, M. J.: An approach for mapping large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial resolution imagery, Canadian Journal of

710   Remote Sensing, 29, 230–240, 2003.

Yang, L., Jiang, L., Lin, H., and Liao, M.: Quantifying Sub-pixel Urban Impervious Surface through Fusion of Optical and InSAR Imagery, GIScience and Remote Sensing, 46, 161–171, 2009.

Yuan, F. and Bauer, M. E.: Mapping impervious surface area using high resolution imagery: A comparison of object-based and per pixel classification, in: Proceedings of ASPRS 2006 annual conference, pp. 1–5, 2006.

715   Zhang, C. and Kovacs, J.: The application of small unmanned aerial systems for precision agriculture: a review, Precision Agriculture, 13, 693–712, 2012.

Zhou, Y. and Wang, Y. Q.: Extraction of Impervious Surface Areas from High Spatial Resolution Imageries by Multiple Agent Segmentation and Classification, Photogrammetric Engineering and Remote Sensing, 74, 857–868, 2008.

Appendix A: Supplementary Material

A1   Remote sensing methods to extract the imperviousness maps

A considerable amount of remote sensing research has been devoted to the problem of mapping impervious surfaces. Here, we review some of the previous studies and evaluate them in respect of the datasets and classification methods. Furthermore, we focus on the studies which use the classified land-use to predict urban rainfall runoff.

Whereas few studies have used low-resolution (GSD > 100m) satellite sensors, such as MODIS (Lu et al., 2008; Boegh et al., 2009), AVHRR (Carlson and Arthur, 2000) and DMSP-OLS (Elvidge et al., 2007; Lu et al., 2008), the large part of the research in this area focused on medium and high spatial resolution satellite data. Because of its exceptional temporal resolution, Landsat is still the most popular satellite platform. A large number of authors used Landsat 5 TM (Civco et al., 2002; Carlson, 2004; Bauer et al., 2008; Yuan and Bauer, 2006; Li et al., 2011; Parece and Campbell, 2013; Dougherty et al., 2004) and Landsat 7 ETM+ data (Civco et al., 2002; Yang et al., 2003; Wu and Murray, 2003; Lu and Weng, 2006; Lee and Lathrop, 2006; Powell et al., 2007; Chormanski et al., 2008; Chabaeva et al., 2009; Van de Voorde et al., 2009) for analysing impervious surface cover. Other examples of using images acquired by high resolution platforms include SPOT (Yang et al., 2009; Li et al., 2011; Tan et al., 2009) and ASTER (Weng and Hu, 2008; Hu and Weng, 2009; Weng et al., 2009).

However, recent developments of remote sensing imaging sensors and platforms gave access to VHR imagery. Examples of VHR satellite sensors application to impervious surfaces mapping include Ikonos (Cablk and Minor, 2003; Lu and Weng, 2009; Mohapatra et al., 2008; Chormanski et al., 2008; Van de Voorde et al., 2009; Mathieu et al., 2007), and QuickBird (Lu et al., 2008; Yuan and Bauer, 2006; Zhou and Wang, 2008). Except of satellite imagery, aerial images are also an important source of information. Many studies used aerial orthophotos only as a reference check to satellite imagery (Yang et al., 2003; DeBusk et al., 2010; Parece and Campbell, 2013). However few attempts to automatically map imperviousness using such data were made (Nielsen et al., 2011; Dougherty et al., 2004; Hodgson et al., 2003; Zhou and Wang, 2008; Fankhauser, 1999; Lee and Heaney, 2003).

One possible way to extract imperviousness from images is to interpret them manually. Even though this is the most reliable method, and has been used in few studies (*e.g.* Lee and Heaney (2003)), it is very costly in terms of time and money. Therefore it is common to automate the process by using image classification. Maybe the simplest method is to assume that only vegetation is pervious and rely on the normalized differential vegetation index (NDVI) (Nielsen et al., 2011; Carlson and Arthur, 2000). Many of the studies use more advanced classification methods, such as object based image analysis (OBIA) (Zhou and Wang, 2008; Hodgson et al., 2003; Nielsen et al., 2011; Mathieu et al., 2007). Other examples include maximum likelihood classifier (Fankhauser,

1999; Hodgson et al., 2003), spectral mixture analysis (SMA) (Small, 2003; Van de Voorde et al., 2009; Weng et al., 2009), artificial neural networks (ANN) (Chormanski et al., 2008; Van de Voorde et al., 2009; Lee and Lathrop, 2006), classification and regression trees (CART) (Yang et al., 2003; Li et al., 2011; Dougherty et al., 2004) and rule-based classifiers (Hodgson et al., 2003). Some of the mentioned methods also use the perviousness maps for urban drainage modelling like we do (Nielsen et al., 2011; Melesse and Wang, 2008; Chormanski et al., 2008; Dougherty et al., 2004; Lee and Heaney, 2003; Fankhauser, 1999). However, to our best knowledge no studies exist, that used UAV-based imagery to extract imperviousness information, and to use it in the field of urban drainage modelling.

A2 UAV platform

The UAV platform used in this study is an autonomous fixed-wing drone produced by senseFly SA (*cf*. http://www.senseFly.com). Table 3 includes detailed information about the platform.

Weight (incl. camera)	ca. 0.69 kg
Wingspan	96 cm
Material	EPP foam, carbon structure and composite parts
Propulsion	Electric pusher propeller, 160 W brushless DC motor
Battery	11.1 V, 2150 mAh
Camera (supplied)	16 MP IXUS/ELPH
Cameras (oprional)	S110 RGB, thermoMAP
Max. flight time	50 min
Nominal speed	40-90 km/h
Wind resistance	Up to 45 km/h (12 m/s)
Radio link range	Up to 3 km
Max. coverage (single flight)	Up to 12 km^2
Cost	ca. 20'000 CHF (Drone + Software)

Table 3. Specifications of the UAV used in the study (source: http://www.senseFly.com)

The imaging unit mounted on a UAV was a customized version of Canon IXUS 127 HS compact camera. Table 4 includes its specifications.

Camera effective pixels	ca. 16.1 million
Lens' focal length	4.3 - 21.5 mm (35 mm equivalent: 24 - 120 mm)
Interfaces	Hi-speed USB, HDMI Output, Analog audio output, Analog video output (NTSC/PAL)
Dimensions	$93.2 \times 57.0 \times 20.0$ mm
Weight	ca. 135 g (incl. battery and memory card)

Table 4. Specifications of the Canon IXUS 127 HS Camera

A3   Exploratory data analysis of the importance of image source and processing method for the surface runoff

A3.1   Regression

775 **Imperviousness**

Please refer to Table 5 and Figure 12.

Here we try to answer a following question: Which has the greater influence/is stronger correlate with a change in imperviousness and surface runoff characteristics, the image source or the process-

780 ing method?

Model and results

Here we present logit-transformation of imperviousness. This was done to constrain the model output to the range between 0 and 1 and not to improve the statistical assumptions regarding the errors

785 of the data generating process.

Description/Interpretation

UAV images seem to be negatively correlated with the imperviousness. The effect is not really strong. Regarding the methods, there seems to be no influence, because the estimated linear relation is prac-

790 tically negligible. In addition, there is no evidence for interactions between the image source and the processing method.

3

Peak runoff

795 *Model and results*

Please refer to Table 6 and Figure 13

Description/Interpretation

UAVdata generally seem to produce slightly smaller peaks, whereas the RQE method is positively

800 correlated to peak hight. However both effects are not significant by any means. There are no inter-

actions of these two. Statistical assumptions are not fulfilled.

Runoff volume

805

Model and results

Please refer to Table 7 and Figure 14

Description/Interpretation

810 UAV data generally seem to produce slightly runoff volumes, whereas the RQE method is positively

correlated to runoff volume. However both effects are not significant by any means. There are no

interactions of these two. Statistical assumptions are not fulfilled.

815 **Time to peak**

Analysis was not performed, because exploratory analysis suggest that the differences between the

different image sources are negligibly small.

820

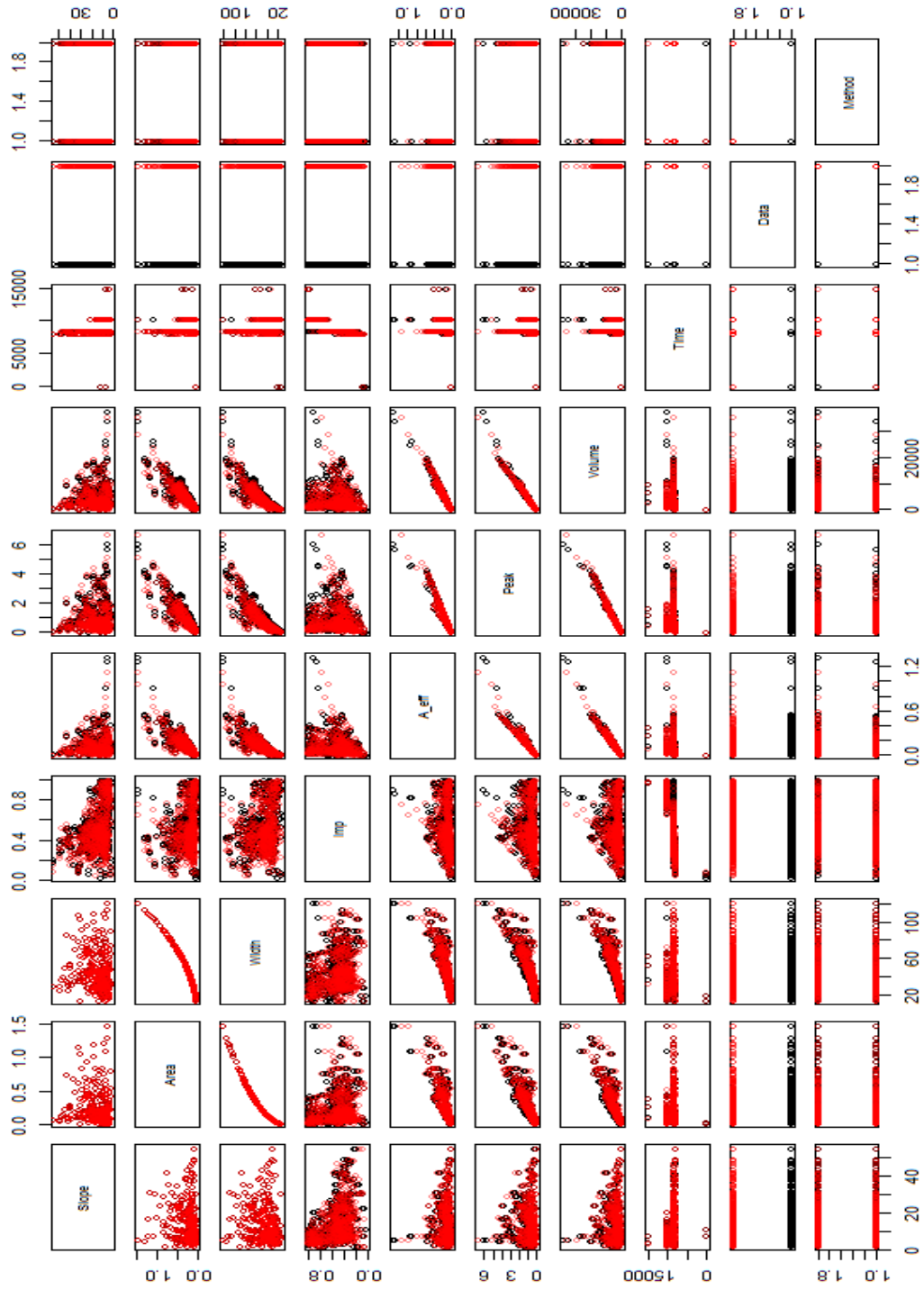**A4   Pipe flow predictions**

Please refer to Figure 15

825

5

Figure 10. Scatterplot of surface runoff characteristics for the 307 individual subcatchments of the Wartegg SWMM model. Black = Orthophotos, Red= UAV images. A_eff: effective area, Imp: imperviousness
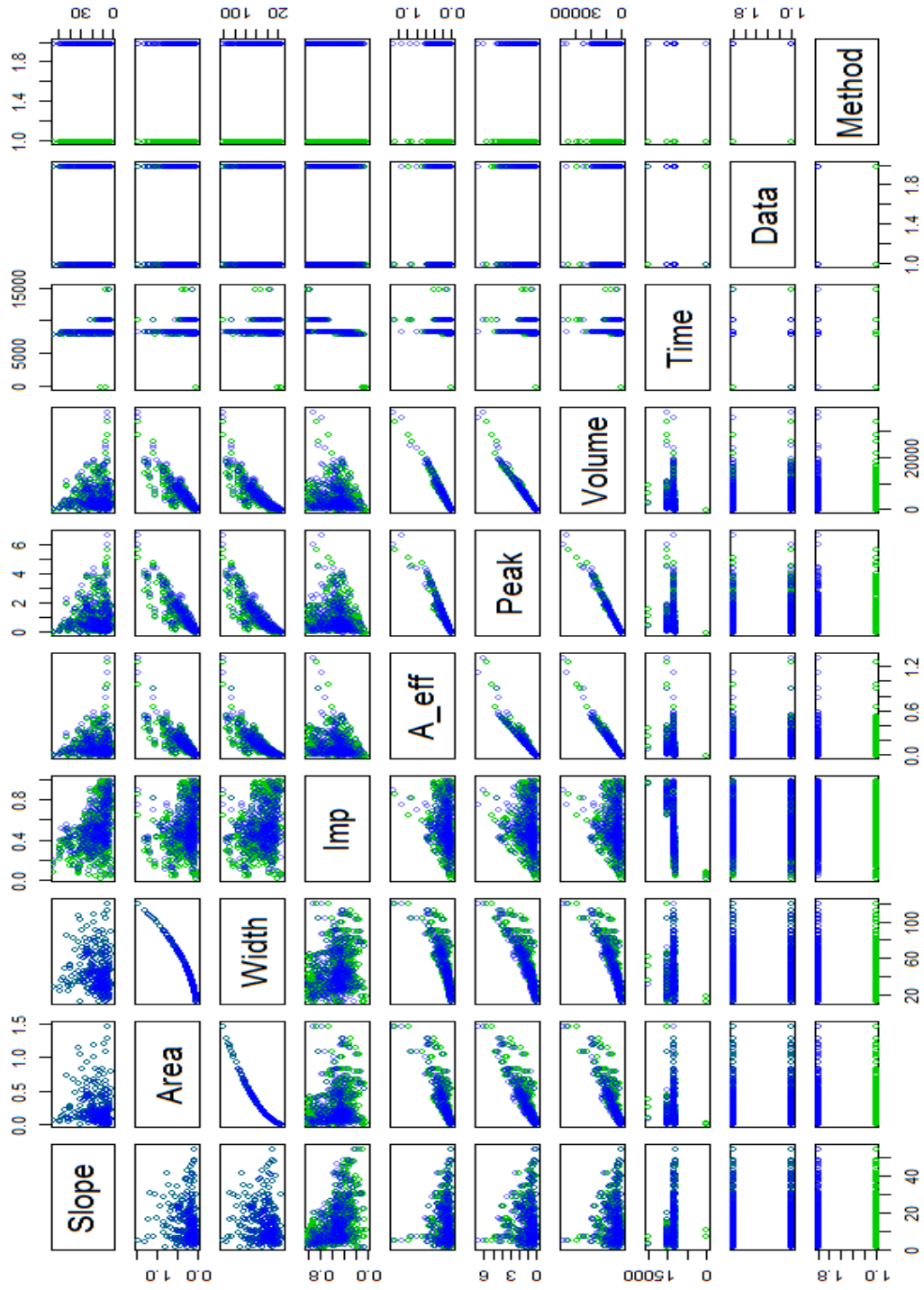
Figure 11. Scatterplot of surface runoff characteristics for the 307 individual subcatchments of the Wartegg SWMM model. Green= ML, Blue = RQE. A_eff: effective area, Imp: imperviousness

Table 5. Summary results of the regression analysis. The negative sign of the estimated slope parameter suggests that the UAV images generally go together with a lower imperviousness. In addition, the influence of the image source seems to be larger than that of the classification method, although the high p-values for all parameters suggest that it is not very likely that the observed values of imperviousness were to have occurred under the given statistical model.

	Dependent variable:
	Volume
DataUAV	-301.699
	(331.033)
MethodRQE	298.671
	(331.033)
DataUAV:MethodRQE	199.362
	(468.151)
Constant	3,893.406***
	(234.075)
Observations	1,228
R^2	0.003
Adjusted R^2	0.001
Residual Std. Error	4,101.333 (df = 1224)
F Statistic	1.274 (df = 3; 1224)

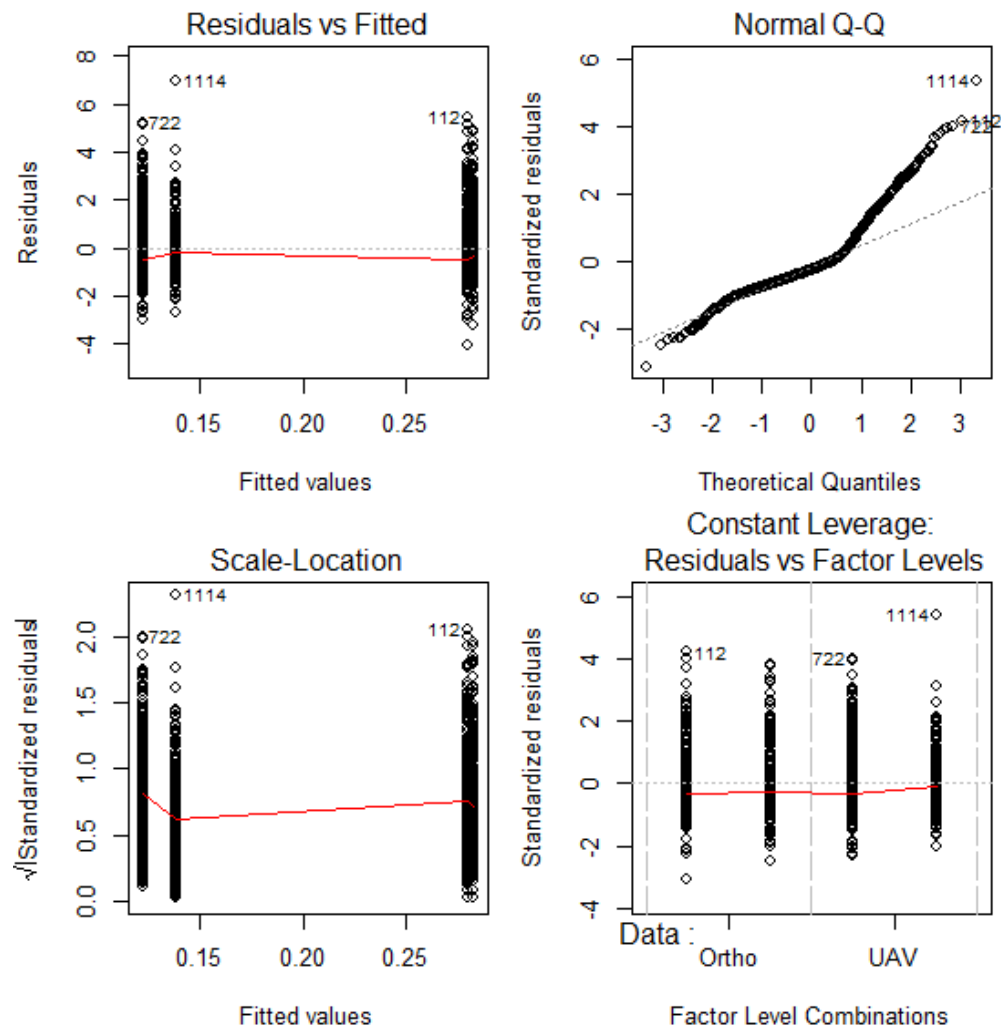| *Note:*	*p<0.1; **p<0.05; ***p<0.01

Figure 12. Diagnostic plots of the regression analysis. It is obvious that the statistical assumptions are not fulfilled very well and that the observe imperviousness is not well explained.

Table 6. Summary results of the regression analysis for peak runoff. The negative sign of the estimated slope parameter suggests that the UAV images generally go together with a lower stormwater peak flow. Here, the influence of the image source seems to be in the same order of magnitude than that of the classification method, although the former is negatively correlated and the latter has a positive correlation with peak runoff. Again, the high p-values for all parameters suggest that it is not very likely that the observed peak runoff values were to have occurred under the given statistical model.

	Dependent variable:
	Peak
DataUAV	-0.065
	(0.067)
MethodRQE	0.068
	(0.067)
DataUAV:MethodRQE	0.038
	(0.094)
Constant	0.826***
	(0.047)
Observations	1,228
R^2	0.004
Adjusted R^2	0.001
Residual Std. Error	0.827 (df = 1224)
F Statistic	1.507 (df = 3; 1224)

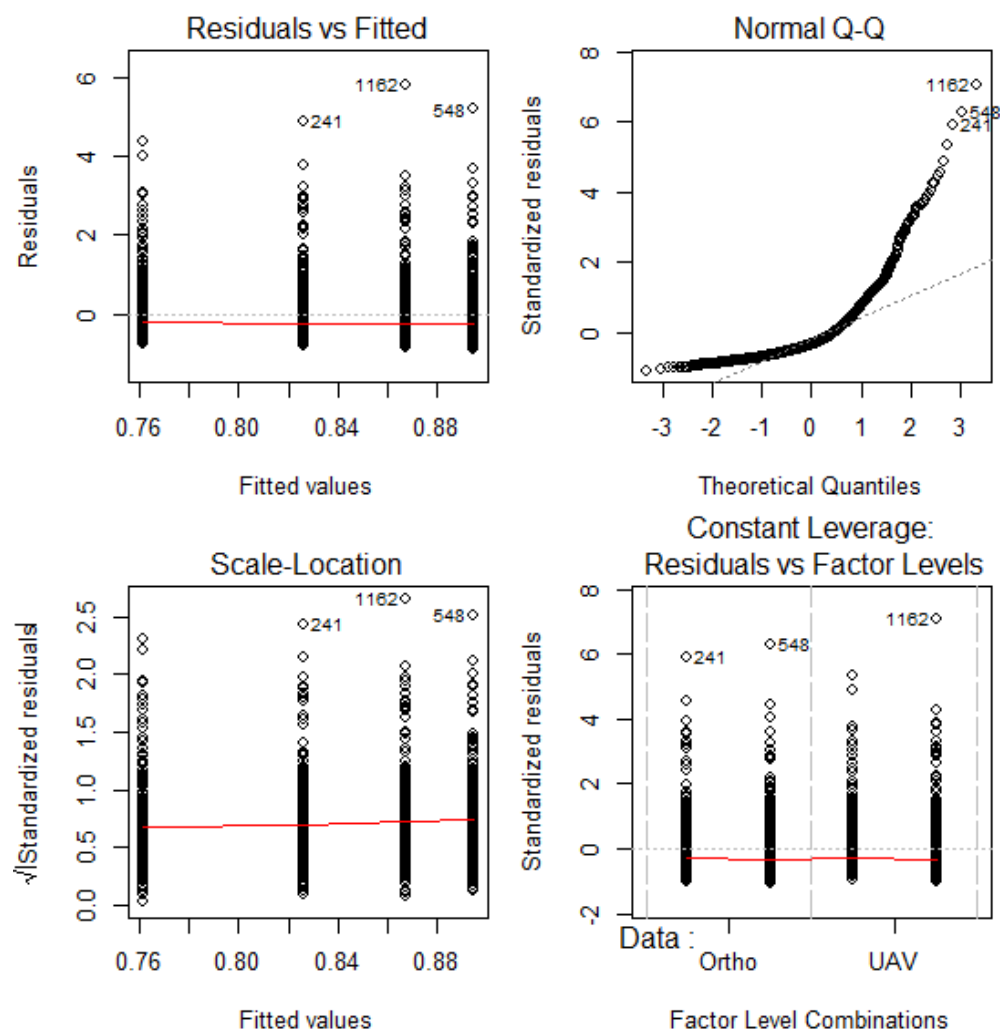*Note:* *p<0.1; **p<0.05; ***p<0.01

Figure 13. Diagnostic plots of the regression analysis. It is obvious that the statistical assumptions are not fulfilled very well and that the observe imperviousness is not well explained.

Table 7. Summary results of the regression analysis for runoff volume.

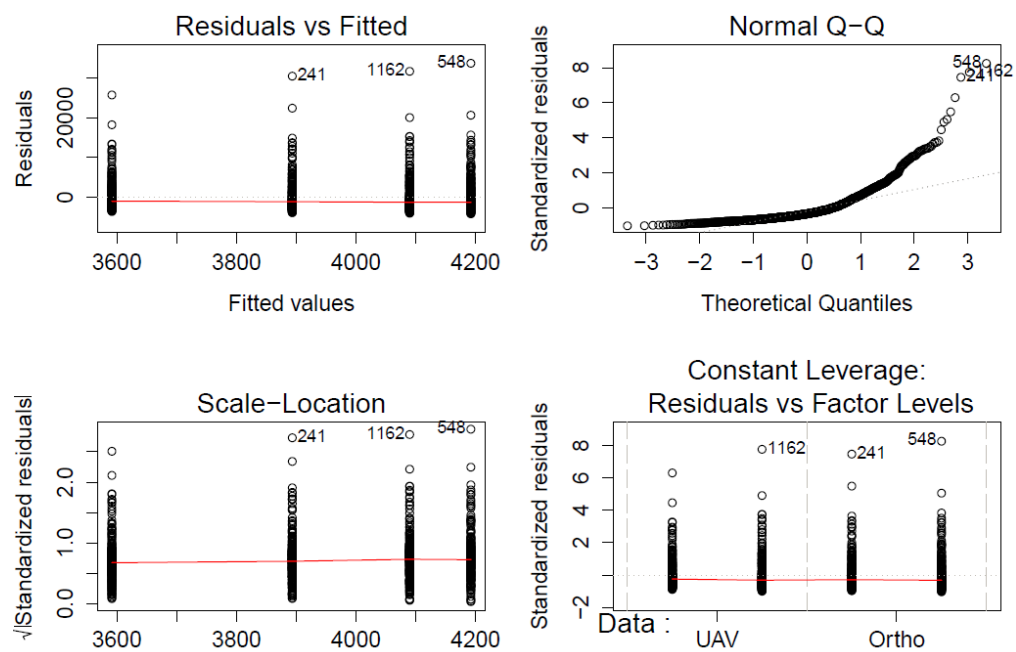|	*Dependent variable:*
	Volume
DataUAV	-301.699
	(331.033)
MethodRQE	298.671
	(331.033)
DataUAV:MethodRQE	199.362
	(468.151)
Constant	3,893.406***
	(234.075)
Observations	1,228
R^2	0.003
Adjusted R^2	0.001
Residual Std. Error	4,101.333 (df = 1224)
F Statistic	1.274 (df = 3; 1224)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 14. Diagnostic plots of the regression analysis. It is obvious that the statistical assumptions are not fulfilled very well and that the observe imperviousness is not well explained.
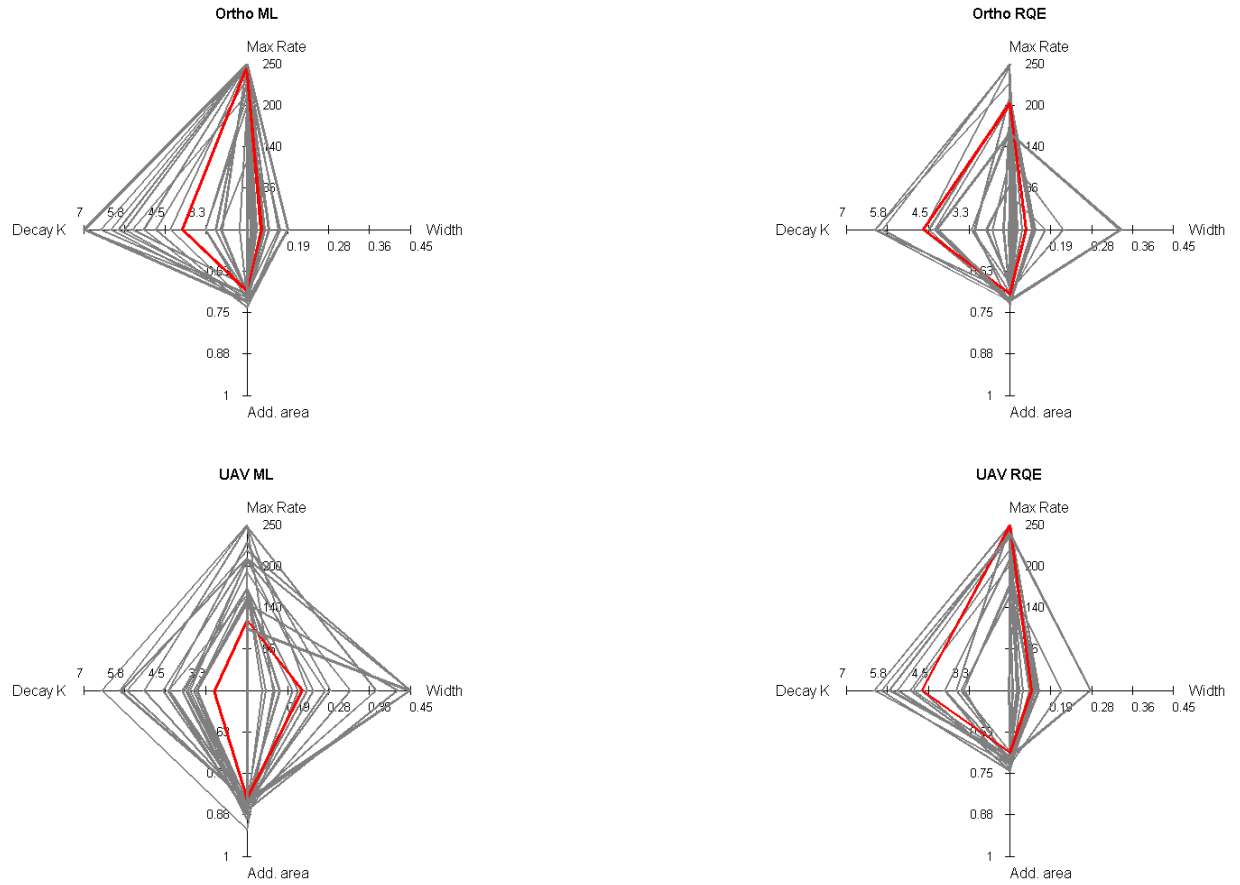
Figure 15. Distribution of calibration parameter (Decay K: infiltration decay rate after HORTON; MaxRate: maximum infiltration rate after HORTON; width: conceptual parameter describing the width of a subcatchment; Add.area: conceptual parameter describing event-based sewer infiltration) values identified during the auto-calibration process. Grey rhombs represent the optimum parameter set identified for each population; the red rhomb represents the final parameter set.