**Natural Hazards
and Earth
System Sciences**

Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data

S. Martinis, A. Twele, and S. Voigt

German Aerospace Center (DLR), Oberpfaffenhofen, Germany

Abstract. In this paper, an automatic near-real time (NRT) flood detection approach is presented, which combines histogram thresholding and segmentation based classification, specifically oriented to the analysis of single-polarized very high resolution Synthetic Aperture Radar (SAR) satellite data. The challenge of SAR-based flood detection is addressed in a completely unsupervised way, which assumes no training data and therefore no prior information about the class statistics to be available concerning the area of investigation. This is usually the case in NRT-disaster management, where the collection of ground truth information is not feasible due to time-constraints. A simple thresholding algorithm can be used in the most of the cases to distinguish between "flood" and "non-flood" pixels in a high resolution SAR image to detect the largest part of an inundation area. Due to the fact that local gray-level changes may not be distinguished by global thresholding techniques in large satellite scenes the thresholding algorithm is integrated into a split-based approach for the derivation of a global threshold by the analysis and combination of the split inherent information. The derived global threshold is then integrated into a multi-scale segmentation step combining the advantages of small-, medium- and large-scale per parcel segmentation. Experimental investigations performed on a TerraSAR-X Stripmap scene from southwest England during large scale flooding in the summer 2007 show high classification accuracies of the proposed split-based approach in combination with image segmentation and optional integration of digital elevation models.

1 Introduction

The demand for near real-time (NRT) information on natural disasters has increased considerably during recent years worldwide. Flooding is considered as the world's most costly type of natural disaster in terms of both property damage and human causalities. Because of their nearly all-weather day-night capabilities, ensuring data-gathering on every overpass, satellite Synthetic Aperture Radar (SAR) sensors are optimally suited for providing reliable information on extensive floods, which usually occur during long lasting precipitation and cloud cover periods. Flood information is needed as quickly and detailed as possible to provide an overview of the situation to improve crisis management and response activities. For that purpose an unsupervised flood detection approach particularly oriented to the analysis of single-polarized very high resolution SAR satellite data is presented.

Among supervised classification (e.g. Townsend et al., 2002), various change detection methods (e.g. Bazi et al., 2005; Brivio et al., 2002; Nico et al., 2000) and the application of active contour models (e.g. Mason et al., 2007), thresholding is one of the most popular image processing techniques to separate flooded from non-flooded areas in SAR imagery (e.g. Chen et al., 1999; Brivio et al., 2002; Matgen et al., 2007). Commonly, this classification is performed by assigning all elements of a SAR intensity image with a scattering cross-section lower than a given threshold to the class "flood". One of the main advantages of this approach is that it is computationally relatively inexpensive and therefore suitable for rapid mapping purposes. Its results are usually reliable and commonly, most of the extent of an inundation area can be derived. The applicability of thresholding procedures for floodplain detection using SAR sensors depends on the contrast between the flood and non-flood areas. Its

results are satisfactory for calm water surfaces, which can be regarded as specular reflectors with low backscatter values. In contrast, the surrounding terrain usually exhibits higher signal return due to increased surface roughness. Elevated surface roughness of water bodies due to the influence of waves, precipitation as well as diffuse and double bounce reflection at vegetation reduces the class-separability and also complicates the determination of an appropriate threshold.

Since a flood separating threshold value depends on many factors such as environmental and, satellite system parameters, it can be highly variable and has to be determined individually for each satellite scene. In most cases thresholding based derivation of inundation areas from remote sensing data is generated by visual inspection of the image histogram and subsequent fine tuning of the threshold. However, such manual trial-and-error operations typically turn out to be relatively time consuming; additionally, the quality of their results highly depends on the visual interpretation and hence subjective impression of the operator. Therefore, such interactive threshold determination lacks objectivity and traceability. Automatic thresholding procedures can overcome these problems. In the context of automatic image thresholding, e.g. Sahoo et al. (1988), Pal and Pal (1993) and Bazi et al. (2007) provide extensive methodological reviews. Thresholding algorithms are either based on global information (e.g. gray-level histogram of the whole data set) or on local information (e.g. co-occurrence matrix) of an image. Independently of the used algorithm, the use of a single threshold for every image element in an entire data set is called global thresholding (Pal and Pal, 1993). On the other hand the partitioning of a data set into subregions and the derivation of a threshold value for each of these subimages is termed local thresholding (Sahoo et al., 1988).

Most parametric thresholding algorithms determine automatically the desired gray value which best separates two semantic classes under the assumption that the a priori probabilities of the class-conditional densities in a histogram are sufficient to properly model the class-distributions. The spatial flood extent, in combination with the spatial resolution and the coverage of the used satellite sensor determine the percentage of the inundated area within a satellite scene. Therefore the classes "flood" and "non-flood" may show quite different a priori probabilities when looking at the entire dataset compared to looking at a subset of it. This may reduce the capability of any global thresholding algorithm to detect an adequate threshold in the histogram, built from the large SAR scenes, because local gray-level changes may not be distinguished.

Due to these limitations a split-based approach (SBA) for unsupervised flood detection in SAR data is applied which originally has been proposed by Bovolo and Bruzzone (2007) for the identification of tsunami induced changes in multitemporal imagery. This approach comprises a tiling of the satellite imagery into smaller subscenes of user-defined size and a successive local thresholding analysis of a limited amount of splits with a high probability to contain large quantities of image elements of the classes "flood" and "nonflood". Three thresholding approaches, all based on the KI minimum error thresholding algorithm (Kittler and Illingworth, 1986) are tested. Different combination techniques are applied to extract one global threshold from the analysis of the selected image splits. The global threshold is then integrated into a multi-scale segmentation process, which connects the advantages of small-, medium- and large-scale per parcel segmentation and hence should lead to a more accurate classification of the SAR data. Additionally, a high resolution digital elevation model (DEM) is used to improve the classification accuracy.

The effectiveness of the SBA in combination with subsequent classification refinement steps by multi-scale segmentation and DEM integration is analyzed using a large TerraSAR-X Stripmap scene from southwest England during a flood situation in July 2007.

The outline of the paper is as follows: in Sect. 2 the general SBA adapted to flood detection in SAR data is descibed. The application of the SBA to a TerraSAR-X scene of the River Severn (UK) and the subsequent classification refinement process is presented in Sect. 3. In Sect. 4, the results of the SBA by the use of three different automatic thresholding approaches as well as quality assessments of different refinement steps are reported. Finally, conclusions are drawn in Sect. 5.

2 A split-based automatic thresholding approach for unsupervised NRT-flood detection

2.1 Image tiling and split selection

The image-tiling technique divides a SAR scene X with dimensions R, C into N quadratic non-overlapping subimages X_n of user-defined size $s*s$. The selection of s and therefore N depends on the extent of the two classes "flood" and "non-flood" within X, and on the spatial resolution of the SAR scene. Due to the fact that the thresholding algorithms only extract proper thresholds values if the histogram is not unimodal, the selection process has to ascertain that the subscenes contain a distinct number of pixels from both classes. A minimum amount of 10% of each class seems to be sufficient for accurate threshold detection up to reasonable overlapped class mixtures according to empirical assumptions of Bazi et al. (2007) in automatic change detection. Appropriate splits are selected, combining the statistical properties of X and its subscenes. The coefficient of variation CV_{X_n} (ratio between the mean μ_{X_n} and standard deviation σ_{X_n} of the gray values of X_n) is used as an appropriate statistical measure to select splits with a wide data range. It serves as a measure of the degree of variation within the data and can therefore be used as an indicator of the probability that the splits contain more than one semantic class.

Erroneous selections may occur if a subset comprises both agricultural crop land and high reflecting urban areas. Thus, similar values of CV_{Xn} like for subsets containing the classes "flood" and "non-flood" are reached. A second measure R_{Xn} is added which selects the image splits in dependence of the ratio between the μ_{Xn} to the global intensity mean μ_X of X. This step prohibits the mis-selection of subsets comprising both agricultural crop land and urban areas characterized by similar values of CV_{Xn} as splits covering large parts of the classes "flood" and "non-flood". Finally, the following data ranges have been derived by the investigation of several SAR data sets exhibiting flood situations for optimal split selection:

$$X_{n'} = \left\{ X_n \,\middle|\, CV_{X_n} \geq 0.7 \wedge R_{X_n} 0.4 \leq \ldots \leq 0.9, n=1,\ldots,N \right\}. \quad (1)$$

Optionally, ancillary data like vector layers containing the position of the land-water boundary at normal hydrological conditions may optionally be integrated into the split-selection process. This information is in most cases already readily available (e.g. SRTM water mask) or can be generated from pre-flood satellite archive scenes. The split-selection process may be restricted to these subimages intersecting the vector file. This reduces the probability of selecting split-elements not containing any water surfaces.

Out of the total set N' of subimages $X_{n'}$ which fulfils Eq. (1) a limited number N'' of subimages $X_{n''}$ is finally selected for the threshold computation. The value of N'' can be manually set by the user before starting the algorithm. This selection is made to constrain and equalize processing times independently from N'. This final selection is carried out by the computation of the Euclidean distance $\Delta CVR_{Xn'}$ within a 2-D feature space determined by $CV_{Xn'}$ and $R_{Xn'}$ between the points of the respective subset and the mean $\mu CVR_{Xn'}$ of the cluster of $X_{n'}$. The N'' elements with the lowest Euclidean distance are selected:

$$X_{n''} = \left\{ X_1, X_{2'}, \ldots, X_{n'} \,\middle|\, \Delta CVR_{X_{1'}} \leq \Delta CVR_{X_{2'}} \leq \right.$$
$$\left. \ldots \leq \Delta CVR_{X_{n'}}, \, n \leq N'' \right\}. \quad (2)$$

If the case occurs that no splits are selected according to the predefined criterion, simultaneously the range of CV_{Xn} is decreased and R_{Xn} is increased stepwise by 0.05 until a minimum set of N'' splits is reached.

2.2 Automatic threshold selection procedure

The principal concept of using global thresholding algorithms for delineating flood zones in SAR intensity data is to assign all elements with a scattering cross-section lower or greater than a threshold τ to the class "flood" or "non-flood", respectively. This means that a $R*C$ dimensioned gray-scale image $X = \{x_{rc} : r=0, 1, \ldots, R-1, c=0, 1, \ldots, C-1\}$ with G possible gray-levels ($x_{rc} = g$, $g \in \{0, 1, \ldots, G-1\}$) will be transformed according to $\tau \in \{0, 1, \ldots, G-1\}$ into a binary form $Y = \{y_{rc} : r=0, 1, \ldots, R-1, c=0, 1, \ldots, C-1\}$ with following values:

$$y_{rc} = \begin{cases} 1 & x_{rc} \leq \tau \\ 0 & x_{rc} > \tau \end{cases}, \quad (3)$$

where 1=flood, 0=non-flood .

Difficulties may arise within the global thresholding procedure if the illumination varies across the image due to incidence angle linked backscatter variances or atmospheric conditions. However, within TerraSAR-X data, especially Spotlight and Stripmap products, the system induced effect is marginally existent due to only small variations of the incidence angle from near- to far-range (3.562° maximum at a ground swath width of 31.638 km). Therefore these system-related backscatter variations can be neglected. Despite the fact that the all-weather imaging capabilities of SAR sensors are often emphasized in the literature, heavy precipitation in combination with the relatively short wavelength of the actively emitted X-band microwaves may cause artefacts which appear as dark patches due to the attenuation of the transmitted signal as well as bright objects due to partial backscattering at rain-cells. However, these effects are very rare and appear just locally. Likewise, these exceptions can be disregarded.

In this study three bi-level thresholding techniques are tested for the distinction between floodwater surfaces and non-flood areas. The first is the original KI algorithm itself. The second one computes the most prominent (global) minimum (GM) of the image histogram starting at the derived minimum error cutting point. The third approach uses an index function, combining the separability of the two distributions and the most dramatic change in the KI's criterion function, to obtain the optimal threshold by the use of a quality index (QI).

2.2.1 Kittler and Illingworth's algorithm

In this work the KI thresholding algorithm (Kittler and Illingworth, 1986) has been adopted, which has been used in remote sensing image analysis predominantly in modified versions for automatic change detection in difference or (log) ratio data (Bazi et al., 2005; Melgani et al., 2002; Bovolo and Bruzzone, 2007; Moser et al., 2006). This is a global parametric thresholding technique which uses a minimum error approach to group the sets of pixels of gray-scale images into object and background classes and assumes that the image histogram $h(g)$ which gives the frequency of occurrence of the various levels of g to be the only available information about the image. The histogram is viewed as an estimate of the class-conditional probability density functions of a mixture of two clusters. We assume that the histograms of the selected subimages $X_{n'}$ can be modelled statistically by two 1-D normal distributions of the semantic classes "flood" and "non-flood" $p(g \mid i)$, $i=\{1:$ flood, $2:$ non-flood$\}$, with param-

eters mean μ_i, standard deviation σ_i and a priori probability P_i, so that

$$p(g) = \sum_{i=1}^{2} P_i \, p(g \mid i) \qquad (4)$$

where

$$p(g \mid i) = \frac{1}{2\pi \sigma_i} e^{-(g-\mu_i)^2/2\sigma_i^2}. \qquad (5)$$

Due to the fact that these parameters are usually not known in advance they have to be estimated at some arbitrary threshold level T which separates the two pixel populations with the following properties (Kittler and Illingworth, 1986):

$$P_i(T) = \sum_{g=a}^{b} h(g) \qquad (6)$$

$$\mu_i(T) = \frac{\sum_{g=a}^{b} g * h(g)}{P_i(T)} \qquad (7)$$

$$\sigma_i(T) = \frac{\sum_{g=a}^{b} (g - \mu_i(T))^2 * h(g)}{P_i(T)} \qquad (8)$$

where $a=1$ if i=flood, $a=T+1$ if i=non-flood, $b=T$ if i=flood and $b=G$ if i=non-flood.

The selection of an appropriate threshold τ is based on the optimization of a given predefined criterion function $J(T)$ that averages the cost function $c(g, T)$ over the histogram:

$$J(T) = \sum_{g=0}^{G-1} h(g)c(g, T) \qquad (9)$$

The cost function measures the cost of classifying pixels by comparing their gray-levels with the threshold T. According to the number of brightness levels of an image up to G different threshold values can be tested in succession. For each brightness value, the fitting criterion $J(T)$ is calculated, which is defined by

$$J(T) = 1 + 2 \left[P_1(T) \log \sigma_1(T) + P_2(T) \log \sigma_2(T) \right] - \\ 2 \left[P_1(T) \log P_1(T) + P_2(T) \log P_2(T) \right]. \qquad (10)$$

As long as threshold T is varied, the models of the Gaussian distributions change. The better the model fits the data, the lower is the criterion of this cost function. Therefore, the brightness value T, which reduces the criterion function, is considered to be the optimal threshold τ_{KI} since at this position, the classification error is minimized according to the Bayes classification rule, i.e., the number of mis-segmented pixels is smallest:

$$\tau_{KI} = \arg \min_{T} J(T) \qquad (11)$$

One has to note that τ_{KI} corresponds to the intersection point of the two Gaussian populations. If an overlap between the classes exists, as it is usually occurs in remote sensing data, the tails of these underlying distributions are truncated by the thresholding procedure and therefore the models $h(g \mid i)$ will be biased estimates of the true mixture components.

2.2.2 Global minimum thresholding

The second strategy used for automatic object-background separation is a bottom of the valley approach which locates, if existent, the most pronounced global minimum (GM) in a subhistogram. This belongs to the position between adjacent peaks of a bimodal histogram which often is chosen as a starting threshold in manual trial-and-error approaches by visual inspection of the image histogram. Based on the position of the threshold value τ_{KI} which is located between the two global maxima of a bimodal histogram the most pronounced valley between the two distributions "flood" and "non-flood" can be derived, primarily computing the direction in which the subhistogram declines in the surrounding of the start value τ_{KI} and moving step wise to this location τ_{GM} where the global minimum occurs.

2.2.3 Quality-index thresholding

The third method proposed for automatic partitioning of bimodal histograms derives the threshold according to a quality measure. For that purpose we use an index function $I[h(g)]$, introduced by Miasnikov et al. (2004) as a measure for the usefulness of bi-partitioning a 1-D projection of a dataset at the position of a computed threshold. $I[h(g)]$ is a composite measure containing a term sep describing the separability between the peaks of two distributions in relation to their means μ_i and variances σ_i^2 as well the term depth which accounts for the largest variation in KI's criterion function $J(T)$:

$$I[h(g)] = \text{sep} * \text{depth} \qquad (12)$$

where

$$\text{sep} = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}. \qquad (13)$$

The depth of $J(T)$ is originally given by the difference between the value of $J(T)$ calculated at the minimum error threshold τ_{KI} and the closest local maximum $J(T_{max})$ of the KI function. In this study $I[h(g)]$ is used for the search of the threshold τ_{QI} which best separates the class distributions "flood" and "non-flood". This is identical to the position where $I[h(g)]$ is maximized:

$$\tau_{QI} = \arg \max I[h(g), T] \qquad (14)$$

The calculation of the closest local maximum is accomplished for each of the two class peaks of $J(T)$ separated by the calculated global minimum at τ_{GM}.

2.2.4 Split combination techniques

To derive a spatially coherent binary flood mask without any discontinuance at the boarders of adaptively thresholded splits one global reliable threshold is used for classification.

For its calculation two approaches can be applied using the split inherent information (Bovolo and Bruzzone, 2007):

mathematical combination techniques of (sub)histogram de-rived thresholds and a (sub)histogram-merging strategy with subsequent thresholding.

The principal concept of the first method is to compute for every (sub)histogram of the selected splits $X_{n''}$ a local threshold τ_{loc} by any of the proposed KI, GM or QI thresholding algorithms. To obtain a global threshold τ_{glob} which is used for the classification of X simple mathematical operations can be applied; e.g. the computation of the median or the arithmetic mean. Due to the fact that all values of the variables are integrated with the same weight within the computations, the arithmetic mean contains more information than the median. However, it is more sensitive to outliers appearing just at one side of the range of values, especially in the presence of just a few samples. This may produce improper final threshold values due to the limited number N'' of selected splits $X_{n''}$, even if a robust split-selection procedure is applied.

The second method avoids the use of mathematical combination techniques by computing τ_{glob} directly from a merged histogram which is a combination of the distributions of the individual splits $X_{n''}$. However, the accuracy of the threshold selection may be reduced by differences in the statistical properties of the class-distributions in the respective splits.

3 Application for high resolution TerraSAR-X data at the River Severn, UK

In this section the unsupervised flood detection approach is applied exemplarily to a TerraSAR-X data set acquired during heavy floods in southwest England at the River Severn in 2007. The proposed workflow (Fig. 1) contains the SBA for automatic threshold computation of the pre-processed SAR scene as well as a classification refinement process including a multi-scale segmentation and an optional DEM integration step.

3.1 Data set

A period of heavy precipitations over the Gloucestershire Region in Southwest England during the summer season 2007 caused severe flooding at the River Severn. The record flood level at Tewkesbury was measured at 5.43 m on 22 July 2007 which was 0.13 m above the previously highest record from the year 1947. The proposed SBA is applied on a TerraSAR-X Stripmap scene (Fig. 2a) with 3.25 m pixel spacing in range and azimuth direction showing the flood situation along a section of about 69 km length of the River Severn on 25 July 2007 (06:34:08 UTC) including the confluence of River Avon coming from northeast. The data are HH polarized, which is the favoured polarization in flood mapping, because it provides the most obvious discrimination between open water and non-flood areas since it is less sensitive to small scale roughness of waves on the water surface than VV like-polarisation or cross-polarisations (Henry et al., 2006).

At the time of the satellite overpass there was no precipitation and low wind speed (16.5 km/h). Additionally, two ancillary data sets are available for this study: orthorectified aerial photographs with a spatial resolution of 0.2 m recorded about half a day prior to the TerraSAR-X overpass (24 July 2007 between 13:10:00 and 15:40:00 UTC) by a Rollei Modular AIC as well as a LiDAR-DEM with a horizontal resolution of 2 m and a vertical accuracy of 0.1 m.

3.2 Pre-processing

A preliminary despeckling was performed by adaptive filtering to reduce salt-and-pepper structures within the data. The speckle appearing on SAR imagery is a natural phenomenon generated by the coherent processing of radar echoes (Lee, 1986). Its presence not only reduces the interpreter's ability to resolve fine details, but also makes automatic segmentation of such noisy images difficult. For this purpose the Gamma-MAP filter (Lopes et al., 1990) was chosen because it comprises a good trade-off between the quality of the filtering result and the computation time. The filtering was accomplished by a mask-size of 3×3 pixels, which seems to be a sufficient size for this data set.

For validation purposes of the results of the proposed SBA and refinement process a reference map was created for a small subset (approx. $5.4 * 10^6$ pixels) of the TerraSAR-X scene by visual interpretation and manual digitisation of the orthophotos, which is covered by 23.8% of the class "flood" and 76.2% of the class "non-flood". The subset of the reference scene (see Figs. 2a and 5a) was chosen based on the overlap between the SAR data, the aerial images and the DEM. It comprises the heavily flooded Tewkesbury region.

3.3 Split-based approach

The SBA is applied to a TerraSAR-X scene X with dimensions of R=14 461, C=20 153 pixels. According to a split size of s=500, X is tiled into 1187 subimages, whereof a quantity N of 637 subimages contains no pixels with no-data values. An amount of N'=7 splits (Fig. 2a and b) were selected out of N according to the criterion defined in Eq. (1). These remaining subscenes are considered as suitable for effective threshold detection. This assumption can be confirmed by the fact that selected subimages intersect the visually detectable land-water boundary (see Fig. 2a). Out of N' a fixed number of N''=5 splits are sorted out according to Eq. (2) to perform the final threshold selection (Fig. 2a and b). The three proposed thresholding algorithms (KI, GM and QI) are applied to the five splits and different combination techniques for the derivation of one global threshold are executed. The performing of the thresholding algorithms as well as the results is presented in Fig. 3 exemplarily by split X''_4. For this split the KI algorithm computes a threshold value τ_{KI}=26 where $J(T)$ reaches its minimum. Based on τ_{KI} the global minimum τ_{GM}=29 of the histogram is localised. The
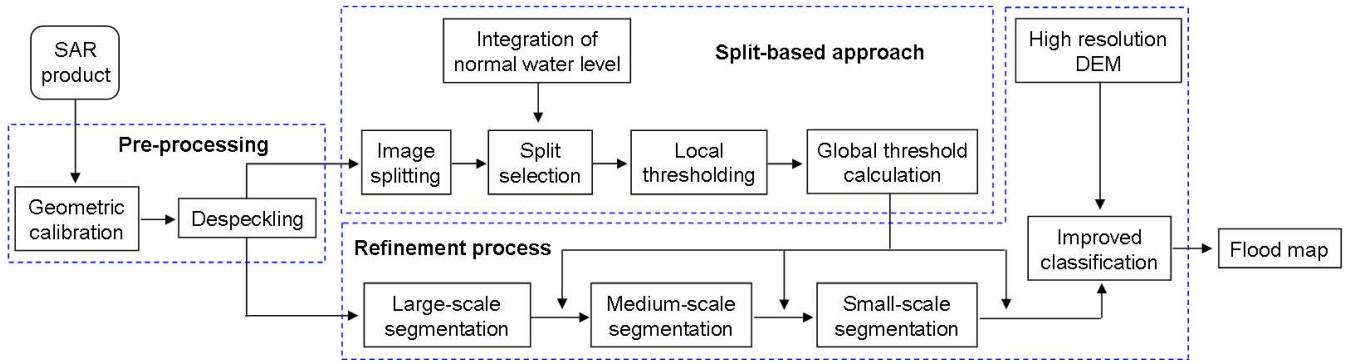
Fig. 1. Block scheme of the proposed workflow for automatic flood detecting within large SAR satellite data.
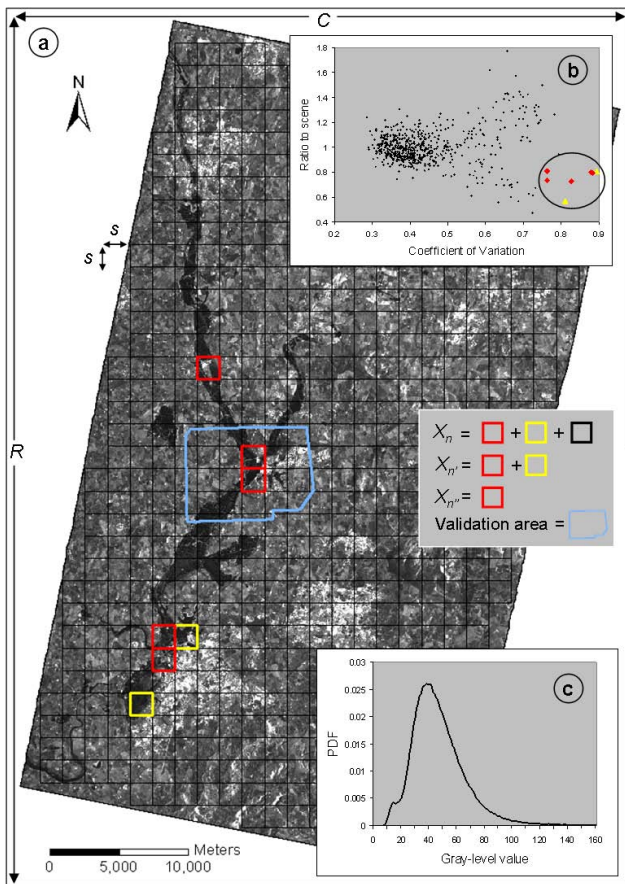

Fig. 2. (a) TerraSAR-X data of the River Severn (25 July 2007) with splits $X_{n'}$, $X_{n''}$ and outline of the validation area, **(b)** Location of X_n, $X_{n'}$ and $X_{n''}$ within the feature space ($CV_{Xn'}/R_{Xn'}$), **(c)** PDF of X.

3.4 Classification refinement process for unsupervised flood detection

The classification refinement process has the purpose to improve the accuracy of the classification result derived by global thresholding. It consists of two main parts: a multiscale segmentation and an optional DEM integration step.

3.4.1 Multi-scale segmentation

Pixel-based and segmentation-based classification techniques can be considered as main concepts for the identification of flooded areas in radar imagery. Conventional classification approaches use pixels as smallest geometrical components of raster data. However, the parameters which can be employed for the classification are limited. Additionally, pixel-based classifiers do not make use of spatial information of the image and are thus not suited in cases where land cover units are composed of spectrally heterogeneous elements. By the use of segmentation techniques, some disadvantages of pixel-based image thresholding can be solved. Image segmentation partitions an image into uniform regions based on similarity criteria of gray values or textural properties (Meinel and Neubert, 2004). The created homogeneous, non-overlapping segments have a strong correlation with real objects or areas of the earth's surface. Due to the increasing spatial resolution of earth observation imagery per-parcel approaches gain in importance in the field of remote sensing image analysis. Especially for data of the new generation of SAR sensors with up to one meter pixel spacing (TerraSAR-X, COSMO-SkyMed) the use of segmentation-based methods appears promising. These images can exhibit very high spectral variances within a given thematic class due to both the reduced mixed pixel phenomenon and the SAR intrinsic speckle effect. In addition to spectral related characteristics of the image objects further parameters such as contextual information, texture and object geometry can be used for improving classification accuracy.

However, the quality of the classification result depends on the properties of the calculated image segments. Gener-
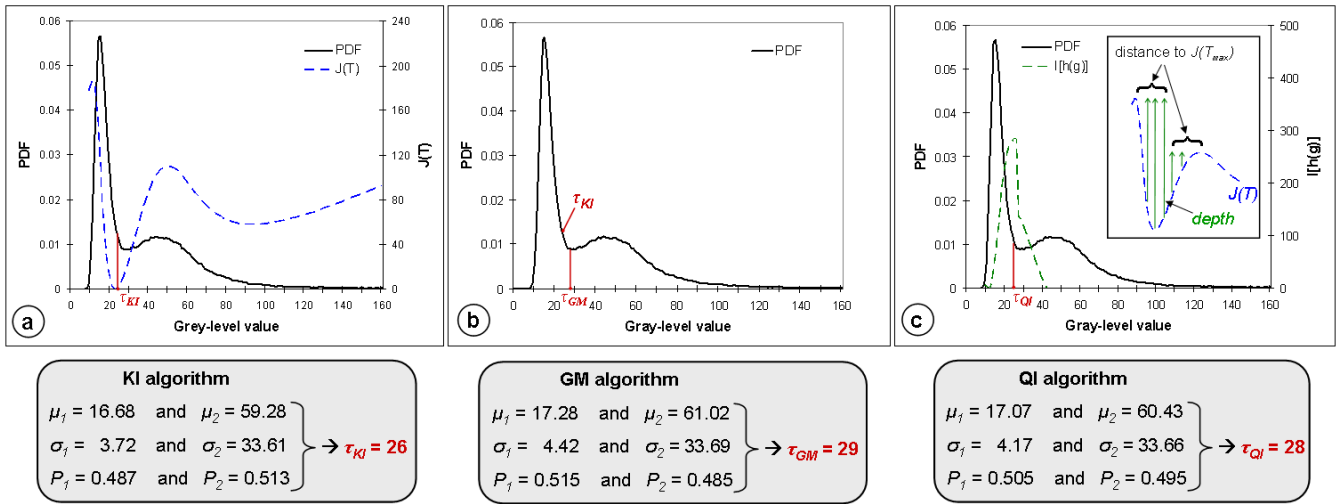
criterion function derived by the KI as well as τ_{GM} are used to calculate the threshold position $\tau_{QI}=28$ where the index function reaches its maximum.

Fig. 3. Histogram of split $X_{4''}$ and derived threshold values by the use of the **(a)** criterion function of the KI algorithm, **(b)** GM algorithm, c) Index function of the QI algorithm.

ally, high homogeneity segmentation parameters cause small segments, which, similar to pixel-based classifiers, produce noisy classification results but are able to detect small objects, whereas low values of the homogeneity criterion generate large segments at the expense of fine detail. Therefore, segmentation algorithms connecting the advantages of multi-scale segmentation and per-parcel classification by automatic thresholding seem promising for improved flood mapping.

Image objects were created using the Fractal Net Evolution concept of the Definiens Developer software (Baatz and Schäpe, 1999; Definiens AG, 2008). An iterative bottom-up segmentation algorithm based on a pair wise region merging technique is employed to segment the image according to the maximum allowable heterogeneity of the image objects, defined by an initially set scale parameter. This scale parameter is a homogeneity criterion h combining spectral h_c and shape homogeneity h_s, with h_s being composed of the smoothness $h_{compact}$ and compactness h_{smooth}. The fusion of the objects within the segmentation process follows the local mutual best fitting algorithm. This leads to the lowest increase of object heterogeneity by merging two adjacent objects which exhibit the lowest distance in a feature space determined by mean spectral values or texture features. For classification refinement we developed an approach containing a threefold processing step which alternates multi-scale segmentation and classification by the SBA-derived global threshold τ_{glob}:

1. Large-scale segmentation step and thresholding by τ_{glob}.

2. Medium-scale segmentation of objects of class "non-flood" and thresholding by τ_{glob}.

3. Small-scale segmentation of objects of class "non-flood" and thresholding by τ_{glob}.

Step 1 partitions the input data X into an initial set of large image objects containing some variations in the spectral properties of the merged pixels. Subsequently, the created objects are thresholded by τ_{glob}. Most of the inundation area is classified by this step, however fine tuning is reached by the following two steps, progressively enforcing the spectral homogeneity constraints of non-flood objects in a defined neighbourhood around flood objects and thresholding by τ_{glob}. A scale parameter of $h=75$ (unitless) was selected for the initial segmentation of X in step 1 based on visual interpretation of the image segmentation results with different scale parameters. This value was considered appropriate as reaching the best compromise between maximal possible heterogeneity and minimal homogeneity within the resultant object primitives, leading to an average object size of 2995 pixels. Steps 2 and 3 were performed by a reduced scale parameter of $h=40$ (average object size: 908 pixels) and $h=5$ (average object size: 16 pixels) respectively to detect fine detailed flood objects. The composition of homogeneity was controlled by both spectral and shape properties. Spectral homogeneity was set to an overall spectral factor of $h_c=80\%$. Shape-homogeneity criteria included an overall shape factor of $h_s=20\%$ which was subdivided into smoothness ($h_{smooth}=10\%$) and compactness ($h_{compact}=10\%$). Medium- and small-scale segmentation was accomplished within a distance of max. 5 objects and 1 object respectively to the class "flood". This restriction prohibits the flood mapping of medium- and small-scale image elements with too large distance from the flood objects and additionally reduces the computational requirements of the segmentation. The settings for the scale parameter h, h_c/h_s and $h_{smooth}/h_{compact}$ are defined as constants so that they need not be set by the user.

3.4.2 DEM integration

We observe that thresholding is capable of detecting the majority of flooded surfaces in X, but there are multiple areas of relatively high backscatter surrounded by flood objects, which are flooded with high probability but interpreted as non-flooded. These areas correspond predominantly to flooded surfaces which are broken by emergent vegetation. With X-band SAR flooded and non-flooded vegetation with a closed canopy can hardly be differentiated, due to very similar radar cross-sections. Additionally, some wrongly labelled flood regions exist with a small surface roughness and therefore low backscatter characteristics similar to calm water like streets, smooth agricultural crop land or radar shadow. For a hydrological plausible improvement of the classification result the topographic information of a high resolution DEM is added into the classification process as it can be used to extract flooding under dense vegetation and forest cover (e.g. Wang et al., 2002; Mason et al., 2007) and within residential areas. The proposed approach is appropriate for floods primarily caused by over-bank flow; predominantly affecting areas connected to the main river channel and also can be applied if no river gauge readings are available for hydraulic floodplain modelling. A prerequisite for this method is that major parts of the flood surface are already detected by the automatic thresholding step, thus a spatially coherent flood surface can be derived. The classification procedure using the DEM can be separated into two parts: an inclusion and an exclusion step. During the inclusion step, topography information can be used in combination with contextual information to classify potentially flooded objects with backscatter intensities higher than the originally defined threshold, e.g. protruding vegetation or urban areas into the flood class. First, all classified flood objects without any conjunction to the core flood area and exceeding a predefined elevation H are temporarily excluded from the classification result:

$$H = \left(\sum_{s_f=1}^{S_f} S_h \bigg/ S_f \right) + 1.5 * \sigma_{S_h} \qquad (15)$$

where S_f is the number of flood objects, S_h the elevation of each single flood object and σ_{sh} the standard deviation of S_h.

After that, an iterative classification process is performed considering local neighbourhood relations: If an object of the main flood area is found next to a non-flood object the height distance between the non-flood objects and the surrounding flood objects is calculated. In the case that the mean altitude of the non-flood object is lower or equal than the elevation of the contiguous flood objects, the non-flood object is recoded as flooded area. This processing step is repeated until no further non-flood object fits to this criterion. In case microtopographic features exist (e.g. embankments, walls) that may prevent flooding of areas at lower elevations, the integration of objects stops if these obstructions with higher elevations are detected as non-flood objects. If the segmentation fails

to delineate such objects or the resolution of the DEM is too coarse classification errors may occur in this processing step.

At the exclusion step, wrongly masked objects in regions with higher elevation than the main flood area can be eliminated. However, due to the fact that most parts of the floodplain are connected to the main river channel which is a body of flowing water, the altitude of the land-water boundary decreases downstream. Therefore using a global threshold for the exclusion step is not effective and a value which is representative for each position along the river is required instead. Accordingly, the altitude of every single flood object without connection to the main expanse of the flood area is compared to the altitude of the nearest object of the core flood area and removed from the mapping result if it exceeds a height difference of 1.0 m. This value was chosen to account for slightly higher objects which should be retained due to local variations in water height caused by the dynamic nature of floods.

4 Results

In this section, the performance of the proposed automatic approach of unsupervised threshold selection as well as the subsequent classification refinement process containing threefold image segmentation and an optional DEM integration are assessed.

The results of the threshold computations by the use of the KI, GM and QI algorithm are presented in Table 1. It was possible to model the histogram of each selected split by a mixture of two Gaussian distributions. Thus, by any of the applied algorithms a threshold value could be derived. In comparison, no decision threshold could automatically be identified from the histogram of X (Fig. 2c) because of the small a priori ratio at the visual detectable global minimum between the class "flood" (P_1=0.023) and "non-flood" (P_2=0.977), and the nearly unimodal distribution of the data set.

The local threshold positions of the selected subscenes computed by the KI, QI and GM algorithms exhibit slightly different values. This results from the varying class-statistical properties in the respective splits. The results derived by the QI algorithm are characterized by the lowest range of threshold values (26.0 to 29.0) which indicates less sensitivity to class statistics. A total amount of nine global threshold values (Table 1) have been derived by combining the three thresholding algorithms (KI, GM, QI) with three different split combination techniques (cf. Sect. 2.2.4), leading to seven different values of τ_{glob} in the range of 22.6 to 29.0. According to this a subset of X (see Fig. 2a), for which the real flood extent is known, has been classified by these seven thresholds with different segmentation adjustments:

– P: chessboard segmentation with an object size of 1 pixel ($\triangleq$ pixel based application);

Table 1. Local thresholds computed by the KI, GM and QI algorithm for splits $X_{n''}$ and global thresholds derived by the use of different split combination techniques.

Thresholding approach	Subscene $X_{n''}$					(sub)histogram threshold combining		(sub)histogram-merging
	$X_{1''}$	$X_{2''}$	$X_{3''}$	$X_{4''}$	$X_{5''}$	Mean	Median	
KI	25	21	17	26	24	22.6	24.0	23.0
GM	29	26	23	29	29	27.2	29.0	29.0
QI	26	27	29	28	26	27.2	27.0	25.0

- L, M, S: uni-scale segmentation at large-, medium- or small-scale;

- L-M: bi-scale segmentation at large- and medium-scale;

- L-M-S: tri-scale segmentation at large-, medium- and small-scale;

- L-M-S-DEM: tri-scale segmentation with subsequent DEM integration.

In Fig. 4 the accuracy assessment of the SBA and the classification refinement steps for each threshold value is illustrated by the missed alarm rate (MDR; i.e. the percentage of flood reference pixels erroneously detected as non-flooded), false alarm rate (FAR; i.e. the percentage of non-flood reference pixels erroneously detected as flood) and overall error rate (OER; i.e. the percentage of erroneously labelled reference pixels). Generally, a trend can be stated that a rise of the threshold value simultaneously is related to an increase of the FAR and a decrease of the MDR. Thus, reduction of false and missed alarms become two partially conflicting objectives. Regarding the classification accuracy of every single threshold derived by the SBA, pixel-based thresholding in all cases produces higher OERs than segmentation-based classification. This observation can be explained by the huge amount of false alarms which cannot by compensated by a low quantity of missed alarms in relation to uni-scale and bi-scale segmentation for τ_{glob}=22.6 to 27.2. Large-scale segmentation (L) yields much better results than small- (S) and medium-scale (M) image decomposition due to the highly reduced number of false alarms which exceed the rise in missed alarms. These missed alarms of L further are strongly reduced by L-M and more considerably by L-M-S, due to the fact that finer details on medium- and small-scale level are identified as flooding in a defined neighbourhood of large-scale flood objects.

L-M-S leads to lowest OERs by classification without ancillary data due to the fact that it combines the advantages of small-scale (low number of missed alarms) and large-scale (low number of false alarms) segmentation.

The integration of topographic information within the classification process of the tri-scale segmentation additionally reduces the percentage of the MDR. This results in the best classification accuracy of L-M-S-DEM out of all segmentation adjustments for all determined threshold values.

When comparing the scale-dependent classification results of the individual threshold values, the minimum pixel-based OER (8.68%) appears at τ_{glob}=23.0, which corresponds to the threshold position derived by the KI algorithm applied to the mixed pixel population of the merged (sub)histograms of $X_{n''}$. In contrast values of $\tau_{glob}<$ and >23.0 (τ_{glob}=22.6) show higher pixel-based OERs.

The difference in OER between P and L-M-S increases with rising threshold values. Additionally, by the use of segmentation, the range of OER between the different threshold values can be reduced. For example the range between the highest and lowest OER at pixel-based classification reaches 3.83% whereas the overall error of L, L-M, L-M-S and L-M-S-DEM just varies between 0.30% and 0.17%. This means that independent of the applied thresholding algorithm and split-combination technique, the variation of the classification accuracy can be reduced by appropriate segmentation. Thus, the method chosen for automatic thresholding is less critical when applied to segmented data. Lowest classification errors for M, L, L-M, L-M-S and L-M-S-DEM are reached when τ_{glob}=25.0, which also produces the best total classification result (OER: 4.55%) at L-M-S-DEM. The threshold value of 25.0 is reached by the QI algorithm applied to the histogram of the merged distributions of $X_{n''}$. The highest of the derived threshold values (τ_{glob}=29.0) also results in the highest OERs for every classification result without DEM integration due to a high FAR. This threshold has been computed by the GM algorithm in combination with (sub)histogram merging. The threshold value is overestimated due to the loss of a clear global minimum in the data set by the combination of the different class populations of $X_{n''}$. However, using τ_{glob}=29.0, the OER comes close to the best classification result at L-M-S-DEM. This is due to the following facts: the relatively high OER of 7.52% at L-M-S is fairly reduced by the exclusion step of the DEM integration process and the largest amount of correctly classified pixels at L-M-S (MDR: 3.46%) generates a more compact core flood area, which helps to identify more easily flooded vegetation and anthropogenic objects during the DEM integration step.
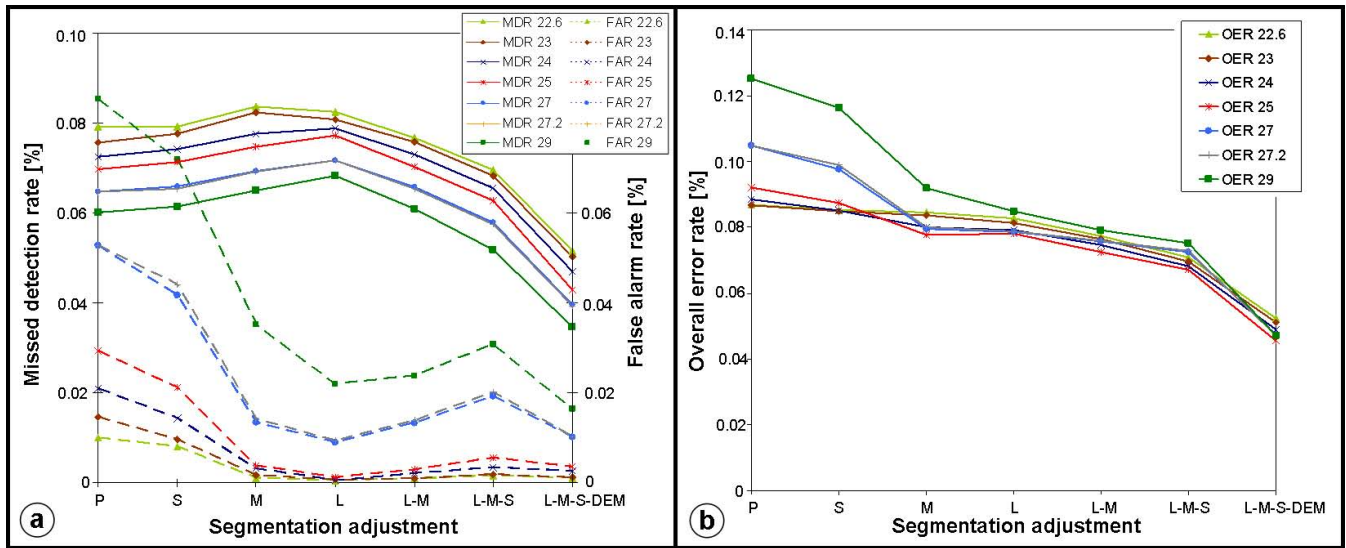
Fig. 4. (a) MDR and FAR for class flood and **(b)** OER calculated for the resultant global threshold values by the use of different segmentation scales.
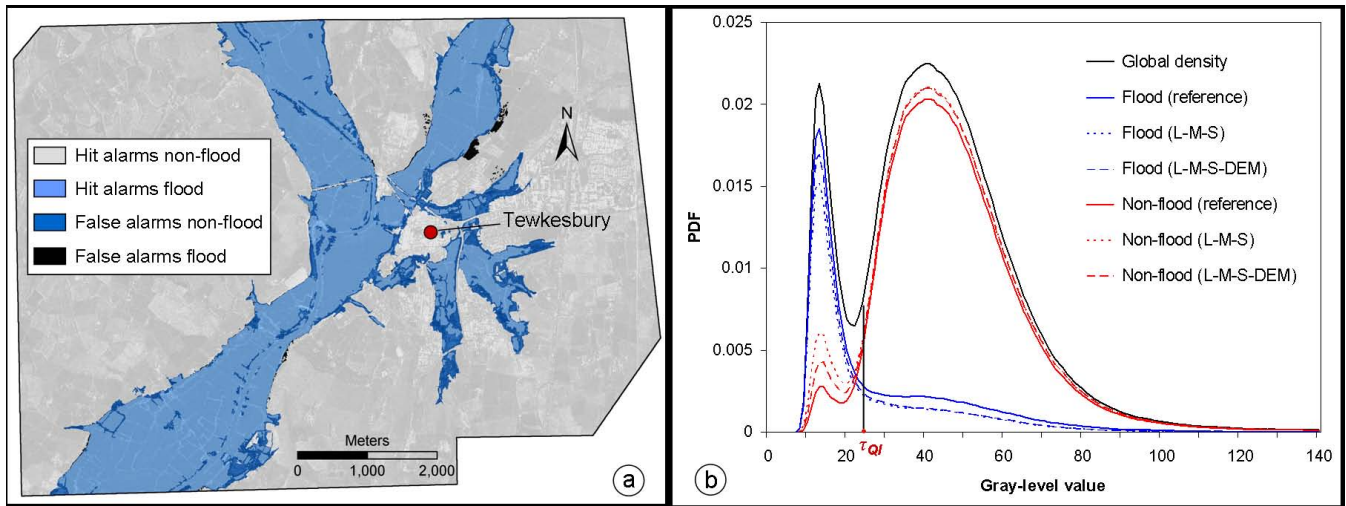

Fig. 5. (a) Classification result by thresholding with $\tau_{glob}=25$ derived from the merged histogram of $X_{n''}$ with L-M-S-DEM by the QI algorithm **(b)** global density function of the reference subset as well as class-conditional density functions of "flood" and "non-flood" of the reference data and of thresholding results P (global density function separated by the vertical black curve), L-M-S and L-M-S-DEM.

The final flood map for the subset of X obtained by L-M-S-DEM with a threshold of $\tau_{glob}=25.0$ is shown in Fig. 5a. The producer's accuracy reaches 82.01%, the user's accuracy 98.65% and the overall accuracy is 95.44%. This is a good result with overall accuracies considerably higher than commonly reported on the literature, which rarely reaches 90%. The result shows a coherent flood surface without salt-and-pepper effects and a good representation of the land-water boundary due to multi-level segmentation. Additionally, compared to pure per-pixel classification, less image el-

ements are classified outside the floodplain. The remaining errors in waterline position found at the riverbanks are either due to emergent vegetation or caused by a slight decrease of the flood extent between the acquisition of the orthophotos (water level at Mythe gauge: 12.49±0.05 m a.s.l.) and the SAR data (water level at Mythe gauge: 12.22±0.05 m a.s.l.), whereas emergent hedges between adjacent flooded fields are predominantly labelled as "flood". Some objects within the floodplain are not detected by the classification with DEM integration due to some remaining artefacts in the LiDAR to-

pography information such as short vegetation, which cause overestimations of the real ground surface elevation. False alarms just appear at the boarder of the extracted flood extent. These are related to crop fields with short vegetation which show a similar spectral characteristic as water areas. Due to the fact that the DEM exclusion step only removes objects at higher elevations without connection to the main flood area, these regions initially labelled as main flood could not be deleted.

Generally, the algorithm performs very well in rural areas. In the partly flooded Tewkesbury town, however, only some larger regions could be detected as flooded. These are areas which are not brightened by double bounce effects as well as contributions from side-lobes of strong reflectors and therefore can be detected due to their intensity value below the derived threshold. Supervised methods with a considerable amount of user interaction are necessary to perform a more accurate classification in urban areas using SAR data (Mason et al., 2008).

It is assumed that the obtained classification accuracy could be slightly increased if the flood extent extracted from the SAR data and from the aerial photos would be fit to one another. An estimation of the flood extent at the satellite overpass time was done by Mason et al. (2008) for the same SAR scene by linearly interpolating between the aerial photo flood extent on 24 July and the flood extent of 27 July derived from aerial photos acquired on a second flight campaign. Due to the fact that the aerial photos of the 27th were not at our disposal during our investigations, no such correction could be accomplished.

The advantage of segmentation-based thresholding is also apparent if the class distributions derived from pixel- and object-based flood detection are compared (Fig. 5b): the class distributions derived by pixel-based thresholding according to $\tau_{glob}=25.0$ show a rigorous truncation of the right tail of the class flood and of the left tail of the class "non-flood" whereas the populations derived by thresholding of the segmented image with L-M-S and especially L-M-S-DEM come closer to the shape of the reference classes' distributions. This is due to the fact that the segmentation process also integrates image elements above the selected threshold into a homogeneous image object in dependence of the chosen homogeneity criterion. Thus, besides pure spectral information derived by histogram thresholding, also topological information is made available for classification.

5 Conclusions

A split based automatic thresholding and classification refinement process has been applied as a test case to a TerraSAR-X Stripmap scene of the River Severn, UK, for unsupervised flood detection in large high resolution single-polarized SAR data sets. This approach already has successfully been applied in operational rapid mapping activities.

The proposed method tiles a SAR image into a set of segments and automatically selects a fixed number of splits according to their probability to comprise adequate portions of both classes "flood" and "non-flood". To derive one global threshold by the analysis of the split inherent information mathematical combination techniques of (sub)histogram derived thresholds as well as (sub)histogram-merging strategies with subsequent thresholding have been analyzed. All three thresholding strategies (KI, GM, and QI) were able to determine a threshold value by the analysis of the histogram of every single selected split and of the merged class-populations of the individual splits, respectively.

The following results could be derived by thresholding a reference subscene decomposed by different segmentation adjustments: object-based context-sensitive thresholding has proven superior to pixel-based context-insensitive procedures due to the addition of spatial information to the pure spectral information derived from histogram thresholding. Lowest OERs could be obtained by a threefold segmentation step which consists of an alternating multi-scale image decomposition and thresholding approach that combines the advantages of small-, medium- and large-scale object generation. By the use of image segmentation, the variation in OER for the classification results derived by different threshold values can be fairly reduced in comparison to pixel-based thresholding approaches. Thus, the classification results of decomposed images are less sensitive to the used threshold algorithm as well as the applied split combination technique. Lowest OERs could be achieved by combining the KI algorithm with the (sub)histogram merging strategy whereas for the image segmentation adjustments M, L, L-M, L-M-S, L-M-S-DEM greatest classification accuracies could be achieved by a higher threshold value derived by the QI algorithm in combination with the (sub)histogram merging strategy.

High resolution topographic information has been successfully combined with multi-scale segmentation to enhance the mapping performance in areas that partly consist of flooded vegetation and anthropogenic objects as well as to remove non-water look-alike areas.

Experimental results of the SBA have shown that it is computationally more efficient than global threshold derivation by the use of the whole SAR scene. This is due to the fact that the most time-consuming step of global threshold detection is the computation of the image histogram which depends on the number of pixels of the analyzed data. In contrast the KI thresholding procedure is very fast, as no iterations are needed, but only the calculations of the criterion function for a defined gray-level range. The processing times of the GM and QI algorithm are nearly identical to the computational speed of the KI algorithm because only some additional calculations are needed to determine τ_{GM} and τ_{QI}, respectively. The computation can be accelerated if the position of the threshold can be restricted a priori. Due to the fact that the criterion function may have local minima at the

boundaries of the histogram meaningless thresholds can be avoided by a constraint of the loop. The computational effort of the SBA (less than 10 s for the entire scene on an a Intel Xeon 5160 CPU with 3.0 GHz and 3 GB of RAM) is much lower than the refinement process by image segmentation, whose processing time rises according to the number of used segmentation-scales (160 s at L-M-S-DEM for the selected validation area). Therefore, the interpreter has to find a compromise between computational effort and classification quality.

The presented thresholding algorithm works successful in most of the cases, due to the fact that there is usually a sufficient contrast between the land and water surface. Future work will be related to extract the flood extent from roughened water surfaces.

Edited by: B. Merz
Reviewed by: G. Schumann and another anonymous referee

References

Baatz, M. and Schäpe, A.: Object-oriented and multi-scale image analysis in semantic networks, in: Proceedings of the 2nd International Symposium on Operationalization of Remote Sensing, Enschede, Netherlands, 16–20 August 1999.

Bazi, Y., Bruzzone, L., and Melgani, F.: An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images, IEEE T. Geosci. Remote, 43, 874–887, 2005.

Bazi, Y., Bruzzone, L., and Melgani, F.: Image thresholding based on the EM alg-orithm and the generalized Gaussian distribution, Pattern Recogn., 40, 619–634, 2007.

Bovolo, F. and Bruzzone, L.: A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment, IEEE T. Geosci. Remote, 45, 1658–1669, 2007.

Brivio, P. A., Colombo, R., Maggi, M., and Tomasoni, R.: Integration of remote sensing data and GIS for accurate mapping of flooded areas, Int. J. Remote Sens., 23, 429–441, 2002.

Chen, P., Liew, S. C., and Lim, H.: Flood detection using multitemporal Radarsat and ERS SAR data, in: Proceedings of the 20th Asian Conference of Remote Sensing, Hong Kong, China, 1185–1189, 22–25 November 1999.

Definiens AG: Defniens Developer 7 User Guide, Document Version 7.0.5.968, Munich, Germany, 2008.

Henry, J. B., Chastanet, P., Fellah, K., and Desnos, Y. L.: ENVISAT multi-polarised ASAR data for flood mapping, Int. J. Remote Sens., 27, 1921–1929, 2006.

Kittler, J. and Illingworth, J.: Minimum error thresholding, Pattern Recogn., 19, 41–47, 1986.

Lee, J. S.: Speckle suppression and analysis for SAR images, Opt. Eng., 25, 636–643, 1986.

Lopes, A., Nezry, E., Touzi, R., and Laur, H.: Maximum A Posteriori speckle filtering and first order texture models in SAR images, International Geoscience and Remote Sensing Symposium (IGARSS), Maryland, USA, 2409–2412, 24–28 May 1990.

Mason, M. S., Horritt, J. T., Dall'Amico, Scott, T. R., and Bates, P. D.: Improving river flood extent delineation from synthetic aperture radar using airborne laser altimetry, IEEE T. Geosci. Remote, 45, 3932–3943, 2007.

Mason, M. S., Speck, R., Schumann, G., Neal, J., and Bates, P. D.: Using TerraSAR-X data for improved urban flood model validation, in: Proceedings of the 3rd TerraSAR-X Science Team Meeting, DLR Oberpfaffenhofen, Germany, 25–26 November 2008.

Matgen, P., Schumann, G., Henry, J. B., Hoffmann, L., and Pfister, L.: Integration of SAR derived river inundation areas, high-precision topographic data and a river flow model toward near real-time flood management, Int. J. Appl. Earth Obs., 9(3), 247–263, 2007.

Meinel, G. and Neubert, M.: A comparison of segmentation programs for high resolution remote sensing data, in: Proceedings of the ISRPS Annual Conference, Istanbul, Turkey, 19-23 July 2004.

Miasnikov, A. D., Rome, V., and Haralick, R. M.: A hierarchical projection pursuit clustering algorithm, in: Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, England, 268–271, 23–26 August 2004.

Melgani, F., Moser, G., and Serpico, S. B.: Unsupervised change-detection methods for remote-sensing images, Opt. Eng., 41, 3288–3297, 2002.

Moser, G. and Serpico, S. B.: Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery, IEEE T. Geosci. Remote, 44, 2972–2982, 2006.

Nico, G., Pappalepore, M., Pasquariello, G., Refice, S., and Samarelli, S.: Comparison of SAR amplitude vs. coherence flood detection methods – a GIS application, Int. J. Remote Sens., 21, 1619–1631, 2000.

Pal, N. R. and Pal, S. K.: A review on image segmentation techniques, Pattern Recogn., 26, 1277–1294, 1993.

Sahoo, P. K., Soltani, S., Wong, A. K. C., and Chen, Y. C.: A survey of thresholding techniques, Comput. Vision Graph., 41, 233–260, 1988.

Townsend, P. A.: Estimating forest structure in wetlands using multitemporal SAR, Remote Sens. Environ., 79, 288–304, 2002.

Wang, Y., Colby, J. D., and Mulcahy, K. A.: An efficient method for mapping flood extent in a coastal floodplain using Landsat TM and DEM data, Int. J. Remote Sens., 23, 3681–3696, 2002.