Unsupervised Extraction of Flood-Induced Backscatter Changes in SAR Data Using Markov Image Modeling on Irregular Graphs

Sandro Martinis, André Twele, and Stefan Voigt

Abstract—The near real-time provision of precise information about flood dynamics from synthetic aperture radar (SAR) data is an essential task in disaster management. A novel tile-based parametric thresholding approach under the generalized Gaussian assumption is applied on normalized change index data to automatically solve the three-class change detection problem in large-size images with small class *a priori* probabilities. The thresholding result is used for the initialization of a hybrid Markov model which integrates scale-dependent and spatiocontextual information into the labeling process by combining hierarchical with noncausal Markov image modeling. Hierarchical maximum *a posteriori* (HMAP) estimation using the Markov chains in scale, originally developed on quadtrees, is adapted to hierarchical irregular graphs. To reduce the computational effort of the iterative optimization process that is related to noncausal Markov models, a Markov random field (MRF) approach is defined, which is applied on a restricted region of the lowest level of the graph, selected according to the HMAP labeling result. The experiments that were performed on a bitemporal TerraSAR-X StripMap data set from South West England during and after a large-scale flooding in 2007 confirm the effectiveness of the proposed change detection method and show an increased classification accuracy of the hybrid MRF model in comparison to the sole application of the HMAP estimation. Additionally, the impact of the graph structure and the chosen model parameters on the labeling result as well as on the performance is discussed.

Index Terms—Automatic thresholding, change detection, flood mapping, generalized Gaussian distribution, hierarchical maximum *a posteriori* (HMAP) marginal estimation, irregular graph, Markov random field (MRF).

I. Introduction

THE WORLDWIDE increase of flood risk [1] and the timely monitoring capability of the recently launched high-resolution space-based synthetic aperture radar (SAR) sensors (TerraSAR-X, COSMO-SkyMed, Radarsat-2, and ALOS-PALSAR) require accurate and automatic change detection methods for the derivation of flood dynamics. This is particularly the case for rapid mapping purposes [2] where accurate information about the extent of a disaster and its spatiotemporal evolution are necessary on a near real-time basis to support decision makers and humanitarian relief organizations.

In comparison to optical remote sensing, the SAR data show the clear advantage of imaging capability even in the case of adverse meteorological and nighttime conditions.

Numerous techniques have been developed for the derivation of changes in remote sensing imagery. The most common change detection techniques are postclassification comparison [3] and classification of some feature maps, such as difference images [4], ratio images [5], and results derived from selective principal component analysis [6].

Several unsupervised change detection methods have been established in the last few years, which derive change maps through the application of automatic thresholding algorithms on feature maps by using minimum-error procedures [5], [7] or expectation-maximization algorithms [8], [9] for the estimation of the data mixture models. These methods are proven to be effective in solving the single-threshold approach between classes "change" and "no change" in small subsets of remote sensing data pairs. However, in large data sets, where local gray-level changes may not be distinguished in the image histogram due to small *a priori* probabilities of some classes, these methods may fail. Furthermore, the change detection problem in feature maps extends to a three-class segmentation problem if different types of changes are to be extracted.

An important feature which improves the quality of classification results is the integration of contextual information into the labeling process. Markov image modeling is proven to be a useful tool in considering the contextual information in remote sensing data. Noncausal lattice-based Markov random fields (MRFs) have been extensively used to model stochastic interactions among classes in remote sensing imagery [4], [10]–[12], and they allow a global Bayesian optimization of the classification result [12]. However, the computational effort is considerable since the classification process must be accomplished iteratively for each image element. This is particularly time consuming in pixel-based applications.

A computationally faster alternative to MRFs are hidden Markov chain (HMC) models [12]–[15], which are applied for spatial image regularization problems to 1-D representations of images, like the Hilbert–Peano Scans, in a noniterative manner. The computational improvement of these methods, however, may come at the expense of lower classification accuracy [12].

In contrast to noncausal MRF and HMC models, the hierarchical Markov models [16]–[18] have the capacity to capture the intrinsic hierarchical nature of remote sensing data. The
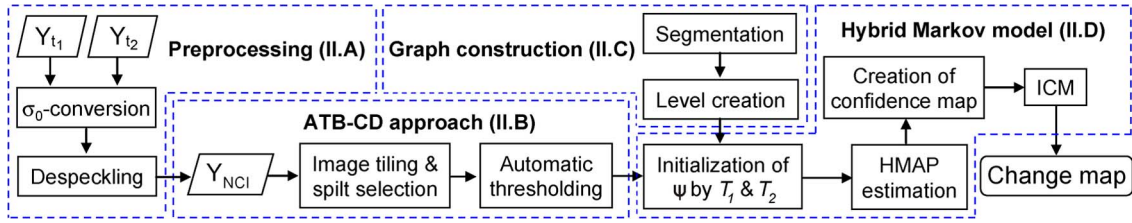
Fig. 1. Block scheme of the proposed method.

image is represented by a hierarchical graph which allows the integration of multiscale information into Markov image models. These models can be solved by fast noniterative inference, e.g., by Markov chains in scale [16], [18].

In most of the cases, hierarchical Markov modeling is conducted on the simple regular graph structure of quadtrees [18]–[21], which is represented by four branches at each node. Due to the fact that image segmentation is accomplished without consideration of the image content, block artifacts may occur in the final estimates. Several methods have been proposed to reduce these undesired effects, e.g., *a posteriori* smoothing [22], the integration of tree structures with overlapping data leaves [23], and the introduction of interlevel edges [16].

On a quadtree, the class label of a node solely depends on the labels of its ancestor and its four descendants. In order to improve the classification, the study in [17] and [19] considered a more complex graph, which additionally integrates a regular spatial neighborhood at each level in the modeling process. However, since the modeling incorporates all elements of each level in an iterative way, the computational demand strongly increases.

This paper presents an automatic method for the near real-time extraction of flood-induced change areas in multitemporal TerraSAR-X data using a hybrid context-based model, combining causal with noncausal Markov image modeling on irregular hierarchical graphs (Fig. 1). The specific objectives are the following: 1) to implement a tile-based automatic change detection approach for the initialization of the Markov image model. Therefore, a parametric thresholding algorithm under the generalized Gaussian (GG) assumption is applied to image tiles to solve the three-class change detection problem in large-size SAR data; 2) to establish an automatic method for the creation of an irregular hierarchical graph structure of the SAR data according to user-defined specifications of average object sizes at each level; and 3) to integrate scale-dependent contextual information by hierarchical maximum *a posteriori* (HMAP) estimation using Markov chains in scale as well as spatiocontextual information by a noncausal MRF model in a partial region of the finest level of the tree, defined by the frequency of occurrence of class labeling in the multiscale representation of the image.

II. METHODOLOGY

A. Data Set

Data Set Description: This study was conducted over the Gloucestershire region in South West England, which was affected by severe flooding at the River Severn in summer 2007.

In the city of Tewkesbury, a record flood level of 5.23 m was reached on July 22, 2007. In order to carry out an experimental analysis that is aimed at assessing the performance of the proposed change detection approach for a real test case, a bitemporal data set that consists of two TerraSAR-X StripMap scenes with 3.0-m pixel spacing in the range and azimuth directions is used. The first image was acquired in HH polarization on July 25, 2007, with an incidence angle of $\sim 22.4°$ along a section of about 69 km long of the River Severn. The second image was recorded roughly one year after (July 22, 2008). It exhibits nearly the same geographic coverage, with a river network at normal water level conditions. Due to the fact that the second image has identical acquisition parameters, shadowing and layover conditions in the data are similar.

For validation purposes of the water level in each scene, two aerial orthophotos are used: The first one shows the flood situation at about half a day prior to the TerraSAR-X scene on July 25, 2007, at a spatial resolution of 0.2 m. For the validation of the water level of the postevent data, no contemporary aerial photographs are available. However, due to the fact that, on July 22, 2008, the SAR scene was recorded at normal water level conditions, the aerial photographs on April 30, 2007, could be used as reference due to a very similar extent of the rivers.

Preprocessing: Initially, the SAR amplitude data are radiometrically calibrated to backscattering coefficients σ_0 to minimize the impact of the topography on the radar signal. To simplify data handling in the following processing steps, all σ_0 values are converted to positive values by adding a constant to the σ_0 value of each date. For a reduction of the SAR data inherent speckle effect, a Gamma-MAP filter [24] of matrix size 3×3 is used. Additionally, the filtering step minimizes the statistical overlap between class distributions and, therefore, leads to a better class separability.

A reference map is created for a subarea $(\sim 6.43 * 10^6$ pixels) of the overlap between the SAR scenes comprising the heavily flooded Tewkesbury region by visual interpretation and manual digitalization of the water areas in both orthophotos. The final flood change mask is created by the intersection of these two water layers.

B. ATB-CD Approach

Problem Formulation: Let us consider two coregistered SAR intensity images Y_{t1} and Y_{t2} acquired over the same geographical areas at two times t_1 and t_2 $(t_1 < t_2)$, respectively. To derive the flood-induced changes over the time interval $[t_1, t_2]$, the information content of Yt_1 and Yt_2 is combined into one

image Y with D elements by the calculation of a normalized change index (NCI) for each pixel at position d ($\sigma_0 > 0$)

$$NCI = \left(\left(\sigma_0^{t_2} - \sigma_0^{t_1}\right) / \left(\sigma_0^{t_2} + \sigma_0^{t_1}\right)\right) + 1. \tag{1}$$

The generated image Y with $y_d = NCI$ has a float data range from $[0, \ldots, 2]$, with values that are equal to one showing the unchanged areas. A high NCI is an indicator for pixels with an increase in backscattering at t_2, which, in this case, is related to the changes caused by a decreasing flood level. For a more efficient data handling, the float values of Y are transformed to 8-bit integers with a possible gray-level range G of 256, i.e., $y_d = g$, $g \in \{0, 1, \ldots, G-1\}$.

Since the histogram of Y can be considered as a mixture of three distributions of classes i ["negative change" ($C-$), "unchanged" (U), and "positive change" ($C+$)], the classification task aims at identifying threshold τ_1 between $C-/U$ and τ_2 and between $U/C+$. This means that y_d could be labeled as $C-$ if $y_d \leq \tau_1$, as U if $\tau_1 < y_d < \tau_2$, and as $C+$ if $y_d \geq \tau_2$.

To automatically derive τ_1 as well as τ_2 and, therefore, to initialize the Markov model in an unsupervised way, an automatic tile-based change detection (ATB-CD) approach is used. This method extends the work of [25], which used an automatic thresholding approach for the extraction of flood areas in high-resolution mono-temporal SAR data in the context of operational rapid mapping activities. In this approach, the SAR data are split into a set of square subimages of a user-defined size, and a fixed number of tiles are automatically selected according to their probability to comprise adequate portions of the classes "flood" and "no flood." In order to derive the best class separation in the (sub)histograms, the Kittler and Illingworth (KI) thresholding algorithm [26] has been used, which models the class-conditional probability density function (pdf) as a mixture of two clusters of two 1-D normal distributions of the semantic classes "flood" and "no flood" and which derives the threshold at the gray-level position where the minimum classification error occurs.

Since change detection is a three-class segmentation problem, if one wants to extract different types of changes, this binary classification method has to be extended to a double threshold selection task. In large NCI images, the *a priori* class probabilities are usually very different. Therefore, in particular, the change classes $C-$ and $C+$ might not be detectable in the histogram of the entire data set. This partly makes it impossible for any thresholding algorithm to compute adequate threshold values. The usage of the ATB-CD approach offers the advantage that the threshold positions are computed from several tiles of the whole NCI image where local gray-level changes, even of classes of low areal representation, may be distinguished in the image (sub)histograms.

NCI Splitting and Tile Selection: Initially, the NCI data Y are split into N square nonoverlapping subimages Y_n of size z^2. In order to prevent unimodality of the selected (sub)histograms, it has to be ensured that a distinct number of elements of both classes $C-$ and U as well as U and $C+$, respectively, are represented within the tiles. Appropriate tiles for the derivation of τ_1 and τ_2 are selected, combining the global statistics of

Y and the local statistics of Y_n by applying the following formulations:

$$Y_{n'_{C-/U}} = \{Y_n | CV_{Y_n} \geq 0.3 \wedge R_{Y_n} \leq 0.9, \quad n = 1, \ldots, N\} \tag{2}$$

$$Y_{n'_{U/C+}} = \{Y_n | CV_{Y_n} \geq 0.3 \wedge R_{Y_n} \geq 1.1, \quad n = 1, \ldots, N\}. \tag{3}$$

CV_{Y_n} is the coefficient of variation (ratio of the square root of the local variance v_{Y_n} of the gray values of Y_n to the local mean μ_{Yn}), and R_{Y_n} is the ratio between μ_{Yn} to the global intensity mean μ_Y of Y. CV_{Y_n} represents the degree of gray-level variation within Y_n. Therefore, it is an appropriate feature to separate subimages with nearly unimodal histograms from tiles that are represented by more than one class. This means that tiles are selected, which contain a significant amount of pixels of the classes $C-$ and U as well as U and $C+$, respectively. Using the additional feature R_{Y_n}, the selected subimages can be divided into tiles which seem to be adequate for the derivation of either τ_1 or τ_2.

Out of the two sets $N'_{C-/U}$ and $N'_{U/C+}$ of subimages $Y_{n'(C-/U)}$ and $Y_{n'(U/C+)}$ which are located in the range given by (2) and (3), respectively, a limited number of tiles are finally used for threshold computation. This selection is accomplished by choosing the quantity $N''_{C-/U}$ and $N''_{U/C+}$ of tiles $Y_{n''(C-/U)}$ and $Y_{n''(U/C+)}$, with the lowest Euclidean distances to the cluster centers of the 2-D feature spaces determined by CV_{Y_n} and R_{Y_n}. The values of CV_{Y_n} and R_{Y_n} in (2) and (3) are determined empirically from the investigation of several NCI data sets. For visualization of the splitting and tile selection process, the reader is referred to [25].

If $N'_{C-/U}$ and/or $N'_{U/C+}$ is zero, the interval of the selection criterion is expanded by iteratively decreasing CV_{Y_n} by 0.01 to a minimum value of 0.25. However, in some cases, in every Y_n of size z^2, the *a priori* probability of a class may be very small. Therefore, if still no tiles are selected according to the enlarged criterion, the initially chosen tile size z is halved, which increases the number of N by a factor of four. If after this step $N'_{C-/U}$ and/or $N'_{U/C+}$ remains zero, it is assumed that the corresponding class $C-$ and/or $C+$ is not represented in Y.

Automatic Threshold Selection Procedure: The KI thresholding algorithm automatically computes the threshold value between the object and background classes of a gray-scale image according to a parametric estimation of the statistical model of the two class pdfs under the Gaussian distribution assumption.

In order to improve the threshold selection process, the study in [7] reformulated this method based on the GG model. This model is well adapted to approximate a large range of symmetric, leptokurtic (ranging from impulsive to Gaussian density), and platykurtic distributions (ranging from Gaussian to uniform density). Their parameterization requires the estimation of only one additional parameter in comparison to the Gaussian model, which is the shape parameter $\beta_i > 0$, which determines the rate of the exponential decay of the pdfs ($\beta_i = 1$ for Laplacian,

$\beta_i = 2$ for Gaussian, and $\beta_i > 8$ for nearly uniform density functions).

In the following, it is assumed that the histograms $h(g)$ of $Y_{n''(C-/U)}$ and $Y_{n''(U/C+)}$ can be parameterized by a pdf $p(g)$, which is a mixture of two GG distributions $p(g|i)$ with parameter mean μ_i, standard deviation σ_i, *a priori* probability P_i, and shape parameter β_i, where

$$i = \begin{array}{ll} (C-, U), & \text{if } Y_{n''} \in Y_{n''_{C-/U}} \\ (U, C+), & \text{if } Y_{n''} \in Y_{n''_{U/C+}} \end{array}$$

$$\text{so that} \quad p(g) = \sum_{i=1}^{2} P_i p(g|i). \tag{4}$$

The expression of the GG distribution for modeling the two class-conditional pdfs of $Y_{n''(C-/U)}$ and $Y_{n''(U/C+)}$ is given by

$$p(g|i) = \frac{b_i \beta_i}{2\Gamma\left(\beta_i^{-1}\right)} e^{-(b_i|g-\mu_i|)^{\beta_i}} \quad \text{with } b_i = \frac{1}{\sigma_i}\sqrt{\frac{\Gamma(3/\beta_i)}{\Gamma(1/\beta_i)}} \tag{5}$$

where $\Gamma(\cdot)$ is the gamma function. The required parameters have to be estimated at some arbitrary decision threshold T which separates the two clusters with the following properties:

$$\mu_i(T) = \frac{1}{P_i(T)} \sum_{g=a}^{b} g h(g) \quad \text{with } P_i(T) = \sum_{g=a}^{b} h(g) \tag{6}$$

$$\sigma_i(T) = \frac{1}{P_i(T)} \sum_{g=a}^{b} (g - \mu_i(T))^2 h(g) \tag{7}$$

$$\text{where} \quad a = \begin{cases} 1, & \text{if } i = C- \wedge Y_{n''} \in Y_{n''_{C-/U}} \\ & \vee \, i = U \wedge Y_{n''} \in Y_{n''_{U/C+}} \\ T+1, & \text{if } i = U \wedge Y_{n''} \in Y_{n''_{C-/U}} \\ & \vee \, i = C+ \wedge Y_{n''} \in Y_{n''_{U/C+}} \end{cases} \tag{8}$$

$$b = \begin{cases} T, & \text{if } i = C- \wedge Y_{n''} \in Y_{n''_{C-/U}} \\ & \vee \, i = U \wedge Y_{n''} \in Y_{n''_{U/C+}} \\ G, & \text{if } i = U \wedge Y_{n''} \in Y_{n''_{C-/U}} \\ & \vee \, i = C+ \wedge Y_{n''} \in Y_{n''_{U/C+}}. \end{cases} \tag{9}$$

The shape parameter is derived by the estimation procedure described in [27]. This technique allows the identification of an estimated value of β_i which best parameterizes the so called GG ratio function $r(\beta_i)$, computed for 450 different values of β_i in the range of $[0.5, \ldots, 5.0[$ with a step size of 0.01.

The computation of the final decision thresholds τ_1 and τ_2 is based on the optimization of a criterion function averaging a cost function $c(g, T)$ over $h(g)$

$$J(T) = \sum_{g=0}^{G-1} h(g) c(g, T). \tag{10}$$

According to the number of gray levels of a tile, the cost at G threshold positions can be evaluated successively to label the pixels by utilizing the generated class statistics by computing the fitting criterion $J(T)$ under the GG assumption [7]

$$\begin{aligned} J(T) = & \sum_{g=0}^{T} h(g) \left[b_{C-}(T) \left| g - \mu_{C-}(T) \right| \right]^{\beta_{C-}(T)} \\ & + \sum_{g=T+1}^{G-1} h(g) \left[b_U(T) \left| g - \mu_U(T) \right| \right]^{\beta_U(T)} \\ & - \sum_{g=0}^{T} h(g) \ln P_U(T) - \sum_{g=T+1}^{G-1} h(g) \ln P_{C-}(T) \\ & - \left[P_U(T) \ln \left(\frac{b_U \beta_U}{2\Gamma(1/\beta_U)} \right)(T) \right. \\ & \left. + P_{C-}(T) \ln \left(\frac{b_{C-} \beta_{C-}}{2\Gamma(1/\beta_{C-})} \right)(T) \right]. \end{aligned} \tag{11}$$

Equation (11) has been adapted for tiles $Y_{n''(C-/U)}$. For tiles $Y_{n'(U/C+)}$, $C-$ has to be replaced with U, and U has to be replaced with $C+$. The gray-level position which minimizes $J(T)$ can be regarded as the optimal decision threshold due to the fact that the lowest classification error occurs according to the Bayes classification rule (i.e., the number of mis-segmented pixels is smallest)

$$\tau_{1/2} = \arg \min_T J(T). \tag{12}$$

For finally obtaining two global thresholds for the classification of Y, the locally derived threshold values of $Y_{n''(C-/U)}$ as well $Y_{n''(U/C+)}$ are combined by their arithmetic mean. Another possibility of threshold computation is the analysis of one histogram that is combined with the local histograms of the tiles $Y_{n''(C-/U)}$ as well as $Y_{n''(U/C+)}$[25].

C. Automatic Graph Construction

Image Segmentation: Segmentation is the basic step in low-level processing of images [28], in which an image is sub-divided into disjoint regions, which are uniform with respect to some homogeneity attributes such as spectral or textural characteristics [29]. During the recent years, segmentation-based image analysis has constantly gained importance in Earth observation. This can particularly be attributed to the strongly increased spatial resolution of remote sensing data, which de-mand image analysis techniques that are specifically adapted to the increased intraclass and decreased interclass variability of images [30]. In particular, for the data of the recently launched high-resolution SAR sensors, the use of the per-parcel methods appears promising. These data are, in comparison to medium-resolution SAR data, characterized by higher variances in backscattering properties of different land-cover classes due to the reduced mixed pixel phenomenon and the salt-and-pepper structures that result from the speckle. This means that semantic image information is less represented in a single pixel but in homogeneous image objects and their mutual relations [31]. The decomposition of the images can be accomplished by several segmentation techniques described in literature [29].

In the presented approach, the decomposition of Y in object primitives is performed using the fractal net evolution concept of the eCognition Developer software [31], [32]. In an iterative way, a bottom-up region-merging technique starting from pixel elements is employed to create image objects of maximum allowable heterogeneity according to user-defined spectral and geometrical constraints. These homogeneity attributes are expressed by the so-called scale parameter H, which combines spectral ($h_{\text{color}} \in [0, \ldots, 1]$) and shape homogeneity ($h_{\text{shape}} \in [0, \ldots, 1]$), with $h_{\text{shape}} = 1 - h_{\text{color}}$ and h_{shape}, being composed of the smoothness h_{smooth} and compactness h_{compact} of the object boundaries ($h_{\text{smooth}} = 1 - h_{\text{compact}}$). The fusion of the adjacent objects follows the local mutual best fitting algorithm, leading to the lowest increase of object heterogeneity within the merging process. The greater the value of H is, the greater is the decrease of the sensitivity for the object fusion, i.e., the obtained object size increases.

The created segments exhibit a strong correlation with real-world objects of the Earth's surface. However, due to the fact that the semantic information is scale dependent, a hierarchical network of image objects is simultaneously constructed, which represents the image content at different spatial scales. By exploiting the relationship between the networked objects, the local contextual information can be integrated into the multi-scale image analysis.

Automatic Level Creation: In the proposed approach, the spatial context of each pixel in image Y is modeled according to a complete hierarchical multilevel representation of the scene, where objects at the same level S^l ($l = 1, \ldots, L$) are related to each neighboring object and where every object at an arbitrary level is hierarchically linked to those represented at lower and/or higher levels. Tree-based hierarchic constraints are used, whereby each object at S^{l-1} ($l < L$) has only one parent node at level S^l, i.e., it cannot be included in more than one adaptive neighborhood at the coarser superlevel. The object size and, therefore, the number of objects O^l at each level are adjusted by the scale parameter H^l, which regulates the homogeneity property at each level. The greater the value of H^l is, the greater is the dimension of the objects obtained at S^l.

In the quadtree decomposition of an image [18], [20], O^l of each level and, therefore, the percentage difference E^l of O^l ($l > 1$) to O^1 are known before the image decomposition is accomplished. The total number of segments is approximately given by $4/3 * O^1$ [16]. However, in the context of irregular hierarchical image segmentation, these parameters are not known in advance, and they must primarily be adjusted in a trial-and-error fashion.

In this section, an approach is presented which automatically generates a multilevel segmentation with the following characteristics: 1) the finest level S^1 is decomposed by an initially intended mean relative object number $\tilde{O}_p^1$, defined in objects per pixel, and 2) the coarser levels S^l ($l \neq 1$) offer an initially defined value of $\tilde{E}^l$. On the one hand, this is important in order to prevent over- and undersegmentation of S^1. On the other hand, the transferability of algorithms is ensured by a fixed difference in the mean object size between levels. The choice of $\tilde{O}_p^1$ depends on the minimum mapping unit of the classification process and, therefore, on the information detail that is to be detected from the classes $C-$, U, and $C+$. The lower $\tilde{O}_p^1$ is, the higher is the probability that the fine detail is lost. Therefore, $\tilde{O}_p^1$ should be approximately equal to the size of the smallest objects to be detected. An increasing value of $\tilde{O}_p^1$ causes a higher processing demand due to the increasing number of nodes of the graph. Additionally, more false alarms may occur due to the reduced smoothing effect that is related to the oversegmented data.

For the creation of level S^1 with a user-intended value of $\tilde{O}_p^1$, we present an approach which uses a presegmentation of a limited area of Y. Therefore, Y is decomposed in N square tiles of size z^2. Out of N, three tiles Y_n of size A_{init}^k are assigned to three classes w_k, which show the minimum ($k = \min$) and maximum ($k = \max$) coefficient of variation (CV) as well as the lowest difference to the mean CV ($k = \text{mean}$) of all Y_n. The three tiles are segmented with an arbitrary scale parameter H_{init}^1. Subsequently, the resulting number of objects O_k^1 of each class is used to estimate $\hat{O}_p^1$, which would result from the segmentation of the entire image Y by H_{init}^1. For this task, the value of O_k^1 is weighted in dependence of an extrapolated spatial representation A_{ext}^k of the classes w_k in relation to the size A_Y of Y. The tiles N are then labeled by the following formulations for $n = 1, \ldots, N$:

$$w_{\min} = \{Y_n | CV_{Y_n} \leq CV_{\min} + \sigma_{CV}\} \tag{13}$$

$$w_{\text{mean}} = \{Y_n | CV_{\text{mean}} - 0.5\sigma_{CV} \geq CV_{Y_n}$$
$$\leq CV_{\text{mean}} + 0.5\sigma_{CV}\} \tag{14}$$

$$w_{\max} = \{Y_n | CV_{Y_n} \geq CV_{\max} - \sigma_{CV}\}. \tag{15}$$

The value of $\hat{O}_p^1$ of the extrapolated class probabilities is calculated by

$$\hat{O}_p^1 = A_Y \left(\sum_{k=1}^{3} \frac{O_i^1 A_{\text{ext}}^k}{A_{\text{init}}^k} \bigg/ \sum_{k=1}^{3} A_{\text{ext}}^k \right) D^{-1}. \tag{16}$$

In the case where $\hat{O}_p^1$ differs from $\tilde{O}_p^1$, the relation between O_p^1 and H^1 is used to approximate the user-intended value of $\tilde{O}_p^1$ by adapting H^1 to the prevailing image content of Y. However, the image content is not known in advance, and a varying number of objects at a given value of H^1 are generated in dependence of the class mixtures.

To model the image content, the relation between O_p and H in the range of [1–115] is calculated (Fig. 2) for six tiles ($z = 500$) of Y containing the following class *a priori* probabilities: a) 25%; b) 50%; c) 75%; and d) $> 90\%$ of the change caused by flooding (the remaining percentage is the unchanged natural area); e) 100% unchanged natural area; and f) 100% unchanged urban area.

The lowest value of O_p in dependence of H [1–115] is created by the tile covering 100% of the natural areas e) due to the large size and spectral homogeneity of these regions. In contrast, small-scale features within urban areas f) induce the highest values of O_p in dependence of H. For clarity reasons, just the trends of e) and f) are shown in Fig. 2. Trend lines a)–d) are situated between e) and f). The trends are well modeled ($R^2 > 0.998$) by second-order polynomial equations of the
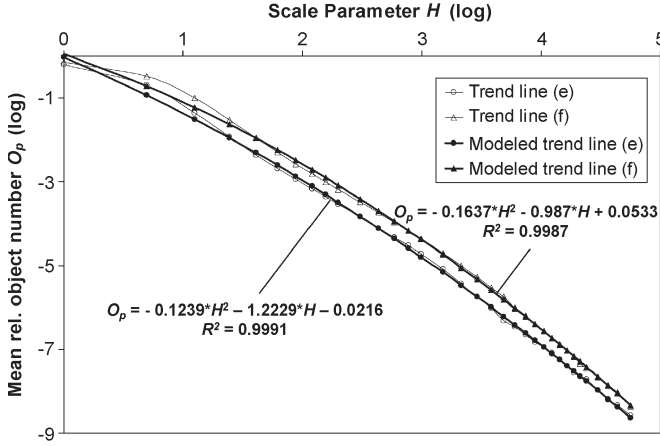
Fig. 2. Dependence of the mean relative object number O_p from the scale parameter H ($h_{\text{color}} = 0.9$ and $h_{\text{compact}} = 0.5$) for two subimages of Y consisting of natural and urban areas, respectively.

form $O_p = a * H^2 + b * H + c$. For each equation of the trend lines a)–f), the respective value of O_p is computed by H^1_{init}, and the function which is most representative for the scene, which generates the lowest difference between O_p and $\hat{O}^1_p$, is used for the subsequent approximation of $\tilde{O}^1_p$. By integrating $\tilde{O}^1_p$ in the respective equation of the chosen function, $\hat{H}^1$ is calculated, which leads to the decomposition of the lowest level S^1 of Y with $\hat{O}_p(\hat{H})^1 \sim \tilde{O}^1_p$

$$\hat{H}^l = e^{\left(-b - \sqrt{b^2 - 4a\left(c - \ln\left(O^l_p\right)\right)}\right)\Big/ 2a}. \qquad (17)$$

Finally, levels S^l ($l > 1$) are decomposed by $\hat{H}^l$, approximating a user-defined percent difference $\tilde{E}^l$ of O^l_p ($l \neq 1$) to $\hat{O}_p(\hat{H})^1$. This is accomplished by integrating $\tilde{O}^l_p$ ($l > 1$) into (17).

D. Markov Image Modeling

Context is a rich source of information for the analysis of remote sensing images. In particular, for the new generation of high-resolution SAR data, objects on the ground are much greater than the pixel size. Thus, the likelihood is very high that the adjacent pixels, although characterized by different scattering properties, belong to the same land cover unit and form a homogeneous region. Markov image modeling is a commonly used approach for the integration of contextual information into the classification scheme of remotely sensed images.

In the following, two different Markov image models are described in the image object domain, i.e., a causal Markov model which uses HMAP estimation on a multiscale tree and a noncausal MRF model on an irregular planar graph associated to an image partition. We furthermore introduce a hybrid method that combines the computational speed of the HMAP modeling approach with the fine spatial modeling of noncausal MRFs on a restricted number of elements. This hybrid Markov model can be combined with every segmentation algorithm, which is able to perform a multilevel representation of an image.

Causal Markov Model: The multilevel decomposition of a SAR scene Y can be represented as a connected graph $\Psi_L = (S, W)$ with L levels that are composed of a set of nodes S and W edges, where each node s, apart from the root r, has a unique parent node s^- (Fig. 3). The superscript "$-$" designates the decrease of the spatial resolution from a child node to its parent node. A descendant node originating from s is denoted as node t. The set of nodes can be partitioned into levels of different scales $S = S^1 \cup S^2 \cup \cdots \cup S^L$ according to the path length from each node to the root. The level decomposed by the lowest scale parameter is S^1, and the coarsest level consists of only one node $S^l = \{r, l = L\}$. In the following, hierarchical Markov image modeling is applied on an irregular hierarchical graph (Fig. 3), leading to a unique statistical inference problem instead of a sequence of multiscale problems that are only loosely related [18].

In this hierarchical classification problem, one attempts to estimate the hidden variable set x through a set of observed variables y. Variables x and y are occurrences of the random vectors X and Y, which are attached to the nodes s of Ψ_L.

In the following, a labeling process is considered, which assigns a label x_s to each node s of levels S^l of Ψ_L, where x_s takes its values in the discrete set x of classes i ($i \in x = \{C-, U, C+\}$):

$$x = \{x^l\}_{l=1}^L \quad \text{with } x^l = \{x_s, s \in S^l\}. \qquad (18)$$

The classification is accomplished by HMAP estimation. The following Bayesian estimator is used:

$$\hat{x} = \arg \max_{x \in x^S} P(x|y) = \arg \max_{x \in x^S} P(x, y) \qquad (19)$$

where x^S is the configuration set of the vectors x of all nodes S.

A number of statistical assumptions about the random vectors (X, Y) are defined in the following [18].

The labeling process X is supposed to be Markovian in scale (i.e., X is a first-order top-down Markov chain)

$$P(x^l|x^\lambda, \lambda > l) = P(x^l|x^{l+1}) \qquad \forall\, l \in \{1, \ldots, L\}. \qquad (20)$$

To simplify the notation, the discrete probability $P(X = x)$ is denoted as $P(x)$ throughout this paper. The interscale transition probabilities can be factorized by the following criteria (21)–(23):

$$P(x^l|x^{l+1}) = \prod_{s \in S^l} P(x_s|x_{s^-}). \qquad (21)$$

This means that, for each S^l, the conditioning in X^{l-1} reduces to a dependence from its parent node only.

Assuming conditional independence, the likelihood of the observations Y conditionally to X is defined by the following model:

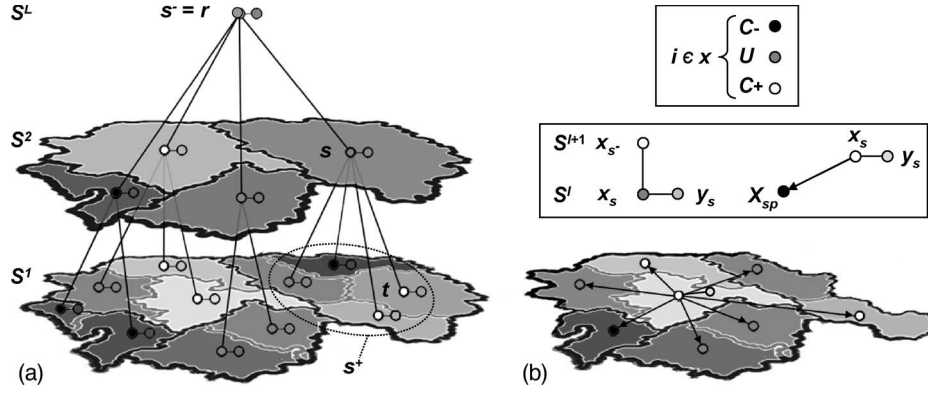$$P(y|x) = \prod_{s \in S^l} P(y_s|x_s). \qquad (22)$$

Fig. 3. (a) Three-level independence graph Ψ_3 corresponding to an irregular tree structure. (b) Irregular planar graph with a second-order clique set in neighborhood system V_{sp}. The left circles represent the labeling according to classes $C-$, U, and $C+$. The right circles represent the observations at each node.

TABLE I
TWO-PASS MAP ESTIMATION ON IRREGULAR HIERARCHICAL GRAPHS

Bottom-up sweep: Functions $P_s(x_{s-})$ (most probable label $\hat{x}_{s-}$ of a parent node s^-) and $x_s^*(x_{s-})$ (MAP estimate of child node s according to MAP component $\hat{x}_{s-}$) are computed and stored in look-up tables for all possible values of x_{s-}

Initialization ($s \in S^1$):
$$P_s(x_{s-}) = \max_{x_s} P(y_s|x_s) P(x_s|x_{s-}) \varphi$$
$$x_s^*(x_{s-}) = \arg\max_{x_s} P(y_s|x_s) P(x_s|x_{s-}) \varphi$$

Recursion ($s \in S^2 ... S^L$):
$$P_s(x_{s-}) = \max_{x_s} P(y_s|x_s) P(x_s|x_{s-}) \prod_{t \in s^+} P_t(x_s) \varphi$$
$$x_s^*(x_{s-}) = \arg\max_{x_s} P(y_s|x_s) P(x_s|x_{s-}) \prod_{t \in s^+} P_t(x_s) \varphi$$

Top-down sweep: Components of the MAP estimate are recovered from the root to the leaves

Initialization (root r):
$$\hat{x}_r = \arg\max_{x_r} P(y_r|x_r) P(x_r) \prod_{t \in r^+} P_s(x_r)$$

Recursion ($s \in S^{L-1} ... S^1$):
$$\hat{x}_s = x_s^*(x_{s-})$$

According to these statistical assumptions, the joint distribution $P(x, y)$ can be expressed by

$$P(x, y) = P(x_r) \prod_{s \neq r} P(x_s|x_{s-}) \prod_{s \in S} P(y_s|x_s) \qquad (23)$$

where $P(x_r)$ is the root prior probability and $\{P(y_s|x_s)\}_{s \in S}$ is the data conditional likelihoods.

One of the advantages of these models lies in the possibility to compute the exact MAP estimate of X, given $Y = y$ at each node of the tree within two passes. This extension of the Viterbi algorithm [18], which is a standard technique for the computation of the MAP estimate of HMCs [33], is summarized in Table I.

To account for the problem of underflow in the Viterbi algorithm, the whole estimation procedure is implemented by computing the logarithm of the probabilities. In the bottom-up sweep, we integrated a weighting factor φ, which takes into account the proportion of the spatial extent A of child node

s to its ancestor s^- in dependence of the number of child nodes t

$$\varphi = (1 - (A_s/A_{s-})) / (\text{card}(t) - 1)). \qquad (24)$$

Equation (24) has been adapted to weight negative values of the probabilities, which result from the logarithmic expression of the probabilities in the HMAP estimation. In contrast to the MAP estimation on the quadtree, where all child objects have the same size of $A_{s-}/4$, φ has to be integrated in the computation of the inference on irregular graphs, where an ancestor is represented by a varying number of child objects at lower levels.

For the conditional likelihoods $P(y_s|x_s)$, a Gaussian model was chosen, defined by the parameter vector $\theta_i = (\mu_i, \sigma_i)$, where μ_i is the mean and σ_i is the standard deviation of class i. Vector θ_i may vary depending on the spatial resolution so that the classes are defined by θ_i^l. The conditional likelihoods are defined by

$$\forall s \in S^l, \quad P(Y_s = y_s|X_s = i) = \frac{1}{2\pi\sigma_i} e^{-(y_s - \mu_i^l)^2 / 2\pi(\sigma_i^l)^2}. \qquad (25)$$

The scale-dependent parameters for each class are defined by the unsupervised change detection method described in Section II-B. For the definition of the prior probability, the Potts-like model is used [16], which favors identical labeling of s and s^-

$$\forall s \in S^l, \quad P(X_s = j|X_{s-} = i) = \begin{cases} \phi, & \text{if } i = j \\ \frac{1-\phi}{M-1}, & \text{if } i \neq j \end{cases} \qquad (26)$$

with the number of classes M, the root prior $P(r) = 1/M$, and the parent prior $\Phi \, [1 \leq \Phi > 1/M]$. A higher value of Φ increases the smoothing effect of the classification result.

The labeling of S^1 is used as the final classification result of the HMAP approach, as during the two pass steps, most contextual information is integrated into the labeling of the lowest segmentation level. The HMAP approach is a fast and effective method for postclassification smoothing on irregular trees. To further improve the results, spatial context is integrated into the labeling process using noncausal MRFs.

Noncausal MRF Model: The principal idea of noncausal MRFs is the modeling of contextual correlations among image elements at the same level according to criteria such as MAP estimation (19). To make MRFs computationally tractable, the conditional prior probability of individual image elements has to be modeled, given their neighboring objects. According to the Hammersley–Clifford theorem [34], [35], the prior $P(x)$ can be expressed by a Gibbs distribution. Using this MRFs–Gibbs equivalence, the global contextual relationship of Y can be modeled by MRFs of local spatial neighborhoods

$$P(x_s|X_{sp}) = Z^{-1} \exp^{-U_{sp}(x_s, X_{sp})} \quad (27)$$

where x_s denotes the class of node s, X_{sp} is a subset of the labeling of the image in a predefined spatial neighborhood system V_{sp}, and Z is a normalizing constant. This simplification is a reasonable approach because the interactions between object labels decrease rapidly with an increasing object distance. Usually, in image analysis, MRFs are applied to regular lattices with first- or second-order neighborhood systems, where each pixel has four or eight neighbors. Since we are dealing with irregular planar graphs associated to an image partition, the number of image elements in V_{sp} varies. The spatial energy function U_{sp} is characterized as

$$U_{sp}(x_s, X_{sp}) = -\sum_{q \in V_{sp}} \gamma \delta(x_s, x_q). \quad (28)$$

Here, δ is the Kronecker delta function, which is, according to a second-order clique system $(s-q)$, equal to one if $x_s = x_q$ and zero, otherwise. The parameter γ controls the influence of the spatiocontextual information of the change detection process, i.e., the greater this term is, the more likely that two adjacent image objects will have the same label. The class-conditional distribution under Gaussian assumptions leads to the following data attach term:

$$U_{\text{data}}(Y_s = y_s|X_s = i)$$
$$= 0.5 \ln |2\pi\sigma_i^2| + 0.5(y_s - \mu_i)^2 (\sigma_i^2)^{-1}. \quad (29)$$

Together with the regularization term U_{sp} and the data term U_{data}, the MAP estimate can be transformed to the following energy minimization problem:

$$\hat{x}_s = \arg \min_{x_s} \{U_{\text{data}}(Y_s = y_s|X_s = i) + U_{sp}(x_s, X_{sp})\}. \quad (30)$$

This energy function can be solved by iterated conditional modes (ICM) [36], which represents a computationally feasible alternative to simulated annealing [37], converging to a local, but usually sufficient, minimum of the energy function. In this proposed method, the ICM algorithm can be computed as follows.

1) Estimate class parameter vectors θ_i^l of the initial tile-based thresholding result.
2) For all elements $s \in S^1$, update X by using the HMAP approach.

3) For all $s \in S^1$, perform the local minimization by using (30).
4) Repeat the preceding step until convergence is reached.

The application of the ICM algorithm to irregular planar graphs compared to pixel-based applications has the advantage that the model is applied to a strongly decreased number of image elements. This is related to a dramatic reduction of computational complexity. However, since the computing time of ICM may still be considerable, particularly if the segmentation of S^1 is accomplished by a small homogeneity parameter, we introduce a hybrid Markov model. According to this model, the ICM algorithm is applied only to a restricted number of image objects, defined by the hierarchical representation of the HMAP labeling result.

Hybrid Markov Model: In the following, a confidence map is created at S^1 by combining the HMAP labeling result from the different tree levels. The objects of S^1 are assigned to different categories according to their frequency $f(x_s)$ of class representation in the coarser super objects of S^l $(l \neq 1, L)$:

$$f(x_s) = \sum_{s^- \in (S^2, \dots, S^{L-2})} \kappa,$$
$$\kappa = \begin{cases} 1, & \text{if } x_{s \in S^1} = x_{s^- \in (S^2, \dots, S^{L-2})} \\ 0, & \text{if } x_{s \in S^1} \neq x_{s^- \in (S^2, \dots, S^{L-2})}. \end{cases} \quad (31)$$

If $f(x_s) = L - 2$ (excluding S^1 and S^L), an object at S^1 is represented by the same class label at all super objects. Therefore, this object shows a high degree of confidence to be labeled to the correct class, and no optimization process by the ICM algorithm is necessary for this object. However, if $f(x_s) < L - 2$ at least at one level S^l $(l \neq 1, L)$, the labeling of a super object differs from the labeling of an object at S^1. Accordingly, the class assignment of all elements $s \in S^1$ with $f(x_s) < L - 2$, is checked by the ICM algorithm in the first iteration. This restricted number of elements $O_{\text{sub-1}}^1$ of the subarea $S_{\text{sub-1}}^1$ can be reduced to $O_{\text{sub-2}}^1$ after the first iteration by the restriction of the ICM algorithms to objects $s \in S^1$ with $f(x_s) < L - 2$, which have a common border to other elements of $s \in S^1$ with $f(x_s) < L - 2$.

For the ICM algorithm, $f(x_s)$ is used for the definition of an adaptive smoothness parameter γ_{adapt} for each object $s \in S_{\text{sub-1}}^1$, with values in a user-defined range $[\gamma_{\text{min}}, \gamma_{\text{max}}]$, according to the following criterion:

$$\gamma_{\text{adapt}} = \gamma_{\text{min}} + (\gamma_{\text{max}} - \gamma_{\text{min}})/f(x_s). \quad (32)$$

Thus, an image element at S^1, whose class label is less represented by the same labeling at its ancestors, is characterized by a higher probability that the classification result of the HMAP approach is incorrect. Therefore, a higher value of γ_{adapt} is assigned to this element, which causes an increased influence of the neighboring objects during the ICM approach.

III. EXPERIMENTAL RESULTS

In this section, the effectiveness of the proposed change detection approach is evaluated using an NCI image of the
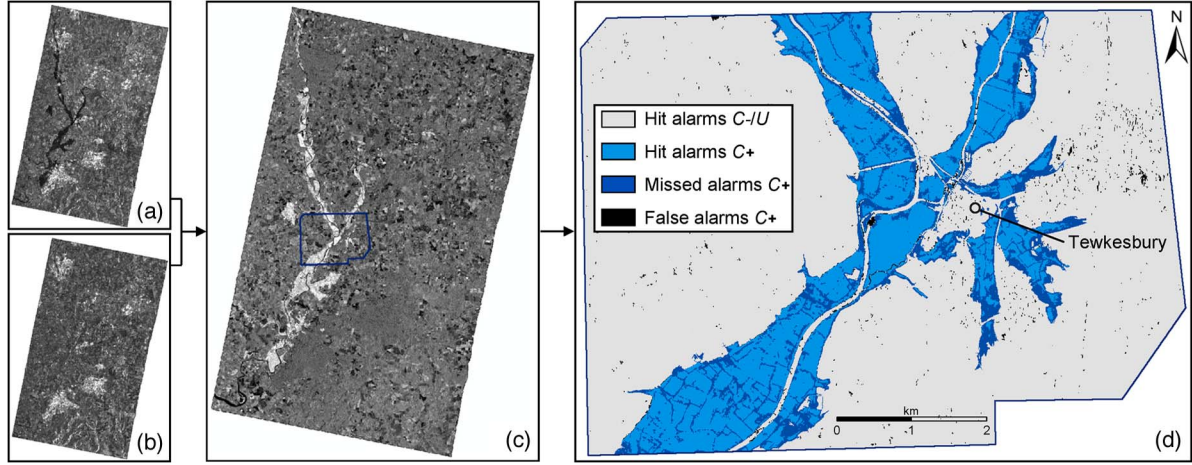
Fig. 4. TerraSAR-X data of the River Severn on (a) July 25, 2007, and (b) July 22, 2008. (c) NCI image. (d) Classification result of Y_{sub} derived by the hybrid Markov model for Ψ_8, with γ_{1-5} and $H = 24$.

TABLE II
GLOBAL THRESHOLDS CALCULATED BY THE ARITHMETIC MEAN OF
LOCALLY DERIVED THRESHOLDS OF TILES $Y_{n''(C-/U)}$ AND $Y_{n''(U/C+)}$

Class-mixture	Local threshold					Global threshold
	$Y_{1''}$	$Y_{2''}$	$Y_{3''}$	$Y_{4''}$	$Y_{5''}$	Mean
$C-/U$	103	101	105	105	101	104.4
$U/C+$	163	164	171	172	168	167.0

Gloucestershire region (Fig. 4): First, the ATB-CD method is applied on Y for the unsupervised initialization of the Markov image models. In the next step, the structure of a hierarchical irregular graph is created automatically, and its properties of fitting to user-intended scaling parameters are evaluated. Finally, the efficiency of the hybrid Markov image model is assessed. Since the reference water mask only covers a small part of Y, a subarea Y_{sub} ($\sim 6.43 * 10^6$ pixels) of the Tewkesbury region is chosen for the evaluation of the method.

A. Tile-Based Automatic Thresholding Approach

The ATB-CD approach is used to derive global thresholds for the initialization of the Markov image models. According to a tile size of $z = 500$, the entire image Y, with $3.5 * 10^8$ elements, is decomposed into $N = 1140$ square tiles. Using the criterion defined in (2) and (3), $N'_{C-/U} = 37$ and $N'_{U/C+} = 32$ tiles are selected, which are suitable for the derivation of the thresholds τ_1 and τ_2. To restrict the computational effort, $N''_{C-/U} = 5$ and $N''_{U/C+} = 5$ tiles are finally chosen for the computation of the local threshold values by applying the GG–KI algorithm. Reasonable global thresholds $\tau_{C-/U} = 104.4$ and $\tau_{U/C+} = 167.0$ are calculated by the arithmetic mean of the locally derived thresholds (Table II). Since these local thresholds are characterized by a low gray-level range of $\sigma_1 = 2.0$ and $\sigma_2 = 4.0$, the proposed method seems to be useful in separating the classes $C-/U$ and $U/C+$. This thresholding approach has the advantage that classes with a small *a priori* probability are also detected in the histogram of Y. Additionally, the computational speed of this methodology is very high since the GG–KI algorithm is only applied on a small number of tiles.

B. Automatic Creation of the Graph Structure

Before the HMAP approach is accomplished, the hierarchical irregular graph structure of the chosen subarea Y_{sub} is generated. In the following, the properties of a graph with $L = 8$ (Ψ_8), created according to user-intended parameters $\tilde{O}_p^1$ and $\tilde{E}^l$, are evaluated.

To approximate a decomposition of S^1 with a mean relative object number $\tilde{O}_p^1$ of 0.015, the following steps are accomplished: First, three square tiles ($z = 500$) of Y_{sub}, labeled to different classes w_k according to (13)–(15), are decomposed by an arbitrary scale parameter $H_{\text{init}}^1 = 15$. The next step encompasses the estimation of $\hat{O}_p^1$ through the extrapolation of the resulting object number O_k^1 of each class by their spatial representation in Y_{sub}. According to (16), $\hat{O}_p^1 = 0.0129$. This is an underestimation of $\tilde{O}_p^1$. In order to approximate $\tilde{O}_p^1$, the scale parameter has to be adapted. According to that, for each of the equations of the trend lines a)–f), describing the contents of image Y in relation to the object size, O_p is calculated by H_{init}^1. For further modeling, trend line b) is chosen since it offers the lowest difference between O_p and $\hat{O}_p^1$ at H_{init}^1. A value of $\hat{H}^1 = 12.9$ is calculated by integrating $\tilde{O}_p^1$ (0.015) in the equation associated to trend line b). The segmentation of Y by $\hat{H}^1$ results in a mean relative object number of $\hat{O}_p(\hat{H})^1 = 0.0171$, which shows a slightly higher value of 0.0021 than $\tilde{O}_p^1$. These results are nearly independent of the initially selected value of H_{init}^1.

The next step encompasses the automatic creation of the coarser superlevels according to user-defined values of the relative percentage $\tilde{E}^l$ of the mean relative object number of level l ($l \neq 1, R$) to $\hat{O}_p(\hat{H})^1$. $\tilde{E}^l$ is set to 50% ($l = 2$), 25% ($l = 3$), 17.5% ($l = 4$), 10% ($l = 5$), 5% ($l = 6$), and 2.5% ($l = 7$). The integration of the related values of $\tilde{O}_p^l$ into the equation describing trend line 2) results in values of $\hat{H}^l$ ranging from 18.7 ($l = 2$) to 67.0 ($l = 7$).

As shown in Fig. 5, the mean relative object number of the levels ($l = 1-7$), decomposed by $\hat{H}^l$, is nearly identical to the intended values. Therefore, the presented method seems to be able to automatically model the parameters for the creation of
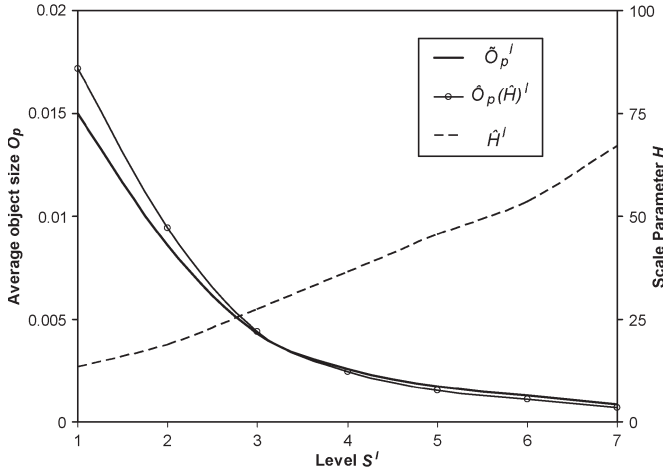
Fig. 5. Fit of the intended value of $\tilde{O}_p^l(= 0.015)$ to the average relative object number $\hat{O}_p^l$ of the segmented levels, decomposed by the modeled scale parameter $\hat{H}^l$.

TABLE III
VALUES OF THE ADAPTIVE SMOOTHNESS PARAMETER γ_{adapt}
IN DEPENDENCE OF $f(x_s)$ FOR Ψ_L ($L = 4, 6, 8$)

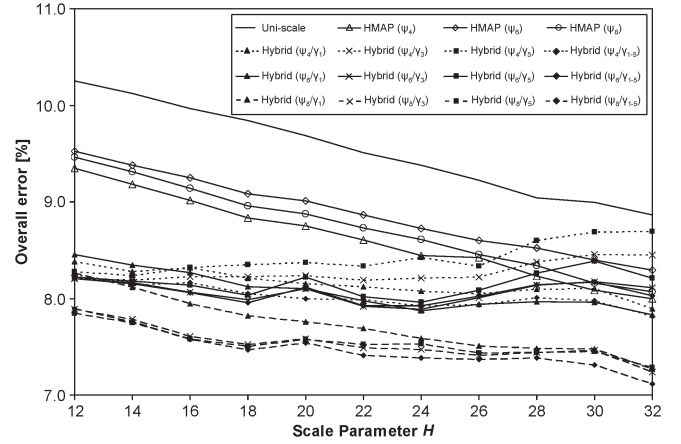| $f(x_s)$	Number of level		
	4	**6**	**8**
5	-	-	1.0
4	-	-	1.8
3	-	1.0	2.6
2	-	2.33	3.4
1	1.0	3.66	4.2
0	5.0	5.0	5.0

Fig. 6. OER of the uniscale classification, the HMAP estimation, and the hybrid Markov model in dependence of $H(= 12-32)$, $L(= 4, 6, 8)$, and different values of γ (1.0, 3.0, and 5.0; 1.0–5.0).

an irregular graph structure according to user requirements. This replaces the time-consuming trial-and-error procedures for the adaptation of H to the image content. Additionally, over- and undersegmentation of an image are automatically avoided. In particular, oversegmentation may lead to serious performance problems in the processing of large data sets due to memory constraints. In contrast, undersegmentation causes a loss of fine detail. The segmentation of the lowest level is the most important step in the generation of the graph structure since the properties of the coarser superlevels depend on the mean object size of S^1.

C. Effectiveness of the Markov Image Models

In order to evaluate the quality of the proposed change detection approach and to derive transferable parameter settings, the HMAP estimation and the hybrid Markov image model are computed with different parameters on a diversified irregular graph structure to analyze the following: 1) the quality of the HMAP algorithm as well as the hybrid Markov image model in comparison to uniscale classification results. In particular, the influence of the hierarchical graph structure and smoothness parameter γ on the classification accuracy is investigated; and 2) the performance of the Markov image models in dependence of the graph structure and γ.

The hybrid Markov model is computed on a graph Ψ_L with a varying number of levels $(L = 4, 6, 8)$ and different scale parameters H in the range of [12–32]. The parent prior was empirically set to 0.9. By the use of Ψ_6, an additional level is integrated between levels S^1 and S^2 as well as S^2 and S^3 of Ψ_4, respectively. H^3 and H^5 of Ψ_6 are, therefore, identical to H^2 and H^3 of Ψ_4. In comparison, using Ψ_8, two coarser levels are generated above S^5 of Ψ_6.

The ICM algorithm is computed with $\gamma = 1.0, 3.0,$ and 5.0 as well as γ_{adapt} in the range of $\gamma_{min} = 1.0$ and $\gamma_{max} = 5.0$ on a restricted region according to the class label representation $f(x_s)$ of a leaf object on its super objects after HMAP estimation. In Table III, the values of γ_{adapt} are shown in dependence of L. The ICM algorithm stops if the percentage of objects with different labels at iterations $1 + n$ and 1 is lower than a predefined threshold of 0.02%.

Accuracy: In Fig. 6, the accuracy assessment of the uniscale classification, the HMAP, and the hybrid Markov model is illustrated by plots of the overall error rates (OER, i.e., the percentage of the erroneously labeled reference pixels, including false positives and false negatives) of class $C+$ in dependence of the altered parameters.

The classification results confirm that the proposed HMAP estimation and the hybrid Markov model applied on irregular graphs always exhibit a much lower OER compared with that obtained using only one decomposition level.

Regarding the results of the uniscale segmentation as well as of the HMAP approach, a nearly linear decrease of the OER with increasing H can be observed in the range of [12–32]. The decrease of the OER resulting from the uniscale classification is about 0.69% per H, starting from 10.25% at $H = 12$ to 8.86% at $H = 32$. The trend lines of the HMAP approach have nearly the same decrease of about 0.62% per H. However, the HMAP trend lines start at about 0.72% (Ψ_6) to 0.90% (Ψ_4) below the OER of the uniscale classification approach at $H = 12$.

The HMAP results show very similar classification accuracies. Therefore, the number of levels does not seem to have a significant impact on the HMAP classification accuracy. Surprisingly, the lowest OERs are calculated on Ψ_4. Slightly higher OERs are obtained on Ψ_6. This is caused by an increased false alarm rate (i.e., the percentage of reference pixels that do not belong to $C+$ that are erroneously detected as $C+$). Therefore, using Ψ_6, created by the integration of additional

layers between S^1 and S^2 as well as S^2 and S^3 of Ψ_4, reduces the smoothing effect of the HMAP approach. The integration of two coarser levels for the creation of Ψ_8 improves the classification accuracy in comparison to the HMAP estimation on Ψ_6 by reducing the missed detection rate (i.e., the percentage of reference pixels of $C+$ that are erroneously detected as U or $C-$). However, the OER is slightly higher in comparison on the HMAP on Ψ_4. In comparison to the number of levels of Ψ, H has a greater influence on the classification accuracy of the HMAP approach, which improves with an increasing H.

Looking at the results of the hybrid Markov image model, the following points can be stated: In comparison to the HMAP results, the hybrid Markov model generates a much lower OER for all parameter settings in the range of $H = [12-26]$. This increase of accuracy is related to the integration of spatial-contextual information into the labeling process using noncausal Markov modeling. However, there are significant differences in the OER caused by the parameter settings and the graph structure. At $H = 12$, the OERs are very similar, and they exhibit values in the interval of 7.89% (Ψ_8/γ_5) and 8.45% (Ψ_6/γ_1). At this position, the greatest increase in accuracy in relation to the HMAP approach can also be stated. With an increasing H, the difference between the maximum and minimum OER increases, and it reaches the highest difference (1.58%) at $H = 32$ between 8.70% (Ψ_4/γ_5) and 7.11% (Ψ_8/γ_{1-5}). Hence, for $H > 26$, the classification accuracy of the hybrid Markov model for Ψ_4 and Ψ_6 is nearly similar to the OERs of the HMAP approach. It is obvious that, with an increasing H, the number of levels of Ψ gains importance and leads to significant differences in classification accuracy. In contrast to HMAP estimation, the accuracy improves with an increasing number of levels.

The decrease of the OER with an increasing H, which has been stated for the HMAP estimation, can only be observed for the hybrid Markov model for Ψ_8. However, the influence is not as pronounced as in the HMAP approach, and the accuracies are nearly identical in the range of $H = [18-28]$ for (Ψ_8/γ_3), (Ψ_8/γ_5), and (Ψ_8/γ_{1-5}), with an interval between 7.39% and 7.58%.

The highest classification accuracies over all trend lines are reached for Ψ_8. Additionally, trend lines for (Ψ_8/γ_3), (Ψ_8/γ_5), and (Ψ_8/γ_{1-5}) offer nearly identical OERs, which are independent from the parameters chosen. This can also be stated for the trend lines for Ψ_6 in the range of $H = [12-26]$. In contrast, the results derived for Ψ_4 are more sensitive to the selection of γ.

The highest overall accuracies are reached between $H = [14-32]$ by (Ψ_8/γ_{1-5}), with OERs in the range of 7.75% ($H = 4$) and 7.11% ($H = 32$). These accuracies are up to 1.5% higher than the HMAP results. Also, for Ψ_4, the lowest OERs can be observed using γ_{adapt} in the whole range of $H = [12-32]$, while at Ψ_6, the generation of the lowest OER by (Ψ_6/γ_{1-5}) is replaced at $\geq H = 24$ by (Ψ_5/γ_1). In contrast, the lowest accuracies are reached using γ_5, particularly for Ψ_4 and Ψ_6.

It can be concluded that an increasing number of levels in the graph enhance the classification accuracy and decrease the influence of the remaining parameters on the results.
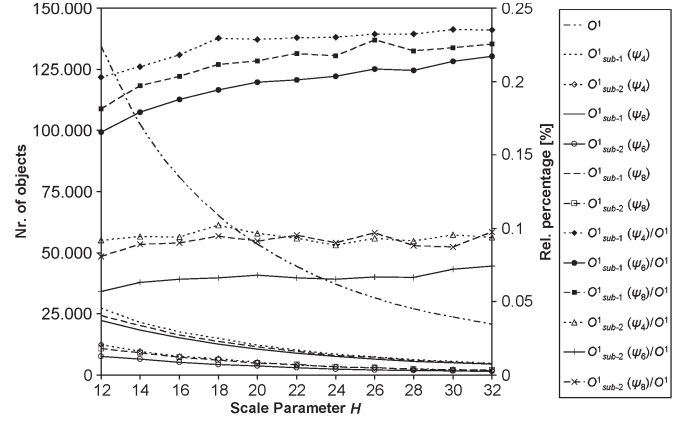

Fig. 7. Performance of the hybrid Markov model in dependence of H and L on all objects of S_1 as well as of $S^1_{\text{sub-1}}$ and $S^1_{\text{sub-2}}$.

The change map of Y_{sub} is shown in Fig. 4, derived for Ψ_8, with $H = 24$ and γ_{1-5}. The map shows a coherent change area caused by the receding flooding. Since no quadtree structure, but an irregular graph, has been used in the processing, no block artifacts are visible in the map. Most errors occur at the border of the flood surface, and they are predominantly misclassified as class U. On the one hand, these are related to the flooded vegetation, which shows nearly similar radar cross sections between t_1 and t_2, and, on the other hand, to a slight decrease of the flood extent between the acquisition of the orthophotos and the SAR data at t_1. False alarms are very sparse, and they occur predominantly in urban areas. These errors could be further reduced by methods integrating elevation information into the labeling process [25], [38].

Performance: The HMAP estimation is a very fast technique due to its noniterative inference on irregular hierarchical graphs. The performance depends on the number of nodes S of Ψ and, therefore, on H^1 as well as on L. This results in a fixed computational complexity per site, i.e., the computational effort increases linearly with a decreasing S. Accordingly, the performance of the HMAP approach ($H^1 = 12$) on Ψ_4 ($S = 181\,250$) is ~37% faster than that on Ψ_8 ($S = 286\,900$) and ~34.5% faster than that on Ψ_6 ($S = 272\,200$). In contrast, the computational efficiency on Ψ_8 at $H = 12$ ($S = 286\,900$) is ~86.2% higher than that at $H = 32$ ($S = 39\,500$).

In comparison, the noncausal MRFs are characterized by an iterative minimization of an energy function. In a per-pixel application, the MAP estimation must be computed on Y_{sub} for all elements (~$6.43 * 10^6$ pixels) in each iteration. In that case, the ICM algorithm achieved convergence after 12 iterations ($V_{sp} = 8$ and $\gamma = 2.0$). Using the ICM algorithm on irregular planar graphs, the performance of this noncausal MRF approach is strongly increased due to the following facts.

The decomposition of an image by H leads to a strongly reduced number of elements integrated into the MAP estimation. When applied to the whole irregular planar graph with O^1, this number decreases with an increasing H from 103 350 ($H = 12$) to 20 780 ($H = 32$) (Fig. 7).

A strong increase in performance is reached by restricting the ICM algorithm to $O^1_{\text{sub-1}}$, derived in Section II-D using the HMAP labeling approach. The value of $O^1_{\text{sub-1}}$ shows ~16.5%–23.5% of O^1. The difference between O^1 and $O^1_{\text{sub-1}}$

decreases with larger values of H (Fig. 7). Therefore, the performance increase is particularly significant on images decomposed by a small value of H. $O_{\text{sub-1}}^1$ decreases continually with an increasing H, which is nearly independent on Ψ_4, Ψ_6, and Ψ_8. The highest values of $O_{\text{sub-1}}^1$ in the range of $27\,250$ ($H = 12$) and 4890 ($H = 32$) are achieved for Ψ_4. In contrast, the lowest values of $O_{\text{sub-1}}^1$ are generated for Ψ_6. This results from the integration of additional intermediate levels into the graph structure, which favors increased identity of the class labeling of a node at S^1 and its ancestors.

Additionally, after the first ICM iteration, $O_{\text{sub-1}}^1$ is further decreased to $O_{\text{sub-2}}^1$ for an increase of the processing time (Fig. 7). Objects with no common border to the other elements of $S_{\text{sub-1}}^1$ are removed from the MAP estimation since they can only change their labeling in the first iteration. This reduces the object number in the subsequent iterations to $\sim 5.5\%$–10.0% in comparison to O^1.

In addition to the number of nodes O^1, checked by the ICM algorithm for a possible label change, the performance of the hybrid Markov model further depends on the size of the neighborhood system V_{sp} and the speed of convergence. In pixel-based applications, usually, four- or eight-neighborhood systems are used. However, on irregular planar graphs, the average size of V_{sp} is nearly similar for all computations, with a range of 3.5–3.9 neighbors per MAP estimation. With an increasing H, the average size of V_{sp} slightly reduces.

The ICM converges independently of the chosen parameters in three to four iterations. This is a strong decrease in comparison to 12 iterations of a per-pixel application of the MAP estimation on Y_{sub}. The mean number of iterations of the ICM with γ (1.0, 3.0, and 5.0) and γ_{adapt} (1.0–5.0) processed on Ψ_4, Ψ_6, and Ψ_8 over $H = [12 - 32]$ shows that the fastest convergence is reached by using γ_{adapt} in 3.30 iterations. This value increases by the use of constant values of γ ($\gamma = 1.0$: 3.45, $\gamma = 3.0$: 3.78, and $\gamma = 5.0$: 3.93). The use of γ_{adapt} shows that the adaptation of γ to confidence maps can be effectively applied to reduce the number of iterations and also to increase the classification accuracy.

When applied to Y_{sub}, the described hybrid Markov approach takes ~ 25 s of processing time on a hierarchical graph with eight levels and a total number of $9 * 10^4$ nodes (CPU details: Intel Xeon 5460 Core Duo @ 3.16-GHz processor with a 3-GB RAM). The HMAP estimator accounts for $\sim 70\%$, and the noncausal Markov model accounts for $\sim 30\%$ of this time. The generation of the graph takes more processing time (~ 60 s). In contrast, the automatic thresholding approach only requires ~ 30 s in deriving thresholds from the entire NCI data.

IV. CONCLUSION

In this paper, an unsupervised method for the extraction of flood-induced change areas in multitemporal SAR data has been presented. The experimental results obtained on a bitemporal TerraSAR-X data set of South West England, with an overall error (OER) of the class $C+$ of $\sim 7.5\%$, confirm the effectiveness of this approach.

A tile-based parametric thresholding approach under the GG assumption is successfully applied to an NCI image to automatically solve the three-class change detection problem in large-size data with small *a priori* class probabilities within a short time. The thresholding result is used in initializing a hybrid context-based model, combining causal with noncausal Markov image modeling on automatically generated irregular hierarchical graphs.

The integration of the hierarchical context in the labeling process is accomplished by the adaptation of HMAP estimation using Markov chains in scale. The supplementary integration of spatial-contextual information in the classification using noncausal MRF modeling leads to an increase in accuracy of up to 1.5% in comparison to the HMAP estimation.

To decrease the computational demand of noncausal Markov models, the energy minimization process is defined as the MRF approach, which is applied on the finest tree level to a restricted region specified by the use of confidence maps.

Due to its high accuracy and fast performance, this unsupervised hybrid change detection approach is useful for operational rapid mapping activities. It is important to note that, although this method has been presented in the specific context of the analysis of flood-induced change areas, it could be used in any change-detection application. Additionally, it has been successfully used for flood detection in single-polarized single-temporal SAR data. For future work, it would be an interesting topic to investigate the integration of elevation information into the process of extracting flood-related changes in feature maps. Therefore, methods using digital elevation models for improved flood detection [25], [38] have to be modified and adapted to this application.

REFERENCES

[1] Z. W. Kundzewicz, U. Ulbrich, T. Bruecher, D. Graczyk, A. Krueger, G. Leckebusch, L. Menzel, I. Pińskwar, M. Radziejewski, and M. Szwed, "Summer floods in Central Europe—Climate change track?" *Natural Hazards*, vol. 36, no. 1/2, pp. 165–189, Sep. 2005.

[2] S. Voigt, T. Kemper, T. Riedlinger, R. Kiefl, K. Scholte, and H. Mehl, "Satellite image analysis for disaster and crisis-management support," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1520–1528, Jun. 2007.

[3] P. Deer, "Digital change detection in remotely sensed imagery using fuzzy set theory," Ph.D. dissertation, Dept. Geograph. Comput. Sci., Univ. Adeleide, Adeleide, Australia, 1998.

[4] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1181, May 2000.

[5] G. Moser and S. B. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2972–2982, Oct. 2006.

[6] J. F. Mas, "Monitoring land-cover changes: A comparison of change detection techniques," *Int. J. Remote Sens.*, vol. 20, no. 1, pp. 139–152, Jan. 1999.

[7] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 874–887, Apr. 2005.

[8] F. Melgani, G. Moser, and S. B. Serpico, "Unsupervised change detection methods for remote sensing images," *Opt. Eng.*, vol. 41, no. 12, pp. 3288–3297, Dec. 2002.

[9] G. Moser and S. B. Serpico, "Unsupervised change detection from multichannel SAR data by Markovian data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2114–2128, Jul. 2009.

[10] Q. Jackson and D. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.

[11] T. Kasetkasem and P. K. Varshney, "An image change detection algorithm based on Markov random field models," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1815–1823, Aug. 2002.

[12] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, "Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 675–686, Mar. 2003.

[13] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 5, pp. 465–475, May 1997.

[14] S. Derrode, G. Mercier, and W. Pieczynski, "Unsupervised change detection in SAR images using a multicomponent HMC model," in *Proc. 2nd Int. Workshop Anal. Multi-Temp. Remote Sens. Images*, Ispra, Italy, 2003, vol. 3, pp. 195–203.

[15] C. Carincotte, S. Derrode, and S. Bourennane, "Unsupervised change detection on SAR images using fuzzy hidden Markov chains," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 432–441, Feb. 2006.

[16] C. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Process.*, vol. 3, no. 2, pp. 162–177, Mar. 1994.

[17] Z. Kato, M. Berthod, and Z. Zerubia, "A hierarchical Markov random field model and multi-temperature annealing for parallel image classification," *Graph. Models Image Process.*, vol. 58, no. 1, pp. 18–37, Jan. 1996.

[18] J. M. Laferté, P. Pérez, and F. Heitz, "Discrete Markov image modeling and inference on the quadtree," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 390–404, Mar. 2000.

[19] R. Wilson and C. Li, "A class of discrete multiresolution random fields and its application to image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 1, pp. 42–56, Jan. 2003.

[20] J.-N. Provost, C. Collet, P. Rostaing, P. Pérez, and P. Bouthemy, "Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps," *Comput. Vis. Image Understanding*, vol. 93, no. 2, pp. 155–174, Feb. 2004.

[21] C. Collet and F. Murtagh, "Multiband segmentation based on a hierarchical Markov model," *Pattern Recog.*, vol. 37, no. 12, pp. 2337–2347, Dec. 2004.

[22] M. Luettgen, W. Karl, and A. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," *IEEE Trans. Image Process.*, vol. 3, no. 1, pp. 41–64, Jan. 1994.

[23] W. Irving, P. Fieguth, and A. Willsky, "An overlapping tree approach to multiscale stochastic modeling and estimation," *IEEE Trans. Image Process.*, vol. 6, no. 11, pp. 1517–1529, Nov. 1997.

[24] A. Lopes, E. Nezry, R. Touzi, and H. Laur, "Maximum *a posteriori* speckle filtering and first order texture models in SAR images," in *Proc. IGARSS*, College Park, MD, 1990, vol. 3, pp. 2409–2412.

[25] S. Martinis, A. Twele, and S. Voigt, "Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data," *Natural Hazards Earth Syst. Sci.*, vol. 9, no. 2, pp. 303–314, 2009.

[26] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recog.*, vol. 19, no. 1, pp. 41–47, Jan./Feb. 1986.

[27] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.

[28] L. Lucchese and S. K. Mitra, "An algorithm for unsupervised color image segmentation," in *Proc. IEEE 2nd Workshop Multimedia Signal Process*, Redondo Beach, CA, 1998, pp. 33–38.

[29] R. M. Haralick and L. G. Sharipo, "Image segmentation techniques," *Comput. Vis., Graph., Image Process.*, vol. 29, no. 1, pp. 100–132, Jan. 1985.

[30] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2587–2600, Sep. 2006.

[31] M. Baatz and A. Schäpe, "Object-oriented and multi-scale analysis in semantic networks," in *Proc. 2nd Int. Symp. Oper. Remote Sens.*, Enschede, The Netherlands, 1999.

[32] *eCognition Developer 8 User Guide*, Definiens AG, Munich, Germany, 2009.

[33] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[34] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. R. Stat. Soc., B*, vol. 36, no. 2, pp. 192–236, 1974.

[35] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. London, U.K.: Springer-Verlag, 2009.

[36] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc., B*, vol. 48, no. 3, pp. 259–302, 1986.

[37] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. PAMI-6, pp. 721–741, Jun. 1984.

[38] D. C. Mason, R. Speck, B. Devereux, G. J.-P. Schumann, J. C. Neal, and P. D. Bates, "Flood detection in urban areas using TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 882–894, Feb. 2010.

Sandro Martinis received the M.Sc. degree in physical geography, physics, and remote sensing from the Ludwig-Maximilians-University of Munich, Munich, Germany, in 2006. He is currently working toward the Ph.D. degree in the German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Wessling, Germany.

From 2006 to 2007, he was a Research Associate with the Institute of Geology, Munich, where he worked on the development of remote sensing methods for glacier monitoring. He is involved in operational and method development activities at DLR/DFD's Center for Satellite-Based Crisis Information (ZKI). His main research activities are in the areas of operational flood extraction and change detection from synthetic aperture radar data.

André Twele received the M.Sc. degree in geography and applied geoinformatics from Westfälische Wilhelms-Universität Münster, Münster, Germany, in 2003.

As part of his M.Sc. thesis research, he spent eight months with the Joint Research Centre (JRC), European Commission, working on remote-sensing-based methods for burned-area assessment and post-fire regeneration monitoring. From 2003 to 2006, he was a Research Associate with the Cartography, GIS and Remote Sensing Department, Göttingen University, Göttingen, Germany, working in an international research program on rainforest monitoring. In 2007, he joined the German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Wessling, Germany, where he is currently coordinating a research project focused on operational methods for flood detection based on synthetic aperture radar remote sensing data. Furthermore, he is involved in method development activities at DLR/DFD's Center for Satellite-Based Crisis Information (ZKI), which is contributing to several national and international activities and projects in the field of disaster mitigation, humanitarian relief, and civil security.

Mr. Twele is a member of the Remote Sensing and Photogrammetry Society of Nottingham (RSPSoc).

Stefan Voigt received the M.Sc. degree in physical geography, physics, and remote sensing from the Ludwig-Maximilians-University of Munich, Munich, Germany, in 1997 and the Ph.D. degree from Berne University, Bern, Switzerland, in 2000, working on remote sensing for runoff/flood forecasting.

Since 2000, he has been with the German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Wessling, Germany, leading a research group and coordinating with the DLR Center for Satellite-Based Crisis Information (ZKI). From 2001 to 2008, he coordinated an international research initiative on innovative technologies for detection, prevention, and monitoring of uncontrolled coal seam fires in China. He is actively involved in numerous national and international research activities and projects in remote sensing for disaster mitigation, humanitarian relief, and civil security matters.