



# A Hidden Markov Tree Model for Flood Extent Mapping in Heavily Vegetated Areas based on High Resolution Aerial Imagery and DEM: A Case Study on Hurricane Matthew Floods

Zhe Jiang and Arpan Man Sainju

Department of Computer Science, the University of Alabama, Tuscaloosa, Alabama, USA

## ARTICLE HISTORY

Compiled May 30, 2021

## ABSTRACT

Flood extent mapping plays a crucial role in disaster management and national water forecasting. In recent years, high-resolution optical imagery becomes increasingly available with the deployment of numerous small satellites and drones. However, analyzing such imagery data to extract flood extent poses unique challenges due to the rich noise and shadows, obstacles (e.g., tree canopies, clouds), and spectral confusion between pixel classes (flood, dry) due to spatial heterogeneity. Existing machine learning techniques often focus on spectral and spatial features from raster images without fully incorporating the geographic terrain within classification models. In contrast, we recently proposed a novel machine learning model called geographical hidden Markov tree (HMT) that integrates spectral features of pixels and topographic constraint from Digital Elevation Model (DEM) data (i.e., water flow directions) in a holistic manner. This paper evaluates the model through case studies on high-resolution aerial imagery from National Oceanic and Atmospheric Administration (NOAA) National Geodetic Survey (NGS) together with DEM. Three scenes are selected in heavily vegetated floodplains near the cities of Grimesland and Kinston in North Carolina during Hurricane Matthew floods in 2016. Results show that the proposed HMT model outperforms several state of the art machine learning algorithms (e.g., random forests, gradient boosted model) by an improvement of F-score (the harmonic mean of the user's accuracy and producer's accuracy) from around 70% to 80% to over 95% on our datasets.

## KEYWORDS

Remote sensing; flood mapping; topography; tree canopies; spatial structured prediction; hidden Markov tree

## 1. Introduction

Flood extent mapping plays a crucial role in addressing grand societal challenges such as disaster management, national water forecasting, as well as energy and food security (Eftealioglu et al. (2016)). For example, during Hurricane Harvey floods in 2017, first responders needed to know where flood water was in order to plan rescue efforts to residents in vulnerable communities, to understand the extent of damage to critical infrastructures (e.g., chemical plants and oil refineries), and evaluate the impact on road networks and transportation. In national water forecasting, National Oceanic and Atmospheric Administration (2018) and Cline (2009) introduce a Super-resolution National Water Model being operated at the National Water Center, which can forecast the flow of over 2.7 million rivers and streams through

the entire continental U.S.. One main issue of the national water model is that it has a large number of parameters in physical models that can only be calibrated and validated based on observations from a few thousands of river gauges. Detailed flood extent maps derived from remote sensing observations provide an alternative way to calibrate and validate the national water model in broad spatial coverage.

Mapping flood extent from earth observation imagery has been extensively studied in the remote sensing community. Existing techniques can be categorized by the types of remote sensors (optical sensor, radar, Light Detection and Ranging (LiDAR)) or the underlying approaches (machine learning, non-machine learning). Fayne et al. (2017) uses the Normalized Difference Vegetation Index (NDVI) from Moderate Resolution Imaging Spectroradiometer (MODIS) optical imagery together with change detection to classify flood extent. McFeeters (1996) and Xu (2006) propose the use of Normalized Difference Water Index (NDWI) and modified NDWI to detect surface water features respectively. Sun, Yu, and Goldberg (2011) proposes to use decision trees to map flood extent from MODIS optical images. Zhang, Zhu, and Liu (2014) proposes the blend of Land Remote-Sensing Satellite (Landsat) and MODIS optical images for urban flood mapping. Lin et al. (2019) proposes algorithms filter noises from terrain shadows to improve National Aeronautics and Space Administration (NASA)/MODIS Near Real-Time (NRT) Global Flood Mapping. Feng, Liu, and Gong (2015) uses random forest algorithms on Unmanned Aerial Vehicles (UAV) imagery for mapping flood extent in China. One major limitation of using imagery from optical sensors is that the scene is often obscured by obstacles (e.g., cloud, tree canopies). Because of this reason, radar sensor data are often used to identify flood extent since the signals can penetrate through clouds and some vegetated areas. Giustarini et al. (2012) proposes a change detection method on radar imagery from TerraSAR-X. Horritt et al. (2003) proposes to use airborne Synthetic Aperture Radar (SAR) imagery to identify flooded vegetation. Townsend (2001) proposes to use multi-temporal SAR imagery to map seasonal flooding in forested wetlands. Oberstadler, Hönsch, and Huth (1997) assessed the capabilities of European Remote Sensing (ERS) satellite ERS-1 SAR data for flood mapping. Yang et al. (2014) proposes change detection algorithms for high-resolution SAR images in the context of flood mapping. Though radar data have the advantage of penetration through clouds and some vegetation, the signals are often very noisy (e.g., speckles due to interference with irregular surface). There are works that propose the integration of optical or radar sensor data together with topographic information from Digital Elevation Model (DEM) to delineate flood extent. Wang, Colby, and Mulcahy (2002) proposes the use of Landsat optical imagery together with topography from DEM to map flood extent. The approach showed significant improvements in identifying the underestimated flood extent in heavily vegetated area, but the approach is not automatic and involves manually thresholding the DEM. Such threshold may vary from one region to another. Matgen et al. (2007) proposes the integration of SAR data, high-precision topographic features, and a river flow model for near real-time flood management. Rahman et al. (2019) proposes to use model-driven soil moisture from the Soil Moisture Active Passive (SMAP) data product for rapid flood progress monitoring to estimate crop loss in several flood events, including the Hurricane Harvey flood in Houston 2017. There are other review articles on machine learning or data science techniques for earth observation imagery data, including Jiang and Shekhar (2017), Jiang (2018), Shekhar et al. (2015), Jiang (2020), and Karpatne et al. (2016). Rahman and Di (2017) provides a review on existing methods and approaches that overcome the constraints in the application of spaceborne remote sensing in flood management, such as bad weather condition and cloud cover.

Mapping flood from high-resolution imagery in heavily vegetated areas poses several unique challenges that are not well addressed in traditional machine learning classification models. First, high-resolution earth imagery often has noise, shadows, clouds, and tree

canopies (Jiang et al. (2012, 2013, 2015)). Second, class confusion exists due to heterogeneous spectral features. For example, the spectral signature of pixels in shadows and flooded areas can be very similar but they are in different classes (Jiang et al. (2017, 2019)). As another example, pixels of tree canopies overlaying flood water have the same spectral features with pixels of tree canopies in dry areas. This is a challenge for machine learning models which classify pixel types largely based on pixel or neighborhood features on the imagery. Shen et al. (2019) finds that though the automation and robustness of non-obstructed inundation mapping have been achieved with acceptable accuracy, they are not yet satisfactory for the detection of beneath-vegetation flood mapping. Third, implicit directed spatial dependency exists between flood pixel locations. Specifically, due to gravity, flood water tends to flow to nearby lower locations following topography (Sainju et al. (2020)). Such dependency is not uniform in all directions (anisotropic). Such topographic constraints are often ignored in the existing machine learning image classification algorithms. Finally, the data volume is huge in high-resolution imagery (e.g., hundreds of millions of pixels), requiring algorithms to be computationally scalable.

To address these challenges, a novel spatial classification model called *geographical hidden Markov tree* (HMT) was recently proposed in the data mining community by Xie, Jiang, and Sainju (2018); Jiang and Sainju (2019); Jiang, Xie, and Sainju (2019). Intuitively, HMT is a spatial machine learning model that integrates local likelihood of pixels' classes derived from their own spectral signatures and the topographic constraint between pixels derived from water flow directions. Specifically, HMT is a probabilistic graphical model that generalizes the common hidden Markov model (HMM) from a one-dimensional (1D) sequence to a two dimensional (2D) geographical map. The hidden class layer contains nodes (pixels) in a reverse tree structure to represent anisotropic spatial dependency of water flow directions. Each hidden class node has an associated observed feature node for the same pixel (local class likelihood). Such a unique model structure can potentially reduce classification errors due to noise, obstacles, and heterogeneity among spectral features of individual pixels. Efficient algorithms exist for model parameter learning and class inference based on the Expectation-Maximization (EM) algorithm with message propagation on tree traversal.

The goal of this paper is to introduce this technique to the remote sensing community. The paper first introduces the statistical formulation and intuition of the HMT approach. It provides three detailed case studies on high-resolution optical imagery from National Oceanic and Atmospheric Administration (NOAA) National Geodetic Survey (NGS) collected during Hurricane Matthew floods in North Carolina 2016. Results show that HMT outperforms several state of the art machine learning algorithms in accurately mapping flood extent in heavily vegetated areas. Note that the paper only focuses on machine learning approaches for flood mapping. Other non-machine learning approaches (e.g., thresholding a water index, simple change detection) are beyond the scope. The HMT model can be further tested on datasets collected by different sensors among different areas and in different flood events in future studies.

## 2. Materials and Methods

### 2.1. Dataset Description

Our real-world high-resolution aerial imagery datasets were acquired by the NOAA Remote Sensing Division to support homeland security and emergency response requirements after each hurricane disaster. The imagery was downloaded from the NOAA NGS at National Oceanic and Atmospheric Administration (2017) related to the Hurricane Matthew in 2016.

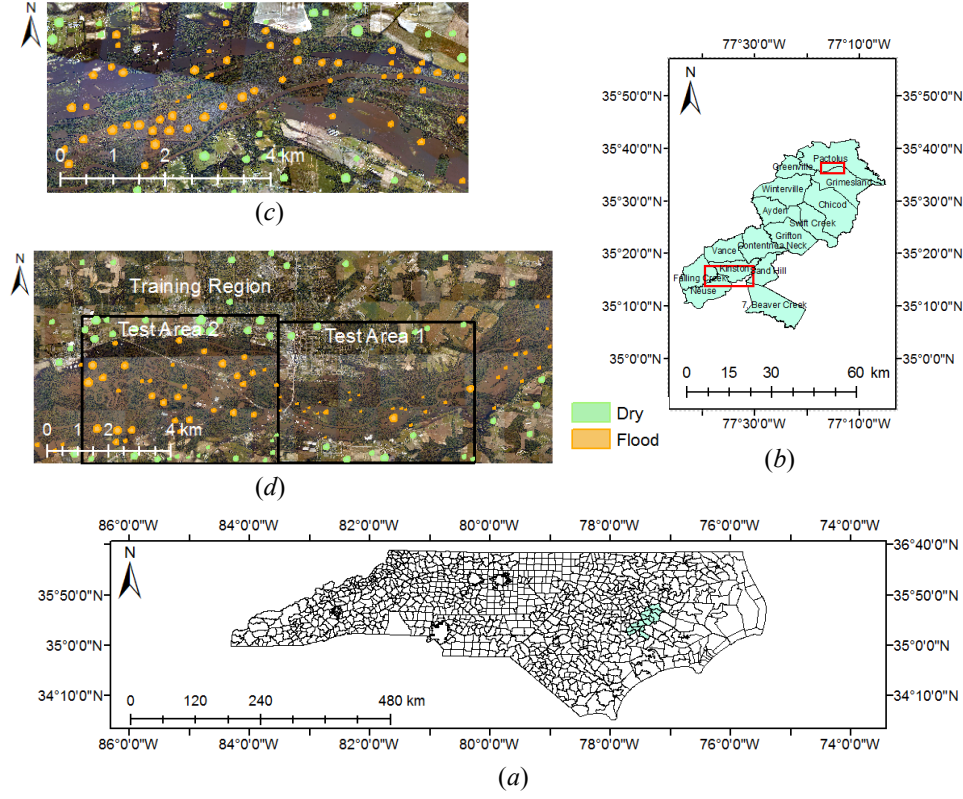


Figure 1.: The map of three study areas (Source: NOAA NGS, Date: 15 October 2016). (a) North Carolina. (b) Study area. (c) Grimesland test region. (d) Kinston region. The same training set in Kinston region is used for three test scenes (test area 1 and test area 2 in Kinston region as well as Grimesland test region).

The imagery was originally at 0.5 metre spatial resolution with red, green, and blue bands only. The limited spectral bands in the visible spectrum and the varying illumination conditions in the imagery were the major challenges in utilizing the data for flood mapping. We used this dataset to test the potential of HMT in leveraging topography to address spectral limitations. The imagery for flood disasters only covered a few scattered scenes at river floodplains in North Carolina and South Carolina. The specific spatial coverage of the complete datasets can be found on the Emergency Response Imagery Viewer at National Oceanic and Atmospheric Administration (2017). The high-resolution imagery was collected only once during the hurricane flood disaster. Due to the large data volume, we cropped three representative scenes near the cities of Grimesland and Kinston in North Carolina. We also resampled the imagery into 2 metre resolution to reduce the data volume and also to make it consistent with the DEM, which was downloaded from North Carolina State University Library (2018) at 2 metre resolution. The HMT model uses DEM data to construct a flow dependency tree structure on all pixels in the a region. This tree structure is the core component of the HMT model in parameter learning and class inference, since inferring pixels' classes are not only based on its own spectral signatures but also the likely classes of other pixels on the topographic surface. The locations of the three scenes are shown in Figure 1.

The locations of the ground truth polygons in the three study areas are also shown in Figure 1. These flood and dry polygons were manually labelled through visual interpreting the aerial imagery together with the DEM. The labeled polygons within the rectangular boundary

Table 1.: Dataset description

| Dataset            | Training set |       | Testing set |         |
|--------------------|--------------|-------|-------------|---------|
|                    | Dry          | Flood | Dry         | Flood   |
| Grimesland, NC     | 5,000        | 5,000 | 49,597      | 99,532  |
| Kinston area 1, NC | 5,000        | 5,000 | 46,245      | 20,331  |
| Kinston area 2, NC | 5,000        | 5,000 | 110,805     | 123,366 |

of the test regions are used for testing. The labeled polygons in Kinston region outside the test areas were used for training. Note that the same training set was used for model learning in all three test regions. For the test region in Grimesland, the training data was far away from it so that we can evaluate the effectiveness of the transductive learning in HMT (initializing model parameters from training data and updating the parameters based on test data that is collected far away). Training and testing samples (pixels) were randomly drawn from the ground truth polygons. Their numbers are listed in Table 1.

## 2.2. Methods

We now introduce the HMT model first proposed by Xie, Jiang, and Sainju (2018), which explicitly incorporates the physical constraints of water flow directions based on topography into image classification process. Figure 2 provides an illustrative example with eight consecutive locations (pixels) in 1D space. We can consider these eight locations as eight consecutive pixels on a 1D intersection of a DEM layer. Figure 2 (a) shows the DEM values of the eight pixel locations with a topography structure. Due to gravity, water tends to flow from a location to nearby lower locations. If location 5 is flooded, then locations 2 to 4 and 6 to 7 must also be flooded. Such dependency structure of pixel classes based on topography can be represented in a tree structure shown by Figure 2 (b), whereby any node being flooded indicates all sub-tree nodes being flooded as well. For example, in Figure 2 (b), if node 5 is flood, all the subtree nodes (2 to 4, and 6 to 7) must be flooded. Similarly, if node 1 is flooded, nodes 2 to 7 must also be flooded. Though the example shows the flow dependency tree based on elevation values in one dimensional space (DEM along a 1D line segment), the idea can be easily generalized to a two dimensional DEM map, which is used in this paper. Figure 3 shows the topography (illustrated by flow directions) on DEM in the three test regions. It is worth noting that the flow dependency tree structure is not merely based on flow directions between adjacent pixels. For example, if we were only using flow directions between adjacent pixels in Figure 2 (a) to construct the flow dependency tree, we can only capture the dependency between pixel 5 and its neighbors 4 and 6 and miss the dependency between pixel 1 and pixel 5 (though in reality, if pixel 1 is flood, pixel 5 should also be flood since water can fill into pixel 5 from pixel 1). The flow dependency tree can be efficiently constructed from a DEM layer based on computational topology algorithms (details are in Xie, Jiang, and Sainju (2018)). Note that in this paper, we use the word *node*, *location*, *sample*, and *pixel* exchangeably.

The HMT model incorporates the class dependency structure from topography into a hidden class layer. As illustrated in Figure 2 (c), an HMT model consists of two layers: a hidden class layer of unknown pixel classes (e.g., flood, dry), and an observation layer of sample spectral features (e.g., spectral band values). Each node corresponds to a spatial data sample (image pixel location). Arrow directions between class nodes (with gray shading) capture class transitional probability based on the topography structure. Arrows between an observation node (without shading) and a hidden class node (with gray shading) captures the con-

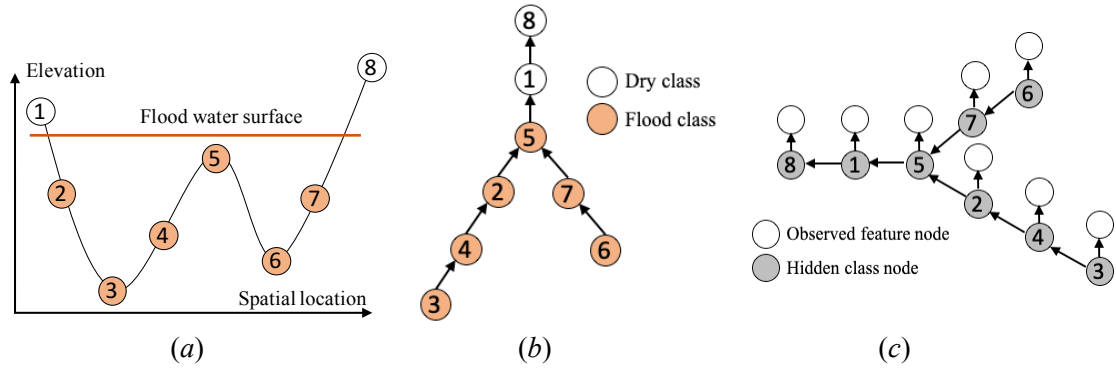


Figure 2.: Illustration of hidden Markov tree framework. (a) Eight consecutive sample locations in 1D space. (b) Partial order constraint in a reverse tree. (c) Hidden Markov tree

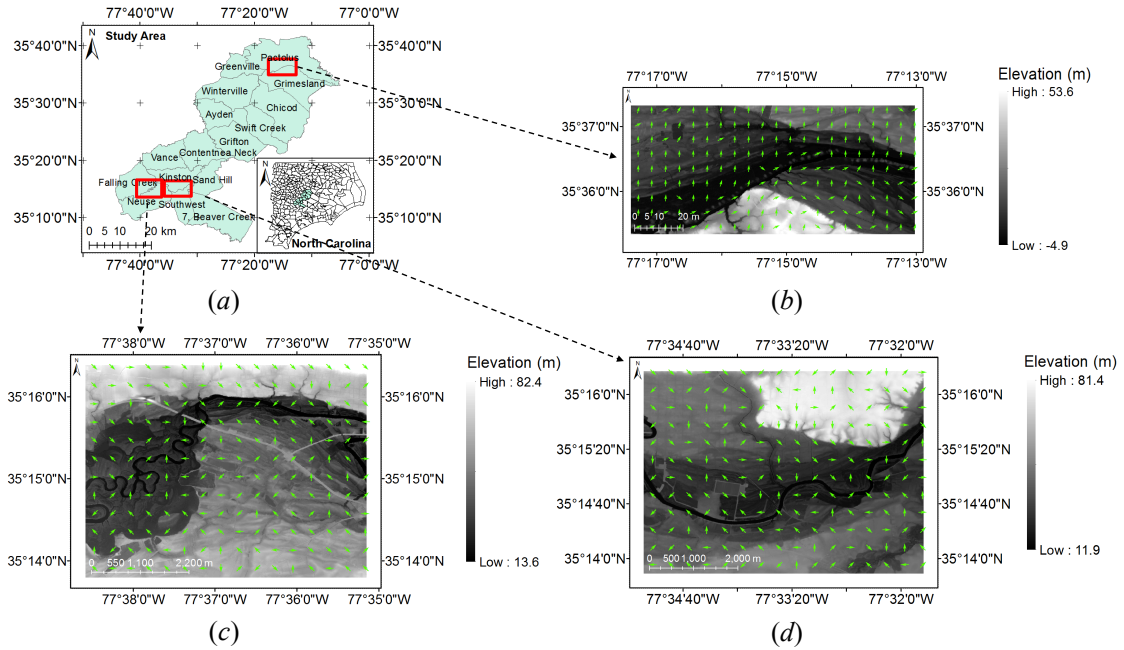


Figure 3.: Flow direction maps based on DEM of the three test regions. (a) Study Area. (b) Grimesland (m). (c) Kinston: Area 2 (m). (d) Kiston: Area 1 (m)

ditional probability of a pixel's spectral signatures given its class (local likelihood). In this way, the class of a pixel is determined by not only the spectral signatures of itself but also the classes of other pixels based on water flow directions in topography surface. Such model structure is important since the spectral signatures of individual pixels or their local neighborhoods can be misleading due to noise, shadows, obstacles (e.g., tree canopies, clouds), as well as spectral confusion (i.e., the same pixel color may correspond to different classes due to varying illustrations). The class dependency structure based on the topography constraint provides an opportunity to overcome these issues.

HMT is a probabilistic graphic model. It generalizes the common hidden Markov model from a total order sequence structure to a partial order tree structure. Based on the dependency structure in the model, the joint distribution of all samples' features and classes are expressed as Equation (1), where  $\mathbf{X}$  and  $\mathbf{Y}$  are matrices of the spectral features and class labels of all samples respectively,  $n$  is the sample (node) index,  $N$  is the total number of samples (i.e., nodes, or pixels),  $\mathbf{x}_n$  and  $y_n$  are the spectral feature and class label of the  $n$ th sample,  $Z_n$  is the set of parent nodes of the  $n$ th node in the dependency tree, and  $M_n = \{y_k | k \in Z_n\}$  is the set of class nodes corresponding to parents of the  $n$ th node. For a leaf node  $n$ ,  $Z_n = \emptyset$ , and  $\Pr(y_n | M_n) \equiv \Pr(y_n)$ .

$$\begin{aligned} \Pr(\mathbf{X}, \mathbf{Y}) &= \Pr(\mathbf{X} | \mathbf{Y}) \Pr(\mathbf{Y}) \\ &= \prod_{n=1}^N \Pr(\mathbf{x}_n | y_n) \prod_{n=1}^N \Pr(y_n | M_n) \end{aligned} \quad (1)$$

Specifically, the conditional probability of a sample feature vector given its class can be assumed independent and identically distributed (i.i.d.) Gaussian for simplicity, as shown in Equation (2), where  $\boldsymbol{\mu}_{y_n}$  and  $\boldsymbol{\Sigma}_{y_n}$  are the mean and covariance matrix of feature vector  $\mathbf{x}_n$  for class  $y_n$  ( $y_n = 0, 1$  with 0 for dry, and 1 for flood). It is worth noting that  $\Pr(\mathbf{x}_n | y_n)$  could be more general than i.i.d. Gaussian. This conditional probability captures the local likelihood of a pixel's class (flood or dry) based on its own spectral features. For example, pixels corresponding to exposed flood water surface often look brown in color. However, for pixels corresponding to noise and obstacles (e.g., tree canopies), their spectral colors may not indicate the true underlying classes. That is why we need to use the class transitional probability to capture the dependency between pixel classes based on topography.

$$\Pr(\mathbf{x}_n | y_n) \sim \mathcal{N}(\boldsymbol{\mu}_{y_n}, \boldsymbol{\Sigma}_{y_n}) \quad (2)$$

Class transitional probability follows the partial order constraint from topography. For example, due to gravity, if any parent node's class is *dry*, the child's class must be *dry*; if all parents' classes are *flood*, then the child has a high probability of being *flood*. For example, in Figure 2, if node 5 is flood, then nodes 2 to 4 and 6 to 7 must also be flood. Consider *flood* as the positive class (class 1) and *dry* as the negative class (class value 0), the transitional probability is actually conditioned on the product of parent classes  $y_{Z_n} \equiv \prod_{k \in Z_n} y_k$ . The specific transitional probability is expressed in Table 2, where  $\rho$  is the probability of a child node being class 1 (flood) given all parents in class 1 (flood). Due to spatial autocorrelation, the value of  $\rho$  is very high (close to 1). For a leaf node  $n$ ,  $Z_n = \emptyset$ . The transitional probability is degraded into simple class prior probability  $\Pr(y_n | M_n) \equiv \Pr(y_n)$  as expressed on the right of Table 3, where  $\pi$  is the prior probability of  $y_n$  being in class 1. The class transitional probability is a key component in the HMT model. The local likelihood of individual pixels based on their own spectral features can be misleading due to noise, obstacles, and spectral confusion. The class transitional probability provides additional dependency constraints between

Table 2.: Class transition probability  $\Pr(y_n|y_{Z_n})$ 

|           | $y_{Z_n} = 0$ | $y_{Z_n} = 1$ |
|-----------|---------------|---------------|
| $y_n = 0$ | 1             | $1 - \rho$    |
| $y_n = 1$ | 0             | $\rho$        |

Table 3.: Class prior probability

|           | $\Pr(y_n)$ |
|-----------|------------|
| $y_n = 0$ | $1 - \pi$  |
| $y_n = 1$ | $\pi$      |

pixels' classes based on topography. In this case, even if some pixels' local class likelihood is wrongly estimated, it can still be corrected by other good pixels (whose local class likelihood is more trustworthy).

The parameters of HMT include the mean and covariance matrix of sample features in each class, prior probability of leaf node classes, and class transition probability for non-leaf nodes. We denote the entire set of parameters as  $\Theta = \{\rho, \pi, \mu_c, \Sigma_c | c = 0, 1\}$ . Learning the set of parameters poses two major challenges: first, there exist unknown hidden class variables  $\mathbf{Y} = [y_1, \dots, y_N]^T$  (T refers to Transpose), which are non-i.i.d.; second, the number of samples (nodes) is huge (up to hundreds of millions of pixels).

To address these challenges, the EM algorithm and message (belief) propagation can be used. The EM-based approach has the following major steps:

- (1) Initialize parameter set  $\Theta_0$
- (2) Compute posterior distribution of hidden classes:  $\Pr(\mathbf{Y}|\mathbf{X}, \Theta_0)$
- (3) Compute posterior expectation of log likelihood:  
 $LL(\Theta) = \mathbb{E}_{\mathbf{Y}|\mathbf{X}, \Theta_0} \log \Pr(\mathbf{X}, \mathbf{Y}|\Theta)$
- (4) Update parameters:  $\Theta_0 \leftarrow \arg \max_{\Theta} LL(\Theta)$   
Return  $\Theta_0$  if it's converged (no updates), otherwise go back to (2)

Among the four steps above, step (2) that computes the joint posterior distribution of all sample classes  $\Pr(\mathbf{Y}|\mathbf{X}, \Theta_0)$  is practically infeasible due to the large number of hidden class nodes that are non-i.i.d. Fortunately, it is not necessary to compute the entire joint posterior distribution of all sample classes. In fact, we only need the marginal posterior distribution of a node's and its parents' classes for non-leaf nodes, as well as the marginal posterior distribution of a node's class for leaf nodes. The reason can be explained through the expression of the posterior expectation of log likelihood in Equation (3). Note that for leaf node,  $Z_n = \emptyset$ , and the last term in the last line of above equation is degraded, i.e.,  $\log \Pr(y_n|M_n, \Theta)\Pr(y_n, M_n|\mathbf{X}, \Theta_0) = \log \Pr(y_n|\Theta)\Pr(y_n|\mathbf{X}, \Theta_0)$ . It can be seen that the joint likelihood can be decomposed into local factors only related to an individual node  $n$  and its parents  $Z_n$ . To compute the expected log likelihood expression, we first need to compute

the local marginal posterior distribution such as  $\Pr(y_n|\mathbf{X}, \Theta_0)$  and  $\Pr(y_n, M_n|\mathbf{X}, \Theta_0)$ .

$$\begin{aligned}
LL(\Theta) &= \mathbb{E}_{\mathbf{Y}|\mathbf{X}, \Theta_0} \log \Pr(\mathbf{X}, \mathbf{Y}|\Theta) \\
&= \mathbb{E}_{\mathbf{Y}|\mathbf{X}, \Theta_0} \log \left\{ \prod_{n=1}^N \Pr(\mathbf{x}_n|y_n, \Theta) \prod_{n=1}^N \Pr(y_n|M_n, \Theta) \right\} \\
&= \sum_{\mathbf{Y}} \Pr(\mathbf{Y}|\mathbf{X}, \Theta_0) \sum_{n=1}^N \{ \log \Pr(\mathbf{x}_n|y_n, \Theta) + \log \Pr(y_n|M_n, \Theta) \} \\
&= \sum_{n=1}^N \sum_{y_n} \Pr(y_n|\mathbf{X}, \Theta_0) \log \Pr(\mathbf{x}_n|y_n, \Theta) \\
&\quad + \sum_{n=1}^N \sum_{y_n, M_n} \Pr(y_n, M_n|\mathbf{X}, \Theta_0) \log \Pr(y_n|M_n, \Theta)
\end{aligned} \tag{3}$$

To compute the marginal posterior distribution  $\Pr(y_n, M_n)$  and  $\Pr(y_n)$  (we omit the condition on  $\mathbf{X}$  and  $\Theta_0$  for brevity), we can use the message propagation method based on the sum and product algorithm. Due to space limit, we do not provide the detailed math formula. The intuition of message propagation along graph (or tree) nodes is a process of marginalizing out those corresponding node variables in a joint distribution. The marginalization process is done over a tree by a traversal order to reduce redundant computation. More details can be found in Kschischang, Frey, and Loeliger (2001); Ronen, Rohlicek, and Ostendorf (1995).

After learning model parameters, we can infer hidden class variables by maximizing the overall probability. A naive approach that enumerate all combinations of class assignment is infeasible due to the exponential cost. We use a dynamic programming based method called *max-sum* from Rabiner (1989). The process is similar to the sum and product algorithm above. The main difference is that instead of using sum operation, we need to use max operation in message propagation, and also memorize the optimal variable values. We omit the details due to space limit. The main intuition is that in class inference, the class label for each node is not only based on the likelihood of its own spectral features, but also other nodes' class labels through transitional probability. In other words, the class labels of all nodes are jointly inferred based on the topography constraint.

Note that our HMT model uses transductive learning, i.e., the model is trained based on not only the features and labels from the training pixels but also the features from the test samples. This is reflected by the fact that the model will create a flow dependency tree based on the topography of the test region. Thus, the model needs to be re-trained for each specific test region. This is different from many existing machine learning algorithms based on inductive learning: once a general model is learned from training pixels, it can be applied to any test set instead of only being applied to a particular test set. Since the computational structure of both the parameter learning and class inference is largely based on tree traversal operations, the time cost is linear to the number of tree nodes. The model training is computationally very fast.

### 3. Results

In this section, we compared the HMT model with several baseline methods in classification performance on three real-world datasets. Unless specified otherwise, we used default parameters in open source tools for baseline methods. Candidate classification methods in-

clude:

- **Non-spatial classifiers:** We tested random forest (**RF**) and gradient boosted tree (**GBM**) in R packages on spectral features (red, green, blue spectral bands).
- **Non-spatial classifier with post-processing label propagation (LP):** We tested **RF** and **GBM** together with label propagation smoothing using 4 neighborhood Zhu and Ghahramani (2002). Label propagation was added on top of the pre-classified map from a classifier through updates based on neighborhood majority.
- **HMT:** We used the model from Xie, Jiang, and Sainju (2018) implemented in C++. We added an additional hidden class layer to accommodate the large scale of tree canopies overlaying flood water.

The evaluation metrics we used include precision ( $P$ ) (user’s accuracy), recall ( $R$ ) (producer’s accuracy), and  $F$ -score ( $F_1$ ) (the harmonic mean between precision and recall). The definition of the metrics are provided in Equation 4, Equation 5 and Equation 6, where TP, FP, and FN are the numbers of true positives (TP) (i.e., pixels truly in the positive class and correctly predicted into the positive class), false positives (FP) (i.e., pixels truly in the negative class but mistakenly predicted into the positive class) and false negatives (FN) (i.e., pixels truly in the positive class but mistakenly predicted into the negative class) respectively. A high precision on a class means that among the pixels predicted into the class, most predictions are correct. A high recall on a class means that among the pixels that are actually in the class category, most of them are correctly identified in the prediction. Quantitative evaluation results on classification accuracy was based on the predictions on randomly sampled pixels from labeled test polygons.

$$P = \frac{(TP)}{(TP)+(FP)} \quad (4)$$

$$R = \frac{(TP)}{(TP)+(FN)} \quad (5)$$

$$F_1 = 2 \times \frac{R \times P}{R + P} \quad (6)$$

### 3.1. Classification Performance Comparison

The comparison of classification performance between different methods on the three datasets are summarized in Table 4, Table 5, and Table 6 respectively.

From the results on Grimesland dataset in Table 4, we can see that traditional classification models such as RF and GBM performance poorly (with overall  $F_1$  scores of 0.6 and 0.68). The GBM is slightly better than RF. Within each method, the precision for the flood class is very high (0.95 and 0.9), meaning that the flood pixels being identified are largely correct. These correctly predicted flood pixels are mostly open exposed flood surface. However, the recall of the flood class is very low (0.42 and 0.59), meaning that there are a significant amount (around half) of flood pixels not being identified. The large amount of actual flood pixels being missed in the prediction results are largely due to the obstacles in the image (e.g., tree canopies). The spectral signature of the tree canopies overlaying flood water is very close to that of the trees in the dry area, and thus they are mistakenly classified into the dry class (though there are flood water beneath). Adding post-processing through label propagation

Table 4.: Classification on real dataset in Grimesland NC

| Classifier | Class | $P$  | $R$  | $F_1$ | Average $F_1$ |
|------------|-------|------|------|-------|---------------|
| RF         | Dry   | 0.45 | 0.96 | 0.61  | 0.60          |
|            | Flood | 0.95 | 0.42 | 0.58  |               |
| GBM        | Dry   | 0.51 | 0.86 | 0.64  | 0.68          |
|            | Flood | 0.90 | 0.59 | 0.71  |               |
| RF+LP      | Dry   | 0.45 | 0.99 | 0.62  | 0.59          |
|            | Flood | 0.99 | 0.40 | 0.57  |               |
| GBM+LP     | Dry   | 0.53 | 0.97 | 0.68  | 0.70          |
|            | Flood | 0.97 | 0.57 | 0.72  |               |
| HMT        | Dry   | 0.99 | 0.98 | 0.99  | 0.99          |
|            | Flood | 0.99 | 0.99 | 0.99  |               |

Table 5.: Classification on real dataset in Kinston NC: Area 1

| Classifier | Class | $P$  | $R$  | $F_1$ | Average $F_1$ |
|------------|-------|------|------|-------|---------------|
| RF         | Dry   | 0.84 | 0.96 | 0.90  | 0.79          |
|            | Flood | 0.87 | 0.57 | 0.69  |               |
| GBM        | Dry   | 0.86 | 0.88 | 0.87  | 0.79          |
|            | Flood | 0.72 | 0.67 | 0.70  |               |
| RF+LP      | Dry   | 0.84 | 0.99 | 0.91  | 0.82          |
|            | Flood | 0.97 | 0.58 | 0.72  |               |
| GBM+LP     | Dry   | 0.87 | 0.96 | 0.91  | 0.83          |
|            | Flood | 0.87 | 0.66 | 0.75  |               |
| HMT        | Dry   | 1.00 | 0.94 | 0.96  | 0.95          |
|            | Flood | 0.87 | 0.99 | 0.93  |               |

Table 6.: Classification on real dataset in Kinston NC: Area 2

| Classifier | Class | $P$  | $R$  | $F_1$ | Average $F_1$ |
|------------|-------|------|------|-------|---------------|
| RF         | Dry   | 0.57 | 0.96 | 0.72  | 0.61          |
|            | Flood | 0.90 | 0.36 | 0.51  |               |
| GBM        | Dry   | 0.61 | 0.84 | 0.72  | 0.67          |
|            | Flood | 0.83 | 0.50 | 0.62  |               |
| RF+LP      | Dry   | 0.56 | 0.98 | 0.71  | 0.59          |
|            | Flood | 0.94 | 0.31 | 0.46  |               |
| GBM+LP     | Dry   | 0.62 | 0.95 | 0.75  | 0.69          |
|            | Flood | 0.91 | 0.48 | 0.63  |               |
| HMT        | Dry   | 1.00 | 0.92 | 0.96  | 0.96          |
|            | Flood | 0.93 | 1.00 | 0.97  |               |

slightly decreases the  $F_1$  of RF (from 0.6 to 0.59) but slightly increases the  $F_1$  of GBM (from 0.68 to 0.7). The reason is that neighborhood smoothing could reduce salt-and-pepper noise errors, but also could mistakenly change a correctly predicted pixel if most of the neighbors were mis-classified. In contrast, the HMT model performed the best, with an  $F_1$  around 0.99. It almost perfectly identified the flood extent. The results of the Area 2 in the city of Kinston are summarized in Table 6. The trends look very close to the results in Table 4.

The results on the dataset collected from the Area 1 in the city of Kinston are summarized in Table 5. The overall comparison between methods looks similar to the results on the other two datasets. Some differences exist though. The classification performance of non-spatial classifiers is generally better compared with the results on the other datasets, with the  $F_1$  scores of RF and GBM around 0.8. The precision on the flood class is generally worse than other datasets, while the recalls is generally better. This means that the non-spatial models identified more flood pixels on this dataset compared with the other datasets, but the accuracy is lower among those flood pixels being identified. We also observe that HMT performed slightly worse on this dataset than other dataset, largely due to a lower precision on the flood class (0.87). HMT is still significantly more accurate than the other methods on this dataset.

### 3.2. Interpretation of Results

We now provide detailed interpretation of classification results through visualized flood maps from different methods. The input aerial imagery and digital elevation model, as well as classified flood maps from methods on three datasets are shown in Figure 4, Figure 5, and Figure 6 respectively.

Figure 4 shows the classified flood maps on the Grimesland dataset (flood is shown in blue while dry is shown in brown). The study area is shown in Figure 4 (a) and the input aerial imagery with red, green, and blue bands are shown in Figure 4 (b). The elevation map is shown in Figure 4 (c), with black color indicating low elevation and white color indicating high elevation. From the aerial imagery, we can found that the central part of the region is largely flooded (black and brown colors). We can also observe that there are a significant amount of vegetation (tree canopies in green color) among the flooded area. These tree canopies have the same color and texture (spectral signature) as the tree canopies in the dry area. Moreover, the aerial imagery has various illumination condition (the same flood surface shows discontinuity in color tones). All these issues pose significant challenges in identifying the correct flood extent. From Figure 4 (d) - (e), we can see that RF and GBM successfully identified part of exposed flood surface on the lower right part of the region. There are some part of exposed flood surface in the lower left part not being identified, probably because the colors of exposed flood pixels in two parts are different due to varying illumination. We also observed that the two models almost all missed the flooded pixels obscured by significant tree canopies. This is because the spectral signatures of the tree canopies overlaying flood water are very close to those signatures in dry areas. This observation is consistent with the low recall observed in Table 4. In addition, we observe a significant amount of salt-and-pepper noise in the classified maps, due to the fact that both models classify each pixel independently (per-pixel classification) without incorporating the spatial autocorrelation effect. Figure 4 (f) - (g) show the results after conducting post-processing through neighborhood label propagation. We can see that post-processing can significantly reduce the amount of salt-and-pepper noise due to the incorporation of neighborhood context. But post-processing with neighborhood level label propagation could not address the errors in missing a large area of flood areas under tree canopies because the errors are systematic beyond a neighborhood scale. Finally, we can see from Figure 4 (h) that HMT successfully identified most of the flooded areas, including the ar-

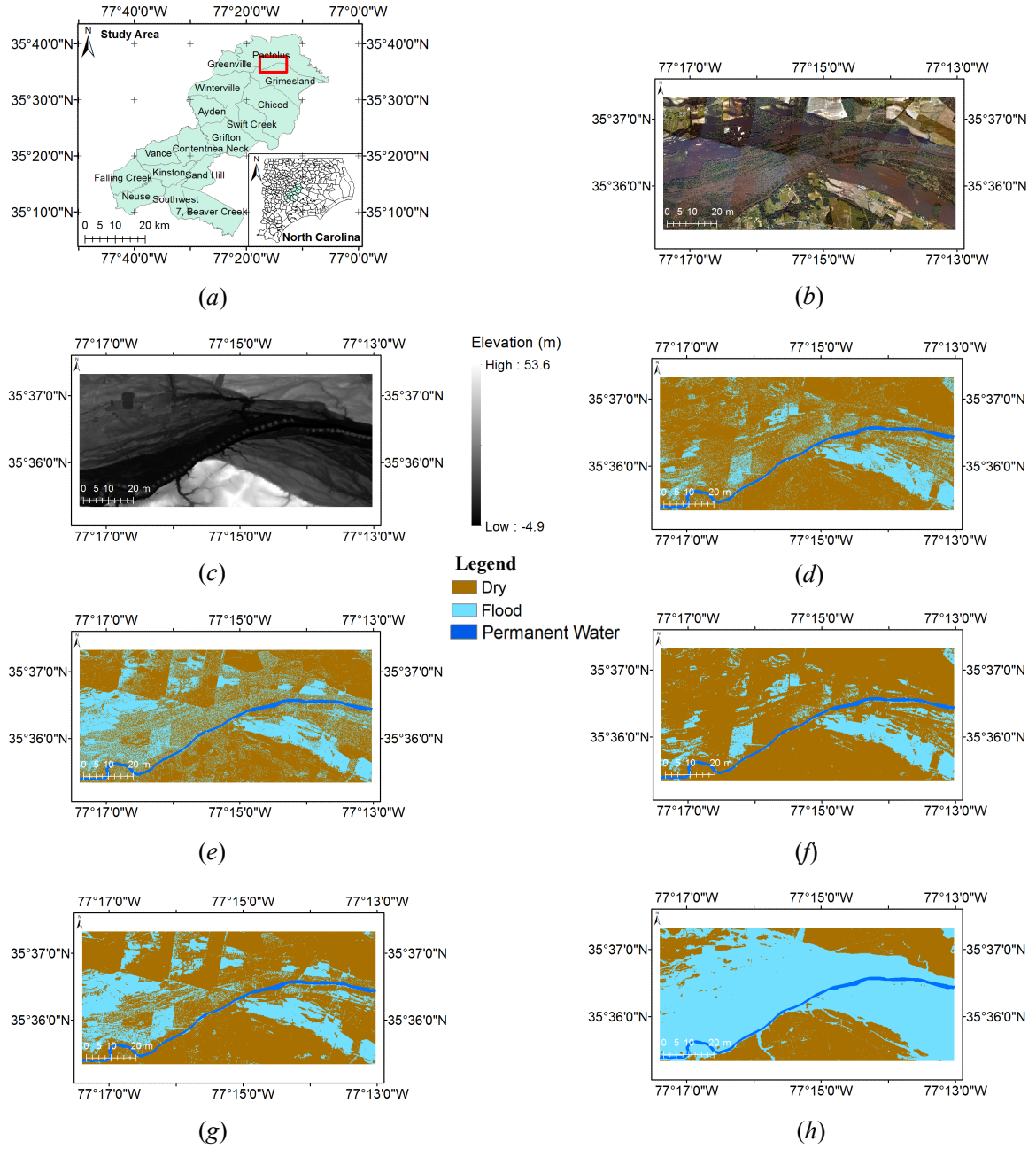


Figure 4.: Results on Matthew flood, Grimesland, NC. (a) Study area. (b) High-resolution aerial imagery in Grimesland (Source: NOAA National Geodesic Survey, Date: 15 October 2016). (c) DEM (m) (d) RF result (e) GBM result (f) RF+LP result (g) GBM+LP result (h) HMT result

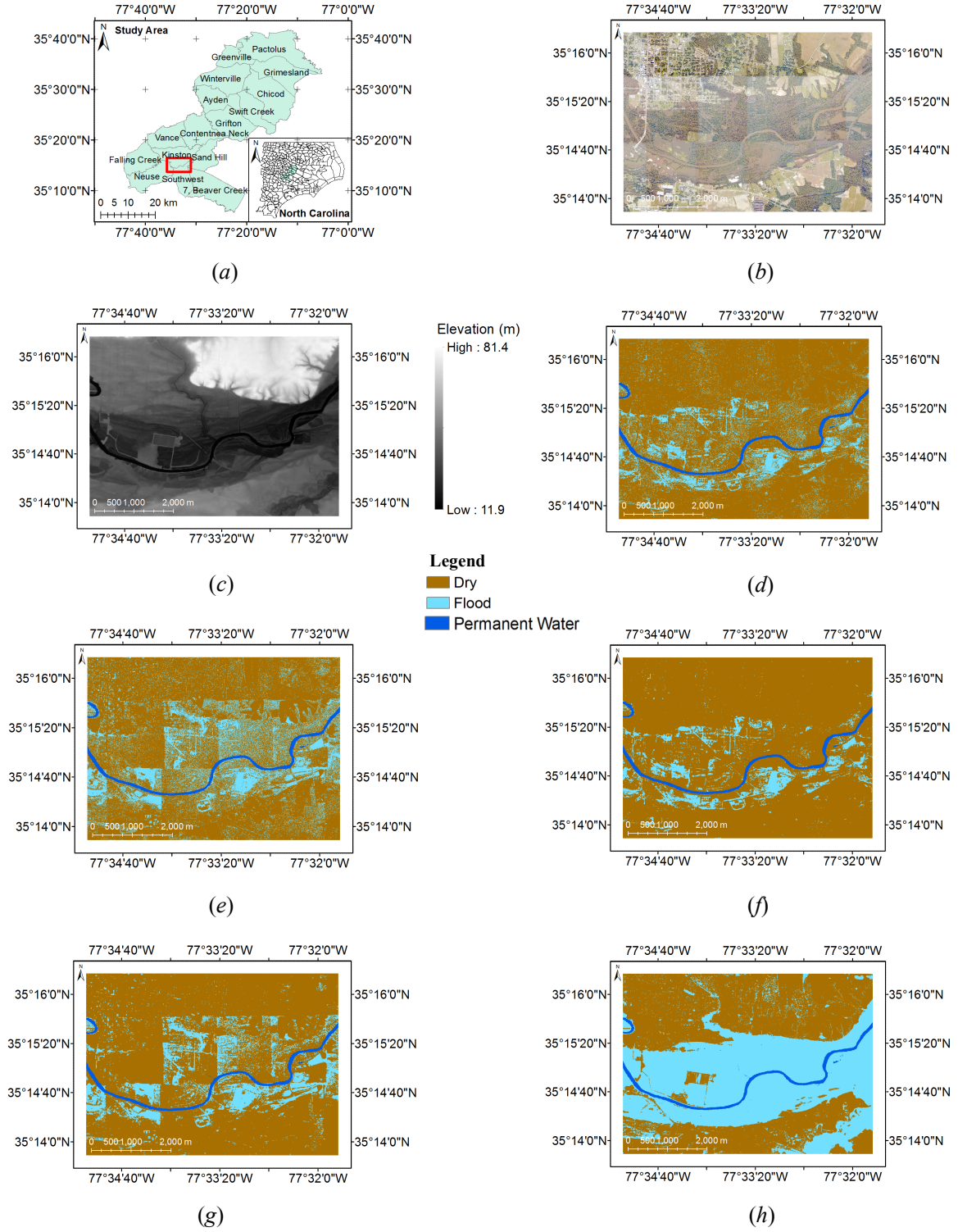


Figure 5.: Results on Matthew flood, Area 1 of Kinston, NC. (a) Study area. (b) High-resolution aerial imagery in Kinston: Area 1 (Source: NOAA National Geodesic Survey, Date: 15 October 2016). (c) DEM (m) (d) RF result (e) GBM result (f) RF+LP result (g) GBM+LP result (h) HMT result

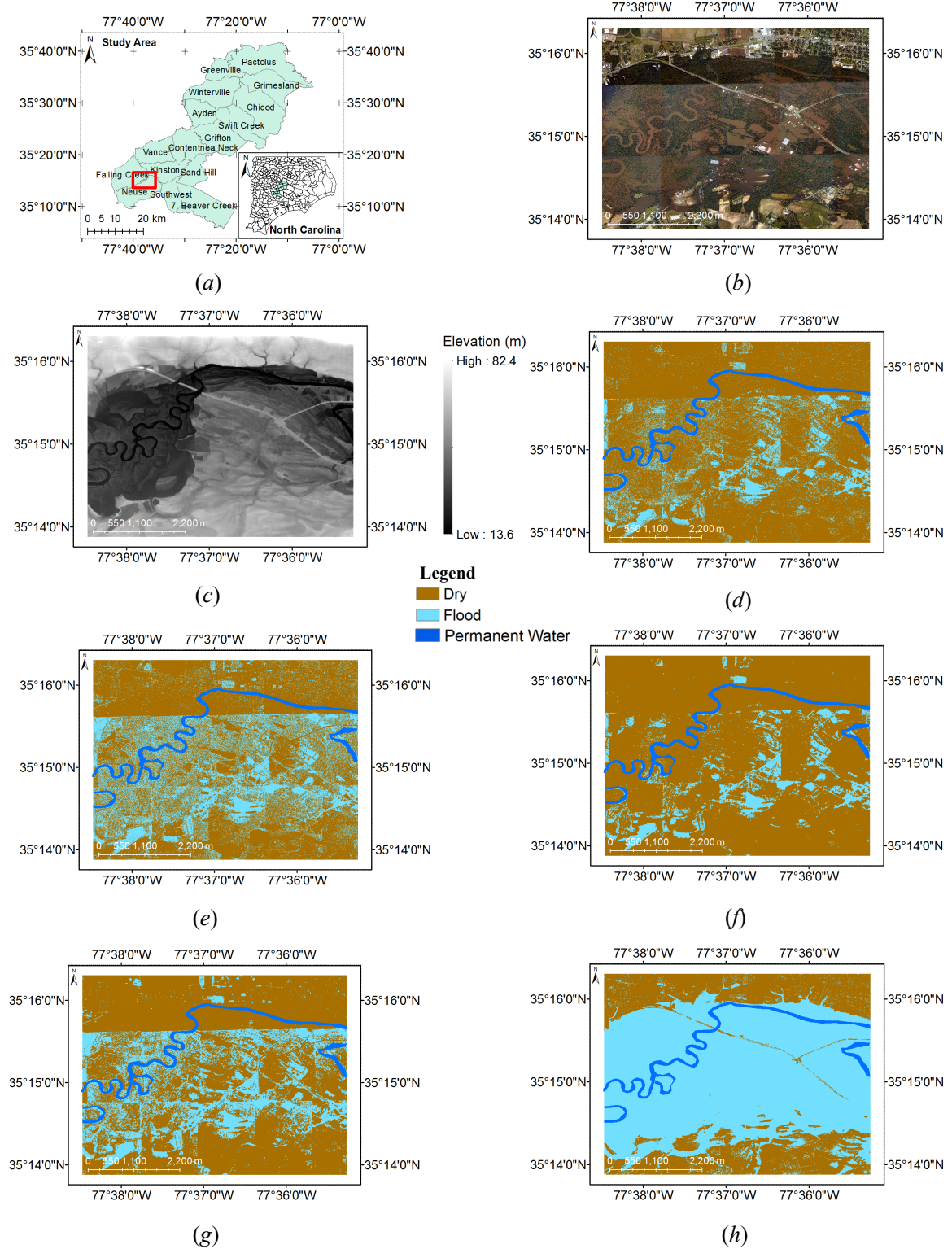


Figure 6.: Results on Matthew flood, Area 2 of Kinston, NC. (a) Study area. (b) High-resolution aerial imagery in Kinston: Area 2 (Source: NOAA National Geodesic Survey, Date: 15 October 2016). (c) DEM (m) (d) RF result (e) GBM result (f) RF+LP result (g) GBM+LP result (h) HMT result

eas under tree canopies. The reason is that HMT incorporate the spatial structural dependency between pixel locations based on the topography. For example, the exposed flood surface on the lower right part has relatively higher elevation on the elevation map in Figure 4 (c). These pixels have a higher elevation than the other flooded areas obscured by tree canopies. Thus, the model is able to fill the gap of missed flood areas under tree canopies since the topography indicates water flow directions.

The results on Area 1 of Kinston are shown in Figure 5. The study area is shown in Figure 5 (a). From the aerial imagery and elevation map in Figure 5 (b) - (c), we can observe that the central part of the region is flooded (with brown flood water). There are a significant amount of tree canopies overlaying flood water. The upper and lower parts are mostly dry. There are also varying illuminations within the region (similar to the Grimesland dataset). This is shown in several vertical strips in the aerial images. From the classified maps of RF and GBM, we can observe a significant amount of salt-and-pepper noise and missed flood areas due to tree canopies as well as spectral confusion due to varying illumination. Adding post-processing reduces the salt-and-pepper noise, as shown in Figure 5 (f) - (g). Finally, the HMT model identifies more flood areas with tree canopies.

The results on Area 2 of Kinston are shown in Figure 6. The study area is shown in Figure 6 (a). From the aerial image and the elevation map, we can see similar challenges to the other two datasets. What is more significant in this region is the varying illumination between different strips of the aerial imagery. We can observe a clear horizontal line separating the top part (much darker color) from the central part (lighter color). This is also reflected by the artifacts (horizontal flood boundary) in the classified flood maps from RF and GBM (Figure 6 (d) - (e)). The results of the two non-spatial models miss a significant amount of flood areas, but GBM is slightly better in identifying flood. This is consistent with the low recalls of the two methods in Table 6 (with recall slightly higher in GBM). Post-processing reduces salt-and-pepper noise, but still could not correct the missing flood areas. In contrast, HMT can identify most of the flood areas due to incorporating the topography.

Note that the HMT prediction output is the entire inundation extent, including the perennial water surface. Perennial water can be separated from the flood water through further post-processing, e.g., change detection. Our reported classification accuracy metrics were based on test polygons that were drawn from the flood area outside the perennial water.

### 3.3. Computational Time Costs

We also conducted computational experiments on the three test regions to measure the time costs of the tree construction, parameter learning and class inference process. Experiments were conducted on a Dell workstation with Intel(R) Xeon(R) Central Processing Unit (CPU) E5-2687w v4 @ 3.00GHz, 64GB main memory, and Windows 10. As mentioned in the earlier, our HMT model was implemented in C++. The codes ran sequentially on one CPU core only. Table 7 summarizes the results. The number of pixels in three test regions are around 7.6 million, 7.4 million and 7.7 million respectively. The tree construction part only took around 3 seconds. The class inference was also very fast, only taking less than 2 seconds for over 7 million pixels. The parameter learning part contributed to the vast majority of the time costs (around 260 seconds). The reason is that the parameter learning step involves multiple iterations, each of which has message propagation over the entire tree. Overall, the HMT algorithm was highly efficient. It could process over 7 million pixels within 5 minutes. Please note that we did not compare the computational time costs of HMT with those of the baseline models such as GBM and RF, since these baseline models were implemented in a different programming language (R instead of C++) and they did not capture the topography

in the classification process.

Table 7.: Computational time costs for three study areas

| Study area | Size (pixels)      | Tree construction (s) | Learning (s) | Inference (s) |
|------------|--------------------|-----------------------|--------------|---------------|
| Grimesland | $1856 \times 4104$ | 2.88                  | 264.57       | 1.94          |
| Kinston 1  | $2313 \times 3207$ | 3.00                  | 257.75       | 1.73          |
| Kinston 2  | $2418 \times 3205$ | 3.26                  | 268.25       | 1.93          |

#### 4. Discussions

This paper introduces a machine learning approach called hidden Markov tree for flood extent mapping from high-resolution optical imagery and DEM data. Existing image classification algorithms for flood mapping often uses spectral and spatial features at the pixel or neighborhood level, and thus cannot fully address the issues of noise and large scale obstacles (e.g., a large area of vegetation over flood water). In contrast, HMT model captures the topography constraint of water flow directions over the entire study area. Water flow directions are incorporated as a directed tree structure in a hidden class layer. In this way, the order of class labels of pixels (flood or dry) are constrained by the topography. Thus, the model can automatically infer the class labels of pixels not only based on its local (or neighborhood) spectral signatures, but also based on other pixels information in a topography surface. Results show that the model can fill the gap of missing flood extent due to dense vegetation. There exist other ways of incorporating topography into flood extent mapping, e.g., adjusting thresholds on elevation map to find the best matching of flood extent (e.g., Wang, Colby, and Mulcahy (2002)). But the HMT model can add value in learning and inferring class labels automatically without the need of manually tuning the thresholds. The HMT model provides a clear statistical framework that makes the model easily interpretable. In addition, the HMT model is also very fast due to the linear time complexity. Classifying the entire area with over 7 million pixels took just a few minutes.

One limitation of the HMT approach is the flow dependency tree derived from the DEM data is quite simple. It assumes that the inundation surface is flat (with equal elevation). Such assumption is valid for river floodplain with relatively flat topography. When the study area is a big watershed with complex inters of flow direction, the flow dependency tree structure in HMT needs to be re-designed to reflect the complex topography (i.e., to allow for non-flat flood surface). Theoretically, the HMT model can work with any flow direction dependency structure as long as it can be represented in a tree or polytree. Otherwise, we may choose to divide the big watershed into smaller pieces such that the flood surface in each piece can be considered as flat. In future work, we plan to address the limitation by integrating a more sophisticated topography model into HMT for vast watershed areas.

#### 5. Conclusion and Future Work

This paper conducts several case studies on using a spatial machine learning model called hidden Markov tree for flood extent mapping based on high-resolution aerial imagery and DEM. The model can automatically incorporate topography constraint (water flow directions) in the image classification process in an efficient manner. Incorporating topography constraint is critically important in addressing noise and obstacles (e.g., tree canopies blocking the view

of flood water) in high-resolution imagery. Results show that the HMT model significantly improves several existing methods in classification accuracy (with  $F_1$  increasing from around 70% to 80% to over 95%) on three study areas in hurricane Matthew floods in North Carolina 2016.

In future work, we plan to further test the generalizability of the HMT model on datasets collected by different remote sensors (e.g., radar image, MODIS or Landsat imagery) in different flood events and study areas. The HMT model can also be potentially improved by adding morphological characters of a watershed (in addition to spectral bands from optical imagery) as model inputs. We plan to evaluate the HMT approach in refining pre-classified flood maps from other sensors (e.g., MODIS Flood Water maps, which have coarser spatial resolution and contain gaps due to feature obstacles such as tree canopies). We will also explore the integration of the HMT with the recent deep learning models.

## Acknowledgement

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. IIS-1850546, IIS-2008973, CNS-1951974 and the University Corporation for Atmospheric Research (UCAR).

## Declaration of Interest

No potential conflict of interest was reported by the authors.

## References

- Cline, Don. 2009. *Integrated Water Resources Science and Services: an Integrated and Adaptive Roadmap for Operational Implementation*. Technical Report. National Oceanic and Atmospheric Administration.
- Eftelioglu, Emre, Zhe Jiang, Reem Ali, and Shashi Shekhar. 2016. "Spatial computing perspective on food energy and water nexus." *Journal of Environmental Studies and Sciences* 6 (1): 62–76.
- Fayne, Jessica V, John D Bolten, Colin S Doyle, Sven Fuhrmann, Matthew T Rice, Paul R Houser, and Venkat Lakshmi. 2017. "Flood mapping in the lower Mekong River Basin using daily MODIS observations." *International journal of remote sensing* 38 (6): 1737–1757.
- Feng, Quanlong, Jiantao Liu, and Jianhua Gong. 2015. "Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—A case of Yuyao, China." *Water* 7 (4): 1437–1455.
- Giustarini, Laura, Renaud Hostache, Patrick Matgen, Guy J-P Schumann, Paul D Bates, and David C Mason. 2012. "A change detection approach to flood mapping in urban areas using TerraSAR-X." *IEEE transactions on Geoscience and Remote Sensing* 51 (4): 2417–2430.
- Horritt, MS, DC Mason, DM Cobby, IJ Davenport, and PD Bates. 2003. "Waterline mapping in flooded vegetation from airborne SAR imagery." *Remote Sensing of Environment* 85 (3): 271–281.
- Jiang, Zhe. 2018. "A survey on spatial prediction methods." *IEEE Transactions on Knowledge and Data Engineering* 31 (9): 1645–1664.
- Jiang, Zhe. 2020. "Spatial Structured Prediction Models: Applications, Challenges, and Techniques." *IEEE Access* 8: 38714–38727.
- Jiang, Zhe, Yan Li, Shashi Shekhar, and Joseph Knight. 2019. "Spatial ensemble learning for heterogeneous geographic data with class ambiguity." *ACM Transactions on Intelligent Systems and Technology*.

- Jiang, Zhe, Yan Li, Shashi Shekhar, Lian Rampi, and Joseph Knight. 2017. "Spatial ensemble learning for heterogeneous geographic data with class ambiguity: A summary of results." In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 23. ACM.
- Jiang, Zhe, and Arpan Man Sainju. 2019. "Hidden Markov Contour Tree: A Spatial Structured Model for Hydrological Applications." In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 804–813.
- Jiang, Zhe, and Shashi Shekhar. 2017. *Spatial big data science*. Springer.
- Jiang, Zhe, Shashi Shekhar, Pradeep Mohan, Joseph Knight, and Jennifer Corcoran. 2012. "Learning spatial decision tree for geographical classification: a summary of results." In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, 390–393. ACM.
- Jiang, Zhe, Shashi Shekhar, Xun Zhou, Joseph Knight, and Jennifer Corcoran. 2013. "Focal-test-based spatial decision tree learning: A summary of results." In *2013 IEEE 13th International Conference on Data Mining*, 320–329. IEEE.
- Jiang, Zhe, Shashi Shekhar, Xun Zhou, Joseph Knight, and Jennifer Corcoran. 2015. "Focal-test-based spatial decision tree learning." *IEEE Transactions on Knowledge and Data Engineering* 27 (6): 1547–1559.
- Jiang, Zhe, Miao Xie, and Arpan Man Sainju. 2019. "Geographical Hidden Markov Tree." *IEEE Transactions on Knowledge and Data Engineering*.
- Karpatne, Anuj, Zhe Jiang, Ranga Raju Vatsavai, Shashi Shekhar, and Vipin Kumar. 2016. "Monitoring land-cover changes: A machine-learning perspective." *IEEE Geoscience and Remote Sensing Magazine* 4 (2): 8–21.
- Kschischang, Frank R, Brendan J Frey, and H-A Loeliger. 2001. "Factor graphs and the sum-product algorithm." *IEEE Transactions on information theory* 47 (2): 498–519.
- Lin, Li, Liping Di, Junmei Tang, Eugene Yu, Chen Zhang, Md Rahman, Ranjay Shrestha, Lingjun Kang, et al. 2019. "Improvement and validation of NASA/MODIS NRT global flood mapping." *Remote Sensing* 11 (2): 205.
- Matgen, Patrick, G Schumann, J-B Henry, Lucien Hoffmann, and Laurent Pfister. 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.
- McFeeters, Stuart K. 1996. "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features." *International journal of remote sensing* 17 (7): 1425–1432.
- National Oceanic and Atmospheric Administration. 2017. "Data and Imagery from NOAA's National Geodetic Survey." <https://www.ngs.noaa.gov>.
- National Oceanic and Atmospheric Administration. 2018. "National Water Model: Improving NOAA's Water Prediction Services." <http://water.noaa.gov/documents/wrn-national-water-model.pdf>.
- North Carolina State University Library. 2018. "LIDAR Based Elevation Data for North Carolina." <https://www.lib.ncsu.edu/gis/elevation>.
- Oberstadler, R, H Hönsch, and D Huth. 1997. "Assessment of the mapping capabilities of ERS-1 SAR data for flood mapping: a case study in Germany." *Hydrological processes* 11 (10): 1415–1425.
- Rabiner, Lawrence R. 1989. "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE* 77 (2): 257–286.
- Rahman, Md, Liping Di, Eugene Yu, Li Lin, Chen Zhang, Junmei Tang, et al. 2019. "Rapid flood progress monitoring in cropland with NASA SMAP." *Remote Sensing* 11 (2): 191.
- Rahman, Md Shahinoor, and Liping Di. 2017. "The state of the art of spaceborne remote sensing in flood management." *Natural Hazards* 85 (2): 1223–1248.
- Ronen, O, JR Rohlicek, and M Ostendorf. 1995. "Parameter estimation of dependence tree models using the EM algorithm." *IEEE Signal Processing Letters* 2 (8): 157–159.
- Sainju, Arpan Man, Wenchong He, Zhe Jiang, and Da Yan. 2020. "Spatial classification with limited observations based on physics-aware structural constraint." In *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 1–8.
- Shekhar, Shashi, Zhe Jiang, Reem Ali, Emre Eftelioglu, Xun Tang, Venkata Gunturi, and Xun Zhou.

2015. "Spatiotemporal data mining: a computational perspective." *ISPRS International Journal of Geo-Information* 4 (4): 2306–2338.
- Shen, Xinyi, Dacheng Wang, Kebiao Mao, Emmanouil Anagnostou, and Yang Hong. 2019. "Inundation extent mapping by synthetic aperture radar: a review." *Remote Sensing* 11 (7): 879.
- Sun, Donglian, Yunyue Yu, and Mitchell D Goldberg. 2011. "Deriving water fraction and flood maps from MODIS images using a decision tree approach." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4 (4): 814–825.
- Townsend, Philip A. 2001. "Mapping seasonal flooding in forested wetlands using multi-temporal Radarsat SAR." *Photogrammetric engineering and remote sensing* 67 (7): 857–864.
- Wang, Y, JD Colby, and KA Mulcahy. 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.
- Xie, Miao, Zhe Jiang, and Arpan Man Sainju. 2018. "Geographical Hidden Markov Tree for Flood Extent Mapping." In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18*, 2545–2554. ACM.
- Xu, Hanqiu. 2006. "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery." *International journal of remote sensing* 27 (14): 3025–3033.
- Yang, Wen, Hui Song, Xiaojing Huang, Xin Xu, and Mingsheng Liao. 2014. "Change detection in high-resolution SAR images based on Jensen–Shannon divergence and hierarchical Markov model." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (8): 3318–3327.
- Zhang, Fang, Xiaolin Zhu, and Desheng Liu. 2014. "Blending MODIS and Landsat images for urban flood mapping." *International journal of remote sensing* 35 (9): 3237–3253.
- Zhu, Xiaojin, and Zoubin Ghahramani. 2002. "Learning from labeled and unlabeled data with label propagation."

## List of Figures

- |   |                                                                                                                                                                                                                                                                                                                      |    |
|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1 | The map of three study areas (Source: NOAA NGS, Date: 15 October 2016). (a) North Carolina. (b) Study area. (c) Grimesland test region. (d) Kinston region. The same training set in Kinston region is used for three test scenes (test area 1 and test area 2 in Kinston region as well as Grimesland test region). | 4  |
| 2 | Illustration of hidden Markov tree framework. (a) Eight consecutive sample locations in 1D space. (b) Partial order constraint in a reverse tree. (c) Hidden Markov tree . . . . .                                                                                                                                   | 6  |
| 3 | Flow direction maps based on DEM of the three test regions. (a) Study Area. (b) Grimesland (m). (c) Kinston: Area 2 (m). (d) Kiston: Area 1 (m) . . . . .                                                                                                                                                            | 6  |
| 4 | Results on Matthew flood, Grimesland, NC. (a) Study area. (b) High-resolution aerial imagery in Grimesland (Source: NOAA National Geodesic Survey, Date: 15 October 2016). (c) DEM (m) (d) RF result (e) GBM result (f) RF+LP result (g) GBM+LP result (h) HMT result . . . . .                                      | 13 |
| 5 | Results on Matthew flood, Area 1 of Kinston, NC. (a) Study area. (b) High-resolution aerial imagery in Kinston: Area 1 (Source: NOAA National Geodesic Survey, Date: 15 October 2016). (c) DEM (m) (d) RF result (e) GBM result (f) RF+LP result (g) GBM+LP result (h) HMT result . . . . .                          | 14 |
| 6 | Results on Matthew flood, Area 2 of Kinston, NC. (a) Study area. (b) High-resolution aerial imagery in Kinston: Area 2 (Source: NOAA National Geodesic Survey, Date: 15 October 2016). (c) DEM (m) (d) RF result (e) GBM result (f) RF+LP result (g) GBM+LP result (h) HMT result . . . . .                          | 15 |

## List of Tables

|   |                                                                |    |
|---|----------------------------------------------------------------|----|
| 1 | Dataset description . . . . .                                  | 5  |
| 2 | Class transition probability $\Pr(y_n y_{Z_n})$ . . . . .      | 8  |
| 3 | Class prior probability . . . . .                              | 8  |
| 4 | Classification on real dataset in Grimesland NC . . . . .      | 11 |
| 5 | Classification on real dataset in Kinston NC: Area 1 . . . . . | 11 |
| 6 | Classification on real dataset in Kinston NC: Area 2 . . . . . | 11 |
| 7 | Computational time costs for three study areas . . . . .       | 17 |