TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Methodik der Fernerkundung

Damage Assessment System of Linear Infrastructure Objects during Flooding using Probabilistic Graphical Models

Daniel Frey

Vollständiger Abdruck
der von der Fakultät für Bauingenieur- und Vermessungswesen
der Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigten Dissertation.

Vorsitzender:

Univ.- Prof. Dr.-Ing. Uwe Stilla

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. habil. Richard H. G. Bamler
2. Univ.-Prof. Dr. sc. techn. Daniel Straub
3. Univ.-Prof. Dr.-Ing. habil. Stefan Hinz
   Karlsruher Institut für Technologie

Die Dissertation wurde am 08.10.2012 bei der Technischen Universität München eingereicht und durch die Fakultät für Bauingenieur- und Vermessungswesen am 30.11.2012 angenommen.

Summary

In the thesis, a damage assessment system for linear infrastructure objects during flooding based on probabilistic graphical models is developed. The knowledge on the damage state of infrastructure objects during flooding is crucial for decision makers for the disaster management and a rapid response. Several information sources, such as remote sensing imagery, digital elevation models, water gauge measurements and information from a GIS about the infrastructural objects can contribute to a comprehensive damage assessment system. One part of the thesis deals with the fusion of these information sources based on probabilistic graphical models.

Probabilistic graphical models can be divided into Bayesian networks, which are directed graphical models, and Markov networks, which are undirected graphical models. The Bayesian network differs from the Markov network in their independence properties and therefore used for different problems. The representation and the independence properties are discussed in this work. In addition, the factor graph is introduced, which is a graphical representation used for inference. Both Bayesian networks and Markov networks can be transformed into factor graphs.

Furthermore, the state of the art of several damage assessment systems during flooding is discussed. The described damage assessment system can be divided into image-based and fusion-based systems. The image-based systems are based solely on image information. Different image analysis methods for assessing the flood scenario are described. The image-based systems can be subdivided into systems dealing with optical or radar imagery, systems combining both sensors and systems which focus on the development of a flood. On the other hand, fusion-based systems are focusing on the combination of different input data such as imagery and DEM. Further fusion-based methods integrate GIS data and hydrological models.

In this thesis, a system is developed which fuses different input data by means of Bayesian networks. The presented Bayesian network models the causal relations between the digital elevation model, the water gauge measurements, the image information and the damage state of the infrastructure object. This proposed causal Bayesian network is called pixel-based model since each pixel in the imagery is assessed individually and the context information is neglected. Beside the pixel-based model, a topology-based model is presented which considers additionally the context information of neighboring pixels. These neighborhood relations are modeled via a Markov chain representing the topology of linear infrastructural objects. The topology-based model combines the causal relation modeled as a Bayesian network and the symmetrical neighborhood relations based on a Markov network. Both networks are transformed into factor graphs, which is the basis for inference. The big advantage of the developed topology-based model is the opportunity to fuse different input data and simultaneously considering context information in one probabilistic model. Finally, a multi-temporal model is presented, which additionally integrates the information sources at different points in time.

All three models presented in this thesis, the pixel-based, the topology-based and the multi-temporal model are evaluated by means of two test scenarios. Since the stated problem can be considered as a detection problem (the infrastructure object is flooded or not flooded) receiver operator characteristic curves are used to demonstrate the performance of the different models. The evaluation shows that the fusion of the data via a Bayesian network can significantly improve the results compared to a flood simulation or a classification using only the imagery. In addition, the integration of the context in the topology-based model can further improve the damage assessment system. In case of timely limited occlusions in the image data such as clouds, the additional information from previous points in time can additionally increase the performance of the detection of flooded infrastructure objects.

Finally, results of the developed systems are shown and existing problems are discussed. Possible solutions of unsolved problems are discussed and further ideas expanding the developed models are described. A general framework for non linear infrastructure objects is given and possible methodological approaches are pointed out.

Zusammenfassung

In dieser Arbeit wird ein System zur Schadensdetektion von linearen Infrastrukturobjekten nach Überflutungen auf der Basis von probabilistischen graphischen Modellen präsentiert. Während Überflutungen ist der Zustand von Infrastrukturobjekten für Entscheidungsträger essentiell, um für ein angepasstes Katastrophenmanagement in kurzer Zeit zu sorgen. Verschiedene Informationsquellen, wie Fernerkundungsdaten, digitale Geländemodelle, Wasserpegelmessungen und Daten aus einem Geoinformationssystem über die zu bewertenden Infrastrukturobjekte können Informationen für ein System zur Schadensdetektion liefern. Ein Teil dieser Arbeit beschäftigt sich mit der Fusion dieser Informationen, basierend auf probabilistischen graphischen Modellen.

Probabilistische graphische Modelle können in gerichtete Bayessche Netze und ungerichtete Markov Netze unterteilt werden. Bayessche Netze und Markov Netze unterscheiden sich in ihren Möglichkeiten, Abhängigkeiten zwischen Zufallsvariablen zu modellieren und kommen somit bei unterschiedlichen Problemstellungen zum Einsatz. Zusätzlich werden Faktorgraphen eingeführt, die zur Inferenzberechnung verwendet werden. Sowohl Bayessche Netze als auch Markov Netze können in Faktorgraphen transformiert werden.

Außerdem wird ein Überblick über den Stand der Forschung von Systemen zur Schadensbewertung nach Überflutungen gegeben. Die beschriebenen Systeme zur Schadensbewertung können in bildbasierte Systeme und Systeme basierend auf Datenfusion unterteilt werden. Die bildbasierten Systeme verwenden ausschließlich Information, die aus den Bildern gewonnen werden kann. Verschiedene Methoden der Bildanalyse werden beschrieben, die es ermöglichen das Ausmaß von Überflutungen zu bewerten. Die bildbasierten Systeme können wiederum in Systeme unterteilt werden, die auf optische Bilddaten, Radardaten, der Kombination von beiden Daten und auf multitemporale Daten ausgelegt sind. Die Systeme, basierend auf Datenfusion, kombinieren hauptsächlich Bildinformation und digitale Geländemodelle. Weitere Systeme, basierend auf Datenfusion, integrieren Daten aus Geoinformationssystemen und hydrologische Modelle.

In dieser Arbeit wird die Fusion von verschiedenen Eingangsdaten mit Hilfe von Bayesschen Netzen durchgeführt. Das entworfene Bayessche Netz modelliert die kausalen Zusammenhänge zwischen dem digitalen Geländemodell, dem Wasserpegel, der Bildinformation und dem Schadenszustand der Infrastrukturobjekte. Dieses kausale Bayessche Netz wird „pixel-basiertes Modell“ genannt, da jedes Pixel im Bild für sich allein bewertet wird und keine Kontextinformation verwendet wird. Neben dem pixel-basierten Modell, wird in dieser Arbeit ein topologie-basiertes Modell vorgestellt, dass zusätzlich Kontextinformation von benachbarten Pixeln einbezieht. Diese Nachbarschaftsbeziehungen werden mit Hilfe von Markovketten modelliert, die die Topologie der linearen Infrastrukturobjekte wiederspiegelt. Somit kombiniert das topologie-basierte Modell die kausalen Beziehungen, die mit einem Bayeschen Netz modelliert sind, mit den symmetrischen Nachbarschaftsbeziehungen, die als Markovkette modelliert sind. Die Transformation der Netze in Fakorgraphen bildet die Basis für die Berechnung der Inferenz. Der große Vorteil des topologie-basierten Modells ist die Möglichkeit, in einem konsistenten statistischen Modell eine Fusion verschiedener Inputdaten durchzuführen und gleichzeitig Kontextinformation zu berücksichtigen. Das dritte entworfene multi-temporale Modell integriert die Informationsquellen von verschiedenen Zeitpunkten.

Alle drei in dieser Arbeit entwickelten Modelle: das pixel-basierte, das topologie-basierte und das multi-temporale Modell werden mit Hilfe von zwei Testszenarien evaluiert. Da es sich bei der Schadensbewertung von Infrastrukturobjekten um ein Detektionsproblem handelt (Infrastrukturobjekte sind überflutet oder nicht überflutet) werden zur Evaluierung Receiver Operating Characteristic Kurven herangezogen, um die Leistungsfähigkeit der entwickelten Modelle zu überprüfen. Die Evaluierung zeigt, dass die Datenfusion mit Hilfe von Bayesschen Netzen die Ergebnisse, im Vergleich zur Simulationen oder zu Klassifikationen, die nur auf den Bilddaten beruhen, deutlich verbessern kann. Zusätzlich führt die Integration von Kontext im topologie-basierten Modell zu Verbesserungen des Systems. Bei Verdeckungen in den Bilddaten, wie zum Beispiel durch Bewölkung, können Informationen von vorhergehenden Zeitpunkten die Leistungsfähigkeit des Systems zusätzlich steigern.

Neben den Ergebnissen werden auch die existierenden Probleme erläutert. Mögliche Lösungsvorschläge für ungelöste Probleme werden diskutiert und weitere Ideen zur Erweiterung der entwickelten Modelle werden beschrieben. Außerdem wird ein allgemeiner Rahmen zur Schadensdetektion für nicht lineare Infrastrukturobjekte unter der Verwendung von möglichen methodischen Ansätzen vorgeschlagen.

Contents

1 Introduction

1.1 Motivation

A *natural disaster*, which is some rapid, instantaneous or profound impact of the natural environment upon the socio-economic system (Alexander, 1993), is a major problem affecting man's lives, security, properties, infrastructure and development. Between 1980 and 2010 natural disasters are responsible for more than 2,275,000 fatalities and an overall loss of 3,000 billion US$ worldwide (MunichRe, 2011). Therefore, the understanding of the complex interrelations of natural disasters is of crucial importance. In addition, science and technology is challenged to develop tools reducing the effects of naturally occurring events (e.g., flood, hurricane, volcanic eruption, earthquake ...) and minimize the destruction they cause. Furthermore, the probability that climate change will exacerbate the situation is more than 66% (IPCC, 2007). Thus, it is important to strengthen our capability for disaster emergency management as well as to intensify scientific disaster mitigation research (Guo, 2009).

Disaster management can be divided into the preparedness before the disaster occurs and the response after disasters (e.g. emergency evacuation). In spite of all efforts to improve the preparedness and to minimize vulnerability to extreme events, these events will occur. Therefore, emergency response will always be a crucial part dealing with natural disasters. Harrald (2006) defines critical success factors that must be met if the operation of emergency response should be effective. One of these critical success factors is the ability to manage the collection, synthesis and analysis of information in a timely manner. However, the rapid and in-situ collection of data in the affected area is often difficult or even impossible due to the destruction of infrastructure, danger of access, remoteness of the disaster area and political restrictions.

Remote Sensing technology, such as earth observing satellites carrying optical and radar sensors, is independent of the destroyed on-site infrastructure and delivers current and area-wide information of the affected areas (Bamler et al., 2006). On one hand the whole process from the selection of satellites to the raw data processing have to be optimized. On the other hand the interpretation of remote sensed imagery, extraction of geometrically precise and semantically correct information as well as the production of maps need to be conducted as fast as possible. The main issue is the interpretation of the implicitly given information in the imagery and making it explicit. In addition, the uncertainties of the given image information have to be considered. Therefore, automatic image analysis methods have to be developed, focusing on the fast interpretation of the imagery.

However, the information obtained from remote sensing data is limited due to low spectral and spatial resolution or occlusions such as cloud coverage in optical imagery. *Additional information*, such as Digital Elevation Model (DEM) or the data from Geographic Information Systems (GIS) might significantly enhance the information value, if they are combined with imagery (De Gunst and Den Hartog, 1994; Baltsavias, 2004). The challenge using different kinds of data is to find methods combing the data in an appropriate way. In case of relief actions during natural disasters only immediately available data could be used, which are usually very diverse concerning the format, the spatial resolution and the information content. Therefore, methods have to be used, which are flexible to heterogeneous input data.

The *damage assessment of infrastructure* during natural disasters is of crucial importance to support emergency relief actions. Especially information about the operational reliability of transportation lifelines such as roads is important for rapid emergency response (Morain and Kraft, 2003). Therefore, fast damage assessment systems are needed for analyzing the trafficability of roads during natural disasters.

1.2 Goals and contribution of the thesis

The goal of the thesis is the development of a damage assessment system of linear infrastructure objects during flooding using probabilistic graphical models. The focus of the damage assessment system is the investigation of the trafficability of roads during flooding. However, also the operational reliability of other infrastructure objects, such as railroads or power supply systems could be assessed. In Fig. 1 assessed road networks are depicted grouped into three categories. The green lines indicate trafficable roads, the red lines indicate flooded roads and the yellow lines indicate the category possibly flooded, which means it is not clear if the roads are trafficable or flooded. Beside the flood state also the water depth of flooded roads is estimated.

Figure 1. Classification of roads during flooding in three categories: red = flooded roads, green = trafficable roads, yellow = possibly flooded road (left: IKONOS image, right: TerraSAR-X image).

The damage assessment system has to fulfill several requirements in order to guarantee a rapid and reliable emergency response.

- ◇ The handling of different kinds of image data (optical and radar) with varying spectral and spatial resolution.
- ◇ The integration of additional information such as DEM, data from GIS and in-situ measurements.
- ◇ The combination of the available data using a consistent statistical framework in order to quantify the uncertainties of the results.
- ◇ The handling of sequential data at different times in order to monitor the development of flooding.
- ◇ The processing time should be as fast as possible to guarantee a rapid response.
- ◇ The transferability of the system to different scenarios.

The goal of the thesis is to develop a system fulfilling all the listed requirements. The system consists of probabilistic graphical models, which provide a consistent statistical framework modeling uncertainties. Probabilistic graphical models enable the fusion of different kinds of information, which leads to an improvement of the analysis, and are flexible concerning the input data. Theoretically well-founded probabilities as well as purely subjective estimates can be used for the same network. This flexibility is important to integrate different kind of information and adapt the model to various scenarios.

The specific focus of the thesis is on the development of a probabilistic graphical model, which reflects the dependencies between the observations (e.g. imagery, DEM, in-situ measurements) and the state of roads (e.g. flooded or trafficable). Therefore a model is needed which combines causal physical relations, the topology of a road network and the dependencies over time. The

challenge is the embedding of all these different types of dependencies in one probabilistic model.

The main contribution of the thesis can be divided into two parts. Firstly, from the methodical point of view, it is shown that probabilistic graphical models provide a good framework to combine spatial information such as imagery and GIS data in consideration of the uncertainties of the data. The thesis points out that causal physical relations, the topology of spatial data and multi-temporal data can be described in one probabilistic graphical model. Secondly, it is shown, that a rapid assessment of roads after flooding is possible using remote sensing data. In this work, the limitations of remote sensing data and the benefit of additional information is discussed concerning the damage assessment system.

1.3 Organization of the thesis

The thesis is structured as follows:

In chapter 2 the basics of probabilistic graphical models are described. Probabilistic graphical models can be divided into two different types: directed and undirected graphical models. The representations as well as the independence properties of both graphical models are discussed separately. Afterwards the relations between directed and undirected models are described. Beside the representation of probabilistic graphical models, the most important step is the inference, which is discussed in Section 2.4. Inference in both types of graphical models can be conducted after the transformation into another graphical model called factor graph. The representation of factor graphs is discussed and finally, an algorithm called the sum-product algorithm is shown conducting inference in factor graphs.

In chapter 3 an overview on damage assessment systems during flooding based on the analysis of satellite imagery is given. The chapter is subdivided into methods focusing only on the analysis of imagery (image-based), and methods that combine the image information with additional data, such as DEM, GIS or hydrological models (fusion-based). Moreover the methods differ concerning the sensor (optical or radar). A lot of damage assessment systems are based on change detection techniques investigating multi-temporal imagery. Finally, the state of the art and the shortcomings of the described methods are discussed.

In chapter 4 the probabilistic graphical models assessing roads concerning the trafficability during flooding are presented. Firstly, a pixel-based model is presented, which combines the observation from remote sensing data with a DEM. The model is represented by means of a directed graph which is transformed into a factor graph. Secondly, the topology-based model takes the statistical dependence among neighboring pixels into account, which is modeled by means of an undirected graph. Thirdly, the time-domain is modeled in the multi-temporal model. The inference in all models is conducted via factor graphs.

In chapter 5 the proposed models are evaluated. Two different test scenarios are used to show the performance of the probabilistic graphical models. For both test scenarios, a reference is generated. For the evaluation of the results, receiver operator characteristic curves are used, since the stated problem can be formulated as a detection problem. The different proposed models are compared with approaches based only on simulation using only the DEM data and classification approaches using only the image data. Furthermore, the sensitivity and robustness of the three presented models in chapter 4 is investigated and advantages and shortcomings are discussed.

In chapter 6 the proposed probabilistic graphical models are discussed considering the stated goals of the thesis. Furthermore, several propositions are given to include further additional data and possible extensions of the models are discussed. Finally, open and unsolved problems and possible further investigations and research are pointed out in the outlook.

2 Basics of probabilistic graphical models

The calculus of probability theory is widely used in modern image analysis as well as in data fusion problems. The probability theory has the opportunity to deal with noisy images, can handle errors in observations and vague a priori information can be utilized. Especially in data fusion problems, the quality of the different kinds of data can be modeled using probabilistic approaches.

Models combining probability theory with graph theory are called probabilistic graphical models. The principal idea of probabilistic graphical models is the diagrammatic representation of complicated probability distributions by means of graphs. The representation of complex probability distributions using probabilistic graphical models has several useful properties:

Firstly, probabilistic graphical models provide a simple mechanism to visualize and exploit the structure of complex multi-dimensional distributions. A complex joint probability can be written down in a tractable way by exploiting the fact that in practice variables tend to interact directly only with few other variables. Therefore probabilistic graphical models enable the intuitive design and the effective construction of new models and facilitate the simple update of existing models.

Secondly, probabilistic graphical models are a transparent representation, which allows the evaluation of its semantics and properties. Especially, insights into the independence properties can directly be obtained by the analysis on the graphical structure.

Thirdly, the graphical structure can be exploited to perform and simplify inference. Only by changing the graph structure, complex computations for inference are carried out implicitly. Once the graph structure is fixed, effective inference algorithms are available for computing posterior probabilities of random variables given evidence on others.

Probabilistic graphical models are used to transform high-dimensional joint distributions in manageable smaller factors of a lower-dimensional space represented by a graph. Instead of modeling the high-dimensional joint distribution only the low-dimensional factors have to be assigned. The dependence structure of the variables determines the dimensions of the factors to be modeled. In many practical cases the variables tend to interact directly only with very few others. The graph represents the independence structure of the used variables.

In general, probabilistic graphical models can be divided into two different types: directed and undirected graphical models. Both types consist of nodes, which correspond to random variables, and links, which represent dependence among the random variables. The random variables can be divided into hypothesis variables and information variables. The hypothesis variables encode the variables we are interested in. However, it is usually not possible to observe them directly or the observation is too costly. The information variables encode the observed information in the graphical model (Jensen and Nielsen, 2007). In the following figures, all information variables are highlighted in blue. In the discrete case, each random variable consists of a finite set of mutually exclusive and exhaustive states. In the following the random variables are denoted as capital letters and the states are denoted as small letters. A set of random variables $\mathbf{X} \in \{A, B...\}$ is denoted as a capital bold letter and $\emptyset$ represents the empty set. The probability distribution of a random variable A with the states $a^1, a^2, ..., a^n$ (also denoted as $Val(A) = \{a^1, a^2, ..., a^n\}$) is expressed by $P(A) = (x_1, x_2, ..., x_n)$. The probability of A being in state a^i is x_i, which fulfills the basic rules:

$$x_i \geq 0, \qquad \sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n = 1. \tag{1}$$

The most established directed graphical models are causal Bayesian Networks, in which the

causal links are represented by conditional probabilities (Jensen and Nielsen, 2007). The second major class of probabilistic graphical models are undirected graphical models, also known as Markov Networks or Markov Random Fields, which are widely used in image processing (Blake et al., 2011). The Markov network theory provides a convenient and consistent way of modeling context-dependent entities such as neighboring image pixels (Li, 2009). Both Bayesian Networks and Markov networks are described in detail in Section 2.1 and 2.2. The relation between the two types of graphical models is discussed in Section 2.3. In Section 2.4 the issue of performing inference in graphical models is described. One possibility to perform inference is to convert directed and undirected graphs into factor graphs, which are a superclass of both types (Bishop, 2006). The representation of factor graphs and algorithms to perform inference in factor graphs are presented in Section 2.5.

2.1 Bayesian Network

Bayesian networks are probabilistic graphical models, which are mainly used to reflect causal relationship between random variables. Although the principle of causality, which reflects the relation between cause and effect, is not necessary, most of the developed models are based on causal Bayesian networks. The second kind beside the causal Bayesian network is the diagnostic Bayesian network. The difference between the two kinds of Bayesian network is limited to the different way of modeling but does not influence the mathematics. In the following, only causal Bayesian networks are considered. They provide a general methodology for data fusion combining theoretically well-founded probabilities, as well as subjective estimates. The combination of noisy measurements of different sensors is performed by Bayesian conditioning. The first applications of Bayesian networks are in the field of medical expert systems (Heckerman et al., 1992). The range of many real-world applications using Bayesian networks increased rapidly (Heckerman et al., 1995). Bayesian networks are well established as decision support systems in many applications. Several networks are developed for risk assessments concerning natural disasters such as hydrological issues (Molina et al., 2005; Park and Stenstrom, 2006), earthquakes (Bayraktarli et al., 2005) or avalanches (Grêt-Regamey and Straub, 2006; Straub and Grêt-Regamey, 2006).

2.1.1 Representation of Bayesian Networks

A Bayesian network is a directed acyclic graph consisting of nodes (also called vertices) and links (also called edges). The links of a directed acyclic graph are directed which represents the causality in Bayesian network. A graph is acyclic, if the graph has no loops. A loop exists in a directed graph if it is possible to start at a node A and follow the direction of the links and come back to A again. In Fig. 2 an example of a directed acyclic graph is shown.

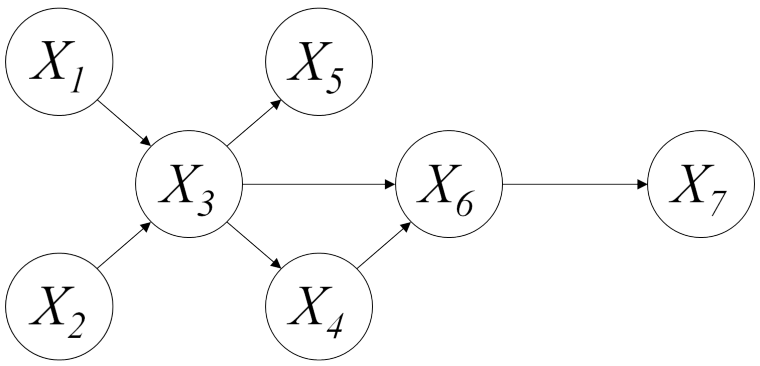

Figure 2. Example of a directed acyclic graph consisting of nodes and directed links

The nodes in a Bayesian network are random variables and the links connecting the nodes represent conditional probabilities. If there is a link directing from one random variable X_1 to another X_2 then X_1 is called parent of X_2, and X_2 is called child of X_1. All computations, which are applied in a Bayesian network, are based on two fundamental rules of probability. Firstly, the sum rule:

$$P(X_1) = \sum_{X_2} p(X_1, X_2) \tag{2}$$

whereas X_1 and X_2 are random variables and $p(X_1, X_2)$ is the joint probability. The distribution $P(X_1)$ is the marginal distribution obtained by marginalizing over the joint probability (marginalization). Secondly, the product rule:

$$p(X_1, X_2) = p(X_2|X_1)p(X_1) \tag{3}$$

with the conditional probability $p(X_2|X_1)$, which describes the probability of X_2 given X_1. The product rule, together with the symmetry property $p(X_1, X_2) = p(X_2, X_1)$ leads to the Bayes' theorem:

$$p(X_1|X_2) = \frac{p(X_2|X_1)p(X_1)}{p(X_2)} \tag{4}$$

The Bayesian network is a compact representation of a joint distribution. Let $p(X_1, X_2, ..., X_k)$ be a joint distribution over K variables, then it can be generally calculated by iteratively applying the product rule:

$$p(X_1, X_2, ..., X_K) = p(X_K|X_1, ..., X_{K-1})p(X_{K-1}|X_1, ..., X_{K-2})...p(X_2|X_1)p(X_1). \tag{5}$$

However, with increasing K the determination of the joint probability becomes rapidly intractably large. Bayesian networks provide the opportunity to describe the joint distribution in compact form exploiting the independence structure of the random variables. Considering the graphical structure of a Bayesian network the joint distribution can be obtained by:

$$p(X_1, X_2, ..., X_K) = \prod_{k=1}^{K} p(X_k|pa(X_k)), \tag{6}$$

whereas $pa(X_k)$ are the parents of X_k. Therefore the joint distribution of the example depicted in Fig. 2 is:

$$p(X_1, ..., X_7) = p(X_1)p(X_2)p(X_3|X_1, X_2)p(X_4|X_3)p(X_5|X_3)p(X_6|X_3, X_4)p(X_7|X_6), \tag{7}$$

Assuming the random variables in Fig. 2 are all binary valued $X_k \in \{x_k^0, x_k^1\}$ for $n = 1, 2...7$, the modeling and computation of the joint distribution by applying Equation 7 requires 16 independent parameters. In contrast, the full joint distribution would require $2^7 - 1 = 127$ independent parameters.

In the discrete case the links in Bayesian networks are conditional probabilities represented by conditional probability tables. The conditional probability tables fulfill the basic rules of probability theory stated in Equation 1.

If the random variables consist of continuous values, they are expressed by a probability density function with the property

$$\int_{-\infty}^{\infty} f(x)dx = 1. \tag{8}$$

The most commonly used continuous probability density function is the Gaussian distribution

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{9}$$

with the mean μ and the standard deviation σ. The D-dimensional Gaussian distribution is given by

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D\,|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{10}$$

where $\boldsymbol{\mu}$ is the D-dimensional mean, $\boldsymbol{\Sigma}$ is the covariance matrix of the size $D \times D$ and $|\boldsymbol{\Sigma}|$ denoting the determinant of $\boldsymbol{\Sigma}$.

Bayesian networks consisting of both discrete and continuous variables are called hybrid Bayesian networks (Lerner, 2002). If a continuous random variable has discrete parents the parameters of the probability density function have to be specified for each state combination of the discrete parents. If the probability density function is a Gaussian, the resulting table consists of different means and variances for each state combination (Jensen and Nielsen, 2007). If a discrete child has continuous parents, the simplest approach is the threshold model. A simple threshold in the continuous domain is selected to define the probabilities of the states for discrete variable. However, the change of probability is discontinuous, which leads to inconvenience from a mathematical perspective (Koller and Friedman, 2009). Another opportunity to circumvent hybrid Bayesian networks is the discretization of all continuous variables. However, on one hand the characteristic structure of the continuous variable is lost and a fine discretization leads rapidly to memory problems.

It is also possible to introduce deterministic relations in a Bayesian network, which often occur in natural cases. A deterministic relation of a random variable to its parents is denoted via an additional cycle in a node as depicted in Fig. 3. In the example the variable C is a deterministic function of its parents A and B (Cobb and Shenoy, 2004). In the binary case the deterministic relation could reflect easy operations such as "or". In the continuous domain, $P(C|A,B)$ is determined by a deterministic function such as C is equal to $A+B$.

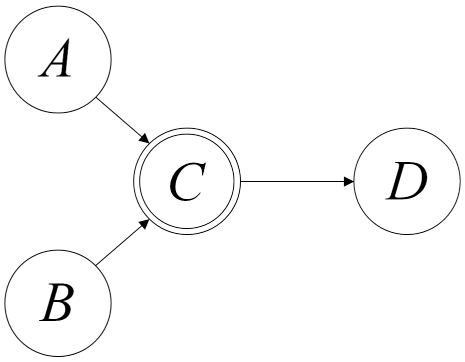

Figure 3. BN with deterministic node C

A further extension of the basic Bayesian network is the additional representation of parameters. The previous representation of a Bayesian network consists only of random variables and the conditional probabilities between them. However, in many cases additional parameters are given, which for example describe the noise variance. In the graphical representation of Bayesian

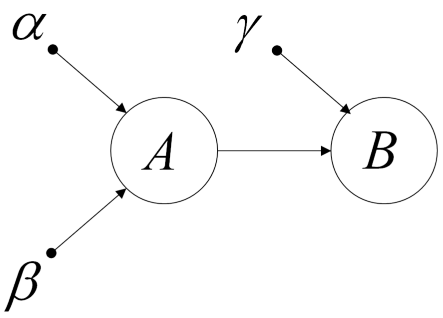

Figure 4. Bayesian network with additional parameters represented as small solid circles

networks, the additional given parameters are visualized by small solid circles as depicted in Fig. 4. The graphical representation of the parameters can also be adapted to the formulas. The Equation 11 shows the joint distribution of the graphical model represented in 4, which explicitly reads as:

$$p(A, B|\alpha, \beta, \gamma) = p(A|\alpha, \beta)p(B|A, \gamma). \tag{11}$$

2.1.2 Independence properties

Beside the property of a Bayesian network to represent a joint distribution, another characteristic of Bayesian networks is the encoding of independence assumptions. A Bayesian network is able to make conditional independence assumptions of several random variables visually readable from a graph. The concept of conditional independence is crucial in unifying many seemingly unrelated random variables (Dawid, 1979). Two random variables A and B given C are conditional independent if

$$p(A, B|C) = p(A|C)p(B|C). \tag{12}$$

An more intuitive interpretation is that the distribution of A given B and C depends only on C and further information about B is irrelevant

$$p(A|B, C) = p(A|C). \tag{13}$$

In the following the conditional independence of A and B given C is denoted in shorthand

$$A \perp\!\!\!\perp B \mid C \tag{14}$$

or if A and B are not conditional independent given C it is denoted as

$$A \not\perp\!\!\!\perp B \mid C. \tag{15}$$

For the analysis of the conditional independence structure the concept of d-separation is introduced, which allows to read the conditional independence properties directly from the graph (Pearl, 1986, 1988). The conditional independence structure of three example graphs is investigated to motivate the concept of d-separation. The three example graphs are depicted in Fig. 5 consisting of a diverging, serial and converging connection. In the following the conditional dependence of A and B given C are investigated.

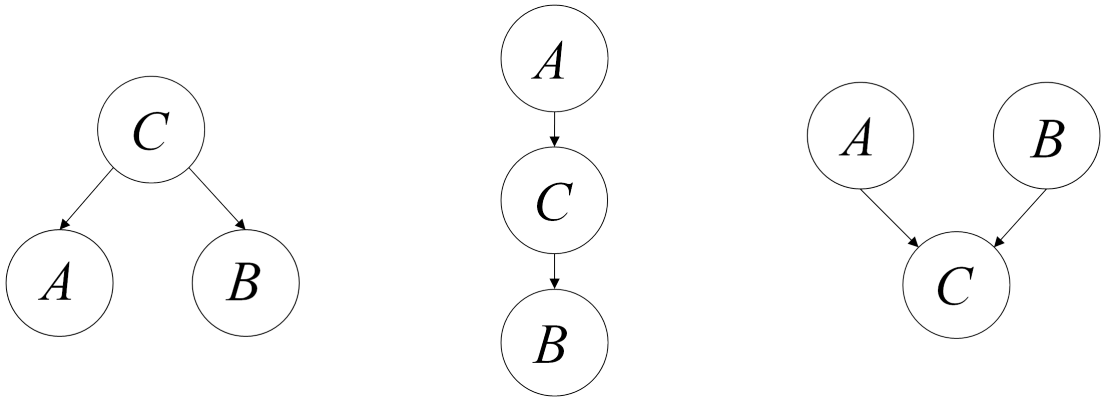

Figure 5. Three example graphs representing the basic connections of a Bayesian network (left: diverging connection, middle: serial connection, right: converging connection).

Considering the diverging connection (Fig. 5 left), the joint probability derived from the graph is

$$p(A, B, C) = p(A|C)p(B|C)p(C). \tag{16}$$

Using the product rule and inserting the joint probability of the diverging connection leads to

$$p(A, B|C) = \frac{p(A, B, C)}{p(C)} = p(A|C)p(B|C), \tag{17}$$

which states the conditional independence property in the diverging connection is

$$A \perp\!\!\!\perp B \mid C. \tag{18}$$

A similar behavior is obtained considering the serial connection (Fig. 5 middle) with the joint probability

$$p(A, B, C) = p(A)p(C|A)p(B|C). \tag{19}$$

Again, the substitution of the joint probability with Equation 19 and applying Bayes' theorem leads to the expression

$$p(A, B|C) = \frac{p(A, B, C)}{p(C)} = \frac{p(A)p(C|A)p(B|C)}{p(C)} = p(A|C)p(B|C), \tag{20}$$

which again shows that

$$A \perp\!\!\!\perp B \mid C \tag{21}$$

in the serial connection. For both example graphs, the diverging and serial connection, it can be shown by marginalizing over C that in general

$$A \not\perp\!\!\!\perp B \mid \emptyset \tag{22}$$

where $\emptyset$ indicates that none of the variables is observed.

Applying the same operations to the converging connection (Fig. 5 right), with the joint probability

$$p(A, B, C) = p(A)p(B)p(C|A, B) \tag{23}$$

leads to

$$p(A, B|C) = \frac{p(A, B, C)}{p(C)} = \frac{p(A)p(B)p(C|A, B)}{p(C)} \neq p(A|C)p(B|C), \tag{24}$$

and therefore

$$A \not\perp\!\!\!\perp B \mid C. \tag{25}$$

However, if C is not given, the joint probability $p(A, B)$ is obtained by marginalizing both sides of Equation 23

$$p(A, B) = p(A)p(B) \tag{26}$$

and therefore

$$A \perp\!\!\!\perp B \mid \emptyset. \tag{27}$$

The independence properties, presented in the three example graphs can be summarized and formulated in the definition of d-separation. Two variables A and B in a causal network are d-separated, if for all possible paths from node A to node B, there is an intermediate variable C such that either

- ◇ the connection is serial or diverging and C is observed or
- ◇ the connection is converging, and neither C nor any of C's descendants are observed.

By means of the d-separation property it is possible to investigate in directed graphs conditional independence between two distinct variables given observed variables. Lauritzen (1996) proofs that the d-separation property is applicable to Bayesian networks. The soundness and completeness of d-separation is shown by Geiger and Pearl (1988). In case of Gaussian distributions the completeness is shown by Geiger and Pearl (1993).

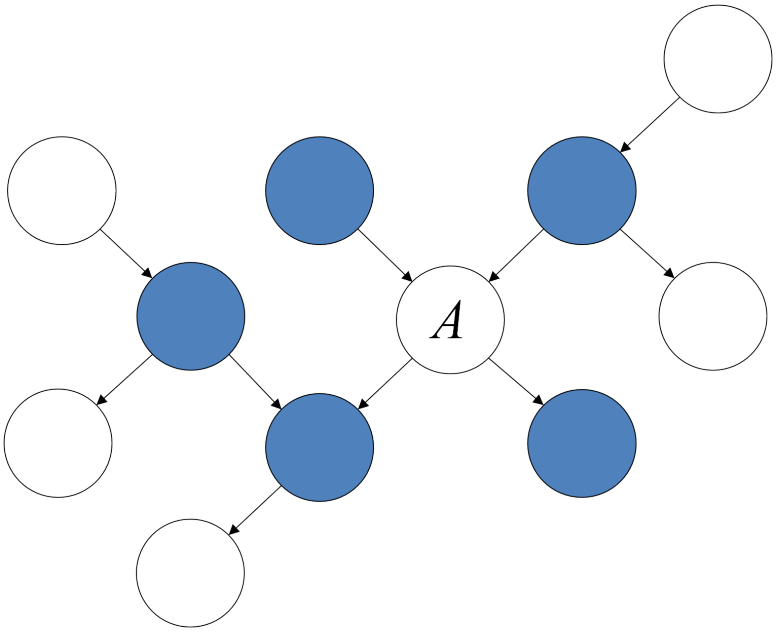

Figure 6. The Markov blanket (colored in blue) of a random variable A.

The d-separation concept describes global independence properties. Beside this global independence property also a local independence property can be defined by means of the Markov blanket. The Markov blanket for a random variable A in a Bayesian network is the set of random variables $\mathbf{MB}(A)$, consisting of the parents of A, the children of A and the variables sharing a child with A. If the Markov blanket of A is given, then A is d-separated from the rest of the network

$$A \perp\!\!\!\perp \mathfrak{A}_{\setminus\{A,\mathbf{MB}(A)\}} \mid \mathbf{MB}(A), \tag{28}$$

where $\mathfrak{A}_{\setminus\{A,\mathbf{MB}(A)\}}$ is the set of all nodes in the graph without A and the set $\mathbf{MB}(A)$. In Fig. 6 the Markov blanket of a random variable A is depicted in blue.

2.2 Markov Network

The second major class of probabilistic graphical models is the Markov Network or the Markov Random Field (MRF), which is undirected graphical model. A Markov network is - like a Bayesian network - a graphical model, which is able to represent the joint probability of a set of random variables and specify conditional independence properties. In contrast to Bayesian Networks, which are used to model causal relation between random variables, the Markov Network is mainly used to model symmetrical interactions between random variables, such as neighboring pixels in imagery. Therefore, Markov Networks are widely used to model imagery (Perez, 1998; Li, 2009). Several tasks, such as image denoising (Geman and Geman, 1984; Besag, 1986), stereo reconstruction (Szeliski et al., 2008) and image segmentation (Bouman and Shapiro, 1994), can be modeled by applying Markov Networks. An overview about Markov Networks in the field of computer vision is given in Li (2009).

2.2.1 Representation of Markov Networks

Markov networks are undirected graphical models consisting of nodes and links. As in Bayesian networks the nodes correspond to random variables. The links are in contrast to Bayesian networks not directed and capture mostly symmetrical relations between random variables. An example of a Markov network is shown in Fig. 7.

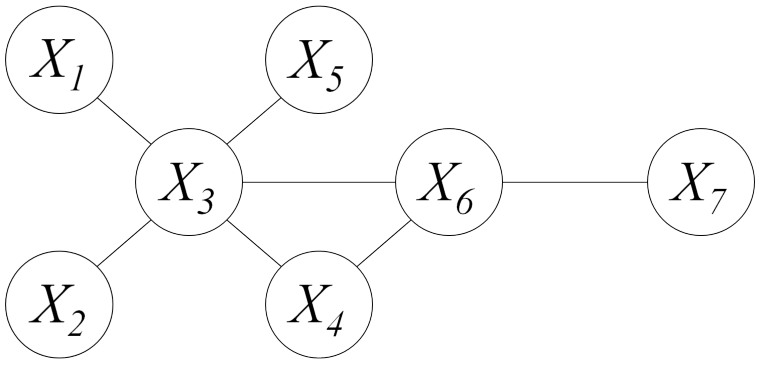

Figure 7. Example of a Markov network

The graphical structure of a Markov Network is as in Bayesian Networks a representation of the joint probability and reflects the independence structure of the random variables. In order to derive the factorization of the joint distribution the graphical concept of cliques is required. A clique is defined as a subset of nodes in a graph such that there is a link between all pairs of nodes in the subset (Bishop, 2006). A maximal clique is a clique that cannot be extended by including one more adjacent node from the graph in the set, meaning it is not a subset of

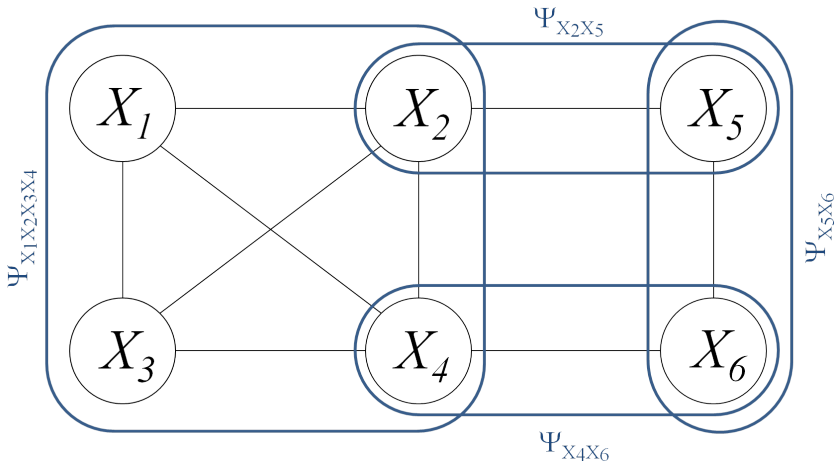

Figure 8. Example of a Markov Network with 14 cliques and 4 maximal cliques highlighted with blue framing

a larger clique. The example in Fig. 8 demonstrates the concepts of cliques. This example consists of 14 cliques and 4 maximal cliques given by $\{X_1, X_2, X_3, X_4\}$,$\{X_2, X_5\}$,$\{X_4, X_6\}$ and $\{X_5, X_6\}$.

By means of the maximal cliques, the joint distribution of a Markov Network can be formulated by means of potential functions $\psi_C(X_C)$ over the maximal cliques

$$p(X_1, X_2, ..., X_K) = \frac{1}{Z} \prod_C \psi_C(\mathbf{X}_C), \tag{29}$$

where the quantity Z is called the partition function and ensures the normalization of the joint distribution. The partition function Z is given by

$$Z = \sum_{X_1} ... \sum_{X_K} \prod_C \psi_C(\mathbf{X}_C). \tag{30}$$

The joint distribution of the example depicted in Fig. 8 is therefore

$$\begin{aligned} p(X_1, X_2, X_3, X_4, X_5, X_6) = \frac{1}{Z} & \psi_{X_1X_2X_3X_4}(X_1, X_2, X_3, X_4) \psi_{X_2X_5}(X_2, X_5) \\ & \psi_{X_4X_6}(X_4, X_6) \psi_{X_5X_6}(X_5, X_6). \end{aligned} \tag{31}$$

The potential functions $\psi_C(X_C)$ are restricted to positive functions. Only if $\psi_C(X_C)$ is positive it is guaranteed that there exists a precise relationship between the factorization shown in Equation 29 and the conditional independence properties discussed in the next section (Clifford, 1990). Since $\psi_C(X_C)$ are positive it is convenient to express the potential functions by means of exponentials

$$\psi_C(\mathbf{X}_C) = \exp\left(-U(\mathbf{X}_C)\right), \tag{32}$$

where $U(X_C)$ is called the energy function. The potential functions capture the affinities of the random variables, which are members of the same clique X_C. In general, they do not correspond to probabilities, which make it difficult to understand them intuitively. However, in special cases the potential functions can be interpreted as conditional probabilities. In principle, it can be assumed the more compatible the variables in a clique, the larger the value of the potential functions. The framework is not restricted to discrete variables. By replacing the sum in Equation 30 with integral also continuous random variables can be treated.

2.2.2 Independence properties

As in Bayesian network global and local independence properties also exist in Markov networks. Corresponding to the d-separation for Bayesian networks, a simpler global independence property in the Markov network can be formulated. Assuming there are three set of nodes **A**, **B** and **C**, then the given set **C** separates **A** and **B**, if there is no path between any node $A \in \mathbf{A}$ and $B \in \mathbf{B}$. The concept of this global independence property is demonstrated in Fig. 9, which leads in the example graph to the conditional independence property

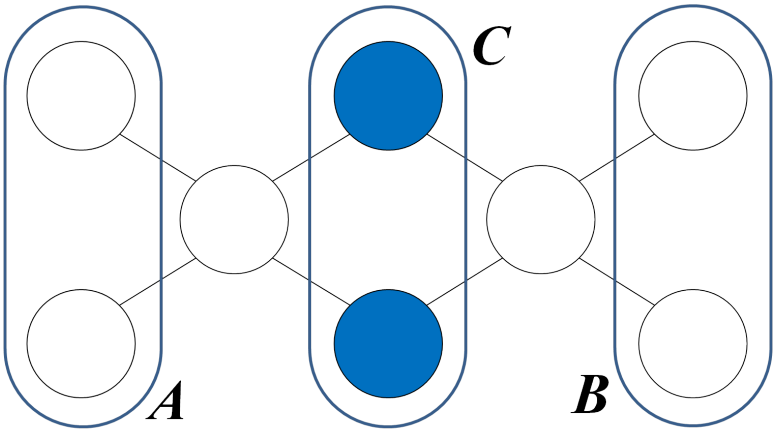

Figure 9. The global independence property state that every node in the set **A** is conditional independent on every node in the set **B** if **C** is given

$$A \perp\!\!\!\perp B \mid \mathbf{C}. \tag{33}$$

The local independence property of a Markov Network can again be expressed by the Markov blanket. In Markov networks the Markov blanket is simply all the neighbors of a node. In Fig. 10 the Markov blanket of a node A is shown. This leads to the local independence assumption, which states that a node is independent of all other nodes in the graph given all its neighbors. Therefore in Fig. 10 the random variable A is independent of all other random variables given the nodes highlighted in blue. If the potential functions are positive, the global independence property and the local independence property are equivalent (Lauritzen, 1982; Koller and Friedman, 2009).

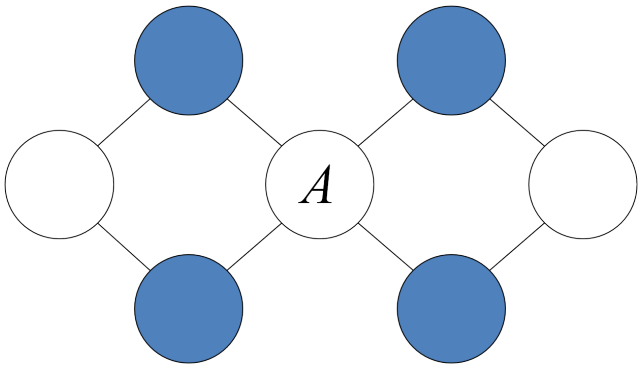

Figure 10. The Markov blanket of A (colored in blue) in a Markov network

2.3 Relation between graphical models

Both Bayesian networks and Markov networks are graphical representations of joint distributions and each reflects different kinds of independence properties. In the following the joint

distribution represented in a Bayesian Network, which is factorized as formulated in Equation 6 should be transformed in a factorization used for Markov Networks (see Equation 29). The transformation is simply done by expressing the potential functions $\psi_C(X_C)$ by the conditional distribution of the directed graph. The scope of the potential functions consists of a random variable X_k and its parents $pa(X_k)$. These potential functions satisfy additional normalization property leading to a partition function of $Z = 1$, which holds not in general. From the graphical perspective all the random variables appearing in the conditional distribution have to be members of at least one clique of the undirected graph. This leads to the following rule for converting a directed acyclic graph in an undirected graph, which is called "moralization". Firstly, all parents, which have a common child, are connected by an undirected link. Secondly, all other directed links are replaced by undirected links. In Fig. 11 an example for moralization is shown. The converted graph is called moral graph.

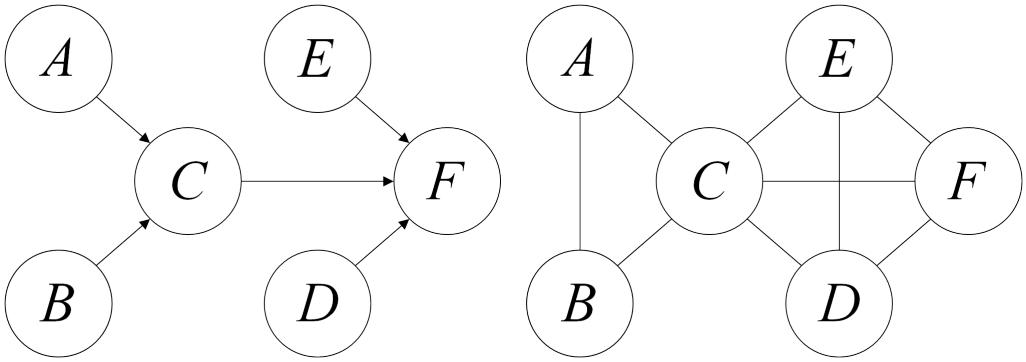

Figure 11. Example of the conversion of a directed graph into an undirected graph via moralization. (left: directed graph, right: moral graph)

In general, some independence properties encoded in Bayesian networks are lost after the transformation. However, using moralization the minimum number of extra links is added and the maximum independence properties are retained. Bayesian networks and Markov networks represent generally different kinds of independence properties. If a graph is able to describe every independence property of a given distribution, then the graph is called a perfect map. For some distribution a prefect map can be obtained only by means of Bayesian networks and for other distributions only Markov networks deliver perfect maps. But there are also distributions for which both graphical models can deliver perfect maps or neither of them is able to represent the independence properties in perfect maps. The set diagram in Fig. 12 illustrates the different possibilities of representing independence properties by the different graphical models and shows three examples of the different sets.

On the left side in Fig. 12 the independence properties $A \perp\!\!\!\perp B|\emptyset$ and $A \not\perp\!\!\!\perp B|C$ are encoded by a Bayesian network. Only a Bayesian Network is able to represent these independence properties. In contrast, the Markov network depicted on the right side in Fig. 12 represents the independence properties $A \not\perp\!\!\!\perp B|\emptyset$, $A \perp\!\!\!\perp D|B \cup C$ and $B \perp\!\!\!\perp C|A \cup D$, which cannot be encoded by a Bayesian network. The independence properties $A \not\perp\!\!\!\perp C|\emptyset$ and $A \perp\!\!\!\perp C|B$ can be represented by means of a Bayesian network as well as by a Markov Network. Replacing the directed links by undirected links in the graph depicted at the bottom of Fig. 12 fulfills the same independence properties.

Beside Bayesian networks and Markov Networks also partially directed models exists, which consist of directed and undirected links. Partially directed acyclic graphs are called chain graphs (Frydenberg, 1990). Similar to Bayesian networks and Markov networks a factorization of the joint probability is given and independence properties can be derived from the graphical structure (Lauritzen, 1996; Studený and Bouckaert, 1998).

Special Markov networks with directed dependencies on a subset of random variables are conditional random fields (Lafferty, 2001) or discriminative random fields (Kumar and Hebert,

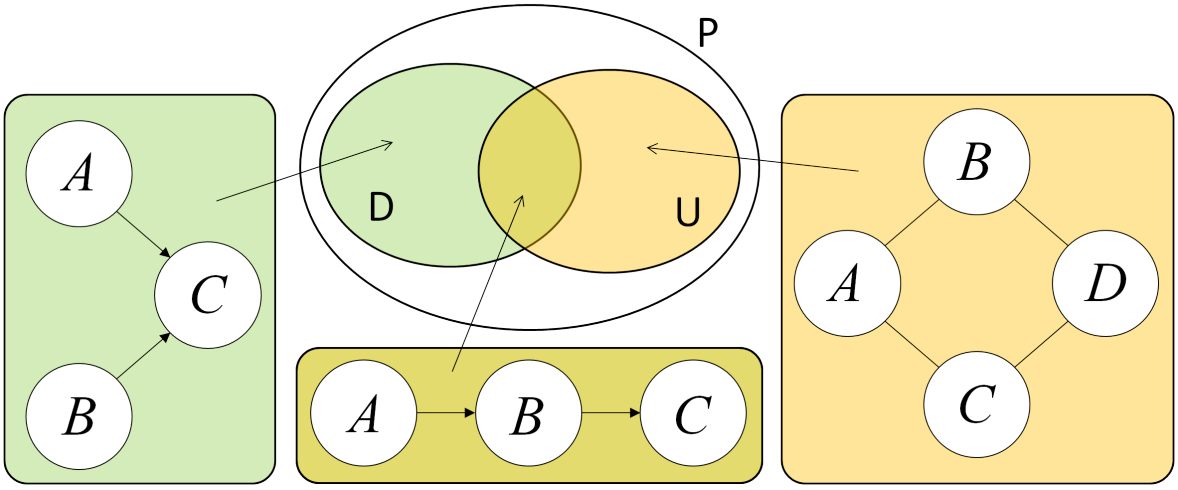

Figure 12. Illustration of the set of all distributions P over a given set of variables. The set D represents the distribution, which can be modeled in a prefect map by Bayesian networks and respectively the set U for Markov networks (Bishop, 2006). In addition three example graphs for different sets are illustrated.

2006). Both models are mainly used for modeling sequential data and can be considered as an undirected alternative to hidden Markov models (Rabiner and Juang, 1986). An overview of conditional random fields and some of its applications can be found in Sutton and McCallum (2007).

2.4 Inference in Bayesian networks and Markov networks

In the previous section Bayesian networks and Markov networks are introduced and the relations of the two probabilistic graphical models are discussed. In the following sections the computation of the posterior distributions of the latent hypothesis variables given the observed information variables is presented. This process is called inference. In general, the algorithms performing inference can be divided in exact inference algorithms and approximate algorithms. Most of the approximate algorithms can be considered as optimization problems and are often formulated as energy functionals (Yedidia et al., 2005). However, in the following, the focus is on methods computing exact inference, since the developed model in the thesis can be solved by exact inference. A general method computing exact inference in Bayesian or Markov networks is the junction tree algorithm (also called clique tree algorithm) (Lauritzen and Spiegelhalter, 1988; Shafer and Shenoy, 1990). The first step of the junction tree algorithm in case of Bayesian networks is the transformation to an undirected graphical model via moralization as described in Section 2.3. The resulting moral graph is then transformed in a junction tree, by means of the maximal cliques. The junction tree has a tree structure, which means that there exists one, and only one, path between any pair of nodes. By means of the junction tree, the elimination ordering of the random variables can be determined. A detailed description of the junction tree algorithm is given in Jensen and Nielsen (2007). In the next section an exact inference algorithm is presented based on factor graphs. A factor graph may be viewed as an alternative approach to junction trees. The results obtained by inference algorithms using junction trees may be translated into equivalent results using inference algorithms in factor graphs, and vice versa (Kschischang et al., 2001). Both Bayesian networks and Markov networks can be transformed into factor graphs. The graphical structure of the factor graphs is used to conduct inference.

2.5 Factor graphs

Factor graphs originate from Tanner graphs, which are graphs describing families of codes (Tanner, 1981). Factor graphs can be considered as a generalization of Tanner graphs (Wiberg et al., 1995) and were firstly introduced by Frey et al. (1997). As Bayesian and Markov networks also factor graphs are graphical representations of joint distributions. Both, Bayesian and Markov networks can be transformed into factor graphs with a corresponding factorization (Loeliger, 2004). In the next section the representation of factor graphs is discussed and in Section 2.5.2 the sum-product rule, which is an algorithm to conduct inference in factor graphs, is presented.

2.5.1 Representation of factor graphs

In contrast to Bayesian and Markov networks, factor graphs are bipartite graphs consisting of two types of nodes: variable nodes corresponding to the random variables and factor nodes describing the mathematical relation between the random variables. An example of a factor graph is depicted in Fig. 13.

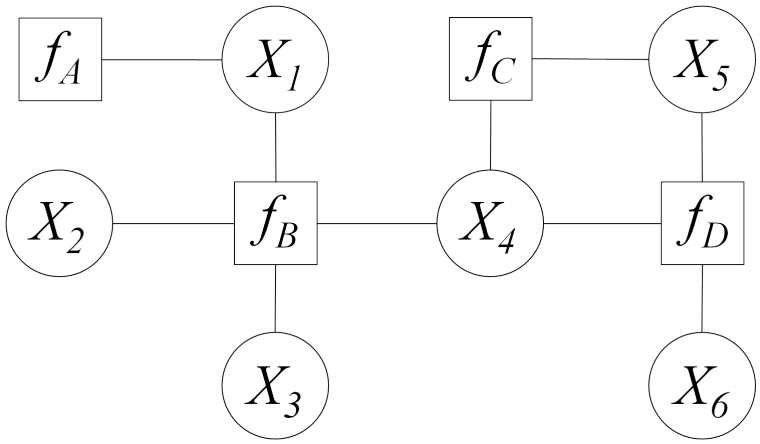

Figure 13. Example of a factor graph

The variable nodes are equivalent to the variable nodes in Bayesian and Markov networks. The factor nodes are local functions, which represent explicitly the mathematical relation between the random variables, which in the following are simply called factors. The joint distribution in a factor graph is simply the product of all factors

$$p(X_1, X_2, ..., X_K) = \prod_s f_s(\mathbf{X}_s), \tag{34}$$

where f_s are the factors dependent on a subset of variables $\mathbf{X}_s$. Referring to Equation 34 the factorization of the example graph depicted in Fig. 13 is

$$p(X_1, X_2, ..., X_6) = f_A(X_1) f_B(X_1, X_2, X_3, X_4) f_C(X_4, X_5) f_D(X_4, X_5, X_6). \tag{35}$$

Factor graphs can be derived from Bayesian and Markov networks. In case of Bayesian networks the directed graph is firstly converted to an undirected graph via moralization. Then the moral graph or respectively the Markov network is transformed into a factor graph, where the factors represent conditional probabilities or respectively the potential functions. In Fig. 14 an example of a transformation from a Bayesian network into a factor graph is shown. The factors correspond to the conditional probabilities.

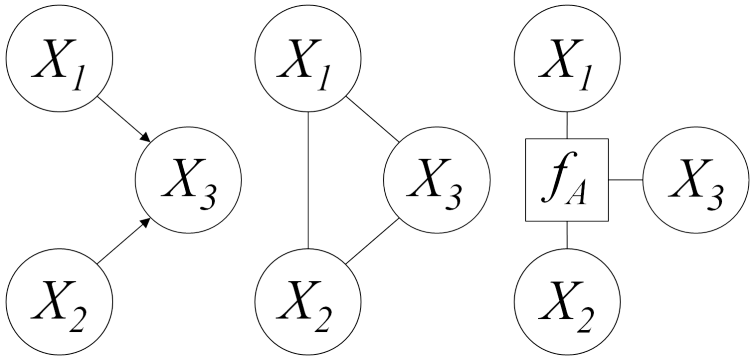

Figure 14. Example of the transformation from a Bayesian network to a factor graph

If a factor graph is tree-structured, which means there exists one, and only one, path between any pair of variable nodes, then exact inference is possible. If a factor graph contains loops, approximate inference is necessary. Approaches to solve approximate inference are variational methods (Sakurai, 1985) or loopy belief propagation (Frey and MacKay, 1998). In the following only tree-structured graphs are considered.

2.5.2 Inference in factor graphs

The most established algorithm to conduct inference in a tree-structured factor graph is the sum-product algorithm. The sum-product algorithm was originally invented by Gallager for low-density parity check codes (Gallager, 1963). The exact inference algorithm for Bayesian networks known as belief propagation (Pearl, 1988; Lauritzen and Spiegelhalter, 1988) can be translated into an instance of the sum-product algorithm, with an equivalent factorization expressed in a factor graph. Kschischang et al. (2001) shows that also the forward-backward algorithm (Rabiner, 1989), the Viterbi algorithm (Viterbi, 1967) and the Kalman filter (Kalman, 1960) are instances of the sum-product algorithm.

Firstly, the sum-product rule is presented to compute the marginal probability of one random variable followed by the description of an efficient computation of the marginal probabilities of a set of random variables. Furthermore the algorithm is presented for the discrete case. However, the algorithm is also applicable for continuous variables, in which the integration replaces performing sums. The following notation to derive the sum-product algorithm is taken from Bishop (2006).

The marginal probability of a random variable $p(X)$ is obtained by marginalizing over all other random variables except X

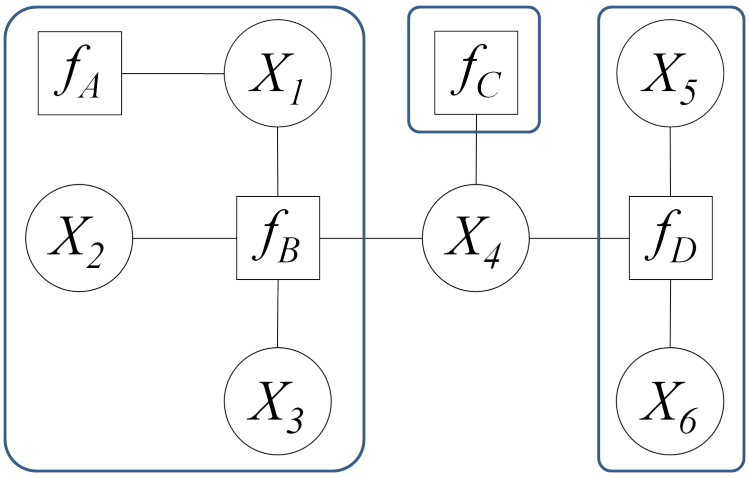

Figure 15. A factor graph divided into 3 variable-subgraphs (blue boxes) referring to variable X_4

$$p(X) = \sum_{\mathbf{X}\backslash X} p(\mathbf{X}), \tag{36}$$

where $\mathbf{X}\backslash X$ denotes all variables in $\mathbf{X}$ without X. The joint distribution in factor graphs is represented as a product of all factors in the graph as formulated in Equation 34. Focusing on a specific hypothesis variable X the factor graph can be divided into several variable-subgraphs as depicted in Fig. 15. The number of variable-subgraphs is equal to the number of neighboring factor nodes of X. Therefore every variable-subgraph can be associated to a factor node, which is a neighbor of the variable. The product of all factors assigned to a variable-subgraph is denoted as $F_s(X, \mathbf{X}_s)$, where $\mathbf{X}_s$ are all the random variables in the variable-subgraph. The joint distribution could therefore be formulated as

$$p(\mathbf{X}) = \prod_{s \in ne(X)} F_s(X, \mathbf{X}_s), \tag{37}$$

where $ne(X)$ denotes the set of factor nodes that are neighbors of X. Using the joint probability given in 37, inserting it in Equation 36 and interchanging the products and sums lead to

$$p(X) = \prod_{s \in ne(X)} \sum_{\mathbf{X}_s} F_s(X, \mathbf{X}_s). \tag{38}$$

In the following a message $\mu_{f_s \to X}(X)$ is defined as the sum of the product of all factors in a variable subgraph

$$\mu_{f_s \to X}(X) \equiv \sum_{\mathbf{X}_s} F_s(X, \mathbf{X}_s), \tag{39}$$

where $\mu_{f_s \to X}(X)$ propagates from the factor node f_s to the random variable X. Therefore the marginal probability $p(X)$ is the product of all incoming messages arriving at node X.

$$p(X) = \prod_{s \in ne(X)} \mu_{f_s \to X}(X). \tag{40}$$

The product of all factors $F_s(X, \mathbf{X}_s)$ in a variable-subgraph can be further divided into the product of the factor f_s and the factors in M factor-subgraphs as depicted in Fig. 16

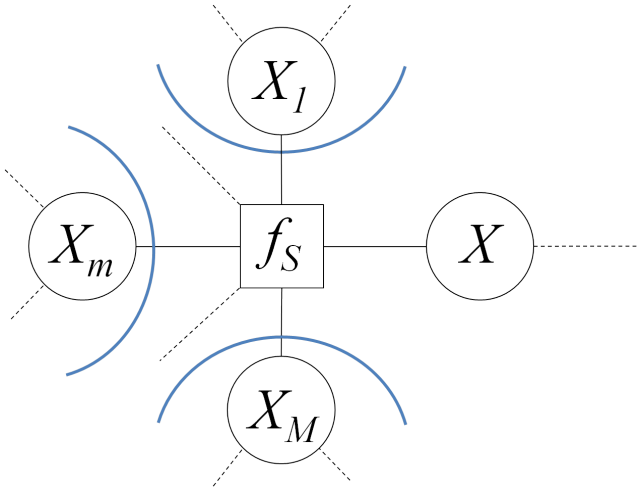

Figure 16. Division of a factor graph in several factor-subgraphs indicated in blue

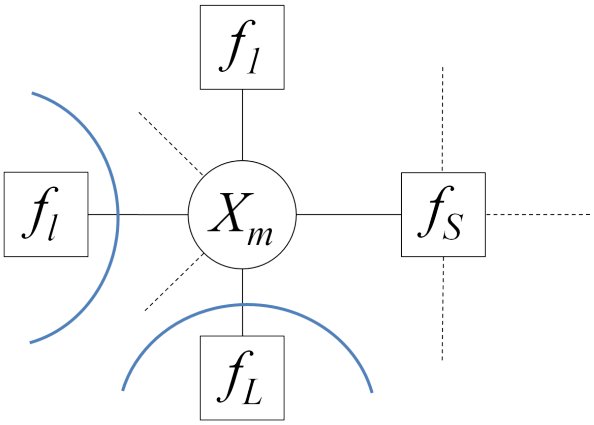

Figure 17. Variable-subgraphs of random variable x_m

$$F_s(X, \mathbf{X}_s) = f_s(X, X_1, ..., X_m, ..., X_M)G_1(X_1, \mathbf{X}_{s1}) \cdot ... \cdot G_m(X_m, \mathbf{X}_{sm}) \cdot ... \cdot G_M(X_M, \mathbf{X}_{sM}). \quad (41)$$

The product of all factors in a factor-subgraph is represented by $G_m(X_m, \mathbf{X}_{sm})$. The random variables $X, X_1, ..., X_m, ..., X_M$ are all the neighbors of the factor f_s as depicted in Fig. 16 and $\mathbf{X}_{sm}$ is the set of all random variables in the factor-subgraph m. Substituting Equation 41 into 39 the message from the factor f_s to X is computed

$$\mu_{f_s \to X}(X) = \sum_{X_1} ... \sum_{X_M} f_s(X, X_1, ..., X_m, ..., X_M) \prod_{m \in ne(f_s) \setminus X} \left[\sum_{\mathbf{X}_{sm}} G_m(X_m, \mathbf{X}_{sm}) \right] \quad (42)$$

where $ne(f_s)$ denotes variable nodes that are neighbors of f_s. Again a message is defined this time from the random variable X_m to the factor f_s as follows

$$\mu_{X_m \to f_s}(X_m) \equiv \sum_{\mathbf{X}_{sm}} G_m(X_m, \mathbf{X}_{sm}), \quad (43)$$

which leads to

$$\mu_{f_s \to X}(X) = \sum_{X_1} ... \sum_{X_M} f_s(X, X_1, ..., X_m, ..., X_M) \prod_{m \in ne(f_s) \setminus X} \mu_{X_m \to f_s}(X_m). \quad (44)$$

The message from a factor node to a variable node is the product of the factor and all incoming messages arriving in the factor node and marginalizing over all of the variables associated with the incoming messages. Therefore the message from a factor node to a variable node cannot be calculated until all incoming messages have arrived. The incoming messages, which are messages from a variable node to a factor node, can again be calculated using the variable-subgraphs. The term

$$G_m(X_m, \mathbf{X}_{sm}) = \prod_{l \in ne(X_m) \setminus f_s} F_l(X_m, \mathbf{X}_{lm}) \quad (45)$$

states that $G_m(X_m, \mathbf{X}_{sm})$ is given by a product of terms which are the products of the variable-subgraphs depicted in Fig. 17.

Substituting Equation 45 into 43 leads to

$$\mu_{X_m \to f_s}(X_m) = \prod_{l \in ne(X_m) \setminus f_s} \sum_{\mathbf{X}_{lm}} F_l(X_m, \mathbf{X}_{lm}) \tag{46}$$

$$= \prod_{l \in ne(X_m) \setminus f_s} \mu_{f_l \to X_m}(X_m), \tag{47}$$

which states that the message from a variable node to a factor node is simply the product of all incoming messages from the factor nodes to the variable node. Again before the message from a variable node to a factor node can be calculated, all the incoming messages have to be computed. Therefore each of the messages can be calculated recursively by applying incoming messages. The messages start at the leafs of the tree-structured factor graphs. If the leafs are variable nodes the corresponding message is initialized by

$$\mu_{X \to f}(X) = 1 \tag{48}$$

and if the leafs are factor nodes the corresponding message is simply the factor itself

$$\mu_{f \to X}(X) = f(X). \tag{49}$$

After initializing the messages starting from the leaf nodes, the formulas in Equation 44 and 47 can be applied to calculate recursively all messages until the hypothesis variable X is reached. The marginal probability is simply the product of all incoming messages. The described algorithm for one hypothesis variable is also called the single-i sum product algorithm, since only one marginal probability is computed.

In the following the single-i sum product algorithm is shown for the example depicted in Fig. 15. As root node the variable node X_4 is chosen. The initial messages starting from the variable nodes are

$$\mu_{X_2 \to f_B}(X_2) = \mu_{X_3 \to f_B}(X_3) = \mu_{X_5 \to f_D}(X_5) = \mu_{X_6 \to f_D}(X_6) = 1 \tag{50}$$

as stated in Equation 48 and the initial messages starting from the factor nodes are

$$\mu_{f_A \to X_1}(X_1) = f_A(X_1) \tag{51}$$

and

$$\mu_{f_C \to X_4}(X_4) = f_C(X_4) \tag{52}$$

as stated in Equation 49. In the following the message

$$\mu_{X_1 \to f_B}(X_1) = \mu_{f_A \to X_1}(X_1) = f_A(X_1) \tag{53}$$

is needed since all incoming messages are required in order to calculate the message from f_B to the root node X_4, which is

$$\mu_{f_B \to X_4}(X_4) = \sum_{X_1} \sum_{X_2} \sum_{X_3} f_B(X_1, X_2, X_3, X_4) \mu_{X_2 \to f_B}(X_2) \mu_{X_3 \to f_B}(X_3) \mu_{X_1 \to f_B}(X_1) \tag{54}$$

$$= \sum_{X_1} \sum_{X_2} \sum_{X_3} f_B(X_1, X_2, X_3, X_4) f_A(X_1). \tag{55}$$

The last message, which is required to calculate the marginal probability of the root node X_4, is the message arriving from f_D

$$\mu_{f_D \to X_4}(X_4) = \sum_{X_5} \sum_{X_6} f_D(X_4, X_5, X_6) \mu_{X_5 \to f_D}(X_5) \mu_{X_6 \to f_D}(X_6) \tag{56}$$

$$= \sum_{X_5} \sum_{X_6} f_D(X_4, X_5, X_6). \tag{57}$$

Since now all incoming messages to the root node X_4 are given, the marginal probability $p(X_4)$ can be calculated.

$$p(X_4) = \mu_{f_B \to X_4}(X_4) \mu_{f_C \to X_4}(X_4) \mu_{f_D \to X_4}(X_4) \tag{58}$$

$$= \sum_{X_1} \sum_{X_2} \sum_{X_3} f_B(X_1, X_2, X_3, X_4) f_A(X_1) f_C(X_4) \sum_{X_5} \sum_{X_6} f_D(X_4, X_5, X_6) \tag{59}$$

If several hypothesis random variables exists and therefore the computation of several marginal probabilities is needed or even the marginal probabilities of all random variables, it is possible to conduct the single-i sum product algorithm for each random variable individually. However, this procedure is not very efficiently. An efficient procedure to compute the marginal probabilities of several random variables leads to the general sum-product algorithm. Firstly, an arbitrary root variable node is selected. Secondly, the single-i sum product algorithm is conducted assuming the selected root variable node as the hypothesis variable. Thirdly, after receiving messages from all neighbors the root node can propagate the messages back to all the leafs. After this step every variable node has incoming messages from all of its neighbors and therefore the marginal can be computed for every random variable in the whole graph.

2.6 Discussion

The presented probabilistic graphical models, Bayesian networks and Markov networks, are able to model different probability distributions. Bayesian models are suited to model causal relations. Therefore many physical processes can be modeled by Bayesian networks. In case of flooding, the causal relation between DEM, flood state and imagery can be modeled by means of a Bayesian network. Furthermore, Bayesian networks are a good tool for the fusion of several observations and are well suited to combine different kinds of data. The combination of DEM and imagery is useful to estimate flooded areas. On the other hand the spatial relation in imagery reflects no causal relation and therefore Bayesian networks are not suited for modeling. As shown in this chapter, Markov networks can model spatial relations in imagery. Therefore,

both models are needed to cope with the requirements stated in the goals of the thesis (see Section 1.2). The transformation of Bayesian networks and Markov networks in factor graphs enables consistent and fast inference algorithms as the described sum-product algorithm.

Beside probabilistic graphical models several other probabilistic frameworks exists, which are alternatives to handle data fusion. The most established models are possibility theory (Zadeh, 1978), which includes the fuzzy theory (Zadeh, 1965), and Dempster-Shafer theory (Dempster, 1967; Shafer, 1976). A deeper discussion about the different alternatives is given in Halpern (2003). However, only the framework of probabilistic graphical models enables the explicit representation of causal relations via Bayesian networks and symmetrical neighborhood relations via Markov networks.

3 State of the art of damage assessment systems during flooding

This chapter gives an overview of image analysis methods, which are used to detect the floodplain delineation during flood events. It is important to distinguish between two kinds of damage assessment systems. The aim of the first kind of systems is the detection of the current state of the flooding, which implies the estimation of the delineation of the floodplain and the water depth (Smith, 1997). Only these systems are suitable for a fast support of the emergency action in the affected areas. The aim of the second kind of systems is the evaluation of the damage after the flood (Bovolo and Bruzzone, 2007; Dhakal et al., 2002). The advantage of the second kind of systems is the possibility of using imagery after the flood event, which is easier to obtain than imagery during the flood. However, the second kind of systems is not suitable for the support of emergency actions.

One goal of the thesis is the development of a system to support the relief actions in the affected area. Therefore, the state of the art focuses on the first kind of systems, which detects the floodplains and water depths during the flood. First approaches using remote sensing data in order to monitor flooding were already studied in the year 1969 (Milfred et al., 1969). An overview on the developed systems and methods is given by Sanyal and Lu (2004) and Smith (1997).

The following subsections are subdivided into image-based systems 3.1 and fusion-based systems 3.2. The image-based systems use solely the information of the imagery. The fusion-based systems use beside imagery additional data such as DEM, water gauge measurements or information from a GIS. The focus of the following subsections is on image analysis methods to detect the floodplain. However, in between of the whole Section 3 further approaches are discussed which are not focused on flooding but are relevant due to the methodology.

3.1 Image-based systems

All methods described in this section are based solely on the information of the imagery. The used image analysis methods differ depending on the used sensor and the number of images available during and before the flood event. Methods, which include imagery after the flood event are neglected since they are not suitable for emergency relief action. The following section is subdivided into four parts. The first part describes image analysis methods for deriving flood parameter from optical imagery. The second part deals with radar data. In the third part multi-sensorial methods are presented, which combine optical and radar imagery. Finally, in the fourth part multi-temporal approaches are shown exploiting the information of several images acquired during and before the flood.

3.1.1 Optical sensors and typical methods

First approaches to map the flood extent with optical sensors were made after the launch of Earth Resource Technology Sensor (ERTS-1) also called Landsat1 carrying the Multispectral Scanner (Hallberg et al., 1973; Morrison and Cooley, 1973; Rango and Salomonson, 1973; Deutsch and Ruggles, 1974). At the same time the ability to monitor floods with the satellites NOAA-2 is demonstrated (Wiesnet et al., 1974). However, the extraction of information from the imagery is based on simple visual interpretation. The interpretation is hampered by the different appearance of flooded areas due to the physical differences in water depth and sediment load (Rango and Anderson, 1974). Chen et al. (1992) investigate the relationship between the spectral reflectance of water and the suspended sediment concentration depicted in

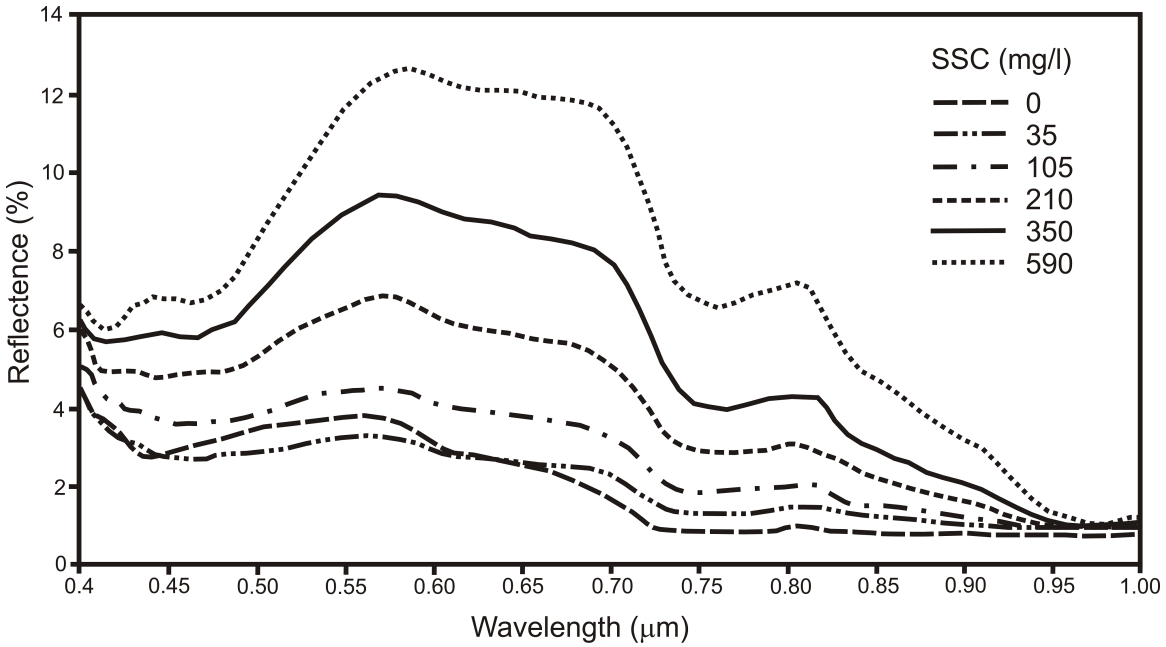

Figure 18. Spectral reflectance R_λ as a function of wavelength λ for six suspended sediment concentrations in water from Chen et al. (1992).

Fig. 18. In case of flooding usually the suspended sediment concentration in water is increased, which leads to a higher reflectance. On the other hand also the soil moisture of not flooded soil increases, which leads to a decrease of spectral reflectance converging to flooded regions with high suspended sediment concentration. Consequently, the reflectance characteristics of ground covers become quite complicated during floods, preventing water and land to be easily distinguished (see Fig. 19) (Sheng et al., 1998). Further factors which cause varying reflectance from water are the change of sun angle and choppy waters. In addition, bottom reflectance can arise if the water is shallow (Wilson, 1997).

In the following methods are discussed extracting the floodplain from optical imagery. In general, the methods can be divided into two different approaches. Firstly, approaches considering only the spectral information of the image for each pixel individually and secondly, methods considering beside the spectral information additionally the spatial information, which takes the relationship of neighboring pixels into account.

Thresholding is one of the first methods, which are applied to optical remote sensing data. The simplest methods use only one spectral band and apply a threshold to detect the flood plain. Usually the near-infrared ($\lambda \approx 1\mu m$), the mid-infrared ($\lambda \approx 2\mu m$) and the far-infrared or thermal bands ($\lambda \approx 11\mu m$) are used due to the low reflectance of water (Work, 1976; Verdin, 1996). In contrast, soil, rock, and vegetation are generally quite reflective (see Fig. 19). Barton and Bathols (1989) state that the far-infrared bands can be used to detect the flooded areas at night under clear condition since the radiative heat loss from the land surface cools it to a temperature well below that of flood. However, the usage of only one band neglects the available information of all the other bands. Several methods are proposed, which combine different bands to enhance the extraction of flooded areas. Wang et al. (2002) propose the summation of the near-infrared and mid-infrared bands and Sheng et al. (1998) calculate the ratio between near-infrared and mid-infrared bands. Since the water body has a high reflectance in the visible band, and a very low reflection in the near-infrared and infrared bands also the Normalized Differenced Vegetation Index (NDVI) is used to detect floodplains (Barton and Bathols, 1989). But usually the NDVI of flooded areas differs from the NDVI of permanent water bodies and hence is closer to the NDVI of soil due to silt and debris in floodwaters (see Fig. 18). The main difficulty of the described thresholding methods is the choice of the correct threshold. In addition, the threshold has to be chosen for each flood scenario individually since

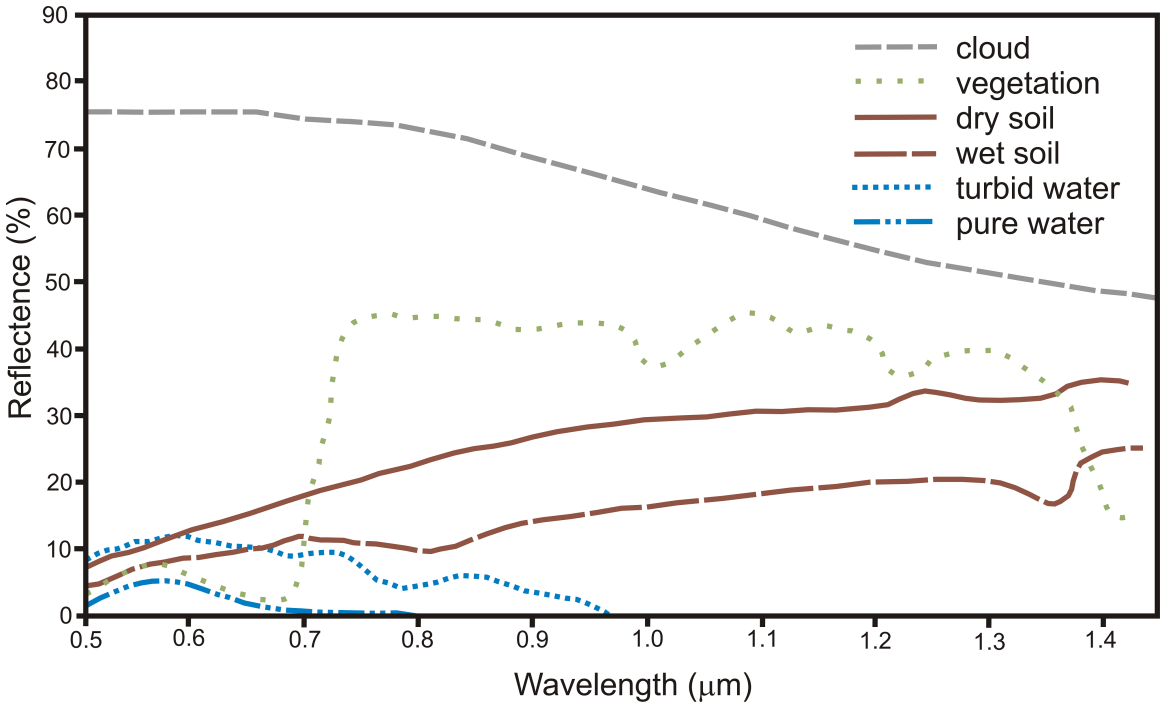

Figure 19. Spectral reflectance R_λ as a function of wavelength λ for ground cover types during floods from Sheng et al. (1998).

the appearance of river flooding varies greatly from place to place.

In contrast to the described thresholding methods the multispectral *maximum likelihood classification* uses usually more than two bands and is based on multivariate probability distributions. Frazier and Page (2000) compares simple thresholding methods with supervised multispectral maximum likelihood classification in case of flood plain delineation. The comparison with reference data leads to a higher overall accuracy using the multispectral maximum likelihood classification than different thresholding methods. In Frey et al. (2009b) a supervised multispectral maximum likelihood classification is used to detect flooded roads. One of the main problems of the supervised multispectral classification is the selection of the training data. The results of the classification are very sensitive concerning the selected training area, especially in cases with large variability within one class, which is often given in flooded areas.

An *unsupervised method* used to detect flooded areas is proposed by Khan et al. (2011) using the Self-Organizing Data Analysis Technique Algorithm (ISODATA), which uses the Euclidean distance to the mean in the feature space in order to assign the pixels to a cluster through a number of iterations. However, the variance of the clusters is neglected unlike the maximum likelihood classification. In addition, the several obtained clusters, which represent the flooded areas, have to be combined manually to one class.

Further approaches to detect flooded areas use *neural networks*. Zhou et al. (2000) proposes a three-layer structured Radial Basis Function Neural Network (RBFNN), in which the output units form a linear combination of the kernel functions computed by the hidden units (Benediktsson et al., 1990; Rollet et al., 1998). Zhou et al. (2000) combines the classification results of the RBFNN with rule based approach. The combination is done via Dempster-Shafer theory (Dempster, 1967; Shafer, 1976).

However, all previously described methods consider only the spectral information of each pixel individually neglecting all kinds of texture and context information. But texture and context information is one of the crucial characteristics identifying objects or regions of interest in imagery (Haralick et al., 1973).

Wilson (1997) uses a *variance filter to describe the texture* of water bodies. The variance filter

provides a measure of local homogeneity. Locally, the water bodies and also flooded areas fulfill this homogeneity. Wilson (1997) conducts a region growing algorithm based on the variance filtered image and further spectral bands. But additional production rules are embedded into the classification schema, which hampers the transferability of the approach.

A hierarchical Markov image model combined with a MRF is proposed by Kersten (2011). The hierarchical Markov model is a directed graphical model defined on a quadtree with the set of all nodes $\mathbf{S}$ and the root node R, which represents the lowest spatial resolution. The joint distribution of the hierarchical Markov image model is given by the factorization

$$P(\mathbf{X}, \mathbf{Y}, R) = p(R) \prod_{s \in \mathbf{S} \setminus R} p(X_s | X_{s^-}) \prod_{s \in \mathbf{S}} p(Y_s | X_s) \tag{60}$$

where $p(R)$ is the root prior, $p(X_s|X_{s^-})$ are the parent child transition probabilities and $p(Y_s|X_s)$ are the data conditional likelihoods with the observations Y_s and the hidden variables X_s, which should be estimated. For inference in the quadtree a similar algorithm as the Viterbi algorithm is applied (Viterbi, 1967; Forney Jr, 1973). The MRF is described by the following energy function

$$U(\mathbf{X}, \mathbf{Y}) = \sum_{1 \leq i \leq I} \sum_{1 \leq j \leq J} U_{data}(Y_{i,j}|X_{i,j}) + U_{context}(X_{i,j}|X_{g,h}, (g,h) \in ne(i,j)) \tag{61}$$

where i, j describes the position of the image and $ne(i, j)$ denotes the neighborhood of the variables. The energy function is divided into a data term U_{data} considering the spectral information and a context term $U_{context}$ considering the neighborhood relations. The ICM algorithm is applied to carry out inference (Kittler and Föglein, 1984; Besag, 1986). Kersten et al. (2010) show that the hierarchical Markov model as well as the MRF clearly outperforms the maximum likelihood classification. The combination of the hierarchal approach with the MRF delivers the best results. However, the proposed system is not fully automatic since user interaction is necessary.

Beside multispectral data also hyperspectral image data with more than 200 bands is used to detect water bodies. One issue working with hyperspectral image data is the reduction of data without losing relevant information for the classification. Especially in time-critical application as the extraction of flood delineation the data reduction plays a decisive part. Hyperspectral satellite missions such as the Environmental Mapping and Analysis Program (EnMAP) can contribute to emergency relief phase during flooding if efficient classification algorithms are available. Braun et al. (2012) suggests support vector machines for the classification of simulated EnMap data and compares support vector machines with independent vector machines and relevance vector machines.

The goal of all the described methods is the delineation of the floodplain or more general the detection of water bodies. A second important parameter in case of flooding is the *water depth*. There are only few approaches which try to extract the water depth solely on the information of optical imagery without topographic information. Lyzenga (1978) proposes a physical model for water depth estimation and discusses the limitations. A segmentation of water depth in sea areas based on a hierarchical Markov model defined on a quadtree is presented by (Provost et al., 2004). The potential to estimate water depth solely from optical imagery is limited to shallow clear water. Therefore in case of flooding a water depth estimation is hardly possible due to the turbidity of water.

The main disadvantage using optical remote sensing data to detect the floodplain delineation are bad weather conditions. Often flood events are accompanied by heavy rains and large cloud coverage. In general, over 50 percent of the Earth's surface is typically covered by clouds at any

time (Paltridge and Platt, 1976). These circumstances is a strong constraint utilizing optical remotely sensed data for emergency relief actions.

3.1.2 Radar sensors and typical methods

The advantage of radar sensors compared to optical sensors is the potential to penetrate clouds. Especially during flooding the probability of heavy rainfall and dense clouds is very high. Therefore the independence of weather conditions and additionally, the independence of daylight are crucial characteristics allowing monitoring floodplains at any time. In contrast to the passive optical sensors, radar sensors are active sensors emitting a series of pulses and recording the return signals. The backscattering is mainly dependent on the surface roughness. In general, the smooth surface of water leads to a nearly specular reflection. However, there are two factors, which complicate the simple specular reflection model. Firstly, the water surfaces could be affected by meteorological condition, such as strong winds or rain cells, which leads to different backscattering patterns with substantial spatial variations. Secondly, the occurrence of emergent vegetation eliminates the specular reflection. In addition, further objects as paved areas or shadowed areas lead to low back-scattering coefficients, which could easily result in misclassification.

Several methods have been developed, which delineate the floodplain from single SAR images. An early attempt was made by Lowry et al. (1981) that used airborne data to map floods by visual interpretation. The simplest image processing approaches deriving the floodplain are as in the optical data *thresholding methods*. Based on the histogram of the processed SAR image a threshold is selected which divides the image into flooded and not flooded areas (Giacomelli et al., 1995; Brivio et al., 2002; Hostache et al., 2009). However, the threshold is often derived by visual interpretation of the histogram and manual subsequent fine tuning. Automatic thresholding procedures can overcome these problems (Sahoo et al., 1988; Bazi et al., 2007). Schumann et al. (2009) uses Otsu's rule in order to find the optimal threshold (Otsu, 1979). Martinis et al. (2009) compares three different automatic thresholding algorithms. Firstly, the KI- algorithm (Kittler and Illingworth, 1986), secondly, an algorithm based on an index function introduced by Miasnikov et al. (2004) and thirdly, the simple global minimum threshold.

The use of *polarimetric SAR* could increase the performance to detect flooded areas (Hess et al., 1994; Henry et al., 2006). Henry et al. (2006) observes that HH polarization provides a more suitable discrimination of flooded areas than HV or VV. However, if multipolarized data is available the combination of different polarizations could improve the detection. Hess et al. (1995) proposes a decision tree classifier based on different polarizations. However, all the described methods consider only the signal return of each pixel individually and neglect the relations of neighboring pixel. All further described methods consider additionally context information.

A common way considering context is the description of *texture* by means of the co-occurrence matrix. Haralick et al. (1973) defines 13 texture measures derived from the co-occurrence matrix. Arzandeh and Wang (2002) suggest using 7 out of the 13 texture measurements (the homogeneity, the contrast, the mean, the variance, the entropy, the angular moment, and the correlation) in order to classify wetlands. The crucial factors affecting the results are beside the used texture measures also the window size the co-occurrence matrix is calculated. Song et al. (2007) investigates different combinations of texture measures at a certain window size in order to detect water areas. He concludes that 3 texture measures (the mean, the contrast and the variance) deliver the best result. Adding additional texture measures do not improve the results for water detection. An increase of the window size leads to a higher overall accuracy of detection but is associated with detection errors at the water land border.

Further segmentation approaches considering context information are methods based on *active*

contour models (Kass et al., 1988). Active contour models are able to combine mathematically the image data with prior modeled object knowledge concerning the shape behavior (Butenuth and Heipke, 2012). The model is formulated as an energy functional consisting of two main components. Firstly, the image energy represents the image forces acting on the spline and secondly, the internal energy modeling the object knowledge concerning the shape. A statistical active contour model especially designed for the segmentation of radar images is presented by Horritt (1999). The method enables the detection of water boundaries and the determination of shorelines in SAR images (Mason and Davenport, 1996; Hahmann and Wessel, 2010). Niedermeier et al. (2000) apply active contour models after edge detection and a wavelet transformation to detect shorelines. In case of flooding the method is used by several authors to detect the boundary between flooded and not flooded areas (Horritt et al., 2001; Mason et al., 2007). A geometric active contour model, also known as level sets, to separate between water and land is proposed by Silveira and Heleno (2009). The advantage of level sets is the ability to automatically change the topology and the independence of parameterization.

An approach which combines the discussed methods, thresholding, texture measures and active contour models is proposed by Schumann et al. (2009). He simply combines the binary results of each method by means of an index which penalizes overprediction of flooding. In addition, a possibility of inundation map based on the Relative Confidence Measure (Romanowicz et al., 1996) is calculated. However, the resulting possibility of inundation map represents not a probability or possibility in any statistical sense, but is only the aggregation of plausible inundation maps.

Another common way to detect flooded areas is to apply *regiongrowing* algorithm (Hess et al., 2003; Schumann et al., 2011). A multiresolution regiongrowing approach is proposed by Baatz and Schaepe (2000). This approach based on the homogeneity of adjacent image objects using different criteria such as distance, texture, spectral similarity and form. The efficiency of this approach detecting flooded areas on TerraSAR-X data is shown by Herrera-Cruz et al. (2010).

Further proposed methods used are *neural networks* (Haykin, 1999). Kussul et al. (2008) applies self-organizing maps to detect the floodplain delineation, which is neural network including a neighborhood function preserving the topological properties of the input space (Kohonen, 2000).

Probabilistic graphical models are used for the extraction of flooding by (Martinis et al., 2011). He combines causal hierarchical graphs with a non causal Markov network. The hierarchical graphical model is similar with the model applied by Kersten et al. (2010) with an equivalent joint probability given in Equation 60. However, the graphical structure is not a quadtree but based on an irregular graphical structure obtained by a preceding segmentation. The energy function corresponding to the Markov network is similar to Equation 61. However, the neighborhood relations are not based on neighboring pixels, but are based on neighboring segmented regions, which leads to an irregular structure.

In addition, SAR has the potential to map water under forested areas (Ormsby et al., 1985). Especially, SAR systems operating with L-band ($\lambda \approx 24cm$) are suited to detect flooded areas due to the capability to penetrate forest canopies (Hess et al., 1990; Pope et al., 1997). But also C-band ($\lambda \approx 6cm$) and P-band ($\lambda \approx 68cm$) SAR data are used to detect flooding beneath the vegetation (Wang et al., 1995). In contrast, Ormsby et al. (1985) found that with X-band ($\lambda \approx 3cm$) no conclusions of flooded forest can be drawn. Beside the wavelength λ also the incidence angel influences the backscattering of flooded vegetation. An investigation of the effect of the incidence angle is done by Lang et al. (2008).

But beside the advantage of the capacity of penetrating clouds there are several disadvantages dealing with radar data to detect floodplain areas. First of all, the side-looking geometry leads to the effects of shadowing, layover and foreshortening. These effects cause problems of assessing flooded areas in urban and mountainous areas. In urban regions substantial areas of ground surface may not be visible due to shadowing and layover causes by buildings or taller vegetation

(Mason et al., 2010). Soergel et al. (2003) found that only one-third of roads were visible in the airborne SAR data of Karlsruhe. In mountainous areas especially the shadowing leads to misclassification. Song et al. (2007) proposes efficient methods for water area classification in a highly relief mountainous environment.

3.1.3 Multi-temporal approaches

The event of a flooding is a highly dynamic process. Therefore, the temporal sampling is crucial to monitor the evolution of a flood event. Several approaches are proposed dealing with multi-temporal images to delineate the floodplain or to describe the development of flooding. In general, two multi-temporal approaches can be differentiated. Firstly, change detection methods, which consider a pre-flood image and an image during the flood, to extract the current floodplain. Secondly, approaches, which dealing with more than two images during the flood to monitor the dynamic process.

Several change detection methods have been developed to obtain information about the changes in between two or more images (Radke et al., 2005). A comprehensive review of change detection methods for optical remote sensing imagery can be found in Lu et al. (2004) and Coppin et al. (2004). The division of different change detection approaches in radar images is discussed in Rignot and van Zyl (1993), Polidori et al. (1995) and Gamba et al. (2006).

A possibility to visualize the extent of flooding could be achieved by color composites generated through the superposition of pre-flood image and an image during the flood (Wang et al., 1995; Badji and Dautrebande, 1997). First color composites were generated to map the extent of flooding by Deutsch and Ruggles (1974).

The simplest change detection approaches are *image differencing and rationing*. The image differencing, which is the difference between the pre-flooded image and the image during the flood (Nico et al., 2000), and the rationing, which is the ratio of the two images (Townsend and Walsh, 1998), are used to detect the floodplain. Badji and Dautrebande (1997) propose a conceptual inundation index based on the image differencing technique.

A *curvelet-based change detection* method to extract changes of flood conditions is proposed by Schmitt et al. (2010). The amplitude radar images are transformed into the curvelet domain, which represent the strength of linear structure apparent in the original image (Candès et al., 2005). The differentiation of the curvelet coefficients in the curvelet domain builds the basis of the change detection algorithm. After the inverse transformation of the differenced and weighted curvelet coefficients back into the space domain the spatial extent of the changes can be seen.

A change detection method for optical imagery which is able to deal with several different bands is proposed by Nielsen et al. (1998). The *Multivariate Alteration Detection* (MAD) method is invariant to linear transformations which imply the insensitivity to linear atmospheric conditions or sensor calibrations at two different times (Nielsen, 2007). The Multivariate Alteration Detection method is based on canonical correlation analysis which was originally introduced by Hotelling (1936). Unlike principal component analysis which identifies patterns of relationship within one set of data, the canonical correlation analysis investigates the intercorrelation between two sets of variables. Frey and Butenuth (2010) apply the Multivariate Alteration Detection method to detect changes during flooding. However, since the results of the Multivariate Alteration Detection method can only be interpreted in a statistical manner there is a need to assign a semantic meaning. Therefore in Butenuth et al. (2011) a supervised classification of the resulting Multivariate Alteration Detection variates is conducted.

All the previously described change detection methods use only two images. Lacava et al. (2010) propose a method, using several pre-flood images to get a more robust description of the common condition without flooding. The *robust AVHRR technique*, originally developed

by Tramutoli (1998), computes the mean value and standard deviation of two features for all pre-flood images. The features are the difference and ratio of two bands in one image. These two reference feature images are subtracted from the feature images computed from the image during the flood and divided by the standard deviation of the reference feature image. The resulting feature differences are the basis for the classification into flooded and non-flooded areas. However, only the flooded area at one point in time can be extracted and no conclusions about the development of the flood can be drawn. In the following, methods are discussed, which investigate the evolution of the dynamic flood process.

Color composites of radar images are used to visualize the dynamic of the flooding in one color image (Long and Trong, 2001). However, the dynamic process of the flood could only be investigated by visual interpretation and is restricted to three distinct points in time. Pulvirenti et al. (2011) proposes an automatic image segmentation method to monitor the flood evolution by means of several radar images. The segmentation technique is based on *morphological profiles* (Benediktsson et al., 2005). For every radar image, which is acquired during the flood, a morphological profile is calculated. The morphological profiles are used as features for a k-means clustering, which results into different classes. The different classes, which consider the temporal evolution in backscattering as well as the texture, build the basis of the flood evolution map.

In general, change detection in radar images can be separated in amplitude change detection and *interferometric coherence change detection*. Usually the amplitude change detection is applied to monitor disasters such as flood due to long time period of pre-flood imagery resulting in disturbing incoherence (Polidori et al., 1995; Schmitt et al., 2010). In addition, the specular reflection over open water yields poor interferometric coherence, and unreliable interferometric phase values (Jung and Alsdorf, 2010). All previously described radar change detection methods could be assigned to amplitude change detection, which analyzes the backscattering coefficient. Nico et al. (2000) proposes a method combining the amplitude and interferometric coherence change detection approaches. However, the integration of the information of coherence is limited to detect open water areas.

Jung and Alsdorf (2010) investigate the coherence in flooded vegetation. Wdowinski et al. (2008) and Hong et al. (2010) use interferometric radar to monitor the water levels in wetlands. Further approaches exist to detect vegetated flooded areas using multi-temporal amplitude information by means of a two-dimensional feature space (Hess et al., 2003) (Martinez and Le Toan, 2007). Hess et al. (2003) use as features simply the backscattering coefficient at different points in time. Martinez and Le Toan (2007) use the mean backscatter coefficient of several radar images as one feature and a temporal change measurement proposed by Quegan et al. (2000).

Beside the discussed approaches with the goal to assess the state of the current flooding, a lot of change detection approaches deal with the damage assessment using imagery before and after the flood e.g. (Yamagata and Akiyama, 1988; Dhakal et al., 2002; Bovolo and Bruzzone, 2007). However, these approaches are not relevant to support relief actions during natural disasters.

3.1.4 Multi-sensorial approaches

Since it is very unlikely that images from different sensors are acquired exactly at the same time, most approaches which use different sensors are inherently multi-temporal. In addition, a flooding is a highly dynamic process, which can change within few hours. Therefore, in case of flooding it is nearly impossible to acquire imagery from different sensors, which map the same state of floodplain delineation.

A common approach is the fusion of radar images showing the current flood situation with pre-flood optical imagery. The radar images are suited for the segmentation of the current

floodplain delineation due to the independence on weather conditions. In contrast, the pre-flood optical images deliver accurate information about the land use of the affected areas. The high classification accuracy of optical imagery compared to radar imagery leads to a more differentiated damage assessment.

In general, image fusion can be divided into three different levels: pixel level, feature level and decision level (Pohl and Van Genderen, 1998). A review of latest research of multi-sensor data fusion is given in Zhang (2010), who updates these three levels of sensor fusion with current developments pointing out the importance of high-level approaches which includes feature and decision level fusion.

A high level or decision level fusion approach is proposed by Kuehn et al. (2002) combining the land use information derived from optical imagery with the floodplain derived from the radar images by means of fuzzy rules (Zadeh, 1965). In contrast, Wang et al. (1995) propose a pixel level fusion by generation of a composite image of the radar image and the blue and infrared band. The classification of the composite image leads directly to flooded and non-flooded land use classes, which are the basis of a damage assessment. Yonghua et al. (2007) propose a further pixel level fusion based on principal component analysis applied to a layer stack generated from current radar images monitoring the flooding and infrared bands from pre-flood optical images. A supervised classification divides the resulting image into flooded areas and areas of natural water expanses. In Frey and Butenuth (2009) a generic system on pixel level is presented, which classifies road objects into flooded and trafficable roads using radar and optical data.

In general, the combination of radar and optical images leads to an improved interpretation capability of remote sensing data. However, the higher the resolution of the imagery the more complicated is the fusion due to the different sensor geometry.

3.2 Fusion-based systems

This section considers approaches which integrate additional information beside the imagery to assess the damage or infer flood parameters, such as water extent and water depth. The additional information can reach from DEMs, hydrological measurements as water gauges or rainfall to GIS data. The only requirement of the additional data is the instant availability during the flooding in order to support the emergency relief actions. The challenge is the appropriate integration or fusion of the additional data with the image information to improve the estimation of the flood parameters. The following section is divided concerning the additional data used in the approaches. In the first Section 3.2.1 approaches are presented embedding a DEM, in Section 3.2.2 hydrological measurements or models are integrated and in Section 3.2.3 further information from GIS such as land cover maps are fused with imagery.

3.2.1 Integration of DEM

In this section, methods to derive flood parameters are discussed using imagery in conjunction with DEMs. As shown in Section 3.1 remote sensing images demonstrate a good performance to derive floodplains. However, optical sensors have severe problems if the flooded areas are occluded by vegetation, clouds and shadows, and radar sensors have to deal with classification uncertainties and the usual effects as shadowing, layover and foreshortening. In case of flooding also a DEM can be used to infer flood parameters. However, simulations based only on DEMs to derive the floodplain lead to correct assignments between only 60% and 80% depending on the resolution and the model (Bates and De Roo, 2000). Since imagery and DEMs deliver complementary information, several methods are based on the fusion of these two information sources.

The simplest models consider each data source individually and combine the results by *logical threshold queries*. Wang et al. (2002) fuse results of the image classification with the results obtained from the DEM by means of the 'OR' operator. The simple rule, that either the classification or the simulation assigns a pixel to flood is basically introduced to overcome the issue of flooded forested areas. In contrast, also the logical 'AND' operator is used to combine the results of classification and simulation (Gianinetto and Villa, 2007). His approach only assigns a pixel to flooded if the classification based on spectral-temporal minimum noise fraction transformation as well as the simulation confirms that the area is affected by the flood. This approach of fusion has the advantage excluding wrongly classified pixels in areas with high altitude. Martinis et al. (2009) divides the combination in an inclusion and exclusion step and incorporates the neighborhood of objects. The inclusion step assigns non-flooded objects to flooded objects, if the considered object is a neighbor of an equal or higher flooded object belonging to the core flood. The exclusion step assigns a flooded object to non-flooded if it is surrounded by non-flooded objects and the altitude is higher than 1 m than the nearest flooded object linked with the flood river.

Beside logical queries also *probabilistic approaches* are used for the fusion. A supervised maximum likelihood classification with two features, the radar backscatter coefficient and the height value from the DEM is proposed by Song et al. (2007). Frey et al. (2009a) propose the multiplication of probabilities derived from imagery via supervised maximum likelihood classification and fuzzy functions describing a simple flood simulation.

Further methods exist based on *least accumulative cost distance matrices* calculated from the DEM (Douglas, 1994). The cost distance matrix describes the energy, which water has to spend to get to a pixel departing from the main stream of the river. Giacomelli et al. (1995) overlay the flooded area obtained from a SAR image with the cost distance matrix. Cost matrix pixels with the largest values corresponding to flooded pixels detected on SAR image are selected as threshold. All pixels with lower cost values are assumed to be flooded Brivio et al. (2002).

Active contour models as mentioned in Section 3.1.2 are also used to combine the information from imagery and DEM information. The active contour models developed to delineate the floodplain in radar images (Horritt, 1999) are extended from 2D to 3D. The algorithm was modified such that the active contour was conditioned not only on the SAR image but also on the DEM, so that the function is also penalized for large height differences (Mason et al., 2007). The internal energy consisting of curvature and tension are adapted to the 3D space rather than the 2D space. The modified active contour model is also investigated in urban areas with TerraSAR-X data (Mason et al., 2010). However, the radar data and the DEM might be acting against each other which lead to the complicating task of finding the suitable weight coefficients.

The incorporation of the DEM enables not only the improvement detecting the floodplain delineation but also allows estimating the flood depth. If no in-situ measurements of water gauges exist, it is possible to derive the water level by overlaying the extracted floodplain with the DEM and pick the height values at the borders of the flooded area. However, it is important to mention that not all borders of the flood extent are trustworthy. For example buildings or vegetation may mask water and cause artificial flood extent limits leading to wrong water level estimates (Hostache et al., 2009).

If the water levels are extracted at the borders of the flood extent, the next step is the estimation of the water depth of the entire flooded area. Several interpolation methods are discussed in the literature. Zwenzner and Voigt (2009) generate cross sections using the centerline of the river. Afterwards the cross sections can be modified concerning the height values at the borders of the floodplain. The region in between two adjacent cross sections can be linearly interpolated. Matgen et al. (2007) compare this cross-section method with a linear regression model, which fits a linear water surface through the point heights extracted at the border of the flooded area.

Further interpolation methods are proposed by Gianinetto and Villa (2007). They compare the linear regression with third-order polynomial regression and composed spline interpolation methods and conclude that the best result is achieved by using the bicubic spline interpolation. Clearly, the level of accuracy that can be expected relies on the spatial resolution and the height uncertainty of the DEM (Matgen et al., 2007).

The described methods show the benefit of using additional DEMs to improve the extraction of the floodplain or to estimate the water depth. However, the availability of a DEM in case of emergency action limits the methods. Currently, only DEMs from the Shuttle Radar Topography Mission (SRTM) are worldwide available. But the spatial resolution of SRTM data is 90m and the relative 90 percent height error is about 10 m (Rodriguez et al., 2006). Therefore the usability of SRTM data for hydrological applications is limited (Bhang and Schwartz, 2008; Karlsson and Arnberg, 2011). However, in future the TanDEM-X mission allows the generation of highly accurate DEMs on a global scale with a relative 90 percent height error between 2 m and 4 m depending on the slope (Krieger et al., 2005).

In addition, the simulation of floodplains using only DEMs is often not a sufficient representation. The modeling of the dynamic process of flood inundation also requires detailed information on hydrological parameters such as backwater curves, flow impedances and roughness coefficients (Overton, 2005).

3.2.2 Integration of hydrological models

The fusion of remote sensing data with hydrological simulations has the potential to improve substantially the understanding of floodplain flow processes. An overview of the contribution of remote sensing data for flood hydrology is given in Bates et al. (1997).

Remote sensing data can deliver several parameters usually used in hydrological flood simulations. Remote sensing data acquired during a flood combined with a DEM offers the opportunity to collect spatially distributed water gauge estimations over large areas without the need for costly hydrological ground surveys (Schultz, 1988). The water gauge estimations are used is input information for the calibration of hydrological flood models (Montanari et al., 2009; Hostache et al., 2009). The calibrated flood models can be used either for the current flood event or for future events. But also the total extracted floodplain from remote sensing data is used for the model calibration (Khan et al., 2011). Profeti and Macintosh (1997) estimate the soil moisture from optical data, which is an input parameter for flood simulation. Another important parameter for flood simulation is the roughness factor, which can be estimated by means of optical pre-flood imagery (Van der Sande et al., 2003).

On the other hand the improved flood simulations again can contribute to the detection of flooded areas, which are not observable by remote sensing due to occlusions such as vegetation, clouds or shadows. However, a big challenge is the integration of all the available knowledge, consisting of remote sensing data, in-situ measurements and hydrological flood simulations in a comprehensive system to support emergency actions.

Brakenridge et al. (1998) show that remote sensing data combined with a DEM have the opportunity to observe flood waves. The comparison of the extracted floodplain from radar data with a simulation based on steady discharge as input parameter reveals the flood wave, which can be used to measure the peak discharges. Nagarajan et al. (2010) estimate the streamflow combining area-wide spatial data derived from remotes sensing with in-situ point measurements via Bayesian networks. The Bayesian network reflects a hydrological process incorporating the DEM, the land cover, soil moisture, measured groundwater levels and rainfall data.

3.2.3 Integration of GIS data

In general, extracted information from remote sensing data can be utilized in conjunction with GIS (Blaschke, 2010). In case of flood management the extracted floodplain from imagery integrated with other data in GIS could improve the decision making in emergency actions. The analysis of the floodplain combined with social economic data from a GIS such as land use, population data, cadastral data, transportation and infrastructure networks leads to an improved performance of the relief management (van Westen and Soeters, 2000). The analysis of the combined data can reveal high risk areas that may be subject to damage. For example, information about the inundated areas of different land use classes, the affected population or the damage to infrastructure objects such as roads, bridges or pipelines can be obtained. Several approaches assess the damage by combining the flood delineation with land use maps (Profeti and Macintosh, 1997; Dewan et al., 2005; Forte et al., 2006). But also current information concerning the state of transportation systems is useful for optimizing the transport of materials for disaster relief (Zhang et al., 2002). Islam and Sado (2000) evaluates the risk if major roads are flooded during a disaster. The integration of all the different kind of thematic spatial information is useful for decision makers in relief actions (Tholey et al., 1997).

All the described methods are based on high level fusion, which combines extracted information from imagery with data from a GIS. However, the information of the GIS data is not used in the process of image segmentation. But the use of knowledge of existing geodata can be used to ease and speed-up the process of extraction (Baltsavias, 2004). The described method in the thesis uses the road information from a road database to improve the image interpretation. Due to the rapid progress of navigation systems for cars the availability of road data is no longer a limiting factor. The assessment of roads without initial road information via automatic road extraction (Wiedemann and Hinz, 1999; Hinz and Baumgartner, 2003) is critical, since only trafficable roads could be extracted and no statement about flooded roads is possible.

3.3 Discussion

One goal of the thesis is the development of a system for the rapid assessment of infrastructure objects during flooding. A basic property of the system must be the handling of different sensors, since the availability of certain sensor systems cannot be guaranteed in a short time frame. And time is the crucial parameter during natural disasters to supply rapid emergency response. Furthermore, the system has to deal with multi-temporal imagery in order to update the analysis. However, if the current imagery lacks quality, the information of previous imagery with higher quality should be used to refine the damage assessment. In addition, the integration of a DEM into the system can significantly improve the estimation of floodplains; especially due to occlusion in the imagery (see Section 3.2.1). However, in case of disaster management DEMs are required which are immediately and worldwide available but usually DEMs with a resolution of 1m as used in several approaches discussed in previous sections are not available. The fusion of the different data sources is one of the main challenges of the system. Probabilistic approaches are able to deal with the different qualities of the sensors. Bayesian networks deliver a probabilistic framework to combine different sensors of different qualities by means of modeling the causal relations between the observations and the hypothesis variables.

In the previous sections many approaches are discussed to detect water bodies from imagery. Statistical models are developed considering the uncertainties in image information. Probabilistic graphical models are one statistical framework, which is used and applied to optical as well as radar sensors. However, the discussed models are restricted to the image information and are not developed to incorporate further information sources. Other approaches focus on the fusion of imagery and other data as DEM or GIS data. However, in case of the detection of flooded areas there is a lack of statistical methods considering the inherent uncertainties of

the different data sources or further measurements. A strategy which combines the statistical image analysis models with additional inherently uncertain data sources for flood delineation is still not present.

All approaches described in this section can be divided into methods, which uses only spectral information and methods using spectral and spatial information. In general, the consideration of both spectral and spatial information leads to a better performance. Markov networks deliver a probabilistic framework to model neighborhood relations in imagery. In addition, the geometry of available objects from a GIS can be used to determine the structure of the Markov network. As described in Section 2.3 both Bayesian and Markov network can be transformed in factor graph, which are the basis for inference. Therefore all available information from different data sources is unified in one probabilistic model, which differs from many approaches conducting successive processing.

4 Probabilistic graphical models for damage assessment during flooding

The following section presents the developed probabilistic graphical models to assess the damage state of linear infrastructure objects during flooding. The strength of the presented models is the integration of all available data in one statistical framework. The statistical framework of probabilistic graphical models enables to consider the varying accuracies of the data (e.g. DEM, water gauge measurements) and quantifies the uncertainties of the results. In general, the true state of a random variable is often uncertain, because the observations are incomplete and noisy. Image data acquired during flooding delivers only partial information due to occlusions or misclassification. Furthermore, each observation contains errors resulting in different accuracies. In case of flood simulations the accuracy of the used DEM is crucial.

In the following subsections three probabilistic graphical models are presented. The models are expanded step by step since they build upon each other. The first presented model is the pixel-based model, which describes the connection between imagery and DEM in case of flooding as a Bayesian network (Frey et al., 2012). The random variables and the conditional probability distributions of the pixel-based model are described in Section 4.1. The second model is the topology-based model, which is based on the pixel-based model. However, additional links are introduced reflecting the topology of the road network (see Section 4.2). In contrast to the pixel-based model, the neighborhood relations are modeled in the topology-based model. The last presented model is the multi-temporal model which integrates the temporal development of a flood. Images from different sensors at several points in time are integrated in one graphical model described in Section 4.3. All models are represented as factor graphs (see Section 2.5), which enable the application of the sum-product rule as described in Section 2.5.2.

4.1 Pixel-based model

The pixel-based model describes the connection between the different input data such as DEM, water gauge measurements and the image data acquired during flooding. The model is designed as a Bayesian network, which represents the causal relation between DEM, damage state and the intensity values of the imagery. The pixel-based model is not restricted to linear infrastructure objects, but could be applied to every pixel in the imagery. However, the spatial relations of neighboring pixels in the imagery are neglected.

4.1.1 Representation as Bayesian Network

The pixel-based model is described by a causal Bayesian network depicted in Fig. 20. The causal Bayesian network of the pixel-based model consists of six random variables the height H, the water gauge G, the water level above ground W, the damage state D, the class C and the intensity values of the imagery $\mathbf{I}$. In addition, several parameters $\mu_H, \sigma_H, \mu_G, \sigma_G, \boldsymbol{\mu}_{c,k}, \boldsymbol{\Sigma}_{c,k}, \omega_{c,k}$ are introduced, which are defined in this section step by step.

The parameters $\mu_H, \sigma_H, \mu_G, \sigma_G$ representing the standard deviation and the mean of H and G. The parameter μ_H is the height value obtained from the DEM. The vertical accuracy of the DEM is expressed by the standard deviation σ_H. Some DEMs deliver accuracy information for each pixel individually and therefore the standard deviation σ_H can vary from pixel to pixel. If no accuracy information for each pixel is available, σ_H is defined as a global parameter for the whole scene. The parameter μ_G represents the measurement of the water gauge, which can be obtained directly from the acquired imagery during flood as described in Section 3.2.1. The

accuracy of the water gauge measurements is represented by σ_G. In the following paragraph each random variable is discussed in more detail.

The random variable H is a continuous variable representing the height value of a pixel related to a specified reference. It is defined by the parameters σ_H and μ_H as a Gaussian distribution

$$p(H|\mu_H, \sigma_H) = \mathcal{N}(h|\mu_H, \sigma_H^2). \tag{62}$$

The random variable water gauge G is also continuous and again defined as a Gaussian distribution by the parameters μ_G and σ_G

$$p(G|\mu_G, \sigma_G) = \mathcal{N}(g|\mu_G, \sigma_G^2). \tag{63}$$

The random variable water level above ground denoted by W is as well a continuous variable and defined by the deterministic function

$$W = G - H. \tag{64}$$

In contrast to H, G and W the random variable D is discrete and reflects the damage state of the road consisting of two states: flooded f and not flooded $\overline{f}$.

$$Val(D) = \{f, \overline{f}\} \tag{65}$$

It has to be mentioned that flooded roads are sometimes still trafficable. Therefore the trafficability of roads can be obtained combining the information of D, W and the properties of the used vehicles in emergency actions. However, this connection is not discussed in the thesis.

A further random variable is the class C which describes all possible classes which could appear in the acquired imagery from space during flooding at the region of the road network. Several different classes are reasonable. The most obvious classes are the class water w and the class road r. Furthermore, occlusions of the road network can exist. The most occurring occlusions are vegetation v and clouds c. In case of radar imagery the class cloud can be neglected apart from extreme cases for X-band. In addition, the class rest e is introduced, which captures all other occlusions beside vegetation and cloud. Therefore the states of the random variable C for optical imagery is

$$Val(C) = \{r, w, v, c, e\} \tag{66}$$

and for radar imagery

$$Val(C) = \{r, w, v, e\}. \tag{67}$$

However, in urban areas and high alpine scenes an additional class "radar shadow" is needed to avoid confusion of dark shadowed pixels and dark flooded pixels. The last random variable $\mathbf{I}$ is the vector of n random variables, describing the intensity values of the imagery

$$\mathbf{I} = \begin{pmatrix} I_1 \\ I_2 \\ \vdots \\ I_n \end{pmatrix}, \tag{68}$$

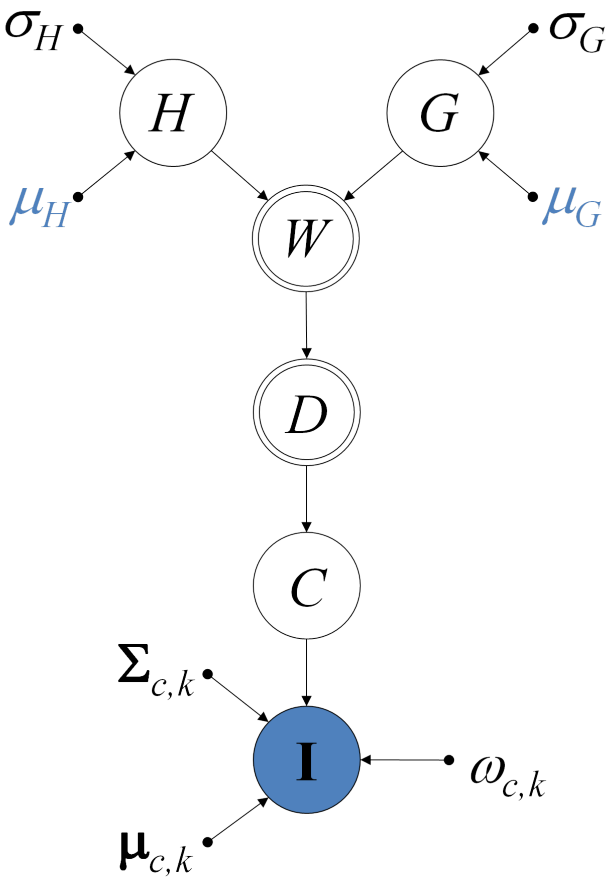

Figure 20. The pixel-based model designed as a Bayesian network combining DEM, water gauge measurement and image data.

where n is the number of bands. In case of high dimensional data such as hyperspectral data a preceding band selection or the reduction of bands for example by means of a principle component analysis would be reasonable. The number of states of I_n depends on the bit depth b of the imagery, which is usually between 8 and 24.

$$Val(I_n) = \{1, 2, ..., 2^b - 1, 2^b\} \tag{69}$$

Therefore the size of the state or feature space of $\mathbf{I}$ is $n \times 2^b$. The parameters $\boldsymbol{\mu}_{c,k}, \boldsymbol{\Sigma}_{c,k}, \omega_{c,k}$ are required to model the likelihood function $p(\mathbf{I}|C)$ discussed in the next section in more detail.

The presented pixel-based model is a hybrid Bayesian network as described in Section 2.1.1, since the random variables H, G and W are continuous and the random variables D, C and $\mathbf{I}$ are discrete variables. The random variable $\mathbf{I}$ is an information variable and therefore highlighted in blue (see Fig. 20). In addition, the observed parameters μ_H and μ_G are highlighted in blue. The hypothesis variables are the damage state D and the water level W.

The links between the described random variables as shown in Fig. 20 reflect the causal relation between them. The links between H, G and W are the deterministic connections. If H and G are given, the water level W can directly calculated deterministically by Equation 64. The water level W is responsible for whether a road is flooded or not. Furthermore, the physical state of the earth surface such as water, vegetation have specific reflection characteristics which cause different intensity measures $\mathbf{I}$ at the sensors. Due to these physical causal relations the links start from H and G in direction to $\mathbf{I}$.

Using the factorization described in Equation 6 the joint probability of the Bayesian network depicted in Fig. 20 is

$$\begin{aligned} &p(H, G, W, D, C, \mathbf{I}|\mu_H, \sigma_H, \mu_G, \sigma_G, \boldsymbol{\mu}_{c,k}, \boldsymbol{\Sigma}_{\mathbf{c},\mathbf{k}}, \omega_{c,k}) = \\ &p(H|\mu_H, \sigma_H)p(G|\mu_G, \sigma_G)p(W|H, G)p(D|W)p(C|D)p(\mathbf{I}|C, \boldsymbol{\mu}_{c,k}, \boldsymbol{\Sigma}_{\mathbf{c},\mathbf{k}}, \omega_{c,k}). \end{aligned} \tag{70}$$

For ease of notation all the parameters are neglected in further equations. This leads to the more concise expression of the joint distribution

$$p(H, G, W, D, C, \mathbf{I}) = p(H)p(G)p(W|H, G)p(D|W)p(C|D)p(\mathbf{I}|C). \tag{71}$$

The goal of the probabilistic graphical model is to calculate the marginal probabilities of the hypothesis variables D and W. Using the product rule, the marginal probabilities of D and W can be expressed as

$$p(W|\mathbf{I}) = \frac{p(W, \mathbf{I})}{p(\mathbf{I})} \tag{72}$$

and

$$p(D|\mathbf{I}) = \frac{p(D, \mathbf{I})}{p(\mathbf{I})}. \tag{73}$$

The probability distributions $p(W, \mathbf{I})$, $p(D, \mathbf{I})$ and $p(\mathbf{I})$ can be expressed by means of marginalization of the joint distribution of all random variables (see Equation 70) using the sum rule.

$$p(W, \mathbf{I}) = \int_H \int_G \sum_D \sum_C p(H, G, W, D, C, \mathbf{I})\ dG\ dH \tag{74}$$

$$p(D, \mathbf{I}) = \int_H \int_G \int_W \sum_C p(H, G, W, D, C, \mathbf{I})\ dW\ dG\ dH \tag{75}$$

$$p(\mathbf{I}) = \int_H \int_G \int_W \sum_D \sum_C p(H, G, W, D, C, \mathbf{I})\ dW\ dG\ dH \tag{76}$$

The substitution of $p(H, G, W, D, C, \mathbf{I})$ in Equation 74, 75 and 76 with the decomposition of the joint probability in Equation 71 and by applying the distributive law leads to

$$p(W, \mathbf{I}) = \int_H p(H) \int_G p(G)p(W|H, G) \sum_D p(D|W) \sum_C p(C|D)p(\mathbf{I}|C)\ dG\ dH, \tag{77}$$

$$p(D, \mathbf{I}) = \int_H p(H) \int_G p(G) \int_W p(W|H, G)p(D|W) \sum_C p(C|D)p(\mathbf{I}|C)\ dW\ dG\ dH \tag{78}$$

and

$$p(\mathbf{I}) = \int_H p(H) \int_G p(G) \int_W p(W|H, G) \sum_D p(D|W) \sum_C p(C|D)p(\mathbf{I}|C)\ dW\ dG\ dH. \tag{79}$$

The expression in Equation 79 is the normalization constant. Inserting Equation 77 and 79 in Equation 72 the final expression for the marginal probability of W given $\mathbf{I}$ is

$$p(W|\mathbf{I}) = \frac{\int_H p(H) \int_G p(G) p(W|H,G) \sum_D p(D|W) \sum_C p(C|D) p(\mathbf{I}|C) \; dG \; dH}{\int_H p(H) \int_G p(G) \int_W p(W|H,G) \sum_D p(D|W) \sum_C p(C|D) p(\mathbf{I}|C) \; dW \; dG \; dH}. \tag{80}$$

Equivalently, inserting the expressions 78 and 79 in Equation 73 the final expression for the marginal probability of D is

$$p(D|\mathbf{I}) = \frac{\int_H p(H) \int_G p(G) \int_W p(W|H,G) p(D|W) \sum_C p(C|D) p(\mathbf{I}|C) \; dW \; dG \; dH}{\int_H p(H) \int_G p(G) \int_W p(W|H,G) \sum_D p(D|W) \sum_C p(C|D) p(\mathbf{I}|C) \; dW \; dG \; dH}. \tag{81}$$

Equations 80 and 81 are straightforward formulas derived by the simple variable elimination algorithm (Koller and Friedman, 2009; Jensen and Nielsen, 2007). However, most of the computations have to be done two times in order to get both marginal distributions. Using the sum-product algorithm multiple computations can be avoided.

In the following sections the conditional probabilities are defined, which are required to compute the marginal distributions and the conditional independence structure of the graphical model is discussed.

4.1.2 Conditional independence properties

In a Bayesian network as described in Section 2.1.2 the conditional independence properties can directly derived from the graphical structure. The pixel-based model consists of serial and converging connections. Focusing on the serial connections depicted in Fig. 20 the following conditional independence properties can be derived. The damage state D and the water level W are conditional independent from the intensity values $\mathbf{I}$ given the class C.

$$\mathbf{I} \perp\!\!\!\perp D, W \mid C \tag{82}$$

Furthermore, if the damage state D is given, the water level W is conditionally independent of C and $\mathbf{I}$.

$$W \perp\!\!\!\perp C, \mathbf{I} \mid D \tag{83}$$

However, in the common configuration only the intensity values of the imagery $\mathbf{I}$ are given and neither D nor C is observed. Therefore, for the configuration depicted in Fig. 20 it holds

$$W \not\perp\!\!\!\perp \mathbf{I}, \tag{84}$$

which is essential, otherwise the information from the imagery would not be utilized. In general, the height H and the water gauge G are conditional independent random variables, if no observations are made.

$$H \perp\!\!\!\perp G \mid \emptyset \tag{85}$$

However, if any of the random variables D, C or $\mathbf{I}$ are given, the height H and the water gauge G are not conditional independent. This leads to the common configuration depicted in Fig. 20 to

$$H \not\perp\!\!\!\perp G \mid \mathbf{I}. \tag{86}$$

4.1.3 Conditional probability distributions

In this section all the conditional probabilities describing the links in the Bayesian network and showing up in the joint distribution of Equation 70 are discussed. The prior probabilities $p(H|\mu_H, \sigma_H)$ and $p(G|\mu_G, \sigma_G)$ are already defined in Equation 62 and 63. The connection between the height H, the water gauge G and the water level W is deterministic and defined in Equation 64. Since $p(H|\mu_H, \sigma_H)$ and $p(G|\mu_G, \sigma_G)$ are Gaussian distributions and the deterministic connection is linear, the resulting conditional probability is also a Gaussian distribution

$$p(W|H, G) = \mathcal{N}(w|\mu_W = \mu_G - \mu_H, \sigma_W^2 = \sigma_G^2 + \sigma_H^2) \tag{87}$$

with the mean $\mu_W = \mu_G - \mu_H$ and the variance $\sigma_W^2 = \sigma_G^2 + \sigma_H^2$. The connection between the water level W and the damage state D is also deterministic and defined by means of the Heaviside step function (Abramowitz and Stegun, 1972), which simply expresses the fact that roads with positive water level are flooded and roads with negative water level are not flooded. Therefore $p(D = f|W)$ is defined as

$$p(D = f|W) = \Theta(w) = \begin{cases} 1, & if \quad w > 0 \\ 0.5, & if \quad w = 0 \\ 0, & if \quad w < 0, \end{cases} \tag{88}$$

and

$$p(D = \overline{f}|W) = 1 - \Theta(w). \tag{89}$$

The conditional probability $p(C|D)$ is defined by a conditional probability table (see Tab. 1). The elements of the conditional probability table are dependent on prior probabilities of the class vegetation pr_v, cloud pr_c and rest pr_e. The prior probability pr_c can be obtained by an estimation of the cloud coverage of the scenario using the imagery. The prior probability of pr_v can be obtained from a GIS or also from the imagery itself. The most difficult part is the determination of the prior probability pr_e, which depends on the complexity of the scenario.

The most obvious entries in Tab. 1 are the probabilities $p(C = w|D = \overline{f})$ and $p(C = r|D = f)$, which is 0. Independent of the damage state, the probabilities $p(C = c|D = f)$ and $p(C = c|D = \overline{f})$ that clouds occlude the flooded or not flooded roads is the prior probability pr_c. The probabilities $p(C = v|D = f)$ and $p(C = v|D = \overline{f})$ are the multiplication of the probability that no clouds exists and the prior probability pr_v, which reflects the fact that clouds occlude the vegetation. In the same way the probabilities $p(C = e|D = f)$ and $p(C = e|D = \overline{f})$ are modeled, which are also independent of the damage state. Finally, the probability $p(C = r|D = \overline{f})$ describes the existence of the road, if there is no flooding and not occluded by clouds, vegetation or something else. Equivalently, the probability $p(C = w|D = f)$ is set in case of flooded roads.

Table 1. Conditional Probability Table $p(C|D)$

	$D = notflooded$	$D = flooded$
road	$(1 - pr_e)(1 - pr_v)(1 - pr_c)$	0
water	0	$(1 - pr_e)(1 - pr_v)(1 - pr_c)$
cloud	pr_c	pr_c
vegetation	$pr_v(1 - pr_c)$	$pr_v(1 - pr_c)$
rest	$pr_e(1 - pr_c)(1 - pr_v)$	$pr_e(1 - pr_c)(1 - pr_v)$

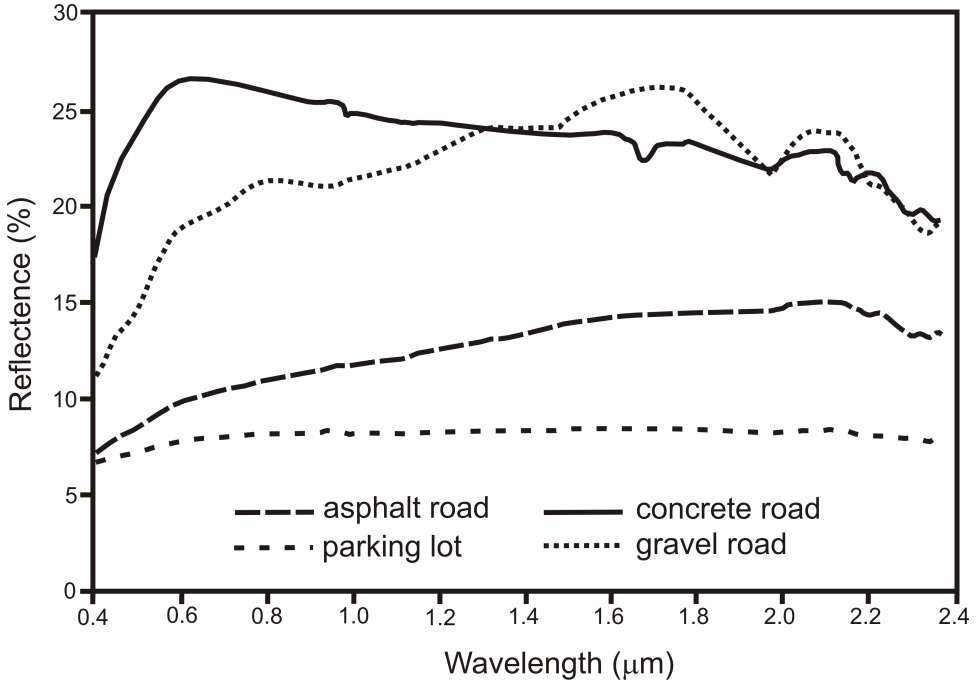

Figure 21. Spectral reflectance R_λ as a function of wavelength λ for road types from Herold et al. (2004).

Finally, the conditional probability $p(\mathbf{I}|C)$ is modeled as a Gaussian mixture model (McLachlan and Peel, 2000) since the radiometric characteristics of the classes road, vegetation and water during flooding have a high variation. The radiometric variability of water during flooding is already shown in Fig. 18. But also the roads show a large radiometric variability. On one hand, different road classes as asphalt, concrete or gravel are responsible for different reflectance curves (see Fig. 21) and, on the other hand, the age of roads leads to varying reflectance characteristics. The older the road, the higher is the reflectance (see Fig. 22). Due to the radiometric variability of the classes, single Gaussian probability functions are not able to describe the complexity sufficiently. Therefore, Gaussian mixture models are used, which combine single functions to a more complex probability function. The resulting probability function is simply a weighted sum of the initial functions. Therefore the conditional probabilities for the classes road, vegetation and water are

$$p(\mathbf{I}|C=r) = \sum_{k=1}^{K_r} \omega_{r,k}\, \mathcal{N}(\mathbf{i}|\boldsymbol{\mu}_{r,k}, \boldsymbol{\Sigma}_{r,k}), \tag{90}$$

$$p(\mathbf{I}|C=w) = \sum_{k=1}^{K_w} \omega_{w,k}\, \mathcal{N}(\mathbf{i}|\boldsymbol{\mu}_{w,k}, \boldsymbol{\Sigma}_{w,k}), \tag{91}$$

and

$$p(\mathbf{I}|C=v) = \sum_{k=1}^{K_v} \omega_{v,k}\, \mathcal{N}(\mathbf{i}|\boldsymbol{\mu}_{v,k}, \boldsymbol{\Sigma}_{v,k}), \tag{92}$$

where K_r, K_w and K_v is the number of centers, which could reflect possible subclasses, $\omega_{r,k}$, $\omega_{w,k}$ and $\omega_{v,k}$ are the weights of the possible subclasses, $\boldsymbol{\mu}_{r,k}$, $\boldsymbol{\mu}_{w,k}$ and $\boldsymbol{\mu}_{v,k}$ are the n-dimensional means and $\boldsymbol{\Sigma}_{r,k}$, $\boldsymbol{\Sigma}_{w,k}$ and $\boldsymbol{\Sigma}_{v,k}$ are the covariance matrices of the possible subclasses. The size of the covariance matrices are $n \times n$ where n is the number of bands. In Fig. 21 possible subclasses for the class road are depicted. The parameters $\boldsymbol{\mu}_{r,k}$, $\boldsymbol{\mu}_{w,k}$, $\boldsymbol{\mu}_{v,k}$, $\boldsymbol{\Sigma}_{r,k}$, $\boldsymbol{\Sigma}_{w,k}$ and $\boldsymbol{\Sigma}_{v,k}$ are obtained by means of training data for every subclass. The weights $\omega_{r,k}$, $\omega_{w,k}$ and $\omega_{v,k}$, also called mixing coefficient, have to fulfill the condition

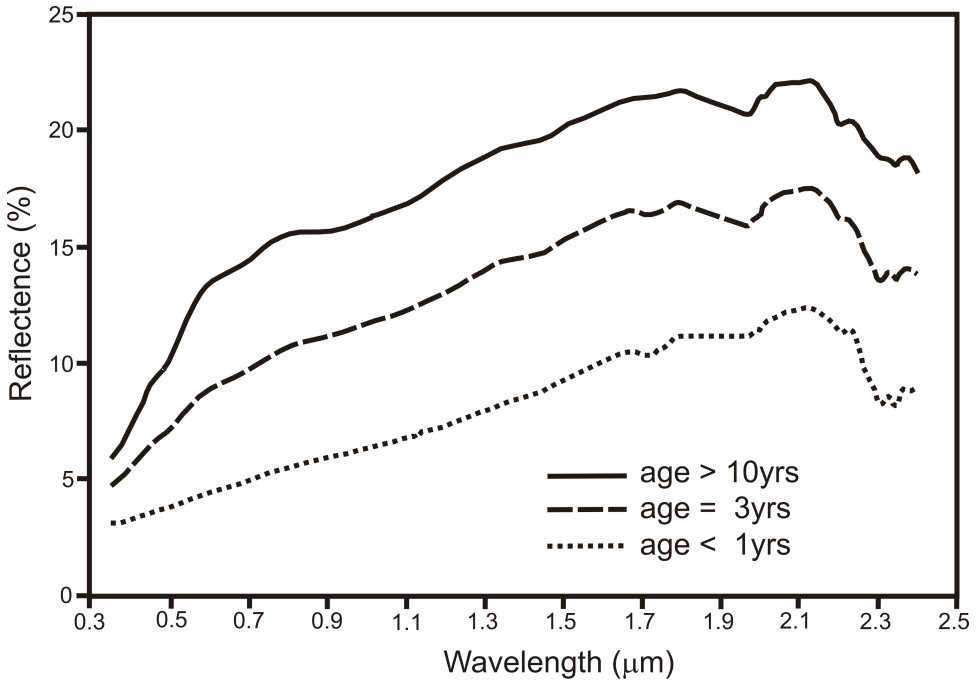

Figure 22. Spectral reflectance R_λ as a function of wavelength λ for different ages of roads from Herold and Roberts (2005).

$$\sum_{k=1}^{K} \omega_k = 1 \tag{93}$$

and

$$0 \leq \omega_k \leq 1. \tag{94}$$

Instead of generating training data for each subclass, the expectation maximization algorithm can be used to determine the weights $\omega_{r,k}$, $\omega_{w,k}$ and $\omega_{v,k}$ and the parameters $\boldsymbol{\mu}_{r,k}$, $\boldsymbol{\mu}_{w,k}$, $\boldsymbol{\mu}_{v,k}$, $\boldsymbol{\Sigma}_{r,k}$, $\boldsymbol{\Sigma}_{w,k}$ and $\boldsymbol{\Sigma}_{v,k}$ (Butenuth et al., 2011). The number of centers K_r, K_w and K_v can be calculated using the minimum message length criterion (Figueiredo and Jain, 2002).

Since there is almost no variation in the radiometric characteristics of clouds, the conditional probability distribution $p(\mathbf{I}|C = c)$ is not modeled as a mixture, but a simple n-dimensional Gaussian distribution.

$$p(\mathbf{I}|C = c) = \mathcal{N}(\mathbf{i}|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c). \tag{95}$$

The conditional probability distribution $p(\mathbf{I}|C = e)$ for the class rest is modeled as a uniform distribution, since no reasonable information about the parameters can be obtained to cover the wide spectrum of the class e. Therefore

$$p(\mathbf{I}|C = e) = \frac{1}{n * 2^b}, \tag{96}$$

where n is the number of bands and b is the color depth of the imagery.

4.1.4 Representation as factor graph

In Fig. 20, the pixel-based model is represented as a Bayesian network. The calculation of the marginal probabilities of the hypothesis variables D and W are given in Equation 80

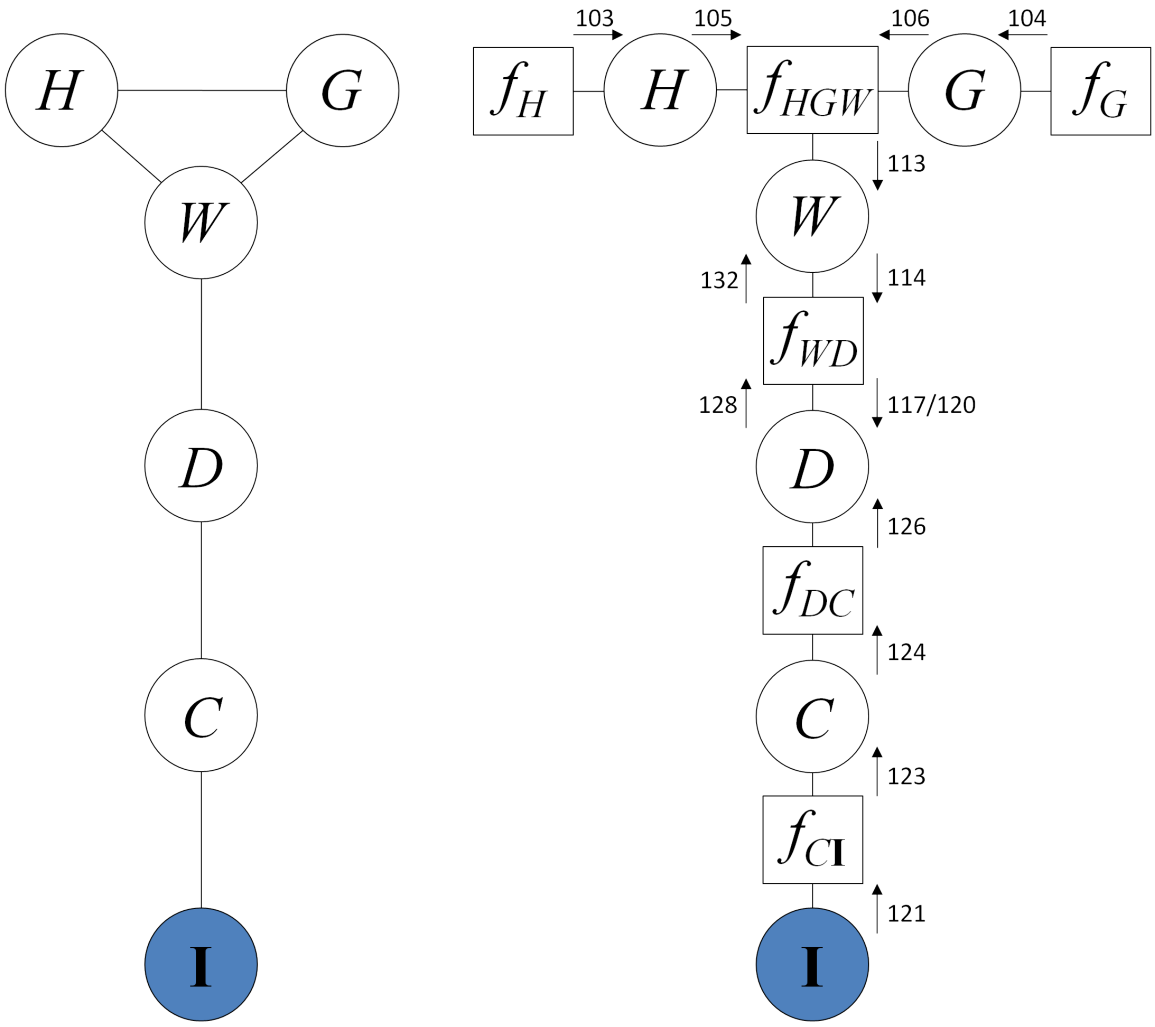

Figure 23. The pixel-based model transformed into a Markov network (left) or factor graph (right). The numbered arrows correspond to the equations in Section 4.1.5

and 81. However, the calculation of these equations is not efficient since it requires multiple computations of the same components. The utilization of factor graphs with the sum-product algorithm for inference enables a more efficient way to calculate the marginal probabilities (see Section 2.5.1 and 2.5.2). In addition, the pixel-based model is part of the topology-based model, which is modeled as a Markov network. Since Bayesian networks and Markov networks can be transformed into factor graphs, the sum-product algorithm applied to factor graphs enables a consistent framework to conduct inference in all presented models.

The transformation of the Bayesian network of the pixel-based model shown in Fig. 20 into a factor graph is depicted in Fig. 23. Firstly, the Bayesian network is transformed into a Markov network by applying moralization described in Section 2.3. The parameters which are depicted explicitly in the Bayesian network are neglected in all further graphical models to maintain clarity. The second step is the generation of the factor graph from the Markov network considering the clique potentials (see Section 2.5).

The joint distribution of the Markov network shown in Fig. 23 is

$$p(H, G, W, D, C, \mathbf{I}) = \frac{1}{Z}\psi_{HGW}(H, G, W)\psi_{WD}(W, D)\psi_{DC}(D, C)\psi_{C\mathbf{I}}(C, \mathbf{I}) \tag{97}$$

with the partition function

$$Z = \int_H \int_G \int_W \sum_D \sum_C \sum_{\mathbf{I}} \psi_{HGW}(H,G,W)\psi_{WD}(W,D)\psi_{DC}(D,C)\psi_{C\mathbf{I}}(C,\mathbf{I})\ dW\ dG\ dH. \quad (98)$$

Using the sum rule and grouping the potential functions, the marginal probability of the hypothesis variable W is

$$p(W) = \frac{1}{Z} \int_H \int_G \psi_{HGW}(H,G,W) \sum_D \psi_{WD}(W,D) \sum_C \psi_{DC}(D,C) \sum_{\mathbf{I}} \psi_{C\mathbf{I}}(C,\mathbf{I})\ dG\ dH \quad (99)$$

and the marginal probability of the hypothesis variables D is

$$p(D) = \frac{1}{Z} \int_H \int_G \int_W \psi_{HGW}(H,G,W)\psi_{WD}(W,D) \sum_C \psi_{DC}(D,C) \sum_{\mathbf{I}} \psi_{C\mathbf{I}}(C,\mathbf{I})\ dW\ dG\ dH. \quad (100)$$

Since the Markov network is derived from a Bayesian network, the potential functions can be substituted by the conditional probabilities defined in Section 4.1.3 (Frey and Butenuth, 2011).

$$\begin{aligned} \psi_{HGW}(H,G,W) &= p(H)p(G)p(W|H,G) \\ \psi_{WD}(W,D) &= p(D|W) \\ \psi_{DC}(D,C) &= p(C|D) \\ \psi_{C\mathbf{I}}(C,\mathbf{I}) &= p(\mathbf{I}|C) \end{aligned} \quad (101)$$

Similarly, the factors described in Fig. 23 can be interpreted as conditional or prior probabilities.

$$\begin{aligned} f_G(G) &= p(G) \\ f_H(H) &= p(H) \\ f_{HGW}(H,G,W) &= p(W|H,G) \\ f_{WD}(W,D) &= p(D|W) \\ f_{DC}(D,C) &= p(C|D) \\ f_{C\mathbf{I}}(C,\mathbf{I}) &= p(\mathbf{I}|C) \end{aligned} \quad (102)$$

Factors which describe general deterministic relations between random variables, such as f_{HGW} can be formulated by means of the Kronecker delta, which is discussed in more detail in the next section.

4.1.5 Inference

After the generation of the factor graph and the determination of the corresponding factors, inference can be conducted using the sum-product rule as described in 2.5.2. The messages sent between the different kinds of nodes are depicted in Fig. 23, where the numbering of the arrows corresponds to the numbering of the following equations.

The first step of the sum-product algorithm is the selection of a root node. It is reasonable to select one of the hypothesis variables W or D as the root node to assure that no needless messages are computed. After the selection of the root node, the forward messages from all leaf nodes to the root nodes are calculated. In the following the equations of the massages are given in case that D is the selected as the root node. The message passing starts at all the leaf nodes of the factor graph, which are a variable node $\mathbf{I}$ and two factor nodes $f_G(G)$ and $f_H(H)$. Considering Equation 49, the first messages from the factor nodes f_H and f_G to the variable nodes H and G are defined as

$$\mu_{f_H \to H}(H) = f_H(H) = \mathcal{N}(h|\mu_H, \sigma_H^2), \tag{103}$$

and

$$\mu_{f_G \to G}(G) = f_G(G) = \mathcal{N}(g|\mu_G, \sigma_G^2). \tag{104}$$

The message from a variable node to a factor node is the product of all incoming messages from other factor nodes to the variable node (see equation 47). Since only one incoming message in H and G exists, the messages from H and G to f_{HGW} are simply

$$\mu_{H \to f_{HGW}}(H) = \mu_{f_H \to H}(H) \tag{105}$$

and

$$\mu_{G \to f_{HGW}}(G) = \mu_{f_G \to G}(G). \tag{106}$$

The general equation of a message passing from a factor node to a variable node is given in Equation 44, which is the product of the factor and all incoming messages arriving in the factor node and marginalizing over all of the variables associated with the incoming messages. Therefore, the message from the factor node f_{HGW} to the variable node W is obtained by

$$\mu_{f_{HGW} \to W}(W) = \int_H \int_G f_{HGW}(H, G, W) \mu_{H \to f_{HGW}}(H) \mu_{G \to f_{HGW}}(G). \tag{107}$$

The factor $f_{HGW}(HGW)$ reflects the deterministic relation $W = f(G, H) = G - H$ discussed in Section 4.1.1. In general, deterministic relations between discrete random variables in factor graphs can be formulated by means of the Kronecker delta function (Loeliger, 2004). In case of continuous random variables the deterministic relations are formulated by means of the Dirac delta function $\delta(x)$, which is 0 for $x \neq 0$ with the condition

$$\int_{-\infty}^{+\infty} \delta(x) dx = 1. \tag{108}$$

A defining property of the Dirac delta function using the convolution is

$$\int_{-\infty}^{+\infty} f(x) \delta(a - x) dx = f(a). \tag{109}$$

Loeliger (2004) shows that a general deterministic relation between the continuous random variables $Z = f(X, Y)$ is represented by the factor

$$f_{XYZ}(X, Y, Z) = \delta(z - f(x, y)). \tag{110}$$

Therefore, the message in Equation 107 can be formulated as

$$\mu_{f_{HGW}\to W}(W) = \int_H \int_G \delta(w-(g-h))\mathcal{N}(h|\mu_H,\sigma_H^2)\mathcal{N}(g|\mu_G,\sigma_G^2), \tag{111}$$

which can be written as the cross-correlation of two Gaussian functions

$$\mu_{f_{HGW}\to W}(W) = \int_{-\infty}^{\infty} \mathcal{N}(\tau|\mu_H,\sigma_H^2)\mathcal{N}(\tau+w|\mu_G,\sigma_G^2)d\tau, \tag{112}$$

which results in the Gaussian as stated in Equation 87. Therefore the message is given by

$$\mu_{f_{HGW}\to W}(W) = \mathcal{N}(w|\mu_W = \mu_G - \mu_H, \sigma_W^2 = \sigma_G^2 + \sigma_H^2). \tag{113}$$

Since again only one incoming message to the variable node W exists, the next message to the factor f_{WD} is

$$\mu_{W\to f_{WD}}(W) = \mu_{f_{HGW}\to W}(W). \tag{114}$$

The following message to the selected root node D is given by

$$\mu_{f_{WD}\to D}(D) = \int_W f_{WD}(W,D)\mu_{W\to f_{WD}}(W), \tag{115}$$

whereas the factor $f_{WD} = p(D|W)$ corresponds to the Heaviside step function $\Theta(w)$ defined in Equation 88. Therefore the message for $D = f$ is

$$\mu_{f_{WD}\to D}(D=f) = \int_W \Theta(w)\mathcal{N}(w|\mu_W,\sigma_W^2). \tag{116}$$

The integral over the multiplication of the Heaviside function with a Gaussian can be calculated by the cumulative distribution function of the Gaussian, which leads to

$$\mu_{f_{WD}\to D}(D=f) = 1 - \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{-\mu_W}{\sqrt{2\sigma_W^2}}\right)\right], \tag{117}$$

whereas

$$\operatorname{erf}(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2}dt. \tag{118}$$

Similar the messages for $D = \overline{f}$ can be calculated as

$$\mu_{f_{WD}\to D}(D=\overline{f}) = \int_W (1-\Theta(w))\mathcal{N}(w|\mu_W,\sigma_W^2) \tag{119}$$

$$= \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{-\mu_W}{\sqrt{2\sigma_W^2}}\right)\right]. \tag{120}$$

The massages starting from the factor leaf nodes f_H and f_G have now arrived at the selected root node D. But the calculation of marginal probability of the root node D requires all

incoming messages. Therefore, the messages starting from the variable leaf node $\mathbf{I}$ in direction to the root node are described in the following. The first message from the variable leaf node $\mathbf{I}$ to the factor $f_{C\mathbf{I}}$ is initialized as described in Equation 48.

$$\mu_{\mathbf{I}\to f_{C\mathbf{I}}} = 1 \tag{121}$$

Therefore the message

$$\mu_{f_{C\mathbf{I}}\to C}(C) = \sum_{\mathbf{I}} f_{C\mathbf{I}}(C,\mathbf{I})\mu_{\mathbf{I}\to f_{C\mathbf{I}}} \tag{122}$$

simplifies to

$$\mu_{f_{C\mathbf{I}}\to C}(C) = \sum_{\mathbf{I}} p(\mathbf{I}|C). \tag{123}$$

The conditional probability $p(\mathbf{I}|C)$ are the Gaussian mixtures defined in equations 90 - 92 or the functions defined in 95 and 96 depending on the class. Since the random variable $\mathbf{I}$ is given, the message $\mu_{f_{C\mathbf{I}}\to C}(C)$ reduces to an m-dimensional vector of probabilities, whereas m is the number of different classes. The inference between the random variable $\mathbf{I}$ and C is equivalent to a multispectral maximum likelihood classification.

The last two forward messages in direction to the root node D are

$$\mu_{C\to f_{DC}}(C) = \mu_{f_{C\mathbf{I}}\to C}(C) \tag{124}$$

and

$$\mu_{f_{DC}\to D}(D) = \sum_{C} f_{DC}(D,C)\mu_{C\to f_{DC}}(C). \tag{125}$$

Inserting the conditional probability $p(C|D)$ corresponding to the conditional probability table defined in Tab. 1, the message arriving at D is

$$\mu_{f_{DC}\to D}(D) = \sum_{C} p(C|D) \sum_{\mathbf{I}} p(\mathbf{I}|C). \tag{126}$$

Now, all incoming messages at the root node D are calculated and therefore the marginal probability of D can be calculated by multiplying all incoming messages, which is

$$p(D) = \mu_{f_{WD}\to D}(D)\mu_{f_{DC}\to D}(D). \tag{127}$$

The message passing from the leaf nodes to the root node D is an example of the single-i sum-product algorithm described in Section 2.5.2. The result for the marginal probability $p(D)$ is equivalent to the result obtained by applying the variable elimination algorithm shown in Equation 81. However, the calculation of further marginal probabilities applying the general sum-product rule is more efficient. This advantage will be apparent if the general sum-product rule is applied to the topology-based model. The calculation of all the other marginal probabilities in the pixel-based model requires all the backward messages starting from the root node D back to the leaf nodes f_H, f_G and $\mathbf{I}$. Since in the pixel-based model only the marginal probabilities of D and W are of interest, only the backward messages starting from the root node D in direction to W are necessary. The first backward massage is

$$\mu_{D\rightarrow f_{WD}}(D) = \mu_{f_{DC}\rightarrow D}(D). \tag{128}$$

The message to the second hypothesis variable W is

$$\mu_{f_{WD}\rightarrow W}(W) = \sum_{D} f_{WD}(W, D)\mu_{D\rightarrow f_{WD}}(D). \tag{129}$$

For $D = f$ the message is

$$\mu_{f_{WD}\rightarrow W}(W) = \sum_{D} \Theta(w)\mu_{D\rightarrow f_{WD}}(D = f) \tag{130}$$

and respectively for $D = \overline{f}$

$$\mu_{f_{WD}\rightarrow W}(W) = \sum_{D} (1 - \Theta(w))\mu_{D\rightarrow f_{WD}}(D = \overline{f}). \tag{131}$$

Therefore the message can be expressed as a function $\Theta^*(w)$ as follows

$$\mu_{f_{WD}\rightarrow W}(W) = \Theta^*(w) = \begin{cases} \mu_{D\rightarrow f_{WD}}(D = f), & if \quad w > 0 \\ 0.5, & if \quad w = 0 \\ \mu_{D\rightarrow f_{WD}}(D = \overline{f}), & if \quad w < 0. \end{cases} \tag{132}$$

The marginal probability $p(W)$ can now be calculated since all incoming messages are given. In the forward message passing the message $\mu_{f_{HGW}\rightarrow W}(W)$ (see Equation 113) and in the backward message passing the message $\mu_{f_{WD}\rightarrow W}(W)$ is calculated, which leads to the marginal distribution

$$P(W) = \mu_{f_{HGW}\rightarrow W}(W)\mu_{f_{WD}\rightarrow W}(W) \tag{133}$$

$$= \Theta^*(w)\mathcal{N}(w|\mu_W, \sigma_W^2). \tag{134}$$

Again the marginal distribution $P(W)$ is equivalent to Equation 80 if $\mathbf{I}$ is given.

The pixel-based model describes the causal physical relation between the information variables and the hypothesis variables. In general, the Bayesian network presented in Section 4.1.1 can be applied to every pixel in an image and not only to linear infrastructure objects. However, if the Bayesian network is used for the whole image, the states of the random variable class C have to be adapted to cover all possibly appearing classes. In the following the focus is on the assessment of linear infrastructure objects such as road networks.

4.2 Topology-based model

The pixel-based model assesses each pixel individually without considering the dependence among neighboring pixels. However, the consideration of neighborhood relations is crucial for robust methods which are able to handle noise and uncertain data. In addition, the inclusion of the local neighborhood leads to an improved estimation of the requested random variables. The topology-based model uses road vector data given by a GIS to model the dependencies of

neighboring pixels. The better the resolution of the input data (DEM and imagery), the more important is a high accuracy of the road vector data. In the following it is assumed that road data with high spatial accuracy is given. The accuracy of the given road vector data is not considered in the model. But due to the large variability of the accuracy of the available road vector data, it should be part of future work.

4.2.1 Representation

The topology of the road network is represented as a Markov network as depicted schematically in Fig. 24. A Markov network is preferred to a Bayesian network since the dependencies between neighboring pixels are symmetrical and not causal. Each node in the Markov network represents a pixel in the image data. Every node has two nodes as neighbors since only the center axis of the road is considered. Exceptions are nodes at crossings. However, in the following only road segments between two crossings are considered, which leads to simple Markov chains. The partitioning of the whole road network into Markov chains between crossings avoids loops, which occur if the whole road network is represented as a Markov network. In the case the whole road network is represented as a Markov network, alternative inference algorithms to calculate the marginal probabilities such as the loopy belief propagation have to be applied. However, the loops in the Markov network representing a complete road network usually consist of a very large number of nodes and therefore, the gain of applying loopy belief propagation is limited. In addition, the loopy belief propagation is very time consuming compared to the sum-product algorithm. The limited gain together with the time consuming calculations leads to a poor cost-benefit ratio. In case the road segments between crossings are represented as Markov chains, the sum-product algorithm can be applied to calculate exact inference.

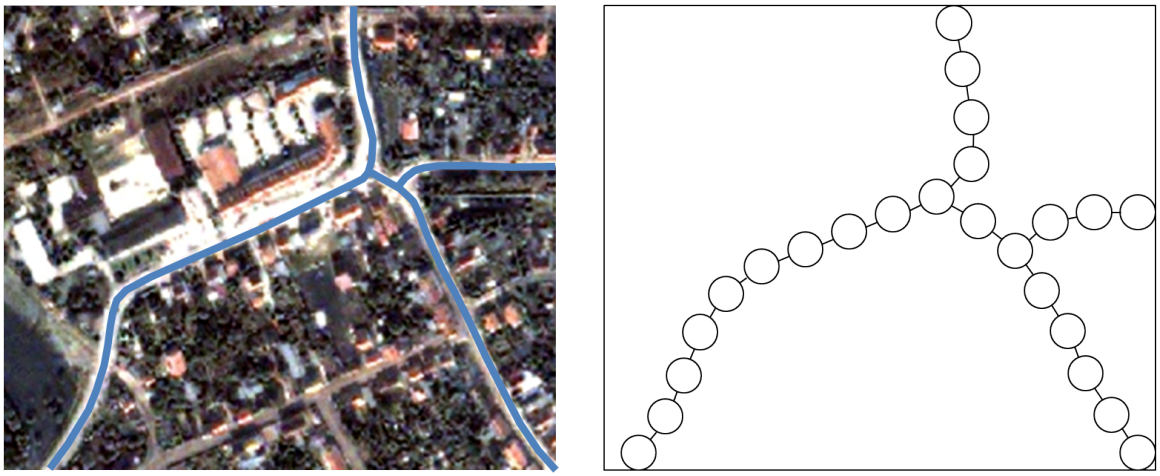

Figure 24. left: IKONOS image overlaid with road information, right: Topology of the road network represented as Markov network.

The combination of the Markov chain with the pixel-based model leads to the proposed topology-based model. The combination is conducted on the basis of factor graphs. In the pixel-based model the neighborhood relations are completely neglected. However, all the variable nodes in the factor graph depicted in Fig. 23 are dependent on the neighboring pixels. For example, the probability that neighboring pixels belong to the same class C is higher than the membership to different classes. Therefore, the variable nodes have to be connected with the corresponding variable nodes of the neighbors given by the Markov chains generated from the given road network. But if all the neighborhood dependencies of the variable nodes are considered the resulting factor graph is not tree-structured and therefore exact inference with the sum-product algorithm is not possible. In the following several models are discussed, in which the dependence of neighborhood is modeled via links between different random variables. Two examples connecting the random variables via G and C are depicted in Fig. 26 or Fig. 27.

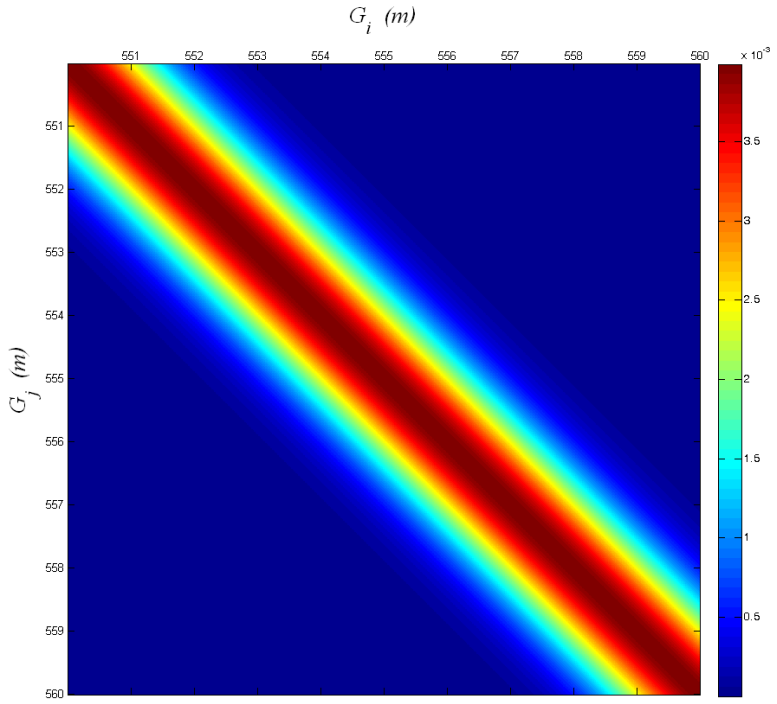

Figure 25. The factor $f_{G_iG_j}$ in the topology-based model

The different topology-based models are denoted as 'topology-based model G', 'topology-based model D' and 'topology-based model C' corresponding to the position of the links describing the neighborhood relation.

In all topology based models only one additional factor which describes the dependence of the neighboring variables G, D or C has to be introduced compared to the pixel-based model. The factor $f_{G_iG_j}$ depicted in Fig. 25 describes the probability that the water gauge changes from a pixel to its neighbor. The factor $f_{G_iG_j}$ is defined as a Gaussian distribution over g_i with the mean g_j

$$f_{G_iG_j} = \mathcal{N}(g_i|g_j, \sigma^2_{GG}), \tag{135}$$

which states that the probability is maximal if the water gauge at the pixel j is equal as the water gauge at the pixel i. In Fig. 25 the factor $f_{G_iG_j}$ is visualized. The standard deviation σ_{GG} depends on the currents, the slope of the flood flow and the resolution of the image. Usually the longitudinal slope of the flood flow is about $1mm/m$ (Mason et al., 2007).

In contrast, the factor $f_{C_iC_j}$ of the 'topology-based model C' depicted in Fig. 27 depends on the two discrete variables C_i and C_j. Therefore the factor can be described by means of a discrete probability table instead of continuous distributions. The probability table is shown

Table 2. Probability Table of factor $f_{C_iC_j}$

	r	w	c	v	e
r	p_{equal}	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$
w	$\frac{1-p_{equal}}{m-1}$	p_{equal}	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$
c	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$	p_{equal}	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$
v	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$	p_{equal}	$\frac{1-p_{equal}}{m-1}$
e	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$	$\frac{1-p_{equal}}{m-1}$	p_{equal}

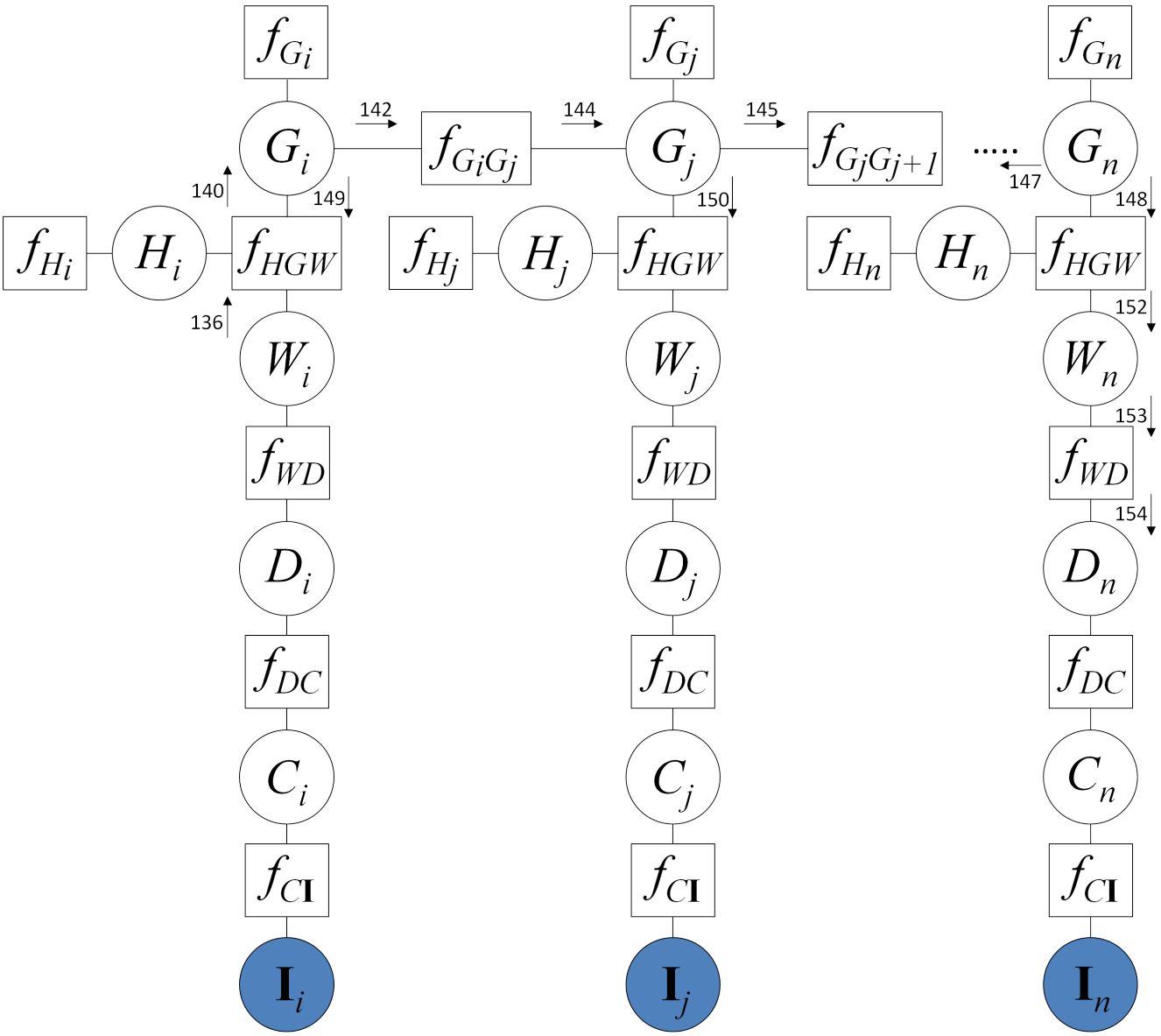

Figure 26. The 'topology-based model G' represented as factor graph. The numbered messages refer to the numbering of the equations in Section 4.2.2

in Tab. 2, where p_{equal} is the probability, that the class of two neighboring pixels is equal and m is the number of classes. Similar in the 'topology-based model D' the factor $f_{D_iD_j}$ describes the probability, that two neighboring pixels belong to the same damage state flooded or not flooded. In contrast to the table depicted in Tab. 2, the conditional probability table of the factor $f_{D_iD_j}$ has the size of 2×2 consisting only of the probabilities that the damage state changes or not.

All the topology based models consider only the center axis of the roads and neglect the pixel values perpendicular to the road axis. However, depending on the resolution and the width of the road the information of the pixels perpendicular to the center axis deliver additional information which should be exploited. Therefore all the different topology-based models can be expanded by considering the neighboring pixels perpendicular to the center axis of the road. As an example the factor graph of the expanded 'topology-based model C' is shown in Fig. 28. In the next chapter the inference of the topology-based models are discussed.

4.2.2 Inference

Again the sum-product rule is used for inference for all the presented topology-based models. In the following, the messages that are part of the pixel-based model are no longer described since they are already shown in Section 4.1. Firstly, the messages of the 'topology-based model G' depicted in Fig. 26 are discussed. As root node the random variable G_n is assumed. Again, the numbering of the equations corresponds to the number aligned to the arrows in Fig. 26. The first forward message not described in the pixel-based model is

$$\mu_{W_i \to f_{HGW}}(W_i) = \mu_{f_{WD} \to W}(W_i) = \Theta^*(w_i), \tag{136}$$

with $\Theta^*(w)$ defined in Equation 132. The subsequent message in direction to G_i is

$$\mu_{f_{HGW} \to G_i}(G_i) = \int_{W_i} \int_{H_i} f_{HGW}(H_i, G_i, W_i) \mu_{W_i \to f_{HGW}}(W_i) \mu_{H_i \to f_{HGW}}(W_i). \tag{137}$$

Applying Equation 110 and substituting the messages leads to

$$\mu_{f_{HGW} \to G_i}(G_i) = \int_{W_i} \int_{H_i} \delta(g_i - (w_i + h_i)) \Theta^*(w_i) \mathcal{N}(h_i | \mu_{H_i}, \sigma^2_{H_i}), \tag{138}$$

which can be written as a convolution

$$\mu_{f_{HGW} \to G_i}(G_i) = \int_{-\infty}^{\infty} \mathcal{N}(g_i - \tau | \mu_{H_i}, \sigma^2_{H_i}) \Theta^*(\tau) d\tau \tag{139}$$

$$= \left(\mathcal{N}(g_i | \mu_{H_i}, \sigma^2_{H_i}) * \Theta^* \right)(g_i). \tag{140}$$

After the calculation of all incoming messages to the random variable G_i the messages can start to propagate from the first pixel in the road segment i to the neighboring road pixel j. In case of the first pixel of the road axis the massage is

$$\mu_{G_i \to f_{G_i G_j}}(G_i) = \mu_{f_{HGW} \to G_i}(G_i) \mu_{f_{G_i} \to G_i}(G_i), \tag{141}$$

whereas $\mu_{f_{G_i} \to G_i}(G_i)$ is defined in Equation 104, which leads to

$$\mu_{G_i \to f_{G_i G_j}}(G_i) = (\mathcal{N}(h_i | \mu_{H_i}, \sigma^2_{H_i}) * \Theta^*) \mathcal{N}(g_i | \mu_{G_i}, \sigma^2_{G_i}). \tag{142}$$

The incoming message at the random variable G_j of the pixel j is

$$\mu_{f_{G_i G_j} \to G_j}(G_j) = \int_{G_i} f_{G_i G_j}(G_i, G_j) \mu_{G_i \to f_{G_i G_j}}(G_i). \tag{143}$$

Inserting the factor as defined in Equation 135 leads to

$$\mu_{f_{G_i G_j} \to G_j}(G_j) = \int_{G_i} \mathcal{N}(g_i | g_j, \sigma^2_{GG}) (\mathcal{N}(h_i | \mu_{H_i}, \sigma^2_{H_i}) * \Theta^*) \mathcal{N}(g_i | \mu_{G_i}, \sigma^2_{G_i}). \tag{144}$$

The outgoing forward message from a pixel j, which has two neighboring pixels, differs from the message shown in Equation 142, which describes the outgoing forward message for a pixel

at the border of the road axis. Instead of the multiplication of two messages the additional message from the previous pixel has to be considered:

$$\mu_{G_j \to f_{G_j G_{j+1}}}(G_j) = \mu_{f_{G_i G_j} \to G_j}(G_j) \mu_{f_{HGW} \to G_j}(G_j) \mu_{f_{G_j} \to G_j}(G_j). \tag{145}$$

The forward message passing is conducted as long as all incoming messages arrived at the root node G_n. Since only the hypothesis variables W and D are of interest, the backward messages starting from G_n have to be calculated till the messages arrive at the hypothesis variables.

Equivalent to the forward message starting from the random variable G_i, also the backward message starting from random variable G_n in direction to G_{n-1} is the multiplication of only two incoming messages.

$$\mu_{G_n \to f_{G_n G_{n-1}}}(G_n) = \mu_{f_{HGW} \to G_n}(G_n) \mu_{f_{G_n} \to G_n}(G_n) \tag{146}$$

$$= (\mathcal{N}(h_n | \mu_{H_n}, \sigma^2_{H_n}) * \Theta^*) \mathcal{N}(g_n | \mu_{G_n}, \sigma^2_{G_n}) \tag{147}$$

This property results from the simplification that the roads are investigated only in between two crossings, which leads to a biasing at the borders since only the information from one direction is considered. This biasing can be reduced if the information of several pixels of the adjacent roads connected to the crossing are considered and embedded into the model. This information is only used for the forward messages since we are not interested in the marginal probabilities belonging to the adjacent roads. However, the information of adjacent roads connected to the crossing is not considered in this work. The biasing can also be observed at the messages from G_n in direction to the hypothesis variables W_n and D_n. Again, only the message from one direction is used:

$$\mu_{G_n \to f_{HGW}}(G_n) = \mu_{f_{G_{n-1} G_n} \to G_n}(G_n) \mu_{f_{G_n} \to G_n}(G_n). \tag{148}$$

Similar for the first pixel i the message $\mu_{G_i \to f_{HGW}}(G_i)$ is calculated by the summation of only two messages.

$$\mu_{G_i \to f_{HGW}}(G_i) = \mu_{f_{G_i G_j} \to G_i}(G_i) \mu_{f_{G_i} \to G_i}(G_i). \tag{149}$$

In contrast, the message $\mu_{G_j \to f_{HGW}}(G_j)$ of a pixel in between the road axis is obtained by the multiplication of three messages

$$\mu_{G_j \to f_{HGW}}(G_j) = \mu_{f_{G_j G_{j+1}} \to G_j}(G_j) \mu_{f_{G_j} \to G_j}(G_j) \mu_{f_{G_i G_j} \to G_j}(G_j). \tag{150}$$

In the following the messages into the direction of the hypothesis variables W_n and D_n are discussed. Similar to Equation 111 and 112 the message to f_{HGW} is

$$\mu_{f_{HGW} \to W_n}(W_n) = \int_H \int_G \delta(w - (g - h)) \mu_{G_n \to f_{HGW}}(G_n) \mu_{H_n \to f_{HGW}}(H_n) \tag{151}$$

but with the difference that the message $\mu_{G_n \to f_{HGW}}(G_n)$ is no longer Gaussian. Again the message can be formulated as a correlation of the two messages.

$$\mu_{f_{HGW} \to W_n}(W_n) = \int_{-\infty}^{\infty} \mu_{G_n \to f_{HGW}}(w_n) \mu_{H_n \to f_{HGW}}(w_n + \tau) d\tau. \tag{152}$$

All further messages in direction two the hypothesis variable D_n are

$$\mu_{W_n \to f_{WD}}(W_n) = \mu_{f_{HGW} \to W_n}(W_n) \tag{153}$$

and

$$\mu_{f_{WD} \to D_n}(D_n) = \int_{W_n} f_{WD}(W, D) \mu_{W_n \to f_{WD}}(W_n). \tag{154}$$

The messages in Equation 153 and 154 are equivalent for all pixels. Furthermore the marginal probabilities of the hypothesis variables W and D is also equivalent for all pixels and can be calculated by the multiplication of one forward and one backward message. The marginal probability for W_n is

$$P(W_n) = \mu_{f_{HGW} \to W_n}(W_n) \mu_{f_{WD} \to W_n}(W_n) \tag{155}$$

and for D_n is

$$P(D_n) = \mu_{f_{WD} \to D_n}(D_n) \mu_{f_{DC} \to D_n}(D_n). \tag{156}$$

The presented inference for the model depicted in Fig. 26 is similar to the inference of the model in Fig. 27, where the connection of the neighboring pixels is modeled by links between the random variable C. Again the sum-product algorithm is applied and as root node the node C_n is selected. The message $\mu_{f_{WD} \to D}(D)$ is already calculated in Equation 117 or 119. Therefore we obtain the forward message

$$\mu_{f_{DC} \to C_i}(C_i) = \sum_{D_i} f_{DC}(D, C) \mu_{D_i \to f_{DC}}(D_i) \tag{157}$$

with

$$\mu_{D_i \to f_{DC}}(D_i) = \mu_{f_{WD} \to D_i}(D_i). \tag{158}$$

Again the numbering of the equations is illustrated in the Fig. 27. The following forward and backward messages are similar to the previous model. Therefore the formulas of the individual messages are omitted. The only important difference can be seen in the message

$$\mu_{f_{C_i C_j} \to C_j}(C_j) = \sum_{C_i} f_{C_i C_j}(C_i, C_j) \mu_{C_i \to f_{C_i C_j}}(C_i) \tag{159}$$

with

$$\mu_{C_i \to f_{C_i C_j}}(C_i) = \mu_{f_{DC} \to C_i}(C_i) \mu_{f_{C\mathbf{I}} \to C_i}(C_i), \tag{160}$$

where the sum replaces the corresponding integral in Equation 143. The summation over the discrete factor shown in Tab. 2 reduces the processing time compared to the summation over the discretized factor $f_{G_i G_j}$. The effects on the results are demonstrated in Section 5. Equivalently, the inference for the 'topology-based model D' in which the neighborhood relations are modeled via links between the hypothesis variables D is calculated.

The model depicted in Fig. 27 is expanded considering the neighboring pixels perpendicular to the road center axis which is denoted as 'topology-based model C+N' in the following

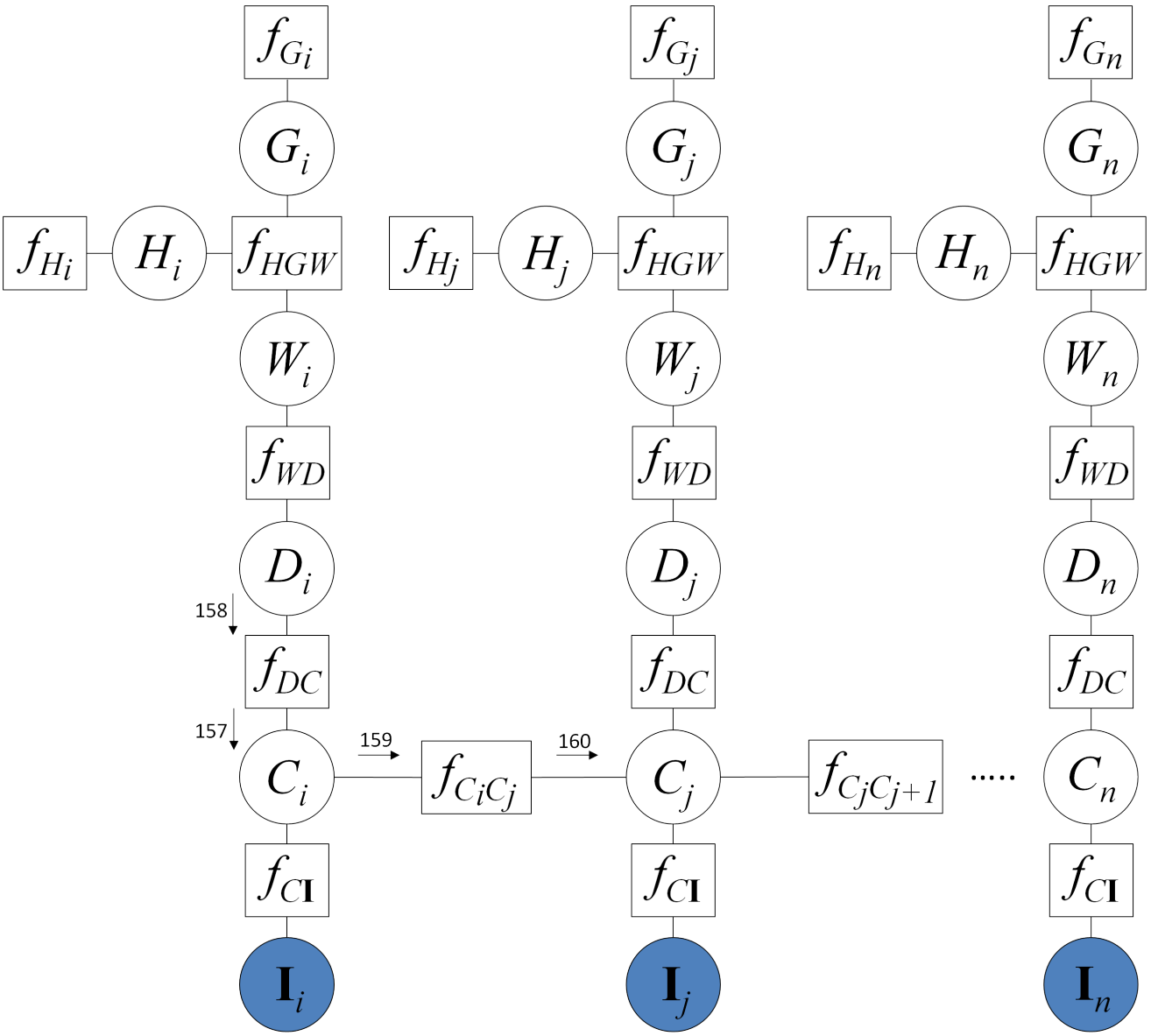

Figure 27. The 'topology-based model C' represented as factor graph linked between the random variables C.

sections. The factor graph of the expanded model is depicted in Fig. 28. Similar to the two models previously discussed, the sum-product algorithm can be applied to calculate inference. The main difference compared to the previous models is apparent by looking at the message $\mu_{C_i \to f_{C_iC_j}}(C_i)$, which is now the multiplication of four messages

$$\mu_{C_i \to f_{C_iC_j}}(C_i) = \mu_{f_{DC} \to C_i}(C_i)\mu_{f_{C\mathbf{I}} \to C_i}(C_i)\mu_{f^+_{CC} \to C_i}(C_i)\mu_{f^-_{CC} \to C_i}(C_i) \tag{161}$$

or for a pixel C_j in between the road axis a multiplication of five messages

$$\mu_{C_j \to f_{C_jC_{j+1}}}(C_j) = \mu_{f_{DC} \to C_j}(C_j)\mu_{f_{C\mathbf{I}} \to C_j}(C_j)\mu_{f_{C_iC_j} \to C_j}(C_j)\mu_{f^+_{CC} \to C_j}(C_j)\mu_{f^-_{CC} \to C_j}(C_j) \tag{162}$$

where

$$f_{C_iC_j} = f^+_{CC} = f^-_{CC}. \tag{163}$$

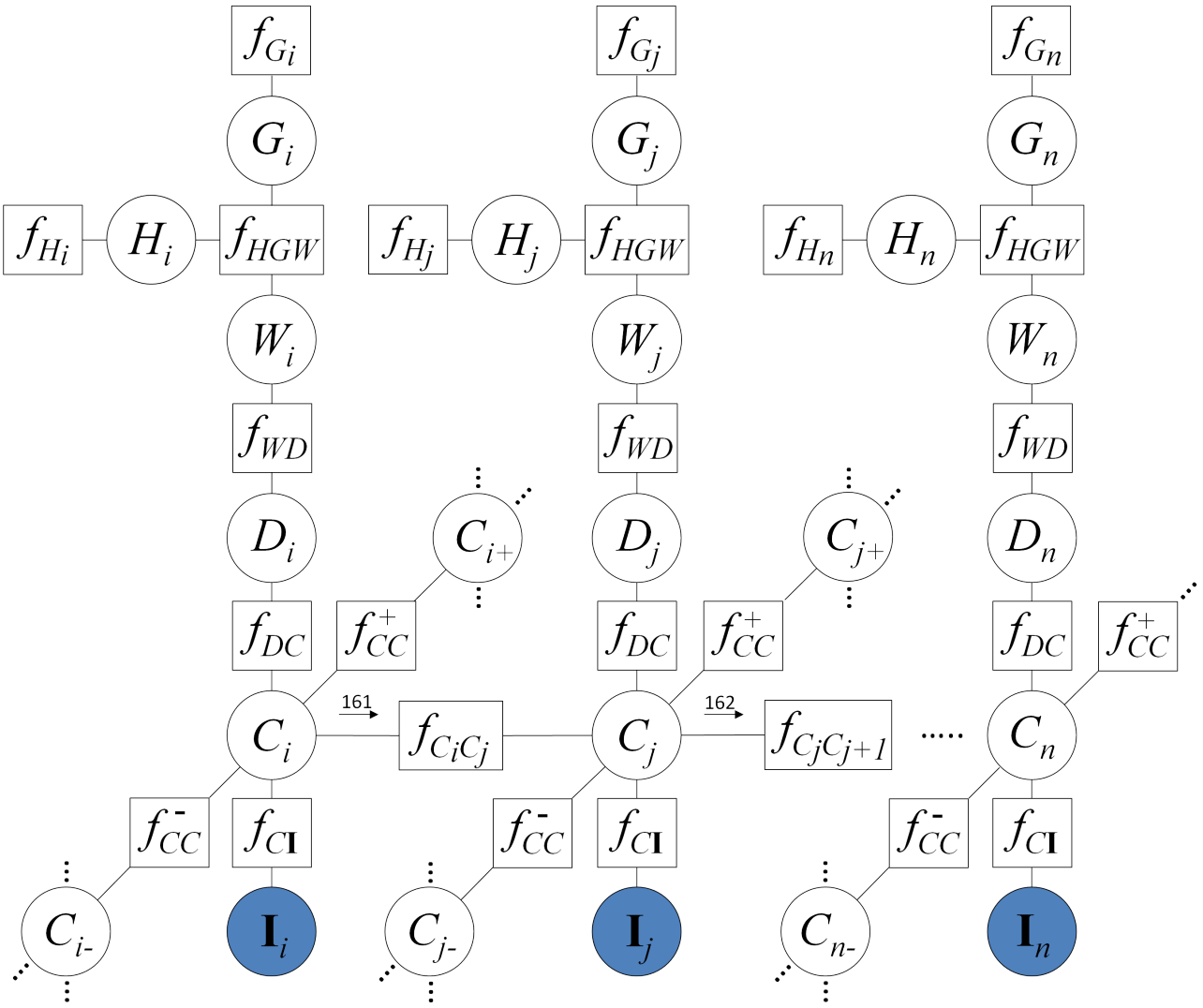

Figure 28. The 'topology-based model C+N' considering the neighbors perpendicular to the road center axis.

The messages $\mu_{f_{CC}^{+} \to C_i}(C_i)$ and $\mu_{f_{CC}^{-} \to C_i}(C_i)$ contain the information of the pixels perpendicular to the road center axis. The number of pixels which are taken into account depends on the spatial resolution of the imagery and the width of the roads which as often given as attribute values in GIS data. However, the problem of the biasing at the borders of the road axis is still present.

4.3 Multi-temporal model

The pixel-based as well as the topology-based models are dealing with information at one specific point in time during a flood event. However, flooding is a highly dynamic process, which can change within short time. One goal of the assessment system is the detection of the flood state at the current point in time. Therefore, all useful information available up to the current time should be exploited. This information includes imagery acquired at a previous state of the flooding and also water gauge measurements at previous times. In particular, if the imagery at a previous point in time has a higher spectral or spatial resolution or if the meteorological condition at a previous point in time is more appropriate to detect the floodplain, the detection of the flood state at the current point in time can be improved. In the following the current point in time is denoted as T and the data at previous point in time as $T-t$, where

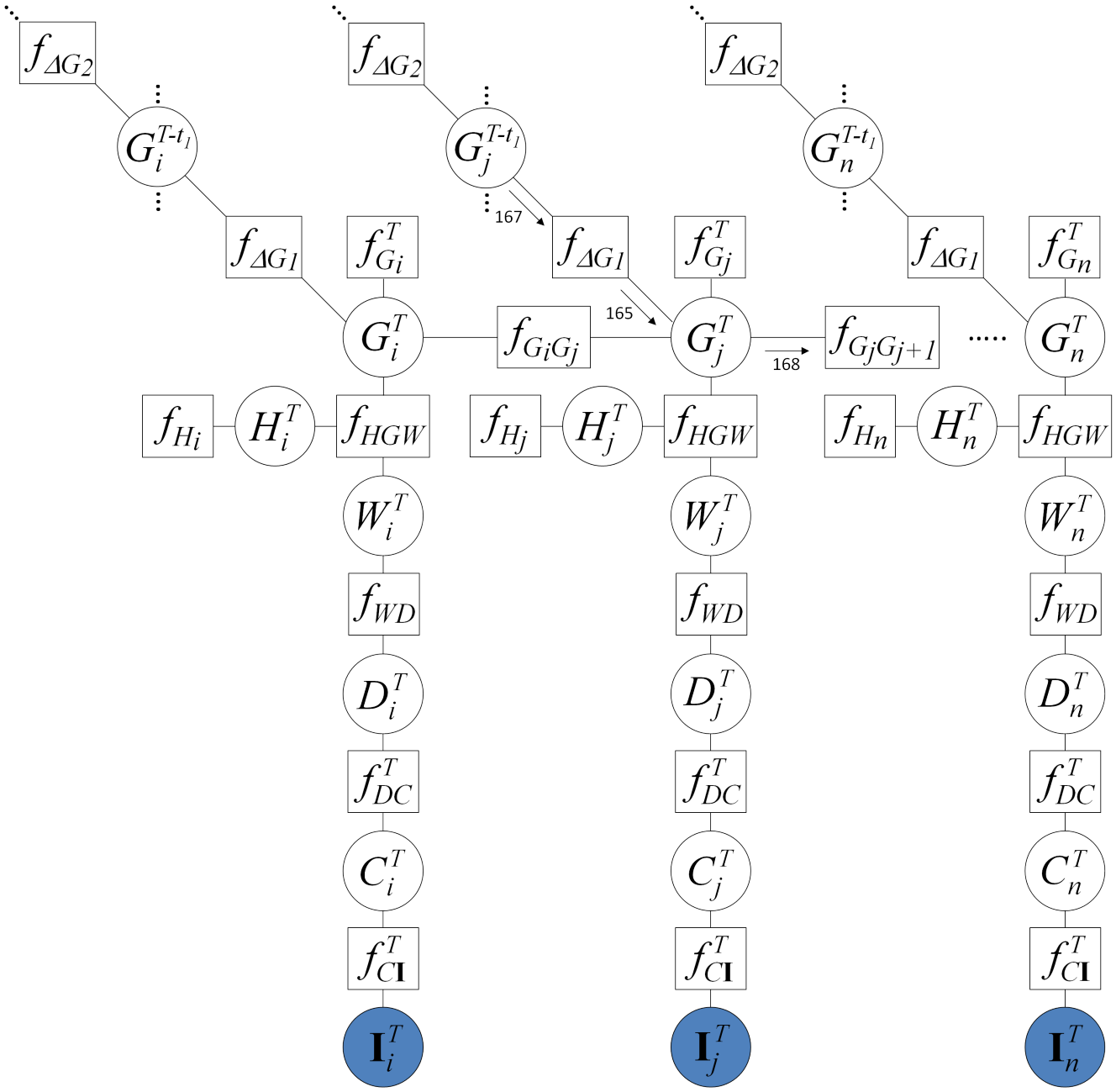

Figure 29. The multi-temporal model represented as factor graph. The numbered messages refer to the numbering of the equations in Section 4.3.2

t denotes the time period between the acquired images.

4.3.1 Representation

The multi-temporal model can be regarded as a combination of the pixel-based model at the time $T - t$ and one of the presented topology-based models at the time T. Again, several possibilities exist to link the pixel-based model with one of the topology-based models. In the following the 'multi-temporal model G' which consists of links between the random variables G at different points in time is exemplarily used to explain the concept. However, the causal dependence between the different points in time can also be modeled via links between D or C as in the topology based models. These models are denoted in the following as 'multi-temporal model C' or 'multi-temporal model D'.

The factor graph of the 'multi-temporal model G' is depicted in Fig. 29. The pixel-based

model is used to calculate the random variables G^{T-t}, which represents the water gauge at the time point $T-t$. The 'topology-based model G' is used for the current point in time but with the additional message from G^{T-t}. The combination of the pixel-based model and the 'topology-based model G' is chosen to retain the tree-structured factor graph. Therefore the spatial dependencies at time $T-t$ are neglected. A more detailed discussion about the disregard of spatial dependencies, which also occur in the topology-based model, is given in Section 4.3.3.

Only the factors $f_{\Delta G}$ are added in the 'multi-temporal model G' compared to the topology-based model, which connects the different points in time. The factor describes the probability of the change of the water gauge in time and can be formulated as a conditional probability $p(G^T|G^{T-t})$. Similar to Equation 135 the factor can be approximated by a Gaussian distribution

$$f_{\Delta G} = p(G^T|G^{T-t}) = \mathcal{N}(g^{T-t}|g^T - \Delta G, \sigma^2_{\Delta G}) \tag{164}$$

where $\sigma^2_{\Delta G}$ describes the uncertainty of $\Delta G = G^T - G^{T-t}$. In contrast, in the 'multi-temporal model C' and the 'multi-temporal model D', the corresponding factors to $f_{\Delta G}$ are the factors $f_{\Delta C}$ or $f_{\Delta D}$ which again can be represented as conditional probability tables. The determination of the factor $f_{\Delta C}$ describing the class changes between the points in time is difficult just as in the 'topology-based model C'. This issue is discussed in more detail in Section 5.6.2. In case of the 'multi-temporal model D' the conditional probability table $p(D^T|D^{T-t})$ describing the factor $f_{\Delta D}$ is easier to determine. Two different scenarios can be distinguished. In case of an ascending water gauge, the conditional probability table is shown in Tab. 3, where the 0 describes that it is not possible that a road is flooded at time $T-t$ and at time T flooded. The knowledge about ascending or decreasing water gauge can directly observed from the water gauge measurements. If no water gauge measurement is available the information can be obtained by the comparison of the imagery at different points in time.

Table 3. Conditional probability table $p(D^T|D^{T-t})$ in case of ascending water gauge

	f	$\overline{f}$
f	1	$p_{\overline{f}f}$
$\overline{f}$	0	$p_{f\overline{f}}$

Equivalently, the conditional probability table for the case of a decreasing water gauge is shown in Tab. 4.

Table 4. Conditional probability table $p(D^T|D^{T-t})$ in case of decreasing water gauge

	f	$\overline{f}$
f	$p_{\overline{f}f}$	0
$\overline{f}$	$p_{\overline{f}\overline{f}}$	1

Since the factor graph is still a tree-structured graph exact inference is possible.

4.3.2 Inference

Again, the sum-product algorithm is applied for inference. Equivalent to the topology-based model, the random variable G^T_n is selected as root variable. In contrast to the topology-based model, an additional message containing information from the previous point in time arrives at the random variable G^T_j

$$\mu_{f_{\Delta G_1} \to G_j^T}(G_j^T) = \int_{G_j^{T-t_1}} f_{\Delta G_1}(G_j^T) \mu_{G_j^{T-t_1} \to f_{\Delta G_1}}(G_j^{T-t_1}), \tag{165}$$

where the outgoing message $\mu_{G_j^{T-t_1} \to f_{\Delta G_1}}(G_j^{T-t_1})$ can be calculated by

$$\mu_{G_j^{T-t_1} \to f_{\Delta G_1}}(G_j^{T-t_1}) = \mu_{f_{HGW} \to G_j^{T-t_1}}(G_j^{T-t_1}) \mu_{f_{G_j^{T-t_1}} \to G_j^{T-t_1}}(G_j^{T-t_1}). \tag{166}$$

If additional data at time $T - t_2$ with $t_2 > t_1$ is available the message is adapted to

$$\mu_{G_j^{T-t_1} \to f_{\Delta G_1}}(G_j^{T-t_1}) = \mu_{f_{HGW} \to G_j^{T-t_1}}(G_j^{T-t_1}) \mu_{f_{G_j^{T-t_1}} \to G_j^{T-t_1}}(G_j^{T-t_1}) \mu_{f_{\Delta G_2} \to G_j^{T-t_1}}(G_j^{T-t_1}). \tag{167}$$

The forward message in direction to the root node results in

$$\mu_{G_j^T \to f_{G_j G_{j+1}}}(G_j^T) = \mu_{f_{HGW} \to G_j^T}(G_j^T) \mu_{f_{G_j^T} \to G_j^T}(G_j^T) \mu_{f_{G_i G_j} \to G_j^T}(G_j^T) \mu_{f_{\Delta G_1} \to G_j^T}(G_j^T). \tag{168}$$

If the width of the road is considered similar as shown in Fig. 28, in addition, the two messages have to be taken into account.

$$\begin{aligned} \mu_{G_j^T \to f_{G_j G_{j+1}}}(G_j^T) = & \mu_{f_{HGW} \to G_j^T}(G_j^T) \mu_{f_{G_j^T} \to G_j^T}(G_j^T) \mu_{f_{G_i G_j} \to G_j^T}(G_j^T) \\ & \mu_{f_{\Delta G_1} \to G_j^T}(G_j^T) \mu_{f_{GG}^+ \to G_j^T}(G_j^T) \mu_{f_{GG}^- \to G_j^T}(G_j^T). \end{aligned} \tag{169}$$

The propagation of the backward messages can be restricted to the random variables at the time T, since only the hypothesis variables W^T and D^T reflect the state of the flooding at the current point in time.

If in the multi-temporal model the spatial neighborhood relations are neglected and only the dependencies in time are considered, the model reflects a dynamic Bayesian network (Hellwich et al., 2000). Since in the time domain only the forward messages are calculated, the dynamic Bayesian network is similar to the Kalman filter equations. However, in case of the Kalman filter, represented as a hidden Markov model, the emission distributions are Gaussian, which is not the case in the presented model.

4.3.3 Discussion

In all presented models, the pixel-based, the topology-based and the multi-temporal model, several neighborhood relations are neglected. Disregarding of some neighborhood relations leads to the presented but also simplified models. For the topology-based model, the consideration of all existing relations would result in a dense 3-dimensional grid structure. The Markov network of the topology-based model is shown in Fig. 30. For ease of visualization the random variable G is not considered in the figure.

In case of the multi-temporal model a dense 4-dimensional grid structure would reflect all existing dependencies. In addition to the links shown in Fig. 30, further links between the corresponding random variables at different points in times exist. Since these grid structures are not solvable by exact inference algorithms such as the sum-product algorithm, alternative energy minimization algorithms could be applied. However, the minimizing of the energy

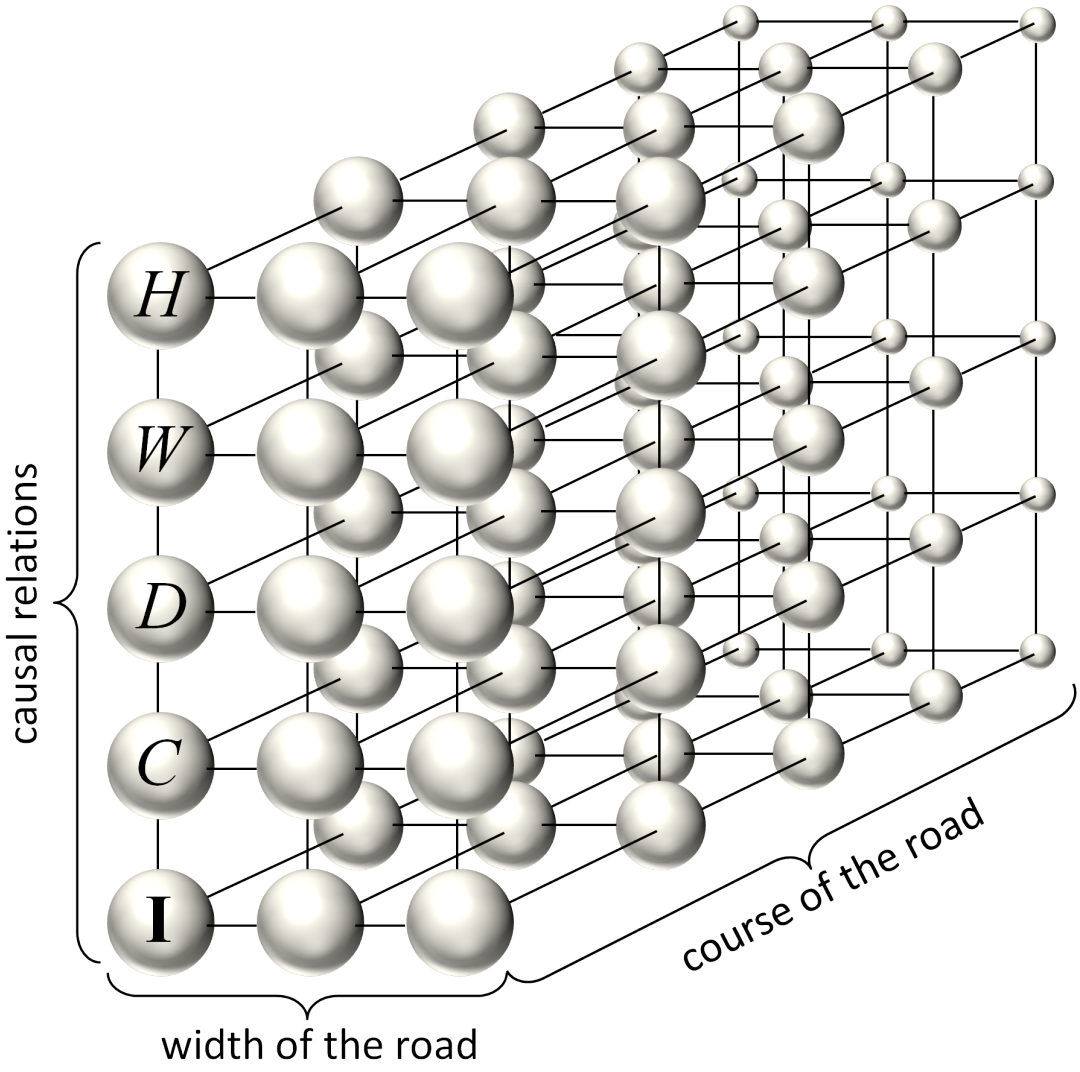

Figure 30. The Markov network of the topology-based model considering all neighborhood dependencies.

functions is not trivial and there is no guarantee to find the global minimum. In addition the energy minimization algorithms are far more time consuming as the proposed exact solutions by the sum-product algorithm. The evaluation of the proposed models is discussed in the following sections.

5 Results and evaluation

In this chapter the three presented probabilistic graphical models are analyzed and evaluated. In Section 5.1, the objectives of the evaluation are discussed in detail. The evaluation of the models is conducted by means of two test scenarios, which are completely different concerning the used input data. The two test scenarios are described in Section 5.2. An important issue for the evaluation is the availability of ground truth. The handling and generation of the ground truth is discussed in Section 5.3. The basic concepts used for the evaluation as the Receiver Operating Characteristic (ROC) curves are introduced in Section 5.4. The ROC curves obtained by the simulation based on the DEM and classification using only the image data are discussed in Section 5.5. In Section 5.6 the three presented probabilistic graphical models are evaluated. The benefit of the fusion of imagery and DEM via the pixel-based model is shown. In addition, the effects of different input data with varying quality and the influence of the factors on the results is presented. Finally, the results obtained by the probabilistic graphical models are discussed in Section 5.7.

5.1 Goals of evaluation

The main goal of the evaluation is the investigation of the developed probabilistic graphical models concerning their benefits and their limitations. Firstly, the benefit of combining the image information with the elevation data is shown. Secondly, the sensitivity concerning the conditional probabilities in the pixel-based model is evaluated. In the topology-based model the influence of the factors $f_{G_jG_{j+1}}(G_j, G_{j+1})$, $f_{C_jC_{j+1}}(C_j, C_{j+1})$ and $f_{D_jD_{j+1}}(D_j, D_{j+1})$ is investigated. In addition, the influence of different DEM data with varying accuracy is evaluated, which leads to a discussion of the required quality of the data in order to enable a rapid and reliable emergency respond.

5.2 Test scenarios

Test scenarios which are suitable for the evaluation of the developed probabilistic graphical models have to fulfill some basic requirements. Firstly, image data during flooding and a DEM have to be available. Secondly, a reference is needed to compare it with the results obtained from the models. The availability of image data during flooding and DEMs is not a limiting factor for the evaluation. In contrast, the existence of a ground truth in cases of flooding at a specific time is rare since the generation is difficult and tedious. Therefore, the evaluation is restricted to two test scenarios with existing reference data. One test scenario is the Elbe flooding in the year 2002 near Dessau (see. Section 5.2.1) and the second test scenario is a flooding at the river Chobe in Namibia in the year 2009 (see. Section 5.2.2).

5.2.1 Elbe river flooding

The first test scenario addresses the one-hundred-year flood in August 2002 at the Elbe near Dessau. The Elbe river flood led to over 20 fatalities and caused damages of more than 25 billion Euros. A detailed discussion of the effects of the flooding with focus on the geosciences is given in (Niedermeyer, 2006). The evaluation is restricted to the area depicted in Fig. 31, which shows a false-color infrared composite of an IKONOS scene. On the left side the city Dessau can be seen which is connected with a bridge over the Mulde at the top left of the image with Rosslau. A small part of the natural stream channel of the Elbe river can be seen in the

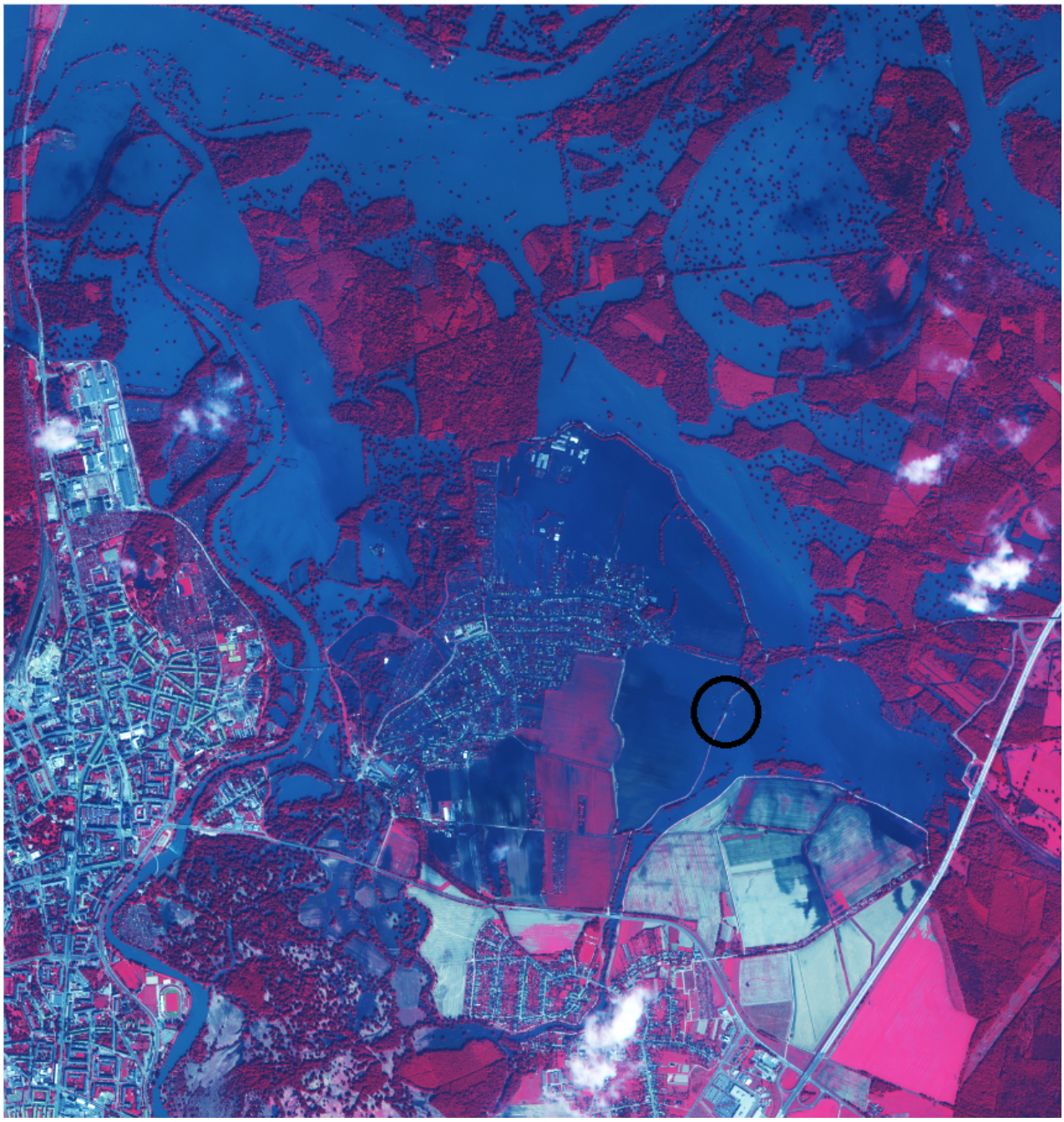

Figure 31. False-color infrared composite of IKONOS showing the river flood on Elbe 2002

middle of the top. In the center of the image, a district of Dessau with the name Waldersse can be seen, which was almost completely flooded due to a dam failure. The breach in the dam occurred east of Waldersee, highlighted by the black circle in Fig. 31. Waldersee and the city of Dessau are separated by the river Mulde, which also contributed to the flood disaster. The river Mulde ranges from the bottom left and flows into the river Elbe at the top left. The water gauges of the river Elbe and the river Mulde are depicted on the top of Fig. 33.

The available IKONOS image was taken on August 21 in 2002, two days after the maximum water gauge of the Elbe and three days after the breach of the dam. The IKONOS image consists of four bands: blue ($\lambda = 0.445 - 0.516\ \mu m$), green ($\lambda = 0.506 - 0.595\ \mu m$), red ($\lambda = 0.632 - 0.698\ \mu m$) and near-infrared ($\lambda = 0.757 - 0.853\ \mu m$). The resampled spatial resolution is 1 m in the panchchromatic band and 4 m for the color bands. The radiometric resolution is 11 bit. An additional IKONOS scene at 26th of August is available which shows the situation one week after the maximum water gauge of the river Elbe. However, no ground

Figure 32. Digital Elevation Model with 1 m resolution showing the same area as Fig. 31

truth at this time was available. Therefore the obtained results for the scene at 26th of August cannot be evaluated quantitatively. For this reason the Elbe scenario is used to evaluate only the pixel-based and topology-based model and is not used to evaluate the multi-temporal model due to the lack of ground truth.

Besides the image data, a high resolution digital terrain model (DTM) generated from airborne laser scanning is used. The DTM is provided by the Federal Agency for Cartography and Geodesy Sachsen-Anhalt. The spatial resolution of the DTM depicted in Fig. 32 is 1m with an accuracy of +/- 0.15m. The river Elbe as well at the river Mulde can be recognized by low elevation, which corresponds to dark grayvalues. Furthermore, the bright linear structures correspond to raised roads or the dam system. In general, it is important to use a DTM and not a digital surface model, since we are interested in the points on the ground in order to simulate the expansion of the flooding. In digital surface models, the highest points are mapped, such as the treetops which lead to an erroneous flood simulation. However, in cases of bridges, the

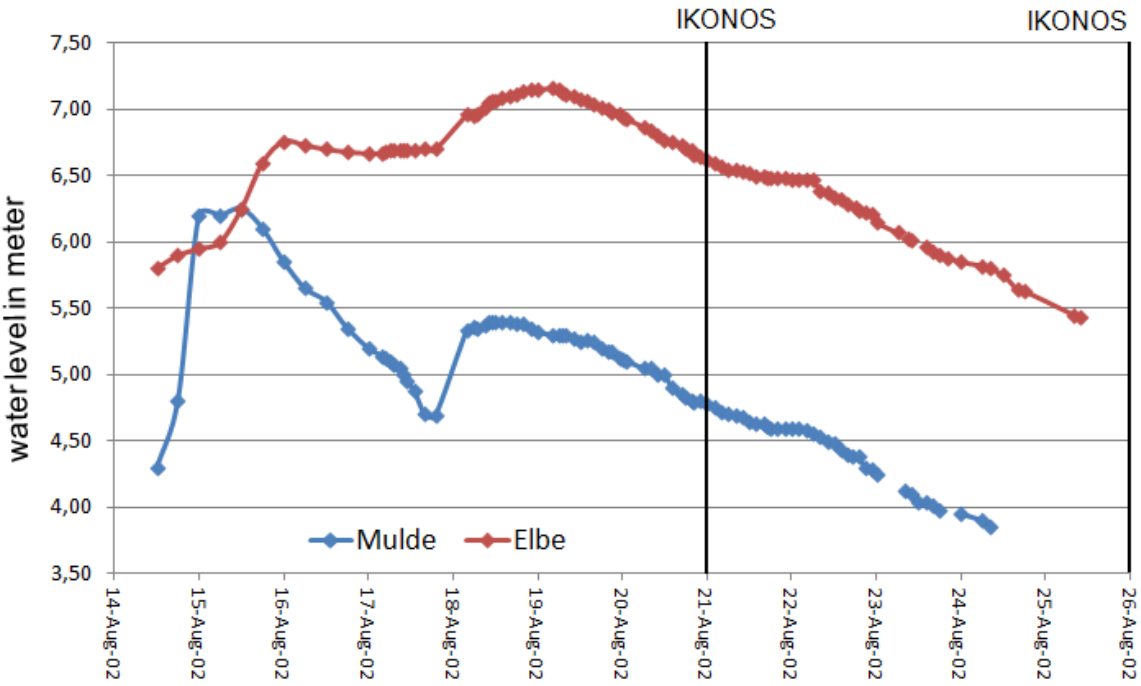

Figure 33. Water level measurements at the river Elbe and Mulde in Dessau and the points in time of the acquisition of IKONOS images.

used DTM leads to erroneous results, since not the elevation of the road is mapped, but the elevation of the points under the road. An example of a bridge can be seen at the top left in Fig. 32. In addition, only in digital surface models the information of the building structures is available, which can be used for an improved flood simulation. Therefore, the optimal elevation information would be a fusion of a DTM with a digital surface model. Although the future TanDEM-X mission allows the generation of highly accurate DEMs on a global scale, it cannot be assumed that a DEM with 1m resolution is worldwide available. Therefore, an additional

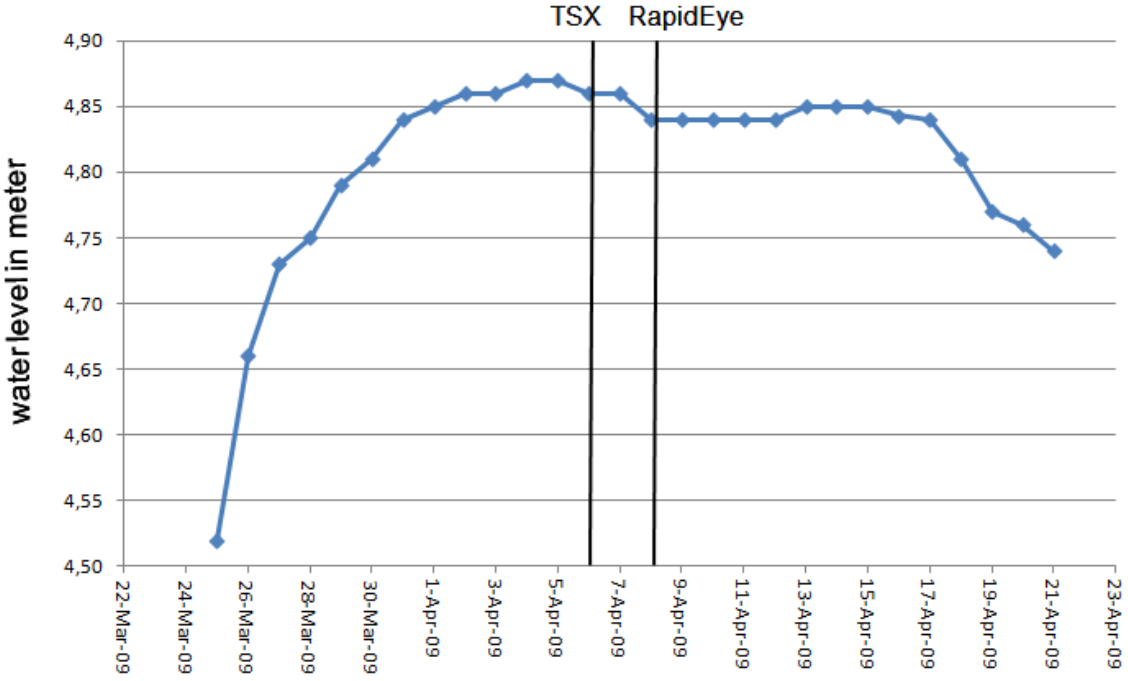

Figure 34. Water level measurements at the river Chobe and the points in time of the acquisition of different satellite images.

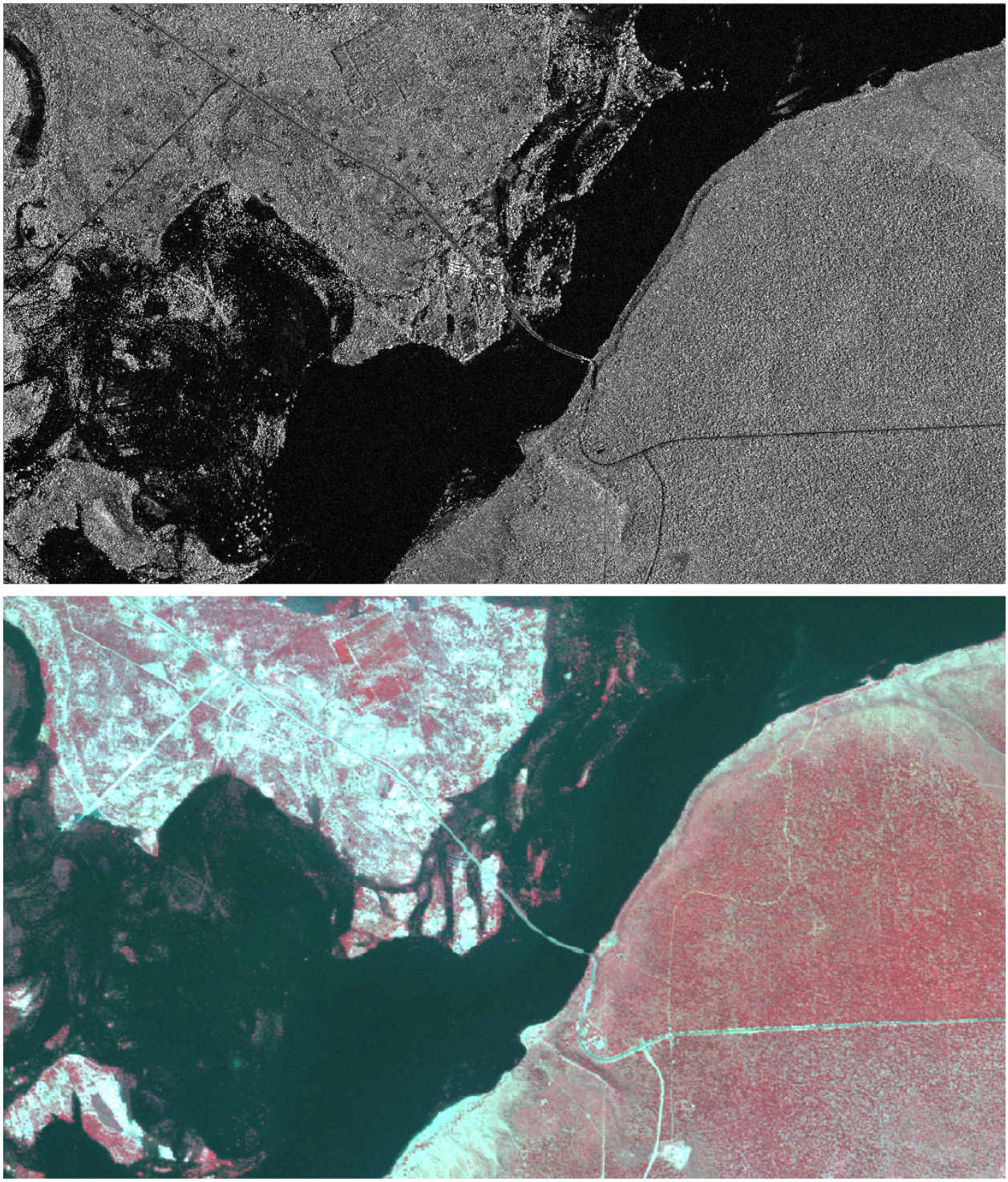

Figure 35. TOP: TerraSAR-X image acquired in SpotLight mode at 6th of April. Bottom: False-color infrared composite of the RapidEye image acquired at April 8th.

DEM obtained from the SRTM mission is used for the evaluation. Furthermore, the used road data have elevation information where the root mean square error of the absolute vertical accuracy is 1m. The influence of the three different datasets for elevation is investigated in Section 5.5.

The used road information is taken from the Intermap road database where the root mean square error of the absolute horizontal accuracy is 3m. The root mean square error of the relative horizontal accuracy which is the accuracy between two neighboring points is 1m. Alternatively, the road information from ATKIS (Amtliches Topographisch-Kartographisches In-

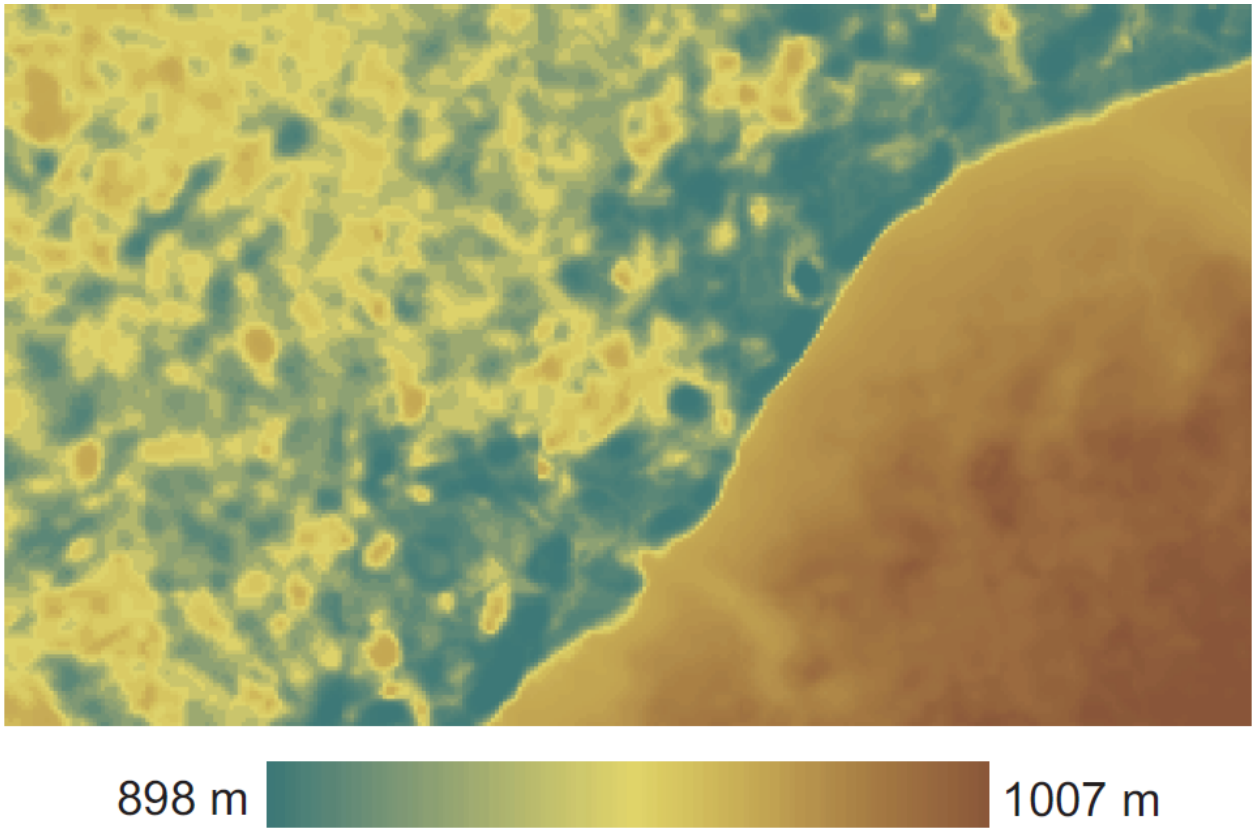

Figure 36. ASTER Digital elevation model for the river Chobe scenario with spatial resolution of 15m.

formationssystem) or from OpenStreetMap can be used.

In Fig. 33 water level measurements from the river Elbe and Mulde between the 14th and 26th of August are plotted. In addition, the points in time of the available IKONOS image data are shown. The second IKONOS image at the 26th of August is not been considered in this study since for the most part the water has receded to the original river channels and only small water basins remained flooded. In addition, it was not possible to generate ground truth for the 26th of August due to missing information. However, a change detection approach between the two IKONOS images based on the MAD algorithm described in Section 3.1.3 is applied in Butenuth et al. (2011).

5.2.2 Chobe river flooding

The second test scenario investigates the flooding that took place at the Chobe river in March 2009, which caused 85 fatalities in the Caprivi region in Namibia alone and forced thousands to evacuate. The river rised in the central plateau of Angola and downstream represents the border between Namibia and Botswana. In Fig. 34 the water level measurements during the flooding between the 25th of March and 21st of April are depicted. The water level rises till March 31th and then remains static between 1st of April and April 17th. Furthermore the times of the acquisition of satellite imagery is shown.

Two satellite images are available showing the flood situation. Firstly, a high resolution TerraSAR-X (TSX) image acquired in SpotLight mode with a spatial resolution of 1 m. The TSX image was acquired on April 6th and reflects the flood situation near the maximum water level (see Fig. 35). Secondly, a RapidEye image was taken two days after the TSX scene at April 8th (see Fig. 34). The RapidEye image consists of five bands (red, green, blue, red edge and infrared) and has a spatial resolution of 6.5 m. Again a false-color infrared composite of the RapidEye image is shown on the bottom on Fig. 35. Since the change of water level between

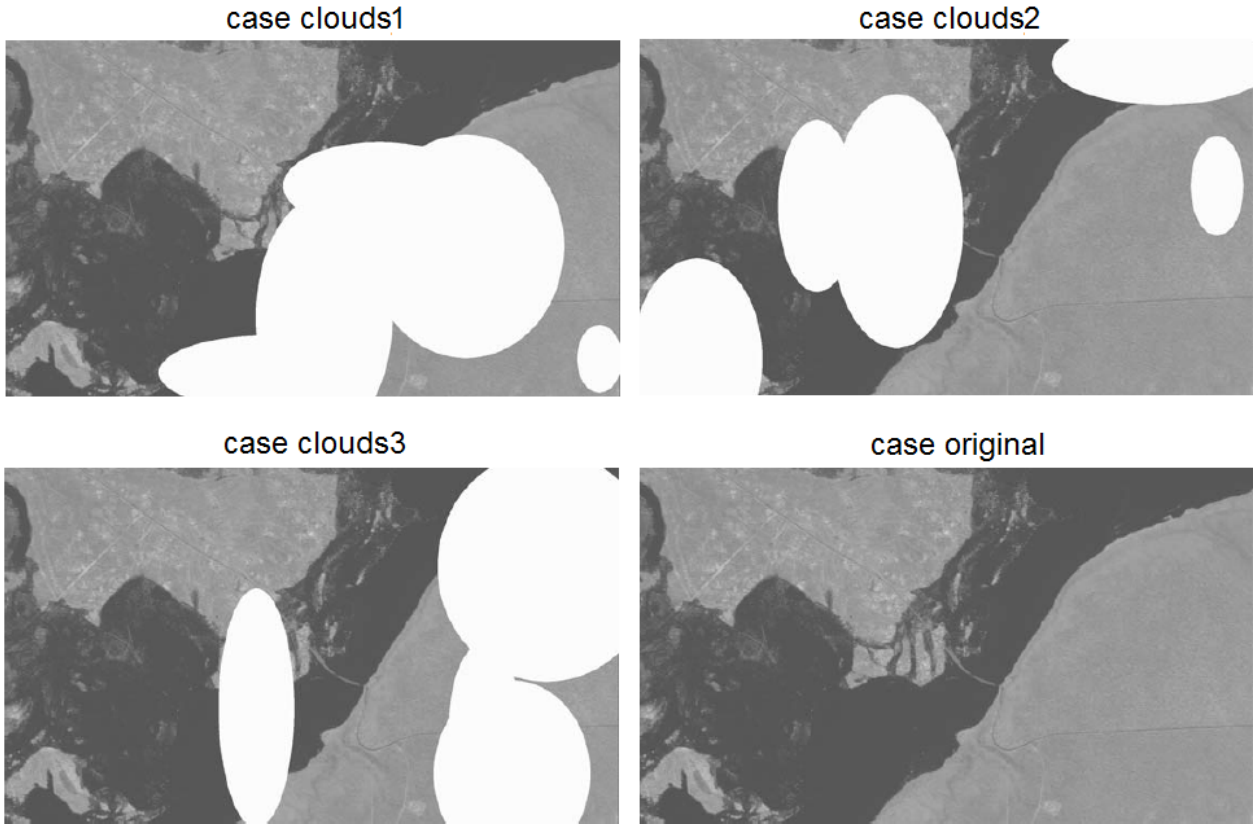

Figure 37. The infrared band of the three cases with different artificial cloud coverage and the original river Chobe scenario

the two acquisitions is only about 3 cm the change of the expansion of flooded areas is limited.

In the Caprivi region, a high resolution DEM as used for the river scenario is not available. Therefore, the only height information available for this region is an ASTER DEM with a spatial resolution of 15 m. However, the difference of the input data of the two test scenarios underlines the transferability and flexibility of the developed method. The DEM of the same area as the two acquired images is depicted in Fig. 36. Since no comprehensive road information was available for the evaluation of the developed models, a dense road network was extracted manually from a high resolution Quickbird scene showing a not flooded situation.

The images available for the river Chobe scenario are on the one hand free of clouds and on the other hand there are almost no occlusions by vegetation. These circumstances lead to the opportunity to generate the reference by means of the imagery itself. However, generally imagery without any occlusions by clouds and vegetations are rare. Therefore, additional cases with artificial clouds are generated in order to demonstrate the performance of the developed models in case of occlusions. The artificial clouds have the form of ellipses and are overlaid to the original image. They are generated randomly and consist of a Gaussian noise, which is trained from clouds in other image data. The original case and the three cases with different artificial cloud coverage are depicted in Fig. 37. In the following it is important to distinguish between the scenarios and cases. The evaluation is conducted by means of two different scenarios (river Elbe and river Chobe scenario), whereas in the river Chobe scenario four different cases are distinguished varying in the artificial cloud coverage.

5.3 Reference data

Reference data is needed to evaluate the results obtained by the usage of the different probabilistic graphical models. The goal of the evaluation is the comparison of results obtained by

Figure 38. The generated reference data of the Elbe scenario with the aid of expert knowledge and the IKONOS image. (white: flooded areas, black: not flooded areas)

the models with the real state of the flooding. Therefore, the best reference data is the data collected directly on location at the same time the image is acquired, which is often called ground truth. However, the generation of ground truth is very time consuming and costly. In case of flooding it is nearly impossible to get a ground truth at the same time of the image acquisitions due to the dynamic processes of flood events. Therefore two different strategies are applied for the two different test scenarios to generate the reference data sets.

The Elbe scenario is a complex scenario with a lot of occlusions through vegetations and clouds. Therefore, the generation of the reference by means of using only the imagery is not very promising and leads to many regions impossible to assign to flooded or not flooded areas. In addition, there are areas which are not occluded by vegetation or clouds and still difficult to classify. An example is discussed in Section 5.7 (see Fig. 57). Therefore the reference is generated on location with the knowledge of experts. Together with the operation commander of the disaster control of Dessau-Rosslau, an on-site inspection has taken place covering the whole test scenario. The expert knowledge combined with the IKONOS imagery leads to the reference data depicted in Fig. 38. However, even by means of expert knowledge it was not possible to reconstruct the flooded areas at the 26th of August when the second IKONOS image

Figure 39. The generated reference data for the RapidEye image of the Chobe scenario with the aid of the image data. (white: flooded areas, black: not flooded areas)

was acquired.

For the second river Chobe test scenario, an on-side inspection was not possible. However, the images (see Fig. 35) are not affected by cloud coverage and only small occlusions by vegetation exist. Therefore it is possible to derive the flooded areas directly from the imagery. The generated reference for the RapidEye image is shown in Fig. 39. Since the evaluation of the river Chobe scenario is restricted to the point in time when the RapidEye image was acquired, no reference data is generated for the other points in time.

Both reference data sets for the two scenarios are of course not free of errors. However, the level of quality of the reference data sets allows a detailed and accurate evaluation of the results. Furthermore, for both scenarios it was not able to generate a satisfactory reference data set considering the water level of the road network. Therefore the evaluation of the water level is restricted to visual interpretation described in Section 5.7.

5.4 Basics of evaluation

In the following the basic concepts are described that are used for the evaluation of the results. The initially stated task, namely assessing the roads if they are flooded or not flooded can be regarded as a binary classification or detection problem. Therefore, the receiver operating characteristic (ROC) curves are applied to demonstrate the performance of the different developed models (Egan, 1975). The comparison of the results with the reference leads to four possible outcomes, which defines the confusion matrix with the size of 2 times 2. In the specific case of the assessments of roads during flooding the entries for the confusion matrix are: The result states "flooded" and the reference states "flooded" which is called true positive (TP) or hit. The result states "flooded" in the reference states "not flooded", which is called false positive (FP) or false alarm. The result states "not flooded" and the reference states "flooded", which is called false negative (FN) or missed hit and the result states "not flooded" and the reference states also "not flooded", which is called true negative (TN) or correct rejection. Using these four outcomes the two characteristic numbers, the true positive ratio (TPR) and the false positive ratio (FPR) can be calculated, which define the axes of the ROC curves. The TPR defines

how many road pixels are assigned correctly to flooded road pixels among all existing flooded road pixels in the reference

$$TPR = \frac{TP}{TP + FN}. \tag{170}$$

On the other hand the FPR defines the number of road pixels assessed to flooded by the model but which are in reality not flooded (FP) among all not flooded road pixels

$$FPR = \frac{FP}{FP + TN}. \tag{171}$$

Beside the TPR and the FPR, the overall accuracy is defined as

$$overall\ accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{172}$$

However, the information content of the overall accuracy is limited, if the occurrence of the existing states is unequally distributed. The confusion matrix, the TPR, the FPR and the overall accuracy can be computed comparing the results with the reference. However, the TPR and FPR of one result reflect only one point in the ROC space. For the computation of the complete ROC curve, a probability threshold is defined, which describes the probability that a road pixel is assigned to flooded or not flooded. Usually the threshold is fixed at 0.5. By shifting this probability threshold from 0 (assigns all road pixels to flooded, $TPR = FPR = 1$) to 1 (assigns no road pixel to flooded $TPR = FPR = 0$) the true positive rate (TPR) and the false positive rate (FPR) for each probability threshold can be calculated and therefore the complete ROC curve can be computed for each detection model.

5.5 Simulation and classification

First of all, the pixel-based model is split in order to evaluate the damage assessment if only the imagery is used or only the DEM is used. The corresponding Bayesian networks are depicted in Fig. 40.

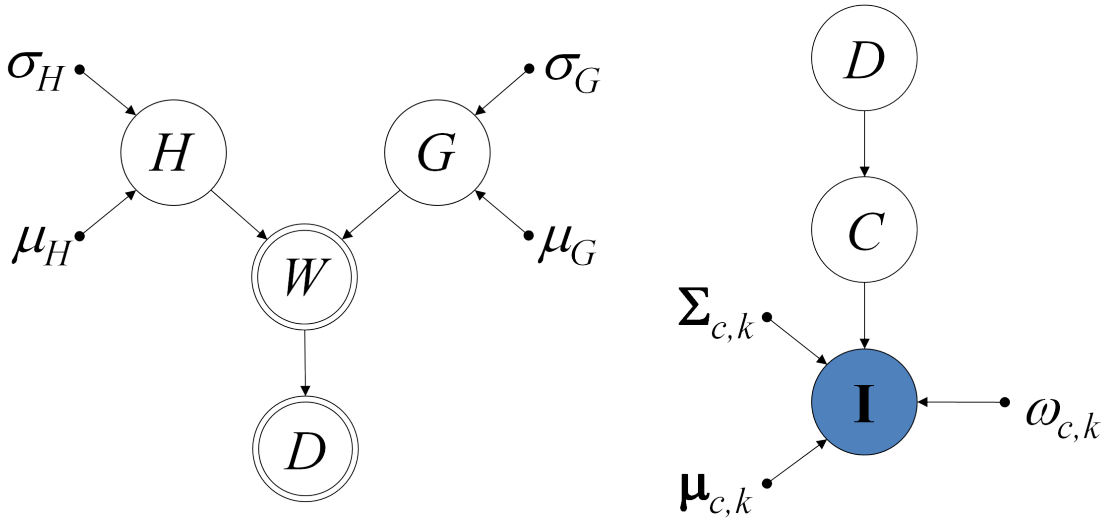

Figure 40. The fragments of the pixel-based model representing the simulation (left) and the multispectral classification (right)

The Bayesian network on the left with the DEM as input information can be considered as a flood simulation. On the right the Bayesian network for a multispectral classification is depicted.

For the river Elbe scenario three different elevation data sets with varying spatial resolution and different accuracy in vertical direction are used to demonstrate the influence on the results. The most accurate data set is a DEM generated from laser scan data distributed by the Federal Agency for Cartography and Geodesy Sachsen-Anhalt with a spatial resolution of 1 m with and a vertical accuracy of +/- 0.15 m. However, the Bayesian model assumes that the height information corresponds to a normal distribution. Therefore it is assumed that 99 percent of the points are within the range +/- 0.15 m, which corresponds to a standard deviation of about 0.06 m. The second elevation information is taken directly from the 3D Intermap road database with a standard deviation of 1 m of the absolute vertical accuracy. The standard deviation of the relative vertical accuracy which is the accuracy between two neighboring points is 0.3 m. The third dataset is a DEM from SRTM with a spatial resolution of 90 m and a relative 90 percent height error of 10 m (Rodriguez et al., 2006), which corresponds to a standard deviation of about 6 m. The ROC curves of the simulation using the different height information are depicted in Fig. 41.

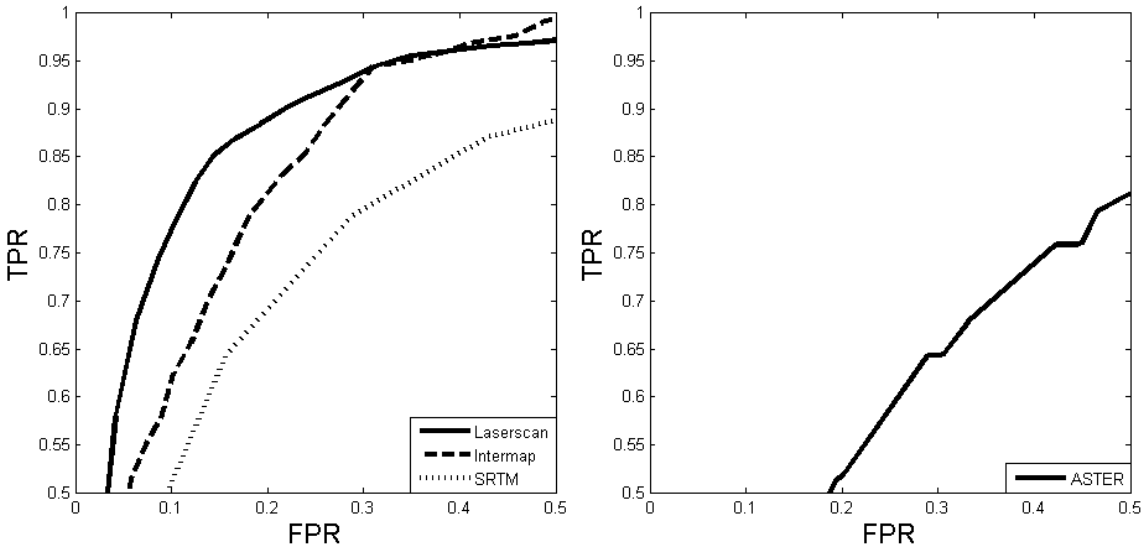

Figure 41. The ROC curves of the flood simulation using different elevation data. (left: Elbe scenario, right: Chobe scenario)

For the river Chobe scenario only the ASTER DEM is available with a spatial resolution of 15 m. The ASTER DEM is generated using several stereo pairs. The vertical accuracy of the ASTER DEM depends on the number of used stereo pairs. Beside the elevation data an additional file is delivered with the number of images used to estimate the height value of each pixel. The relation between the standard deviation of a pixel and the number of used images n for the estimation of the height value is investigated by Jacobsen (2010) which results in the equation

$$\sigma_{H_{ASTER}} = 12.43m - n \cdot 0.35m. \tag{173}$$

In the river Chobe scenario the number of used images n varies from 7 to 18, which leads to standard deviations ranging from 9.98 m to 6.13 m. The ROC curve of the simulation in the river Chobe scenario using the ASTER DEM is shown in Fig. 41. The ROC curves shows that the results of the simulation increase the higher the accuracy of the dataset. In general,

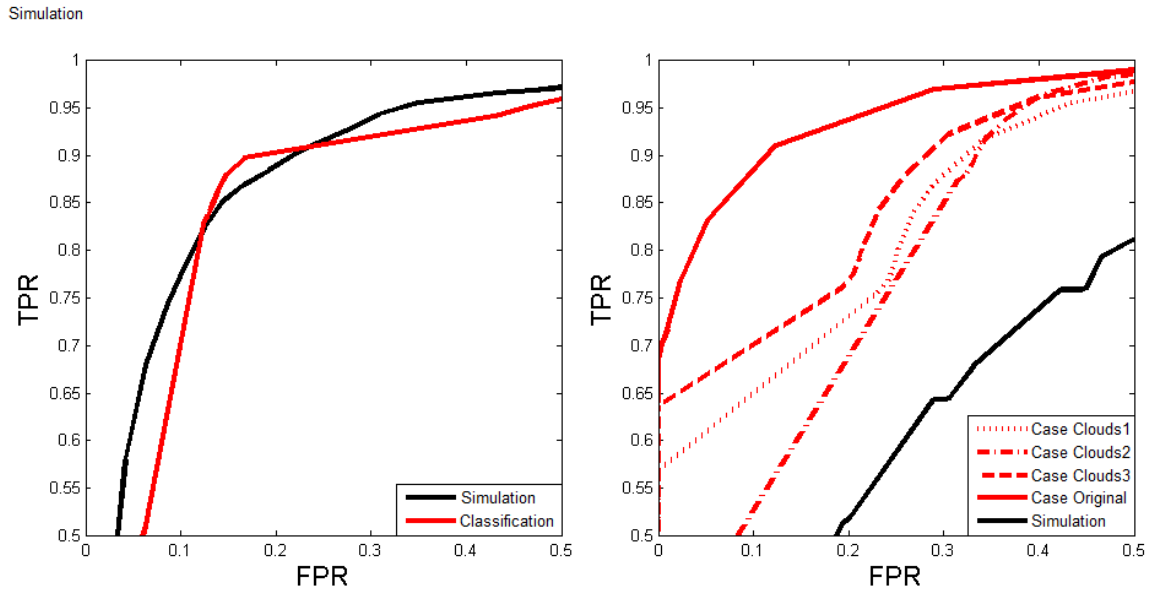

Figure 42. The ROC curves for a multispectral classification (red) compared with the simulation (black). (left: Elbe scenario, right: all cases of Chobe scenario)

it can be stated the closer the ROC curve is to the upper left corner, the higher is the overall accuracy.

The simulation is compared with a classification represented by the Bayesian network depicted on the right of Fig. 40. The estimation of the damage state D is based on a multispectral classification using solely the image information. In Fig. 42 the ROC curves of the classification (red) and the ROC curves obtained from the simulation (black) are compared. In the Chobe

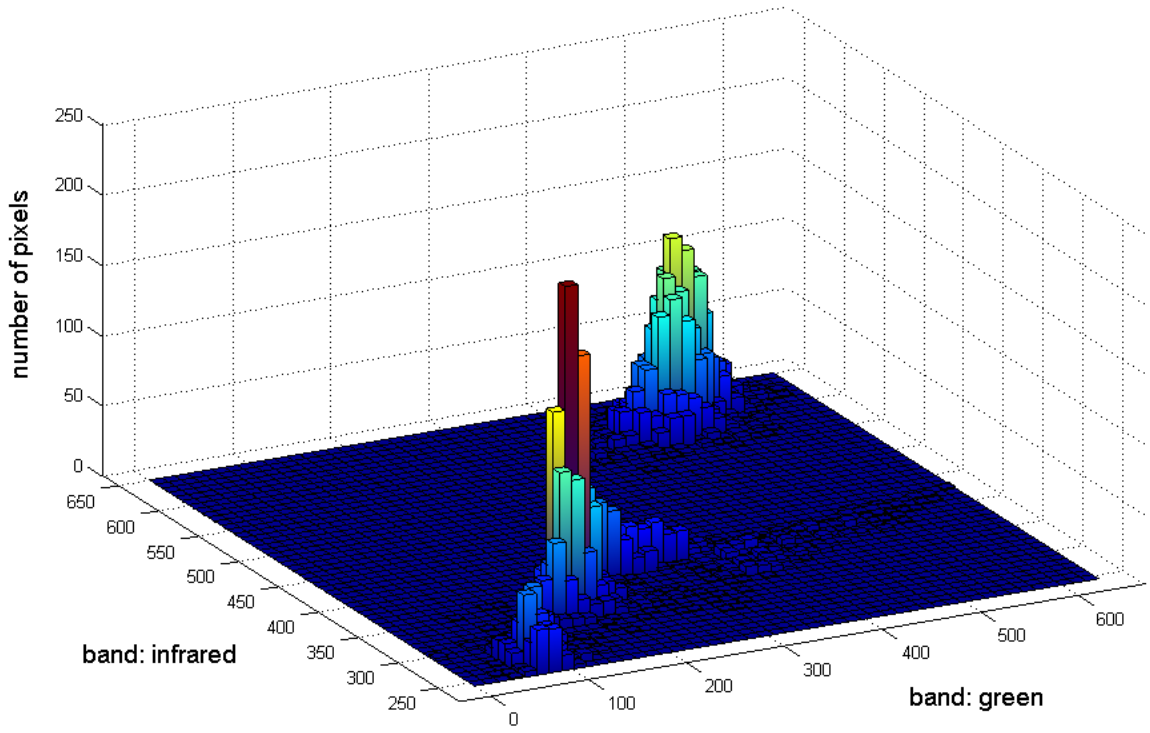

Figure 43. The scatter plot of five different road subclasses in the green and infrared band

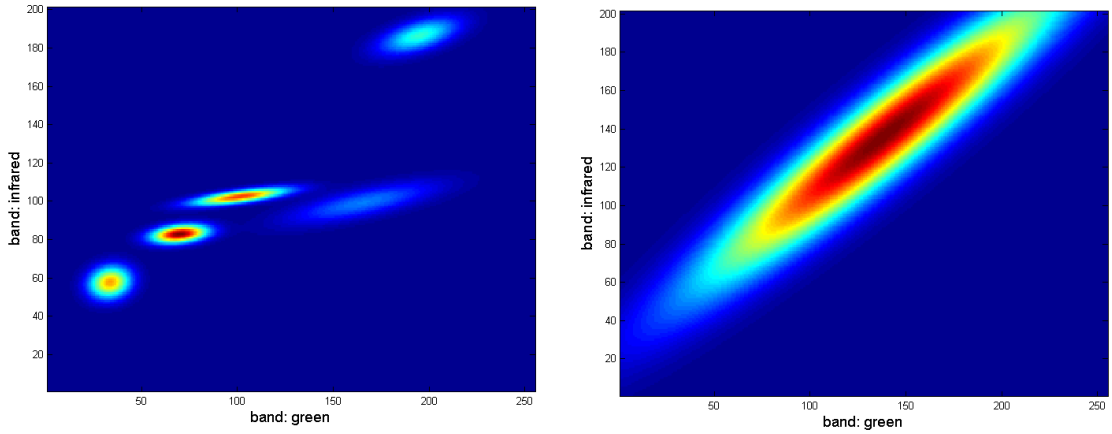

Figure 44. The conditional probability $p(\mathbf{I}|C = r)$ modeled as a Gaussian mixture model (left) and modeled as one Gaussian function (right) shown in the 2-dimensional case using the infrared and green band.

scenario the results of the classification is shown for the original case and for the different cases with varying artificial cloud coverage as depicted in Fig. 37. The ROC curves obtained from the different cases and the original river Chobe case show a large difference in accuracy. Of course, the ROC curve of the original case without cloud coverage shows the most accurate results. The accuracy of the other cases depends on the coverage of roads by the artificial clouds. However, in all cases the classification result is by far more accurate as the simulation using the ASTER DEM. In contrast, in the Elbe scenario the simulation and the classification leads to more equivalent ROC curves due to the high accuracy of the laser scan information (see. Fig. 42).

The multispectral classification is a supervised classification approach. The parameters $\boldsymbol{\Sigma}_{c,k}, \boldsymbol{\mu}_{c,k}, \omega_{c,k}$ have to be learned by training data for each class C. The radiometric characteristics of the broad classes road, water and vegetation are usually not Gaussian distributed if different subclasses such as asphalt road, concrete road and gravel road are combined in one broad class road (see. Fig. 21 and 22). Therefore Gaussian mixtures are used for the multispectral classification as defined in Equations 90 to 92. In Fig. 43 a scatter plot for five different road subclasses (motorway, country road, two different roads in the city and gravel road) is shown for the Elbe scenario. Exemplarily, the green and the infrared band are used for

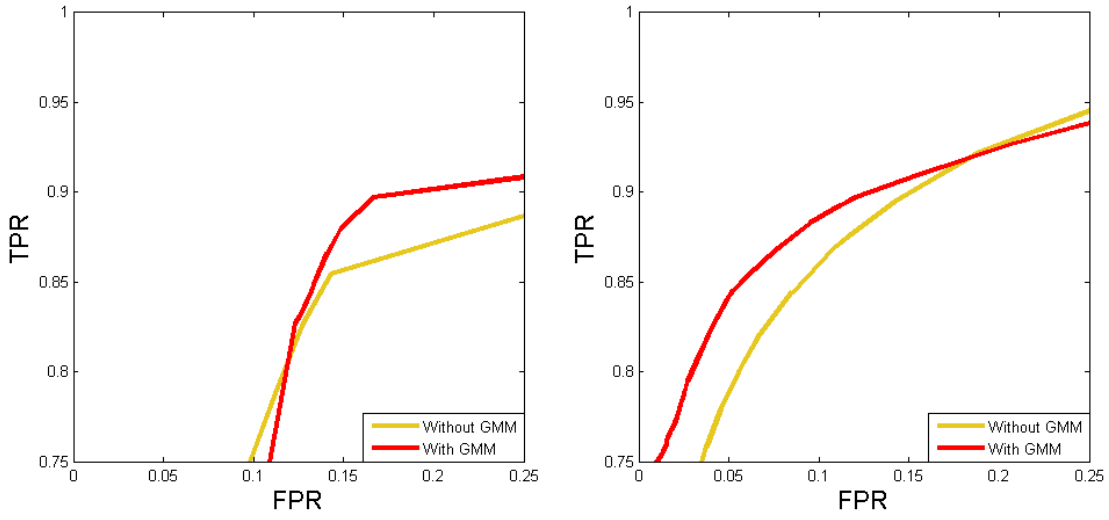

Figure 45. The ROC curves comparing the multispectral classification using Gaussian mixture models (red) and one Gaussian distribution (yellow)

the scatterplot with the original intensity values of the 16 bit image. In Fig. 44 the class road is modeled on one hand by a Gaussian mixture model (left) and on the other hand by only one multivariate Gaussian probability density function. Again only the two bands green and infrared are visualized. In case of Gaussian mixture models it is more important to cover all appearing road subclasses in the imagery. If all road subclasses are used as training data the Gaussian mixture model delivers a more detailed description of the radiometric appearances of the roads in the imagery than a single Gaussian probability density function. A more detail discussion of the modeling of the roads by Gaussian mixture models is given in Butenuth et al. (2011).

In both scenarios the benefit of modeling the radiometric characteristics by Gaussian mixture models can be seen in the resulting ROC curves depicted in Fig. 45. The left image shows the comparison of the ROC curves for the river Elbe scenario and on the right the different ROC curves are plotted for the original case in the river Chobe scenario. In both cases the use of Gaussian mixture models delivers ROC curves closer to the upper left corner, which is equivalent with a higher overall accuracy. In contrast to the previous ROC plots the plot in Fig. 45 shows only the detail between 1 to 0.75 for the TPR and 0 to 0.25 for the FPR. In the following these more detailed visualization of the ROC curves are used to demonstrate smaller differences in ROC curves.

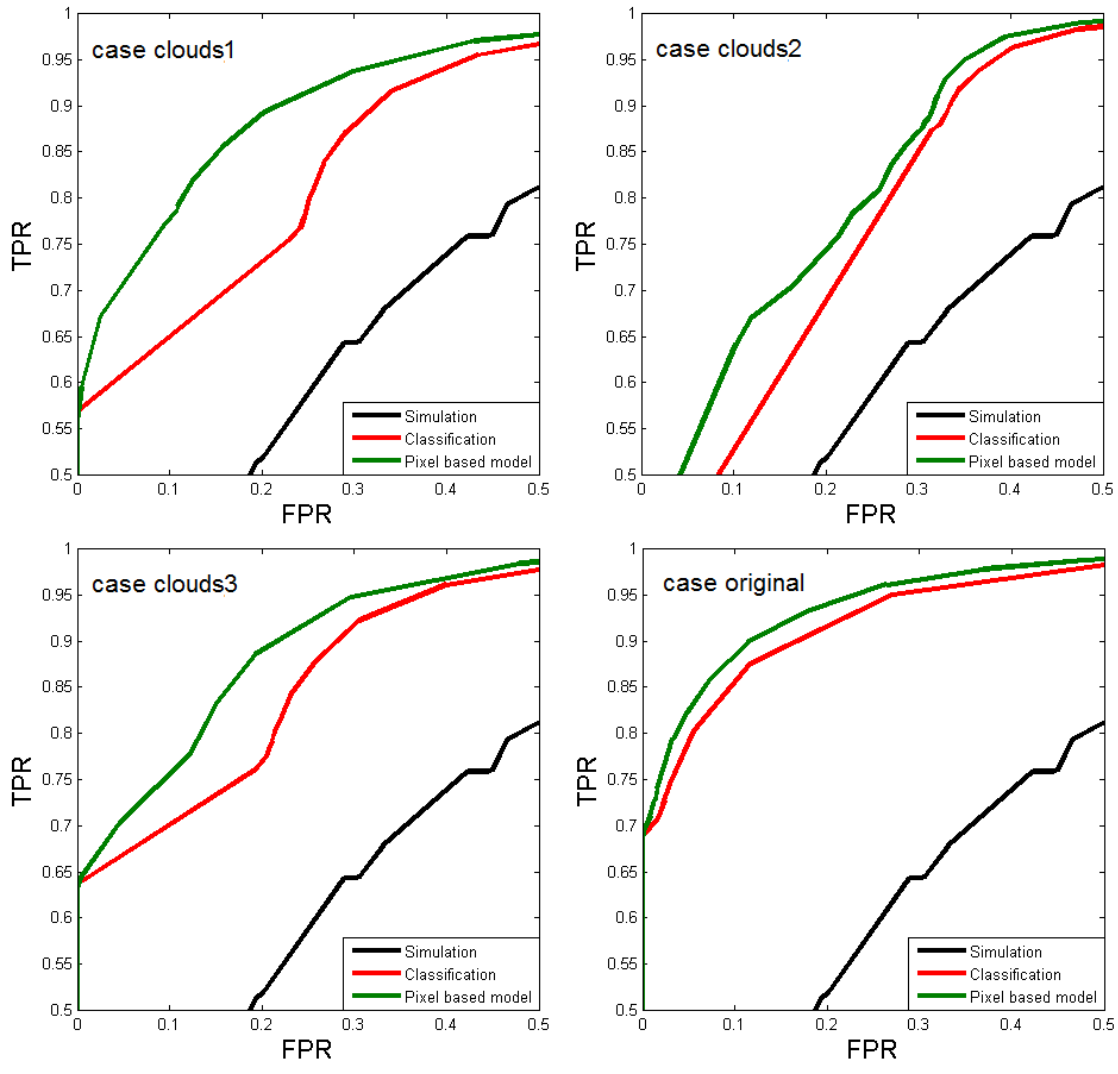

Figure 46. Comparison of the ROC curves obtained by the pixel-based model (blue), the classification (red) and the simulation (black) for the four different cases of the river Chobe scenario.

5.6 Evaluation of the presented models

5.6.1 Pixel-based model

The combination of the classification and simulation described by the Bayesian networks in Fig. 40 is the pixel-based model discussed in detail in Section 4.1. In both scenarios the combination of elevation data and image data by means of the Bayesian network depicted in Fig. 20 leads to more accurate results as the use of each dataset individually. In Fig. 46 the ROC curves for the river Chobe scenario obtained by the pixel-based model (blue), the classification (red) and the simulation (black) are compared for the three cases with artificial cloud coverage and the original case. In all four cases the ROC curve of the pixel-based model shows a higher accuracy than the classification and simulation. In particular, the benefit is obvious in the cases with additional artificial cloud coverage. The estimation of the flood state under the artificial clouds is therefore totally based on the elevation information since no image information concerning the flood state is available and neighborhood relations are not considered in the pixel-based model. But also in the original case without artificial cloud coverage the benefit of the data fusion can be observed. In particular, road pixels with very high or low elevation values with uncertain classification results can be assigned to the correct flood state by means of the combination of the data.

The results obtained for the river Elbe scenario behave similar as in the river Chobe scenario. The combination of classification and simulations leads again to ROC curves describing results with higher accuracy. The ROC curves are depicted in Fig. 47. Two different elevation data sets are used for the pixel-based model. On one hand the laser scan data (solid lines) and on the other hand the worldwide available SRTM DEM is used (dotted lines). In both cases the combination leads to better results.

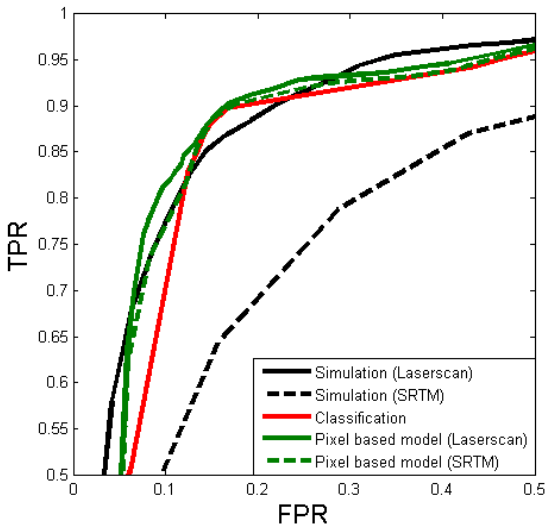

Figure 47. Comparison of the ROC curves obtained by the pixel-based model (blue), the classification (red) and the simulation (black) for the river Elbe scenario. The dotted lines shows the ROC curves obtained by the simulation with SRTM data or the pixel-based model using the SRTM data as elevation information.

In both scenarios it is shown that the fusion of the imagery and the DEM by means of the Bayesian network leads to better results, even if the elevation information has low accuracy as the DEM from ASTER or SRTM. Therefore it can be stated that the applicability of the pixel-based model is ensured as soon as any image information of the affected flooded area is available since elevation data with low accuracy as the DEM from the SRTM is worldwide available. The applicability of the pixel-based model will even increase when a more accurate

global DEM derived from the TanDEM-X mission is available.

5.6.2 Topology-based models

In the following the effects of modeling the neighborhood relations between pixels are evaluated. Several different topology based models discussed in Section 4.2 are compared. Firstly, the 'topology-based model G' depicted in Fig. 26 which models the neighborhood relations by means of the probability of water gauge changes between neighboring pixels. The neighborhood relation is expressed in the factor graph via the links between the random variables G. This 'topology-based model G' is the only topology-based model in which the neighborhood relation is modeled by a factor describing a physical behavior.

Secondly, the 'topology-based model C' depicted in Fig. 27 and the 'topology-based model D' is evaluated. In addition, the expanded models are evaluated which consider pixels perpendicular to the road center axis. These models are denoted in the following as 'topology-based model G+N', 'topology-based model C+N' and 'topology-based model D+N'. The factor graph of the 'topology-based model C+N' is shown in Fig. 28.

In Fig. 48 the three topology-based models are compared with the pixel-based model for all

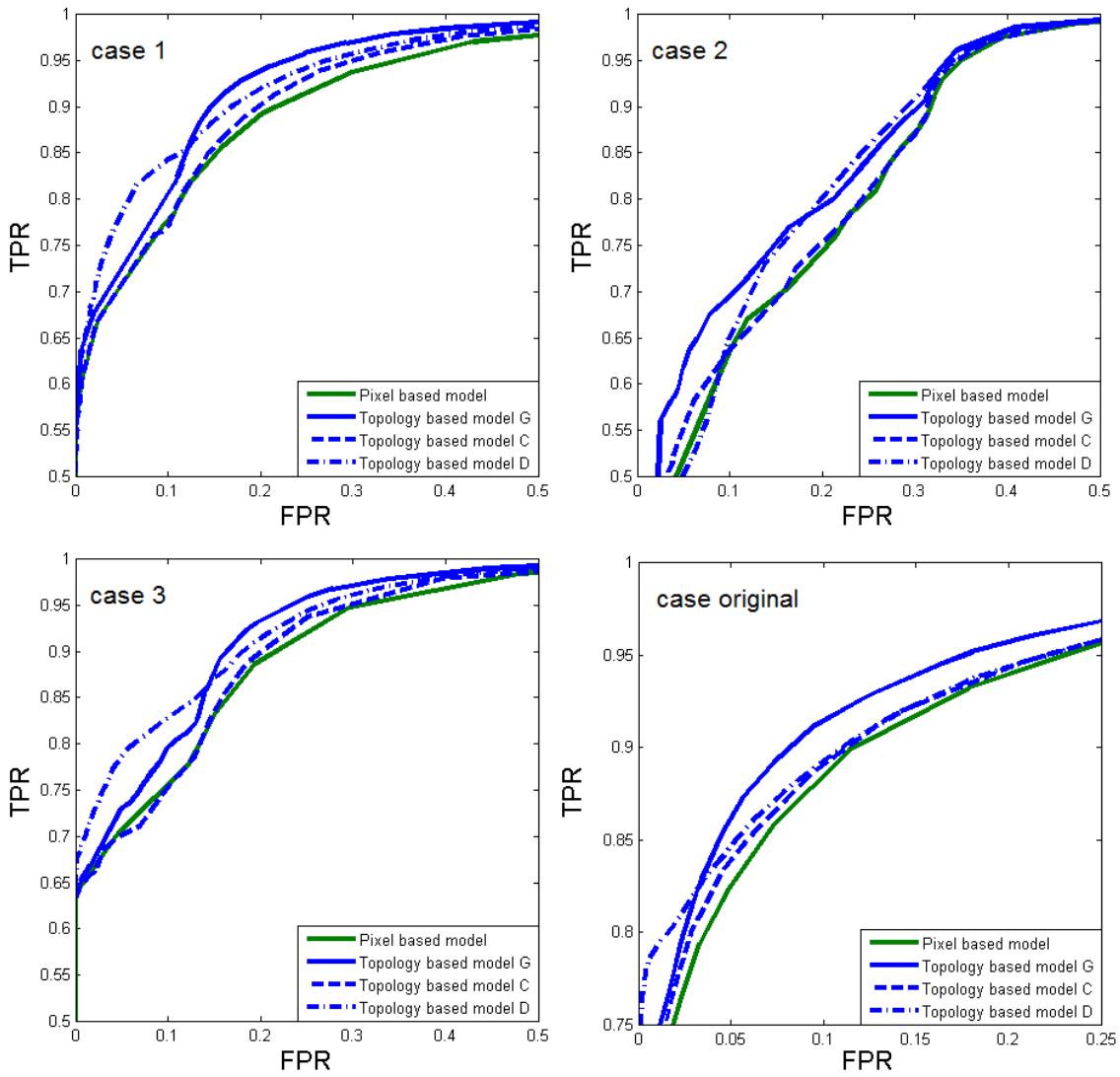

Figure 48. Comparison of the topology-based models differing in the links between neighboring pixels for the river Chobe scenario

the cases of the river Chobe scenario. In all cases with artificial cloud coverage as well as in the original case all the topology-based models (blue) lead to better results as the pixel-based model (green). In all topology-based models the smoothness of neighboring pixels is modeled with different kinds of factors which lead to the reduction of noise existing in the observations. In all the cases the 'topology-based model G' and the 'topology-based model D' show a better performance as the 'topology-based model C'.

A similar result is obtained for the river Elbe scenario (see Fig. 49). Again, the ROC curves of the three topology-based models (blue) are compared to the pixel-based model (green). The 'topology-based model G' and the 'topology-based model D' lead to better ROC curves than 'topology-based model C' and the pixel-based model. A problem of the 'topology-based model C' is the allocation of the conditional probability table shown in Tab. 2. Only one probability p_{equal} is used to describe all the changes between all kinds of existing classes. The features of the individual classes are not considered. A detailed factor $f_{C_jC_{j+1}}$ should consist of five different probabilities describing the probability of no change for every specific class. An even better description would be the determination of all the 25 probabilities describing all kinds of different class changes. However, these probabilities can vary from scenario to scenario and are therefore very difficult to learn. In contrast, in the 'topology-based model D' only the probabilities that the random variable D changes the state from flooded to not flooded and vice versa have to be defined. In case of the 'topology-based model G' the variation of the water gauge is only one physical quantity, which is in general easier to quantify for varying scenarios. Therefore, the 'topology-based model G' and the 'topology-based model D' are the models with higher potential concerning the transferability to different scenarios.

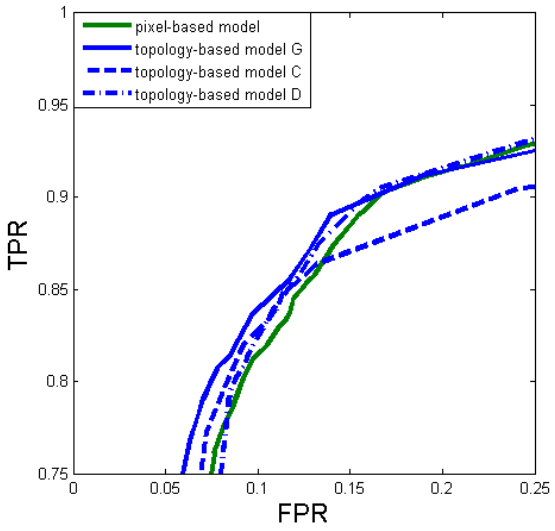

Figure 49. Comparison of the different topology-based models (blue) with the pixel-based model (green) for the river Elbe scenario

In Fig. 50 the factors describing the neighborhood of pixels are evaluated. In the 'topology-based model G' the factor $f_{G_jG_{j+1}}$ describes the probability that the water gauges changes between two neighboring pixels. The factor is modeled as a normal distributions as stated in Equation 135 with the standard deviation σ^2_{GG}. The σ^2_{GG} can be interpreted as a smoothing factor, which describes the smoothness of the water surface. The lower the standard deviation σ^2_{GG} the higher the smoothness. In the extreme case, if $\sigma^2_{GG} \to 0$ the factor $f_{G_jG_{j+1}}$ results into Dirac delta functions since

$$\delta(x) = \lim_{\sigma \to 0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right). \tag{174}$$

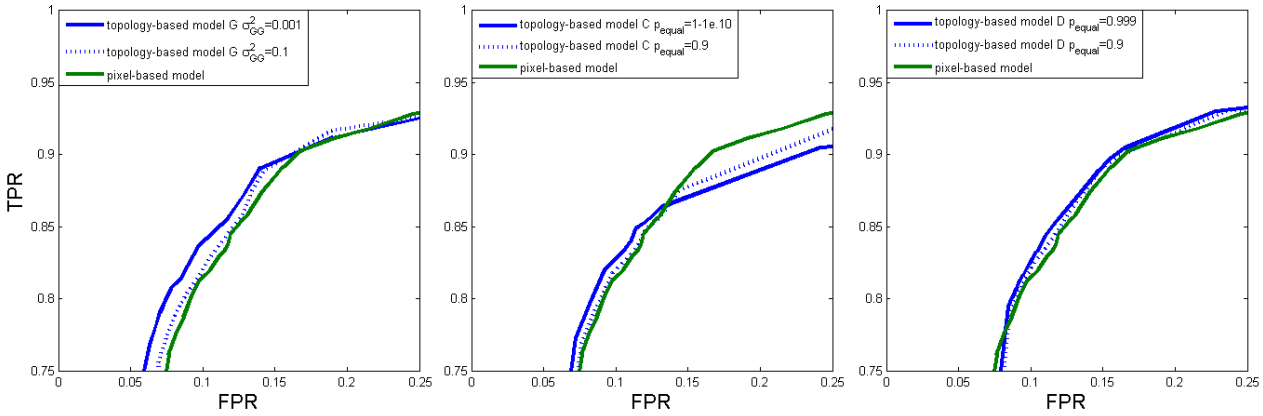

Figure 50. The ROC curves of the 'topology-based model G' (left), 'topology-based model C' (middle) and the 'topology-based model D' (right) with varying parameters defining the factors $f_{G_jG_{j+1}}$, $f_{C_jC_{j+1}}$ and $f_{D_jD_{j+1}}$.

Replacing the normal distribution describing the factor $f_{G_jG_{j+1}}$ stated in Equation 135 by the delta functions and applying the defining property of the delta function in Equation 109 leads to the simplification of the messages

$$\mu_{f_{G_jG_{j+1}}\to G_{j+1}}(G_{j+1}) = \mu_{G_i\to f_{G_jG_{j+1}}}(G_j). \tag{175}$$

The effect of σ^2_{GG} is shown in Fig. 51 depicting the messages $\mu_{G_j\to f_{G_jG_{j+1}}}(G_j)$ for one road axis. The lower the parameter σ^2_{GG} the more similar are the neighboring distributions describing the water gauge of a road pixel. The corresponding ROC curves are shown in Fig. 50 on the left. If $\sigma^2_{GG}\to\infty$ the Gaussian distributions describing the factor $f_{G_jG_{j+1}}$ become uniform distributions which leads to the pixel-based model. The ROC curves with different smoothing

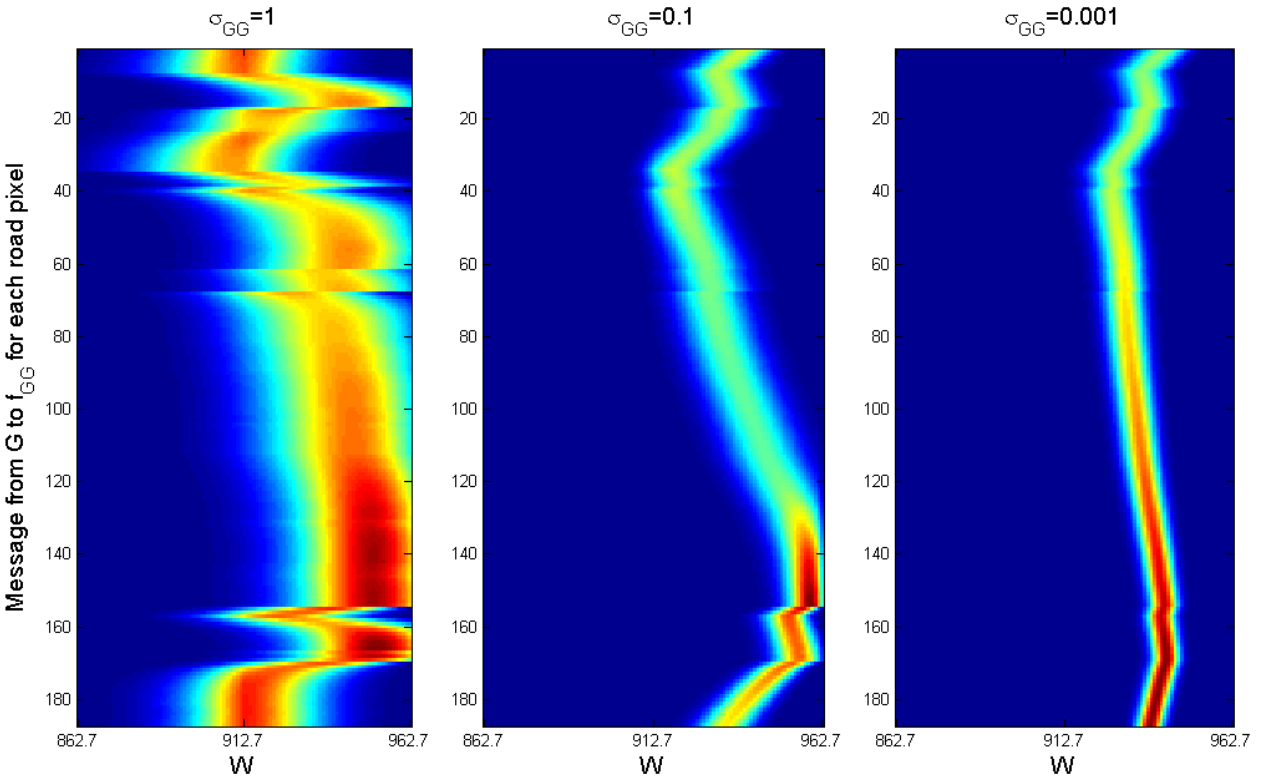

Figure 51. The messages $\mu_{G_j\to f_{G_jG_{j+1}}}(G_j)$ shown in Equation 142 and 145 for different σ^2_{GG} describing the smoothness of the water gauge between neighboring road pixels.

properties for the 'topology-based model C' and the 'topology-based model D' are also shown in the middle and on the right of Fig. 50. In the 'topology-based model C' the factor $f_{C_jC_{j+1}}$ is defined by p_{equal} describing the probability that neighboring pixels belong to the same class C. The lower the probability p_{equal} the more the ROC curve of the 'topology-based model C' converges to the pixel-based model. If $p_{equal} = 0.2$ the 'topology-based model C' is equal to the pixel-based model. The effects are in all three models the same. If the smoothing is very low the ROC curves of the topology-based models is close to the pixel-based model.

An example of the influence of the smoothing on the results assessing a road segment is shown in Fig. 52. The result of the pixel-based model leads to some erroneous assignment to flooded pixels due to shadow areas. The 'topology-based model G' with $\sigma^2_{GG} = 0.1$ leads to better results smoothing out the most erroneous pixel except one segment of consecutive erroneous pixel. The 'topology-based model G' with a $\sigma^2_{GG} = 0.001$ smooths out all erroneous pixels which leads to the correct assignment of all pixels to not flooded (green). However, the smoothing property can also deteriorate the results if only few pixels in the road center axis shows evidence for the correct assignment.

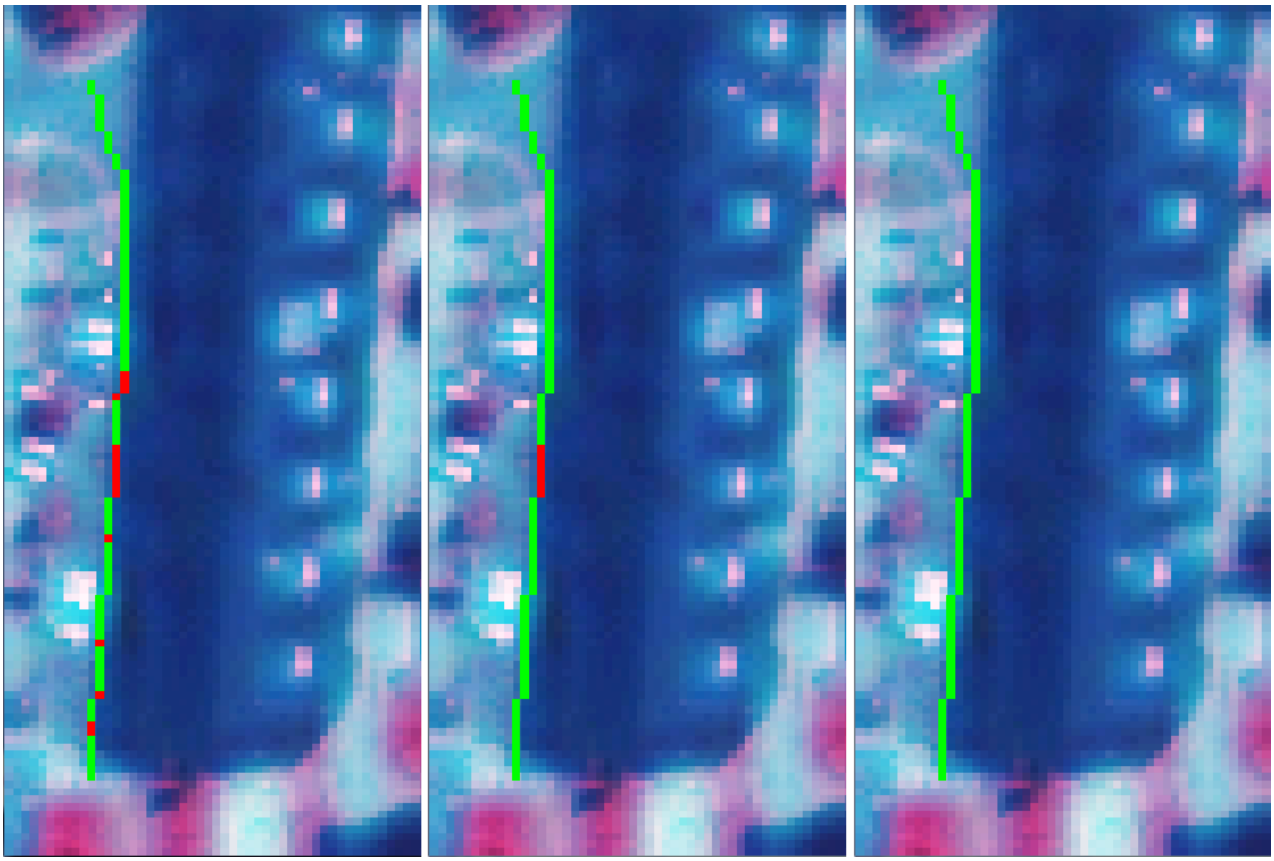

Figure 52. Example which demonstrates the smoothing property of the 'topology-based model G' depending on σ^2_{GG}. Some pixels are assigned to flooded (red) due to shadow but are in reality not flooded. Left: pixel-based model, Middle: 'topology-based model G' with $\sigma^2_{GG} = 0.1$, Right: 'topology-based model G' with $\sigma^2_{GG} = 0.001$.

The expansion of the topology-based models incorporating the pixel information perpendicular to the road center axis is evaluated in Fig. 53 for the river Elbe scenario. Only for the 'topology-based model D' a significant improvement of the ROC curves can be observed. In case of the 'topology-based model G' and 'topology-based model C' the incorporation of neighboring pixels perpendicular to the road axis leads to no improvements.

An important property of an automatic damage assessment system is the processing time in order to guarantee a fast response. Therefore the processing time of the different developed models is investigated. In Tab. 5 the processing time of the different presented models are shown. For the investigation an Intel(R) Core(TM) i7 with a CPU of 3.40 GHz was used. The

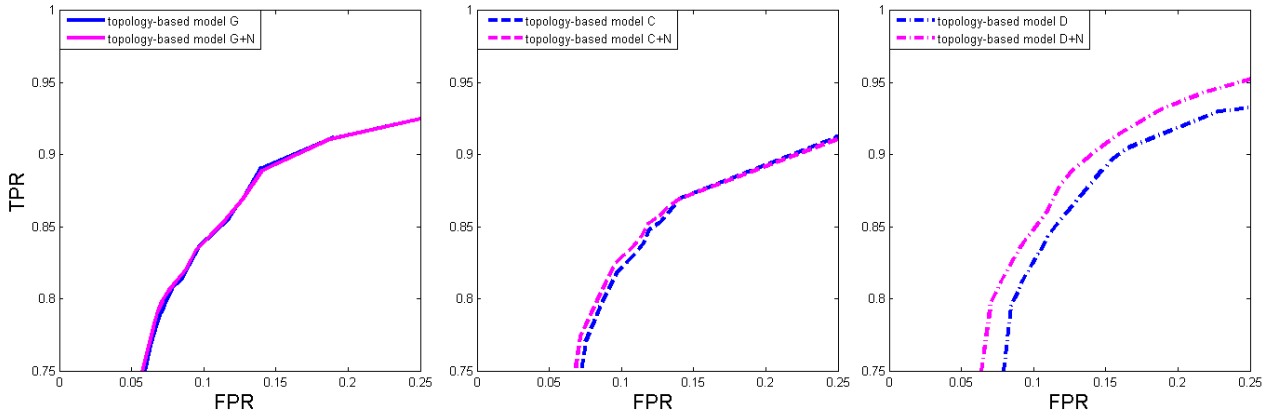

Figure 53. Comparison of the ROC curves of obtained from the 'topology-based model G+N' (left), 'topology-based model C+N' (middle), 'topology-based model D+N' (right) with the topology based models neglecting the neighborhood relations perpendicular to the road center axis.

crucial factors for the processing time are $f_{G_jG_{j+1}}$, $f_{C_jC_{j+1}}$ and $f_{D_jD_{j+1}}$. Therefore the three topology-based models and their expansions including neighboring pixels perpendicular to the road axis are investigated.

Table 5. Processing time of the different topology-based models for both scenarios (Intel(R) Core(TM) i7 CPU: 3.40 GHz)

	$G+N$	G	$C+N$	C	$D+N$	D
Elbe	$9min\ 50sec$	$4min\ 8sec$	$40sec$	$20sec$	$40sec$	$19sec$
Chobe	$1min\ 52sec$	$57sec$	$10sec$	$4sec$	$10sec$	$4sec$

The most time-consuming models are the models with the links between the random variables G. Compared to the factors $f_{C_jC_{j+1}}$ and $f_{D_jD_{j+1}}$ which are modeled by small probability tables the factor $f_{G_jG_{j+1}}$ is a 2-dimensional function stated in Equation 135. In the implementation of the 'topology-based model G' and the 'topology-based model G+N' this function is discretized by a 100 times 100 matrix. Small steps of discretization are necessary to deal with the varying accuracies of the observations and therefore to avoid aliasing effects. The substitution of the 2-dimensional function stated in Equation 135 by a delta Dirac functions as described in Equations 174 would lead to simplified approximated model which reduces the processing time enormously. The parameterization of the messages between the random variable G_j and G_{j+1} is not possible since the outgoing messages from G_j to $f_{G_jG_{j+1}}$ cannot be described by a parameterized function as can be seen in Equation 142. The message $\mu_{G_j \to f_{G_jG_{j+1}}}(G_j)$ is not symmetrical and the skewness depends on Θ^*.

5.6.3 Multi-temporal based model

The evaluation of the multi-temporal models is restricted to the river Chobe scenario since the lack of ground truth in the river Elbe scenario. The best results are obtained if the hidden hypothesis variables are linked as in a hidden Markov model, which corresponds to the 'multi-temporal model D'. In Fig. 54 the multi-temporal model is compared with the topology-based model and the pixel-based model for all four cases for the river Chobe scenario.

In the original scenario the performance of the topology-based model and the multi-temporal model is similar because almost no occlusions, whether by vegetation or clouds exist at the time T. Therefore the additional information at the previous time $T-t$ from TSX image cannot contribute much information for the damage assessment. However, in all other scenarios with artificial cloud coverage the multi-temporal model outperforms the topology-based model.

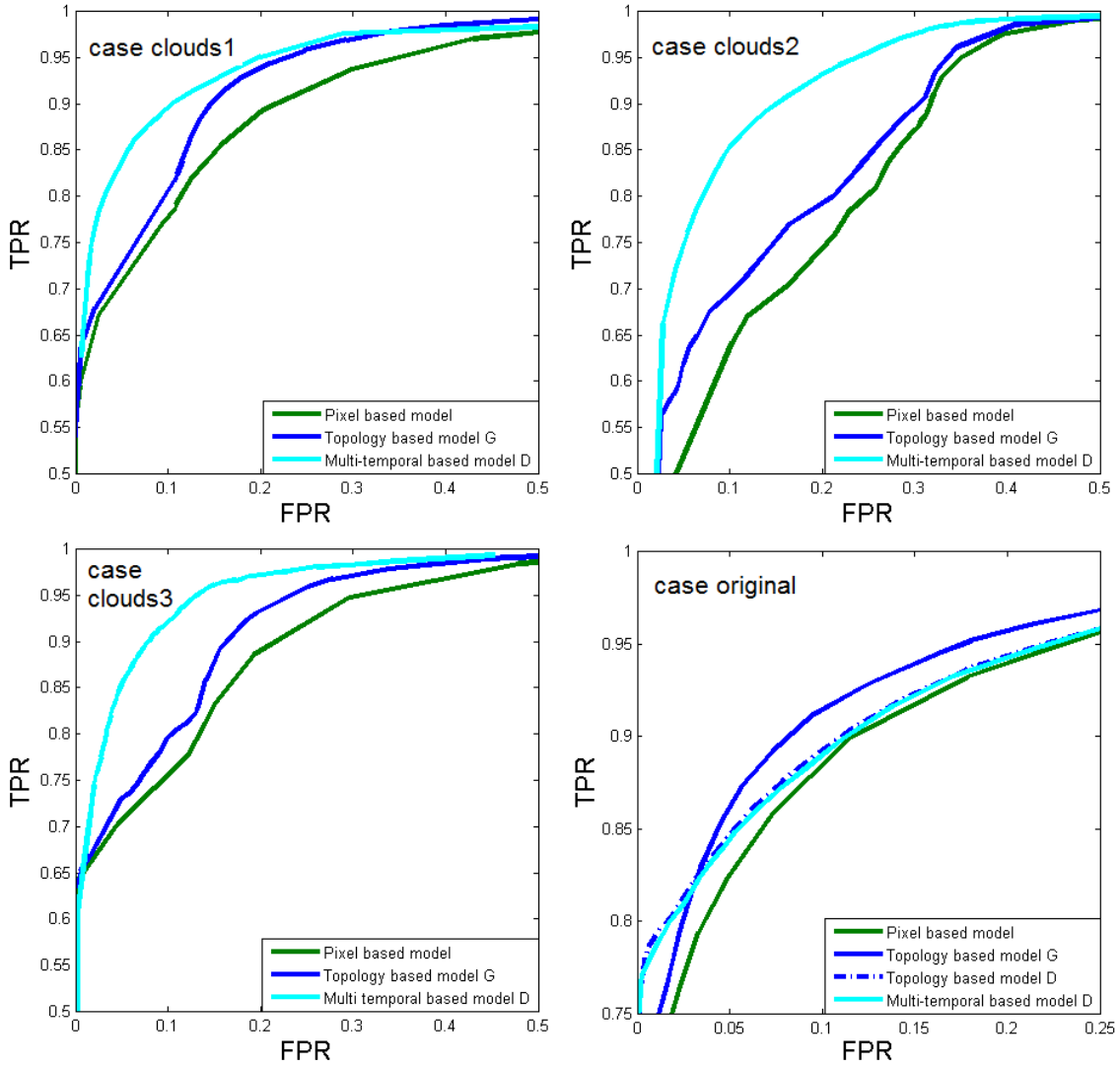

Figure 54. Comparison of the pixel-based, topology-based and mutli-temporal based model for the river Chobe scenario

5.7 Discussion of the Results

In this section results of both scenarios for the damage assessment are shown and discussed. In addition, the error sources which lead to wrong assignments are presented. The outcome of all presented probabilistic graphical models are probability distribution for the hypothesis variables D and W. In case of the hypothesis variable D, the probability distribution consists only of the two probabilities that a road pixel is flooded or not. The probability distribution of W is a continuous distribution with only one global maximum.

A possible representation of the obtained probabilities of D could be a continuous color coding. However, the interpretation of color coded probabilities is time consuming and not suitable for rapid decisions. But in case of emergency actions rapid decisions are crucial. Therefore the road pixels are assigned to three categories: flooded roads, possibly flooded roads and not flooded roads. In the following the three different categories are color coded. Flooded roads are red, not flooded roads are green and roads, which are not assigned to flooded or trafficable due to a marginal probability $p(D)$ less than a threshold p_t are yellow. The threshold describes the balance between wrong assignments and no assignments and was manually chosen. In future work, the estimation of an optimal threshold should be considered.

Figure 55. Result of 'topology-based model G' for river Elbe scenario. (Green: not flooded, Yellow: possibly flooded, Red: flooded)

In Fig. 55 the assigned road pixels to the three categories are shown for the river Elbe scenario. The result is obtained by the 'topology-based-mode G'. A graphical assessment of the results is shown in Fig. 56, which shows the correct assignments in green, the roads assigned to possible flooded in yellow and the wrong assignments in red.

In the following several effects in the river Elbe scenario are discussed, which lead to wrong assignments. The most severe problems can be observed in the urban region on the left of the scene. The shadows lead to erroneous classification since the reflectance of shadow is more similar to the radiometric characteristics of water than of one of the road subclasses used in the Gaussian mixture model. Furthermore, the altitude of the northern part of the urban area is low and was protected from the flooding by a dam (see Fig. 32). These two properties lead to several wrong assignments which are more often in the northern part.

Another reason for wrong assignments which is very case specific is illustrated in Fig. 57. On the right a detail of the IKONOS images is depicted showing a flooded road. However, the reflectance properties in the IKONOS image are similar to a trafficable road. Only by means

Figure 56. Assessment of the result shown in Fig. 55. (Green: correct assignment, Yellow: no assignment, Red: wrong assignment)

of the expert knowledge the reflectance patterns could be assigned to a temporary dam, which was built on top of the road. A photo of the same road as depicted in the IKONOS images is shown in Fig. 57 on the left side.

A last error source, which leads to wrong assignments is the used DEM in the river Elbe scene. The high resolution DEM maps always the points on the ground. Therefore, bridges are not mapped in the DEM but the ground points under the bridges. Of course this effect leads to an erroneous simulation part. Alternatively, the 3-dimensional road information from the Intermap road database can be used to avoid this effect. However, the accuracy of the Intermap data is less than the used DEM as discussed in Section 5.2.1.

For the river Chobe scenario in Fig. 58 a result is shown of the 'multi-temporal model D' for case clouds1. Again the roads are assigned into the three categories flooded (green), possibly flooded (yellow) and not flooded (red). In this case a large area is affected by artificial clouds

Figure 57. A temporary dam built on the road (left: in-situ photo, right: IKONOS imagery)

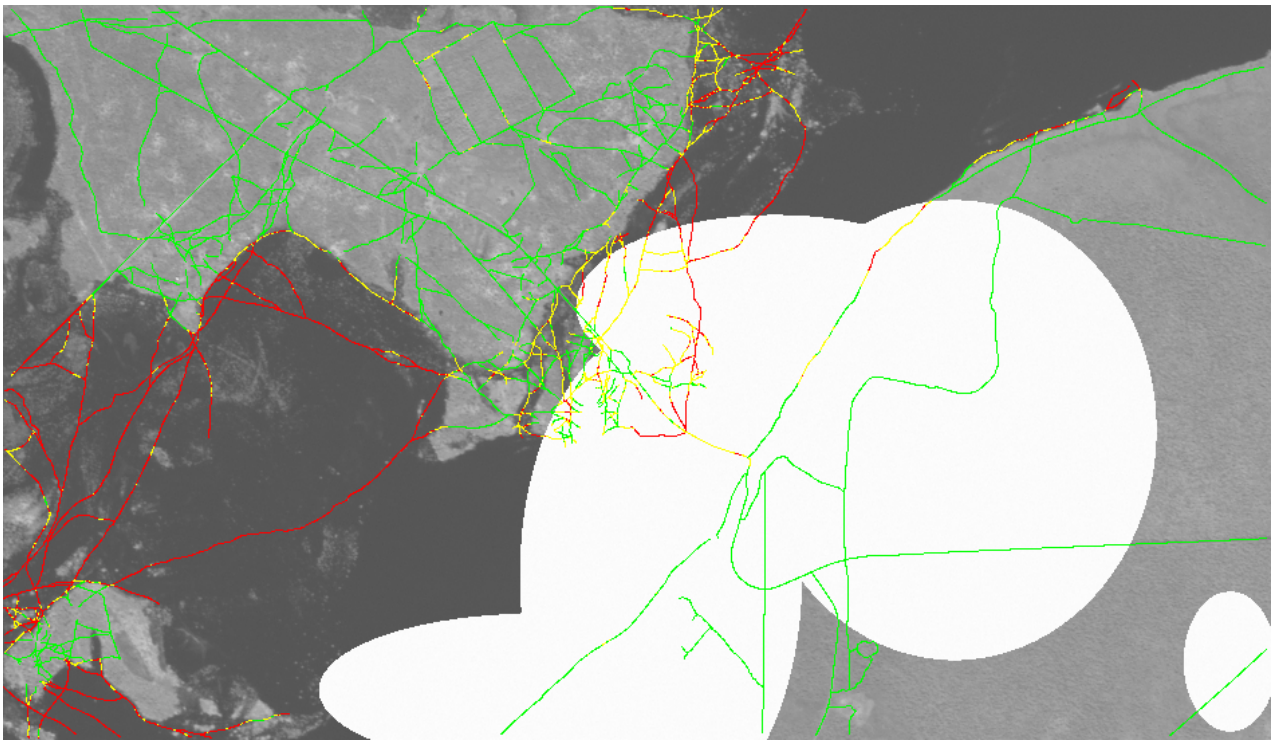

Figure 58. Result of 'multi-temporal model D' for river Chobe scenario for case clouds1. (Green: not flooded, Yellow: possibly flooded, Red: flooded)

which occlude flooded and trafficable roads. In Fig. 58 the positive effect of combining the different information sources as the DEM and the TSX imagery at the time $T - t$ can be seen. The roads on the bottom right occluded by artificial clouds are assigned to not flooded (green) mainly due to the evidence resulting from the DEM. In contrast, the assignments of the roads occluded by artificial clouds in the center of the imagery can be ascribed to the information from the TSX imagery at the previous point in time $T - t$.

Beside the hypothesis variable D also the state of the hypothesis variable W is estimated. The estimation of the water level is based on a simple maximum likelihood estimation considering the probability distribution of W. However, a quantified evaluation of the estimation is not possible due to the lack of reference data. Instead of the comparison with a reference only a visual interpretation of the results is possible. In Fig. 59 on the top the estimated water level is shown using only the DEM information. On the bottom the water level is shown using the 'topology-based model G'. Negative values for W can be interpreted as the vertical distance

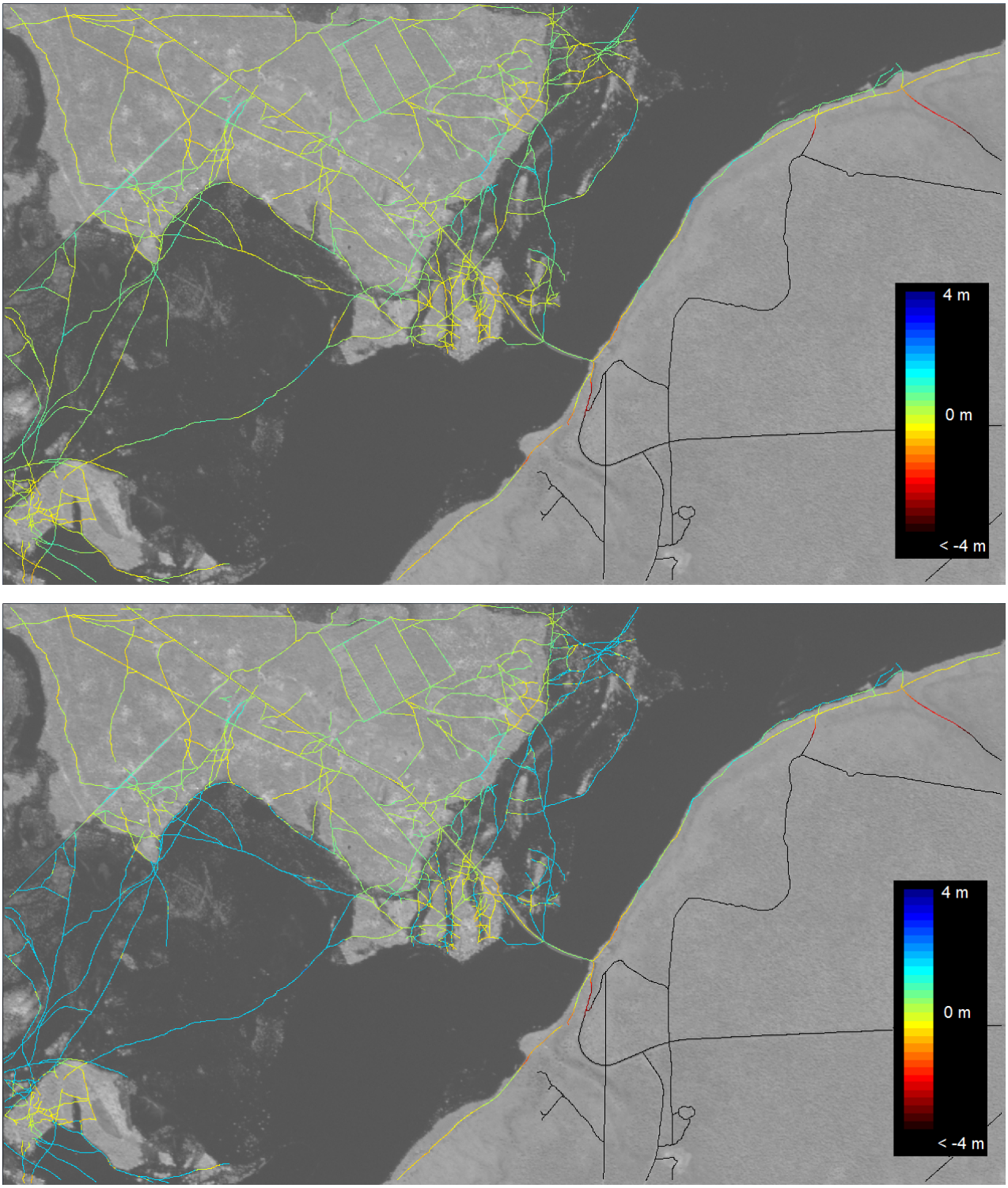

Figure 59. The estimated water level W (Top: only the ASTER DEM is used, Bottom: the 'topology-based model G' is used for the estimation

of the road pixel to the current water gauge. The comparison of the two results by visual inspection shows the more reasonable results in case of using the 'topology-based model G'.

In this work, models are developed to assess the roads concerning the damage state. Although the effects of decisions based on the results are not part of the thesis, the most important effects for wrong assignments are shortly discussed. If a road is assigned to flooded but in reality it is trafficable the relief workers have to make a detour. On the contrary, if a road is assigned to not flooded but it is in reality flooded the navigation would lead to a dead end. These two events reflect two entries in the confusion matrix (false positive and false negative). Costs can

be assigned to every entry in the confusion matrix and the minimization of the expected costs leads to an optimal decision. Graphical approaches which are able to model such costs are influence diagrams, which are a generalization of Bayesian networks.

6 Discussion and outlook

In this chapter the developed probabilistic graphical models to assess the damage of linear infrastructure objects after flooding as well as the general capability of the developed approach are discussed. The discussion focuses on whether the goals stated in Section 1.2 are achieved. Furthermore, problems of the presented models are emphasized and possible future research is pointed out, which focus on the expansion of the developed models as well as new strategies to assess infrastructure objects after flooding.

6.1 Discussion

The main goal of the thesis was the development of a damage assessment system of linear infrastructure objects during flooding. In this chapter, it is discussed if the requirements of a damage assessment system stated in 1.2 are fulfilled in order to guarantee a rapid and reliable emergency response. In the following each stated goal is repeated and comments are added.

◇ *The handling of different kinds of image data (optical and radar) with varying spectral and spatial resolution.*

The developed models enable the utilization of every kind of image information. The general framework of probabilistic graphical models allows the embedding of optical and radar imagery. Depending on the image data the Bayesian network can be adapted to enable different approaches for the classification part. In this thesis, Gaussian mixture models are used, which deliver good results. However, in case of radar imagery other classification approaches considering the specific properties of the radar images are better suited than Gaussian mixture models. If the classification approach is modeled by means of probabilistic graphical models as in the case of Martinis et al. (2011) there is no obstacle to substitute the proposed classification by a model which is designed especially for radar imagery.

In addition, it is shown that the model is able to deal with imagery with varying spectral and spatial resolution. The only constraint for the spatial resolution is the visibility of the infrastructure objects. The factors $f_{G_jG_{j+1}}$, $f_{C_jC_j+1}$ and $f_{D_iD_{j+1}}$ have to be adapted corresponding to the spatial resolution.

◇ *The integration of additional information such as DEM, data from GIS and in-situ measurements.*

The presented model combines the image information with the DEM, GIS data and in-situ measurements. It has been shown that every kind of DEM can improve the damage assessment. Even DEMs from SRTM or ASTER with very coarse spatial resolution can contribute to the improvement of the results. The integration of information about the water gauge can be integrated by in-situ measurements or, if not available, methods can be used to derive the water gauge directly from the used imagery as described in Section 3.2.1. From a GIS the location of the roads is obtained. However, beside the location of the roads a lot of more information is available in road databases, which are currently not used. A discussion about the integration of further road information is given in the Section 6.2.

◇ *The combination of the available data using a consistent statistical framework in order to quantify the uncertainties of the results.*

The framework of probabilistic graphical models is used to model the dependencies of the input information and the hypothesis variables. The combination is carried out by modeling the causal relations between the input data using Bayesian networks. Markov chains are used to model the topology or the road network. Both models are transformed into factor graphs building the

basis for exact inference. However, the dependencies between the random variables modeled in Bayesian networks and Markov networks are different. In this work this issue is circumvented by the transformation of Bayesian Networks into Markov networks via moralization which leads to the good results as presented in this thesis. However, one probabilistic model which is able to model the causal relation and the symmetrical relations simultaneously would be desirable. A possible solution is given in Section 6.2.

◇ *The handling of sequential data at different times in order to monitor the development of flooding.*

The proposed multi-temporal models enable the incorporation of imagery at previous points in time. In Chapter 5.6.3 it has been shown that the information from previous points in time can significantly improve the results of the assessment of the current point in time. The development of flooded infrastructure objects can be obtained by simple change detection approaches applying on the results of different points in time.

◇ *The processing time should be as fast as possible to guarantee a rapid response.*

The performance of the processing time is demonstrated in Tab. 5. The processing time of the 'topology-based model C' and the 'topology-based model D' are in both scenarios less than 1 minute and therefore the requirements are clearly fulfilled. In case of the 'topology-based model G' the computation is more demanding, but still acceptable for a rapid respond. The proposed approximation in Equation 175 is a further possibility to decrease the processing time. However, it is important to mention that the system is a supervised approach which includes the determination of training data. This time has to be added to the processing time.

◇ *The transferability of the system to different scenarios.*

Two different test scenarios are used to evaluate the performance of the presented models. Both scenarios consist of different image information sources varying in spectral resolution, spatial resolution and sensors. The used road information is obtained from different sources and the different DEM are used and evaluated even within one scenario. The effects of the different information sources are evaluated. Furthermore the classification is based on a supervised approach with the disadvantage that the system is not fully automatically. However, the supervised approach is one crucial factor to guarantee that the transferability of the system to further scenarios is given. Therefore the requirement of the transferability to different scenarios is fulfilled.

6.2 Outlook

The outlook is divided into two parts. In the first part the expansion of the developed model is discussed integrating additional information. The second part focuses on further promising methodical approaches.

The presented probabilistic graphical model combines crucial information to assess linear infrastructural objects after flooding. The model combines imagery, DEM, in-situ measurements and road information. This information builds the basis for the assessment. In this work it was focused on information, which is worldwide available within a short time frame. However, the framework of Bayesian networks allows the adaptation of the model without destroying the basic structure if additional information is available. In the following several examples are listed, which could contribute to an improved assessment of infrastructure objects.

In this work the road databases are only used to obtain the information about the spatial distribution of the road network and no attribute information is used. However, many road databases deliver additional attribute information as the road category, the width of the road

or information about the spatial accuracy. The road category or the surface property can be used to select training areas, the width of the roads can be used to estimate the number of pixels which should be integrated in the system perpendicular to the road axis and the accuracy or the road data should be used to estimate the probability that the vector data from the GIS matches with the road information in the imagery.

The flood simulation used in the presented model based only on the DEM. However, a flooding depends on many other parameters as currents, surface structure and infiltration. The substitution of the simple simulation depicted in Fig. 40 on the left by an enhanced flood simulation model developed by hydrologists based on Bayesian networks could improve the complete system. Furthermore, the state of a dam system is in many scenarios a crucial factor for the flood simulation. The integration of the dam system and their state during the flooding could be obtained from a GIS and easily integrated in the Bayesian network.

As mentioned in the previous section the results obtained from the model builds the basis for decision makers. However the possible decisions and the effects of the decisions are not modeled. Influence diagrams, which are a generalization of Bayesian networks delivers the framework to model the decisions and their impacts by means of a graph structure. The developed model should be expanded and generalized by means of influence diagrams to show the decision makers the consequences of their decisions.

A challenging task in this work is the modeling of causal relations and symmetrical neighborhood relations in one probabilistic model. In this thesis the task was solved by the transformation of the Bayesian network and Markov network into factor graphs. Since in general independence properties are lost by the moralization step it would be desirable to have a model, which can simultaneously model causal and symmetrical relations. Chain graphs are graphical models which combine directed and undirected models. However, inference in chain graphs is still part of current research.

Furthermore, the presented model is restricted to linear infrastructure objects. However, a lot of infrastructure objects which are important for relief actions are not linear, such as buildings. Therefore a general probabilistic graphical model to assess the damage of every kind of infrastructure objects could not be based on Markov chains as in this work and therefore the generation of tree structured factor graphs to calculate exact inference is not possible. Therefore frameworks have to be used which are able to model linear objects as well as grid structured objects such as buildings. Conditional random fields are a promising undirected graphical model, which could be a suitable framework to generate general damage assessment systems for all kinds of infrastructure objects (Kumar and Hebert, 2006). However, the inference in conditional random fields is usually solved by the minimization of an energy function and not by exact inference as proposed in this thesis.

Finally, the thesis points out that probabilistic graphical models provide a suitable framework to combine spatial information such as imagery and GIS data in consideration of the uncertainties of the data. It is shown that causal physical relations, the topology of spatial data and multitemporal data can be described in one probabilistic graphical model. Furthermore, the thesis demonstrates that remote sensing data combined with data from a GIS is suitable for a rapid assessment of linear infrastructure objects after flooding.

References

Abramowitz, M., Stegun, I., 1972. Handbook of mathematical functions with formulas, graphs, and mathematical tables. National Bureau of Standards Applied Mathematics, New York.

Alexander, D., 1993. Natural disasters. Kluwer Academic Publishers.

Arzandeh, S., Wang, J., 2002. Texture evaluation of RADARSAT imagery for wetland mapping. Canadian Journal of Remote Sensing 28 (5): 653–666.

Baatz, M., Schaepe, A., 2000. Multiresolution segmentation: an optimization approach for high quality multiscale image segmentation. In J. Strobl, and T. Blaschke (Eds.), Angewandte Geographische Informationsverarbeitung 12: 12–23.

Badji, M., Dautrebande, S., 1997. Characterisation of flood inundated areas and delineation of poor drainage soil using ERS-1 SAR imagery. Hydrological Processes 11 (10): 1441–1450.

Baltsavias, E., 2004. Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. ISPRS Journal of Photogrammetry and Remote Sensing 58 (3-4): 129–151.

Bamler, R., Reinartz, P., Riedlinger, T., Schroeder, M., 2006. Moderne Raumfahrttechniken für Prävention bei Naturkatastrophen und das Krisenmanagement. In: Future Security, 1st security research conference, Germany, Karlsruhe.

Barton, I., Bathols, J., 1989. Monitoring floods with AVHRR. Remote Sensing of Environment 30 (1): 89–94.

Bates, P., De Roo, A., 2000. A simple raster-based model for flood inundation simulation. Journal of Hydrology 236 (1-2): 54–77.

Bates, P., Horritt, M., Smith, C., Mason, D., 1997. Integrating remote sensing observations of flood hydrology and hydraulic modelling. Hydrological Processes 11 (14): 1777–1795.

Bayraktarli, Y., Ulfkjaer, J., Yazgan, U., Faber, M., 2005. On the application of Bayesian probabilistic networks for earthquake risk management. In: 9th international conference on structural safety and reliability, ICOSSAR'05.

Bazi, Y., Bruzzone, L., Melgani, F., 2007. Image thresholding based on the EM algorithm and the generalized gaussian distribution. Pattern Recognition 40 (2): 619–634.

Benediktsson, J., Palmason, J., Sveinsson, J., 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. IEEE Transactions on Geoscience and Remote Sensing 43 (3): 480–491.

Benediktsson, J., Swain, P., Ersoy, O., 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data. IEEE Transactions on Geoscience and Remote Sensing 28 (4): 540–552.

Besag, J., 1986. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society. Series B. Methodological 48 (3): 259–302.

Bhang, K., Schwartz, F., 2008. Limitations in the hydrologic applications of C-band SRTM DEMs in low-relief settings. IEEE Geoscience and Remote Sensing Letters 5 (3): 497–501.

Bishop, C., 2006. Pattern recognition and machine learning. Vol. 4. Springer, New York.

Blake, A., Kohli, P., Rother, C., 2011. Markov random fields for vision and image processing. MIT Press.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65 (1): 2–16.

Bouman, C., Shapiro, M., 1994. A multiscale random field model for Bayesian image segmentation. IEEE Transactions on Image Processing 3 (2): 162–177.

Bovolo, F., Bruzzone, L., 2007. A split-based approach to unsupervised change detection in large-size multitemporal images: Application to tsunami-damage assessment. Geoscience and Remote Sensing, IEEE Transactions on 45 (6): 1658–1670.

Brakenridge, G., Tracy, B., Knox, J., 1998. Orbital SAR remote sensing of a river flood wave. International Journal of Remote Sensing 19 (7): 1439–1445.

Braun, A., Weidner, U., Hinz, S., 2012. Classification in high-dimensional feature spaces - assessment using SVM, IVM and RVM with focus on simulated enmap data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5 (2): 436–443.

Brivio, P., Colombo, R., Maggi, M., Tomasoni, R., 2002. Integration of remote sensing data and GIS for accurate mapping of flooded areas. International Journal of Remote Sensing 23 (3): 429–441.

Butenuth, M., Frey, D., Nielsen, A., Skriver, H., 2011. Infrastructure assessment for disaster management using multi-sensor and multi-temporal remote sensing imagery. International Journal of Remote Sensing 32 (23): 8575–8594.

Butenuth, M., Heipke, C., 2012. Network snakes: graph-based object delineation with active contour models. Machine Vision and Applications 23 (1): 91–109.

Candès, E., Demanet, L., Donoho, D., Ying, L., 2005. Fast discrete curvelet transforms. Multiscale Modeling & Simulation 5: 861–899.

Chen, Z., Curran, P., Hansom, J., 1992. Derivative reflectance spectroscopy to estimate suspended sediment concentration. Remote Sensing of Environment 40 (1): 67–77.

Clifford, P., 1990. Markov random fields in statistics. Disorder in physical systems: A volume in honour of John M. Hammersley : 19–32.

Cobb, B., Shenoy, P., 2004. Inference in hybrid bayesian networks with deterministic variables. In: Proceedings of the Second European Workshop on Probabilistic Graphical Models, PGM'04, 57–64.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E., 2004. Digital change detection methods in ecosystem monitoring: a review. International Journal of Remote Sensing 25 (9): 1565–1596.

Dawid, A., 1979. Conditional independence in statistical theory. Journal of the Royal Statistical Society 41 (1): 1–31.

De Gunst, M., Den Hartog, J., 1994. Knowledge-based updating of maps by interpretation of aerial images. In: IEEE12th IAPR International Conference on Pattern Recognition, ICPR'94, Vol. 1, 811–814.

Dempster, A., 1967. Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics 38 (2): 325–339.

Deutsch, M., Ruggles, F., 1974. Optical data processing and projected application of the ERTS-1 imagery covering the 1973 mississippi river valey floods. Water Resource Bulletin 10 (5): 1032–1039.

Dewan, A., Alam, M., Nishigaki, N., 2005. Evaluating flood damage using GIS and RADARSAT data-A case of the 1998 catastrophe in Greater Dhaka, Bangladesh. Journal of the Faculty of Environmental Science and Technology 10 (1): 67–74.

Dhakal, A., Amada, T., Aniya, M., Sharma, R., 2002. Detection of areas associated with flood and erosion caused by a heavy rainfall using multitemporal Landsat TM data. Photogrammetric Engineering and Remote Sensing 68 (3): 233–239.

Douglas, D., 1994. Least-cost path in GIS using an accumulated cost surface and slopelines. Cartographica: The International Journal for Geographic Information and Geovisualization 31 (3): 37–51.

Egan, J., 1975. Signal detection theory and ROC-analysis. Academic Press.

Figueiredo, M., Jain, A., 2002. Unsupervised learning of finite mixture models. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (3): 381–396.

Forney Jr, G., 1973. The viterbi algorithm. Proceedings of the IEEE 61 (3): 268–278.

Forte, F., Strobl, R., Pennetta, L., 2006. A methodology using GIS, aerial photos and remote sensing for loss estimation and flood vulnerability analysis in the Supersano-Ruffano-Nociglia Graben, southern Italy. Environmental Geology 50 (4): 581–594.

Frazier, P., Page, K., 2000. Water body detection and delineation with Landsat TM data. Photogrammetric Engineering and Remote Sensing 66 (12): 1461–1467.

Frey, B., Kschischang, F., Loeliger, H., Wiberg, N., 1997. Factor graphs and algorithms. In: Proceedings of the Annual Allerton Conference on Communication Control and Computing, Vol. 35, 666–680.

Frey, B., MacKay, D., 1998. A revolution: Belief propagation in graphs with cycles. In M.I. Jordan, M.J. Kearns, and S.A. Solla (Eds.), Advances in Neural Information Processing Systems 10: 479–485.

Frey, D., Butenuth, M., 2009. Classification system of GIS-objects using multi-sensorial imagery for near-realtime disaster management. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII(3/W4), 103–108.

Frey, D., Butenuth, M., 2010. Assessment system of GIS-objects using multi-temporal imagery for near-realtime disaster management. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII (Part A), 43–48.

Frey, D., Butenuth, M., 2011. Trafficability analysis after flooding in urban areas using probabilistic graphical models. In: Joint Urban Remote Sensing Event, JURSE'11, IEEE, 345–348.

Frey, D., Butenuth, M., Hinz, S., 2009a. Analysis of road networks after flood disasters using multi-sensorial remote sensing techniques. In: Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation, Vol. 18, 69–77.

Frey, D., Butenuth, M., Hinz, S., 2009b. Automated detection and classification of intact road networks in multi-sensorial spaceborne imagery for near-realtime disaster management. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS'09.

Frey, D., Butenuth, M., Straub, D., 2012. Probabilistic graphical models for flood state detection of roads combining imagery and DEM. IEEE Geoscience and Remote Sensing Letters 9 (6): 1051.

Frydenberg, M., 1990. The chain graph markov property. Scandinavian Journal of Statistics 17: 333–353.

Gallager, R., 1963. Low-density parity-check codes. MIT Press.

Gamba, P., Dell'Acqua, F., Lisini, G., 2006. Change detection of multitemporal SAR data in urban areas combining feature-based and pixel-based techniques. IEEE Transactions on Geoscience and Remote Sensing 44 (10): 2820–2827.

Geiger, D., Pearl, J., 1988. On the logic of causal models. In: Uncertainty in Artificial Intelligence, UAI'88, Vol. 4, North-Holland, 3–14.

Geiger, D., Pearl, J., 1993. Logical and algorithmic properties of conditional independence and graphical models. The Annals of Statistics 21 (4): 2001–2021.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence (6): 721–741.

Giacomelli, A., Mancini, M., Rosso, R., 1995. Assessment of flooded areas from ERS-1 PRI data: An application to the 1994 flood in northern Italy. Physics and Chemistry of the Earth 20 (5-6): 469–474.

Gianinetto, M., Villa, P., 2007. Rapid response flood assessment using minimum noise fraction and composed spline interpolation. IEEE Transactions on Geoscience and Remote Sensing 45 (10): 3204–3211.

Grêt-Regamey, A., Straub, D., 2006. Spatially explicit avalanche risk assessment linking Bayesian networks to a GIS. Natural Hazards and Earth System Sciences 6: 911–926.

Guo, H., 2009. Natural Disaster Mitigation: A scientific and Practical Approach. Inter Academy Panel on International Issues. Science Press, Beijing.

Hahmann, T., Wessel, B., 2010. Surface water body detection in high-resolution TerraSAR-X data using active contour models. In: Proceedings of 8th European Conference on Synthetic Aperture Radar, EUSAR'10, 897–900.

Hallberg, R., Hoyer, B., Rango, A., 1973. Application of ERTS-1 imagery to flood inundation mapping. In: Proceedings of the Symposium on Significant Results Obtained from the Earth Resource Technology Satellite, Vol. 1, 745–753.

Halpern, J., 2003. Reasoning about Uncertainty. MIT Press.

Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics 3 (6): 610–621.

Harrald, J., 2006. Agility and discipline: Critical success factors for disaster response. The Annals of the American Academy of Political and Social Science 604 (1): 256.

Haykin, S., 1999. Neural networks: A comprehensive foundation. Prentice Hall.

Heckerman, D., Horvitz, E., Nathwani, B., 1992. Toward normative expert systems: Part I. The Pathfinder project. Methods of Information in Medicine 31 (2): 90.

Heckerman, D., Mamdani, A., Wellman, M., 1995. Real-world applications of Bayesian networks. Communications of the ACM 38 (3): 24–26.

Hellwich, O., Günzl, M., Wiedemann, C., 2000. Fusion of optical imagery and sar/insar data for object extraction. International Archives of Photogrammetry and Remote Sensing 33 (B3/1; PART 3): 389–396.

Henry, J., Chastanet, P., Fellah, K., Desnos, Y., 2006. ENVISAT multi-polarized ASAR data for flood mapping. International Journal of Remote Sensing 27 (10): 1921–1929.

Herold, M., Roberts, D., 2005. Spectral characteristics of asphalt road aging and deterioration: Implications for remote-sensing applications. Applied Optics 44 (20): 4327–4334.

Herold, M., Roberts, D., Gardner, M., Dennison, P., 2004. Spectrometry for urban area remote sensing - development and analysis of a spectral library from 350 to 2400 nm. Remote Sensing of Environment 91 (3): 304–319.

Herrera-Cruz, V., Müller, M., Weise, C., 2010. Flood extent mapping based on TerraSAR-X data. Photogrammetrie-Fernerkundung-Geoinformation 2010 (6): 475–488.

Hess, L., Melack, J., Davis, F., 1994. Mapping of floodplain inundation with multi-frequency polarimetric SAR: Use of a tree-based model. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS'94, Vol. 2, 1072–1073.

Hess, L., Melack, J., Filoso, S., Wang, Y., 1995. Delineation of inundated area and vegetation along the amazon floodplain with the SIR-C synthetic aperture radar. IEEE Transactions on Geoscience and Remote Sensing 33 (4): 896–904.

Hess, L., Melack, J., Novo, E., Barbosa, C., Gastil, M., 2003. Dual-season mapping of wetland inundation and vegetation for the central amazon basin. Remote Sens-

ing of Environment 87 (4): 404–428.
Hess, L., Melack, J., Simonett, D., 1990. Radar detection of flooding beneath the forest canopy: a review. International Journal of Remote Sensing 11 (7): 1313–1325.
Hinz, S., Baumgartner, A., 2003. Automatic extraction of urban road networks from multi-view aerial imagery. ISPRS Journal of Photogrammetry and Remote Sensing 58 (1-2): 83–98.
Hong, S., Wdowinski, S., Kim, S., Won, J., 2010. Multitemporal monitoring of wetland water levels in the florida everglades using interferometric synthetic aperture radar (InSAR). Remote Sensing of Environment 114 (11): 2436–2447.
Horritt, M., 1999. A statistical active contour model for sar image segmentation. Image and Vision Computing 17 (3-4): 213–224.
Horritt, M., Mason, D., Luckman, A., 2001. Flood boundary delineation from synthetic aperture radar imagery using a statistical active contour model. International Journal of Remote Sensing 22 (13): 2489–2507.
Hostache, R., Matgen, P., Schumann, G., Puech, C., Hoffmann, L., Pfister, L., 2009. Water level estimation and reduction of hydraulic model calibration uncertainties using satellite SAR images of floods. IEEE Transactions on Geoscience and Remote Sensing 47 (2): 431–441.
Hotelling, H., 1936. Relations between two sets of variates. Biometrika 28 (3/4): 321–377.
IPCC, 2007. Climate Change 2007: Synthesis Report.
Islam, M., Sado, K., 2000. Satellite remote sensing data analysis for flood damaged zoning with GIS for flood management. Journal of Hydraulic Engineering 44: 301–306.
Jacobsen, K., 2010. Vergleich von ASTER GDEM- mit SRTM-Höhenmodellen. Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation 19: 581–588.
Jensen, F., Nielsen, T., 2007. Bayesian networks and decision graphs. Springer, New York.
Jung, H., Alsdorf, D., 2010. Repeat-pass multi-temporal interferometric SAR coherence variations with amazon floodplain and lake habitats. International Journal of Remote Sensing 31 (4): 881–901.
Kalman, R., 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82 (Series D): 35–45.
Karlsson, J., Arnberg, W., 2011. Quality analysis of SRTM and HYDRO1K: a case study of flood inundation in Mozambique. International Journal of Remote Sensing 32 (1): 267–285.
Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. International Journal of Computer Vision 1 (4): 321–331.
Kersten, J., 2011. Ein Rahmenwerk zur interaktiven Klassifikation hochauflösender optischer Satellitenbilder mittels graphenbasierter Bildmodellierung. Ph.D. thesis, Universitätsbibliothek TU Berlin.
Kersten, J., Gähler, M., Voigt, S., 2010. A general framework for fast and interactive classification of optical VHR satellite imagery using hierarchical and planar Markov random fields. Photogrammetrie-Fernerkundung-Geoinformation 2010 (6): 439–449.
Khan, S., Hong, Y., Wang, J., Yilmaz, K., Gourley, J., Adler, R., Brakenridge, G., Policelli, F., Habib, S., Irwin, D., 2011. Satellite remote sensing and hydrologic modeling for flood inundation mapping in lake victoria basin: Implications for hydrologic prediction in ungauged basins. IEEE Transactions on Geoscience and Remote Sensing 49 (1): 85–95.
Kittler, J., Föglein, J., 1984. Contextual classification of multispectral pixel data. Image and Vision Computing 2 (1): 13–29.
Kittler, J., Illingworth, J., 1986. Minimum error thresholding. Pattern recognition 19 (1): 41–47.
Kohonen, T., 2000. Self-Organizing Maps. Springer, New York.
Koller, D., Friedman, N., 2009. Probabilistic graphical models: Principles and techniques. MIT Press.
Krieger, G., Fiedler, H., Hajnsek, I., Eineder, M., Werner, M., Moreira, A., 2005. TanDEM-X: mission concept and performance analysis. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Vol. 7, 4890–4894.
Kschischang, F., Frey, B., Loeliger, H., 2001. Factor graphs and the sum-product algorithm. IEEE Transactions on Information Theory 47 (2): 498–519.
Kuehn, S., Benz, U., Hurley, J., 2002. Efficient flood monitoring based on RADARSAT-1 images data and information fusion with object-oriented technology. Vol. 5, 2862–2864.
Kumar, S., Hebert, M., 2006. Discriminative random fields. International Journal of Computer Vision 68 (2): 179–201.
Kussul, N., Shelestov, A., Skakun, S., 2008. Grid system for flood extent extraction from satellite images. Earth Science Informatics 1 (3): 105–117.
Lacava, T., Filizzola, C., Pergola, N., Sannazzaro, F., Tramutoli, V., 2010. Improving flood monitoring by the robust AVHRR technique (RAT) approach: The case of the april 2000 Hungary flood. International Journal of Remote Sensing 31 (8): 2043–2062.
Lafferty, J., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of 18th International Conference on Machine Learning, ICML'01.
Lang, M., Townsend, P., Kasischke, E., 2008. Influence of incidence angle on detecting flooded forests using C-HH synthetic aperture radar data. Remote Sensing of Environment 112 (10): 3898–3907.
Lauritzen, S., 1982. Lectures on contingency tables. Aalborg University Press.
Lauritzen, S., 1996. Graphical models. Oxford University Press.
Lauritzen, S., Spiegelhalter, D., 1988. Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society. Series B : 157–224.
Lerner, U., 2002. Hybrid Bayesian networks for reasoning about complex systems. Ph.D. thesis, Stanford University.
Li, S., 2009. Markov random field modeling in image analysis. Springer, New York.
Loeliger, H., 2004. An introduction to factor graphs. IEEE Signal Processing Magazine 21 (1): 28–41.
Long, N., Trong, B., 2001. Flood monitoring of Mekong River delta, Vietnam using ERS SAR data. In: Proceeding of the 22nd Asian Conference on Remote Sensing, ACRS'01, Vol. 5.
Lowry, R., Mudry, N., Langham, E., 1981. A preliminary analysis of SAR mapping of the Manitoba flood, May 1979. Satellite Hydrology : 316–323.
Lu, D., Mausel, P., Brondizio, E., Moran, E., 2004. Change detection techniques. International Journal of Remote Sensing 25 (12): 2365–2401.
Lyzenga, D., 1978. Passive remote sensing techniques for mapping water depth and bottom features. Applied Optics 17 (3): 379–383.
Martinez, J., Le Toan, T., 2007. Mapping of flood dynamics and spatial distribution of vegetation in the Amazon floodplain using multitemporal SAR data. Remote Sensing of Environment 108 (3): 209–223.
Martinis, S., Twele, A., Voigt, S., 2009. Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data. Natural Hazards and Earth System Sciences 9 (2): 303–314.
Martinis, S., Twele, A., Voigt, S., 2011. Unsupervised extraction of flood-induced backscatter changes in

SAR data using Markov image modeling on irregular graphs. IEEE Transactions on Geoscience and Remote Sensing 49 (1): 251–263.

Mason, D., Davenport, I., 1996. Accurate and efficient determination of the shoreline in ERS-1 SAR images. IEEE Transactions on Geoscience and Remote Sensing 34 (5): 1243–1253.

Mason, D., Horritt, M., Dall'Amico, J., Scott, T., Bates, P., 2007. Improving river flood extent delineation from synthetic aperture radar using airborne laser altimetry. IEEE Transactions on Geoscience and Remote Sensing 45 (12): 3932–3943.

Mason, D., Speck, R., Devereux, B., Guy, J., Neal, J., Bates, P., 2010. Flood detection in urban areas using TerraSAR-X. IEEE Transactions on Geoscience and Remote Sensing 48 (2): 882–894.

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

McLachlan, G., Peel, D., 2000. Finite mixture models. Wiley-Interscience, New York.

Miasnikov, A., Rome, J., Haralick, R., 2004. A hierarchical projection pursuit clustering algorithm. In: Proceedings of the IEEE 17th International Conference on Pattern Recognition, ICPR'04, Vol. 1, IEEE, 268–271.

Milfred, C., Parker, D., Lee, G., 1969. Remote sensing for resource management and flood plain delineation. In: 24th Midwestern States Flood Control and Water Resources Conference.

Molina, M., Fuentetaja, R., Garrote, L., 2005. Hydrologic models for emergency decision support using Bayesian networks. In: L. Godo (Ed.) Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Computer Science, Vol. 3571, Springer, 468–468.

Montanari, M., Hostache, R., Matgen, P., Schumann, G., Pfister, L., Hoffmann, L., 2009. Calibration and sequential updating of a coupled hydrologic-hydraulic model using remote sensing-derived water stages. Hydrology and Earth System Sciences 13 (3): 367–380.

Morain, S., Kraft, W., 2003. Transportation lifelines and hazards: Overview of remote sensing products and results. In: Proceedings of Remote Sensing for Transportation, Vol. 29, 39–46.

Morrison, R., Cooley, M., 1973. Assessment of flood damage in Arizona by means of ERTS-1 imagery. In: Proceedings of the Symposium on Significant Results Obtained from the Earth Resource Technology Satellite, Vol. 1, 755–760.

MunichRe, 2011. Natural catastrophes worldwide 1980 - 2010.

Nagarajan, K., Krekeler, C., Slatton, K., Graham, W., 2010. A scalable approach to fusing spatiotemporal data to estimate streamflow via a Bayesian network. IEEE Transactions on Geoscience and Remote Sensing 48 (10): 3720–3732.

Nico, G., Pappalepore, M., Pasquariello, G., Refice, A., Samarelli, S., 2000. Comparison of SAR amplitude vs. coherence flood detection methods - a GIS application. International Journal of Remote Sensing 21 (8): 1619–1631.

Niedermeier, A., Romaneessen, E., Lehner, S., 2000. Detection of coastlines in SAR images using wavelet methods. IEEE Transactions on Geoscience and Remote Sensing 38 (5): 2270–2281.

Niedermeyer, R., 2006. Das Elbe-Hochwasser 2002: Geowissenschaftliche Auswertungen. Schweizerbart'sche Verlagsbuchhandlung.

Nielsen, A., 2007. The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. Image Processing, IEEE Transactions on 16 (2): 463–478.

Nielsen, A., Conradsen, K., Simpson, J., 1998. Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. Remote Sensing of Environment 64 (1): 1–19.

Ormsby, J., Blanchard, B., Blanchard, A., 1985. Detection of lowland flooding using active microwave systems. Photogrammetric Engineering and Remote Sensing 51 (3): 317–329.

Otsu, N., 1979. A threshold selection method from gray-level histograms. Automatica 11: 285–296.

Overton, I., 2005. Modelling floodplain inundation on a regulated river: integrating GIS, remote sensing and hydrological models. River Research and Applications 21 (9): 991–1001.

Paltridge, G., Platt, C., 1976. Radiative processes in meteorology and climatology. Developments in Atmospheric Science 5: 32–34.

Park, M., Stenstrom, M., 2006. Using satellite imagery for stormwater pollution management with Bayesian networks. Water Research 40 (18): 3429–3438.

Pearl, J., 1986. Fusion, propagation, and structuring in belief networks. Artificial Intelligence 29 (3): 241–288.

Pearl, J., 1988. Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann.

Perez, P., 1998. Markov random fields and images. CWI Quarterly 11 (4): 413–437.

Pohl, C., Van Genderen, J., 1998. Multisensor image fusion in remote sensing: Concepts, methods and applications. International Journal of Remote Sensing 19: 823–854.

Polidori, L., Caillault, S., Canaud, J., 1995. Change detection in radar images: Methods and operational constraints. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS'95, Vol. 2, 1529–1531.

Pope, K., Rejmankova, E., Paris, J., Woodruff, R., 1997. Detecting seasonal flooding cycles in marshes of the Yucatan peninsula with SIR-C polarimetric radar imagery. Remote Sensing of Environment 59 (2): 157–166.

Profeti, G., Macintosh, H., 1997. Flood management through Landsat TM and ERS SAR data: A case study. Hydrological Processes 11 (10): 1397–1408.

Provost, J., Collet, C., Rostaing, P., Pérez, P., Bouthemy, P., 2004. Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps. Computer Vision and Image Understanding 93 (2): 155–174.

Pulvirenti, L., Chini, M., Pierdicca, N., Guerriero, L., Ferrazzoli, P., 2011. Flood monitoring using multitemporal COSMO-SkyMed data: Image segmentation and signature interpretation. Remote Sensing of Environment 115 (4): 990–1002.

Quegan, S., Le Toan, T., Yu, J., Ribbes, F., Floury, N., 2000. Multitemporal ERS SAR analysis applied to forest mapping. IEEE Transactions on Geoscience and Remote Sensing 38 (2): 741–753.

Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77 (2): 257–286.

Rabiner, L., Juang, B., 1986. An introduction to hidden Markov models. ASSP Magazine, IEEE 3 (1): 4–16.

Radke, R., Andra, S., Al-Kofahi, O., Roysam, B., 2005. Image change detection algorithms: a systematic survey. IEEE Transactions on Image Processing 14 (3): 294–307.

Rango, A., Anderson, A., 1974. Flood hazard studies in the Mississippi river basin using remote sensing. Water Resource Bulletin 10 (5): 1060–1081.

Rango, A., Salomonson, V., 1973. Repetetive ERTS-1 observations of surface water variability along rivers and

other low-lying areas. In: Proceedings of International Symposium on Remote Sensing and Water Management, 191–208.

Rignot, E., van Zyl, J., 1993. Change detection techniques for ERS-1 SAR data. IEEE Transactions on Geoscience and Remote Sensing 31 (4): 896–906.

Rodriguez, E., Morris, C., Belz, J., 2006. A global assessment of the SRTM performance. Photogrammetric Engineering and Remote Sensing 72 (3): 249–260.

Rollet, R., Benie, G., Li, W., Wang, S., Boucher, J., 1998. Image classification algorithm based on the RBF neural network and K-means. International Journal of Remote Sensing 19 (15): 3003–3009.

Romanowicz, R., Beven, K., Tawn, J., 1996. Bayesian calibration of flood inundation models. Floodplain Processes : 333–360.

Sahoo, P., Soltani, S., Wong, A., 1988. A survey of thresholding techniques. Computer Vision, Graphics, and Image Processing 41 (2): 233–260.

Sakurai, J., 1985. Modern quantum mechanics. Addison Wesley Publishing Company.

Sanyal, J., Lu, X., 2004. Application of remote sensing in flood management with special reference to monsoon Asia: A review. Natural Hazards 33 (2): 283–301.

Schmitt, A., Wessel, B., Roth, A., 2010. Curvelet-based change detection on SAR images for natural disaster mapping. Photogrammetrie-Fernerkundung-Geoinformation 2010 (6): 463–474.

Schultz, G., 1988. Remote sensing in hydrology. Journal of Hydrology 100 (1-3): 239–265.

Schumann, G., Di Baldassarre, G., Bates, P., 2009. The utility of spaceborne radar to render flood inundation maps based on multialgorithm ensembles. IEEE Transactions on Geoscience and Remote Sensing 47 (8): 2801–2807.

Schumann, G., Neal, J., Mason, D., Bates, P., 2011. The accuracy of sequential aerial photography and SAR data for observing urban flood dynamics, a case study of the UK summer 2007 floods. Remote Sensing of Environment 115 (10): 2536–2546.

Shafer, G., 1976. A mathematical theory of evidence. Vol. 1. Princeton University Press.

Shafer, G., Shenoy, P., 1990. Probability propagation. Annals of Mathematics and Artificial Intelligence 2 (1): 327–351.

Sheng, Y., Su, Y., Xiao, Q., 1998. Challenging the cloud-contamination problem in flood monitoring with NOAA/AVHRR imagery. Photogrammetric Engineering and Remote Sensing 64 (3): 191–198.

Silveira, M., Heleno, S., 2009. Separation between water and land in SAR images using region-based level sets. IEEE Geoscience and Remote Sensing Letters 6 (3): 471–475.

Smith, L., 1997. Satellite remote sensing of river inundation area, stage, and discharge: A review. Hydrological Processes 11 (10): 1427–1439.

Soergel, U., Thoennessen, U., Stilla, U., 2003. Visibility analysis of man-made objects in SAR images. In: 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, 120–124.

Song, Y., Sohn, H., Park, C., 2007. Efficient water area classification using Radarsat-1 SAR imagery in a high relief mountainous environment. Photogrammetric Engineering and Remote Sensing 73 (3): 285.

Straub, D., Grêt-Regamey, A., 2006. A Bayesian probabilistic framework for avalanche modelling based on observations. Cold Regions Science and Technology 46 (3): 192–203.

Studený, M., Bouckaert, R., 1998. On chain graph models for description of conditional independence structures. Annals of Statistics 26: 1434–1495.

Sutton, C., McCallum, A., 2007. An introduction to conditional random fields for relational learning. MIT Press.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C., 2008. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30 (6): 1068–1080.

Tanner, R., 1981. A recursive approach to low complexity codes. Information Theory, IEEE Transactions on 27 (5): 533 – 547.

Tholey, N., Clandillon, S., De Fraipont, P., 1997. The contribution of spaceborne SAR and optical data in monitoring flood events: Examples in northern and southern France. Hydrological Processes 11 (10): 1409–1413.

Townsend, P., Walsh, S., 1998. Modeling floodplain inundation using an integrated GIS with radar and optical remote sensing. Geomorphology 21 (3-4): 295–312.

Tramutoli, V., 1998. Robust AVHRR techniques (RAT) for environmental monitoring: Theory and applications. In: Society of Photo-Optical Instrumentation Engineers, Vol. 3496, SPIE Bellingham, WA, 101–113.

Van der Sande, C., De Jong, S., De Roo, A., 2003. A segmentation and classification approach of IKONOS-2 imagery for land cover mapping to assist flood risk and flood damage assessment. International Journal of Applied Earth Observation and Geoinformation 4 (3): 217–229.

van Westen, C., Soeters, R., 2000. Remote sensing and geographic information systems for natural disaster management. Natural disasters and their mitigation: A remote sensing and GIS perspective : 31–76.

Verdin, J., 1996. Remote sensing of ephemeral water bodies in western Niger. International Journal of Remote Sensing 17 (4): 733–748.

Viterbi, A., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory 13 (2): 260–269.

Wang, Y., Colby, J., Mulcahy, K., 2002. An efficient method for mapping flood extent in a coastal floodplain using Landsat TM and DEM data. International Journal of Remote Sensing 23 (18): 3681–3696.

Wang, Y., Hess, L., Filoso, S., Melack, J., 1995. Understanding the radar backscattering from flooded and nonflooded Amazonian forests: Results from canopy backscatter modeling. Remote Sensing of Environment 54 (3): 324–332.

Wdowinski, S., Kim, S., Amelung, F., Dixon, T., Miralles-Wilhelm, F., Sonenshein, R., 2008. Space-based detection of wetlands' surface water level changes from L-band SAR interferometry. Remote Sensing of Environment 112 (3): 681–696.

Wiberg, N., Loeliger, H., Kotter, R., 1995. Codes and iterative decoding on general graphs. European Transactions on Telecommunications 6 (5): 513–525.

Wiedemann, C., Hinz, S., 1999. Automatic extraction and evaluation of road networks from satellite imagery. International Archives of Photogrammetry and Remote Sensing 32 (3; SECT 2W5): 95–100.

Wiesnet, D., McGinnis, D., Pritchard, J., 1974. Mapping of the 1973 Mississippi river floods by the NOAA-2 satellite. Water Resource Bulletin 10 (5): 1040–1049.

Wilson, P., 1997. Rule-based classification of water in Landsat MSS images using the variance filter. Photogrammetric Engineering and Remote Sensing 63 (5): 485491.

Work, E., 1976. Utilization of satellite data for inventorying prairie ponds and lakes. Photogrammetric Engineering and Remote Sensing 42: 685–694.

Yamagata, Y., Akiyama, T., 1988. Flood damage analysis using multitemporal Landsat Thematic Mapper data. International Journal of Remote Sensing 9 (3): 503–514.

Yedidia, J., Freeman, W., Weiss, Y., 2005. Construct-

ing free-energy approximations and generalized belief propagation algorithms. IEEE Transactions on Information Theory 51 (7): 2282–2312.

Yonghua, S., Xiaojua, L., Huili, G., Wenji, Z., Zhaonin, G., 2007. A study on optical and SAR data fusion for extracting flooded area. In: Proceedings on Geoscience and Remote Sensing Symposium, IGARSS'07, 3086 – 3089.

Zadeh, L., 1965. Fuzzy sets. Information and control 8 (3): 338–353.

Zadeh, L., 1978. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1 (1): 3–28.

Zhang, J., 2010. Multi-source remote sensing data fusion: status and trends. International Journal of Image and Data Fusion 1 (1): 5–24.

Zhang, J., Zhou, C., Xu, K., Watanabe, M., 2002. Flood disaster monitoring and evaluation in China. Environmental Hazards 4 (2): 33–43.

Zhou, G., Luo, J., Yang, C., Li, B., Wang, S., 2000. Flood monitoring using multi-temporal AVHRR and RADARSAT imagery. Photogrammetric Engineering and Remote Sensing 66 (5): 633–638.

Zwenzner, H., Voigt, S., 2009. Improved estimation of flood parameters by combining space based SAR data with very high resolution digital elevation data. Hydrology and Earth System Sciences 13: 567–576.