ORBIT (Ordering Based Information Transfer): A Physics Guided Machine Learning Framework to Monitor the Dynamics of Water Bodies at a Global Scale

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Ankush Khandelwal

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Prof. Vipin Kumar, Advisor

May, 2019

Acknowledgements

The fruitful journey of my PhD became possible through the support and patience of a number of people in my life, to whom I am greatly indebted. Foremost, I am deeply thankful to my advisor, Prof. Vipin Kumar, who has not only taught me how to pursue research and solve problems but also set a great example of being respectful and supportive of people's goals and constraints. His intensity towards research, and his humility in spite of his accomplishments, will always be a guiding force in my life. I have been fortunate to work under his guidance. He always challenged me to not let my own self criticism hinder my work and to focus on bringing forth the best of what I accomplished. I am also thankful to my master's thesis advisor Prof. K.S. Rajan who championed for me and made it possible for me to have this life changing experience.

I would also like to thank Professors Snigdhansu Chatterjee, Arindam Banerjee, and Michael Steinbach for taking the time to serve on my thesis proposal and final exam committees.

I would also like to thank Dennis Lettenmaier, Miriam Marlier, Paul Hanson, Hilary Dugan, and Juan Carlos Castilla-Rubio for working together on exciting research collaborations. I am grateful to my fellow lab-mates and friends, Anuj Karpatne, Varun Mithal, Saurabh Agrawal, Guruprasad Nayak, Xiaowei Jia, Xi Chen, Zhihao Wei and Rahul Ghosh for creating a productive work environment. I have also had the opportunity to interact with some very bright undergraduate interns on our research projects: Reid Anderson, Mace Blank, Robert Leunberger, Connor Tarczon, Samuel Hanson, and Huangying Kuang. Their contributions have significantly helped me in my research and my pursuit to make my research accessible and useful to different stakeholders.

Finally, everything that I have accomplished could not have been possible without the love, immense sacrifice, and support of my parents, siblings and my wife. I would

also like to thank all my friends in graduate school, Naveen Elangovan, Shanka Seal, Sourabh Agarwal, Gowtham Atluri, Rahul Saladi, and Akash Agarwal, with whom I have shared numerous stressful and joyful moments during my PhD journey.

Dedication

To my parents, and my wife whom I met during this journey.

Abstract

Predictive learning approaches along with vast amounts Earth Observation data offer a great opportunity to track changes on the earth's surface. However, due to data quality issues (sensor anomalies and atmospheric disturbances) and heterogeneity in the land surface, even state of the art machine learning algorithms perform poorly when applied on a global scale. Furthermore, due to inherent trade-off in sensor design, a single source does not provide both high spatial and temporal resolution required by various scientific applications.

This thesis focuses on developing new machine learning algorithms that can leverage physical principles governing geo-physical processes to overcome these challenges, in the context of monitoring surface water changes at global scale. The thesis introduces a new framework, ORBIT (Ordering Based Information Transfer) that uses an implicit ordering constraint among instances to address the aforementioned challenges. For this application, the topography (the elevation structure) enforces such an ordering. This elevation constraint, however, is not available explicitly in almost all the cases. This thesis introduces a new rank aggregation approach to infer the inherent ordering from the noisy labels. This thesis also introduces a new approach that makes use of this elevation constraint to enforce temporal consistency in surface area variations of water bodies. Finally, this thesis introduces a new approach to downscale low resolution land/water masks to a higher spatial resolution using elevation ordering available at high resolution.

The ORBIT framework was applied to approximately 200,000 lakes globally to create a new dataset, GLADD (Global Lake Dynamics Database) that provides monthly surface dynamics for these water bodies at 30m spatial resolution during the period 1984-2015. The thesis provides a summary of a specific subset of this dataset, GLADD-R, that provides information on new reservoirs created between 1986 and 2013. GLADD-R provides a significant improvement over the existing database on reservoirs, as it detects an additional 1467 reservoirs across the globe that were not known to the scientific community.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Freshwater is one of the most important resources required by humans, and is essential for industry, agriculture, and ecosystem functions. About 97 % of the total amount of water on Earth is found in the oceans and 2 % is stored in the Greenland and Antarctic ice sheets. Thus, only the remaining 1 % of water is available for the earth's processes and for all human needs. Water availability is undergoing tremendous change throughout the world due to both climatic and human-induced impacts. According to an estimate, over one third of the world's population is not served by adequate supplies of clean water [1]. Despite the importance of information about the space-time variations and long-term trends in surface freshwater resources, the current understanding at the global scale is remarkably limited [2]. Improved spatio-temporal quantification of changes in these freshwater resources, especially lakes and reservoirs (they hold 87 % of total surface freshwater [3]) is critical for understanding both anthropogenic impacts, and climate change influences [4], and for better representation of water management effects in hydrological and other Earth System models to prepare effectively for its inevitable future variability.

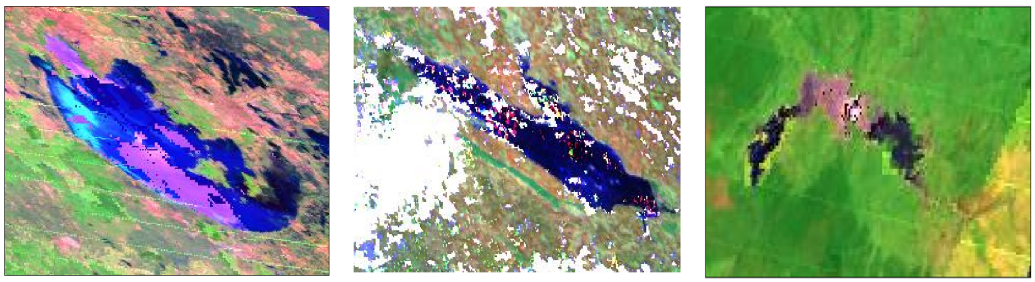
The twenty-first century is considered by many as the golden age of data science. Our ability to collect, store, and access large volumes of data has increased tremendously, along with significant improvements in statistical methods for analyzing this data. Data-driven predictive learning approaches have found tremendous success in

various scientific, as well as commercial applications, e.g., face and object detection, autonomous driving, filtering spam messages, and forecasting health risks.

Similarly, predictive learning approaches have shown promise for analyzing the vast and continuously growing Earth Observation (EO) data acquired through various remote sensing satellites and thus provides an opportunity to monitor freshwater resources. Remote sensing satellites observe a wide variety of physical attributes about the Earth's surface at various spatial resolutions and time intervals. For example, data from MODIS sensor on NASA's Earth observing satellites is available daily at 500 meters spatial resolution, whereas data from the LANDSAT-7 satellite is available every 16 days at 30 meters spatial resolution. Such spatially and temporally explicit data can be used to label each location on Earth as either land or water at any given timestep. Changes in the class labels can then be used to track changes in surface water at a location or a region of interest. This information is especially valuable (and usually the only viable option) in parts of the world where in-situ monitoring networks are lacking [5].

1.2 Challenges

Despite their promise, machine learning methods suffer from a number of challenges when applied to global scale remote sensing data, leading to imperfect class labels. First, remote sensing datasets are generally plagued with noise, outliers (Figure 1.1a) and missing data, (Figure 1.1b) due to sensor anomalies and atmospheric disturbances such as clouds, aerosols and sun angle. Second, even without the above data acquisition related issues, remote sensing data might not be able to distinguish certain classes, such as algae on water, as they appear similar to land (Figure 1.1c). Third, these challenges become even more severe at global scale due to high heterogeneity in the data, as locations with same input values can belong to either land or water depending on their geographical context. Hence, the state-of-the-art machine learning methods for creating water extent maps show unsatisfactory performance especially in the context of identifying dynamics of water bodies at global scale [6].



(a) Outliers (b) Missing Data (c) Algae

Figure 1.1: Examples of data related challenges

Another issue with analyzing satellite imagery datasets is the inherent trade-off in different resolutions. As an example, consider the Earth Observation (EO) data acquired through various remote sensing satellites. MODIS sensor onboard TERRA and AQUA satellites capture earth's surface every day at a coarse spatial resolution (500m). On the other hand ETM+ sensor onboard LANDSAT 7 satellite captures the earth's surface every 16 days but at a high spatial resolution of 30m (see Figure 1.2). Thus, a single sensor is not enough to provide both high spatial and temporal detail required in many earth science applications, for example, calibration of hydrological models [7]. Hence, there is a need to develop methods that can transfer information across scales as well as across time to effectively use the rich complementary information available in these datasets.

(a) Sep 13, 2004 from MODIS product (b) Sep 16, 2004 from Landsat 7

Figure 1.2: An illustrative example showing False Color Composite images from two different sensors onboard Earth Observation satellites.

1.3 Thesis Contributions

This thesis presents a new framework, **ORBIT** (Ordering Based Information Transfer) that aims to address the aforementioned challenges in global surface water monitoring of lakes and reservoirs. Specifically, **ORBIT** aims to incorporate physical principles that govern water body dynamics to improve the quality and resolution of labels obtained from existing satellite imagery derived products. Following are the main contributions and the organization of the thesis.

- Chapter 2 presents a new approach, **ORBIT-E** (Ordering Based Information Transfer using Elevation) to improve label accuracy of existing noisy multi-temporal classification maps by incorporating an implicit ordering constraint among locations due to their topography. The approach formulates the problem as a rank aggregation task to simultaneously estimate the ordering and physically consistent labels [8, 9].

- Chapter 3 presents a new approach, **ORBIT-T** (Ordering Based Information Transfer across Time) to incorporate another physical property of surface water dynamics to further improve the accuracy of labels. Specifically, **ORBIT-T** uses temporal auto-correlation property in surface area variations of water bodies to handle scenarios where individual time steps could have large amounts of errors or missing data [10]. The key idea here is that in most real situations, a water body grows and shrinks smoothly (except sudden events such as floods) i.e. surface extents of nearby dates are likely to be very similar. **ORBIT-T** uses elevation ordering to enforce this temporal consistency in total area values which is more robust than enforcing temporal consistency in labels of individual pixels (as done by most current approaches).

- Chapter 4 presents a new approach, **ORBIT-S** (Ordering Based Information Transfer across Scales). This approach also uses the elevation constraint available at high spatial resolution to downscale land/water maps available at low spatial resolution (but at high temporal resolution) to create maps that have both high spatial and temporal resolution [10]. The key idea here is that in order to create maps at high spatial resolution, we need to find labels for only some of the locations

and then elevation constraint can be used to infer the labels for the remaining locations.

- The **ORBIT** framework was applied to approximately 200,000 lakes globally to create a new dataset, GLADD (Global Lake Dynamics Database) that provides monthly surface dynamics for these water bodies at 30m spatial resolution during the period 1984-2015. Chapter 5 provides the summary of a specific subset of this dataset, GLADD-R, that provides information on new reservoirs created between 1986 and 2013. GLADD-R provides a significant improvement over the widely used database on reservoirs as it detects an additional 1467 reservoirs across the globe that were not known to the scientific community.

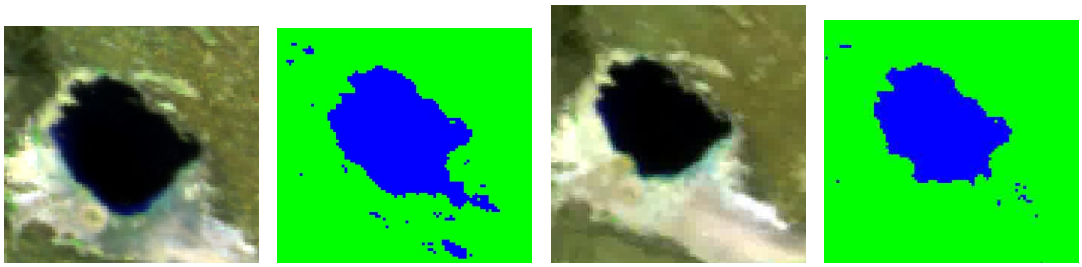
Chapter 2

ORBIT-E: Ordering Based Information Transfer using Elevation

2.1 Introduction

Supervised classification of remote sensing images has been widely used for monitoring water dynamics [11], [12], [13] and other land cover changes on earth. Specifically, we are considering the setting in which different snapshots of a region of interest are classified independently to assign each pixel a land or water label. These individual classification maps are then used for reporting application specific queries such as extent of lakes, extent variation over time etc. However, the accuracy of these classification maps is limited due to noise and outliers, lack of representative training data, and appropriate classification models that can handle the high spatial and temporal heterogeneity at global scale [14]. Figure 2.1 shows a real world example. Figure 2.1(a) and 2.1(c) show false color composite images of lake Abbe in Brazil at two different timesteps. Figure 2.1(b) and 2.1(d) show the corresponding classification maps. As we can see, both maps have classification errors.

Different post classification refinement methods have been proposed to improve the accuracy of these labels [15], [16], [17], [18]. Post classification refinement methods take



(a) (b) (c) (d)

Figure 2.1: False Color Composites and classification maps on two timesteps for lake Abbe in Brazil. a) FCC of July 1, 2003. b) Classification map of July 1, 2003. c) FCC of August 10, 2008. d) Classification map of August 10, 2008.

these noisy labels as input and aim to improve the accuracy of the labels by making certain assumptions about the application at hand and other data related characteristics. To better understand how inherent constraints in water body dynamics can be used for label improvement, consider the following cases in which reality can exist:

Lake's extent is static: In this case, the only information available is that the lake's extent is static i.e., it does not grow or shrink over time. If we assume that label accuracy for a pixel across various snapshots is better than random, then a pixel's label can be estimated as the majority label that it gets across snapshots.

Lake's extent is static and labels tend to be spatially correlated: Given the extra information of spatial structure, snapshot based methods such as spatial majority filtering and other related techniques can be used to remove salt and pepper noise in labels. The problem with these techniques is that they will perform poorly when noise is also spatially auto-correlated. However, since the extent is static, information from different snapshots can be combined to create probability map representing the relative occurrence of a particular label in a pixel. Markov Random Field based methods that effectively model the individual pixel's label probability along with the spatial structure, are more suitable.

Lake's extent is dynamic: In this case, the lake's extent changes across different snapshots. In this scenario, even Markov Random Field based methods would not be suitable because information from different snapshots cannot be aggregated as they potentially represent different reality (potentially different extents of the lake).

In this chapter, we present a new approach, **ORBIT-E** (Ordering Based Information Transfer using Elevation) that uses inherent physical characteristics of water body dynamics to overcome the challenges faced by traditional approaches in improving label accuracy. Specifically, the approach makes use of the inherent ordering constraint among instances/pixels due to the earth's topography/elevation, which determines the surface area dynamics of the water body.

Elevation of a geographical location on earth is its height above a certain fixed point (eg. elevation above sea level). Earth's surface is highly uneven and has various bowl shaped depressions called basins. Water bodies are formed when water fills these basins. Hence, locations inside and around the water have varying elevation/depth. For example, Figure 2.2 shows this elevation information (called bathymetry) of Medicine Lake in California. This elevation information of locations introduces an inherent ordering in the locations. This ordering constraint determines how a water body grows or shrinks. The key idea is the following - if a location is filled with water then by laws of physics all the locations in the basin that have lower elevation should also be filled with water. This constraint has been explained in more detail in section 2.3.2. Thus, if we have elevation information then we can detect inconsistent class labels that do not adhere to this physical constraint. Similarly, if we are given perfect class labels i.e., they perfectly agree to the physics constraint, then the growing and shrinking of a lake across various snapshots can be used to extract correct elevation ordering. In other words, by combining information from multiple timesteps, instances can be ordered according to their elevation.

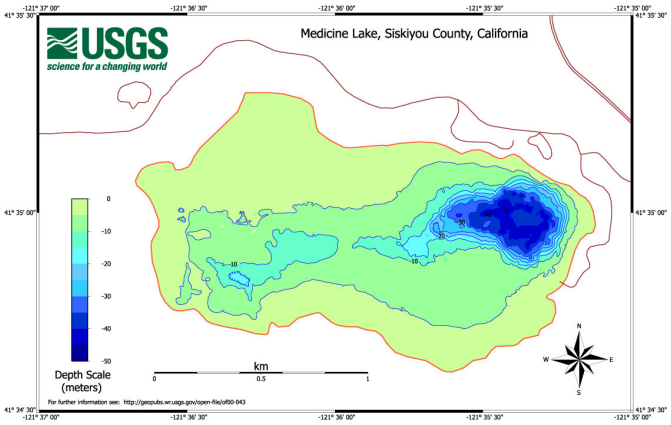


Figure 2.2: A bathymetric map of Medicine Lake in Califorinia

The idea of using ordering among instances is more general and can be used for any application that has following characteristics:

- There exists a bipartite grouping of instances based on a criteria (for example, pixels belonging to either water or land group/class) and these groupings are available from multiple sources of similar category. For example, different snapshots of the same region on earth.
- The bipartite grouping varies between sources due to different characteristics of the sources. For instance, varying extent of a lake across different snapshots.
- There exists an inherent ordering among instances that can be utilized to detect and subsequently correct the inconsistencies in labels. For instance, the ordering through elevation.

As an illustrative example of another application, consider a set of 50 students (instances) taking a chemistry test of 100 questions (sources). Each question has binary (1/0) grading, so each question creates a bipartite grouping of students. The groupings are noisy because some students might have answered a question correctly by chance. The groupings vary across questions because questions are of different difficulty level. If we assume that there exists an ordering among students based on their overall knowledge of chemistry that determines their ability to answer questions, then for any given question it can be estimated which students have answered it correctly by chance based

on how students with higher rank have performed on that question. For instance, let's say for a given question, only the top 10 students and the 30^{th} ranking student answered it right. Because of this available ranking information it can be said that the 30^{th} ranking student might have answered this question correctly by chance because the other 19 students, who had better skills than the 30^{th} ranking student, were not able to answer the question.

In real-world applications, however, ordering among instances is often not available. For example, bathymetry information is available only for a very small number of lakes in the world. Similarly, in the student example, obtaining inherent ordering is the final goal itself. Thus, in real-world situations both pure bipartite groupings and inherent ordering are often unknown. To overcome this challenge, the proposed approach uses an EM framework that iterates between estimating inherent ordering from noisy labels, and estimate correct labels from the ordering. Since the goal is to estimate the inherent total ordering among instances using partial rankings (bipartite groupings) from individual time steps, our problem setting can also be considered as a rank aggregation problem.

In this chapter, without the loss of generality, we assume that higher elevation locations get higher ranking. Hence, a shallow location will have a higher rank through elevation than a deeper location. Under this notation, each pure bipartite grouping can be considered as coarse ordering information where all land pixels get rank 1 and all water locations get rank 2.

The rest of the chapter is structured as follows: Section 2.2 describes the related work on post classification label refinement that has been proposed in the remote sensing community as well as related rank aggregation methods proposed in various disciplines; Section 2.3 explains various aspects of the **ORBIT-E** approach; Section 2.4 describes the theoretical properties of **ORBIT-E**; Section 2.5 describes various baseline algorithms and results on synthetic and real-world datasets. Section **??** highlights some of the limitations of the approach and finally Section 2.7 provides the summary of the work with possible future directions.

2.2 Related Work

Various post classification label refinement methods have been proposed in remote sensing literature that aim to detect inconsistencies in class label assignments and subsequently correct them by making certain assumptions about the application at hand. These assumptions are derived from two main contexts, namely *data context and process context*. Methods that exploit the data context use constraints from input labeled data or ancillary data. For instance, majority filter based methods [15], [19] do label refinement by assuming spatial and temporal auto-correlation in datasets. Recently, more sophisticated methods using Markov Random Fields [17], [20] and edge preserving smoothing have been proposed [21]. The major challenge with these approaches is that errors in labels tend to be spatially and temporally correlated as well which makes these approaches less effective.

Methods that exploit the process context use constraints derived from the underlying process. For instance, [16] defined knowledge based refinement rules e.g., change from urban to vegetation is not allowed as it is highly unlikely. More sophisticated algorithms using process context have been proposed in [22] and [18]. In [22], the authors used transition probability matrices to code domain knowledge about compatible ecological changes in their MAP-MRF model. In [18], the authors used Hidden Markov Models based methodology to dynamically learn and use transition and emission probabilities of various land cover classes for the application of urban mapping. However, since surface water dynamics depends heavily on external factors such as climate, precipitation, and human use, it becomes challenging to model transition probabilities for our application.

With respect to the aforementioned categorization of the existing approaches, **ORBIT-E** belongs to the category of process context based methods since we are exploiting the lake growing and shrinking process as a constraint to improve the accuracy of labels.

A key element of the proposed approach is to estimate the inherent elevation ordering in order to correct inconsistent labels. The problem of obtaining a global ordering of instances by aggregating partial information about the instances has been explored in various areas such as sports [23], collaborative filtering [24], meta search [25] and database middle-ware [26]. Various rank aggregation methods have been proposed that aggregate partial information that is available in different forms. There are three major

forms of partial information that have been studied in the literature. Partial information can be available in the form of pairwise preference relations [27]; rankings on different subsets of instances [28]; top-k lists from various sources [29]. In our setting, the partial information is available in a new form where bipartite grouping from various sources (snapshots) are available. This bipartite information on instances can also be represented as a set of pairwise preference information between instances. In other words, this form of preference information can be considered as a special case of pairwise preference information from different sources. Hence, methods that use pairwise preference relationships can be potentially used to estimate the total ordering. For instance, Borda Count is a well known rank aggregation method used in elections to obtain the winner by aggregating votes from a set of voters. Variations of this method have been proposed for scenarios when pairwise preference information is given [30]. Some approaches [27] obtain total ordering by aggregating pairwise preference information collected from different sources in the form of a preference graph. Similarly, [28] defined Markov Chain based methods to estimate the ordering using probabilities of the stationary distribution of the markov random walk on the preference relationship graph. Bradley-Terry [31], Thrustone [32] and Mallows [33] are well known rank aggregation models that have been proposed that make certain assumptions about the distribution from which preference relationships have been sampled. The key difference between the proposed method and these methods is that they do not exploit the extra higher level information that instances have a bipartite grouping in them as well. Bipartite ranking algorithms [34] are also related to this type of information but bipartite ranking algorithms work in a supervised setting and also require access to feature space of instances as they learn a ranking/scoring function on the feature space. But in our setting, both the feature space and training data are not available.

Since, in the application of water monitoring, methods that assume spatial and temporal auto-correlation in labels have been used, we have implemented a traditional spatial majority filter and a temporal majority filter for comparison. As mentioned before, due to correlation in noise, these methods tend to perform poorly as demonstrated in the results section. It is noteworthy that ordering based approach does not assume any spatial or temporal auto-correlation in instances and hence should have better performance even when noise is auto-correlated. In order to compare the proposed ordering

based method with other ordering based methods, we have implemented a variation of Borda Count method described in [35] and a preference graph based methodology proposed in [27].

2.3 Proposed Approach

Now, we describe the **ORBIT-E** approach in the context of lake extent monitoring. From here on, the terms snapshot and time step will be used interchangeably.

2.3.1 Input

We are given classification maps of a region across multiple time steps. The matrix $A_{R\times C\times T}$ represents the input data in the 3-dimensional form where $A_{i,j,t} \in \{0, 1\}$ represents the class label (1 means land, 0 means water) of a pixel at location (i, j) on the grid at time t. Alternatively, matrix A can be represented as a 2-dimensional matrix $D_{N\times T}$ where N is the total number of locations $(R * C)$ and $D_{l,t}$ is the class label of pixel at location index l at time t. From here on, the cells of the grid will be referred to as locations and a (location, timestep) pair defines a pixel. Each row of matrix D represents the temporal sequence of class labels of a location. Since each time step is a binary classification map with noise, each column of D represents a noisy bipartite grouping of the given N instances. Figure 2.3 shows an illustrative example of these two data representations. For the proposed approach, we will be using the 2-dimensional data representation as the approach does not use spatial auto-correlation. Given this input, our aim is to detect inconsistent labels and correct them.

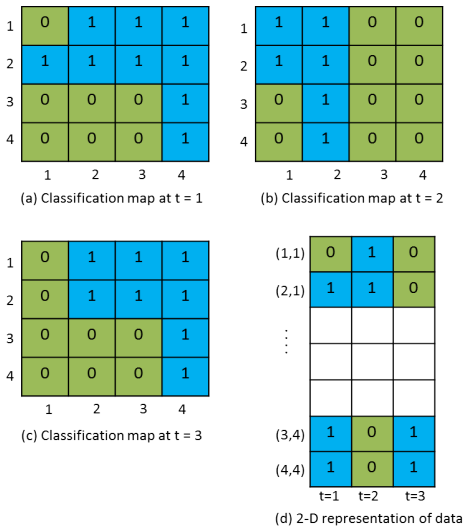


Figure 2.3: Illustrative example of 3-D and 2-D representation of the data. (a)-(c) show the 2-D representation of the data for 3 different timesteps, (d) show the 2-D representations of the complete data.

2.3.2 Physical Constraint through Elevation

As mentioned earlier, **ORBIT-E** uses elevation constraint to improve the accuracy of labels. Locations have an inherent ordering based on their elevation. Specifically, if location l is filled with water then all locations that are deeper (lower elevation) than location l should also be filled with water. Given this constraint through ordering, imperfect labels can be estimated and subsequently corrected. Figure 2.4(a) shows a pure synthetic input matrix D_p with $N = 100$ locations and $T = 200$ timesteps. Blue represents water pixels and green represents land pixels. As stated earlier, each row of input matrix D_p represents a temporal sequence of class labels for a location. Figure 2.4(b) shows the matrix D_p^{π} that is obtained by ordering locations in D_p in increasing order of their elevation. The bottom of the matrix represents lowest elevation and top of matrix represents highest elevation. This ordering information through elevation is referred to as π. Each column (timestep) of D_p exhibits pure bipartite grouping i.e., for any given time step t all water pixels have lower elevation than land pixels. Figure 2.4(b) also shows the varying nature of bipartite groupings. In our application this varying amount of water pixels in different timesteps corresponds to growing and shrinking of a

water body across time. Figure 2.4(c) shows the matrix D_n which is the noisy version of D_p, i.e., bipartite groupings are impure. Figure 2.4(d) shows matrix D_n^π which is obtained by ordering locations using π. We can see that elevation based ordering has reasonably partitioned water and land labels for different timesteps. However, there are some location pairs that are showing physically inconsistent behaviour in some timesteps. For instance, in timestep t a lower elevation (deeper) location k has been labeled as land where a higher elevation (shallow) location i has been labeled as water. Hence, using the elevation ordering these inconsistencies can be detected. However, an issue here is that we still cannot determine whether the inconsistency is due to location i being incorrectly labeled as water or location k being incorrectly labeled as land.

2.3.3 Estimation of Correct Labels from Elevation

If the elevation ordering (π) and size of one partition in bipartite grouping of a timestep t (henceforth referred as θ_t) are available, then inconsistent labels can be estimated and corrected for that timestep. Without the loss of generality, we consider the partition that contains water pixels. For example, $\theta_t = k$ would mean that for timestep t, bottom k locations in π are filled with water. This would automatically mean that for timestep t, top $N - \theta_t$ locations in π are land where N is the total number of locations. In other words, θ_t represents water level at timestep t in our application or θ_t induces the true bipartite grouping in that timestep. Now, if there are locations in the bottom θ_t in π that are labelled as land at timestep t, then they have been incorrectly labelled as land in that timestep. Similary, if there are locations that are in top $N - \theta_t$ and are labelled as water in timestep t, then they have been incorrectly labelled as water. Hence, given π and θ_t incorrectly labeled pixels in the given timestep can be detected.

Unfortunately, in our application θ_t (water level at timestep t) is a latent variable and thus it has to estimated as well. Here, we consider the maximum likelihood interpretation of θ_t. Specifically for a given ordering, θ_t is chosen such that the corrected labels have the least number of disagreements with the input labels. In order words, the θ_t that best describes the given bipartite grouping on that timestep will be selected.

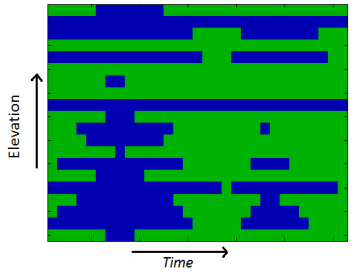
Mathematically, $\hat{\theta}_t$ is estimated as -

$$\hat{\theta}_t = \underset{k \in [0,N]}{\arg\max}\, Acc(k, t, \hat{\pi}) \tag{2.1}$$

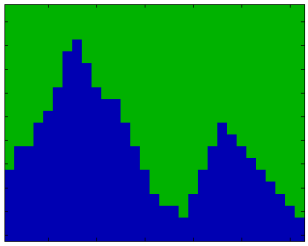
where,

$$Acc(k, t, \hat{\pi}) = \sum_{i=0}^{k} \mathbb{1}(D_n^{\hat{\pi}}(i, t) == 0) + \sum_{i=k+1}^{N} \mathbb{1}(D_n^{\hat{\pi}}(i, t) == 1) \tag{2.2}$$

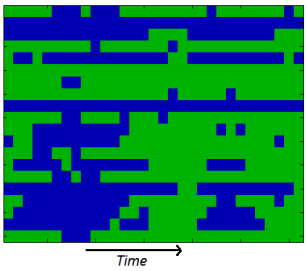
The first term in equation 2.2 measures the agreement of estimated water partition with the noisy water partition and the second term measures the agreement of the estimated land partition with the noisy land partition. Note that the maximum likelihood estimate of $\hat{\theta}_t$ is applicable only when a majority of the labels are correct. In other words, if a majority of the labels are wrong then no method has any chance of correcting the errors. Figure 2.5 shows an illustrative example demonstrating estimation of $\hat{\theta}_t$. In this example, $k = 3$ leads to maximum agreement with the noisy input partition with *Acc* value of 6.



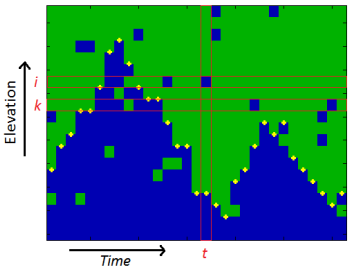
(a) Input Matrix D_p



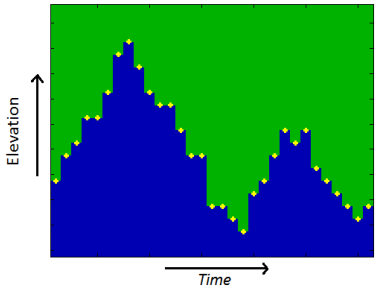
(b) Ordered Input Matrix D_p^{π}



(c) Noisy Input Matrix D_n



(d) Ordered Noisy Matrix D_n^{π}



(e) Partition Matrix $P^{\pi,\hat{\theta}}$

Figure 2.4: Synthetic data demonstrating utility of elevation constraint for label correction

Figure 2.4(e) shows the matrix $P^{\pi,\hat{\theta}}$ that represents estimated bipartite groupings using equation 2.1 on D_n^{π}. Yellow dots on each timestep mark the partition boundary for which estimated partitions showed maximum agreement with noisy partitions. Hence, by definition, for any given timestep, locations below yellow markers are labelled as water and locations above the marker are labelled as land. Now, given matrix $P^{\pi,\hat{\theta}}$, we can resolve the ambiguity mentioned at the end of section 2.3.2. Specifically, from $P^{\pi,\hat{\theta}}$

we can say that in D_n^{π} on timestep t, location i was incorrectly marked as water and not location k.

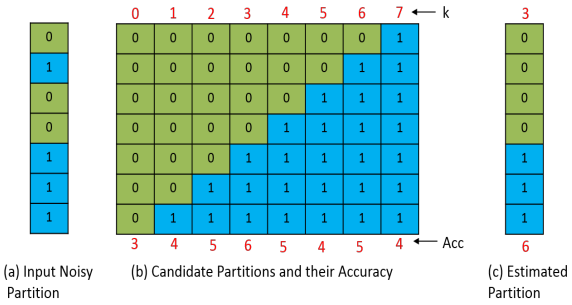


Figure 2.5: Illustrative example showing candidate partitions and their *Acc* values for different values of k

2.3.4 Estimation of Elevation from Perfect labels

In the previous section, we described how accurate elevation information can be used to estimate correct labels. Similarly, accurate class labels (bipartite groupings) can also be used to estimate inherent elevation information. Physically consistent variation in class labels due to dynamics in water body can be used as a proxy to estimate the ordering. Specifically, over any given time period a deeper location will be labeled as water more frequently than a shallow location. This is due to the physics described in section 2.3.2 above. Whenever a shallow location is water then the deeper will strictly be water. But if a deeper location is water then a shallow location may or may not be water depending on the current extent of the water body. Mathematically, the elevation of a location is directly proportional to the number of time steps in which the location is labeled as water. Specifically, we define the this relationship as -

$$\hat{\pi_l} = T - \sum_{t=1}^{T} \mathbb{1}(D_p(l,t) == 1) \tag{2.3}$$

where,

$\hat{\pi_l}$ is the estimated elevation of the location l,

T is the number of time steps,

2.3.5 Estimation of Elevation Ordering from Noisy Labels

Till now, we have assumed that either accurate elevation information is available, or accurate labels are available. But in our application setting, labels are noisy and even elevation information is not available. $\hat{\pi}$ will only be an approximation of the true hidden elevation information under the noisy labels scenario. In later sections we will show that this a poor approximation of the elevation information.

Hence, we need a better way to aggregate these noisy bipartite groupings so that a better approximation of inherent ordering can be made, which will subsequently lead to better label correction. Here, we define the objective function that will guide the search for the better approximation of the true ordering as:

$$\arg\max_{\hat{\pi}} \sum_{t=0}^{T} Acc(\hat{\theta}_t, t, \hat{\pi}) \tag{2.4}$$

As mentioned before, *Acc* is calculated with respect to a given ordering and for a timestep. It measures the agreement between the estimated bipartite grouping and input bipartite grouping. So the above objective function would prefer the ordering that leads to overall better agreement between estimated and input bipartite groupings. In other words, we would prefer an ordering that overall describes the data well. This objective function extends the maximum likelihood interpretation from a single timestep to the whole data. This objective function is very similar to a well known objective function defined as Kemeny optimal aggregation in [36] for finding a global ordering that maximally agrees with input rankings on a subset of instances coming from multiple sources. In Kemeny optimality, instead of *Acc*, a variation of Kendall Tau correlation was used. Kendall Tau correlation is not suitable for our setting because we do not have ranking, but bipartite groupings. Kemeny optimal aggregations satisfy extended Condercet Criterion which makes them suitable for detecting outliers or inconsistencies in partial rankings [28]. As explained in section 2.3.4, if elevation is given, $\hat{\theta}$ can be estimated. But here elevation is also a variable. Hence, in the above function there are two latent variables, $\hat{\theta}$ and $\hat{\pi}$.

Now, to obtain an ordering that best fits the given data or maximizes the above objective function is a NP-Hard problem. Hence, we would need different heuristic

solutions to estimate the approximate ordering. Next, we describe the proposed Expectation Maximization framework to optimize the aforementioned objective function.

2.3.6 Optimization

As explained in section 2.3.3, if an ordering (π) is available, water levels ($\hat{\theta_t}$), and subsequently correct labels can be estimated for each time step. On the other hand if correct labels are available (section 2.3.4) then elevation information can be accurately estimated. **ORBIT-E** iterates between estimating water levels from a given ordering, and estimating new ordering from the water levels, such that the new ordering has an improved value of the objective function over the previous ordering. The two steps are explained below -

Estimating Water Levels from Ordering (Maximization Step)

This step is very similar to the step explained in section 2.3.3. The only difference is that here we consider the current estimate of ordering ($\hat{\pi}$) instead of true ordering (π) because it is not available to us. To initialize the process, any random ordering can be provided ($\hat{\pi}^0$). So, water levels at iteration i are calculated as:

$$\hat{\theta_t^i} = \underset{k \in [0,N]}{\arg\max}\, Acc(k, t, \hat{\pi}^{i-1}) \tag{2.5}$$

Estimating ordering from water levels (Expectation Step)

This is the key step of the approach. First, we describe the notion of *location profiles* and *elevation profiles*. As mentioned in section 2.3.2, location profile is basically the temporal sequence of class labels of the given location. Now, consider the matrix shown in 2.4(e). As mentioned earlier $P_n^{\theta,\pi}$ is obtained with respect to a given ordering. This matrix can also be viewed from a different perspective. Each column (timestep) of $P_n^{\theta,\pi}$ represents the estimated bipartite partition for the given column (timestep). On the other hand, each row of $P_n^{\theta,\pi}$ represents the ideal temporal sequence of labels for that row (elevation). To re-emphasize, i^{th} row of $P_n^{\theta,\pi}$ represents the ideal label sequence of i^{th} ranking elevation for the given ordering π. Note that we are distinguishing between location and its elevation. We ideally want a location to be assigned to an elevation

such that the location's profile and the corresponding elevation's profile has maximum agreement. For instance, in Figure 2.6, location f has 6 mismatches with the 5^{th} ranking elevation's profile but it has 0 mismatches with 1st ranking elevation's profile. Hence, it makes sense to assign the location f to elevation 1 rather than elevation 5. This transformation of ordering process to the assignment process is the key aspect of the approach. To summarize, we would assign each location to an elevation with which it has maximum agreement. If there are multiple elevations for which a location is showing maximum agreement, then ties can be broken arbitrarily. Similarly, multiple locations can be associated to a single elevation. Since elevation profiles are already elevation ordered by definition, this assignment process produces partial ordering where instances associated with a higher ranked elevation profile will be ranked higher than instances associated with lower ranked elevation profile. Now, to break ties within locations that are assigned to the same elevation, we rank them according to the number of water labels in each location because here is no other information available for those locations. This gives a new complete ordering of instances.

g	0	0	0	1	0	0	1	0
f	1	1	1	1	1	1	1	1
e	1	0	1	1	1	1	1	0
d	0	0	0	0	0	0	1	0
c	0	0	0	0	0	0	0	0
b	1	1	1	1	1	1	1	1
a	0	0	1	1	1	0	1	0

(a) Input Matrix D_n

7	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	1	0
5	0	0	0	1	0	0	1	0
4	0	0	1	1	1	0	1	0
3	1	0	1	1	1	1	1	0
2	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1

(b) Estimated Partition Matrix $P_n^{\hat{\pi}^{i-1},\hat{\theta^i}}$

Figure 2.6: Illustrative example demonstrating the assignment operation

The algorithm iterates between these two steps until there is no further increase in the value of the objective function.

2.4 Properties of ORBIT-E

2.4.1 Proof of Convergence

Next we describe the qualitative understanding of the convergence of the approach. The first step of the approach aims at estimating water levels using a given ordering. This step trivially minimizes the objective function as it uses the function itself to find the optimal cuts. The second step aims to assign each location to an elevation profile that best matches the location's profile. Therefore, in principle this step also reduces the disagreement between the input labels and estimated labels using the current set of parameters. Thus, qualitatively we can see that after every iteration the value of the objective function will decrease and the algorithm will stop when there is no further decrease in the value of the objective function. Also note that the algorithm converges to a local optimum and not a global optimum as any another EM based method.

2.4.2 Time and Space Complexity

In any given iteration, only two NxT size matrices (the input data matrix and the current partition matrix) are maintained where N is the number of instances and T is the number of timesteps. The second step uses a Nx1 vector to store the index of the associated elevation for each location.

In step 1 of the method, it takes $\Theta(N)$ operations to estimate the water level of a timestep. Hence, time complexity of step 1 is $\Theta(NT)$. In step 2, to find the best elevation profile for a location, it takes $\Theta(NT)$ operations. Hence, the time complexity of this step is $\Theta(N^2T)$.

2.5 Results

Next, we describe different baseline algorithms, synthetic and real-world datasets used for evaluation of **ORBIT-E** approach.

2.5.1 Baseline Algorithms

the following algorithms have been used for comparison:

Spatial Smoothing (SS): This is just a basic spatial majority filter with a predefined window size. Here, the window size of 5 was chosen. Since, it is a spatial majority filter, it is applied on each timestep individually.

Temporal Smoothing (TS): This is another majority filter based smoothing. Here, the smoothing is done across time. Hence, each location is smoothed individually. The temporal length of the filter was chosen to be 5.

The idea behind analyzing SS and TS is to demonstrate that noise in the data is not trivial and hence cannot just be removed by trivial majority filters. These two algorithms do not use elevation information.

Borda Count (BC): Borda Count is a very widely used rank aggregation method mostly used for aggregating votes in an election. Specifically, each voter gives a ranking to the candidates. From each vote, candidates get a score which is equal to the number of candidates that are ranked below them. For instance, in a N candidate election, if a candidate gets a rank r in a vote, then that candidate gets a score $N - r$ from the vote. The candidate whose has the highest total score over all votes is called a Borda winner. In our application, we do not have total ranking on candidates (locations in our case) from different voters (timesteps in our case). But bipartite information can be converted into score information. Specifically, for a given timestep all water pixels have the same rank and all land pixels have the same rank. But a land pixel is ranked higher than all water pixels. Hence, each land pixel gets a score which is equal to the number of water pixels in that timestep. Similarly, each water pixel in that timestep gets a score of 0. In this way total scores over all timesteps can be calculated and can be used for ordering. Once the ranking is obtained, the labels are corrected by using the strategy described in section 2.3.3. Also, Borda Count for bipartite groupings turns out to be same as $\hat{\pi}$ in equation 2.3.

Preference Based Ordering (PB): This method uses a preference relationship graph to calculate the total ordering among locations. $PREF_{i,j}$ denotes the number of timesteps when location j was land and location i was water. From the graph point of view, an outgoing edge from location/node i to location j with weight k means location j was land for k timesteps when location i was water. Similarly, $PREF_{j,i}$ denotes the number of timesteps when location i was land and location j was water. The location

that maximizes following function becomes the highest ranking location:

$$\arg\max_{j} \sum_{i} PREF_{i,j} - \sum_{i} PREF_{j,i} \tag{2.6}$$

The first term in the above equation is the sum of incoming edges to location j and the second term is the sum of outgoing edges from location j. Once the location that maximizes the above function is selected, the outgoing and incoming edges from the selected location are removed from the graph and the graph is searched again for the top ranking location. In this way, (iteratively) all locations get ranked and a total order is obtained. Once the total ordering is obtained, the labels are corrected using the strategy described in section 2.3.3.

To summarize, TS and SS are majority filter based method which do not use elevation constraint. **BC**, **PB** and **ORBIT-E** are methods that use elevation constraint to obtain total ordering of locations and to correct the labels.

2.5.2 Evaluation Measure

To evaluate the performance of algorithms, the following evaluation measure has been considered:

Error: If ground truth labelling is available then the quality of corrected labels generated from different algorithms can be evaluated with respect to the ground truth labelling. Here, we consider classification error as the metric, which is the fraction of pixels whose labels do not match with the ground truth.

2.5.3 Synthetic Datasets

Here, we describe synthetic data generation process:

Extent and Dynamics

First, the extent for different timesteps are created such that the dynamics in the lake are physically consistent, i.e. the synthetic water body grows and shrinks according to the predefined inherent ordering of locations. This set of extent maps are the ideal maps that we intend to recover after label correction. Hence, they will be used as ground

truth to compare the performance of various algorithms.

Noise Structure

Now, noise is introduced in the ground truth extents to create the dataset that will be provided as input to different algorithms for correction. Noise can have different characteristics and hence will impact algorithms differently. Here, we have analyzed five different types of noise structures -

Random Noise (RN): (location, timestep) pairs i.e. pixels are randomly selected and noise is added in those pixels.

Spatial Noise (SN): Pixels are randomly selected as seed pixels around which spatially auto-correlated noise is added. The spatially auto-correlated noise is added only into the timestep to which that pixel belongs. The strength of spatial auto-correlation is again randomly chosen from a predefined list of possible auto-correlation strength values. Strength is nothing but the size of the window that determines the amount of noise that will be added around the seed pixel. For instance, a strength value of 7 means that 49 (7x7) noisy pixels should be added around the seed pixel. Noise is added around the seed pixel with a region growing strategy. Specifically, the region is initialized as the seed pixel and then immediate neighbors of the region are considered for addition into the region. A subset of the immediate neighbors are chosen at random and added to the region in order to give a more realistic nature to noise.

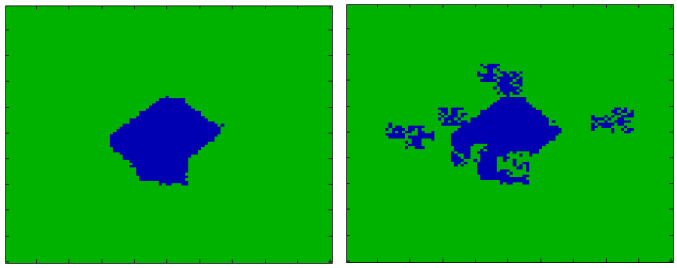
Temporal Noise (TN): Pixels are randomly selected as seed pixels around which temporally auto-correlated noise is added. The strength of the auto-corelation is randomly selected from a predefined list of possible auto-correlation strength values.

Spatio-temporal Noise (STN): Pixels are randomly selected as seed pixels around which spatially and temporally auto-correlated noise is added. First, the strength of temporal auto-correlation is randomly selected. This determines how many timesteps around the timestep of the seed pixel would be affected by noise. Then for each of those timesteps, spatially auto-correlated noise is added around the seed location using the strategy described before.

Location Specific Noise (LN): In the application of water monitoring, it happens sometimes that there are some locations that are more noisy than others due to various

reasons. This noise structure aims to capture that phenomenon. Here, instead of randomly selecting (location,timestep) pairs, only locations are randomly selected first. Then spatio-temporal noise is added in only the selected locations. The amount of noise to be added is again randomly chosen from predefined noise levels.

Figure 2.7 shows the ground truth and different noisy datasets that have been used for analysis. The locations in these datasets have been ordered according to the predefined elevation based ordering. We can see from figure 2.7(a) that there are lot of locations which are land throughout (top green part of the image). Similarly the bottom part of the image has all the locations which are water all through out. Finally in between there are dynamic pixels which have been water or land at different points of time. In the noisy datasets, we can see that most of the noise has been added in the dynamic locations in order to keep noise sufficiently complex and therefore not too easy for ordering based methods to correct. As an example, Figure 2.8 shows the ground truth extent map of a time step and the corresponding noisy classification map of that time step when spatial noise is added. We can see that some blobs of spatially auto-correlated noise have been added.

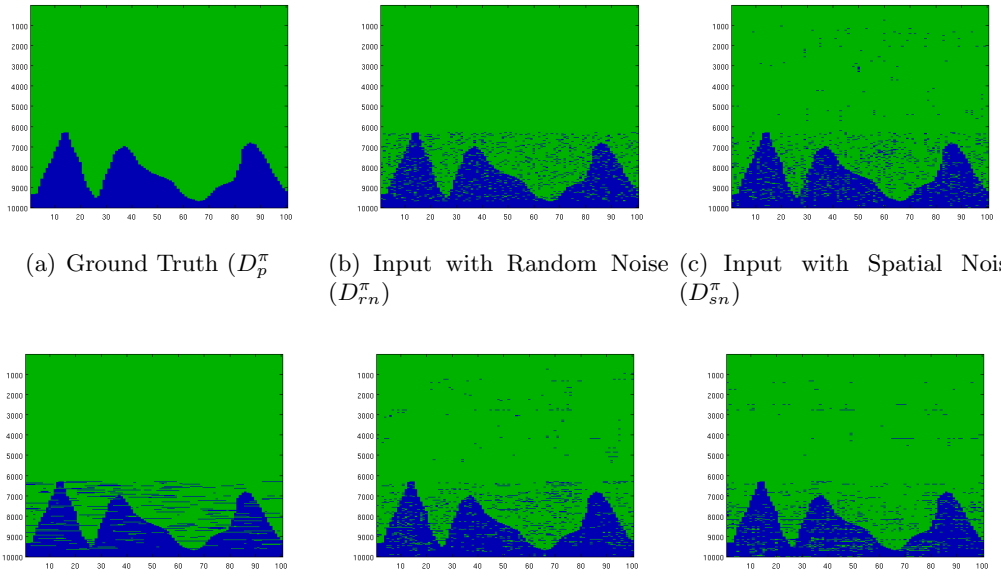


(a) (b)

Figure 2.8: An example of a pure snapshot with corresponding corrupted snapshot with spatial noise. a) Ground Truth. b) Spatio-temporal Noise

Next, we provide the analysis of different algorithms on this synthetic dataset.

Impact of Noise: The performance of the algorithms depends on the type of noise structure present in the data as well as the amount of noise in the data. First, we compare different algorithms using the error metric. Tables 2.1 to 2.5 show the error rate of different algorithms for different noise structures. The first column of each table reports the amount of noise introduced.



(a) Ground Truth (D_p^π) (b) Input with Random Noise (D_{rn}^π) (c) Input with Spatial Noise (D_{sn}^π)

(d) Input with Temporal Noise (D_{tn}^π) (e) Input with Spatio-temporal Noise (D_{stn}^π) (f) Input with Location Specific Noise (D_{ln}^π)

Figure 2.7: Elevation ordered Synthetic Ground Truth and input datasets with various noise structures for noise amount=20 %

In the presence of random noise (Table 2.1), all algorithms are able to remove significant amount of noise. **ORBIT-E** shows the best performance at all noise levels. As expected, the performance of all algorithms reduces with an increase in amount of noise. But **ORBIT-E** shows more robustness to noise than other algorithms. TS has the worst performance. This can be attributed to the high dynamics in the lake and hence poor temporal auto-correlation structure.

Table 2.1: Error in presence of Random Noise

	SS	TS	BC	PB	**ORBIT-E**
height1 %	1.86	0.53	0.57	0.56	0.05
5 %	2.04	1.73	1.62	1.61	0.24
10 %	2.30	3.63	2.60	2.58	0.60
20 %	3.14	9.89	4.69	4.66	1.86
40 %	17.32	34.19	15.85	15.60	14.54

In the presence of spatial noise (Table 2.2), the performance of SS degrades significantly. This explains that the spatial noise structure is not trivial and cannot be removed by just using traditional majority filtering. On the other hand, TS and all other total ordering based methods (BC, PB and **ORBIT-E**) show performance that is very similar to the random noise case. This explains that ordering based methods are much more robust to spatial noise.

Table 2.2: Error in presence of Spatial Noise

	SS	TS	BC	PB	**ORBIT-E**
1 %	2.51	0.51	0.54	0.54	0.03
5 %	5.38	1.43	1.54	1.54	0.16
10 %	9.32	3.06	2.43	2.41	0.34
20 %	18.22	8.51	4.20	4.14	1.09
40 %	39.21	29.52	16.07	16.11	14.73

In the presence of temporal noise (Table 2.3), the performance of TS degrades significantly due to the non trivial nature of temporal noise. Also, note that the performance of total ordering algorithms show more degradation as compared to the spatial noise case. This shows that total ordering algorithms are more sensitive to temporal noise than spatial noise. This happens because the ranking of a location is loosely related to the amount of water in that location, so, if the noise has strong temporal structure then it becomes difficult for ordering based methods to recover the true ordering of locations. Due to this phenomenon, SS shows better performance than all total ordering based algorithms when the amount of noise is high.

Table 2.3: Error in presence of Temporal Noise

	SS	TS	BC	PB	**ORBIT-E**
1 %	1.85	1.27	0.95	0.94	0.07
5 %	2.02	5.36	4.02	4.01	0.50
10 %	2.25	10.46	6.75	6.74	1.41
20 %	2.97	20.69	11.85	11.79	5.88
40 %	20.25	41.21	35.70	35.56	32.39

In the presence of spatio-temporal noise, both SS and TS show poor performance which demonstrate the non trivial nature of the noise. This noise structure is most

representative of the noise that occur in the water monitoring application. Here also **ORBIT-E** shows the best performance and high robustness to noise.

Table 2.4: Error in presence of Spatio-temporal Noise

	SS	TS	BC	PB	**ORBIT-E**
1 %	2.53	0.73	0.79	0.78	0.04
5 %	5.43	3.02	2.53	2.56	0.25
10 %	9.25	6.17	4.13	4.10	0.48
20 %	18.01	13.87	6.82	6.77	1.47
40 %	38.82	34.60	20.78	20.92	19.40

Location specific noise is a special case of spatio-temporal noise where noise is more concentrated on some randomly selected locations as opposed to spatio-temporal noise structure where noise is randomly added across locations and time. We can see that **ORBIT-E** still shows very good performance and performs much better than other algorithms.

Table 2.5: Error in presence of Location Specific Noise

	SS	TS	BC	PB	**ORBIT-E**
1 %	2.66	0.89	0.60	0.59	0.08
5 %	5.56	3.37	2.78	2.73	0.38
10 %	9.61	6.75	5.26	5.26	0.89
20 %	18.74	14.91	8.98	8.97	3.41
40 %	39.56	35.57	26.14	26.08	22.97

In summary, the performance of TS and SS across different noise structures demonstrate the non trivial nature of noise. Ordering based methods perform significantly better than traditional majority filters. **ORBIT-E** shows very good and robust performance across different noise structures and at different noise levels. Total ordering based methods are relatively more sensitive to temporal noise than spatial noise. The performance of BC and PB is very similar because in PB also the difference of weights of incoming and outgoing edges in a location (equation 2.6) is loosely related to the amount of water in the location.

Impact of Initial Ordering on ORBIT-E: Since **ORBIT-E** is an EM based algorithm, the quality of the initial ordering could have an impact of the quality of the

local minima obtained by the algorithm. In this experiment, we analyze the robustness of **ORBIT-E** to initial ordering. Here, we considered datasets with all 5 different types of noise structures with a noise amount of 20 %. Table 2.6 shows the mean error rate and standard deviation in error rate when **ORBIT-E** was run 100 times, each time with a different random initial ordering for the same given input dataset. The other columns show the error values of other algorithms on the same dataset. As we can see that **ORBIT-E** still has better mean error values than other algorithms while having very low standard deviation values. This shows that the algorithm is very robust to the choice of the initial ordering.

Table 2.6: Impact of Initial Start on **ORBIT-E** for different noise structures at noise level = 20 %

	SS	TS	BC	PB	**ORBIT-E**
RN	3.14	9.90	4.75	4.70	2.04 ±0.11
SN	18.36	8.61	4.04	3.97	1.17 ±0.12
TN	2.99	20.70	11.77	11.70	6.56 ±0.37
STN	18.11	14.27	6.76	6.78	1.24 ±0.11
LN	19.08	15.08	8.74	8.79	4.10 ±0.19

Importance of Iterations in ORBIT-E: In this experiment, we analyze the convergence properties of **ORBIT-E**. Figure 2.9 shows the percentage improvement in the value of the objective function consistency for different datasets with different noise levels. Percentage improvement is calculated with respect to the value of the objective function in the final iteration compared to the value in the first iteration. From Figure 2.9, we can say that for all datasets, the algorithm achieves more than 90 % of the improvement in the first iteration itself. By the 6^{th} iteration, the algorithm converged for all datasets. This tells us that the first iteration itself contributes the maximum to the overall improvement. However, other remaining iterations do add some improvement in the objective function's value. This demonstrate the value of the iterative procedure in improving the value of the objective function.

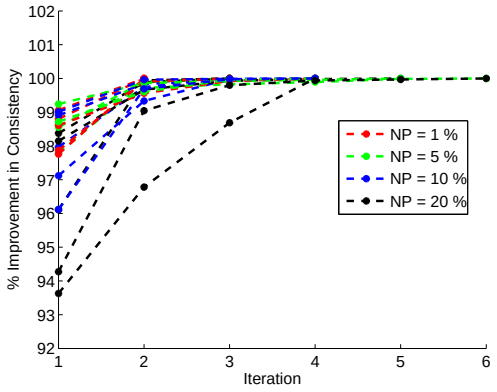


Figure 2.9: Percentage Improvement in consistency with iterations for different amount of noise

2.5.4 Evaluation on Real World Dataset

Since obtaining gold standard land/water labels for multiple lakes at various timesteps is challenging, we used satellite alitmetry databases to compare surface area variations with relative height variations obtained from the altimetry database. Surface area and height (or depth) of a water body are monotonically related, i.e, when area increases, height will either increase or will remain the same but can never decrease. Here, we use Pearson's Correlation as the measure of agreement between the area and height times series. Noisy multi-temporal temporal maps were obtained using the methodology described in [9]. They used MODIS sensor based products (MOD09A1, MCD43A4) [37] which are available at 500m spatial resolution and time interval of 8 days. A SVM classification model was trained to classify each pixel as land or water to obtain multi-temporal maps that are both noisy and incomplete due to missing data. We obtained a altimetry time series data (10 day interval) for 103 lakes from the GREALM database [38–40]. For example, Figure 2.10 shows the surface area variation and the corresponding altimetry time series for Lake Angostura.

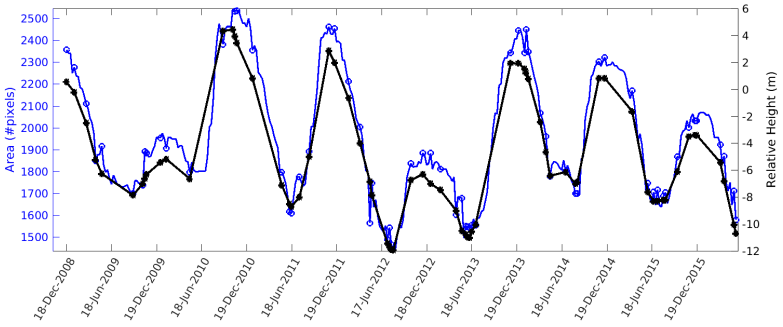


Figure 2.10: Comaprison of surface area time series and relative height time series for Lake Angostura in Mexico. Blue line represents the area time series and the black line represents height time series

Input (Noisy) Labels: We used a SVM based binary classification approach to create 8-day land/water masks at 500m spatial resolution using multispectral data obtained from the MODIS sensor onboard TERRA and AQUA satellites [9]. These masks have errors as well missing labels due to classification errors and missing data. Please refer to the paper [9] for more details on the classification methodology.

We applied **ORBIT-E** on 103 lakes belonging to different parts of the world. Figure x shows the geographical location of these lakes. To evaluate the performance of **ORBIT-E** we used Pearson's correlation between surface area time series and altimetry time series. Figure 2.11 shows the comparison of correlation values before and after label correction using **ORBIT-E**. In Figure 2.11, the Y-axis represents correlation between area estimates (obtained after label correction using **ORBIT-E**) and relative height time series. The X-axis represents the correlation between area estimates (using noisy input labels) and relative height time series. As we can see, **ORBIT-E** is able to improve the correlation for a majority of lakes. Also, note that for some of the lakes with poor correlation before correction, **ORBIT-E** is not able to improve the performance significantly because the error in the initial labels is too high for **ORBIT-E** to recover labels effectively. In other words, this highlights the fact that if the majority of labels are incorrect then **ORBIT-E** will not be able to improve the performance. For correlation values greater than 0.3, we can see that **ORBIT-E** is showing good performance.

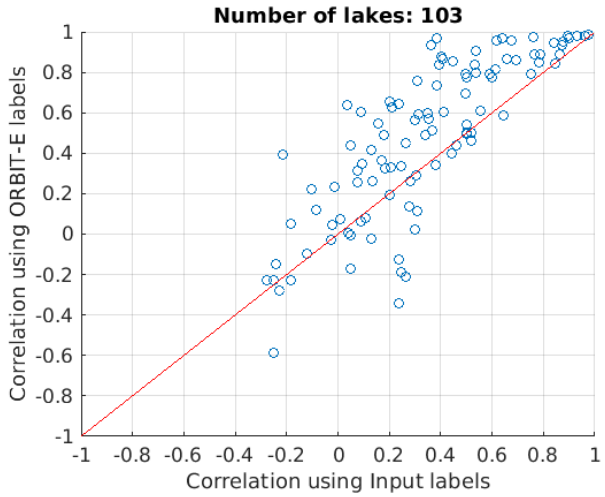


Figure 2.11: Comparison of correlation area values with height variation before and after label correction

2.5.5 Illustrative Examples

Here, we provide some illustrative examples to highlight the performance of **ORBIT-E** under different challenges.

Impact of Algae

Correctly classifying algae can be challenging because they appear as vegetation and cover the water surface. Figure 2.12 demonstrates a scenario in which the errors due to algae can be corrected by **ORBIT-E**. Regions highlighted by red circles show the presence of algae. The presence of algae was confirmed by visual inspection of MODIS multispectral data at daily scale and the corresponding Landsat images. Even though some pixels were incorrectly labelled as land, there were locations which were shallower than the algae locations but still filled with water. In other words, errors due to algae created an inconsistency in water body dynamics where some of the locations that are deep are labelled as land while the shallower locations are labelled as water. This inconsistency in labels and elevation/depth was exploited by the elevation based label correction method to correct the labels.

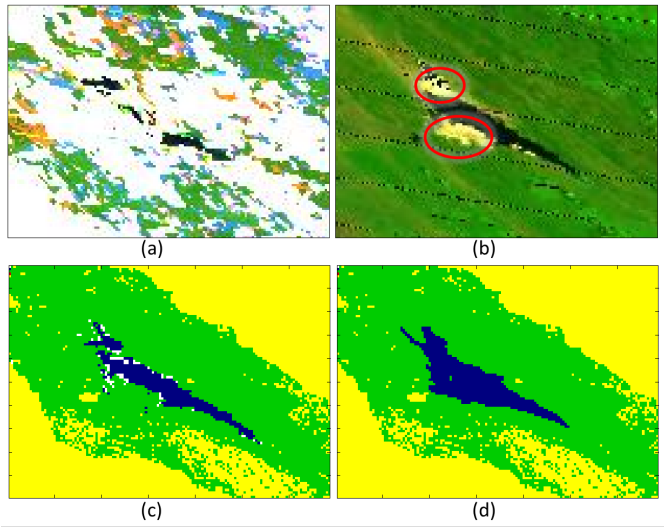


Figure 2.12: An illustrative example showing the performance of **ORBIT-E** in the presence of algae

Impact of Classification Errors

Due to noise in the data, spectral inseparability in feature space, or lack of training data, certain locations can get misclassified by the classification methodology. Figure 2.13 show an illustrative example where the proposed approach successfully corrects the labels. Since the approach does not depend on spatial and temporal auto-correlation, it can correct even large spatially contiguous regions of erroneous labels for both wide and narrow water bodies.

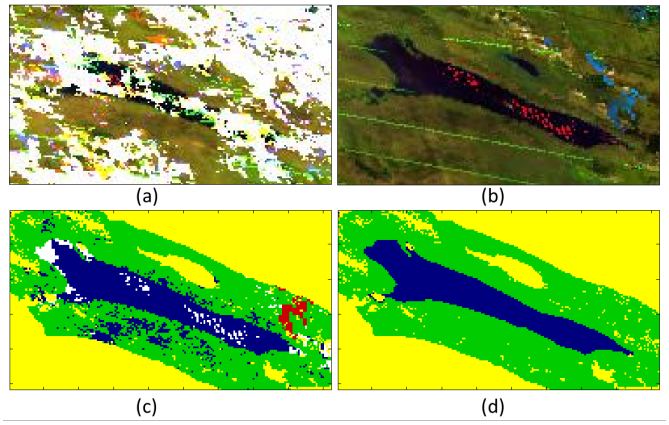


Figure 2.13: An illustrative example showing the performance of **ORBIT-E** in the presence of spatially auto-correlated errors

2.6 Limitations of ORBIT-E

?? As mentioned earlier, the proposed approach aims to find the parameters that maximize the likelihood of observing the given labels. Due to this, **ORBIT-E** expects that a majority of the labels are correct otherwise **ORBIT-E** would make poor corrections. Currently, all sources are considered as having similar weight but in real world applications sources have different quality and hence it should be taken into account. In our application of water monitoring, some timesteps can have very poor label accuracy due to noise from clouds and other environment factors. Such timesteps will not get effectively corrected by **ORBIT-E** and might lead to sudden spurious changes in surface area extents because it does not take into account information from nearby timesteps while making the corrections. Since the extents should follow the physical constraint, it is necessary that the region on interest has a single connected body of water. In other words, all the water bodies in a region should be impacted in the same way by external forces that cause the water bodies to grow and shrink. As mentioned before, **ORBIT-E** and all ordering based methods are more robust to spatial noise than temporal noise.

2.7 Conclusion and Future Work

In this chapter, we proposed a novel approach for improving label accuracy of input class labels in scenarios where there exists an inherent ordering among instances. We demonstrated the utility of the approach for the application of monitoring the extent of lakes and reservoirs. We showed how using elevation constraint can allow the development of much more robust algorithms than existing post classification refinement techniques. Similarly, the proposed method demonstrated that it does a better job in exploiting the bipartite information than existing rank aggregation methods. We demonstrated the efficacy of the method on synthetic datasets with varying level of noise complexity.

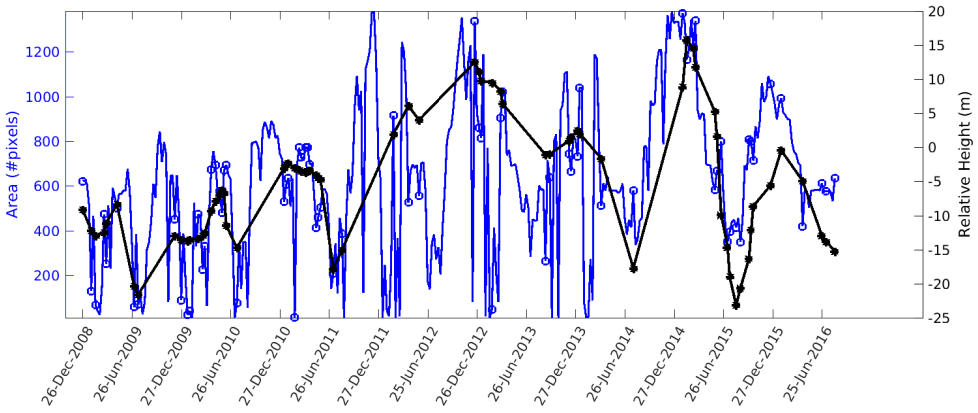
For future work, the proposed approach can be extended for other scenarios such as working with probabilistic labels rather than binary labels, modelling the quality of sources of bipartite information, uncertainty quantification of correct labels. Since, the algorithm converges to a local optimum, it is important to determine bounds on the quality of aggregation with respect to the optimal aggregation. The proposed approach is just one way to effectively utilizing the bipartite information, new methods can be developed for other suitable applications.

Chapter 3

ORBIT-T: Ordering Based Information Transfer across Time

3.1 Introduction

In the previous chapter, we described **ORBIT-E** which uses the elevation constraint and corrects each time step individually to obtain physically consistent land/water labels. However, in real-world scenarios, a single time step might contain large amounts of errors or missing data which could lead to spurious correction in those time steps. For example, Figure 3.1 demonstrate this limitation for Xiaolangdi reservoir in China. As we can see, there are large sudden changes in surface area time series while the altimetry time series is varying smoothly.



(a)

Figure 3.1: An illustrative example showing the limitation of **ORBIT-E** in presence of large errors and missing labels in individual time steps. Blue line represents surface area time series of Xiaolangdi Dam in China. The black line represents the relative height variation.

In most situations, however, a water body grows and shrinks smoothly (except sudden events such as floods) i.e. surface extents of nearby dates are likely to be very similar. Hence, incorporating the temporal context in the label correction process can potentially lead to better performance. Consider a toy example shown in Figure 3.2. Figure 3.2(a) shows ground truth labels for 8 locations across 3 time steps. Figure 3.2(b) shows the observed labels that have errors as well as missing data (shown in red). Figure 3.2(c) shows corrected labels when each time step is corrected independently which leads to abrupt increase in number of water locations at time step t. Figure 3.2(d) shows another possible label configuration that is physically consistent but also has more temporal consistency at the expense of only one more disagreement between observed and estimated labels. Thus, in certain situations, information from nearby time steps can be utilized to avoid spurious corrections.

Current state-of-the-art methods mainly enforce the temporal consistency either for each pixel individually (e.g. majority filters in time or both in space and time) to obtain temporal consistent labels. As shown in the previous chapter, these methods perform poorly when noise and missing data is also spatially and temporally auto-correlated which is very common in our application. For example, in Figure 3.2, the errors in

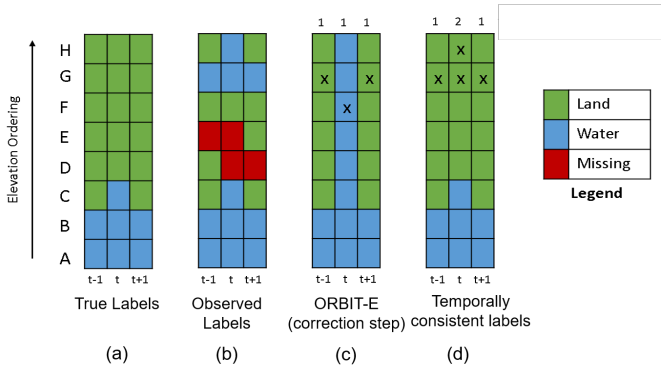


Figure 3.2: A toy example showing the utility of incorporating temporal context in elevation based label correction process.

locations G are likely to remain uncorrected by existing methods. Moreover, existing methods tend to remove real changes in labels as well as they enforce labels in nearby time steps to be same.

In this chapter, we present a new approach, **ORBIT-T** (Ordering Based Information Transfer using Time) that uses elevation ordering to enforce the temporal consistency. Specifically, **ORBIT-T** uses elevation ordering to enforce temporal consistency in total area values instead of consistency in labels of individual pixels. Temporal consistency in total area (water levels) is a more direct constraint and does not penalize real dynamics unlike existing methods. The criteria of enforcing consistency in total area can be incorporated very easily in **ORBIT** framework because estimating water levels for each time is one of the steps of the framework.

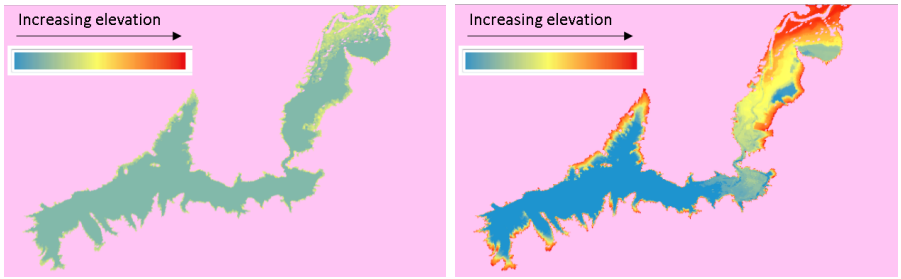
3.2 Related Work

Various post classification label refinement techniques have been proposed that aim to exploit the spatial and temporal context (spatio-temporal auto-correlation) in class labels to detect inconsistent labels and subsequently correct them. A vast majority of work has been done in developing spatial-temporal consistency models using Markov Random Fields to remove "salt and pepper" noise in classification maps. These methods tend to perform poorly in our application because the errors in class labels is also spatially and temporally auto-correlated, as shown in the last chapter. Other approaches

that incorporate process information have also been developed. For example, in [22], authors used transition probability matrices to code domain knowledge about compatible ecological changes in their MAP-MRF model. However, transition probability matrix cannot be assumed or estimated in our application because growing and shrinking of a water body depends largely on external factors such as precipitation, evaporation, in flow, out flow, and ground seepage, etc. that can not be modelled easily.

3.3 ORBIT-T Approach

For this approach, we assume that the elevation ordering is already available. The ordering could either be learned from the data using **ORBIT-E** or can be obtained from existing sources. For example, Figure 3.3 shows elevation map for Kajakai Reservoir from two different sources. Figure 3.3 (a) shows the elevation ordering obtained from SRTM's elevation dataset. Figure 3.3 (b) shows the relative elevation ordering learned from the data itself using **ORBIT-E** on noisy labels available from JRC-Google product. [41].



(a) SRTM based elevation ordering (b) **ORBIT-E** based elevation ordering

Figure 3.3: Comparison of relative elevation orderings at 30m spatial resolution for Kajakai Reservoir, Afghanistan. (a) Relative elevation ordering from SRTM's Digital Elevation data. (b) Relative elevation ordering learned using monthly scale classification maps available from JRC-Google product.

3.3.1 Problem Formulation

Given, an elevation ordering (π) and a set of noisy classification maps $\mathbf{C_i}$, our goal is to estimated physically consistent maps $\mathbf{C_o}$. The estimated labels should be such that they lead to low number of disagreements (henceforth referred to as $Cost_{mismatch}$) with the input labels as well low number of spurious variations in area values (henceforth referred to as $Cost_{transition}$). The objective of reducing $Cost_{mismatch}$ follows from the maximum likelihood assumption made in the previous chapter for **ORBIT-E**. In particular, we assume that the majority of labels are correct and thus the configuration of physically consistent labels that shows maximum agreement (minimum disagreement) with the input label is preferred.

However, both of these costs conflict with each other. The configuration that leads to best consistency with input labels is obtained when each time step is corrected independently but would lead to least temporal consistency in water levels. On the other hand, the most temporally consistent configuration is obtained when all the levels are forced to be same (no dynamics in surface extents) but would lead to very high number of disagreements with input labels. In other words $Cost_{mismatch}$ and $Cost_{transition}$ are inversely proportional to each other.

To capture these observations mathematically, we define the following objective function:

$$Cost_{total} = Cost_{mismatch} + \alpha \; Cost_{transition} \tag{3.1}$$

where, α is tuning parameter.

As α is increased, the above criteria will favour water levels that are more temporally consistent. Hence, $Cost_{transition}$ decreases as α is increased. As mentioned before, $Cost_{mismatch}$ is minimum when water levels for each time steps are estimated independently (i.e. $\alpha = 0$). Hence, as α is increased, $Cost_{mismatch}$ increases.

If a water body has N locations then there are $N + 1$ possible water levels for each time step. For example, Figure 3.4 shows 4 possible label configurations when $N = 3$. Further, if the water body has been observed for T time steps, then there are exponential (N^T) possible configurations of water levels for the given water body. Thus, it would be infeasible to evaluate every possible configuration to find the configuration that minimize the above objective function. Hence, we need to define both cost functions such that

the total objective function can be optimized globally.

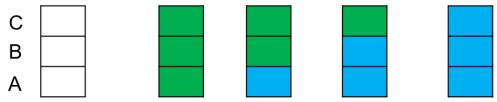


Figure 3.4: A toy example showing number of possible label configurations

We define the measure of mismatch between input labels and estimated labels to be same as used by the **ORBIT-E** approach

$$Cost_{mismatch}(T) = \sum_{t=0}^{T} Err_{\pi}(\theta_t) \tag{3.2}$$

where, $Err_{\pi}(\theta_t)$ represents number of mismatches between input labels and physically consistent labels at timestep t when the water level is chosen to be θ_t

Various methods such as total variation, decomposition based roughness measures, lag auto-correlation, and goodness of fit can be used to quantify the smoothness of a time series. **ORBIT-T** uses total variation as the measure which is defined as -

$$Cost_{transition}(T) = \sum_{t=0}^{T-1} |(\hat{\theta_t} - \hat{\theta_{t+1}})| \tag{3.3}$$

Equation 3.3 represents an aggregate measure of abrupt changes in water levels till time T where $|(\hat{\theta_t} - \hat{\theta_{t+1}})|$ represents absolute difference in water levels between consecutive time steps t and $t+1$. The above criteria enforces temporal consistency as it favours similar water levels in nearby time steps. The choice of total variation as the transition cost function makes it possible to optimize the overall cost function as described in the following section.

3.3.2 Optimization

The goal of **ORBIT-T** is to find the set of water levels $(\hat{\theta_{1...T}})$ such as that the overall objective is minimized -

$$\arg\min_{\hat{\theta_{1...T}}} Cost_{total}(T) \tag{3.4}$$

where,

$$Cost_{total}(T) = Cost_{mismatch}(T) + \alpha\ Cost_{transition}(T) \tag{3.5}$$

and $\alpha \in [0,]$ is the trade-off parameter between the two costs.

We propose a dynamic programming formulation to optimally solve the above objective for any given value of α. The objective function in Eqn 3.5 can be written in expanded form as:

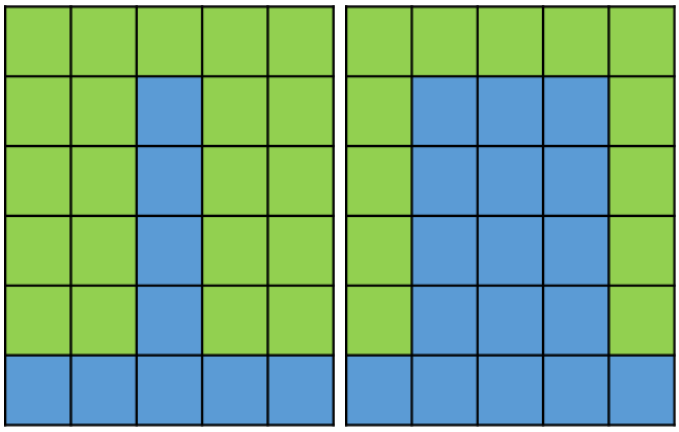
$$\begin{aligned} Cost_{total}(T) &= \sum_{t=0}^{T} Err_{\pi}(\theta_t) + \alpha \sum_{t=0}^{T-1} |(\hat{\theta}_t - \hat{\theta_{t+1}})| \\ &= \sum_{t=0}^{T-1} Err_{\pi}(\theta_t) + \alpha \sum_{t=0}^{T-2} |(\hat{\theta}_t - \hat{\theta_{t+1}})| \\ &+ Err_{\pi}(\theta_T) + \alpha|(\hat{\theta_{T-1}} - \hat{\theta_T})| \\ &= Cost_{total}(T-1) + Err_{\pi}(\theta_T) \\ &+ \alpha|(\hat{\theta_{T-1}} - \hat{\theta_T})| \end{aligned} \tag{3.6}$$

Eqn 3.6 shows the recursive form of the objective function. This recursive equation can be optimized using dynamic programming. The set of water levels that lead to the minimum value of the objective function are chosen as the estimated water levels. Thus, **ORBIT-T** provides the ability to effectively model and control the impact of temporal consistency using the parameter α.

3.3.3 Physical Interpretation of α

As α is increased, the temporal smoothness in water level variations will increase to reduce $Cost_{transition}$. However, over-smoothing of water level variations would not only remove spurious changes but also some real dynamics. In particular, if the cost of removing the dynamics ($Cost_{mismatch}$) is less than the $\alpha * Cost_{transition}$, then the dynamics will be removed. For example, consider the two different dynamics shown in Figure 3.5. The dynamics in Figure 3.5 (a) represents the most abrupt dynamics that

could be present in our data. This dynamics will get removed for $\alpha > 0.5$. In the case of Figure 3.5 (b), $\alpha > 1.5$ would remove the dynamics. To summarize, as α is increased, the smoothing process will impact more number of timesteps.



(a) (b)

Figure 3.5: Relationship between α and persistence of area dynamics. (a) $Cost_{transition} = 8$, $Cost_{mismatch} = 4$ (b) $Cost_{transition} = 8$, $Cost_{mismatch} = 12$

The relationship between $Cost_{mismatch}$ and $Cost_{transition}$ can be used to guide the search for the right value of α. $Cost_{mismatch}$ and $Cost_{transition}$ are inversely related i.e. as α is increased, $Cost_{mismatch}$ decreases and $Cost_{transition}$ increases. However, both these costs can change at different rates. Specifically, if there are spurious changes in total area values, removing them would lead to large drop in $Cost_{transition}$ without increasing $Cost_{mismatch}$ significantly. On the other hand if the change in area values is not spurious then both cost will be very similar.

Another key observation here is that different timesteps may require different values of α depending on the spatial and temporal auto-correlation of errors around that time step. Hence, if a very high value of α is chosen, it might address errors with high spatial and temporal auto-correlation but at the expense of removing true dynamics from other pixels. Therefore, we need to assess the importance of different values of α for any given time step to select the right value of α for that time step. Here, we define a change ratio, δ

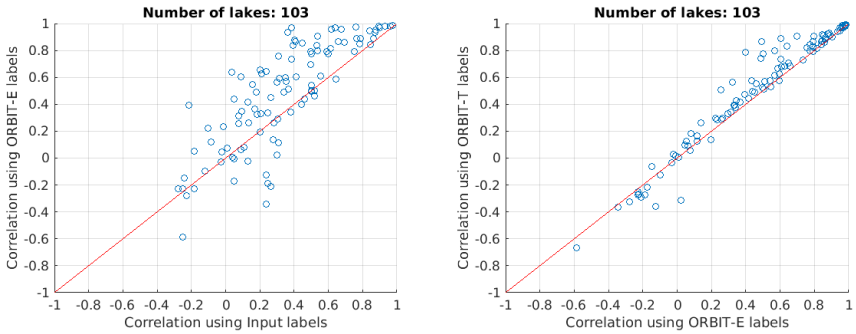
$$\delta_t = \frac{Err_\pi(\theta_t^i) - Err_\pi(\theta_t^0)}{|\theta_t^i - \theta_t^0|} \tag{3.7}$$

where, i represents the i^{th} iteration in the process of changing α. θ_t^0 represent the water level in timestep t when α is 0. The numerator term of the above equation captures the increase in the number of mismatches as α is changed. The denominator term represents the change is water levels due to increase in smoothing effect by the change in α. The small value of δ means that the change in α lead to removal of a spurious transition because it did not lead to comparable increase in the number of mismatches for that time step. The δ value of 1 represents the case where completely consistent labels are corrected and thus potentially true is removed; hence the increase in the value of α is not favourable for this time step.

Although δ provides a way to incorporate timestep specific α values, it is non-trivial to extend the current objective function to include this aspect as it will lead to exponential run-time of the objective function. Thus, in this approach, we will use only a single α to avoid this issue. The α value is chosen to be small so that it can remove some spurious changes without removing the true dynamics. In future, different cost functions could be considered to incorporate timestep specific α values.

3.4 Results

In this section, we evaluate the **ORBIT-T** approach using the altimetry dataset created for evaluating **ORBIT-E**. Figure 3.6 (a) shows the correlation comparison between input labels and labels corrected using **ORBIT-E** (same as Figure 2.11 in Chapter 2). Figure 3.6 (b) shows the correlation comparison between labels corrected using **ORBIT-E** and labels corrected using **ORBIT-T**. Here, **ORBIT-T** was executed using a single value of $\alpha = 0.3$. Since, we chose to use only a single value of α, we chose a smaller value to avoid removing true dynamics while still removing some spurious changes. We can see that **ORBIT-T** was able to futher improve correlation value of some of the lakes.



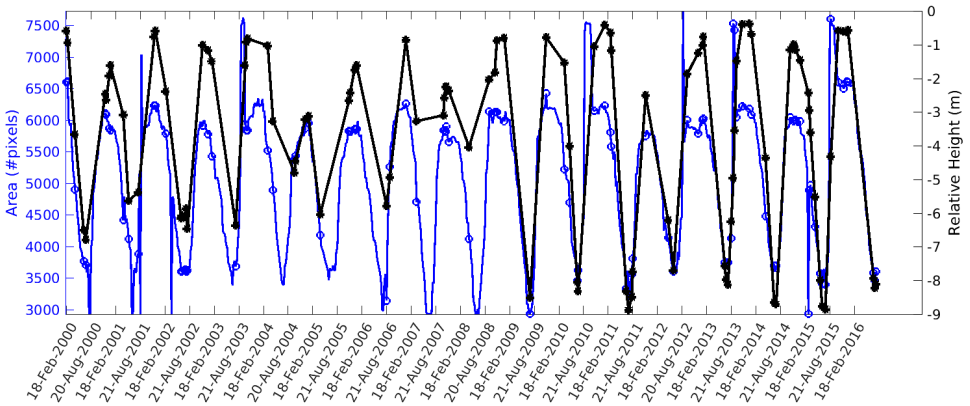
(a) (b)

Figure 3.6: Comparison of correlation area values with height variation before and after label correction. (a) comparison of area-height correlation using input labels and labels after **ORBIT-E**. (b) comparison of area-height correlations using labels after **ORBIT-E** and labels after **ORBIT-T**.

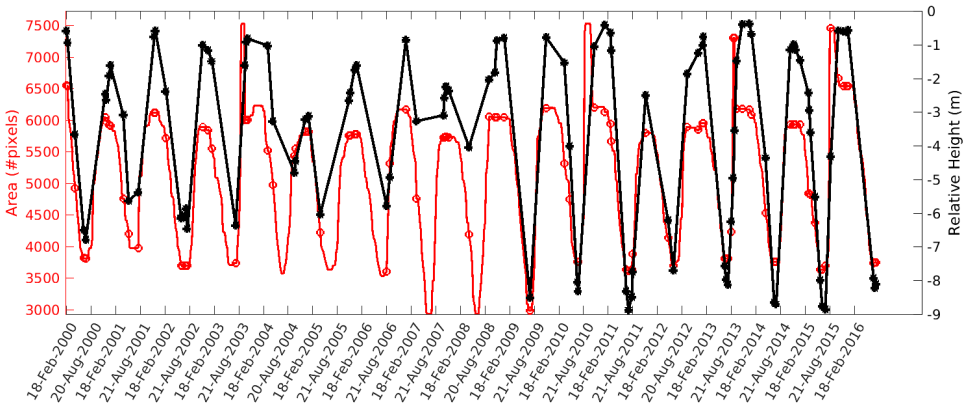
Next, we present some illustrative examples to provide some insights into the performance of the algorithm. Figure 3.7 shows the results for Kainji Dam in Nigeria. In this case, the time series obtained after **ORBIT-E** itself matches very well with the altimetry time series (correlation = 0.78). As we can see, the time series obtained after **ORBIT-T** is very similar to **ORBIT-E**. **ORBIT-T** was able to remove some spurious changes but at the same time retained some of the sudden changes in the area values (correlation = 0.84). This highlights the ability of **ORBIT-T** to preserve sudden changes depending on the label quality.

Figure 3.8 shows the results for Xiaolangdi reservoir in China. In this case, the time series obtained after **ORBIT-E** has a lot more suddent transitions and thus poor correlation of 0.63. On the other hand, **ORBIT-T** based time series is much more smooth and generally matches well with altimetry time series (correlation = 0.83).

Finally, Figure 3.9 shows the results for Sabkhat al-Jabbul in Syria. In this case, the time series obtained after **ORBIT-E** already had relatively smooth area variations but had poor correlation of 0.43 with the altimetry time series. Even after applying **ORBIT-T**, the correlation value increased to only 0.57. This highlights the fact that there can be scenarios in which errors are not due to sudden transitions, and thus **ORBIT-T** will not be able to improve the performance.

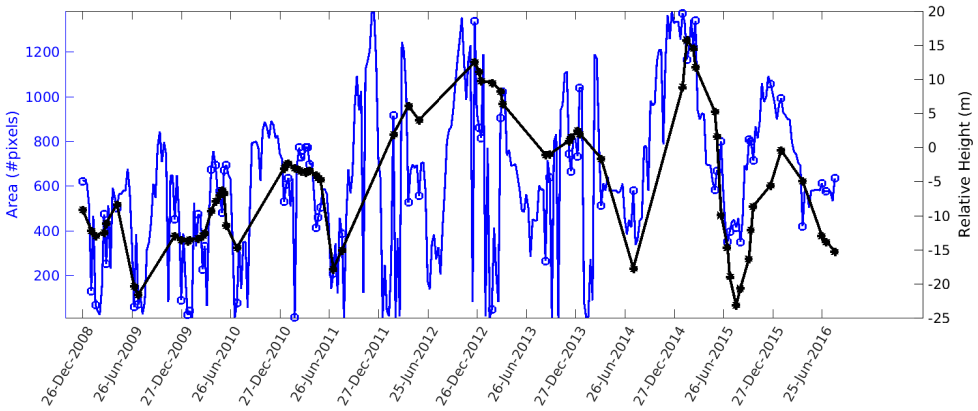


(a)

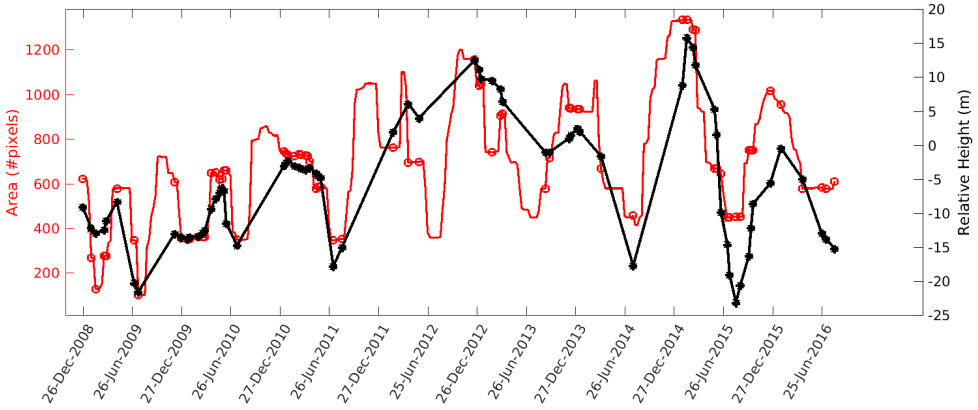


(b)

Figure 3.7: An illustrative example showing the comparison between area time series obtained from **ORBIT-E** and **ORBIT-T**. The black line represents the altimetry time-series. (a) Blue line represents surface area time series of Kainji Dam in Nigeria after **ORBIT-E**. (b) The red line represents surface area time series after **ORBIT-T**.

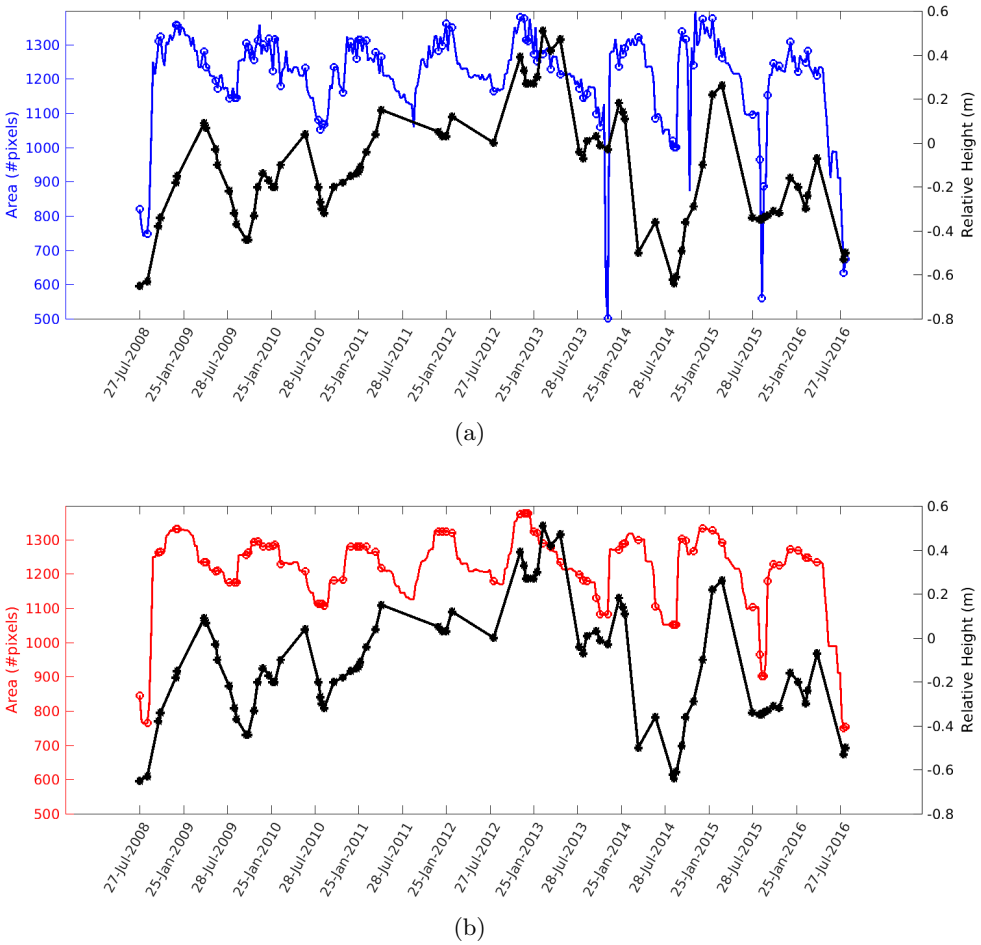


(a)



(b)

Figure 3.8: An illustrative example showing the comparison between area time series obtained from **ORBIT-E** and **ORBIT-T**. The black line represents the altimetry time-series. (a) Blue line represents surface area time series of Xiaolangdi Dam in China after **ORBIT-E**. (b) The red line represents surface area time series after **ORBIT-T**.



(a)

(b)

Figure 3.9: An illustrative example showing the comparison between area time series obtained from **ORBIT-E** and **ORBIT-T**. The black line represents the altimetry time-series. (a) Blue line represents surface area time series of Sabkhat al-Jabbul Lake in Syria after **ORBIT-E**. (b) The red line represents surface area time series after **ORBIT-T**.

3.5 Conclusions and Future Work

In this chapter, we presented a new approach to incorporate temporal consistency in surface area variations of water bodies. Enforcing temporal consistency on total area values instead of individual pixels not only correct the labelling errors but also preserves real changes in labels across time. We showed the utility of the proposed approach for

the surface monitoring application by comparing area variations with height variations obtained from altimetry datasets.

For future work, new functions for the transition cost could be investigated. Furthermore, adapting the cost function to incorporate timestep specific α values would allow more robust smoothing process. Currently, **ORBIT-T** assumes that an elevation ordering is already given. But in future new objective functions can be explored that could simultaneously estimate elevation order and temporally consistent labels.

Chapter 4

ORBIT-S: Ordering Based Information Transfer across Scale

4.1 Introduction

In many applications involving spatio-temporal phenomena, data sets are available at multiple spatial and temporal scale. Often times, due to physical limitations in sensor design and cost issues, there is a trade-off in different resolutions of these datasets. As an example, consider the Earth Observation (EO) data acquired through various remote sensing satellites. MODIS sensor onboard TERRA and AQUA satellites capture earth's surface every day at a coarse spatial resolution (500m). On the other hand ETM+ sensor onboard LANDSAT 7 satellite captures the earth's surface every 16 days but at a high spatial resolution of 30m (see Figure 4.1). Thus, a single sensor is not enough to provide both high spatial and temporal detail required in many earth science applications. For example, hydrological model that aim to characterize and predict floods would benefit from a high fidelity observation data of calibration, assimilation and evaluation [42]. Hence, there is a need to develop methods that can transfer information across scales as well as across time to effectively use the rich complementary information available in these datasets.

Various methods have been proposed that aim to learn a mapping between low spatial resolution (LSR) instances and high spatial resolution (HSR) instances (exploiting local relationships between instances) at timesteps when both datasets are available.

(a) Sep 13, 2004 from MODIS product (b) Sep 16, 2004 from Landsat 7

Figure 4.1: An illustrative example showing False Color Composite images from two different sensors onboard Earth Observation satellites.

Such a mapping can then be used to generate the data at HSR for timesteps when only LSR data is available. These methods in general have limited performance because: (a) a single snapshot in time can have large amount of noise and missing data and hence might not have enough information to learn robust mapping between two resolutions, and (b) the transfer of information across scales is possible only in the duration when both datasets are available.

In this paper, we propose a new approach, **ORBIT-S**(Ordering Based Information Transfer across Scales), that aims to exploit the inherent time-independent relationships that exist between all instances (instead of using local relationships) to transfer information across these complementary datasets. The use of these high-level, and time-independent relationships not only make the method more robust to noise and missing values but also allows for information transfer across datasets available from different duration (without necessarily having overlapping time steps). Specifically **ORBIT-S** uses the elevation constraint that govern surface water dynamics to transfer information across these noisy classification maps at different scales to produce maps that are not only accurate but also produced at both high spatial and temporal resolution (30m maps at daily scale for the case of MODIS and LANDSAT). The key idea here is that in order to create maps at HSR, we need to estimate labels of only some HSR pixels.

The elevation constraint can then be used to estimate labels of other pixels.

The remainder of the paper is as follows: Section 4.2 provides summary of the related work, In Section 4.3, we describe the **ORBIT-S** approach. Section 4.4 provides insights into conditions that are required for effective information transfer across scale in **ORBIT-S** approach. Section 4.5 provides results and discussion. Finally, Section 4.6 provides the summary and some promising future research directions.

4.2 Related Work

Various data fusion techniques have been proposed in remote sensing literature that aim to improve the temporal resolution of fine spatial resolution data by combining information from sensors with differing spatial and temporal characteristics. The most widely used approach is to learn a mapping between raw feature values from two different scales and use the mapping to create synthetic HSR image from the corresponding LSR image [43–47]. However these methods have several limitations. First, data from each scale should be precisely calibrated and spectrally normalized to common wavebands which limits their use to datasets with similar multispectral bands. Second, these models learn relationship between pixels at two scales using pair of images at the same date and use it in nearby dates to create synthetic high resolution images. Due to various atmospheric disturbances (noise, outliers and missing data), a single image pair might not have enough information to learn a robust relationship. Furthermore, these methods cannot handle land cover changes that are not captured in the image pair used to learn the mapping. Finally, these methods are not applicable in situations where there are no overlapping timesteps (e.g. sensors with few days of offset in their revisit cycles). Finally, in our scenario, we have access to only label information and raw features of pixels are not available. According to the standard terminology used in the information fusion literature, our case corresponds to the decision level fusion category where outputs or decisions from two different sources are combined together.

A second set of data fusion algorithms are based on unmixing techniques. Spectral unmixing methods use linear (or non-linear) spectral mixture model to extract end members and their proportions on a sub-pixel scale. These proportions/fractions are

then downscaled using different strategies. For example, [48–54] use statistical techniques such as particle swarm optimization to distribute labels at high resolution based on the fraction values. A major challenge with these algorithms is to find the appropriate end-members for unmixing. Some methods [55–58] have used existing elevation datasets (e.g. SRTM) to convert fraction value to labels at high resolution by filling pixels with water label in the increasing ordering of their elevation. These set of methods are not applicable to our setting because we are given binary labels and not fractions.

4.3 Approach

Here, we describe the **ORBIT-S** approach. The key idea behind this approach is to use high level, and time-independent relationships between instances (elevation ordering in our application) to transfer information across scales. The global nature of elevation ordering constraint enables a more effective transfer of information compared to using local relationship between instances. Specifically, if we can confidently estimate the label of an instance at HSR, then due to the ordering constraint the labels of other instances can be estimated as well. Furthermore, elevation ordering is an inherent property of locations of the basin and hence does not change with time in most cases and thus enables information transfer even when two datasets are from different time period.

4.3.1 Problem Formulation

We define the noisy binary classification maps at HSR (and low temporal resolution) as ($\mathbf{H_i}$) and noisy binary classification maps at LSR (and high temporal resolution) as ($\mathbf{L_i}$). Given these inputs, our goal is to estimate accurate classification maps at HSR and high temporal resolution ($\hat{\mathbf{H}}_\mathbf{o}$). Furthermore, we define the set of instances at HSR as $\mathbf{P_h}$ and the set of instances at LSR as $\mathbf{P_l}$. The mapping grid $\mathbf{G}$ defines the mapping between a LSR instance and its corresponding instances at HSR. Specifically, $\mathbf{G_j}$ represents the set of instances at HSR that belong to the LSR instance P_l^j. To reduce notations, we assume that each LSR instance has same number of HSR instances within it (defined as g for grid ratio). For example, Figure 4.2 shows a square grid of 600m spatial resolution overlaid on top of the elevation ordering for Kajakai Reservoir at 30

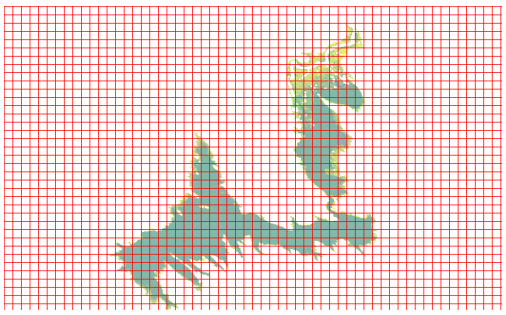


Figure 4.2: Relative elevation ordering for KajaKai Reservoir, Afghanistan from SRTM's Digital Elevation data at 30m spatial resolution. Cells shown in red color corresponding LSR pixels at 600m spatial resolution.

m spatial resolution. Thus, in this case, the grid ratio (g) is 400.

In order to transfer information across the two scales, we need to establish the relationship between the label of a LSR instance and the labels of its corresponding HSR instances. In this approach, we make following assumptions:

Assumption A1: The label of a LSR instance is generated by aggregating the labels of its corresponding HSR instances. Specifically, we assume that a LSR instance will be labelled as water if it contains more than a certain number of HSR instances filled with water (cut-off threshold k).

Assumption A2: In the most general setting, each LSR instance can have a different cut-off threshold which can also vary with time due to varying spectral properties of the land cover types enclosed within the LSR instance. In this approach, we make an assumption that there exists a single cut-off threshold across all pixels and all timesteps. In the results section, we show the impact of this assumption on the performance of the approach.

Assumption A3: The true shape of the water body and the low resolution data grid are independent of each other. In other words, a lake's extent can be considered arbitrary with respect to the low resolution data grid. In the results section, we provide some insights about the validity of this assumption.

The **ORBIT-S** approach has five main steps: 1) Estimate elevation ordering at HSR. 2) Estimate elevation ordering at LSR. 3) Estimate physically consistent classification maps at LSR. 4) Use elevation ordering at HSR from Step 1 and good quality

classification maps at LSR from Step 3 to estimate confident labels at HSR and finally 5) Estimate remaining labels at HSR using the elevation constraint. Next, we describe these steps in detail.

4.3.2 Method

Step 1. Estimate Elevation ordering at high spatial resolution ($\hat{\pi}_h$)
In this step, $\mathbf{H_i}$ is used to estimate elevation ordering at HSR. Specifically, use **ORBIT-E** to estimate the relative ordering from noisy classification maps at HSR. Since, this step is independent from LSR maps, if a high quality elevation based ordering is available from any external source **ORBIT-S** can directly use it in the subsequent steps.

Step 2. Estimate Elevation ordering at low spatial resolution ($\hat{\pi}_l$)
One way to estimate $\hat{\pi}_l$ would be to use **ORBIT-E** with $\mathbf{L_i}$ similar to Step 1. However, we estimate $\hat{\pi}_l$ using $\hat{\pi}_h$ and $\mathbf{L_i}$ as it allows us to 1) estimate the threshold k together with $\hat{\pi}_l$ and 2) ensure that ordering learned at LSR is coherent with ordering at HSR.

As mentioned before, each LSR pixel contains g number of HSR instances. Furthermore, each of the g instances have a local ranking based on the global ranking at the high resolution. For example, Figure 4.3 shows the local and global ordering relationship of HSR pixels within a LSR pixel. In this case, g is 4.

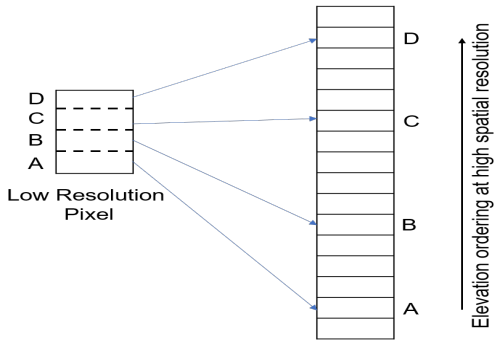


Figure 4.3: An illustrative example showing local and global ordering of a LSR pixel. In this case the grid ratio is 4.

As mentioned before, we have made an assumption that binary labels at LSR can be generated by aggregating labels at HSR and there exists a single threshold k that

determines the label of an LSR pixel. For example, Figure 4.4 shows relationship between labels at LSR and HSR when k is 1. Specifically, if k is 1, and the LSR pixel is labelled as water, then we can be certain that atleast the bottom k (in this case, 1) HSR pixels in the LSR pixel should be labelled as water. In other words, if a LSR pixels is labelled as water then the HSR pixel with local rank k within that LSR pixel should also be water (including all the HSR pixel with local rank $\leq k$). Similarly, if k is 1, and the LSR pixel is labelled as land, then we can be certain all the HSR pixels in the LSR pixels should be land (in general, all the HSR pixels with local rank $\geq k$ should be land). In other words, for any threshold k, we can estimate the label of the HSR pixel with local rank k for each LSR pixel. Figure 4.5 shows the illustrative example for the case $k = 3$. To summarize the above observation, each HSR pixel with local rank k will have the same label as its parent LSR pixel.

Another key observation that can be made here is that if the HSR pixels with local rank k have physically consistent labels, then all other labels inferred using these pixels will also be physically consistent. In other words, if we correct the LSR labels using **ORBIT-E** or **ORBIT-T** by ranking the LSR pixels based on global ranking of the corresponding HSR pixels with local rank k, we can ensure that there will be no physically inconsistent label at HSR during the downscaling process. For example, in Figure 4.4, if we consider HSR pixels with local rank $k = 1$ within each LSR pixel, they can be ordered as c>b>d>a (high elevation to low elevation). Given, this ordering and estimated labels of the corresponding HSR pixels, we can see that there is a physical inconsistency in labels at LSR because a deeper HSR pixel (D) is labelled as land whereas shallow HSR pixels (B and C) are labelled as water. Similarly, Figure 4.5 shows the label and ranking relationship for $k = 3$. In this case, the LSR pixels are ordered as d>c>b>a and the corresponding label estimates do not show any physical inconsistency.

Thus in general, using π_h, LSR instances can be ranked in number of ways. Specifically, we define a LSR ordering π_l^k as the ordering obtained by using HSR instances with local rank k within each LSR instance. Thus, there can be g possible LSR orderings that can be generated from π_h.

In the absence of any external information about the correct labels, we select the LSR ordering that leads to the least amount of corrections in $\mathbf{L_i}$ (using the maximum

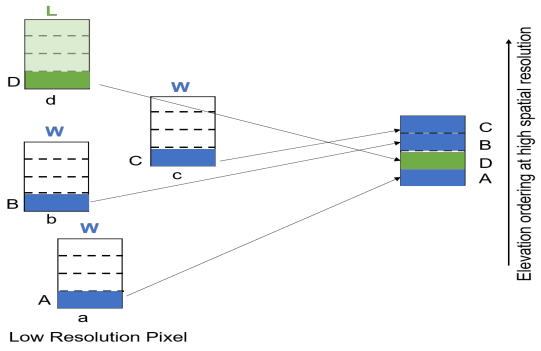


Figure 4.4: An illustrative example showing relationship between labels at LSR and HSR for $k = 1$

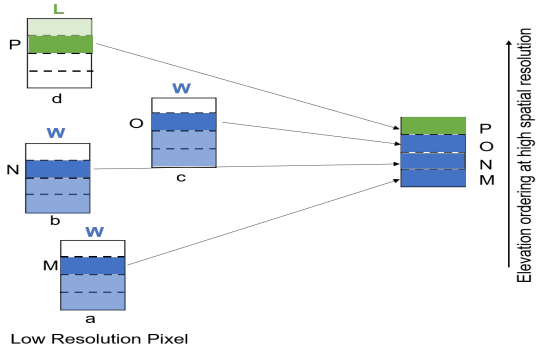


Figure 4.5: An illustrative example showing relationship between labels at LSR and HSR for $k = 3$

likelihood assumption as used for **ORBIT-E**). This would also automatically provide an estimate of k which will be used in next steps of the algorithm. Specifically, k value corresponding to the selected LSR ordering is chosen as the cut-off threshold. Based on our toy example in Figure 4.4 and Figure 4.5, threshold $k = 3$ will be selected as it leads to no physical inconsistencies in the LSR labels.

Step3. Estimate accurate and physically consistent classification maps at LSR ($\mathbf{L_o}$)

Once the LSR ordering is selected in the previous step, we can use it to correct each classification map (L_i^t) using **ORBIT-E** or **ORBIT-T** to obtain physically consistent classification maps at LSR (L_o^t). This step plays an important role because if the

information in these LSR maps is of bad quality then the errors will get propagated in the estimated HSR maps as well.

Step 4. Estimate confident labels in $\hat{\mathbf{H}}_\mathbf{o}$

Once, physically consistent labels are obtained at LSR, we can use them to estimate labels of some HSR pixels using the strategy described in Step 2. Specifically, if a LSR instance is labelled as water then all HSR pixels with local rank $\leq k$ should be water (Assumption A1 and A2). Similarly, if a LSR pixel is labelled as land then all HSR pixels with local rank $\geq k$ should be labelled as land. Using this knowledge, we can confidently estimate physically consistent labels for some of the HSR instances within each LSR instance.

Step 5. Estimate remaining labels in $\hat{\mathbf{H}}_\mathbf{o}$

After Step4, there will be a lot of unknown labels in $\hat{\mathbf{H}}_\mathbf{o}$. For example, if k is $g/2$, then half of the labels in $\hat{H}_o$ would be unknown after Step 4. In this final step, we use $\hat{\pi}_h$ to estimate the labels of remaining instances. Specifically, we first find the shallowest HSR pixel that is labelled as water ($Pivot_w$) and label all the instances deeper than it as water as well due to the physical constraint. Using the same rationale, we find the deepest HSR instance labelled as land ($Pivot_l$) and label all instances shallower than it as land. This step significantly reduce the number of unknown labels. Finally, instances that are between pivots $Pivot_w$ and $Pivot_l$ remain unlabeled. Note that the elevation of $Pivot_l$ will always be higher that the elevation of $Pivot_w$ because Step 4 ensures that only physically consistent labels are estimated. Ideally, the gap between $Pivot_l$ and $Pivot_w$ should be as small as possible. In section 4.4 we provide some insights about scenarios where this gap would be small.

Figure 4.6 shows an illustrative example that demonstrates some of the steps of the **ORBIT-S** approach for Red Fleet Reservoir in USA. Figure 4.6 (a) shows the bathymetry information of the reservoir at 10m spatial resolution available from an elevation dataset [59]. To simulate a possible surface extent, the bathymetry was filled upto a certain elevation starting from the deepest location, as shown in Figure 4.6 (b). To simulate the corresponding LSR extent, we created a 200m square grid (g=400) and created binary labels at low resolution by aggregating HSR labels using the threshold $k = 200$. Figure 4.6 (d) shows the estimated HSR extent after Step 4 and Figure 4.6 (e) shows the extent after Step 5. We can see that the approach was able to reconstruct

the HSR extent well. In this case, the approach was able to estimate labels of all pixels.

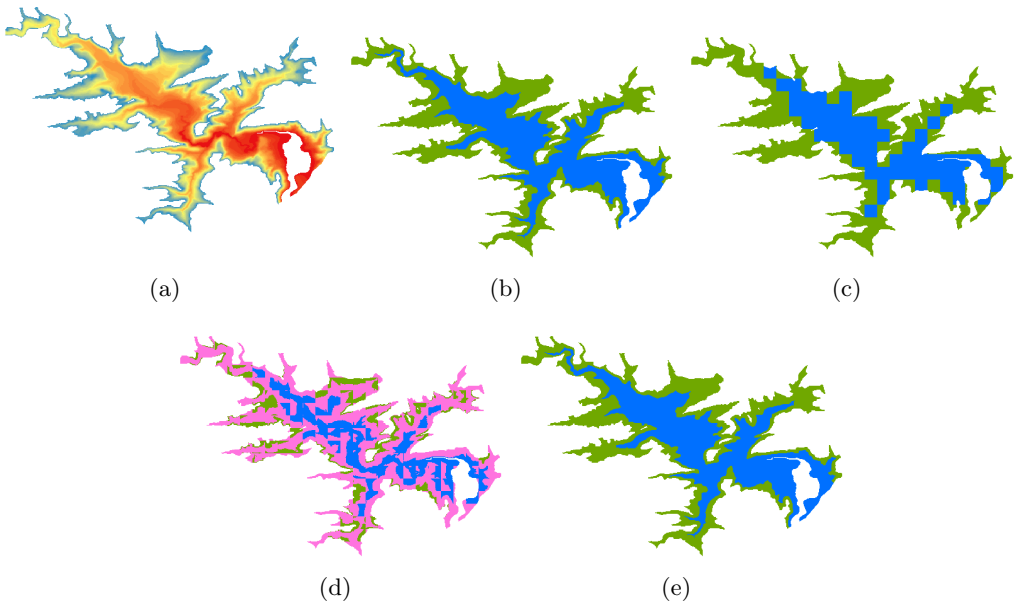


Figure 4.6: An example showing improvement in spatial resolution of surface extent maps for Red Fleet Reservoir, USA. (a) High Quality bathymetry at 10m spatial resolution. Synthetic HSR extent map created for reference using the bathymetry. (c) Perfect LSR map at 200m. (d) Estimated HSR map at 10m after Step 4 (e) Estimated HSR map at 10m after Step 5. In this example, **ORBIT-S** was able to estimate labels of all the pixels (no unknown labels).

4.4 Theoretical Analysis of ORBIT-S

As mentioned earlier, after the final step of **ORBIT-S**, there can be two types of issues with the estimated HSR maps - 1) propagation of error from erroneous LSR labels to HSR labels and 2) number of instances with unknown label (gap between two pivots ($Pivot_l - Pivot_w$). The first type of error depends on the quality of input LSR labels and hence will reduce as their quality increases. However, number of unknown labels depend on factors that are independent of data such as shape and size of the lake, its elevation structure and difference in the resolutions of the two scales. In this section, we provide insights into the impact of these characteristics on the number of unknown

labels produced by **ORBIT-S** approach. Here we assume that LSR labels are perfect.

In order to understand the properties of unknown labels, it is useful to consider the growing and shrinking process of a water body. In particular, a water body grows and shrinks in layers (contours). For example, Figure 4.7 shows the contour like bathymetry for Medicine Lake in California.

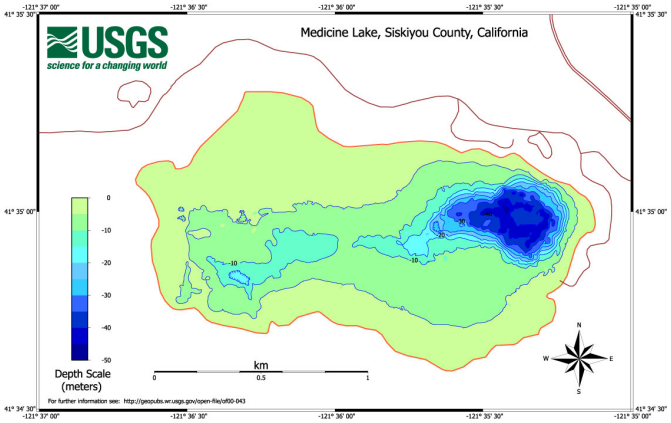


Figure 4.7: A bathymetric map of Medicine Lake in Californinia

Given this nature of the elevation structure, if we can identify a water pixel in any contour at HSR, then we can be certain that all the deeper contours should be filled with water. Similarly if we can identify a land pixel in any contour, then we can be certain that all the shallower contours will have no water. If we can identify one of the water pixel in the last contour filled with water (henceforth referred to as boundary water pixel, B_w) and similarly a land label in the first empty contour (henceforth referred to as boundary land pixel, B_l), then the errors will be bounded to the two edge contours.

Here, we provide a probabilistic bound on ability of **ORBIT-S** approach to identify boundary water and land pixels.

Identification of a boundary water pixel: Step 4 of **ORBIT-S** approach estimate labels at HSR using the assumption A1. Specifically, if a LSR instance is labelled as water then all the HSR instances within it with local rank $\leq k$ are labelled as water. Consider a boundary water pixel at HSR, B_w that belongs to the LSR pixel P_l^j. According to the assumption A1, B_w will be labelled as water in Step 4 only if 1) P_l^j is labelled as water and 2) B_w has local rank $\leq k$. However, both these conditions conflict

with each other because boundary water pixels are the last set pixels that get filled with water by definition and hence determine label of the corresponding LSR pixel. For example, if B_w has local rank $< k$ then the total number of water pixels in P_l^j will be less than k and hence the LSR pixel cannot get a water label. In fact, both conditions satisfy simultaneously only when B_w itself has local rank $= k$. In this case, the P_l^j will have **exactly** k water pixels and hence will get a water label. To summarize, a boundary water pixel can be identified by **ORBIT-S** approach if the corresponding LSR pixel has exactly k HSR water pixels in it. Note that this is not the only condition in which boundary pixels can be detected. In order to do the worst case analysis, here we assumed that each LSR pixel will have only one boundary water pixel at HSR. But in reality, a LSR pixel can have more than one HSR boundary water pixels and which implies that there can be boundary water pixels that have local rank $< k$ also belong to a LSR pixel labelled as water.)

Further, a LSR pixel can have $g+1$ possible number of states (no HSR water pixels to all HSR pixels being water), where g is the number of HSR instances in a LSR instance). Thus, the probability of a LSR pixel to have exactly k HSR water pixels can be defined as $\frac{1}{(g+1)}$. Since, in general the shape of the lake is not related to the imposed mapping grid (Assumption A3), local rank of boundary pixels in a LSR pixel does not determine local rank of boundary pixels in other pixels. Hence the probability of detecting at least one boundary water pixels can be defined as:

$$1 - (1 - \frac{1}{g+1})^W \tag{4.1}$$

where, $(1 - \frac{1}{g+1})$ is the probability that a LSR pixel does not have exactly k HSR water pixels and W is the number of LSR water pixels through which the last filled contour passes.

Identification of a boundary land pixel: Following the above discussion, B_l that belongs to the LSR pixel P_l^j will be detected as land only if P_l^j has **exactly** $k-1$ water pixels. In this case, step 4 of **ORBIT-S** approach will label all the remaining HSR pixels $(g-(k-1))$ as land which will include B_w by definition.

Thus, the probability that the unknown labels will be restricted within the two edge

contours is

$$(1-(1-\frac{1}{g+1})^{W})*(1-(1-\frac{1}{g+1})^{L}) \tag{4.2}$$

where, L is the number of LSR land pixels through which the last filled contour passes. As we can see, the probability that the unknowns will be localized to boundary contours will increases exponentially as the perimeter at LSR increases. On the other hand, the probability decreases as ratio between scales is increased. Note that the probability does not depend on the value of k due to assumption A3. In the results section, we will provide empirical evidence for the validity of assumption A3.

4.5 Results

In this section, we evaluate different aspects of the **ORBIT** framework using a semi-synthetic dataset as well as a real-world datasets. The proposed framework has not been compared with other baselines because to the best of our knowledge there are no existing methods that can do information transfer across scale using classification maps.

4.5.1 Synthetic Data Experiments

Impact of Shape and Size

In this experiment, we will provide the empirical analysis of the bound on the number of unknown labels using a semi-synthetic dataset. The goal of this experiment is to analyze the amount of unknown labels produced during the information transfer process when perfect labels at low spatial resolution are given. The dataset contains 620 lakes of varying shapes and sizes. Figure 4.8 shows the geographical locations of the 620 lakes in the dataset. First, we obtained the elevation structure (bathymetry) for these lakes from USGS's National Elevation Dataset which is available at 10m spatial resolution for USA [59]. Figure 4.9 shows a sample of 12 lakes of different shapes and size from this dataset. Note that some of the water bodies have a flat blue region. This is due to the flattening artifact that happens in using radar instruments to map the terrain.

Since, the microwave signal cannot penetrate the water surface, there is no elevation information under the surface of the water body at the time of data collection.

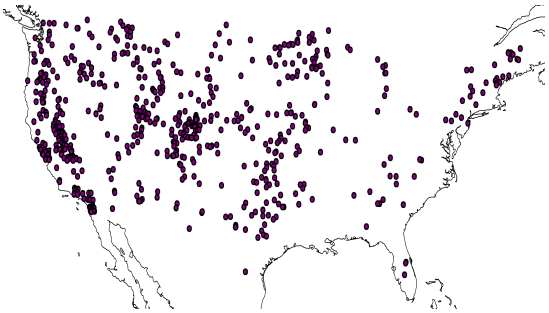


Figure 4.8: Geographical location of the 620 lakes in the synthetic dataset.

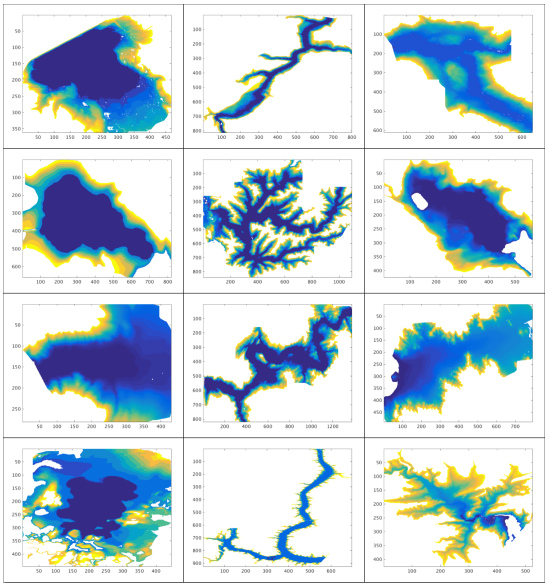


Figure 4.9: Elevation structure of a sample of 12 lakes from the dataset.

Given the bathymetry, multiple ground truth surface extents (H_{gt}) can be created at HSR for these lakes. Specifically, for each lake we created surface extents by filling the lake's bathymetry at different levels. To simulate maps at the LSR scale, we first created synthetic square grids of 200m ($g = 400$) and 400m ($g = 1600$) spatial resolution and used them to create the corresponding LSR surface extent maps (L_{gt}) for two different values of k ($0.5 * g$ and $0.75 * g$). Specifically, if a LSR pixel contains more than k HSR

water pixels, then it is labelled as water. Otherwise, it is labeled as land. We then used the elevation ordering and L_{gt} to estimate HSR maps (H_{est}). These HSR maps will have unknowns in them but no erroneous labels because L_{gt} have no errors. In this experiment, we made use of perfect LSR maps so that we can analyze the impact of shape and size only.

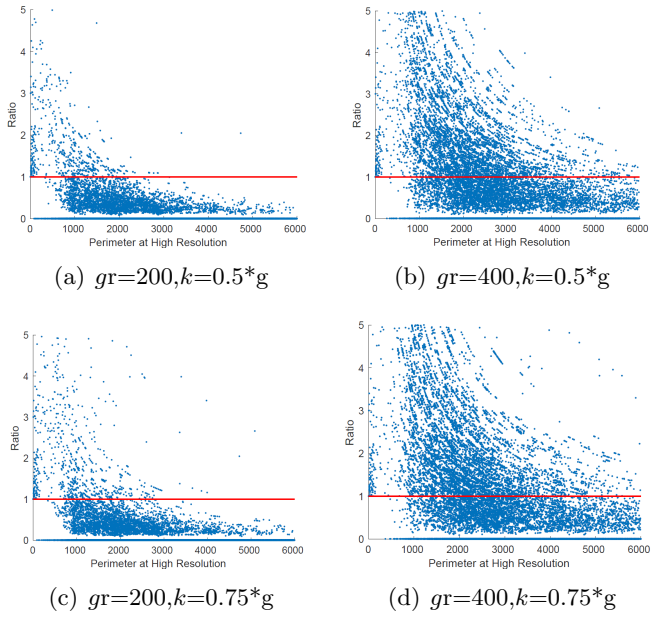
Since, the number of unknowns will be impacted by the shape and size of the water body, it should be normalized appropriately so that the performance can be compared across different lakes. Here, we define a metric, U_{ratio} to normalize the number of unknowns with respect to the length of last water contour (perimeter of ground truth extent at high spatial resolution). This metric is more informative than using size of the water body (total number of pixels in the water body) for normalization because errors on the boundary will be very small compared to the full size of the body.

$$U_{ratio} = \frac{\#\ unknown\ pixels}{perimeter\ at\ high\ resolution} \tag{4.3}$$

Thus, U_{ratio} can be seen as the approximate measure of number of unknown layers around the true water boundary contour. U_{ratio} value less than 1 signifies that the number of unknowns are less than the number pixels on the boundary. Thus, lower U_{ratio} values would mean better information transfer.

Figure 4.10 shows the value of U_{ratio} as a function of extent size for 18000 extents derived from 620 lakes for 4 different settings of and g and k. Note that the probability in Eqn 4.2 is defined with respect to the perimeter at low resolution (P) whereas, in the scatter plots, we have used perimeter at high resolution. This was needed to keep the perimeter same (X-axis) across different mapping for comparison. As described in section 4.4, as the perimeter increases, the probability that unknowns will be localized around the boundary contours increase exponentially which can be empirically observed in Figure 4.10 (a). Since, the probability is inversely proportional to the grid ratio, we observe increase in U_{ratio} values when g is increased from 400 to 1600 (Figure 4.10 (b)). On the other hand, as k is changed from $0.5 * g$ to $0.75 * g$ (Figure 4.10 (c) and (d)), we observe no significant change in the relationships which provides empirical evidence that the independence assumption (Assumption A3) stated in 4.4 holds true in real-world situations. In other words, the grid structure and the lake shape do not introduce bias

towards any particular value of k.



(a) gr=200,k=0.5*g (b) gr=400,k=0.5*g

(c) gr=200,k=0.75*g (d) gr=400,k=0.75*g

Figure 4.10: Relationship between lake extent size and number of unknown pixels labeled by **ORBIT-S** for two different values of g and k. Each point in the plot represents U_{ratio} for an extent. The X-axis represents the perimeter at HSR.

4.5.2 Real Data Experiments

In this section, we present the evaluation of the approach on a real-world reference dataset. Here, we will assume that we have access to neither perfect elevation ordering nor perfect LSR labels. We used noisy multi-temporal classification maps at 30m (available from a product created by European Space Agengy's Joint Research Commission in collaboration with Google [41]) as the source for high spatial resolution input data. This product was created by analyzing the entire LANDSAT archive from March 1984 till October 2015. For each month a global land/water mask is available where pixels are labeled as either land, water or unknown. Currently, this is the state-of-the-art product for obtaining global multi-temporal land/water masks at 30m spatial resolution. From here onwards, we will refer to this dataset as JRC. We used this dataset to learn the relative elevation ordering at HSR using **ORBIT-E** approach. For noisy low resolution

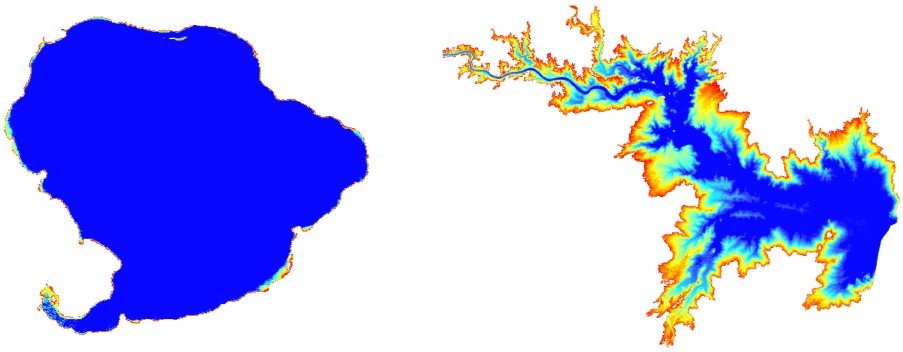
maps, we used daily multi-temporal classification maps at 500m [60]. These labels were created using the 500-m MODIS surface reflectance product, MOD09GA (version 6). Each 500m pixels was labelled as land, water or unknown. This is the most recent product that provides daily scale maps globally at 500m spatial resolution. From here onwards, we will refer to this product at J500. For this pair of datasets, the grid ratio is 256.

To evaluate the different aspects of the **ORBIT** framework, we used the high resolution reference maps created for different lakes for a date in the year 2016. Figure 4.11 shows the geographical location of the 56 lakes used in this dataset.

Figure 4.11: Geographical location of the 56 lakes in the real-world dataset.

Reference Dataset Description

First, we randomly selected lakes that showed sufficient dynamics during the JRC data period (1984-2015). This criteria was needed because if the lake did not show change in its area over the period, then it is not possible to learn the relative ordering of the locations in that lake. Thus, these type of lakes are out of the scope of the **ORBIT** framework. For example, Figure 4.12 shows estimated elevation ordering of two different lakes using **ORBIT-E** approach. Dark blue region in these images represent region that was always water. The color gradient changes from dark blue to red as the elevation increases. As we can see, Lake Apoka did not change much over the JRC data period and thus contains only few dynamic pixels. On the other hand Nath Sagar lake showed a lot of variation over the JRC period and thus have ordering information available for a lot of pixels.



(a) Lake Apoka, USA (b) Lake Nath Sagar, India

Figure 4.12: Examples of elevation orderings estimated from JRC data using **ORBIT-E**. Blue color corresponds to the lowest elevation and red color corresponds to the highest elevation.

A high resolution satellite image was selected for each lake from the year 2016. Specifically, we used Sentinel-2 optical imagery which is at available 10m spatial resolution to create the reference land/water masks for these lakes. Since, elevation ordering is estimated at 30m spatial resolution, we converted 10m resolution reference maps to 30m resolution maps by aggregating using the majority voting strategy. The use of 10m reference maps to obtain 30m reference maps ensured more robustness at the boundary pixels at 30m. In order to create the reference maps, we first trained lake specific SVM classification model by using training data from the lake itself. The initial map thus obtained was then manually verified by the human interpreter to make corrections wherever necessary. The regions where the human interpreter could not make a decision were marked as unknown.

Evaluation Metric

In order to quantify the performance, we have used the following metric (henceforth referred to as $Error$) -

$$Error = \frac{\#\ errors}{lake's\ perimeter\ in\ the\ reference\ map} \tag{4.4}$$

$Error <= 1$ would mean that the number of errors are less than the length of the last contour itself. This metric is more informative in our setting than using the traditional

metric, error rate (normalizing by total number of pixels), because errors on the boundary in general will be very small compared to total number of pixels. Unless stated explicitly, unknown labels will be treated as errors in the following experiments.

Performance of ORBIT-E

In this experiment, we aim to evaluate the quality of the elevation ordering estimated by **ORBIT-E** approach. Here, we provide reference maps as input to the **ORBIT-E** approach to correct physical inconsistencies in the reference maps. If the elevation ordering learned from JRC data perfectly matches the reality then it should not introduce any corrections to the reference map. Thus, lower the value of *Error*, better is the quality of the estimated ordering. Figure 4.13 shows the performance of **ORBIT-E** in estimating the relative elevation ordering. The X-axis represent the perimeter of the water body in the reference maps. The Y-axis represents the amount of error introduced in the reference maps by the **ORBIT-E** approach. As we can see, a vast majority of the lakes have *Error* less than 1.

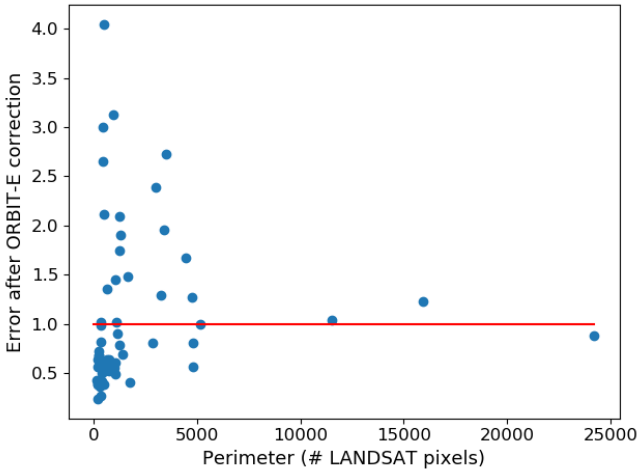


Figure 4.13: Relationship between lake perimeter and performance of **ORBIT-E**.

Performance of ORBIT-S

In this experiment, we aim to analyze the ability of the framework to create high quality (and high resolution) maps from noisy low spatial resolution maps. Here, we used the

low resolution labels available from the J500 product as input labels. Specifically, for each lake we obtained land/water labels for all the days in 2016. **ORBIT-T** was then applied on these labels by setting $k = 128$ (half the size of total number of HSR pixels in a LSR pixel). The physically consistent labels thus obtained were then downscaled using **ORBIT-S**.

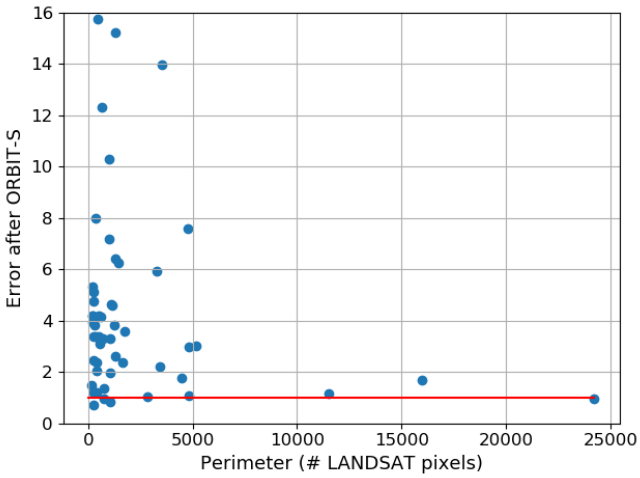


Figure 4.14: Relationship between lake perimeter and the performance of **ORBIT**.

Figure 4.14 shows the performance of the **ORBIT** framework under this scenario. The X-axis represent the perimeter of the water body in the reference maps. The Y-axis represents the *Error* after applying **ORBIT-T** (using $\alpha = 0.3$), followed by **ORBIT-S**. We can see that the few large lakes that are part of the dataset are showing good performance. On the other hand, the smaller lakes are showing more variation in the performance. This highlights the impact of size of the lake. Furthermore, the variance in the performance of the smaller lakes also suggests that for the same perimeter value, the performance can vary depending on both the shape of the lake and the amount of error in the input labels. In order to better understand the impact of errors in input labels, we compared the *Error* values using input labels and labels after correction. Figure 4.15 shows the comparison of *Error* values using input labels directly and labels after correction. The X-axis represents the *Error* when erroneous binary labels at LSR are used directly (replicated to match the high spatial resolution). The Y-axis represents the *Error* after **ORBIT** based correction. As we can see, the framework leads to much

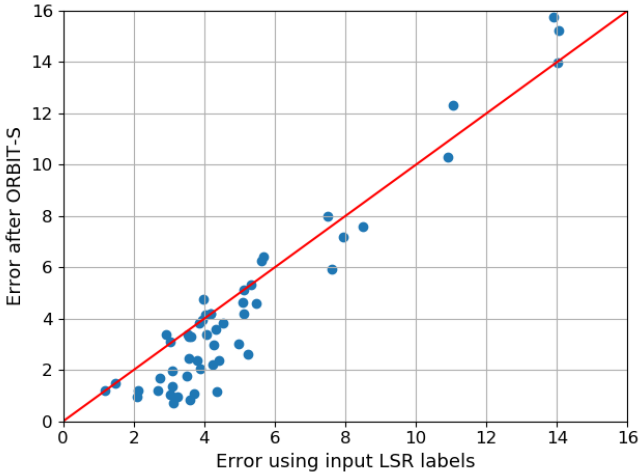


Figure 4.15: Comparison of ratio values from input labels and labels after **ORBIT** based correction.

better improvement when errors in the input labels are less. The amount of improvement decreases for lakes which have high error in the input labels. This matches with our understanding that as the amount of error increases in the input labels, the ability of the **ORBIT** framework to correct labels would decrease. Also, as showed in the previous experiment that estimation ordering would also introduce errors in the estimated HSR maps.

Figure 4.16 shows an illustrative example for Rio Grande Reservoir at the US-Mexico border on June 14, 2016. The **ORBIT** framework performed well in this case. The *Error* when reference map were used as input is 1.04 which demonstrates the good quality of the estimated ordering. The *Error* value is 1.17 when noisy LSR labels were used as input which demonstrates that **ORBIT** framework is able to get the performance very similar to the best possible scenario. Figure 4.16 (a) shows the relative ordering learned from JRC data using **ORBIT-E**. Dark blue color corresponds to the lowest elevation and red color corresponds to the highest elevation. Figure 4.16 (b) shows the reference map for June 14^{th}, 2016. In order to easily visualize the spatial patterns, we show only the bottom part of the lake in Figures 4.16 (c)-(g). Figure 4.16 (c) shows the noisy classification map at LSR. In this map, we can see that the northern part of the lake is being incorrectly labelled as land. Figure 4.16 (d) shows

the corrected classification map at LSR where the labels were corrected to capture the missing part of the lake. Also note the block pattern in the LSR maps due to the low spatial resolution. Figure 4.16 (e) show the estimated HSR classification map. As we can see, the estimated HSR map has much better spatial detail. In Figure 4.16 (f) and (g) we show the estimated LSR and HSR extents overlaid on Sentinel-2 true color composite image on June 14, 2016. The estimated HSR extent matches very well water boundary as seen from the Sentinel-2 data.

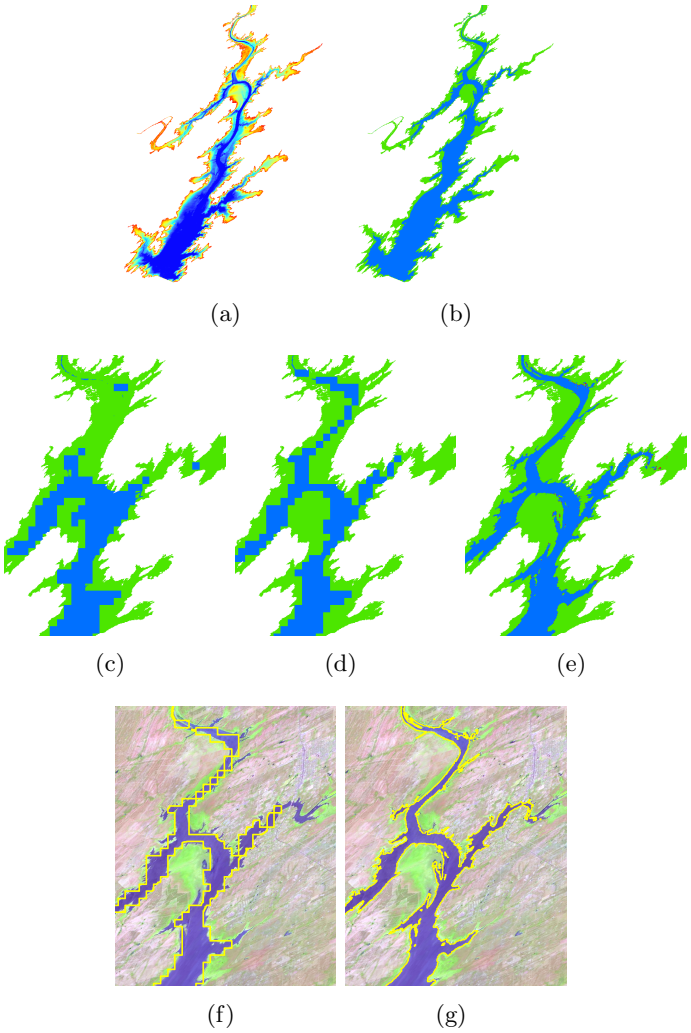
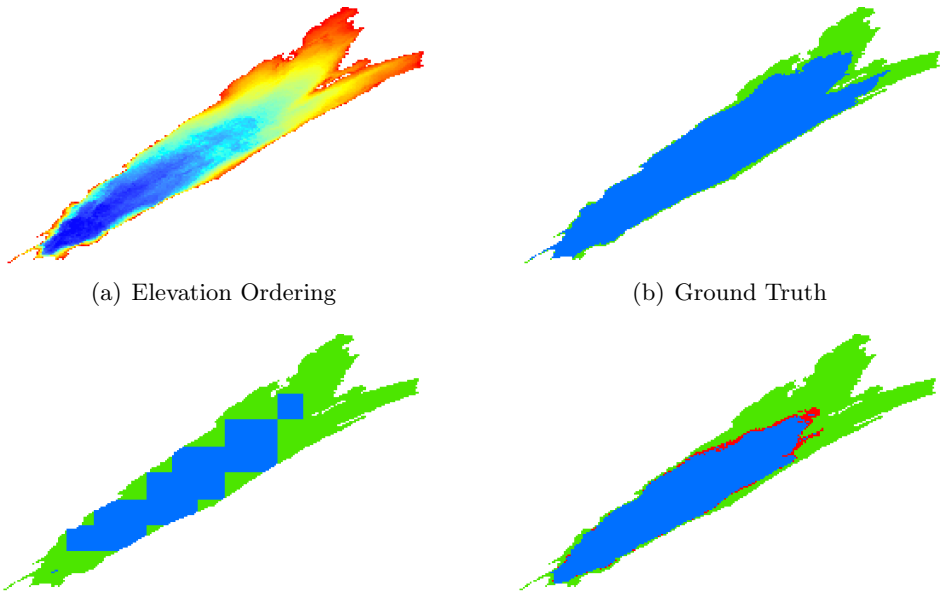


Figure 4.16: Illustrative example to demonstrate the performance of **ORBIT** framework for Rio Grande Reservoir at the US-Mexico border on June 14, 2016. (a) Relative Elevation Ordering (b) Ground Truth (c) Input LSR extent (d) Corrected LSR extent (e) Estimated HSR extent (f) Corrected LSR extent boundary overlaid on top of the high resolution Sentinel-2 image for June 14, 2016. (g) Estimated HSR extent overlaid on top of the high resolution Sentinel-2 image for June 14, 2016.

Figure 4.17 shows an illustrative example for Beulah lake, USA on July 1, 2016. In this case, the **ORBIT** framework was not able to create a high quality map. The *Error* when reference map were used as input is 0.39 which demonstrates good quality of the estimated ordering. However, the *Error* value is 4.21 when noisy LSR labels were used

as input. As we can see, the northern part of the lake is being underestimated as water. This example highlights two key limitations of the approach. First, with very few pixels available at the low resolution, ability of label the correction step is limited. In this case, the whole lake itself is covered by only 61 LSR pixels. Second, if the errors are also physically consistent, then it is difficult for the framework to correct those errors. In this case, some of the incorrect land pixels at low resolution were indeed the shallow pixels instead of deeper pixels and hence the framework was not able to correct them.



(a) Elevation Ordering (b) Ground Truth

(c) LSR Input Map (d) HSR Corrected Map

Figure 4.17: Illustrative example to demonstrate the performance of ORBIT framework for Lake Beulah, USA on July 1, 2016. (a) Relative Elevation Ordering (b) Ground Truth (c) Input LSR extent (d) Corrected LSR extent (e) Estimated HSR extent. The red color represents unknown labels.

Impact of quality of LSR labels

In all the above experiments, we considered J500 as the source of noisy LSR data, as it is currently the most recent global product that provides maps at 500m spatial resolution for every day. Through our analysis across different geographies, we can see that there is considerable amount of error in some of the lakes which is leading to poor performance

of **ORBIT** framework for some lakes. In this experiment, we compared results using another MODIS based product [9]. This product was created using MOD09GA data as well. Specifically, a SVM classification model was trained using data from February 2000 for 99 lakes belonging to different parts of the world. Using this trained model, we obtained daily scale classification maps at global scale for the year 2016. Figure 4.18 shows the comparison between *Error* values from these two different sources of low resolution labels. As we can see, with labels from Khandelwal et al. dataset [9], the performance improved significantly for some of the lakes while it also led to an increase in the error in some other lakes. This highlights the fact that there is still scope for improvement in the task of creating daily scale binary labels using MODIS data.

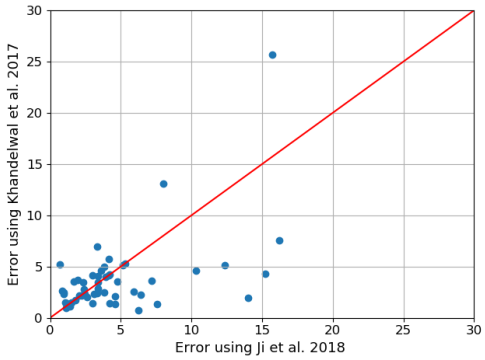


Figure 4.18: Impact of LSR label quality on the performance. Comparison of *Error* values using J500 and Khandelwal et al. daily scale datasets at MODIS scale

Impact of Unknown labels after ORBIT-S

So far, we have considered unknown labels as errors of the approach. In this experiment, we aim to analyze the contribution of unknown labels to the total error. Figure 4.19 shows the comparison of the total error and unknown labels when labels from J500 were used as noisy LSR input. The X-axis represents the *Error* and Y-axis represents the contribution of unknown labels in the total error. So, if a lake has most of the errors due to unknown labels, the point corresponding to that lake should lie on the top of the scatter plot. As we can see, for the majority of lakes, error in labelling is a bigger factor than unknown labels (points lie close to 0). This also suggests that using a single threshold of 128 is leading to more incorrect labels than unknown labels. In the next

experiment, we analyze the impact on the performance by relaxing this assumption.

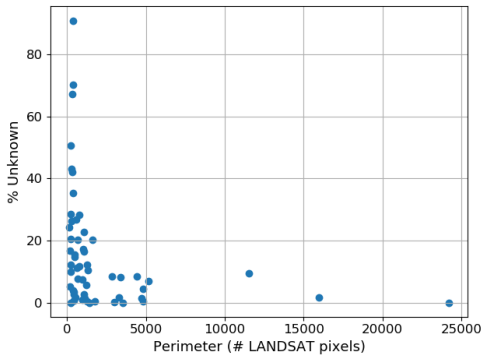
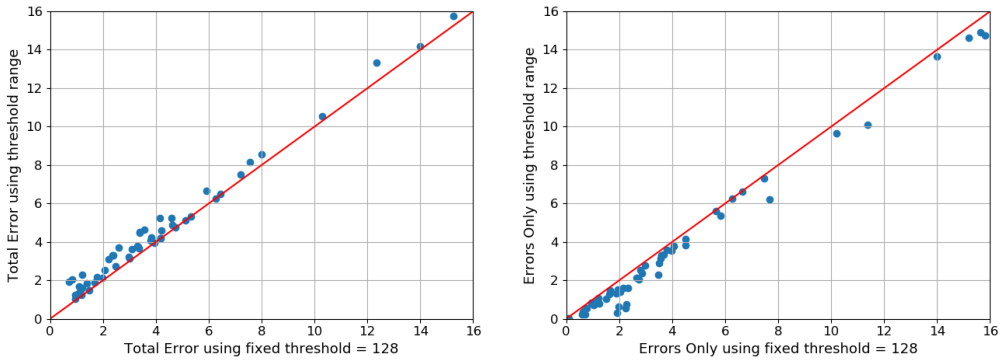


Figure 4.19: Contribution on unknown labels in total error. Fraction of errors due to unknown labels when noisy LSR maps from J500 are used.

Impact of Assumption A2

In **ORBIT-S** approach, we make an assumption that there exists a single threshold k for each pixel. In real world situations, different pixels might have different thresholds which could also be one of the sources of errors reported in Figure 4.14. In this experiment, we aim to analyze the impact of relaxing the assumption on the performance. Specifically, in Step 4 of the approach, if a LSR pixel is labelled water, we labelled only HSR pixels with local rank ≤ 102 (instead of 128) as water. Similarly, if a LSR pixel is labelled as land, we labelled only HSR pixels with local rank ≥ 153 (instead of 128) as land. In other words, we assume that k could be anywhere between 0.4 (102 for this dataset) to 0.6 (153 for this dataset) instead of exact 0.5 (128 for this dataset). By definition, this should lead to more unknowns as there we will be less information transfer. Figure 4.20 (a) compares the performance in this setting with the performance when a single threshold of 128 is used. As we can see, the error values have increased slightly for some lakes but Figure 4.20 (b) demonstrates that in this setting, the contribution of errors due mismatch labels (excluding unknown labels) have decreased. In other words, by using a threshold range, the approach was able to convert some incorrect labels into unknown labels. Thus, using a threshold range provides an effective way to achieve a trade-off between unknown labels and incorrect label estimation.



(a) (b)

Figure 4.20: Impact of relaxing the assumption A2 on the performance. (a) Comparison of *Error* values when a single threshold is used compared to *Error* values using threshold range. (b) Contribution of mismatch labels only (unknown are excluded) in the *Error* values when a threshold range is used.

4.6 Conclusions and Future Work

In this chapter, we presented a new approach for downscaling low spatial resolution land/water masks using elevation ordering constraint available at high spatial resolution. We showed that under certain scenarios, the approach can do the information transfer very effectively. The efficacy of the approach was demonstrated through both synthetic as well real-world datasets from the surface water monitoring application.

For future work, the approach can be extended from several aspects. In the current approach, we assumed a single cut-off threshold (k) for all pixels and timesteps. We aim to develop iterative methods to estimate k separately for each pixel using its multi-temporal class label information. Currently, the step for obtaining ordering at high spatial resolution does not incorporate any information from low spatial resolution multi-temporal maps. New multi-scale objective functions could be defined that could execute **ORBIT-E** and **ORBIT-S** in an iterative fastion. The complementary label information from low resolution label could potentially help in improving the ordering at high resolution. The key idea is that if the errors in two datasets are not correlated to each other then bringing information from the low resolution extent maps can improve ordering quality. Using high quality elevation structure information for a wide variety of

water bodies, we showed that the performance of the framework depends on the extent perimeter. In future, we aim to make the framework more effective for smaller lakes by making use of noisy fractional labels rather than binary labels. The current approach would play a key role in collected high resolution land/water labels that could be used to jump-start the training process of regression models to obtain fractional maps.

Chapter 5

GLADD-R: Global Lake Dynamics Database for Reservoirs created using the ORBIT Framework

5.1 Introduction

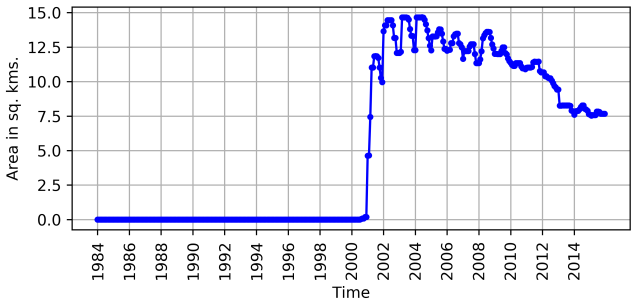
Reservoirs (dams and impoundments) have been constructed for millennia to provide persistent fresh water for agriculture and aquaculture, industry, human consumption, flood mitigation, reliable navigation, energy production, waste disposal, and recreation. The need for accessible and high-quality surface water has grown with the changing needs of civilization. Few human alterations of the Earth's water cycle rival the impacts of reservoir construction, including the unintended negative effects on water quality and contamination, habitability to native species, fish migration, and flooding disasters when infrastructure fails. A global database of reservoirs that provides their location and dynamics can be of great importance to the ecological community as it can enable the study of the impact of human actions and climate change on fresh water availability.

Currently, GRanD database [61] is the largest database that provides information on reservoirs globally. The first version of GRanD (v1.1) was released in 2011 which

provided the locations of 6862 reservoirs and a static snapshot of reservoir's attributes such as dam height, depth of the reservoir, average discharge , average surface area, and reference shape. A new version of the database (v1.3) was released in February 2019 and it provides information for an additional 458 reservoirs (7320 reservoirs in total). The database was created through manual curation effort which impact its completeness and makes it difficult to update over time. Moreover, the database does not provide temporal information about their surface level dynamics.

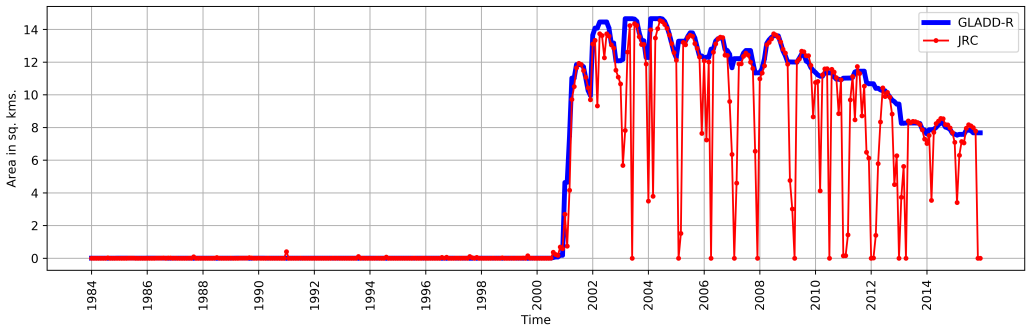
In the chapter, we present a new database, GLADD-R (Global Lake Dynamic Database for Reservoirs) that has been created by analyzing satellite imagery datasets using the **ORBIT** framework. Specifically, the current version, GLADD-R-1.0, was created using LANDSAT based land/water label provided by the JRC product [41] for the period 1984 to 2015 at monthly temporal scale. These labels were then corrected using the **ORBIT** framework. Furthermore, surface area time series of individual water bodies were analyzed to automatically identify candidate reservoirs that show a sudden increase in their surface area soon after their construction (a key time series characteristic of surface area time series of reservoirs). Finally, all the candidate reservoirs were manually verified using high resolution imagery to select the final set of reservoirs across the globe. The current version of the database provides information for reservoirs of size between 1 and 100 square kilometers. We are in the process of analyzing reservoirs with size between 0.1 and 1 sq. kms. and those larger than 100 sq. kms. These will be available in the next version of GLADD-R.

As an illustrative example, Figure 5.1 (a) shows surface area dynamics of a reservoir on the Sambito River in Brazil. The sudden increase in area is evident as the surface area of the reservoir increased from 0 to approximately 12 square kilometers in just three months. The surface area time series also makes it easy to identify seasonal changes and reduction in the size of the reservoir over time. Figure 5.1 (c) highlights the utility of the label correction step. The red line represents surface area time series created using JRC labels without correction. The area variations show a lot of spurious fluctuations due to missing data and labeling errors which were removed in the corrected time series.



(a) Lake x

(b) Lake y



(c) Lake y

Figure 5.1: An illustrative example of surface area dynamics of a reservoir on Sambito river in Brazil (latitude: -6.180322 , longitude: -41.978494). (a) Surface area time series using GLADD-R methodology. (b) High resolution aerial imagery of the reservoir. The zoomed-in inset shows the dam of the reservoir. (c) Comparison of GLADD-R surface area time series with surface area time series created using JRC labels.

5.2 GLADD-R: Highlights

GLADD-R-1.0 provides location and surface area dynamics of 1882 reservoirs built between January 1984 and December 2012 globally. Out of 1882 reservoirs reported in GLADD-R, only 415 were also reported in GRanD. Thus, GLADD-R provides information about an additional 1467 reservoirs that are not in GRanD. This highlights the utility of the automated machine learning approach to creating such a database on a global scale with minimum manual effort.

Figure 5.2 (a) shows the distribution of these reservoirs across different continents

while Figure 5.2 (b) shows the cumulative distribution of the number of reservoirs constructed after 1985 in different continents. The majority of dam construction has occurred in Asia and South America since 1986, and the rate of construction in North America has declined significantly. Figure 5.2 (c) shows the continent-wide distribution of the 1467 reservoirs that are unique to GLADD-R and Figure 5.2 (d) shows the corresponding cumulative distribution.

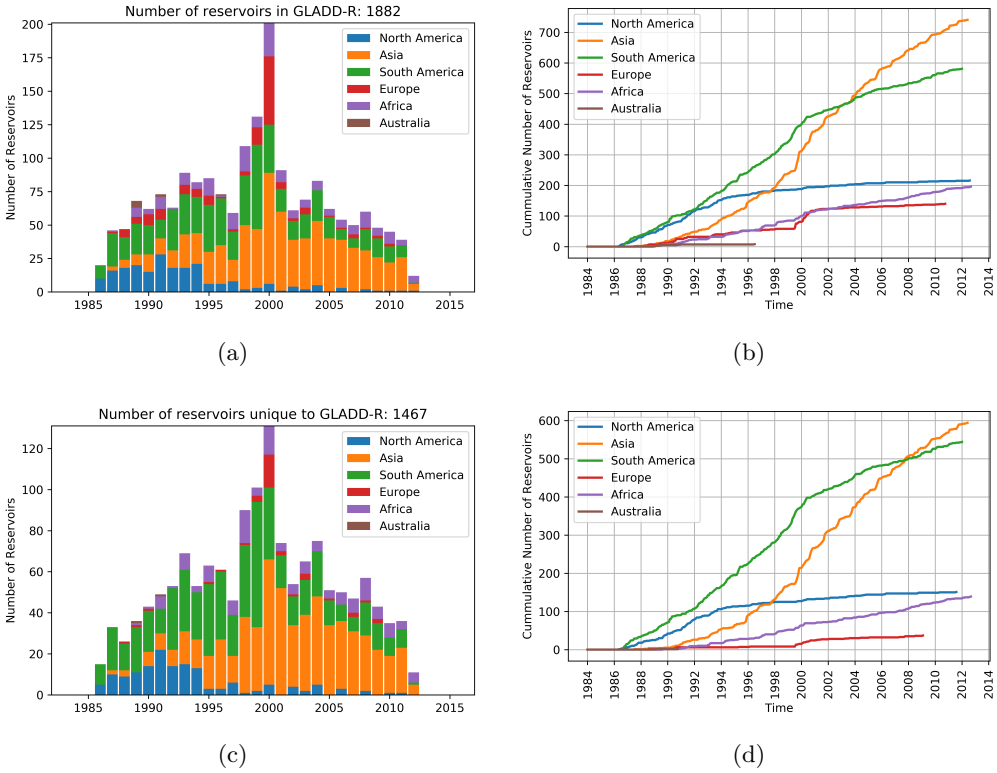


Figure 5.2: Distribution of reservoirs in GLADD-R. (a) Year-wise distribution of reservoirs in GLADD-R across different continents. (b) Time series of cumulative count of reservoirs in GLADD-R across different continents. (c) Year-wise distribution of reservoirs that are unique to GLADD-R across different continents. (d) Time series of the cumulative count of reservoirs that are unique to GLADD-R across different continents.

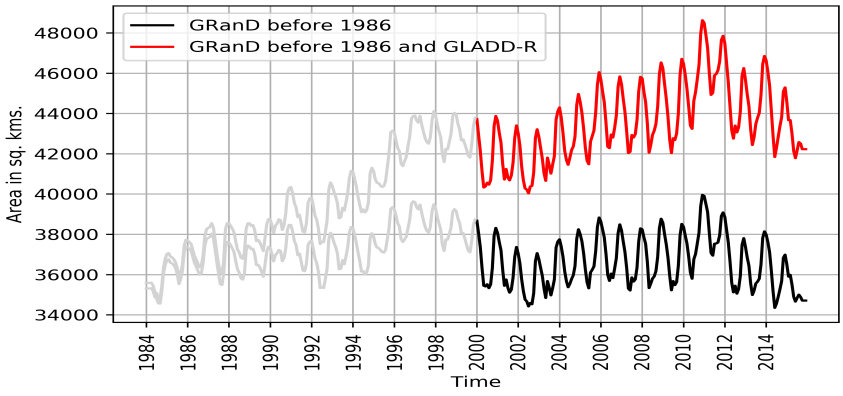


Figure 5.3: Aggregate surface area dynamics of reservoirs globally. The black line represents aggregate surface area of a subset of reservoirs (4142) reported in GRanD that were built before 1986 with size between 1 and 100 sq. kms. The red line represents the aggregate surface area of 4142 old reservoirs and additional 1882 reservoirs created after 1985 that are part of GLADD-R.

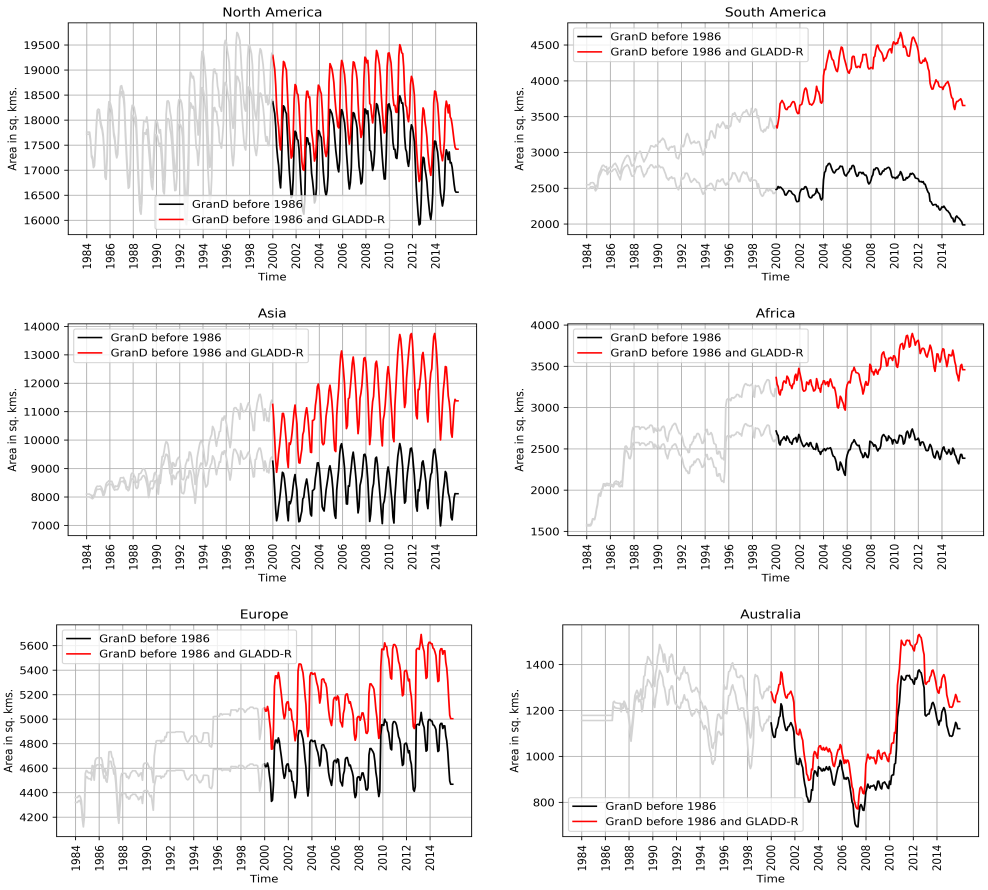


Figure 5.4: Aggregate surface area dynamics of reservoirs across different continents.The black line represents aggregate surface area of a subset of reservoirs reported in GRanD that were built before 1986 with size between 1 and 100 sq. kms. The red line represents the aggregate surface area of the old reservoirs and additional reservoirs created after 1985 that are part of GLADD-R.

While GRanD database only provides static information about the extent of reservoirs. In contrast, GLADD-R-1.0 also provides surface area at monthly scale from March 1984 to October 2015. Figure 5.3 shows the aggregate surface area variation of reservoirs globally. Due to the high prevalence of missing data in JRC label before 2000, the surface area at different time steps during this period can be much lower than the actual area. Hence, the dynamics before 1999 are shown in light grey color to signify less data availability. To provide a baseline of reservoir storage from reservoirs created prior to 1986, we processed a subset of 4142 reservoirs that were reported in GRanD and were created before 1986. At global scale, we can see that surface area in reservoirs continued to increase after 2000 as more reservoirs were constructed and approximately 8,000 sq. kms. of surface area has been added. Furthermore, there has been a reduction in surface area after 2012 until the end of the study period.

Figure 5.4 shows the aggregate surface area dynamics for each continent separately. Different continents show very different variations in surface area over the study period. Asia has the most number of dams and also the largest aggregate surface area. Even though South America has a greater number of dams, reservoirs in North America have more total surface area. All continents show strong seasonality in area, and all continents other than Europe show the decreasing trend from 2011-2015.

5.3 Processing Pipeline

GLADD-R was created using the **ORBIT** framework through the following processing pipeline

5.3.1 Pixel based land/water label generation

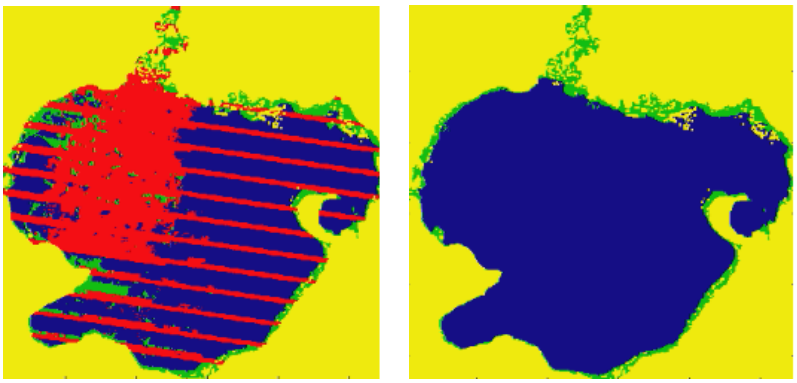
This step involves classification of satellite imagery data to produce land/water label at different timesteps. In the current version, we used the JRC dataset [41] that was created by analyzing the entire LANDSAT archive from March 1984 till October 2015. For each month a global land/water mask is available where pixels are labeled as either land, water or unknown.

5.3.2 Lake Polygons Database Generation

To identify locations and reference shape of lakes around the world, we performed connected component analysis on the JRC dataset's "occurrence" layer. The "occurrence" layer provides a number between 0 and 100 for each pixel, which represents the percentage of months the pixels was observed as water. We first binarized the layer by selecting pixels with percentage value greater than 10. Using the value 10 ensured that spuriously labelled pixels will be not selected. Once the binary layer is obtained, we performed a connected component analysis and assigned each connected component as a water body in our database. Using these reference shapes, we extracted pixel-based land/water label at monthly scale for each lake individually. To avoid including other nearby lakes in the buffer, we further prune the buffer region using an automated approach as described in [3].

5.3.3 Label Correction

If the pixel-based land/water label were accurate and complete, just counting the number of water pixels for each month would have provided area and its variation at the lake level. However, these maps tend to suffer from large amounts of missing data and labelling errors. Thus, these land/water label cannot be used directly to obtain robust surface area dynamics. To overcome this challenge, we used the **ORBIT** framework to improve the label accuracy. Specifically, we used **ORBIT-E** and **ORBIT-T** to improve the quality of land/water labels. This step is the most crucial step for achieving robust land/water labels. As an illustrative example, Figure 5.5 shows labels before and after correction for lake Naivasha in Kenya in February 2012.



(a) JRC based labels (b) Labels after correction

Figure 5.5: An illustrative example showing utility of the label correction step. Blue color represents water, green represents land, yellow represents pixels out of buffer region, and red represents missing labels

5.3.4 Identification of Candidate Reservoirs

Once the improved land/water labels are obtained for each lake, we count the number of water pixels for each month to create surface area time series for each water body. For each timestep in a time series, a score is computed which reflects the sudden and persistent increase in surface area values around that timestep. The maximum score across all timestep of a timeseries is used as an indicator of the reservoir construction activity. All the water bodies that had the score greater than a certain threshold are considered as candidate reservoirs. To ensure reliable estimation of sudden increase in the area of the water body, a minimum time-window of two years is used before and after the timestep under consideration. Due to this constraint, GLADD-R reports dam construction activity between January 1986 and December 2012 even though the surface area dynamics is available from March 1984 till October 2015.

5.3.5 Manual Verification of Candidate Reservoirs

All candidate reservoirs were manually verified by visual inspection using high resolution satellite imagery. Specifically, we looked for a dam wall or an impoundment wall to prune our candidate set of reservoirs. In some cases, especially reservoirs built for mining, agriculture, or just as a lake in a residential neighborhood, such a barrier is not

visible. But even in these cases, we were able to verify the sudden appearance of the reservoir using Google Timelapse.

5.4 Ecological and Hydrological Significance

There are thousands of reservoirs on Earth, and very little empirical data for making sound management decisions based on understanding the ecological dynamics of these systems. Just like large rivers, there are significant physical, biological, and economic consequences of changing water levels in reservoirs. Floods destroy infrastructure and deteriorate water quality, while droughts reduce habitat for important fisheries and diminish water supplies. The GLADD-R database provides the necessary data to build mechanistic, and predictive models for estimating reservoir water level. With associated reservoir depth information, gathered from GRanD, water level models can be leveraged to predict water volume and residence time. In this sense, GLADD-R is complementary to GRanD. Also, GLADD-R has identified new reservoirs, for which GRanD may be interested in collecting relevant attributes from respective countries and/or local authorities.

5.5 Data Availability

Location and time series information of these reservoirs is available at http://umnlcc.cs.umn.edu/GlobalReservoirDatabase. This online interface provides locations and shapes of all the reservoirs in GLADD-R. For each reservoir, its surface area time series can be visualized by clicking on the point on the map. The viewer also makes it easy to see the time lapse view of the reservoir, which allows instant visual verification of the year of reservoir construction. The viewer also provides the surface area time series of 4142 reservoirs reported in GRanD that were built before 1986 with size between 1 and 100 sq. kms.

Chapter 6

Conclusion and Future Directions

The thesis introduced a problem of improving accuracy and resolution of labels created by predictive learning approaches in the context of global surface water monitoring of lakes and reservoirs. Traditional classification approaches have limited performance in analyzing global satellite imagery due to noise, outliers, rich spatial and temporal heterogeneity in data. The thesis presented a new framework that incorporate robust physical principles governing water body dynamics to address the aforementioned challenges. The framework provided approaches to not only use the elevation constraint to improve performance but also provided a new approach to estimate it from the the data itself. Using both synthetic as well as real-world data from lakes belonging to different parts of the world, the thesis provided an evaluation of the framework and insights into the impact of different factors on the performance.

For future work, several advancements can be made to further extend the framework. Currently, all three different approaches (**ORBIT-E**, **ORBIT-T**, **ORBIT-S**) have separate optimization routines. New approaches could be developed to simultaneously optimized some or all these aspects to better leverage the complimentary information in the multi-scale land/water masks available from different sources. Currently, the framework works with binary labels but all the approaches could be modified easily to work with fractional labels. This would extend the applicability of the framework to smaller lakes as well. The use of fractional labels would be crucial for improving the ability to monitor rivers and flood events at global scale. In this regard, the current framework could be used to jump start the training process of the regression model by

creating high resolution binary labels for moderate to large lakes. Due to seasonal nature of water bodies, the framework does not need all the time steps to infer the elevation order. Thus, different variations of the elevation ordering could be learned from only a subset of time steps. This could potentially help applications such as detection of sediment deposition in reservoirs by highlighting variations in the elevation structure over time.

References

[1] Peter H Gleick. Global freshwater resources: soft-path solutions for the 21st century. *Science*, 302(5650):1524–1528, 2003.

[2] Petra Döll, Hervé Douville, Andreas Güntner, Hannes Müller Schmied, and Yoshihide Wada. Modelling freshwater resources at the global scale: Challenges and prospects. *Surveys in Geophysics*, 37(2):195–221, 2016.

[3] Alka Singh, Florian Seitz, Annette Eicker, and Andreas Güntner. Water budget analysis within the surrounding of prominent lakes and reservoirs from multi-sensor earth observation data and hydrological models: Case studies of the aral sea and lake mead. *Remote sensing*, 8(11):953, 2016.

[4] Ingjerd Haddeland, Jens Heinke, Hester Biemans, Stephanie Eisner, Martina Flörke, Naota Hanasaki, Markus Konzmann, Fulco Ludwig, Yoshimitsu Masaki, Jacob Schewe, et al. Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences*, 111(9):3251–3256, 2014.

[5] Douglas E Alsdorf, Ernesto Rodríguez, and Dennis P Lettenmaier. Measuring surface water from space. *Reviews of Geophysics*, 45(2), 2007.

[6] Anuj Karpatne, Ankush Khandelwal, Xi Chen, Varun Mithal, James Faghmous, and Vipin Kumar. Global monitoring of inland water dynamics: State-of-the-art, challenges, and opportunities. In *Computational Sustainability*, pages 121–147. Springer, 2016.

[7] Sanghoon Shin, Yadu Pokhrel, and Gonzalo Miguez-Macho. High-resolution modeling of reservoir release and storage dynamics at the continental scale. *Water Resources Research*, 55(1):787–810, 2019.

[8] Ankush Khandelwal, Varun Mithal, and Vipin Kumar. Post classification label refinement using implicit ordering constraint among data instances. In *Data Mining (ICDM), 2015 IEEE International Conference on*, pages 799–804. IEEE, 2015.

[9] Ankush Khandelwal, Anuj Karpatne, Miriam E Marlier, Jongyoun Kim, Dennis P Lettenmaier, and Vipin Kumar. An approach for global monitoring of surface water extent variations in reservoirs using modis data. *Remote Sensing of Environment*, 2017.

[10] Ankush Khandelwal, Anuj Karpatne, and Vipin Kumar. Orbit: Ordering based information transfer across space and time for global surface water monitoring. *arXiv preprint arXiv:1711.05799*, 2017.

[11] Fangdi Sun, Wanxiao Sun, Jin Chen, and Peng Gong. Comparison and improvement of methods for identifying waterbodies in remotely sensed imagery. *International Journal of Remote Sensing*, 33(21):6854–6875, 2012.

[12] Charles Verpoorter, Tiit Kutser, David A. Seekell, and Lars J. Tranvik. A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, 41(18):6396–6402, 2014. 2014GL060641.

[13] ZHU Xiao-ge. Remote sensing monitoring of coastline changes in pearl river estuary [j]. *Marine Env. Science*, 2:004, 2002.

[14] Dengsheng Lu and Qihao Weng. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5):823–870, 2007.

[15] Frank Paul, Christian Huggel, and Andreas Kääb. Combining satellite multispectral image data and a digital elevation model for mapping debris-covered glaciers. *Remote Sensing of Env.*, 89(4):510–518, 2004.

[16] Patrick Griffiths, Patrick Hostert, Oliver Gruebner, and Sebastian van der Linden. Mapping megacity growth with multi-sensor data. *Remote Sensing of Env.*, 114(2):426–439, 2010.

[17] Desheng Liu, Maggi Kelly, and Peng Gong. A spatial–temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. *Remote sensing of Env.*, 101(2):167–180, 2006.

[18] Varun Mithal, Ankush Khandelwal, Shyam Boriah, K Steinhauser, and Vipin Kumar. Change detection from temporal sequences of class labels: Application to land cover change mapping. In *SIAM International Conference on Data mining, SDM. SIAM*. SIAM, 2013.

[19] Offer Rozenstein and Arnon Karnieli. Comparison of methods for land-use classification incorporating remote sensing and gis inputs. *Applied Geography*, 31(2):533–544, 2011.

[20] Xuehong Chen, Jin Chen, Yusheng Shi, and Yasushi Yamaguchi. An automated approach for updating land cover maps based on integrated change detection and classification methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71:86–95, 2012.

[21] Bassel Solaiman, Raphaël K Koffi, M-C Mouchot, and Alain Hillion. An information fusion method for multispectral image classification postprocessing. *Geoscience and Remote Sensing, IEEE Transactions on*, 36(2):395–406, 1998.

[22] Desheng Liu, Kuan Song, John RG Townshend, and Peng Gong. Using local transition probability models in markov random fields for forest change detection. *Remote Sensing of Env.*, 112(5):2222–2231, 2008.

[23] James P Keener. The perron-frobenius theorem and the ranking of football teams. *SIAM review*, 35(1):80–93, 1993.

[24] David M Pennock and Eric Horvitz. Analysis of the axiomatic foundations of collaborative filtering. *Ann Arbor*, 1001:48109–2110, 1999.

[25] M Elena Renda and Umberto Straccia. Web metasearch: rank vs. score based rank aggregation methods. In *Proceedings of the 2003 ACM symposium on Applied computing*, pages 841–846. ACM, 2003.

[26] Ronald Fagin. Combining fuzzy information from multiple systems. In *Proceedings of the fifteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, pages 216–226. ACM, 1996.

[27] William W Cohen Robert E Schapire and Yoram Singer. Learning to order things. In *Advances in Neural Information Processing Systems 10: Proceedings of the 1997 Conference*, volume 10, page 451. MIT Press, 1998.

[28] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622. ACM, 2001.

[29] Nir Ailon. Aggregation of partial rankings, p-ratings and top-m lists. *Algorithmica*, 57(2):284–300, 2010.

[30] Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial hodge theory. *Mathematical Programming*, 127(1):203–244, 2011.

[31] Ernst Zermelo. Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 29(1):436–460, 1929.

[32] Herbert Aron David. *The method of paired comparisons*, volume 12. DTIC Document, 1963.

[33] Tyler Lu and Craig Boutilier. Learning mallows models with pairwise preferences. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 145–152, 2011.

[34] Massih-Reza Amini, Tuong-Vinh Truong, and Cyril Goutte. A boosting algorithm for learning bipartite ranking functions with partially labeled data. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2008.

[35] Jean C de Borda. Mémoire sur les élections au scrutin. 1781.

[36] H Peyton Young. Condorcet's theory of voting. *American Political Science Review*, 82(04):1231–1244, 1988.

[37] Land Processes Distributed Active Archive Center. `http://lpdaac.usgs.gov`.

[38] Charon M Birkett and B Beckley. Investigating the performance of the jason-2/ostm radar altimeter over lakes and reservoirs. *Marine Geodesy*, 33(S1):204–238, 2010.

[39] CM Birkett. The contribution of topex/poseidon to the global monitoring of climatically sensitive lakes. *Journal of Geophysical Research: Oceans*, 100(C12):25179–25204, 1995.

[40] National Elevation Dataset. http://www.pecad.fas.usda.gov/cropexplorer/global_reservoir, 2018.

[41] Jean-François Pekel, Andrew Cottam, Noel Gorelick, Alan S Belward, et al. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633):418–422, 2016.

[42] Sadiq I Khan, Yang Hong, Jiahu Wang, Koray K Yilmaz, Jonathan J Gourley, Robert F Adler, G Robert Brakenridge, Fritz Policelli, Shahid Habib, and Daniel Irwin. 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, 2011.

[43] Feng Gao, Jeff Masek, Matt Schwaller, and Forrest Hall. On the blending of the landsat and modis surface reflectance: Predicting daily landsat surface reflectance. *IEEE Transactions on Geoscience and Remote sensing*, 44(8):2207–2218, 2006.

[44] Xiaolin Zhu, Jin Chen, Feng Gao, Xuehong Chen, and Jeffrey G Masek. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 114(11):2610–2623, 2010.

[45] Thomas Hilker, Michael A Wulder, Nicholas C Coops, Nicole Seitz, Joanne C White, Feng Gao, Jeffrey G Masek, and Gordon Stenhouse. Generation of dense

time series synthetic landsat data through data blending with modis using a spatial and temporal adaptive reflectance fusion model. *Remote Sensing of Environment*, 113(9):1988–1999, 2009.

[46] V Heimhuber, MG Tulbure, and M Broich. Addressing spatio-temporal resolution constraints in landsat and modis-based mapping of large-scale floodplain inundation dynamics. *Remote sensing of environment*, 211:307–320, 2018.

[47] Julia Amorós-López, Luis Gómez-Chova, Luis Alonso, Luis Guanter, Raúl Zurita-Milla, José Moreno, and Gustavo Camps-Valls. Multitemporal fusion of landsat/tm and envisat/meris for crop monitoring. *International Journal of Applied Earth Observation and Geoinformation*, 23:132–141, 2013.

[48] Raul Zurita-Milla, Jan GPW Clevers, and Michael E Schaepman. Unmixing-based landsat tm and meris fr data fusion. *IEEE Geoscience and Remote Sensing Letters*, 5(3):453–457, 2008.

[49] Xiaolin Zhu, Eileen H Helmer, Feng Gao, Desheng Liu, Jin Chen, and Michael A Lefsky. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sensing of Environment*, 172:165–177, 2016.

[50] Guiping Wu and Yuanbo Liu. Downscaling surface water inundation from coarse data to fine-scale resolution: Methodology and accuracy assessment. *Remote Sensing*, 7(12):15989–16003, 2015.

[51] Wenbo Li, Xiuhua Zhang, Feng Ling, and Dongbo Zheng. Locally adaptive super-resolution waterline mapping with modis imagery. *Remote sensing letters*, 7(12):1121–1130, 2016.

[52] Linyi Li, Tingbao Xu, and Yun Chen. Improved urban flooding mapping from remote sensing images using generalized regression neural network-based super-resolution algorithm. *Remote Sensing*, 8(8):625, 2016.

[53] Linyi Li, Yun Chen, Tingbao Xu, Chang Huang, Rui Liu, and Kaifang Shi. Integration of bayesian regulation back-propagation neural network and particle swarm optimization for enhancing sub-pixel mapping of flood inundation in river basins. *Remote sensing letters*, 7(7):631–640, 2016.

[54] Milad Niroumand-Jadidi and Alfonso Vitti. Reconstruction of river boundaries at sub-pixel resolution: Estimation and spatial allocation of water fractions. *ISPRS International Journal of Geo-Information*, 6(12):383, 2017.

[55] Filipe Aires, Leo Miolane, Catherine Prigent, Binh Pham, Etienne Fluet-Chouinard, Bernhard Lehner, and Fabrice Papa. A global dynamic long-term inundation extent dataset at high spatial resolution derived through downscaling of satellite observations. *Journal of Hydrometeorology*, 18(5):1305–1325, 2017.

[56] Sanmei Li, Donglian Sun, Mitchell Goldberg, and Anthony Stefanidis. Derivation of 30-m-resolution water maps from terra/modis and srtm. *Remote Sensing of Environment*, 134:417–430, 2013.

[57] Alka Singh, Ujjwal Kumar, and Florian Seitz. Remote sensing of storage fluctuations of poorly gauged reservoirs and state space model (ssm)-based estimation. *Remote Sensing*, 7(12):17113–17134, 2015.

[58] Juan Diego Giraldo Osorio and Sandra Gabriela García galiano. Development of a sub-pixel analysis method applied to dynamic monitoring of floods. *International journal of remote sensing*, 33(7):2277–2295, 2012.

[59] National Elevation Dataset. http://ned.usgs.gov/, 2018.

[60] Luyan Ji, Peng Gong, Jie Wang, Jiancheng Shi, and Zhiliang Zhu. Construction of the 500-m resolution daily global surface water change database (2001–2016). *Water Resources Research*, 54(12):10–270, 2018.

[61] B Lehner, C Reidy Liermann, C Revenga, C Vorosmarty, B Fekete, P Crouzet, P Doll, M Endejan, K Frenken, J Magome, et al. Global reservoir and dam database, version 1 (grandv1): reservoirs, revision 01. *NASA Socioeconomic Data and Applications Center (SEDAC), Palisades*, 2011.