Hydrol. Earth Syst. Sci. Discuss., 12, 1205–1245, 2015
www.hydrol-earth-syst-sci-discuss.net/12/1205/2015/
doi:10.5194/hessd-12-1205-2015

Hydrology and Earth System Sciences
Open Access
Discussions

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

P. Tokarczyk[1], **J. P. Leitao**[2], **J. Rieckermann**[2], **K. Schindler**[1], **and F. 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 Management, ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zürich

Received: 17 December 2014 – Accepted: 8 January 2015 – Published: 27 January 2015

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

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 area.

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 unavailable. Modern unmanned air vehicles (UAVs) allow acquiring 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 is tested and applied in a state-of-the-art urban drainage modelling exercise. In a real-life case study in the area of Lucerne, Switzerland, we compare imperviousness maps generated from a consumer micro-UAV and standard large-format aerial images acquired by the Swiss national mapping agency (*swisstopo*). After assessing their correctness, 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 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 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 using UAV imagery processed with modern classification methods achieve accuracy comparable with standard, off-the-shelf aerial imagery. In the examined case study, we find that the different imperviousness maps only have a limited influence on modelled surface runoff and pipe flows. 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 superior quality and a competitive price. Our analyses furthermore suggest that spatially more detailed urban drainage models can even better benefit from the full detail of UAV imagery.

1 Introduction

In the last century we have witnessed a massive migration of people from rural areas to the cities. Today, a majority of the human population live in the 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 cabable to cope with a constantly increasing number of its users. Accordingly, ensuring water supply for the people is important, but being able to safely direct stormwater away from populated areas, in order to avoid flooding, is also a challenging task. It requires predicting the hydraulic behaviour of the given drainage infrastructure using reliable hydrological models. 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 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

HESSD
12, 1205–1245, 2015

not only cause floods, but also lead to extensive erosion events and increase the risk of loading waterbeds with sediments, and its associated constitents (e.g. phosphorus, nutrients and pesticides). Growing level of surface imperviousness has a negative impact on water quality, because the pollutants will be more easily washed out to the nearby waterbodies.

Many different methods have been developed and applied to map impervious areas. Amongst these are manual methods, which for example use existing built-up zone plans, or manually process remote sensing images (Krejci et al., 1994). An important step towards automatization of these processes 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 Supplement). In general, most of the studies on extraction of impervious surfaces from remote sensing data focused on satellite images. Examples include low-resolution sensors, such as MODIS (Boegh et al., 2009), AVHRR (Carlson and Arthur, 2000) or DMSP-OLS (Lu et al., 2008); medium-resolution, such as Landsat 5 TM (Parece and Campbell, 2013) and Landsat 7 ETM+ (Van de Voorde et al., 2009); or high-resolution: SPOT (Li et al., 2011) and ASTER (Weng et al., 2009). 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 Ikonos (Chormanski et al., 2008) and Quick-Bird (Zhou and Wang, 2008) or VHR aerial images (Fankhauser, 1999; Nielsen et al., 2011) were quickly adopted for impervious surfaces mapping. Ravagnani et al. (2009) attempted to use impervious surfaces extracted from VHR satellite and 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 0.31 m GSD in panchromatic

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

HESSD
12, 1205–1245, 2015

channel) and have short revisit periods, are still 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). In comparison to a standard, off-the-shelf satellite or aerial remote 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 information by describing object's local neighbourhood. Here, in order to properly exploit high level of detail of UAV imagery, we propose to use a randomized quasi-exhaustive (RQE) feature bank (Tokarczyk et al., 2015), which consists of a multitude of multiscale textural features describing both, spectral and height information. To sidestep manual selection of features from this exhaustive feature set, we use a boosting classifier to only choose the optimal features during training[2].

In this study, we investigated the feasibility of using imagery acquired with UAVs for urban drainage modelling. Specifically, we present three main aspects:

1. we evaluate whether such low-cost monitoring data of land-use can be used to assess the performance of urban drainage systems,

2. we suggest using a boosting classifier in conjunction with RQE feature bank, to properly exploit high level of detail of UAV imagery, and

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

HESSD
12, 1205–1245, 2015

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.

We demonstrate the usefulness of our approach on a case study from a small urban area in Lucerne, Switzerland. First, we compare the UAV data with standard airborne imagery using a maximum likelihood classifier and the RQE method on both image sources. Second, we use a hydrodynamic model to show the consequences of different land-use information on urban drainage performance indicators (see Fig. 1).

In general, our results are promising because we are able to classify land-use using UAV imagery as accurately as from standard aerial images. We find that the different imperviousness maps only have a limited influence on surface runoff and pipe flows. Interestingly, this indicates that lumped models might become a bottleneck in detailed rainfall–runoff studies. In our view, a major advantage of UAVs in practical applications is the possibility to flexibly acquire up-to-date and detailed aerial images at a good quality and a competetive price, at least for small areas.

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 Materials and methods

2.1 Overview

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. In addition, we explore their usefulness also in relation to predicting surface runoff and pipe flows, which are the ultimately decisive processes for the urban drainage analysis (see Fig. 1). In a case study in the area of Lucerne, Switzerland we evaluated the two remote sensing datasets to show following:

- assess the efficiency of a recent high-performance classification method (RQE) and compare it to a standard classifier (ML) commonly used for perviousness mapping applied to images acquired with an UAV in relation to standard off-the-shelf aerial images (1), and

- perform an end-to-end comparison, in which the maps from different data sources processed through different classification methods were used as input for a hydraulic sewer model predicting surface runoff (2) and in-sewer channel flow (3), whereas the latter is compared to a measured reference.

2.2 Case study and datasets

2.2.1 Case study

For our case study we used a residential area, called Wartegg catchment, in the city of Lucerne, Switzerland (see Fig. 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 centres and sports grounds.

Storm- and wastewater is drained by separate and combined sewers (see Fig. 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 partly interlinked with the storm water network.

HESSD
12, 1205–1245, 2015

2.2.2 Remote sensing datasets

Image data

In this study we used two image (see Fig. 3) and two height 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 acuired on-demand (standard *swisstopo* orthophotos have a GSD of 0.25 m), images acquired by *swisstopo* are publically 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 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 Mpix sensor, mounted on a fixed-wing micro-UAV platform (see Sect. S2 in the Supplement for details). The flight was performed during leaves-off conditions in March 2014. Orthophotos (RGB channels) generated from the acquired images have a GSD of 0.10 m. The main advantage of UAVs, when compared to manned aircraft with large-format mapping cameras, lies in their flexibility, in terms of place and time of deployment, and their low cost for small areas.

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 step reduces the time needed for training the classifier.

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

HESSD
12, 1205–1245, 2015

Height model

In this study we used two different height models. Classification of the swisstopo dataset was performed using the swissALTI3D product (swisstopo, 2014), whereas for UAV imagery we used a height model extracted using dense image matching. The swissALTI3D product is a digital terrain model (DTM) and it features a grid size of 2 m. For impervious surfaces classification the model has been upsampled to the resolution of corresponding image dataset. The second model is a normalized digital surface model (nDSM), and was generated by subtracting a digital surface model (DSM) extracted from UAV images, and a DTM provided by the swisstopo. For urban drainage modelling we used the swisstopo height model, because of its empirically proven quality (swisstopo, 2014).

2.2.3 Rainfall

Precipitation data was collected from a meteo station located 2 km away from the Wartegg catchment area, operated by the Swiss Meteorological Intsitiute (MeteoSwiss). Recordings were taken in a 10 min interval using a tipping bucket rain gauge with a precision of 0.1 mm – readings started at 1981 and last until today. 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 was not considered in the study due to the short distance between the meteo station and the study area.

2.2.4 Sewer flow reference data

Two flow data sets were obtained from in-sewer flow monitoring located at the outlet of the subcatchment (see Fig. 2). Over a period of four months (17 July 2014 to 18 November 2014) the sewer flow was monitored with two different sensors, (i) Sigma 950 (HACH-LANGE) – 1 min monitoring frequency and (ii) FLO-DAR (Marsh Mc Bir-

HESSD
12, 1205–1245, 2015

ney) – 15 min monitoring frequency, to provide redundant high quality measurements. Correlation analysis between the two reference signals show a high agreement and confirm the high quality of data.

2.2.5 Urban drainage model

Urban drainage models are tools to simulate surface runoff and sewer pipe flow. They can be used to analyze the hydraulic behaviour of the urban drainage system, and to support the analysis of flood risk and pollution of receiving water bodies. In general, these models include two main compartments: the hydrological model and the hydraulic model. The hydrological model calculates the initial precipitation losses, and resultant time and space distribution of the direct runoff. The output is then used as input for the hydraulic model to simulate surface and sewer network flows.

Like hydraulic models, hydrological models implemented in urban drainage modelling software are based on simplifying, conceptual formulations of transport phenomena that occur in the catchment. Generally, these models assume 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 infiltration process which is a function of the soil water saturation level. Surface runoff ends when the rainfall is smaller than the verified rainfall losses. Impervious surfaces are those where no infiltration occurs; the catchment imperviousness degree and the catchment imperviousness spatial distribution are then expected to have a great impact on surface runoff and urban drainage system modelling results.

To describe the hydraulic behaviour of the Wartegg catchment area during dry weather and storm events we developed a hydrodynamic sewer model implemented on the EPA SWMM modelling platform (US-EPA, 2010). The modelling platform SWMM is chosen as represents a standard, well-established and freely available urban drainage model. The surface runoff is described by a conceptual approach; pipe flow through the conveyance system is described with the Saint Venant approach.

2.3 Methodology

2.3.1 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, which is de-facto a standard classification method in the field of urban hydrology, is a simple generative model which assumes that the image features within each target class follow a Gaussian 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. An important limitation of ML is that it is not well suited for high-dimensional data; typically its performance degrades beyond a dozen or so feature dimensions due to the "curse of dimensionality" (Hughes, 1968). For a medium-resolution imagery, where objects are generally spectraly consistent, 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 (e.g. roof consists of many pixels and substructures 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 neighbourhood 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 spectral intensities at the pixel itself as features.

Boosting

As an alternative to ML we chose 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 neighbourhood.

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 Fig. 4). 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 ground truth. 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, 14 000 and 36 000 pixels). Thereby we can evaluate how the size of the training data has an influence on the overall classification accuracy.

HESSD
12, 1205–1245, 2015

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. Similarily 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.

2.3.2 Assessing the importance of input data for surface runoff

To assess the importance of input data and the processing method on the surface runoff, we predicted the surface runoff for a medium-size rain event. Then, we analysed the runoff of the 307 individual subcatchments regarding relevant attributes, such as peak runoff and volume. As it is very challenging to directly observe surface runoff that can be compared to the 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. S3 in the Supplement for details).

Prediction of surface runoff

To predict surface runoff, we selected a rain event lasting from 10 August 2014 at 22:00 to 11 August 2014 at 03:00. This was a moderate event with a total volume of 29.7 mm and a peak rainfall intensity of 2.9 $\mathrm{L\,s^{-1}}$. Compared to other events registered for this area, it was an average event, thus we believe that general rainfall–runoff characteristics remain the same. We characterized the hydrographs of all 307 subcatchments with the following attributes: (i) peak flow (Q_{peak}) and (ii) volume of the peak (V_{peak}).

HESSD
12, 1205–1245, 2015

Performance assessment

Exploratory data analysis of surface runoff characteristics

To summarize the important characteristics of the surface runoff, we visualized important aspects using boxplots and scatterplots (see Fig. 7). Main research questions were:

- Which differences in imperviousness (Δ_{Imp}) result for each subcatchment: (i) for the two data sources and (ii) for the two classification methods?
- Does the the image source have a substantial influence on the predictions of surface runoff? 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 (Montgomery 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 classifier. The model for a dependent variable y is:

$$y_i = \beta_0 + \beta_1 I_i^{\mathrm{Data}} + \beta_2 I_i^{\mathrm{Method}} + \beta_3 I_i^{\mathrm{Data\times Method}} + \epsilon_i \qquad (1)$$

where y_i is the ith 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). $\beta_0 \ldots \beta_3$ are the parameters to be estimated and ϵ_i is a random error term. If ϵ_i is normally distributed and independent, i.e., $\epsilon_i \sim N(0, \sigma^2)$, this model is

HESSD
12, 1205–1245, 2015

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

$$z = 2 \cdot \operatorname{arctanh}(2 \cdot \mathrm{Imp} - 1) \quad (2)$$

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 S3–S5 in the Supplement) would be tend to be larger. Here, however, we are not really interested in the magnitude or statistical significance of the individual effect, but just would like to see whether they are very different or not.

2.3.3 Prediction of pipe flows

To assess the model's capability to predict the resulting in-sewer flow (decisive for planning and design of urban drainage infrastructure), we compared the modelling result with flow data measured at the catchment outlet (see Sect. 3.3). To do so, we evaluated the model performance regarding the volume of the total runoff and the flow dynamics, particularly regarding the prediction of the peak flows. Main driving questions for the analysis were:

- How do differences in imperviousness affect pipe flow predictions?
- To what extend may differences regarding input data (imperviousness) be compensated by the model calibration procedure?

Model calibration

To adress 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 data set in a calibration (July to September 2014) and a validation period (September to November 2014). In 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.

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 four implementations, in which 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) the Nash–Sutcliffe-Efficiency (NSE, Nash and Sutcliffe, 1970), (ii) the total flow balance, and (iii) the deviation regarding the peak flows – all with respect to the flow at the catchment outlet. The input parameter "%imp" is 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.

Peformance assessment: flow balance and flow dynamics

In a first step, we evaluated the match between modelled hydrographs and reference flow data using the *Simulation Bias* and the Nash–Sutcliffe-Efficiency (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:

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

HESSD
12, 1205–1245, 2015

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:

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

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 peaks flows from observed and modelled hydrographs using a event filter that identifies independent rainfall–runoff events with an, at least, 6 h preceeding dry weather period.

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. Figures 5 and 6 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,

HESSD
12, 1205–1245, 2015

was not feasible. Post-processing of the imperviousness map might improve overall accuracy, but carries the danger of introducing unwanted biases.

3.2 Prediction of surface runoff

3.2.1 Exploratory analysis

We used boxplots and scatterplots to investigate the effect for the four combinations of data sources and processing methods on (i) the imperviousness and the surface runoff characteristics, (ii) peak flows, and (iii) runoff volumes (see Fig. 7).

- *Imperviousness (Imp)*: the boxplot shows that the overall distributions of imperviousness for 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 seems to deliver slightly larger imperviousness values for both data sources than the ML method.
- *Peak runoff (Peak)*: similar as for the imperviousness, the different image sources lead to very similar peak runoff values. In general, boosting seems to lead to slightly higher peak flows, which also have a slightly larger variance and slightly higher extreme values for a couple of catchments. Regarding the suitability of UAV images in rainfall–runoff modelling, there are no relevant 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 probably 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

HESSD
12, 1205–1245, 2015

seem to be associated with smaller runoff volumes. This is consistent, as they also seemed to be related to a lower imperviousness in the subcatchments.

In general, the relative differences between the different alternatives are very small, with average values of a few percent (see Fig. 7). 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.

Furthermore, the scatterplots of the different explanatory and dependent variables also 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 Fig. S1 in the Supplement). For the boosting classifier, we observe a weak positive correlation with the degree of imperviousness (see Fig. S2), which means that catchments which are rather impervious (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.

3.2.2 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 S3–S5).

For the *imperviousness*, as expected neither the image source nor the classifier are strongly correlated. The negative sign of the estimated slope parameter for the image source ($\beta_1 = -0.16$) 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 ($\beta_2 = 0.003$), 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.

HESSD
12, 1205–1245, 2015

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 ($\beta_1 = -0.6$) suggest that UAV images generally correlate with a lower imperviousness. Here, the influence of the image source seems to be euqally important as the classification method ($\beta_2 = -0.6$), 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 ($\beta_1 = -302$), whereas the RQE method shows a positive correlation ($\beta_2 = 298$), again, neither the two effects nor their interaction seem to be important.

In summary, the analysis suggests that the resulting surface runoff is not different for the different datasources or classification. 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.

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.

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 flow data.

1. Focusing on the results prior calibration, it becomes obvious that uncalibrated models, among each other, differ particularly regarding the peak flow performance (see boxplot in Fig. 8). This clearly corresponds to the findings of the surface runoff analysis (see Sect. 3.2) in which, for instance the implementation “UAV

HESSD
12, 1205–1245, 2015

ML" with the lowest 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 performance of the implementation of which the imperviousness is derived from UAV data classified the ML method (UAV ML) is assumed to occur by chance.

2. Results from calibrated models (see Fig. 9 and Table 2, right) show that conducting a detailed calibration, as expected, leads to an improved model performance (NSE increase, bias reduction) and interestingly compensates the imperviousness mapping deviations among the four implementations. This equalization becomes evident through a visual assessment of simulated hydrographs (see Fig. 9). Even though the results from the UAV ML implementation after calibration still shows slightly different results (see Fig. 9, right), a peak flow analysis comparing the absolute maxima of in-sewer flow for the 13 most intense rain events leads to very similar scatter patterns when cross-comparing the peak flow performance with reference data (see Fig. 10).

However, when analyzing the variation of final calibration parameter sets (see Fig. S6), it becomes clear that the best fit for each of the four model implementations is achieved by a significantly different parameter set. Particularly the parameter "width", "maximum infiltration rate" and "Decay K" (influencing the peak flow) vary significantly within the a priori defined parameter ranges. Ultimately, 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.

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 unconclusive 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 runoff prediction, the differences between different imaging platforms and between different processing methods are small. Even though the classification accuracy between data sets 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 possible reasons is the spatial size of

HESSD
12, 1205–1245, 2015

subcatchments. Each of them consists hundreds of image pixels, but the amount of impervious surfaces per subcatchment used in the hydrological model, is a sum of all impervious pixels belonging to this subcatchment. Thus, even if 20 % of pixels were classified incorretly, it might happen that it does not change the amount of impervious surfaces within a subcatchment. A further observation is that the differences in classification accuracy are 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 input.

4.2 Prediction of surface runoff

4.2.1 Exploratory analysis of surface runoff

While there are substantial differences when the images are compared pixel-by-pixel, these are largely lost for the predicted surface runoff. In our view, this is explained by the SWMM surface runoff model. It is a lumped model, which aggregates the pixels and thus is smoothing out the differences, already on this tiny 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 with flow measurements will also be larger. Future experiments that investigate the continuous downscaling of images may reveal when differences fully disappear.

4.2.2 Model structure as a bottleneck?

Obvious differences in the input data may be assimilated due to the simplified, conceptual representation of the surface runoff in SWMM. In case a pixel-based modelling approach for surface runoff is used, results might be different. In future, this might be even more important considering the increasing popularity of coupled 2-D-overland/1-D-channel flow models including more detailed overland-flow modelling using raster/pixel-based approaches (cf. Austin et al., 2014). Traditional models are not

HESSD
12, 1205–1245, 2015

ready yet to fully process the amount of detail (pixel basis) provided by such aerial images.

4.2.3 High-resolution images provide added value in urban drainage

In future investigations, the aspect of differentating between three or more land-use classes should be investigated. The effect on surface runoff and pipe hydraulics using the current lumped models may not be as immense. However, for the assessment of pollutant loads, which is usually strongly dependent of land-use characteristics, the accurate and up to date monitoring of land-use, i.e. feature recognition is more important. Relevance increases even more against the background the difficulty to obtain adequate reference data for pollution load modelling. It is generally harder to calibrate such models (cf. Dotto et al., 2014) implying the risk of making predictions without calibration.

Also, other urban drainage tasks would greatly benefit from detailed land-use maps, e.g. precise and justified stormwater fees due to exactly delineated roofs/impervious areas (see Fig. 5). An improved feature (gully pots, sewer inlets, curbstone structures) identification would provide valuable input data for network generation approaches 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 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. The analysis of the final calibration parameter values however reveals that the best fit for each of the implementations is achieved by differing parameter sets (see Fig. S6). 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 proce-

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

12, 1205–1245, 2015

Enabling high-quality observations of surface imperviousness

P. Tokarczyk et al.

dure, 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

HESSD
12, 1205–1245, 2015

be explained by input data variations. But still predictions are inaccurate. Conversly, 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.

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 spatial aggregation linked to the hydrological model approach (ensemble modelling).

We furthermore suggest using imperviousness maps consisting of three land-use classes as more differentiated 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 might more clearly show the potential of UAV datasets and advanced classification methods for more accurate urban drainage and pollution load modelling.

Acknowledgements. This publication is an outcome of a fruitful interdisciplinary collaboration between photogrammetrists and hydrologists, which was triggered by joint supervision of Matthew Moy de Vitry's Master thesis titled: "Improving Urban Flood Management with Autonomous Mini-UAVs". We would like to thank Matthew for providing us with a UAV dataset. Also, we are very grateful to MeteoSwiss and the city of Lucerne for providing us with the precipitation and infrastructure data, and Holinger company for providing us with the *swisstopo*

aerial image data. Last but not least, we would like to thank Philippe Gerber for helping us with an automatic calibration of pipe flow model.

References

Austin, R. J., Chen, A. S., Savić, D. A., and Djordjević, S.: Quick and accurate Cellular Automata sewer simulator, J. Hydroinform., 16, 1359–1374, 2014. 1227

Benbouzid, D., Busa-Fekete, R., Casagrande, N., Collin, F.-D., and Kégl, B.: MultiBoost: a multi-purpose boosting package, J. Mach. Learn. Res., 13, 549–553, 2012. 1216

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, J. Hydrol., 377, 300–316, 2009. 1208

Carlson, T. N. and Arthur, S. T.: The impact of land use – land cover changes due to urbanization on surface microclimate and hydrology: a satellite perspective, Global Planet. Change, 25, 49–65, 2000. 1208

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, Sensors, 8, 910–932, 2008. 1208

Dotto, C., Kleidorfer, M., Deletic, A., Rauch, W., and McCarthy, D.: Impacts of measured data uncertainty on urban stormwater models, J. Hydrol., 508, 28–42, 2014. 1228

Eisenbeiß, H.: UAV photogrammetry, Institut für Geodäsie und Photogrammetrie, Eidgenössische Technische Hochschule Zürich, Switzerland, 2009. 1209

ESRI: ArcMap 10.2.1, Environmental Systems Resource Institute, Redlands, California, 2013. 1216

Fankhauser, R.: Automatic determination of imperviousness in urban areas from digital orthophotos, Water Sci. Technol., 39, 81–86, 1999. 1208

Freund, Y. and Schapire, R.: A desicion-theoretic generalization of on-line learning and an application to boosting, in: Computational Learning Theory, vol. 904, Lecture Notes in Computer Science, Springer Verlag, Heidelberg, 23–37, 1995. 1216

Hughes, G.: On the mean accuracy of statistical pattern recognizers, IEEE T. Inform. Theory, 14, 55–63, 1968. 1215

Krejci, V., Fankhauser, R., Gammeter, S., Grottker, M., Harmuth, B., Merz, P., and Schilling, W.: Integrierte Siedlungsentwaesserung: Fallstudie Fehraltorf, Eidg. Anstalt für Wasserversorgung, Abwasserreinigung und Gewässerschutz EAWAG, Dübendorf, Switzerland, 1994. 1208

Li, W., Ouyang, Z., Zhou, W., and Chen, Q.: Effects of spatial resolution of remotely sensed data on estimating urban impervious surfaces, J. Environ. Sci., 23, 1375–1383, 2011. 1208

Lu, D., Tian, H., Zhou, G., and Ge, H.: Regional mapping of human settlements in southeastern China with multisensor remotely sensed data, Remote Sens. Environ., 112, 3668–3679, 2008. 1208

Montgomery, D. C. and Runger, G. C.: Applied Statistics and Probability for Engineers, Wiley, Hoboken, NJ, 2007. 1219

Montgomery, D. C., Peck, E. A., and Vining, G. G.: Introduction to Linear Regression Analysis, vol. 821, John Wiley & Sons, Hoboken, NJ, USA, 2012. 1218

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I: A discussion of principles, J. Hydrol., 10, 282–290, 1970. 1220

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. 1208

Parece, T. E. and Campbell, J. B.: Comparing urban impervious surface identification using Landsat and high resolution aerial photography, Remote Sens., 5, 4942–4960, 2013. 1208

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 Resour. Manag., 23, 1893–1915, 2009. 1208

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. 1209

Stankowski, S. J.: Population density as an indirect indicator of urban and suburban landsurface modifications, US Geological Survey Professional Paper, Washington D.C., USA, 800, 219–224, 1972. 1207

swisstopo: SWISSIMAGE: Das digitale Farborthophotomosaik der Schweiz. Eidgenössisches Departament für Verteitigung, Bevölkerungsschutz und Sport VBS, available at: http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/images/ortho/swissimage.

parsysrelated1.76752.downloadList.50684.DownloadFile.tmp/infosi201003deu.pdf (last access: 22 January 2015), 2010. 1209, 1212
swisstopo: swissALTI3D – Ausgabebericht. Eidgenössisches Departament für Verteitigung, Bevölkerungsschutz und Sport VBS, available at: http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/height/swissALTI3D.html (last access: 22 January 2015), 2014. 1213
Tokarczyk, P., Wegner, J., Walk, S., and Schindler, K.: Features, color spaces, and boosting: new insights on semantic classification of remote sensing images, IEEE T. Geosci. Remote, 53, 280–295, 2015. 1209, 1216, 1217
UN: World Population Prospects 1950–2050. The 2012 Revision. Key Findings and Advance Tables. United Nations Population Division, available at: http://esa.un.org/wpp/Documentation/pdf/WPP2012_%20KEY%20FINDINGS.pdf (last access: 22 January 2015), 2013. 1207
US-EPA: Storm Water Management Model (SWMM), Version 5.0.022. United States Environmental Protection Agency, available at: http://www2.epa.gov/water-research/storm-water-management-model-swmm (last access: 22 January 2015), 2010. 1214
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, Int. J. Remote Sens., 30, 4785–4806, 2009. 1208
Vrugt, J. A. and Robinson, B. A.: Improved evolutionary optimization from genetically adaptive multimethod search, P. Natl. Acad. Sci. USA, 104, 708–711, 2007. 1220
Weng, Q., Hu, X., and Liu, H.: Estimating impervious surfaces using linear spectral mixture analysis with multitemporal ASTER images, Int. J. Remote Sens., 30, 4807–4830, 2009. 1208
Zhang, C. and Kovacs, J.: The application of small unmanned aerial systems for precision agriculture: a review, Precis. Agric., 13, 693–712, 2012. 1209
Zhou, Y. and Wang, Y. Q.: Extraction of impervious surface areas from high spatial resolution imageries by multiple agent segmentation and classification, Photogramm. Eng. Rem. S., 74, 857–868, 2008. 1208

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).

	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

HESSD
12, 1205–1245, 2015

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

Model implementation	Model performance: Bias [–] / NSE [–] (prior calibration)	Model performance: Bias [–] / NSE [–] (after complex auto-calibration)
Ortho ML	2.0/0.54	3.16×10^{-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

HESSD
12, 1205–1245, 2015

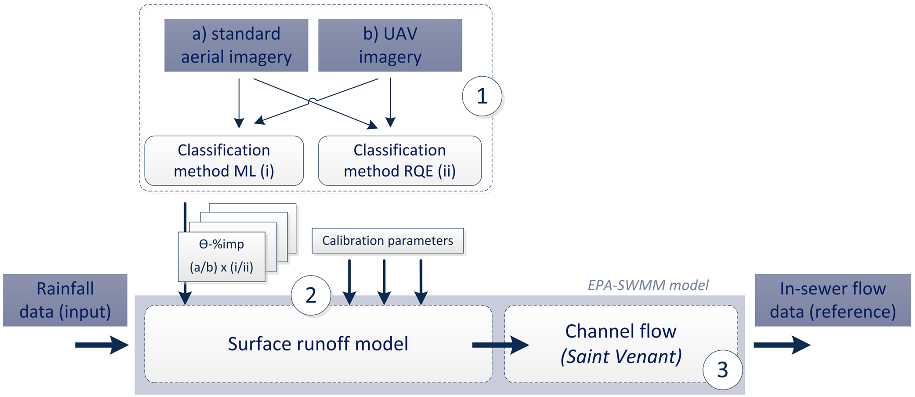

Figure 1. Overall analysis approach (⊖-%imp: model paramater "degree of imperviousness"; ML: maximum likelihood; RQE: boosting with randomized quasi-exhaustive feature bank).

HESSD
12, 1205–1245, 2015

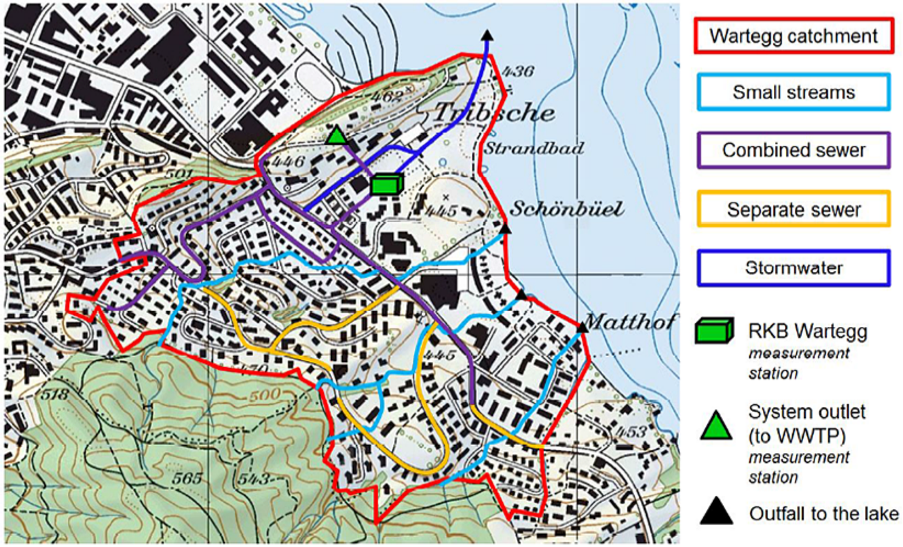

Figure 2. Case study catchment area situated in Lucerne.

HESSD
12, 1205–1245, 2015

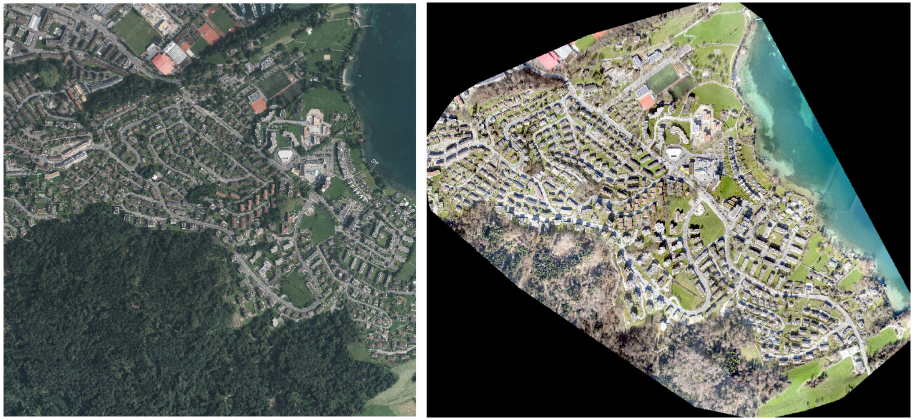

Figure 3. Image datasets. *swisstopo* (left) and UAV (right).

HESSD
12, 1205–1245, 2015

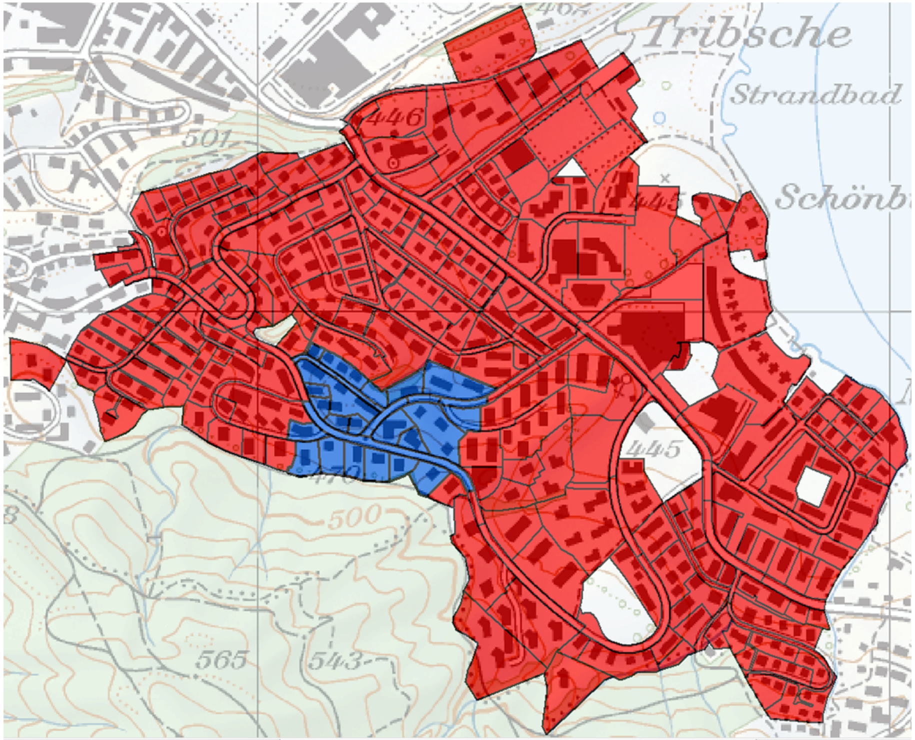

Figure 4. Wartegg area containing 307 subcatchments (red polygons including blue polygons) overlayed on a topographic map. The performance of classifiers was assessed on a subset depicted in blue.

HESSD
12, 1205–1245, 2015

Enabling high-quality observations of surface imperviousness

P. Tokarczyk et al.

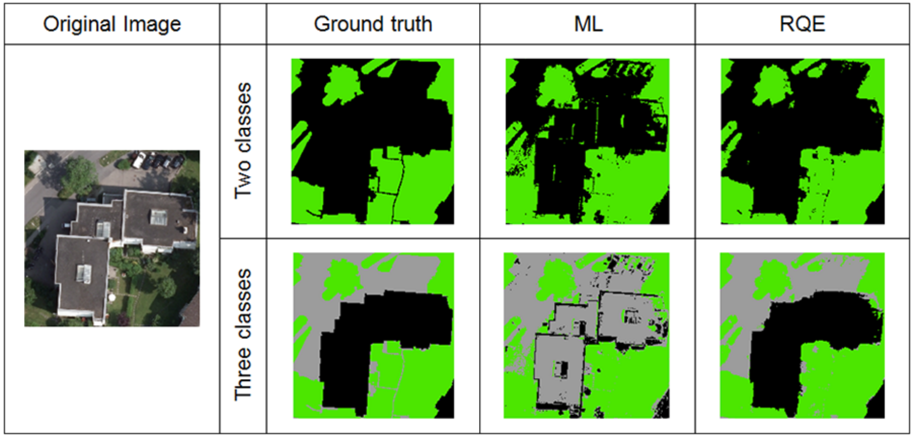

Figure 5. Cutouts of the swisstopo image: original image, manually labeled ground truth, and classification results using ML and RQE (two and three classes). In a case of two classes impervious surfaces are black and pervious are green. In a case of three classes rooftops are black, streets/sidewalks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

HESSD
12, 1205–1245, 2015

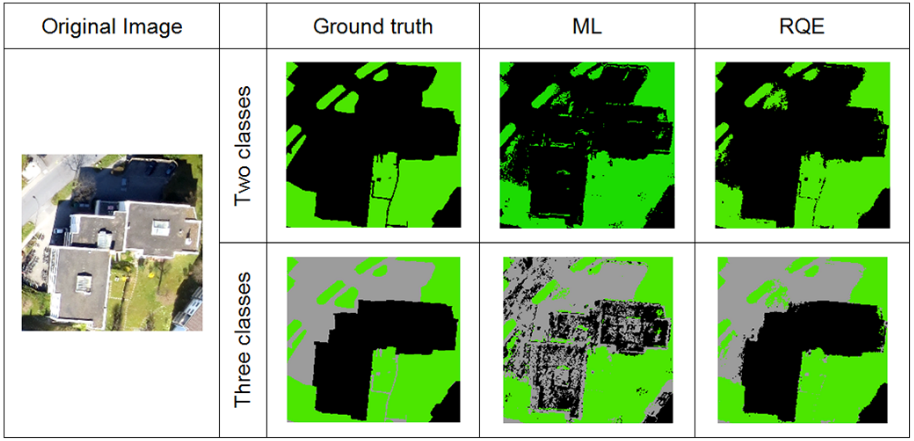

Figure 6. Cutouts of the UAV image: original image, manually labeled ground truth, and classification results using ML and RQE (two and three classes). In a case of two classes impervious surfaces are black and pervious are green. In a case of three classes rooftops are black, streets/sidewalks are grey and vegetation is green.

HESSD
12, 1205–1245, 2015

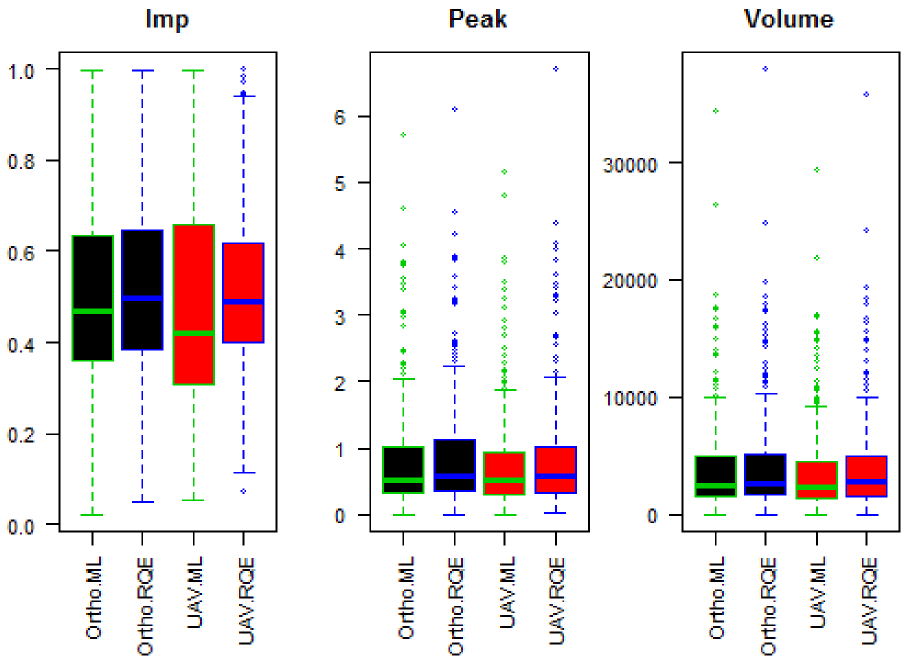

Figure 7. Boxplots of the imperviousness and surface runoff characteristics for the 307 subcatchments for the four combinations of data sources and processing methods. Black = Ortho, Red = UAV, Green = ML, Blue = RQE.

HESSD
12, 1205–1245, 2015

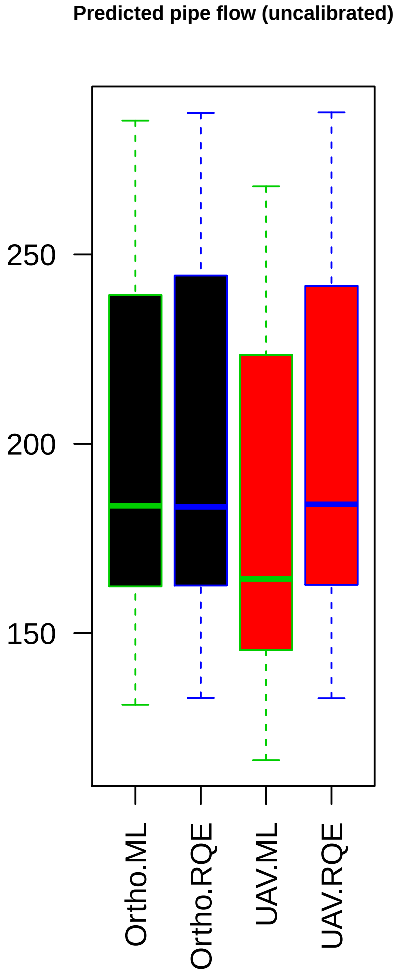

Figure 8. Evaluation of peak flows for the 13 most intense rain events (prior calibration).

HESSD
12, 1205–1245, 2015

Enabling high-quality observations of surface imperviousness

P. Tokarczyk et al.

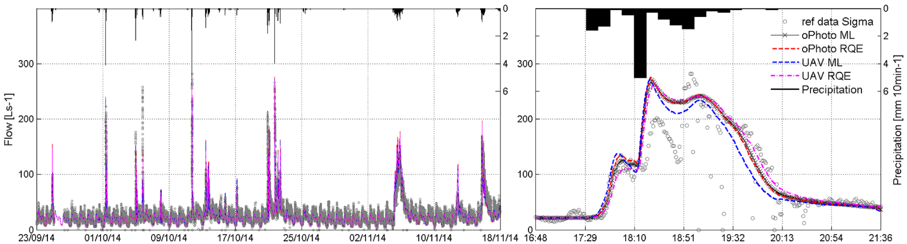

Figure 9. Observed reference and simulations (prior calibration) for the full validation period September to November 2014 (left) and a selected event on 11 October 2014 (right).

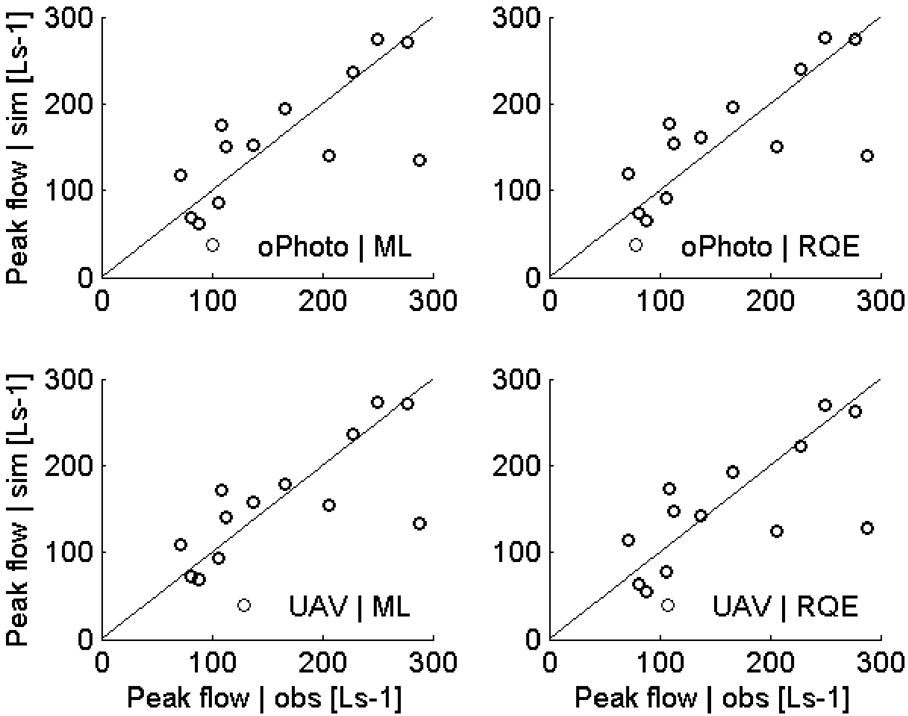

Figure 10. Evaluation of the peak flows for the 13 most intense rain events in the validation period (after calibration).

HESSD
12, 1205–1245, 2015