Advanced Processing of Multispectral Satellite Data for Detecting and Learning

Knowledge-based Features of Planetary Surface Anomalies

by

Srija Chakraborty

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved November 2019 by the
Graduate Supervisory Committee:

Antonia Papandreou-Suppappola, Co-Chair
Philip Christensen, Co-Chair
Christ Richmond
Alexander Maurer

ARIZONA STATE UNIVERSITY

December 2019

ABSTRACT

The marked increase in the inflow of remotely sensed data from satellites have transformed the Earth and Space Sciences to a data rich domain creating a rich repository for domain experts to analyze. These observations shed light on a diverse array of disciplines ranging from monitoring Earth system components to planetary exploration by highlighting the expected trend and patterns in the data. However, the complexity of these patterns from local to global scales, coupled with the volume of this ever-growing repository necessitates advanced techniques to sequentially process the datasets to determine the underlying trends. Such techniques essentially model the observations to learn characteristic parameters of data-generating processes and highlight anomalous planetary surface observations to help domain scientists for making informed decisions. The primary challenge in defining such models arises due to the spatio-temporal variability of these processes.

This dissertation introduces models of multispectral satellite observations that sequentially learn the expected trend from the data by extracting salient features of planetary surface observations. The main objectives are to learn the temporal variability for modeling dynamic processes and to build representations of features of interest that is learned over the lifespan of an instrument. The estimated model parameters are then exploited in detecting anomalies due to changes in land surface reflectance as well as novelties in planetary surface landforms. A model switching approach is proposed that allows the selection of the best matched representation given the observations that is designed to account for rate of time-variability in land surface. The estimated parameters are exploited to design a change detector, analyze the separability of change events, and form an expert-guided representation of planetary landforms for prioritizing the retrieval of scientifically relevant observations with both onboard and post-downlink applications.

To Columbia (STS-107)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Motivation

The marked increase in the inflow of remotely sensed data from satellites has transformed the Earth and Space Sciences with data rich repositories for domain experts to analyze. These include observations acquired by active and passive sensors on airborne as well as spaceborne platforms at varying spatial, temporal and spectral resolutions. These observations offer rich insight into various Earth and Planetary system processes and shed light on a diverse array of disciplines including Earth system monitoring, natural disaster detection, agriculture, geology, biodiversity, climate modeling, and planetary exploration by highlighting expected trends and patterns in the data. However, the complexity of these patterns on local to global scales, coupled with the volume of this ever-growing repository, necessitate the design of techniques to represent the data, sequentially analyze the observations to model the underlying process and update the representations to highlight anomalous trends. Monitoring the short-term patterns of such representations, informs domain scientists of changes and anomalous events such as natural hazards or a planetary landforms in an unexpected locations. Long-term variations highlight the response of natural systems to a changing climate and adaptively refine our knowledge of the of the underlying process.

The primary challenge in defining such representations arises from the spatiotemporal variability of these processes. Spatial variability is a natural characteristic of geospatial datasets and can be attributed to regional geographic variations. Temporal

variabilityon the other hand, arises from two primary sources. The first source is the dynamic nature of planetary processes and the time-variation of their underlying states. For example, land surface reflectance time-series of Earth's land cover can be categorized into multiple states such as a *stable or constant phase* with very minor deviations from seasonal trends, *change periods* with significant deviations from the expected trend; and *post-change regrowth phases* due to gradual recovery from the changes. Representations of time-series must therefore be adaptable over time. The second source of temporal variability arises due to the lack of sufficient data available at the initial stages to accurately represent the process or feature of interest. For example, planetary observations from the initial stages of an instrument's lifespan has limited surface coverage and any models based on it has to be verified and updated as additional data observations are acquired over time.

This work introduces adaptive representations of satellite observations to model the underlying process or feature of interest by sequentially adapting the representation over time. The emphasis is specifically laid on adapting to the temporal variability for modeling dynamic processes as well as on building a representation of feature of interest that is iteratively updated over the lifespan of an instrument. In addition, these representations also build on existing models to account for the spatial variability of geospatial datasets. The resulting models are aimed to provide domain scientists with adaptive analysis tools for assisting them with understanding planetary surface processes.

1.2   Background

1.2.1   Time-Varying Modeling Of Multispectral Satellite Image Time-Series

Earth observation data collected from space is vital for capturing dynamic changes in land, atmospheric, oceanic, and cryosphere on a global scale and is crucial for monitoring and understanding the ever-changing environment. These datasets form an important input data source for analyzing land surface image time-series. Land surface image time-series are obtained from repeated imaging of the Earth's surface and encompasses all categories such as forests, vegetation, agriculture, and natural and urbanized land cover on a global scale. Processing of these datasets further influences research in understanding seasonal, annual, and long-term trends as well as detecting changes in land cover due to both natural and anthropogenic causes. Changes in natural land cover due to human-induced factors such as deforestation, natural hazards such as wildfires and droughts, can trigger environmental perturbations with unknown consequences. For example, deforestation has been identified to be the second largest anthropogenic source of carbon dioxide release into the atmosphere [1]. In addition, deforestation also impacts local weather, modifies cloud cover and precipitation distribution between cleared land and unaltered forests [2], [3]. Changes in land cover due to wildfire depletes forests. This results in release of carbon dioxide [4], alters carbon sinks, radiation budget and increases land surface temperature in the burned area [5]. The released smoke particles during these events affects regional climate, suppress cloud and precipitation, and induce anomalies and droughts [6]. The altered soil in the burned area also affects regional hydrology and soil water retention capability and can lead to subsequent natural disasters such as floods and mudslides [7]. Conversion of natural land cover to urbanized land also alters the surface energy budget due to the effects of urban heat islands, influences land surface temperature, energy use, and

3

adversely affects human health.

The projected rise in events such as wildfires, growth in urban population are expected to contribute to increasing land cover changes. Accurately capturing land cover change is essential for modeling the global carbon budget, quantifying its impact on surface energy, and for incorporating this effect into climate models for accurate predictions [8]. Additionally, identifying land cover change events can play a crucial role in the detection of natural disasters, mitigation efforts, and in forecasting further change events, for example, detection of beetle infestation increases susceptibility to wildfires. Thus, there is a pressing need to analyze land cover time-series at a global scale for detecting changes and anomalies.

Uniform global spatial coverage of satellite observations make them ideally suited to study land cover change. Although the variabilities of observations in terms of spatial and temporal resolution represent complimentary information, moderate spatial resolution short revisit period satellite observations are most relevant for studying time-varying attributes of land cover. In order to maximally leverage this repository, it is essential to be able to extract meaningful patterns relating to trends and anomalies from these observations. However, such patterns are often complex, which renders the process of extracting it from a continually-growing archive non-trivial and consequently manually discovering such patterns is infeasible even by domain experts.

The aforementioned advantages of satellite observations necessitates for designing algorithms that sequentially analyze incoming satellite observations for monitoring and detecting land cover change events. Studies in this area have primarily emphasized on detecting change between two or more satellite images acquired on different dates. However, such approaches have several drawbacks such as not accurately identifying the change point and not capturing land cover seasonality.

Time-series approaches, on the other hand, do not suffer from these limitations

4

and provides the necessary information to quantitatively assess the impact of land cover change on the carbon cycle [9]. The high temporal coverage, makes it a vital dataset for monitoring agricultural productivity, responses of sustainable development practices and detecting natural hazards. Various approaches have been investigated that use time-series data for detecting land cover change and broadly fall into two categories. The first type, which is direct analysis of the time-series data include [10], [11], [12], [13]. The second category is a model based formulation of the land cover time-series that models its seasonality using harmonic analysis including [14], [15], [16], [17], [18], [19], [20], [21]. The model based approaches are based on estimation and monitoring of meaningful parameters related to land cover state and have been shown to be more effective at detecting change than direct time-series approaches. As these model extracted parameters are characteristic to a given land cover, it also allows classifying and mapping different land cover classes in a given region. In addition, the interpretability of parameters namely, the mean, variability about the mean, and time taken for the time-series value to reach its peak renders this approach more suitable for monitoring and analyzing land cover dynamics.

Some existing model based approaches, rely on estimating before and after-change parameters of a given land cover and determining an optimal threshold that maximizes separability of these parameters. The drawback of such approaches is that the model does not adaptively adjust to change and hence requires knowledge of future observations, making it ineffective for online change point detection that is essential for near-real time analysis. This is a severe drawback as Earth system components including land cover, exhibit state transition over time. For example, a stable segment of the time-series (with minor inter-annual variations) could be followed by a gradual or abrupt change that could be permanent or could subsequently be followed by a state of regrowth. Not only are the change points of these transitions unknown

5

beforehand, but also the evolution or future temporal trajectory of the land cover and its type of variability following the change event. To address this challenge, the land cover time-series algorithms have to not only *estimate the land cover parameters, but also adapt the model itself to the land cover state.*

Additionally, as the current approaches estimate parameters from stable segments of land cover time-series, an assumption of constant frequency or land cover growth rate is made, which fails to hold true under the sequential analysis formulation of land cover time-series with multiple stages. Frequency or vegetation growth rate remains nearly constant during stable segments, but varies during change and regrowth phases. The model used for such cases also needs to incorporate a *time-varying frequency parameter relating to land cover growth rate and sequentially estimate it over the different land cover states.*

Most model-based change detection algorithms from satellite image time-series employ supervised change detection, where the class behavior before and after change are represented using some characteristic pre-change and post-change parameters. These parameters are then used to train a machine learning classifier. Evaluating the detection ability of such approaches involves testing on unseen test locations of the same change event in the same region. Other approaches determine the threshold to identify the change point by simultaneous minimization of the false alarm, missed alarm and detection delay, which also requires observations from both pre-change and post-change phases to identify the change point. Thus, the machine learning classifier is aware of the pre and post change characteristics of the region. However, a more realistic framework should be capable of detecting any change type without the knowledge of change from future observations. Additionally, these approaches consider only one type of change, limiting the generalizability of such methods as land cover change events can belong to multiple classes. This limitation in current

approaches arises because the knowledge of the post change class is encoded using some representation as an input to the machine learning classifier. Consequently, the classifier utilizes the class specific information to detect change. Thus, it is essential to formulate a generalized change detection approach which can detect multiple types of change events without specific knowledge of the post change class and is independent of geographic variability. This not only presents a more realistic set up, but also allows the possibility to identify a previously unobserved change event.

The existing time-series approaches are based on vegetation index time-series. Although, this is a good indicator of change, it is limited in its ability to distinguish different types of changes beyond the distinction between gradual (for example beetle infestation, and drought) and abrupt changes(for example wildfire, floods), nearly permanent(deforestation, wildfires that severely alter the ecosystem and require decades of regrowth) and temporary(floods) changes. However, it cannot identify the cause of change and group similar change events, such as distinguishing between floods and fire events as it does not exploit the wealth of information associated with the spectral response during and post change.

The models thus formed to address the aforementioned limitations, should be utilized in detecting land cover change with minimum delay and monitor the time-series post change for learning subsequent trends and distinguish different land cover change events. This has the potential to offer improved insight into changes and anomalies in land cover compared to current approaches that primarily emphasize on change detection alone. Overall, these approaches should be i) adaptive the regional dynamics over time, ii) unsupervised, as gathering accurately labeled training data is expensive, iii) online, to ensure sequential analysis and decision making, and iv) globally scalable and robust to geographic variations.

Although these observations from short revisit period satellites are rich in informa-

7

tion for analyzing and monitoring planetary processes and change, the volume of the continually-growing repositories amounts to petabytes (`https://earthdata.nasa.gov/modis-in-gibs`). Due to the huge deluge of data from these sources, it is infeasible for a human expert to sift through all observations to find the most interesting or relevant subset, which is only a fraction of the total data acquired. This relevant subset comprises of novel instances that are indicative of change or anomalies and are more significant for notifications or for domain experts to analyze. However, a setting where expert inspection of this vast dataset is necessary to identify the scientifically interesting observations is not only inefficient but also infeasible. Consequently, this further necessitates the introduction of approaches that can intelligently identify this interesting subset for human experts to analyze further.

1.2.2 *Time-Varying Representations Of Geological Features Of Interest In Planetary Images*

The large volume of data collected by spacecraft in deep space missions has created a repository which is infeasible to be analyze manually by an expert. This has resulted in a data deluge where *data is downlinked to Earth at a much faster rate than it can be carefully analyzed by human expert.* As expert analysis is related to the relevance or importance of an observation, the objective of this study is to determine the significance of the incoming observations and highlight the most interesting subset (which is a small fraction of the entire data volume and hence made feasible for manual analysis) to human experts for further analysis. The definition of relevance is formulated from expert guidance and is based on *how relevant the observation is to answer a science question of interest* or *how novel the observation is with respect to the expected trend*, which can lead to knowledge discovery. The objective of this work is to utilize an expert-guided definition of relevance to retrieve the subset of instances

for an expert query and to detect and highlight the subset of novel instances for knowledge discovery [22]. These representations are learned with respect to geological classes of interest and are *adaptive to newly discovered information to operate over years throughout the operational phase of an instrument.*

As expert-interest in observations is often determined by the presence of geological landforms of significance, in this work the definition of relevance is centered around the presence or absence of such landforms. In addition, the spatial location, season of prevalence, association with other landforms and spectral properties are considered as well in order to determine the novelty and significance of the observation with respect to the trend learned from the past data, adding to the context of the observations. Studies in this area have focused on the detection of dune fields [23] and other geological features [24], [25], that are solely based on the detection of detection of geological features and does not incorporate the image context. Additionally, highlighting novel instances is also significant for informing planning operations and for targeting future observations, as such operations are also performed by human experts in a short span of time. In addition, such approaches are also useful to support onboard data analysis operations.

Efforts in onboard data analysis have focused on detecting application specific classes of interest such as Martian polar ice caps [26], dust devil tracks and clouds [27] to determine scientific value. Prioritization has been studied with planetary rover data to select rocks with a specific target signature [28], for autonomous exploration by integrating past orbital datasets with a science hypothesis [29]. In this work, the context of the image is utilized such that it incorporates expert preferences to ensure that the prioritized set or the retrieved observations are acceptable to experts.

There are several challenges in identifying scientifically relevant observations by context. A significant challenge is defining what an *interesting* observation is as it

is subjective by nature and varies based on the scientific question at hand and from expert to expert. The criterion for an interesting observation may include observations that are relevant to answer specific scientific questions or to observations that are indicative of novelty based on existing knowledge, leaving scope for exploration. *Thus there needs to be a explicit definition of relevance and novelty.*

Secondly, Incorporating expert knowledge and expert preferences to ensure that the relevant subset and its ranking is acceptable to experts. While identifying relevant observations to be retrieved, *inclusion of expert knowledge for guidance is necessary to ensure that the relevant subset and its ranking is acceptable to experts.* This is particularly important for including *expert preferences*, favoring one class over the other based on interest or when inferring class specific properties is not possible directly from the dataset. In such cases, expert guidance allows scope to direct the automated prioritization approach such that the identified relevant subset matches expert interest.

Thirdly, Expert interests are often varying in nature which results in multiple definitions of relevance or novelty and diverse features of interest. It is expected that there will be multiple questions of interest whose relevant observation set will be mutually exclusive. It is important for the prioritization approach to be such that the *selected subset includes relevant observations for all investigation and caters to a diverse range of questions without introducing bias towards a question or group of questions.*

Fourthly, observations are to be identified as relevant due to its context that might be be useful for answering a scientific question of interest or provides a novel insight. Expert identification of an interesting observation is based on integrated interpretation of its spatial context and location, spectral, seasonal characteristics and topography. In order to achieve expert like understanding of an observation, the

prioritization methodology should adopt a combined analysis of all such factors.

Finally, while identifying observations that are suitable for answering questions of scientific interest, it is also crucial for the prioritization approach to *detect any previously unseen context of a known geological class.* This exploration favoring approach ensures that planetary knowledge bases are improved with time. Thus it is important for the remote module to independently identify novel instances.

1.3   Contributions

1.3.1   *Time-Varying Modeling Of Multispectral Satellite Image Time-Series*

This work proposes to address the challenges described before by designing an time-varying model of land cover time-series. The model aims to sequentially determines the land cover state for a given observation, select the most appropriate model governing the current state and estimates the model parameters relevant at various stages of land surface state. Unsupervised change detection and post-change analysis is made possible by analyzing the dissimilarity in the estimated parameters. The estimation step is also extended to all land cover spectral bands to identify the most responsive bands for a given change event. This is used as an interpretable change signature. Overall as these approaches fundamentally identify the most anomalous instances generated from an aberration in the underlying data generating process, they are ideally suited at mining the vast, continually growing repositories to determine the most unusual observation that warrants further investigation by human experts. Figure 1.1 provides an overview of the proposed approach. Here the sequentially available observations are analyzed by the proposed model based change detector. When a change point is identified, further insight is obtained by analyzing the change signature to categorize the change event and estimate post-change parameters. This
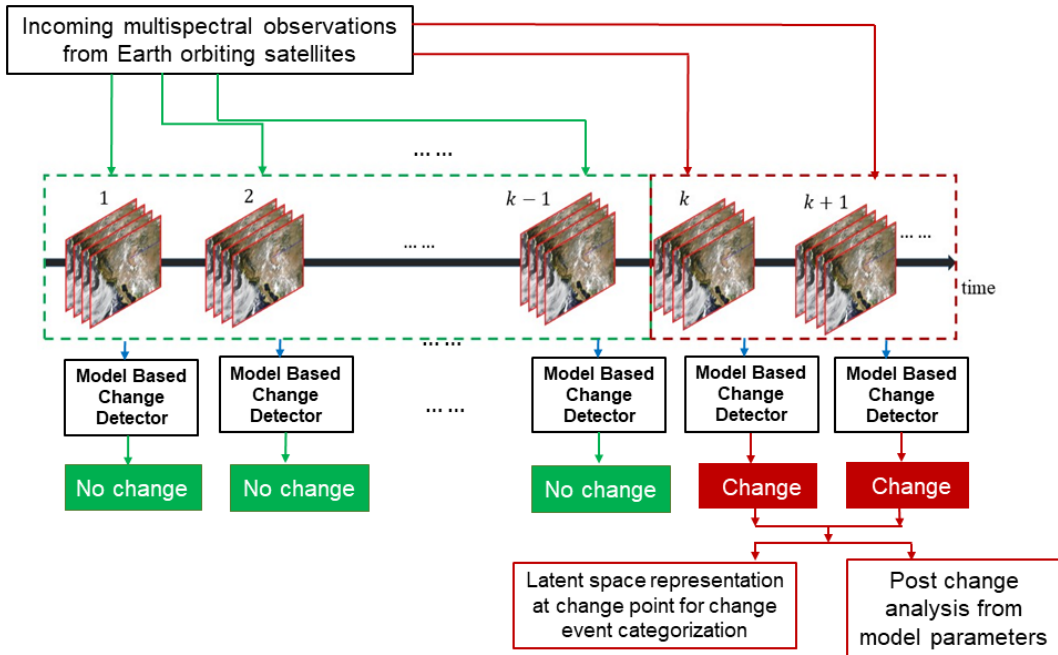
Figure 1.1: Overview of the Proposed Modeling, Change Detection and Monitoring Approach

additional insight helps domain experts for further analysis. Also, by detecting and highlighting the observations with change, a smaller fraction of the dataset requires expert analysis and is a feasible task. Ultimately it identifies the most relevant instances for an expert to analyze. Further analysis by domain experts is made feasible due to the small size of this subset of relevant instances.

We have made the following contributions in this area:

1. *Time-Varying representations of land cover time-series using multiple models:* Based on the observation that land cover state and hence its underlying model varies over time, an time-varying formulation of land cover image time-series is proposed. Here, not only are the parameters time-varying in nature, but also underlying model that best describes the given state. In the proposed work, the model that describes the state variation most accurately is determined sequentially with each observation, which is followed by model parameter estimation.

12

In addition, by adopting this approach the propose work is not only concerned with change detection alone, but with determining post-change trends.

2. *Sequential estimation of land cover variation rate and time-varying frequency parameter:* A novel time-varying land cover frequency parameter is introduced which is also sequentially estimated with every observation. Land cover time-series variation rate, or its frequency is nearly constant during stable time-series segments and is observed to vary due to change, disturbances or regrowth. In addition to the other model parameters, the introduced time-varying frequency is also physically interpretable for analyzing land cover trend. This parameter is estimated using appropriate models to represent the instantaneous frequency of land cover time-series and can be estimated through model selection.

3. *Analysis of estimated parameters in multispectral space:* Estimated parameters for a land cover state are extended to all land cover spectral bands. The expected behavior of the parameters across all bands is learned over time and is contrasted with new observations. At change points, high dissimilarity is observed and by analyzing the deviation pattern across all spectral bands, the change signature and consequently the type of change is clustered. Additionally, this is also a step towards explainable change detection algorithms, that not only detect change, but also explains the features (or spectral bands) with respect to which the given observation is anomalous to a domain expert.

4. *Unsupervised change detection and change event separability:* The sequentially estimated parameters are utilized in unsupervised change detection with a threshold learned from past observation variability. The multispectral change signature is also learned in a self-supervised manner and is therefore not limited by the knowledge and existence of accurately labeled training data. The

13

approaches developed are also geographically scalable as the initial parameters are estimated from harmonic analysis applicable any land cover class.

1.3.2   Time-Varying Representations Of Geological Features Of Interest In Planetary Images

This work proposed a context based analysis of multispectral planetary images for retrieval and prioritization of anomalous and highly novel observations. We introduce a methodology that also considers contextual attributes, such as relevance of a class or landform to experts, co-occurrence of landforms and its spatial and temporal properties to determine its relevance. This resembles expert like interpretation more closely instead of determining relevance from the presence or absence of interesting features alone. We then determine the relative importance of the instances based on these attributes, where it is crucial for the module to *understand* the content in the image like an expert. Figure 1.2 shows the overview of the proposed adaptive representation learning and prioritization approach that incorporates expert feedback. We have made the following contributions in this area:

1. *Rule based representation of novelty:* We introduce a rule based representation of interesting landmarks on a planetary surface with respect to expert selected attributes. This rules are learned from available past data and represents the expected properties of each class of interest with respect to the attributes.

2. *Expert guided feedback:* As the rules are learned from the data, it may not represent domain specific knowledge of relevance or novelty. We propose an approach to obtain this information from domain experts and incorporate it in the rules to emulate expert-like decision making.

3. *Rule update over time:* The proposed methods follow an iterative approach
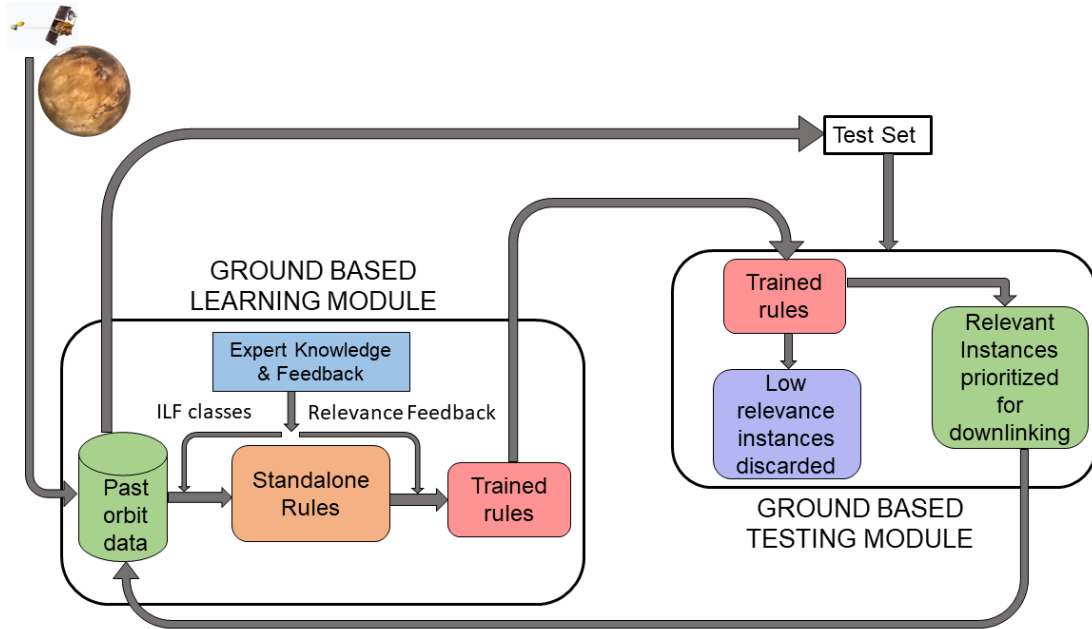
Figure 1.2: Overview of the Proposed Prioritization and Retrieval Approach by Learning Adaptive Representations of Geological Landforms of Interest (ILF) for Detecting Scientifically Relevant Observations.

where the rules are updated over time with the most recently available data to represent the current knowledge with respect to each class of interest. This ensures that the rules are made aware of any detected observation that is anomalous and does not incorrectly detect similar observations in the future as anomalous. Moreover, this update will be used in subsequent iterations and presents an approach to create a planetary knowledge base using the adaptive representations throughout the lifespan of an instrument.

In addition, the proposed approach is developed such that it can operate independently using a feature recognition technique and identify novelty in planetary observations. The feature recognition approach along with the rule based retrieval approach is essential for the proposed work to operate independently over long time periods with minimum expert involvement. Overall, this not only assists in retrieval and prioritization tasks, but also highlights the most interesting or anomalous observations

to experts and provides them with a reduced data volume to analyze further.

1.4   Organization

The organization of this proposal is summarized as follows. Chapter 2 summarizes the fundamental principles that are utilized to develop the proposed approaches. Chapter 3 provides a description of the time-series data and the preprocessing steps. Chapter 4 proposes the modeling of time-series and presents the Bayesian parameter estimation using the proposed time-varying frequency land cover model and change detection results using these models. In Chapter 5 we propose a model selection approach that sequentially selects the best matched model of land cover for change and post change stages. Chapter 6 proposes an analysis on spectral separability of change events and Chapter 7 presents the proposed technique for learning adaptive representations of planetary images for prioritization and retrieval of novel images. We conclude in Chapter 8 and discuss possible future extensions.

Chapter 2

SIGNAL PROCESSING TECHNIQUES

2.1 Approaches

The fundamental principles and approaches that are utilized in the subsequent chapters are introduced here. These include sequential state estimation techniques, efficient representation learning or encoding techniques, image description generators for novelty detection.

2.1.1 Sequential Bayesian Parameter Estimation

Dynamic state space models represent the time-evolution of parameter states that are not known and need to be estimated at each time step. In order to define and make inference about such systems at a given time step k, two models are essential. This consists of a transition model (equation) and an observation or measurement model.

1. Process model: The transition equation provides the relationship of the unknown state $\mathbf{x}_k$ at time step k with the state at the previous time step $k-1$. This defines the evolution of the state vector $\mathbf{x}_k$ from time step $k-1$ to k as:

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \mathbf{v}_k), \tag{2.1}$$

where f_k is the state transition function and $\mathbf{v}_k$ is a random process noise that represents the error in this equation.

2. Observation Model: This model relates the noisy measurement or observation

$\mathbf{z}_k$ at time step k with the state vector as

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \boldsymbol{\eta}_k), \tag{2.2}$$

where the function h_k is the observation function and $\boldsymbol{\eta}_k$ is the observation noise.

In order to estimate $\mathbf{x}_k$, we need to estimate the posterior distribution $p(\mathbf{x}_k|\mathbf{z}_k)$, which is the distribution of $\mathbf{x}_k$ given the measurement $\mathbf{z}_k$. The process and observation models describing a dynamic system can be used for determining the state vector sequentially with every available observation using a Bayesian approach [30].

Using Bayes theorem we can obtain the posterior using

$$p(\mathbf{x}_k|\mathbf{z}_k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{k-1})}{p(\mathbf{z}_k|\mathbf{z}_{k-1})}, \tag{2.3}$$

where the normalizing constant is given by

$$p(\mathbf{z}_k|\mathbf{z}_{k-1}) = \int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{k-1})d\mathbf{x}_k. \tag{2.4}$$

Here, the prior distribution of the state vector of state k is obtained using the following equation applied to the process model

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{z}_{k-1}) &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{k-1})d\mathbf{x}_{k-1}, \\ &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{k-1})d\mathbf{x}_{k-1}, \end{aligned}$$

where $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_{k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$ as the process model describes a Markov Chain of order one.

If both functions $f_k(\mathbf{x}_{k-1}, \boldsymbol{v}_{k-1})$ and $h_k(\mathbf{x}_k, \boldsymbol{\eta}_k)$ are known and are linear, and $\boldsymbol{v}_k$ and $\boldsymbol{\eta}_k$ are both Gaussian random processes, the state vector can be optimally estimated using the Kalman filter. However, when of these two models are nonlinear, the nonlinearities in the system, require suboptimal approaches for state estimation. Extended Kalman filter (EKF) tackles such nonlinearities by local linearization of the

two models f_k and h_k (along with the assumption that the process and observation models are Gaussian). The unscented Kalman filter was proposed to reduce the suboptimal performance and divergence of the EKF [31].

Without any restrictions in the model grid based filters produce optimal solutions for discrete systems with finite number of state spaces [30]. Using the grid based approach, the posterior probability at each time step is estimated by considering the weighted conditional probability of all states and are based on deterministic numeric integration techniques that consider all possible configurations of the system to obtain accurate estimates. However, the drawbacks of these methods include the difficulties in implementation as well as the computational complexity for high dimensional problems. For highly nonlinear models the EKF does not provide accurate results. In such cases a suboptimal solution can be obtained using sequential Monte Carlo methods such as the particle filter.

2.1.2   Particle Filter

The Particle filter (PF) represents the posterior distribution of a time-varying system using a large number of samples or particles and its associated weights to estimate the state vector. By considering sufficiently large number of particles, the posterior distribution at every time step is approximated. This does away with the need to linearize any nonlinearities in the system and also allows computation with process and observation models and noise that are non-Gaussian.

At time step k of the PF algorithm, the posterior distribution is approximated as

$$p(\mathbf{x}_k|\mathbf{z}_k) \approx \sum_{i=1}^{N_s} w_k^i \delta(\mathbf{x}_k - \mathbf{x}_k^i), \tag{2.5}$$

where $\mathbf{x}_k^i$ and w_k^i correspond to the ith particle or sample and its corresponding weight, N_s is the total number of particles under consideration for approximating the

posterior distribution and $\delta(.)$ is the Dirac delta function.

The particle filter consists of the predict step and the update step that uses a process and observation model to sequentially estimate the state vector at every step using equation 2.5. The initial conditions of the state vector $p(\mathbf{x}_0|\mathbf{z}_0) \equiv p(\mathbf{x}_0)$ are assumed to be known. The predict step causes the state vector to evolve over time using the process model by relating the state vector estimated at $k-1$ to that at k to form a prior distribution represented by all the N_s particles. As the measurement at time step k becomes available, the particle weights or likelihoods are updated to represent the most probable particle $\mathbf{x}_k^i$ that relates to the observation $\mathbf{z}_k$. The posterior distribution at k is approximated using equation 2.5. This estimate is then used for the predict step at $k+1$ by using the process model. By repeating the predict and update steps with every observation the state vector is approximated sequentially.

Degeneracy is an undesirable condition encountered with particle filters where as the time step increases, all but one particle will be associated with weights very close to zero. It has been shown that the variance of the particle weights will only increase over time further worsening the degeneracy [32]. In order to avoid degeneracy, a resampling technique called the sequential importance resampling (SIR) is used that eliminates particles with very small weights and emphasizes on the particles with large weights. This is performed by resampling with replacement and generating a new set of particles are drawn with replacement N_s times from the approximate distribution of $p(\mathbf{x}_k|\mathbf{z}_k)$. The Sequential Importance Resampling particle filter used in this work performs the resampling operation at every iteration to avoid degeneracy. This resampling operation is performed after the update step at every iteration, which is then used for the predict step in the consecutive step.

2.2 Seasonality Analysis Of Land Cover: A Time-Frequency Approach

Land cover time-series exhibits periodicity at annual and sometimes at intra-annual scales as the reflectance over time is representative of seasonality. This seasonal characteristic is intrinsic to a given land cover and hence has been extensively used in classification of different land cover classes. These seasonal parameters are extracted using harmonic analysis and relate to the vegetation or land cover state in a given annual cycle. Land surface reflectance seasonality and the rate of signal variation rate as highlighted by time-frequency analysis is a key motivation for developing the adaptive models. In addition, the variation rate shown by time-frequency approaches also highlights the nature of this variation, which will be utilized in designing and selecting the image time-series models for the various land cover stages (stable, change, regrowth).

The discrete Fourier Transform (DFT) has been utilized in several past studies to extract temporal land surface features by decomposing the time-series into its constituent periodic signals. The frequency components relate to different dynamic processes in land cover. The FT of a given time-series $y[n]$ can be represented as

$$X_f = \frac{1}{N} \sum_{n=0}^{N-1} y[n] e^{-2\pi j f n/N}, \qquad (2.6)$$

where f represents the frequency in the DFT component X_f, n is the time index and N is the number of samples in the time-series [33]. Equation (2.7) contains real and imaginary parts which can be represented as cosine and sine waves using Euler's notation. The frequency f represents the number of cycles completed by the given DFT component, for example, the 7th component completes 7 cycles in 7 years and hence can be related to the annual component in a 7 year time-series. The time-series can be represented in the polar form using cosine terms with amplitude and phase shifts [33]. Past studies have used these components (mean, amplitude and

phase; amplitude and phase corresponding to the annual component) as characteristics land cover features relating to the average land cover state, it amplitude of variation about this mean and the time taken for the time-series to reach it peak respectively for classification as these represent intrinsic temporal properties of the land cover [33], [34], [35], [14]. Time-varying estimation of these parameters were also introduced that expressed these features using time-varying state vector which was shown to improve land cover separability over DFT parameters which does not capture this time-varying attribute of the parameters [14].

These parameters are extracted under the assumption that the land cover variation rate or frequency remains constant and is comprised of repeating periodic cycles extracted by the DFT. However, this assumption fails to capture intra-annual variability due to disturbances or events such as forest fire. For example, the change in growth rate due to a forest fire in a given year followed by subsequent years of regrowth is not detected by DFT analysis [33]. Such a variation in land cover change rate is shown in Figure 4.4 due to repeated drought in the region. Transitional periods where the land cover growth rate varies is expected in Earth system components and this calls for adaptive estimation of the time-varying growth rate as this can further shed light on land cover dynamics.

2.2.1 Time-Frequency Representations For Land Cover Variation Rate Analysis

Most real data are non-stationary, in the sense that their frequency content changes with time. Using the DFT as an analysis tool for non-stationary signals implies that there is only a single frequency at each time step. In addition, it is not possible to know which frequency occurred at each time. Time-frequency representations, on the other hand, are functions of both time and frequency and allow us to analyze the signal jointly with respect to time and frequency.

22

The varying change rate of land surface reflectance time-series causes the it to be non-stationary. To study this non-stationary frequency, time-frequency approaches are explored, that determines the frequency variation of a given signal with time. Fundamentally, these approaches are designed with the objective of extracting the instantaneous frequency of the time-series, which is the derivative of the time-varying phase. However, this is restrictive to a time-series with a single frequency component existing at a given time interval. To further overcome this limitation, time-frequency analysis is carried out over a 2-D surface in the time-frequency plane yielding different time-frequency representations (TFR), that temporally localize the spectra of the time-series. The TFR considered for analyzing the varying growth of land cover time-series are introduced.

2.2.2 Short-Time Fourier Transform And The Spectrogram

The short-time Fourier transform is a simple TFR as it uses the FT of windowed segments of the data. It is assumed that windowed data is stationary during the length of the window and provides time localized information of frequency variation in the data. To analyze time-variation of frequency the given time-series $y[n]$ is emphasized at the time-duration of interest by multiplication with a window function over the time-series segment and suppressed at other times. This can be expressed as

$$\text{SFT}(t, f) = \sum_{n=0}^{N-1} y[n]h[n - t]e^{-2\pi jfn/N} \tag{2.7}$$

where h represents the window and $y[n]$ represents a segment of the data whose frequency information is being localized. Here $\text{SFT}(t, f)$ represents the short-time Fourier transform (STFT) time-frequency representation of the time-series [36]. The STFT of a simulated linear chirp is shown in Figure 2.1 which shows the linear increase in frequency over time. However, the time-frequency localization of the signal is poor.

23

The energy density spectrum or the spectrogram is then calculated as

$$\text{SP}(t, f) = |\text{SFT}(t, f)|^2, \qquad (2.8)$$

where $\text{SP}(t, f)$ represents the time-varying spectrum of the of the time-series and the distribution of this spectrum over time is referred to as the spectrogram time-frequency (TF) representation of the time-series [36]. The spectrogram representation of a simulated linear chirp is shown in Figure 2.2 which shows the linear increase in frequency over time. However, the time-frequency localization of the signal is poor and the representation does not localize the signal behavior well.

2.2.3 Reassignment Method As Applied To The Spectrogram

However, time-frequency representations such as the STFT and spectrogram are limited by the uncertainty principle which does not permit high frequency and time resolution of the time-series simultaneously. Thus jointly the time-duration σ_t and bandwidth σ_f are bound by $\sigma_t \sigma_f \geq \frac{1}{2}$. Therefore, although short-time windows are taken that emphasizes on the time-series energy centered around the window, due to the uncertainty principle, this does not allow high localization in frequency. Conversely, to obtain a high frequency localization, the long time window is required resulting in poor time localization.

To tackle this, several time-frequency distributions have been formulated that optimize this tradeoff between time and frequency localization. However, to jointly achieve high resolution in both time and frequency, these methods rely on assumptions that are valid only for specific classes of signal. The reassignment method (RM), on the other hand, is an alternative to such approaches, that sharpen TF distributions, while maintaining temporal localization [37], [38], [39]. This can be achieved by considering the Wigner Ville distribution (WVD) and the spectrogram representation of

24
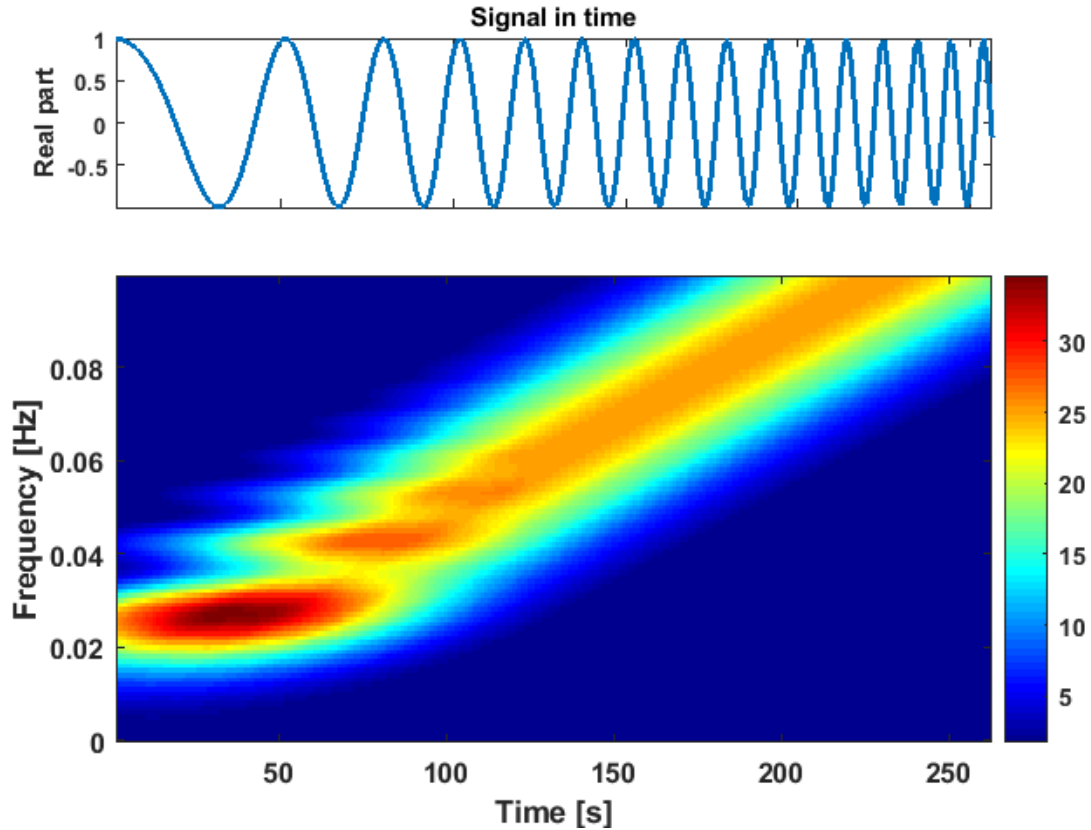
Figure 2.1: Short-Time Fourier Transform Time-Frequency Representation of a Linear Chirp.

a time-series. The WVD is known to localize certain classes of signals very well in the TF plane, but it also introduces interference terms or cross terms. The spectrogram on the other hand, smooths the sharpened TF distribution, which can also the effects of the cross terms. A tradeoff is required between the sharpness or localization of TF distribution and the interference in the TF- distribution due to the cross terms. The RM improves TF localization by adjusting for the shifts in TF distribution due to smoothing by the spectrogram by i) determining the appropriate TF location where the principle energy distribution lies, also referred to as the centroid, ii) shifting the TF distribution in the spectrogram from its original location to the centroid of the distribution.
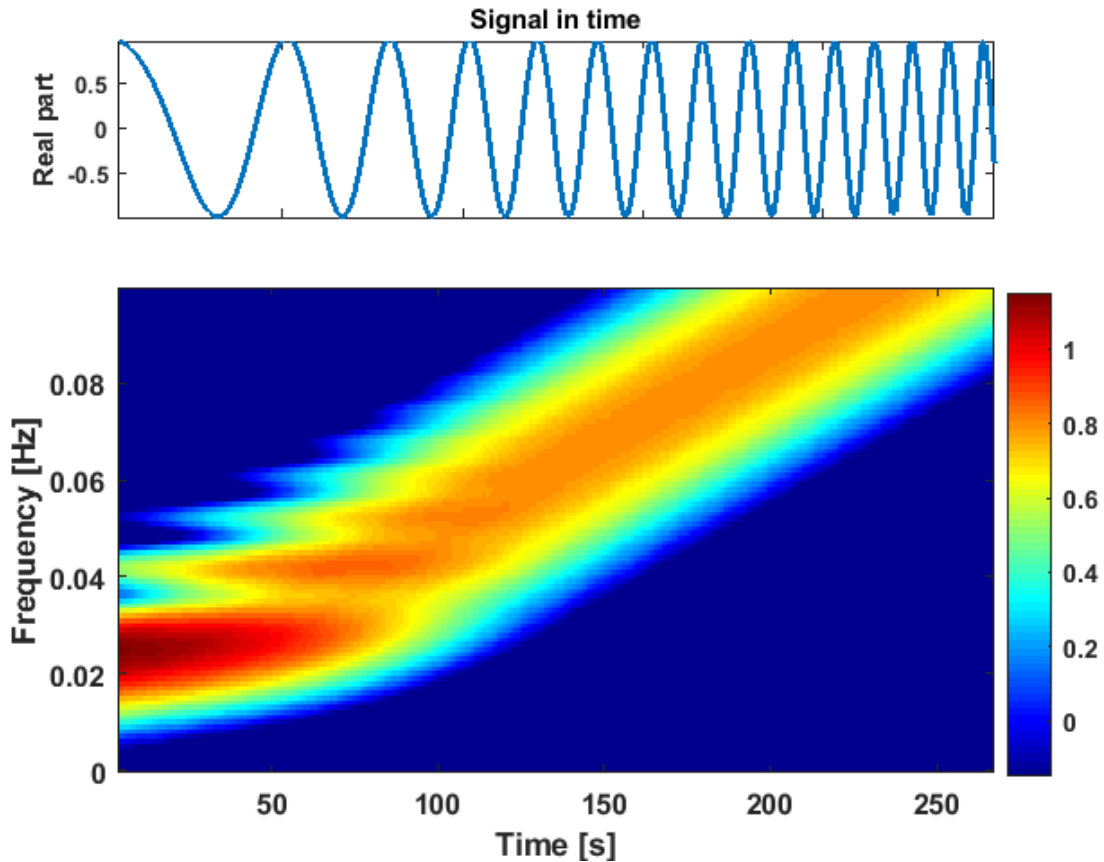
Figure 2.2: Spectrogram Time-Frequency Representation of a Linear Chirp.

Reassignment of a time-frequency representation such as the spectrogram is also effective for analyzing multicomponent signals with separate time-varying frequencies for each component. The reassigned spectrogram representation of a simulated linear chirp is shown in Figure 2.3 which shows the linear increase in frequency over time. Here, the improvement in time-frequency localization over the STFT and spectrogram can be seen making it more ideal to accurately capture time-frequency behavior of a non-stationary signal.
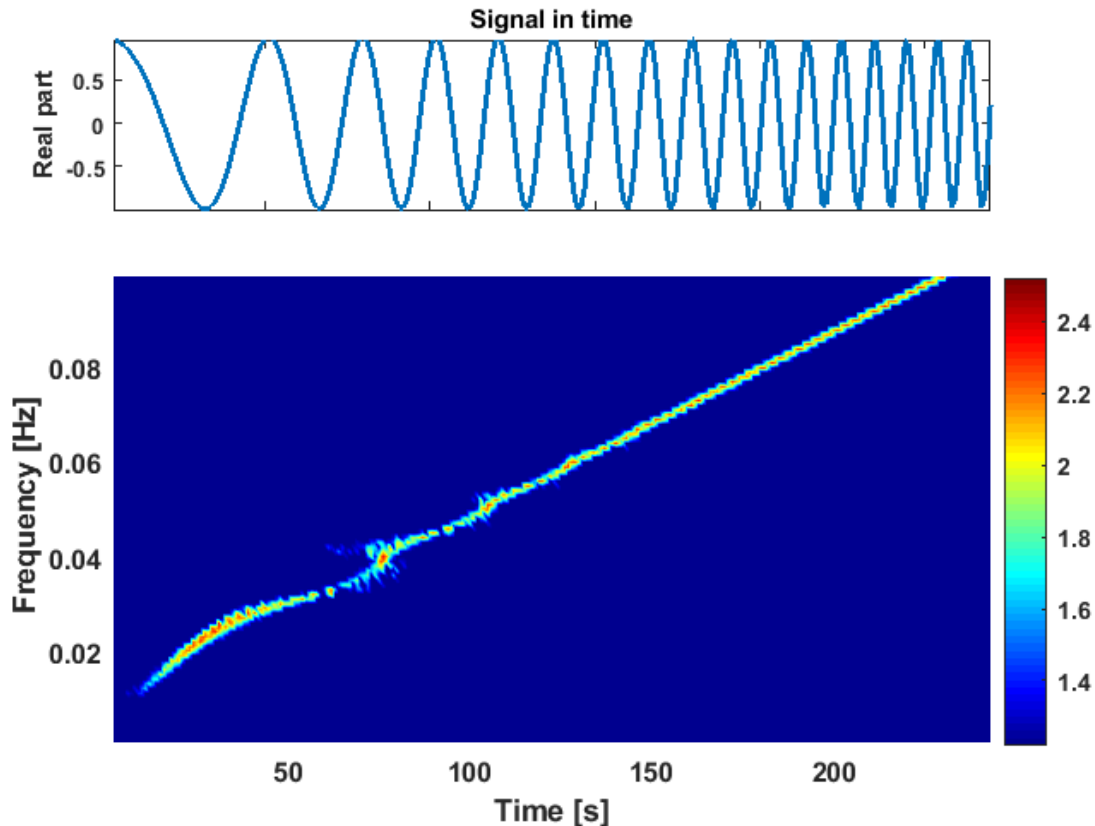
Figure 2.3: Reassigned Spectrogram Time-Frequency Representation of a Linear Chirp.

2.3 Latent Space Representation And Reconstructions: Anomaly Detection And Interpretation

While particle filtering models and estimates the temporal properties of a dynamic system, learning the expected features of the time-series at every observation sheds light on changes and anomalies in the system. The spectral properties of satellite images across multiple bands act as features that define the expected spectra or composition of the dynamic system. In this work, an autoencoder is used to model the expected spectral representation of frequently acquired multispectral satellite time-series.

Let the spectral features of each observation z_k in n dimensions be represented

27
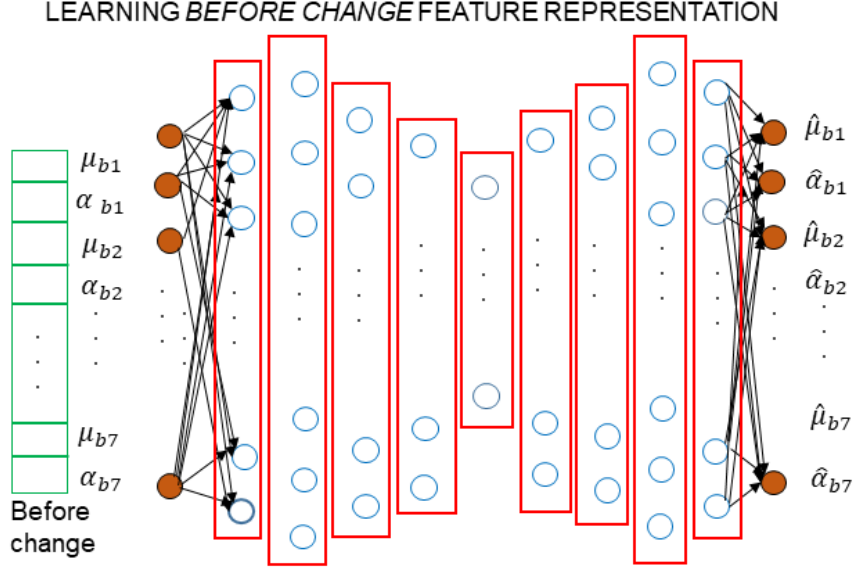
Figure 2.4: Autoencoder Approach for Learning Expected Spectral Representation in Stable Stages of Land Cover

as $z_{k,\lambda} = [z_{k,1}, z_{k,2}, \ldots, z_{k,n}]$. The autoencoder approach is adopted to learn the latent space representation from time-series spectra $z_{k,\lambda}$ of all samples in a training dataset. Autoencoders are neural network architectures where the input is transformed to produce the output with minimum distortion and forms an unsupervised learning approach through which the input data is reconstructed [40], [41]. The input $z_{k,\lambda} \in \mathcal{R}^n$, is transformed to produce the output $\tilde{z}_{k,\lambda} = \mathbb{D}(\mathbb{E}(z_{k,\lambda})) \in \mathcal{R}^n$, such that $\mathbb{E}$ is a class of functions forming the encoder that maps the input to its encoded representation $\mathbb{E}(z_{k,\lambda}) \in \mathcal{R}^m$, and $\mathbb{D}$ is a class of functions forming the decoder that maps the encoded input representation to its output representation $\mathbb{D}(\mathbb{E}(z_{k,\lambda})) \in \mathcal{R}^n$, where $m < n$. The corresponding output $\tilde{z}_{k,\lambda}$ is essentially a reconstruction of the input and the tilde notation is used to represent a feature reconstructed by the autoencoder. Thus, this form of the autoencoder (undercomplete) performs a compression, where it learns the most salient characteristics of the data in the latent space and reconstructs itself as shown in Figure 2.4. The learning process can be described by minimizing a
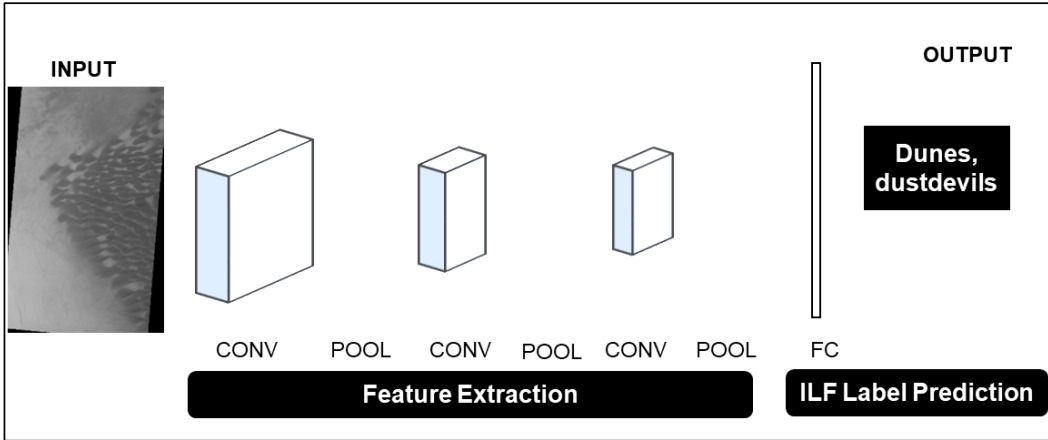
Figure 2.5: Convolutional Neural Network for Detecting Planetary Features of Interest.

loss function

$$L(z_{k,\lambda}, \mathbb{D}(\mathbb{E}(z_{k,\lambda}))), \tag{2.9}$$

where L is a loss function that penalizes $\mathbb{D}(\mathbb{E}(z_{k,\lambda}))$ for being dissimilar with $z_{k,\lambda}$ using metrics such as mean squared error or mean absolute error. In this work, the training dataset for the autoencoder are selected from the stable phase of satellite image time-series for a given region. In this process, the autoencoder learns to reconstruct the most salient features of the time-series of nearly stable phase across all spectral channels.

2.4 Novelty Detection From Planetary Image Descriptors

For the objective of retrieval and prioritization of planetary observations, detection of interesting and relevant geological classes is essential to summarize the image for determining its novelty. Detection of these classes is performed using convolutional neural networks (CNN). CNNs are neural networks that are specially suited for processing data that has a grid-like topology or 2-D structure and have been extremely
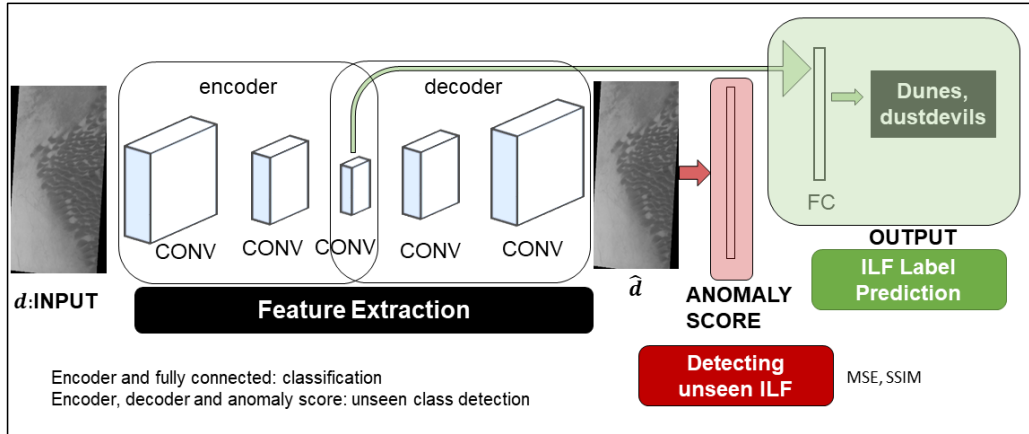
Figure 2.6: Convolutional Autoencoder for Detecting Planetary Features of Interest and Previously Unseen Features.

successfully at feature extraction, classification and similar image analysis techniques in general, which has provided the motivation for utilizing these architectures for detecting spatial features of geological landforms.

These neural networks are comprised of convolutional layers which takes in the image as input and performs convolutional operation across the image depth or volume with spatial filters or kernels to produce feature maps [40] as shown in Figure 2.5. By stacking hierarchies of such layers across the feature maps from previous layers, the weights of the filters are learned in the training phase to identify the different classes in the training dataset. The testing phase exploits the learned weights to determine the appropriate class. In addition, such architectures have a vast number of associated hyperparameters and the most appropriate combination for a given dataset is determined through extensive experimentation. In addition, convolutional autoencoder architectures shown in Figure 2.6 are also being utilized for image labeling. In addition to detecting features of interest, these architectures also enable detecting anomalous observations that are previously unseen or are rare in the training data. These labels are used to summarize or describe the images which are then

to be prioritized.

Chapter 3

DATASET DESCRIPTION AND PREPROCESSING

3.1 Land Surface Reflectance: Instruments And Data Collection

The data acquisition technique by instruments onboard satellites (both airborne and satellite platforms) described as remote sensing, refers to the process of gathering of information about a surface of interest without being in physical contact with it. This information is noted from the interactions and changes induced by the surface in surrounding fields (electromagnetic, acoustic, potential), but *remote sensing* is primarily concerned with changes and interactions with the electromagnetic field [42]. This includes surface-field interactions over the entire electromagnetic spectrum ranging from the radio waves to the gamma ray regions of the spectrum.

Remote sensing has applications in a wide array of disciplines and the class of electromagnetic spectrum utilized is specific to the phenomenon of interest. Depending on source of electromagnetic radiation incident on the surface, the mode of data acquisition can be categorized as active or passive. Passive remote sensing on the other hand, is dependent on the electromagnetic waves from the incident solar radiation, which interacts with the surface and gets reflected back through the atmosphere and is recorded at the sensor. The datasets used in this work are collected using passive remote sensing and hence the subsequent discussions is concerned with this mode of data acquisition only.

The data recorded by the passive mechanism is mostly over an area and therefore these datasets have a spatial dimension to it. In addition, the given area may also be studied with several spectral channels or bands depending on the phenomenon of

interest and the instantaneous field of view is observed with each of these spectral filters. Remote sensing datasets have three fundamental information dimensions which are:

1. Spatial: This relates to the resolving power and separability of different components on the observed surface and is representative of area on the surface captured by a unit element (pixel) in the dataset. Higher spatial resolution enhances the sensor's ability to resolving surface components.

2. Spectral: This relates to the ability of the instrument to resolve a given scene using different channels or bands and the selection of bands is application-specific. The electromagnetic radiation incident on the sensor after reflection is split or separated into these bands by corresponding filter elements, which is then recorded on the detecting medium of the instrument as individual band observations. Higher number of bands (spectral resolution) is important for identifying spectral characteristics of the observed surface.

3. Temporal: The temporal dimension of the dataset relates to how frequently the instrument revisits the same region and obtains observations. The lower the time lag between two subsequent revisits, the higher is the instrument's temporal resolution resulting in more frequent observations of the same region. This is ideal for monitoring dynamic processes of different Earth system components.

The solar radiation incident on the surface carries radiant energy Q, which is proportional to the frequency ν ($\nu = 1/\lambda$) of the electromagnetic wave and can be expressed as $Q = h\nu$, where h is Planck's constant. This incident energy is reflected back in the specular direction (angle between incident light and normal to the surface is equal to the angle between the reflected light and the normal) or scattered in all directions (lambertian surface) from the incident surface or is transmitted through

the material on which the radiation is incident. Satellites operating in passive modes record the radiance which is the reflected radiation from the surface incident on the sensor. This is used to derive reflectance, which is a ratio of the energy reflected by the surface to the total energy incident on the surface. When the energy of the incident wave corresponds to the difference between the energy levels of the electrons in the elements of the surface (electronic spectroscopy) or the frequency of the incident wave corresponds to the harmonic frequency (fundamental and higher-order overtones) of the surface mineral (vibrational spectroscopy), the energy from the wave is absorbed by the surface and this visible as absorption features which are reductions or dips reflected energy. These absorption features are characterized by its depth (reduction in reflectance), shape (pure or mixed compositional unit), and position (corresponding band). This interaction between the surface material and the wave of a given wavelength, resulting in absorption features is dependent on the molecular structure of the material and is therefore intrinsic to it.

Thus this acts as a signature that is specific to the observed surface. The presence of such signatures enable inferring surface composition as well as in detecting surface change and hence this reflected energy plays a crucial role in detecting and monitoring surface properties and its changes, which in turn has a vast range of applications. The resolution of a sensor in each of these dimensions are designed to meet application specific requirements and involves trade off between the dimensions. For example, satellites monitoring the weather, or studying dynamic properties of the Earth's surface, ocean and atmosphere have short revisit period, which in turn reduces the spatial resolution of the observations. On the other hand, instruments capturing high spatial resolution observations (useful for understanding and mapping various Earth system processes in details) trade off the spatial details for a lower temporal resolution.

In addition to resolution, the orbit of a satellite carrying a given instrument or sensor is also dependent on the application. Geostationary satellites orbit the Earth such that the its revolution around the Earth is of the same duration as that of the earth and hence appears to be stationary from the Earth. These satellites orbit the Earth at an altitude of approximately 36000 km in the same plane as the equator and images the same region of the Earth's surface at regular intervals at very coarse resolution for applications such as weather monitoring. Satellites in polar or near-polar orbit revolve around the Earth from pole to pole within a few hours imaging a portion of the surface in the field of view depending on the swath. On the other hand, in sun-synchronous orbits, satellite image the Earth in polar or near-polar orbits at a constant local solar time.

In this study, the primary source of data is obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS), which images the Earth's surface from the Terra and Aqua satellites orbiting the Earth in near-polar (ascending and descending modes) sun-synchronous orbits. The data is collected from the sun-lit side with a revisit period of 1-2 days. This data is collected in 36 spectral bands that are tuned to observe the land, ocean, clouds, aerosol properties, biogeochemistry, phytoplankton, ocean color, land, cloud and atmospheric temperature, and water vapor. The land cover bands used in this study are MODIS bands 1-7 ranging from 620 - 2155 nm. These include red (band 1), blue (band 3), green (band 4), near-infrared (bands 2 and 5) and shortwave infrared bands (bands 6 and 7), where bands 1 and 2 have a spatial resolution of 250 m while bands 3 - 7 are observed at 500 m.

After transmission to Earth, the observations are calibrated so that the observed sensor radiances are converted to meaningful physical quantities for scientific analyses. Further preprocessing steps are applied for atmospheric correction to remove the effects of gaseous and aerosol scattering and absorption, contamination due to clouds,

surface adjacency effects (remove effects due to reflection from contiguous pixels) [43] to create the MODIS MOD09 dataset [44]. An additional correction or adjustment called the Bidirectional Reflectance Distribution Function (BRDF) is applied to the dataset to adjust for the distribution of the reflectance, surface geometry and viewing angle as a function of the light (incident solar radiation) and view (radiation reflected back to the sensor) vectors [45].

3.2 Dataset Description

Multispectral observations acquired by short revisit period satellites image a region in multiple spectral channels capturing the reflected energy from the region. The images represent the reflected energy in the scene along each channel and each pixel represents the reflected energy from smaller sub-regions in the region, where the spatial extent of the sub-regions are defined by the spatial resolution. Therefore, each pixel captures the reflected energy from a sub-region whose area can be determined from the spatial resolution. A low spatial resolution (large area represented in each pixel) causes the reflectance at that pixel to be a mixed response from all the composite units in the region, while a high spatial resolution (smaller area represented in each pixel) indicates that the reflected energy is from a relatively uniform composite unit spatially and is more precise due to a more homogeneous spatial unit.

In this study, multispectral images relating to land surface and land cover captured by MODIS has been used, where the total area represented by a given image tile (adjacent, non-overlapping spatial blocks) is very large (10 degrees by 10 degrees at the equator) and a pixel represents a sub-unit in a given tile. A pixel in a given image tile is useful for monitoring the sub-region in that tile. Moreover, changes in land surface and land cover occur at a scale that more closely matches the spatial resolution of a pixel. Thus for datasets with moderate spatial resolution, monitoring

Earth system changes are carried out on a pixel level. By tracking the pixel level reflectance over time a sub-region can monitored for changes and anomalies. This pixel based reflectance over time forms the time-series that is utilized in the work to form the datasets.

3.2.1 Preprocessing Frequently Acquired Multispectral MODIS Time-Series

The MODIS dataset used in this study, MCD43A4 corrects for atmospheric interference, viewing angles and surface geometry in this way to retrieve the reflectance in each land surface bands from both Terra and Aqua platforms. These composite datasets are averaged over 16 day periods and released every 8 days at a spatial resolution of 500 m and consists of seven surface reflectance bands [46]. This forms the MCD43A4 dataset which is accompanied by a Quality Assurance (QA) dataset (version 5: MCD43A2, version 6: accompanies MCD43A4) that indicates the quality of each retrieved observation due to contaminating atmospheric effects and is used to approximate missing or low quality pixel values using linear temporal interpolation. In addition, anomalous spikes and drops (sudden deviations greater than 20% of the pre-change reflectance followed by a return to the pre-change values) in the time-series are also temporally interpolated. These are aberrations due to sensor artifacts or disturbances and are not associated with known change or regrowth phases.

In this thesis, we considered multiple multispectral MODIS time-series datasets from various spatial locations and these are used to validate the following stages of algorithms:

1. The four land cover classes (labeled as LC1–LC4 in Table 3.1) were selected in North America `https://ti.arc.nasa.gov/m/project/planetary/kml/lcc.kmz` belonging to urban, evergreen needleleaf, open shrubland and deciduous broadleaf classes. For each class, we selected 50 MODIS pixels over a 5-year period, over

seven MODIS surface reflectance bands, and created a NDVI time-series using Red and Near-Infrared bands. No change occurred in these datasets during the selected period. These datasets are used to model the pre-change nearly constant phase and evaluate the performance of the proposed work during these stages.

2. Detecting changes (forest fire examples) and evaluating dissimilarities in before and after change land cover stages. Three different time-series over a 10-year period from regions with land cover that experienced change due to forest fires in Southwestern United States (listed as CR1–CR3 in Table 3.1). These forest fire examples are used to compare the land surface dissimilarity before and after land cover change and are used to evaluate the performance of the proposed methods when the time-series consists of change points.

3. Evaluating spectral separability of different classes of change events: For the evaluation of class separability of land cover change events from spectral signatures at change point, three forest fires, three flood, one drought and one coastal land gain example has been considered from different geographic locations. The considered time-series length is varied and is indicated in the table along with its geographic locations. The pre-change spectral similarity is determined from the time-series spectra during the first three years and the subsequent observations are used as a test set. These examples are selected from known MODIS detected change events that are highlighted at `https://earthobservatory.nasa.gov/topic/natural-event` as many smaller scale changes may not be distinguishable at MODIS spatial resolution. Additionally, seasonal disturbances were observed to be particularly pronounced at the study region in the MODIS v6h10 tile and hence a further noise removal step was applied in addition to the general preprocessing steps defined above using the Savitsky-Golay filter using the TIMESAT toolbox to the noisy reflectance time-series in the region [47].

Table 3.1: Description of Land Cover Datasets

Land Cover	Description	# of Pixels
LC1	Evergreen Needleleaf	50
LC2	Open Shrubland	50
LC3	Urban	50
LC4	Deciduous Broadleaf	50
CR1	Wallow Fire	95
CR2	Horseshoe 2 Fire	60
CR3	Zaca Fire	45
CR4	Sava River Flood 1 (Croatia)	35
CR5	Sava River Flood 2 (Bosnia and Herzegovina)	25
CR6	Coastal Marsh Flood (Louisiana)	30
CR7	Coastal Land Gain (Atchafalaya Basin)	30
CR8	Drought (Northern California)	30

All pixels (line and sample) in the datasets described above were located in the respective image tiles using the MODLAND tile calculator `https://landweb.modaps.eosdis.nasa.gov/cgi-bin/developer/tilemap.cgi`. Time-series of each pixel over every band was extracted and the low quality observations indicated by the quality assurance dataset and missing values were flagged using the remote sensing toolbox Davinci `http://davinci.asu.edu/`. These flagged values were interpolated using linear temporal interpolation. We also considered the Landsat derived burn severity map at 30 m spatial resolution for each of the regions available from [48] for validating the fire severity mapping.

3.3   Preprocessing Multispectral THEMIS Images

The THEMIS images considered in this study comprise of the Visible (VIS) images which are preprocessed on being received by decompression, radiometric calibration and noise removal to arrive at the 8-bit digital numbers (DN) [49]. The adopted calibration steps for the VIS data are described in `http://static.mars.asu.edu/pds/ODTSDP_v1/document/sdpsis.pdf` to derive the VIS-Apparent Brightness Record (VIS-ABR), which are scaled versions of the scene radiance. The VIS-ABR data is a single band image derived from the calibrated VIS-band-3.

Chapter 4

TIME-VARYING MODEL OF LAND COVER CHANGE

4.1   Constant Frequency Model of Land Cover Change

Satellite image reflectance time-series (mostly Normalized Difference Vegetation Index NDVI time-series) have been used to model land cover behavior by capturing annual and intra-annual trends. These models are characterized by hidden time-varying parameters that denote the state of the land cover and deviation in these parameters indicates anomalies or changes in land cover. Thus land cover modeling enables monitoring, detecting changes and analyzing regrowth dynamics of the given region.

Modeling of satellite image time-series is based on harmonic analysis of the time-series to adequately capture its most important characteristics due to seasonality. Fourier analysis of land cover time-series is used to extract parameters mean, amplitude and phase that are characteristic to the type of land cover class [33, 34]. Although modeling of land cover time-series using Fourier analysis characterizes its seasonality from the extracted parameters, the assumption of repeating periodic cycles in time limits such models from capturing the time-varying nature of these parameters. A constant frequency (CF) model with time-varying parameters was proposed to address this limitation of parameters extracted by Fourier analysis and was shown to improve pre and post change class separability over Fourier transform parameters [14]. In addition, the estimated parameters were also shown to be an improved change detector (higher accuracy and lower detection delay) than direct metric based approaches [15].

As the model is nonlinear, the extended Kalman filter (EKF) is used to estimate the dynamic parameters by linearizing the observation model [14, 30]. The constant frequency in the CF model assumes that land cover has an annual vegetation growth cycle that remains constant with time.

The CF model parameter EKF estimation was adapted to improve land cover change detection using repeated sequential probability ratio test statistics and training data [18, 50]. The detection threshold was based on minimizing false negatives, false positives and the mean detection delay performance metrics. This approach was shown to work well for beetle infestation and non-Gaussian time-series data. However, this model cannot capture the time-varying frequency during and after change and hence may not be adequate to represent ecosystem behavior.

4.1.1 Dynamic Parameter Estimation For Land Surface Reflectance State Estimation

Due to the periodicity of land cover, Discrete Fourier Transform (DFT) analysis of the time-series highlights harmonic parameters characteristic to it [14, 33–35]. These parameters consist of mean μ, the average land cover state, amplitude α, the annual component of the time-series representing the maximum variation of the time-series about the mean, phase $\phi \in (0, 2\pi)$, the time taken for the time-series to reach its peak annually [33–35]. However, these parameters cannot capture variations in land cover with time, such as inter-annual differences in seasonality [14, 33]. To overcome this, a constant frequency (CF) cosine model (assuming a constant land cover change rate or frequency) that allows the state parameters to vary over time had been proposed which expresses the $k - th$ observation z_k as

$$z_k = \mu_k + \alpha_k \cos(2\pi f_0 T_s k + \phi_k) + \eta_k, \ k = 1, \ldots, K, \tag{4.1}$$
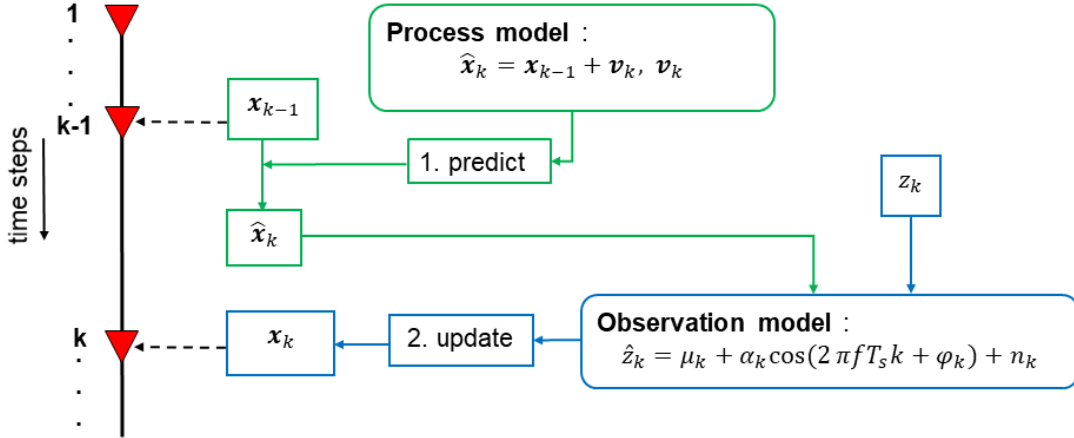
Figure 4.1: Predict and Update Steps for Dynamic State Estimation.

where μ_k represents the average land cover state, α_k represents the variation of the time-series about the mean, and ϕ_k represents the time taken for the time-series to reach its peak annually, $f_0 = 1/365$ represents land cover change rate and T_s is the sampling period (in days per time step), $K = Y\lceil \mathcal{D}/T_s \rceil$, Y is the number of observed years, $\mathcal{D} = 365$ days per year and η_k is the observation noise [14]. This time-varying state vector $\mathbf{x}_k$ captures the parameter variation due to seasonality. This work proposes a particle filter estimation of the state vector as the linearization by the Extended Kalman Filter is expected to introduce errors in the nonlinear observation model.

4.1.2 Bayesian Parameter Estimation

Bayesian filtering methods, such as the Extended Kalman Filter (EKF), for estimating time-varying state parameters are based on tracking the state probability distribution function (pdf) conditioned on available observations [30]. The parameters of the state vector $\mathbf{x}_k$ have been estimated sequentially using the equation (4.4) as the process model governing state vector evolution over time and equation (4.1) as the observation model that relates $\mathbf{x}_k$ with the noisy observation z_k using an EKF

43

Particle Filter: Predict Step

Parameter distribution representation using N_s particles, $x_k^{n_s}$, with weight $w_k^{n_s}$

$$\boldsymbol{x}_k = [\mu_k, \alpha_k, \varphi_k, f_k]$$

$\boldsymbol{x}_{k-1}$, (N_s particles at $\boldsymbol{x}_{k-1}$)

applying process model
$$\widehat{\boldsymbol{x}}_k = \boldsymbol{x}_{k-1} + \boldsymbol{v}_k$$

$\widehat{\boldsymbol{x}}_k$

Particle representation of state vector, size ≡ weight
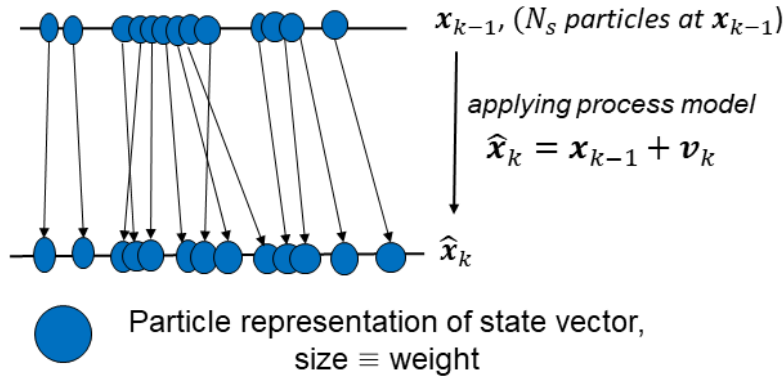
Predict step to obtain prior distribution $p(\hat{x}_k \mid z_{k-1})$

Figure 4.2: State Vector Prediction

formulation due to the non-linearity of the observation model[14, 15, 50]. The EKF is a modification of the Kalman filter (KF). The KF assumes all models involved are linear, and it is based on Bayesian inference and linear quadratic estimation. The EKF uses the same KF computation steps by first linearizing any nonlinear functions in the observation model [30]. However, when a model function is highly nonlinear, such as in the case of a sudden land cover change, the EKF linearization may not be sufficient to avoid large estimation errors. This is shown in Figure 4.1.

4.1.3 Bayesian Parameter Estimation

A sequential Monte Carlo method, such as the Particle Filter (PF), has been adopted to estimate $\mathbf{x}_k$, as the CF model described in equation (4.1) has a nonlinear observation function [20, 21]. The PF estimates $\mathbf{x}_k$ by numerically estimating the
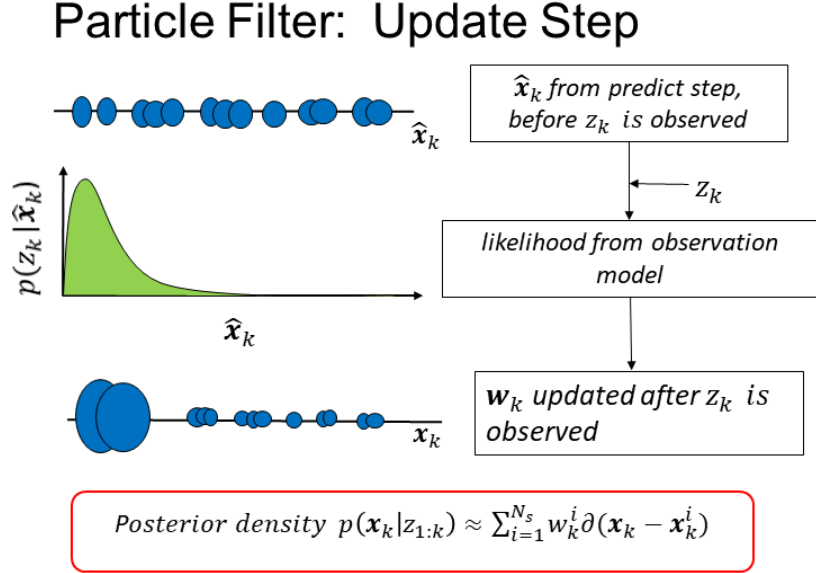
Figure 4.3: State Vector Update

posterior pdf sequentially as

$$p(\mathbf{x}_k \mid z_k) \approx \sum_{n=1}^{N_s} \mathbf{w}_k^{(n)} \, \delta(\mathbf{x}_k - \mathbf{x}_k^{(n)}) \,, \tag{4.2}$$

where $\delta(\cdot)$ is the delta function [30]. The pdf is approximated as a weighted linear combination of N_s particles (or samples) $\mathbf{x}_k^{(n)}$, $n = 1, \ldots, N_s$, that are drawn from a proposal density. At time step k, the nth particle $\mathbf{x}_k^{(n)}$ is drawn from the prior pdf $p(\mathbf{x}_k^{(n)}|\mathbf{x}_{k-1}^{(n)})$ using equation (4.4) (prediction step). The nth particle weight $\mathbf{w}_k^{(n)}$ is computed from the likelihood $p(z_k|\mathbf{x}_k^{(n)})$ using equation (4.1) (update step). This is shown in Figures 4.2 and 4.3. Using the posterior pdf in equation (4.2), the estimated model parameter is $\hat{\mathbf{x}}_k \approx \sum_{n=1}^{N_s} \mathbf{w}_k^{(n)} \mathbf{x}_k^{(n)}$. As expected, the estimation accuracy increases with the particle number N_s [30].

4.2 Time-Varying of Models Representing Intra-annual Behavior of Crops

A nonlinear cosine model that takes into consideration the variation of vegetation growth with time has been considered in [16]. The proposed model consists of a constant annual frequency, that represents an annual vegetation growth cycle, as well
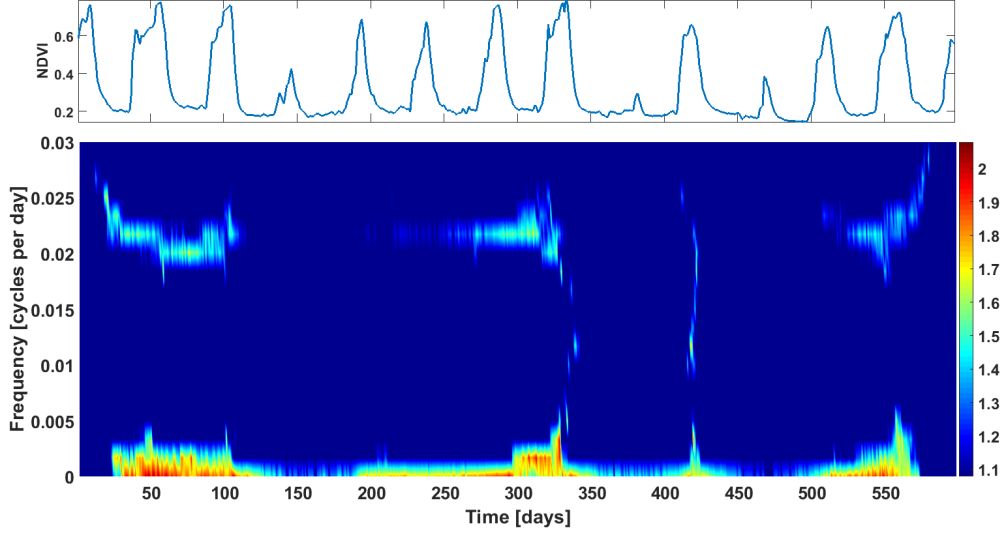
Figure 4.4: NDVI Time-Series and its Time Frequency Representation (TFR) Using Reassigned Spectrogram of a Region in Northern California Affected by Drought in Years 2007, 2012 and 2014 (Approximately Around Time Steps 150, 368, 414).

as a nonlinear cosine time-varying phase function that represents a rate of change that accounts for intra-annual cycles of different land cover types. The model was used for identifying and predicting various phenological attributes of land cover classes that included various types of forests, different types of crops and shrublands. The method worked well, even by using time incomplete set of observations and was also shown to match land cover with periodically varying rate of frequency changes. This model has not been tested in identifying abrupt changes in land cover.

4.3   Time Frequency Analysis of Land Cover Time-Series

4.3.1   *Land Cover Variation Rate As Analyzed By The Time-Frequency Representations*

All time-frequency representations (TFR) detect a constant frequency for land cover (using the Normalized Difference Vegetation Index (NDVI) time-series) that

46
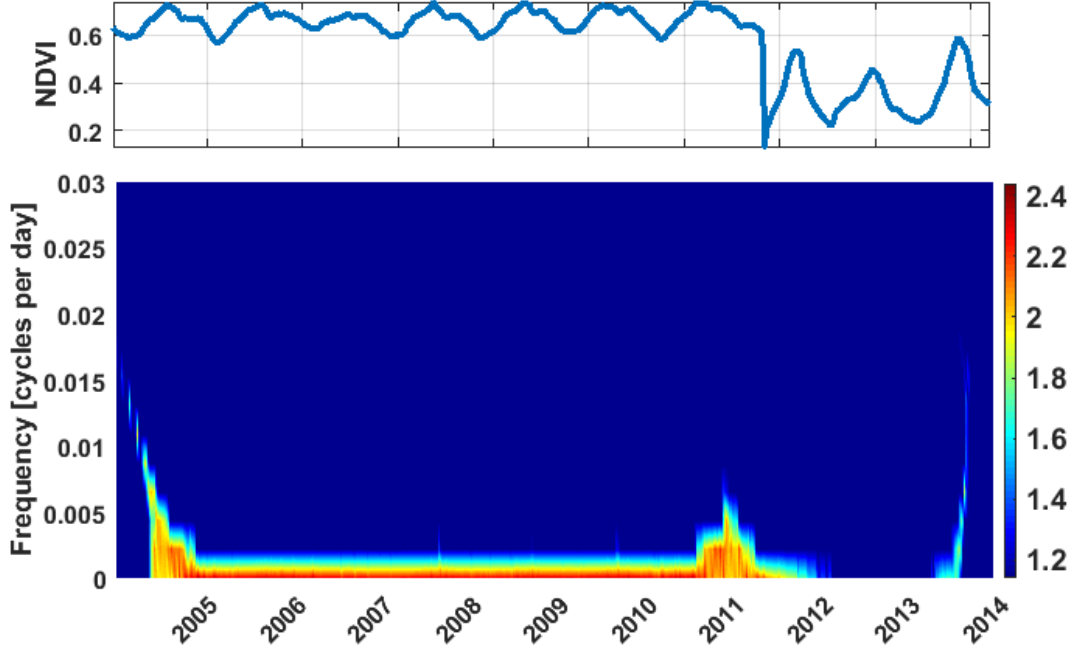
Figure 4.5: Reassigned Spectrogram Time-Frequency Representation of NDVI Time-Series of a Subset of Pixels Affected by the Wallow Fire (CR-1) in North America in Logarithmic Scale.

does not exhibit significant change (only minor disturbances) during the period of study, representing the nearly constant growth rate. On the other hand, when the land cover under consideration undergoes major changes due to drought, wildfire or floods, it alters the land cover or vegetation growth rate which is also captured by all TFRs. As the localization properties of the reassigned spectrogram are more accurate, TFR estimated using this technique are shown in Figures 4.4, 4.5, 4.6 and 4.7. Figures 4.5, 4.6 and 4.7 show single major change in frequency, while Figure 4.4 shows multiple change points in frequency in the given region in Northern California that is affected by drought in years 2007, 2012 and 2014 (approximately at time steps 150, 368, 414). TFR shows constant change rate of the time-series followed by change during drought phases which are followed by regrowth [20, 51]. Thus, for such events the constant frequency or time-series variation rate does not hold true and it is important to

47
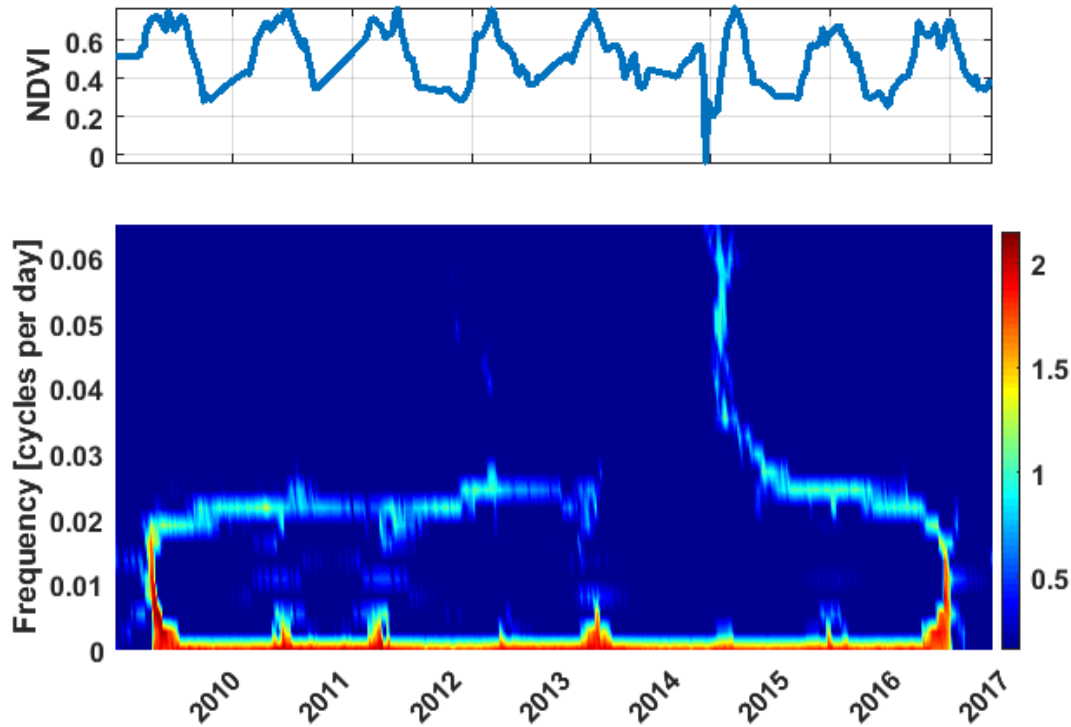
Figure 4.6: Reassigned Spectrogram Time-Frequency Representation of NDVI Time-Series of a Subset of Pixels Affected by the Flooding of the Sava River (CR-4) in Europe in Logarithmic Scale.

incorporate this variability in land cover models for a more accurate representation of ecosystem dynamics due to change. Additionally, the TFR also highlights the regrowth trend in land cover once the effect of change begins to subside and eventually declines. The degenerative effect of change (sharp decrease in frequency or variation rate) is more pronounced for a longer period for an event like the wildfire, and the regrowth (increase in vegetation growth rate) only becomes evident after a prolonged period. Floods, on the other hand induces a sharp reduction in land cover variation rate, however, this effect is short-lived, as usually the effects of flood (for example, the Sava river flood) is brief, after which the land cover is nearly restored to its pre-change conditions.

Land cover change as well as post-change dynamics cause the vegetation growth

48
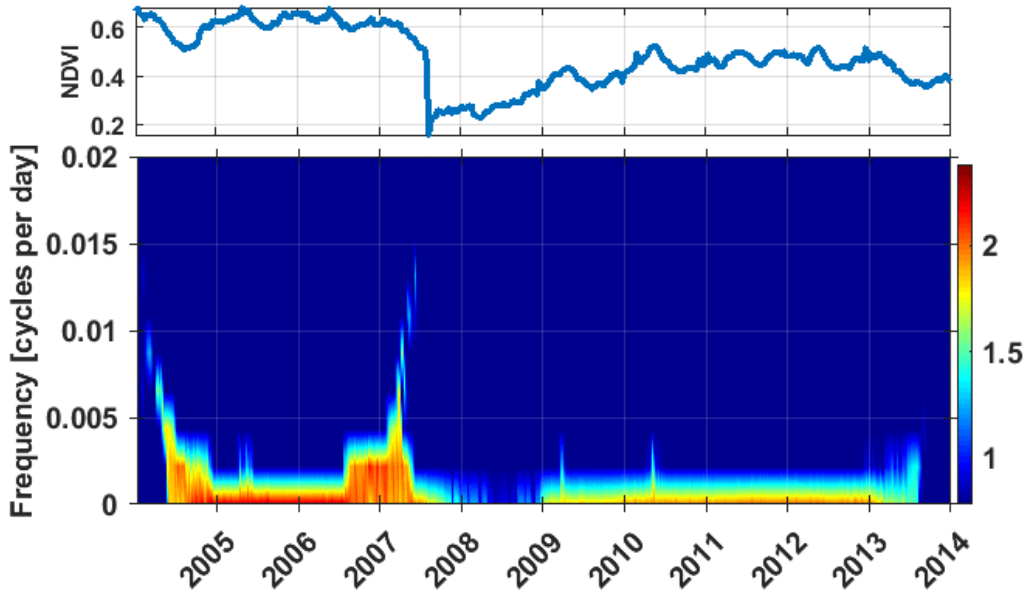
Figure 4.7: Reassigned Spectrogram Time-Frequency Representation of NDVI Time-Series of a Subset of Pixels Affected by the Zaca Fire (CR-3) in North America in Logarithmic Scale.

rate to vary with time. Thus, time-varying growth rate or frequency is a vital trend to capture, especially for Earth system components such as land cover which frequently exhibits transitioning from one phase to another over time. As the CF model does not account for this variation, the use of the model may restrict accurate determination of ecosystem behavior due to change. Thus the use of the CF model for a near real time land cover change detection does not capture the state parameter dynamics, requiring an improved model that can accurately model the shifts in time-varying state parameter even through phases of change in land surface.

4.4   Time Varying Frequency Model

The constant frequency of such models cannot estimate the dynamic land cover change rate, which is seen to vary over time using TFR, specially due to change and post-change regrowth. Moreover, the CF model is designed for parameter estimation

49

during near-constant land cover states and consequently, change point decision making follows a retrospective approach where an optimum class separability threshold is determined from analyzing the before and after change time-series in batch mode by jointly minimizing false positives, false negatives and detection delay. This limits its ability to sequentially estimate the state vector during transition between stages due to change as well as post-change non-stable stages when the land cover is undergoing regrowth. In addition, as shown by the time-frequency representations, land surface reflectance time-series (hence land cover time-series) undergoes several shifts in underlying land cover state and the growth rate or change rate in land cover. Current change detection approaches cannot accurately estimate and detect this variation in parameters near-real time as the models are formed from Fourier based approaches that assume constant growth rate.

As the CF model does not account for this variation, the use of the model may restrict accurate determination of land surface behavior due to change. Thus the use of the CF model for a near real time land cover change detection does not capture the state parameter dynamics, requiring an improved model that can accurately model the shifts in time-varying state parameter even through phases of change in land surface.

In this section, a new land cover model and sequential state estimation approach is proposed that estimates this variation and accounts for varying vegetation growth rate during and after change with a time varying frequency parameter. In addition, to tackle the higher non-linearity of the proposed model, a Sequential Monte Carlo estimation of the posterior distribution of the state vector is also introduced, which can accurately estimate the parameter variation of the proposed model

Let $\mathcal{Z} = \{z_1, z_2, \ldots, z_K, \ldots\}$ be the set of sequential multispectral observations acquired over Y years over the same geographic region with time. Each observation

z_k, is a pixel in a multispectral image acquired in Λ bands or spectral channels. As the time-series under consideration has been acquired over Y years, each year is said to be comprised of $\lceil \mathcal{D}/T_s \rceil$ observations, where T_s is the sampling period (in days per time step) and $\mathcal{D} = 365$ days per year. As z_k varies over time, the observation along each band is characterized by state parameter vector $\mathbf{x}_k$ that represents the underlying land surface reflectance state and the observation and state vector are related by a function or model such that $z_k = h(\mathbf{x}_k)$. The observation set $\mathcal{Z}$ is expected to be comprised of several land cover stages with varying length, where the observations at each stage are best represented by a state vector distribution. Moreover, with the variation in land surface state, the function or model describing the land cover also changes.

The objective of time-varying modeling is to sequentially determine the model that best defines the current land surface state by relating z_k with $\mathbf{x}_k$ with minimum error and detecting a shift in the approximated distribution of $\mathbf{x}_k$ as land cover change. The model parameter estimation process is essentially performed for every pixel across each channel (or band ratio) in the image and subsequently all references to z_k refers to the pixel observation along a single channel.

This is to be followed by change event classification, which aims to analyze the multispectral characteristics during such shifts to distinguish between different types of change events. For this objective, the model parameter estimates across each band are then considered simultaneously. In this work, a total of C change events were considered. When the proposed modeling technique is applied to a set of regions $\{r_1, \ldots, r_R\}$ and land cover change is detected at a subset of these locations, the additional objective is to identify the type c (where $c \in C$) of event in each of these locations based on its spectral characteristics.

4.4.1 Proposed Time Varying Frequency Model for Land Cover Variation Rate Estimation

A new land cover model with a time-dependent frequency has been proposed to capture changes and regrowth dynamics due to forest fires [20, 21]. When a given land cover experiences a sudden change, its model should represent the dynamic disturbance observed in time-series data. This disturbance is observed in the time-frequency analysis in Figures 4.5, 4.6, 4.7 of time-series data from satellite images of three regions that underwent land cover change due to forest fires and flood [51]. Specifically, the reassigned spectrogram time-frequency representation of the data in Figure 4.7 shows the decrease in growth cycle after about 3.5 years (as a result of the Zaca fire in July 2007) and the slow increase in vegetation growth starting after about 5.5 years (almost two years after the fire). To allow for a time-dependent vegetation growth cycle f_k, the time-varying frequency (TVF) model is used to represent the time-series as

$$z_k = \mu_k + \alpha_k \cos(2\pi f_k T_s k + \phi_k) + \eta_k, \ k = 1, \dots, K. \tag{4.3}$$

For the TVF model, the transition state equation in (4.4) now also includes the time-varying frequency f_k, resulting in the state vector $\mathbf{x}_k = [\mu_k \ \alpha_k \ \phi_k \ f_k]$ [20, 21]. The state transition between time steps is normally modeled by first order Markov chain as

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{v}_k. \tag{4.4}$$

Here, $\mathbf{v}_k$ is a random vector that represents possible errors in the transition model [14, 20, 21]. The CF model with time-varying mean, amplitude and phase parameters is shown to improve estimation of land cover dynamics and used for land cover class separation and change detection [14, 15]. These models are characterized by hidden
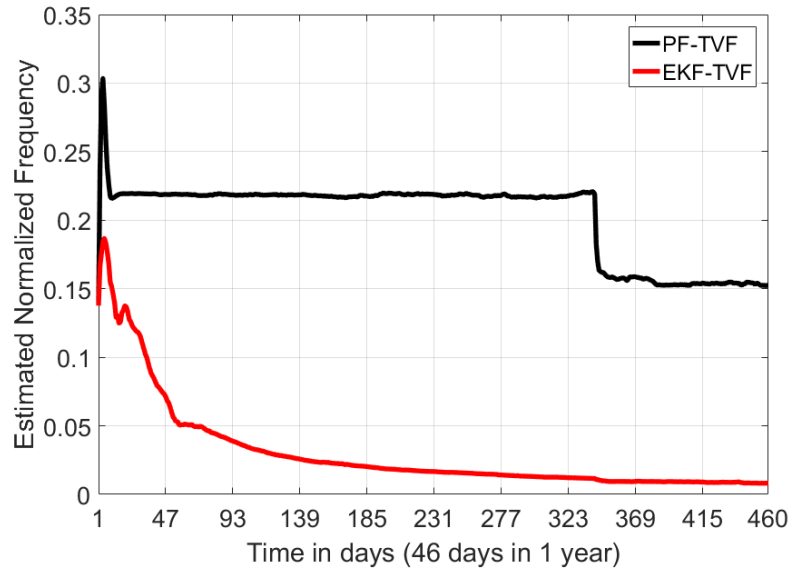
Figure 4.8: Estimated TVF for CR-1 with Change Point Around Observation 341

time-varying parameters that denote the state of the land cover and deviation in these parameters indicate anomalies or changes in land cover. The state parameter estimation for the TVF model is obtained using both the EKF and PF with the expectation that PF will outperform the EKF due to the higher nonlinearity in the TVF model. While the process model remains the same as that of the CF model, the observation model is modified to equation 4.3.

4.5 Evaluation Metrics and Results

4.5.1 Estimation of Instantaneous Frequency

The frequency f_k in equation (4.3) represents the vegetation growth rate as it varies due to changes in land cover, as seen in Figures 4.8 and 4.9 for CR1 and CR3. Using the PF to estimate the TVF parameters reflects this change for all datasets and the PF-TVF estimation can capture important trends related to land cover change and regrowth dynamics when compared to PF-CF estimation. As the EKF-TVF resulted in higher RMSE, subsequent analysis of the TVF model for dissimilarity and
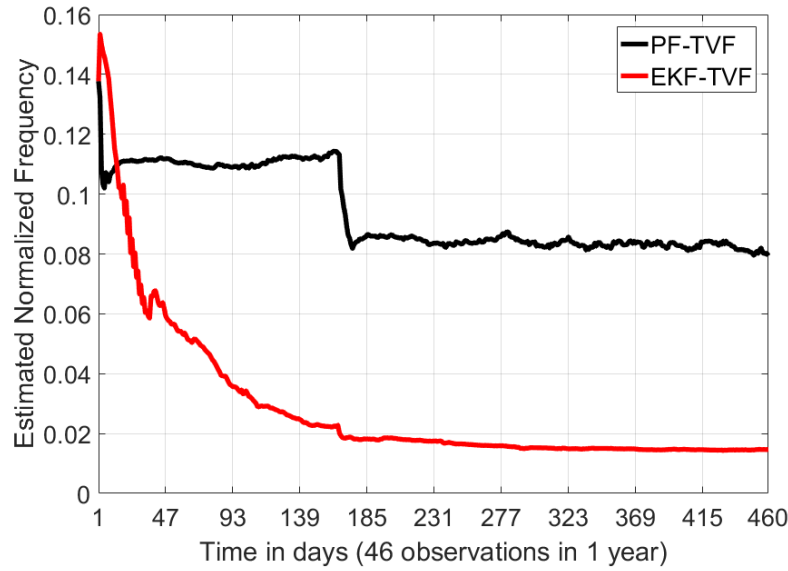
Figure 4.9: Estimated TVF for CR-3 with Change Point Around Observation 185

detection delay is only performed using the PF.

Frequency Variation Due to Land Cover State Transition: The TVF model attempts to capture the frequency variation or change rate variation in land cover by introducing a dynamic f_k. This allows the instantaneous frequency or land cover change rate to vary. Using the modified observation function of the TVF model in equation 4.3, the f_k parameter responds to sudden changes in land cover. However, the spectrogram and reassigned spectrogram representations show that this change rate of land cover growth is nearly constant during the stable phases and shows a sudden shift due to change and recovers during the regrowth phase. From this, it can be hypothesized that the frequency associated with land cover growth exhibits several transitions which can be modeled more accurately with multiple functions defining the observation model. Hence, for every observation of land cover a model selection approach is proposed that chooses the most appropriate model for the given observation based on instantaneous phase functions relating to the instantaneous frequency or varying land cover growth rate.

4.5.2 Model Parameter Estimation Accuracy

Parameter Initialization: The state parameters, transition and observation random processes for the PF and EKF were selected following the approach in [14], using 5% of the time-series pixels from each region for each band. For the regions with change, we considered only years before the change for parameter initialization. The initial state vector $\mathbf{x}_0$ for each model was selected by considering a uniform distribution about the average mean and annual periodic FT components of the pixels. The TVF model parameter ω_0 was initialized to $2\pi 8/365$. Changes in the initial values can influence the performance of the EKF [52, 53] and that of the PF. However, we ensured that the model parameters were initialized to the same values for both methods for fair comparison. The number of particles for the PF implementation was 2,000, and 2,000 Monte Carlo iterations were used. Also, sequential importance resampling was used to avoid degeneracy in the PF [30].

The model parameter estimation accuracy by the PF and EKF is evaluated using the root mean-squared error (RMSE). RMSE is averaged over all pixels and is computed between z_k and $\hat{z}_k$. It shows the error between the true observed time-series at k and the estimated time-series value from $\mathbf{x}_k$ using equation (4.1) or (4.3). This comparison between the observation and estimate is made as enhanced reflectance retrieval techniques as well as input from quality assurance datasets reduces the noise in the observations to provide a refined input for particle filter estimation. However, RMSE is not the best metric as the ground truth values of the parameters are not available. Moreover, it quantifies the model performance by relating the best fit estimate with the real time-series observation and does not capture the performance with respect to estimation of instantaneous frequency. Alternatively, the Normalized RMSE (NRMSE) metric can be used instead where the errors are expressed as a

percentage [54], [55].

However, to explain the behavior of the state vectors before and after change, we define dissimilarity measures. In the absence of ground truth values of the state vector this also serves as an indicator of how well the land surface variability before and after change is captured. We consider time-series observations of land cover that underwent change to analyze the dissimilarity in land cover using the estimated parameters before and after change. Thus, RMSE serves as a performance measure for the estimators, while the dissimilarity metric validates accurate parameter estimation and justifies parameter behavior by capturing the variation in land cover parameters before and after change. The RMSE averaged over all pixels for each band and each land cover class is computed between the observed z_k and the particle filter estimate $z_{k,\text{PF}}$ as

$$\epsilon = \left(\frac{1}{K} \sum_{k=1}^{K} \mid z_k - z_{k,\text{PF}} \mid^2\right)^{1/2}. \tag{4.5}$$

For each land cover LC1-LC4 and CR-1 to CR-3, the EKF and the PF is used to obtain the RMSE of the estimated parameters of the CF land cover models. For ease of reference, the following notation has been used to denote the method selected, followed by the model used: EKF-CF, PF-CF, EKF-TVF and PF-TVF. Table 4.1 shows the RMSE we evaluated for seven different bands, $\lambda = 1, \ldots, 7$, and the NDVI time-series for LC1–LC4. Results show that the PF achieved the lowest time-averaged RMSE for all bands. This is expected as the PF is designed to sequentially estimate nonlinearities. Thus, the parameter estimation RMSE for the PF-CF is lower than that for the EKF-CF. A similar trend was observed for the highly nonlinear TVF model. The high estimation error by the EKF-TVF model can attributed to the high nonlinearity of the TVF model, which cannot be estimated accurately due to the linearization step of the EKF.

Parameter Estimation Accuracy During Change: To evaluate model performance

Table 4.1: CF and TVF Model Parameter Estimation RMSE for Different MODIS Bands λ and Land Cover Classes.

Land Cover	Estimation Method	RMSE $\lambda = 1$	RMSE $\lambda = 2$	RMSE $\lambda = 3$	RMSE $\lambda = 4$	RMSE $\lambda = 5$	RMSE $\lambda = 6$	RMSE $\lambda = 7$	RMSE NDVI
LC1	PF-CF	0.0006	0.0057	0.0003	0.0005	0.0036	0.0028	0.0029	0.0152
	EKF-CF	0.0147	0.0245	0.0106	0.0132	0.0198	0.0175	0.0175	0.0402
	PF-TVF	0.0019	0.0128	0.0012	0.0014	0.0093	0.0062	0.0067	0.0191
	EKF-TVF	0.1076	0.0601	0.0639	0.0841	0.0872	0.1131	0.1128	0.0647
LC2	PF-CF	0.0024	0.0052	0.0013	0.0015	0.0073	0.0073	0.0074	0.0124
	EKF-CF	0.0180	0.0225	0.0150	0.0167	0.0256	0.0260	0.0263	0.0313
	PF-TVF	0.0049	0.0124	0.0022	0.0033	0.0209	0.0187	0.0189	0.0189
	EKF-TVF	0.0973	0.1011	0.0530	0.0906	0.1509	0.1411	0.1423	0.0608
LC3	PF-CF	0.0013	0.0027	0.0004	0.0014	0.0036	0.0028	0.0029	0.0053
	EKF-CF	0.0163	0.0196	0.0136	0.0162	0.0196	0.0187	0.0189	0.0257
	PF-TVF	0.0033	0.0050	0.0018	0.0030	0.0076	0.0070	0.0084	0.0092
	EKF-TVF	0.0610	0.0480	0.0533	0.0559	0.0678	0.0722	0.0734	0.0920
LC4	PF-CF	0.0003	0.0032	0.0025	0.0003	0.0003	0.0004	0.0024	0.0500
	EKF-CF	0.0046	0.1373	0.0047	0.0049	0.0048	0.0049	0.0051	0.0827
	PF-TVF	0.0012	0.0067	0.0010	0.0014	0.0013	0.0015	0.0006	0.0703
	EKF-TVF	0.0295	0.0610	0.0490	0.0260	0.0027	0.0031	0.0610	0.1971

and parameter estimation accuracy during land cover change events, the CR-(1-3) datasets are considered. These pixel time-series have been collected from regions that underwent change and consists of phases with abrupt and gradual deviation. It is expected that parameter estimation accuracy using the CF and TVF models is considerably affected by the land cover state even with the nonlinear PF estimation. During phases of stable land cover state, the posterior distribution of $\mathbf{x}_k$ can be modeled sufficiently well with the first order Markov chain assumption governing the process model along with the associated process noise. However, during change, specially during abrupt change phases due to events such as floods and forest fires,

Table 4.2: PF-CF, EKF-CF, PF-TVF & EKF-TVF RMSE: Overall, Before, During and After Change

Land Cover	Estimation Method	RMSE NDVI	RMSE before	RMSE during	RMSE after
CR1	PF-CF	0.0235	0.0080	0.0676	0.0070
	EKF-CF	0.0453	0.0228	0.1151	0.0185
	PF-TVF	0.0358	0.0138	0.0980	0.0130
	EKF-TVF	0.1155	0.0685	0.2079	0.0953
CR2	PF-CF	0.0110	0.0087	0.0193	0.0056
	EKF-CF	0.0294	0.0217	0.0595	0.0132
	PF-TVF	0.0160	0.0138	0.0248	0.0091
	EKF-TVF	0.0757	0.0501	0.1436	0.0542
CR3	PF-CF	0.0135	0.0086	0.0337	0.0074
	EKF-CF	0.0365	0.0272	0.0849	0.0198
	PF-TVF	0.0211	0.0145	0.0539	0.0101
	EKF-TVF	0.0750	0.0435	0.1713	0.0428

this assumption fails to model the sudden deviation in state vector distribution and consequently results in relatively larger estimation error. However, the PF estimation can adapt to this variation effectively and the adjusts to the new posterior distribution representing the changed land cover state. Thus, the estimation error is lowered to accurately track the shifted land cover temporal trajectory.

This is can be seen in Table 4.2. For Regions CR1–CR3 in Table 3.1 that underwent a sudden change, only the NDVI time-series has been considered, as it represents vegetation content in a region. In Table 4.2, the RMSE has been obtained per year before, during and after land cover change, and the overall error ($\text{RMSE}_{\text{NDVI}}$). For all the considered examples, the estimation error is lowest for PF-CF. Low $\epsilon^{\text{PF-CM}}$

before change shows that even in the absence of change, PF-CM obtains improved estimates. Low $\epsilon^{\text{PF-CM}}$ during change and $\epsilon^{\text{PF-CM}}$ after change indicates more accurate estimation of abrupt change due to the forest fire and post change trends by PF-CM. Although the PF-TVF RMSE is higher than the PF-CF RMSE, the proposed TVF model can estimate the time-varying frequency due to land cover changes.

4.5.3 Dissimilarity Measures

Model derived land cover and land surface reflectance parameter estimates characterize land cover or surface reflectance state and hence deviate due to any changes in the underlying region. Class separability of land cover state vector is a well studied problem to distinguish before and after change classes in batch mode with the knowledge of future observations. Here, the model parameters estimated sequentially using the CF and TVF observation models and the state vector estimated by the particle filter $\hat{\mathbf{x}}_k$ has been utilized as an indicator of dissimilarity between before and after change classes. The model that maximizes this dissimilarity measure between pre and post change class deviations responds better to the event and is therefore a better estimator of change.

In the subsequent sections, dissimilarity measures are used to assess the CF and TVF model response to change. As these parameters are estimated sequentially, they have been utilized in creating a binary hypothesis change detector that decides on the possibility of the current observation being a change point based on its deviation with respect to a past temporal window. Here, dissimilarity measures are compared, within fixed data windows, to sequentially decide whether a change has occurred based on current observations and trends from past observations. The detection threshold is adaptively determined from the variability of past observations, thus performing the change decision process near real-time. Using cost-sensitive mea-

sures to penalize delay when detecting change, we demonstrate an improvement in change detection performance with PF estimation. Although change detection could be performed directly from statistics derived from the NDVI time-series, using the estimated model parameter to detect change enables studying post change regrowth dynamics from parameters such as time-varying frequency. Additionally, it was suggested in [18], [50], [17] that detecting change from estimated parameter time-series of models achieved better performance over approaches that detected change directly from the NDVI time-series [56], [10] ,[11]. Additionally, most change events exhibit varied severity within the spatial change boundary. The dissimilarity estimates are also utilized as measure of the degree of change to distinguish pixels of different severity. This has been explored in the context of mapping fire severity.

4.5.4 Dissimilarity Measures from Parameter Estimates

Significant deviations in the estimated model parameters can provide crucial information for early detection of sudden land cover changes due to events such as forest fires. Model parameters have previously been used for land cover class separability and change detection studies [14, 15]. We consider land cover time-series collected over Y years in regions that underwent change due to forest fires in CR1-CR3. Moreover, in the absence of ground truth values of state parameter vector $\mathbf{x}_k$, the behavior of the state parameter vector estimated by the particle filter $\hat{\mathbf{x}}_k$, is used to analyze and explain the behavior and correctness of the parameters estimates, specifically through the dissimilarity measures of the state vector in the presence and absence of change. Specifically, we use measurements between years p and q, $p, q = 1, \ldots, Y$, at time step k in a given year, $k = 1, \ldots, L$, $(L = \lceil 365/T_s \rceil)$, to estimate land cover model parameter vectors $\hat{\mathbf{x}}_k^q$ and $\hat{\mathbf{x}}_k^p$. Using the ith element of each vector, $\hat{x}_{k,i}^q$ and $\hat{x}_{k,i}^p$, $i = 1, \ldots, M$, where M is the number of parameters in each model, the dissimilarity

60

measure between the pth and qth years is obtained as

$$d_{k,r}^{p,q} = \Big(\sum_{i=1}^{M} |\hat{x}_{k,i}^{p} - \hat{x}_{k,i}^{q}|^{r} \Big)^{1/r}. \tag{4.6}$$

The overall dissimilarity measure between the pth and qth years is $D_r^{p,q} = \sum_{k=1}^{L} d_{k,r}^{p,q}$. Selecting r yields different measures, such as Manhattan distance $D_{\mathrm{MD}} = D_1^{p,q}$ for $r = 1$ and Euclidean distance $D_{\mathrm{ED}} = D_2^{p,q}$ for $r = 2$ [57]. As seasonal variations can change the dissimilarity measure between parameters at different times of the year [14], another important metric is the maximum dissimilarity $D_{\mathrm{max}}^{p,q} = \max_k d_{k,1}^{p,q}$ between the parameters corresponding to the two years p and q.

4.5.5 Binary Hypothesis Change Detector

In the absence of change, low dissimilarity measure values are expected, but the value is expected to suddenly increase at the time of change. Assuming that there was no land cover change at the $q = (p-1)$th year, $p = 2, \ldots, Y$, we design a change detector for the pth year, at time step k (corresponding to the kT_s day of the year). The dissimilarity measure $d_{k,r}^{p,p-1}$ at time step k in the pth year is then compared to the dissimilarity measure within a window of length ℓ in the same year to sequentially make a change decision for each observation. This binary hypothesis change detector is based on a given critical threshold γ_k and the window of length ℓ. Specifically, at the pth year, forest fire is detected at the kth time index if

$$\mathcal{H}_{k,\ell,j} \geq \gamma_k, \text{ for every } j = k - \ell, \ldots, k - 1 ,$$

$$\text{where } \mathcal{H}_{k,\ell,j} = d_{k,r}^{p,p-1} - d_{j,r}^{p,p-1} . \tag{4.7}$$

The change detector depends on two important factors: the window length ℓ and the critical threshold γ_k. Both factors are determined by considering the overall climatic variations in the region in the past and current years. In particular, the threshold is

61

selected based on the variation in dissimilarity over all time steps in the previous year where no change was detected. Also, with every available observation, the decision window and threshold can be updated to consider temporally local data variations. Previous studies have used a threshold optimization step based on knowledge from future observations that involve change [14], [18], [50], which may not be amenable to detect change with every available observation without knowledge of the future change event.

The performance of the change detector depends on the estimation accuracy of the model parameters that are used to compute the dissimilarity measure in equation (4.6). As the PF estimation accuracy is higher than that of the EKF, the detection performance of equation (4.7) increases when a PF is used to estimate the parameters of the nonlinear land cover model.

4.5.6 Cost-Sensitive Evaluation Metrics For Change Detection

Disturbances caused by forest fires are associated with changes in the local ecosystem, resulting in a negative environmental impact. Detection of such changes with minimum delay is therefore important, and we expect the cost of missing a change event (false negative (FN) error) to be higher than the cost of incorrectly detecting change (false positive (FP) error) [58]. We evaluate the performance of the proposed change detection algorithm using the following cost-sensitive metrics [59, 60]. The detection delay metric is the number of time steps required to detect change after the actual change has occurred. Two other metrics are the true positive rate, TPR=TP/(TP+FN) (or sensitivity (S)), and the true negative rate, TNR=TN/(TN+FP), where TP and TN correspond to the number of correct change and correct no-change detections, respectively. Related to these two metrics are the cost sensitive metrics geometric mean, GM=(TNR)(TPR)$^{1/2}$ [61] and the measure $F_\beta = (\beta^2+1)\mathrm{P}\,\mathrm{S}/(\beta^2\mathrm{P}+\mathrm{S})$

(detection precision, P=TP/(TP+FP)) [62, 63]. Higher misclassification cost is assigned to FN errors by selecting the weight parameter $\beta > 1$.

4.5.7   Fire Severity Mapping Methodology

We also formulate an unsupervised method for fire severity mapping that automatically annotates regions with burn severity labels. The Normalized Burn Ratio, computed from high spatial resolution data, is often used to estimate burn severity for post fire land management [48]. However, as such datasets have lower temporal coverage, it may not be acquired during a forest fire and thus can be unsuitable for early fire detection. The proposed approach utilizes Moderate Resolution Imaging Spectroradiometer (MODIS) observations to estimate burn severity, providing preliminary mapping from rapidly acquired data. Change statistics from NDVI can be used to estimate fire severity as it has high correlation with biomass loss [64, 65]. The PF estimates result in minimum detection delay and are better suited for near-real time mapping. We thus consider the improved PF parameter estimates to obtain preliminary burn severity maps within the fire perimeter. In particular, we assume that a forest fire is detected using equation (4.7) at time step k_0 and monitored until fire containment at time step K_T. The overall dissimilarity measure during this disturbance period is given by $D_{\text{CH}} = D_r^{p,q} = \sum_{k=k_0}^{K_T} d_{k,r}^{p,q}$. The measure of all pixels in the region are clustered using *k-means* with three clusters, with the initial seeds selected using *k-means++* [66]. The cluster means are ranked in descending order to assign cluster labels; the cluster with the largest mean is labeled *high severity* whereas the one with the smallest mean is labeled *low severity*. The labels are then used to evaluate the accuracy of fire severity classification over the region.

4.5.8   Change Detection From Estimated Parameter Time-Series

We consider the PF and EKF methods to estimate the CF and TVF model parameters for use in computing the dissimilarity and detection delay metrics. The metrics provide a way to evaluate the change detector performance that depends on the estimation method accuracy. The dissimilarity measure is expected to be low before land cover change and then to increase following the change. For all the datasets, we compare the dissimilarity, during the year of change $D_{\text{CH}}^{p,q}$, after change $D_{\text{ACH}}^{p,q}$ and during gradual regrowth D_{GRG}.

Dissimilariy Metric Comparison: Table 4.3, shows the dissimilarity result before and after change. We observe that using the CF model, the PF shows maximum D_{CH} and D_{ACH} indicating better estimates of change dynamics than the EKF. The dissimilarity of the land cover classes over the years of no change, D_{BCH} between consecutive years before change is observed to be low and comparable for PF and EKF. We also observe that *while D_{BCH} is lower or comparable for the PF-TVF than the PF-CF, D_{CH}, D_{ACH} and D_{GRG} is higher for the PF-TVF than the PF-CF. This indicates that the TVF and CF estimates agree during the no-change period, but the TVF model shows higher dissimilarity during and after change. This is due to the additional time-varying frequency parameter of the TVF model, which shows a sharp deviation due to the change. Overall, this implies that the TVF model better captures the state parameter variations due to the forest fire.* Additionally, using the same CF model, the nonlinear estimation by the PF captures this variation better than EKF, where the parameter estimation suffers due to linearization.

Change Detection Performance: For both the PF and EKF, we computed the change detection delay using equation (4.7) and evaluated the detectors using the metrics described in Section 4.5.6. The aim is to have minimum detection delay while limiting

64

the FP errors within an acceptable bound. From Table 4.4, we observe that for all change datasets (CR1-CR3), the detection delay is minimum using PF-CF estimates. *While the PF-TVF has comparable detection delay as the PF-CF, the PF-TVF has lower FP errors.* This indicates that overall, the PF is more sensitive to NDVI dynamics than the EKF and more accurately detects the time of change. The PF performs better than the EKF with respect to the change as shown by the TPR. This is achieved without increasing the FP errors considerably, which is reflected in the F_β and GM. Hence, the PF estimation is identified as a better change detector than the EKF. Additionally, the proposed method detects change by selecting a threshold based on overall regional variations in land cover. This is shown to be effective in detecting land cover change across multiple regions, which indicates that the method is independent of geographic variability. Window size for each region is chosen to be the length at which both the FP and FN errors are minimum. For evaluation, we set $\beta = 2$ to assign higher cost to FN errors. Change caused by the Zaca Fire has a high detection delay even with the PF because of a strong unaccounted degradation trend in land cover that preceded the fire. Thus, disturbances before a change can affect detection delay. The dissimilarity metric used was D_{MD} as the squared measure of D_{ED} is not well-matched to noisy observations.

As PF-CF detects the considered land cover change events with minimum delay, it is used to map the burn severity of the region using the approach described in Section 4.5.7. This achieves an accuracy of 92.31% in CR-1, 94.73% in CR-2 and 100% in CR-3. High accuracy in CR-3 can be attributed to higher uniformity in burn severity within the fire perimeter. Validation of the clusters with Landsat maps is challenging in regions where fire severity is spatially heterogeneous due to the difference in resolution between MODIS and Landsat. For such regions, we considered the most prevalent class in the Landsat burn area map to be ground truth.

4.5.9 Proposed Change Detection Framework
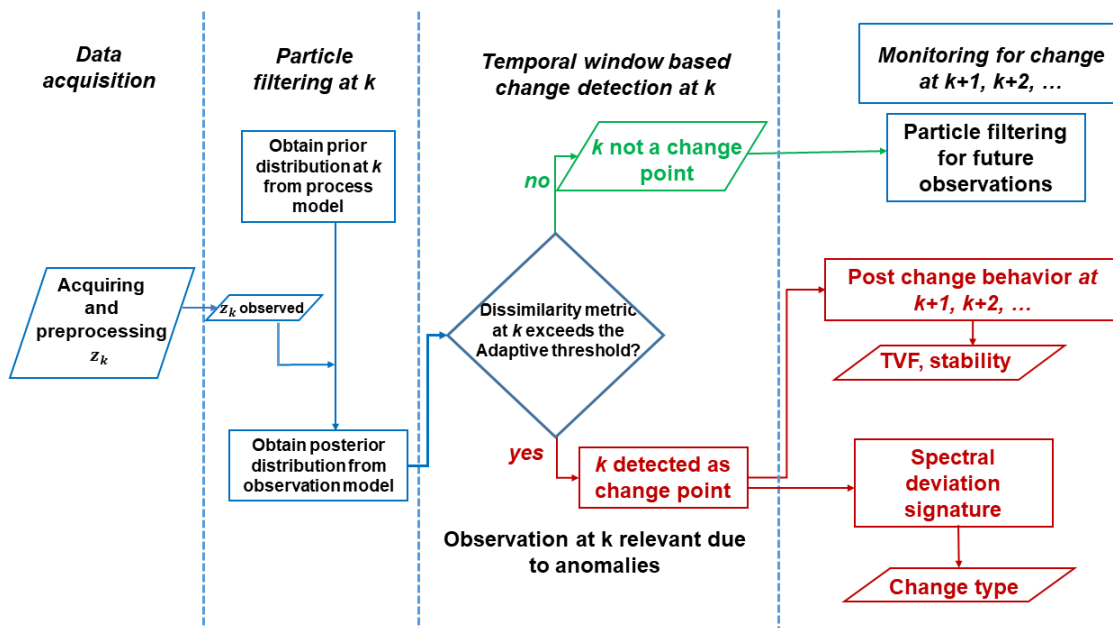
Figure 4.10: Proposed Land Surface Reflectance Time-Series Monitoring for Change Detection and Post-Change Analysis.

The sequentially estimated model parameters are utilized in developing a change detection framework that can be described using Figure 4.10. Here, the model for every region is monitored using CF or TVF models or based on best matched model selected from multiple models. Using the proposed model and the dissimilarity measures, the binary hypothesis land cover change detector is leveraged for sequential decision making. When an observation is detected as a change point, further analysis such as change event identification and post change monitoring is performed. Additionally, by automatically flagging change points, it also directs experts to observations that are most anomalous and rare and this reduced dataset volume (which is a fraction of the incoming data volume) is feasible for experts to analyze.

Table 4.3: Dissimilarity for PF-CF, EKF-CF and PF-TVF: Before, During, After Change, and Due to Gradual Regrowth.

Land Cover	Metric	Method	D_{BCH}	D_{CH}	D_{ACH}	D_{GRG}
CR1	D_{ED}	PF-CF	0.3271	1.3405	1.9019	2.0730
		EKF-CF	0.2856	1.3164	1.8029	2.1383
		PF-TVF	0.9858	1.9528	3.1652	3.0401
	D_{MD}	PF-CF	0.3941	1.7802	2.4556	2.6980
		EKF-CF	0.3370	1.7422	2.3221	2.8371
		PF-TVF	0.4069	3.0313	6.3737	6.1372
	D_{max}	PF-CF	1.0782	5.3691	6.6427	8.5613
		EKF-CF	1.3259	4.8701	5.5945	8.2168
		PF-TVF	2.8577	4.2921	6.2851	11.0191
CR2	D_{ED}	PF-CF	4.5210	9.5547	10.0718	9.6233
		EKF-CF	4.0742	8.4838	9.5083	9.0148
		PF-TVF	2.454	9.2983	11.9867	11.2821
	D_{MD}	PF-CF	5.1879	12.0484	13.0704	12.6582
		EKF-CF	4.5773	10.3555	12.5600	11.786
		PF-TVF	3.5855	14.5094	19.0234	18.8913
	D_{max}	PF-CF	2.0673	4.3691	3.9692	4.0669
		EKF-CF	2.7544	3.7469	3.7729	3.3803
		PF-TVF	1.4951	4.1993	4.0532	4.3752
CR3	D_{ED}	PF-CF	1.1423	7.7829	9.2174	6.0630
		EKF-CF	1.1967	7.4495	8.8881	6.2481
		PF-TVF	1.0665	10.3801	22.5949	21.5359
	D_{MD}	PF-CF	1.4471	9.9621	11.8053	7.4169
		EKF-CF	1.4930	9.2027	11.6256	7.7926
		PF-TVF	1.6455	16.8382	33.4065	30.6727
	D_{max}	PF-CF	0.2339	1.8849	1.3116	0.9791
		EKF-CF	0.2185	1.7589	1.2851	1.0031
		PF-TVF	0.2341	1.7759	2.3739	2.2546

Table 4.4: Cost-Sensitive Evaluation Metrics for Change Detection Using PF-CF, EKF-CF and PF-TVF.

Land Cover	Method	Detection Delay	TPR	TNR	GM	F_β
CR1	PF-CF	0	1	0.9647	0.9822	0.9211
	EKF-CF	1	0.9643	0.9706	0.9674	0.906
	PF-TVF	1	0.9643	0.9735	0.9689	0.9122
CR2	PF-CF	1	0.95	0.9728	0.9036	0.9613
	EKF-CF	3	0.9	0.9763	0.9374	0.871
	PF-TVF	2	0.9333	0.9763	0.9546	0.8974
CR3	PF-CF	4	0.8261	0.9702	0.8121	0.8910
	EKF-CF	6	0.7166	0.9848	0.7420	0.8400
	PF-TVF	4	0.8261	0.9814	0.9004	0.8333

Chapter 5

INSTANTANEOUS FREQUENCY VARIATION TO MODEL LAND COVER
DYNAMICS

5.1   Modeling Land Cover Changes

Multiple approaches have been considered to develop models that match land cover
dynamics using satellite image time-series [14, 15, 20, 21, 33, 34, 56, 67]. The constant
frequency (CF) model has been used to detect land cover changes [15, 56]. However,
as it is a Fourier-based model, it is matched to annual and intra-annual growth cycle
trends that remain constant with time. Variation in the CF model time-dependent
parameters suggests changes in the growth cycle; however, the model's assumption of
periodic cycles limits it use from capturing non-seasonal changes. In addition to the
CF part to represent annual vegetation growth, the authors in [16] proposed a model
with a nonlinear cosine time-varying phase function. This model allows for a peri-
odic time-frequency variation in the time-series that represents various intra-annual
cycles of different land cover types. It was specifically shown to match phenological
attributes of rice crops. Our recently proposed model, as discussed in Chapter 4, is
designed to match vegetation growth rate or frequency variations with time [20, 21].
This time-varying frequency (TVF) model also depends on various time-dependent
parameters that match harmonic characteristics of the time-series. These parameters
are used to represent average land cover state, maximum variation about the mean,
and time required to reach its annual peak. Large deviations in these parameters
is used to detect change. Our improved parameter estimation performance, when
compared to that of the CF model, has resulted in increased dissimilarity and hence

69

class separation between pre and post change classes and improved detecting land cover change with comparable delay and reduced false positives.

We propose a new approach to model land cover dynamics from satellite image time-series. The new model allows for variations in instantaneous frequency (IF) of vegetation growth with respect to time. Detection of land cover change is thus no longer dependent on large variations of harmonic parameters but on changes in the instantaneous frequency variation within the same datasets.

5.1.1 Instantaneous Frequency Variation Model

The IF variation model represents time-series data as a linear combination of M component states, each with a different instantaneous phase function. The number of component states correspond to the number of possible land cover state changes within the same dataset. For example, the number of component states is $M = 3$ for time-series data representing three different vegetation growth states before, during and after a forest fire. The states are periodic annual growth, sudden vegetation loss caused by the fire, and slow vegetation growth spurt.

The IF variation model is developed as follows. We consider the time-series row vector $\mathbf{z} = [z_1 \ z_2 \ \ldots \ z_K]$, where the number of data samples $K = (T_d/T_s)$ depends on the sampling period T_s and the duration T_d of the times-series. We assume that the time-series includes M component states, and that the duration of each state is T_m, such that $\sum_{m=1}^{M} T_m = T_d$; the number of samples of the mth component is $K_m = (T_m/T_s)$, $m = 1, \ldots, M$. The time index sequence of all the samples in the time-series consisting of M stages is given by

$$\{1, 2, \ldots, K\} = \{1, 2, \ldots, K_1, K_1 + 1, \ldots, K_2, K_2 + 1, \ldots, K_{M-1}, K_{M-1} + 1, \ldots, K\}.$$

If a given model m describes the land cover behavior during $K_{m-1} + 1, \ldots, K_m$, then

70

Table 5.1: Instantaneous Frequency Variation Model of Land Cover Changes

Rate of change λ_m	Instantaneous phase function $\lambda_m \, \xi_m[k]$	Instantaneous frequency $\lambda_m \, \nu_m[k]$	Time-frequency variation of land cover state
I) f_0 Hz	$f_0 \, T_s \, k$	f_0	constant
II) α (unitless)	$\alpha \cos\left(2\pi f_0 \, T_s \, k + \phi\right)$	$\alpha \, f_0 \, T_s \, \sin\left(2\pi f_0 \, k + \phi\right)$	periodic (crops)
III) b Hz^2	$b \, (T_s \, k)^2$	$2 \, b \, T_s \, k$	linear increase
IV) c (unitless)	$c \ln\left(T_s \, k\right)$	$c / (T_s \, k)$	hyperbolic decrease

the IF of a sample k in this stage is given by

$$z_{m,k} = A_{m,k} \, \cos\left(2\pi \lambda_m \, \xi_m[k]\right). \tag{5.1}$$

Here, $A_{m,k}$ is the amplitude time-variation, $\xi_m[k]$ is the instantaneous phase function, and λ_m is the rate of change of the mth component state at time k. Note that $K_0 = 0$ and $K_M = K$ in (5.1). The instantaneous frequency $\nu_m[k]$ corresponds to the rate of change with respect to time. For continuous-time functions, the instantaneous frequency $\nu_m(t) = \frac{d}{dt}\xi_m(t)$ is obtained as the derivative of the instantaneous phase function. Table 5.1 summarizes various instantaneous phase functions and the corresponding change in land cover state that they match.

5.1.2   Adaptive Model Selection

When a time-series under consideration consists of varying stages, the most appropriate model that best matches the instantaneous frequency of the given stage is to be selected causing the observation model to switch. For example, in the time-series described above, the instantaneous frequency governing the given land cover stage changes at $K_1 + 1$, $K_2 + 1$ and at $K_{M-1} + 1$. Therefore, the appropriate land cover model is to be selected at these time steps. Model switching is not only indicative

71

of a change point but also of a shift in the land cover state and its parameters. For example, a switch from the CF model with IF defined in I (described in Table 5.1) to the periodic crop model defined in II (described in Table 5.1) due to deforestation of forested land for agriculture allows a more accurate representation of the land cover, obtains improved estimates of the state vector and allows capturing parameters characteristic to that land cover stage.

Essentially, having multiple models allows the observation model to vary and select the function that best describes the current state. This section describes the proposed approach that modifies the PF estimation described before to select the most appropriate observation model and enable model switching. This is achieved by monitoring the particle filter likelihood pdf and exploiting its behavior between successive observations. When the PF estimates the state vector correctly, the particle weights w_k^i and hence the likelihood pdf remain high and similar between successive observations. However, if the same observation model as the pre-change steps are used even at and after a change point, the estimated state vector from the previous step $\mathbf{x}_k$ does not match the observation well after the predict and update steps. As a result, the particle weights or likelihoods are reduced drastically.

We monitor the likelihood pdf between successive steps which serves an indicator for model switching. In the absence of change, the likelihood pdf does not change significantly and its pdf at k w_k^i, is similar to its pdf w_{k-1}^i at $k-1$. On the other hand, when the k is a change point the likelihood probability density function between $k-1$ and k differ significantly when the pre-change stage observation model is retained. In the example described above, this is true at K_1+1, K_2+1 and at $K_{M-1}+1$ when the most appropriate model switches. We compute this shift using the Kullback Leibler

72

divergence (KLD) measure at k κ_k, given as

$$\kappa_k = \sum_{i=1}^{N_s} w_k^i (\log w_k^i - \log w_{k-1}^i), \qquad (5.2)$$

where the KLD is indicative of how dissimilar the two densities $p(\mathbf{x}_{k-1}|\mathbf{z}_{k-1})$ and $p(\mathbf{x}_k|\mathbf{z}_k)$ are. The proposed approach is designed to perform model switching at the change point and estimate the respective state vector sequentially. The appropriate model is selected at the change point by choosing the model whose instantaneous frequency best matches the variation rate observed in the data. This process can be described as follows:

1. As the stable stage is the most dominant land cover state, for the initial observations, the observation model h is assumed to be the constant frequency CF model, and the previously described particle filtering approach is used to estimate the posterior density $(\mathbf{x}_k^i, w_k^i)$ of the state vector, where $i = \{1, \ldots, N_s\}$ represents each particle.

2. For subsequent observations, the observation model from the selected (best) model M_{k-1} in the previous stage $k-1$ is used to estimate the posterior distribution of the current state $(\mathbf{x}_k^i, w_k^i)$ and the KLD is computed between particle weight distribution of consecutive steps (k and $k-1$) using equation 5.2. As the particle weights are nearly similar between successive observations in the absence of change, the KLD value is expected to be low in the near-constant stage. However, the posterior distribution and hence the weights are a poor approximation of the state when k is a change point, which causes the weights to be drastically different (lower likelihood) causing the KLD κ_k to increase.

When $\kappa_k - \kappa_{k-1} > \tau_\kappa$, model switching is performed by performing particle filtering from the estimates at $k-1$ using the alternate observation function M_k whose likelihood pdf is most similar to the likelihood pdf of the previous correctly estimated

Figure 5.1: Case 1: Simulated Signal Formed from Two Models. Model 1: Constant Instantaneous Frequency, Model 2: Sinusoidal Instantaneous Frequency.

step. Here τ_κ is a threshold determined empirically from the standard deviation of the KLD from past observations. The estimates from the alternate model is used for the prediction step at $k + 1$.

The model selected at k is used to obtain the state estimates using particle filtering and steps 2) and 3) are repeated sequentially for the length of the time-series.

5.1.3   Simulation Results

By designing appropriate phase functions for each stage, this approach is formulated to select the observation function that best describes the current observation and estimate the instantaneous frequency.

To evaluate the proposed approach, time-series with different models are simulated with $M = 2$. The model governing the time-series upto $k = 230$ is simulated using the CF model with IF I (described in Table 5.1) to form Model 1, while the subsequent stages are simulated using $z_k = \mu_k + A_k \cos(2\pi f_c T_s k + \alpha \cos(2\pi f_i T_s k + \phi)) + \eta_k$, with IF similar to II (described in Table 5.1) to form Model 2 (Case 1). This is shown in

74

Figure 5.2: a) Model Switching on Simulated Signal (Case 1) at 25 dB, b) Mean Squared Error Between Estimated and Simulated Signal, c)Mean Squared Error in Logarithmic Scale

Figure 5.1.The state vector to be estimated for Model 2 is $\mathbf{x}_k = [\mu_k, A_k, \phi, f_c, \alpha, f_i]$, where f_c and f_i are unknown constants. The model selection (between the estimated model and predicted model) is shown in Figure 5.2. Here, n_k was simulated such that the SNR of the simulated data is 25 dB. The model switching is observed to happen correctly at the change point which results in the low mean squared error even due to an abrupt change and does not fluctuate due to change as shown in Figure 5.2. Thus, the correct instantaneous phase is selected which can be used to represent the IF of crops. Moreover, this would also avoid the higher error in estimation using the single model approach immediately following change.

A second case, where the instantaneous phase functions of Model 1 and Model 2 are markedly different have been simulated (Case 2). Here Model 1 is the constant frequency model with IF I (described in Table 5.1), while Model 2 is defined as
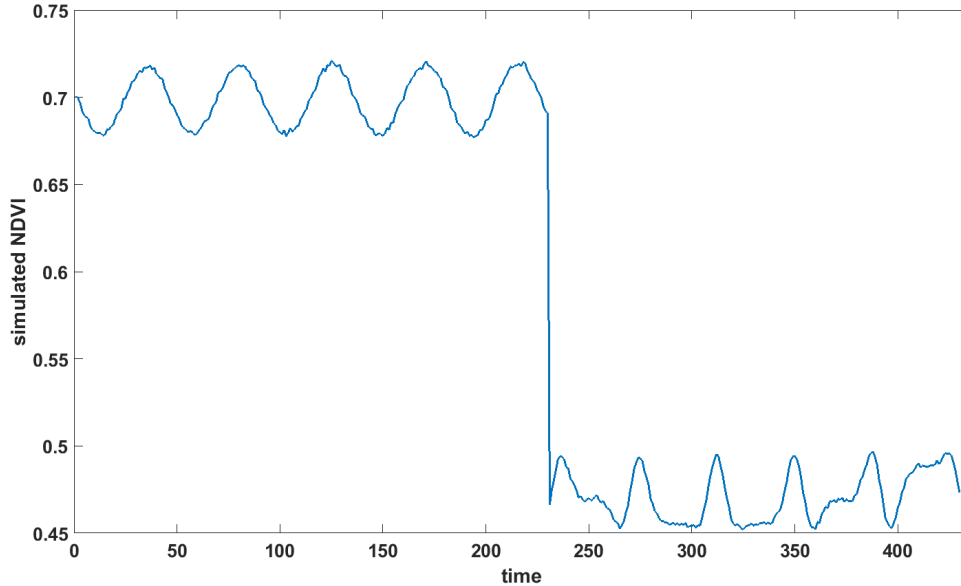
75
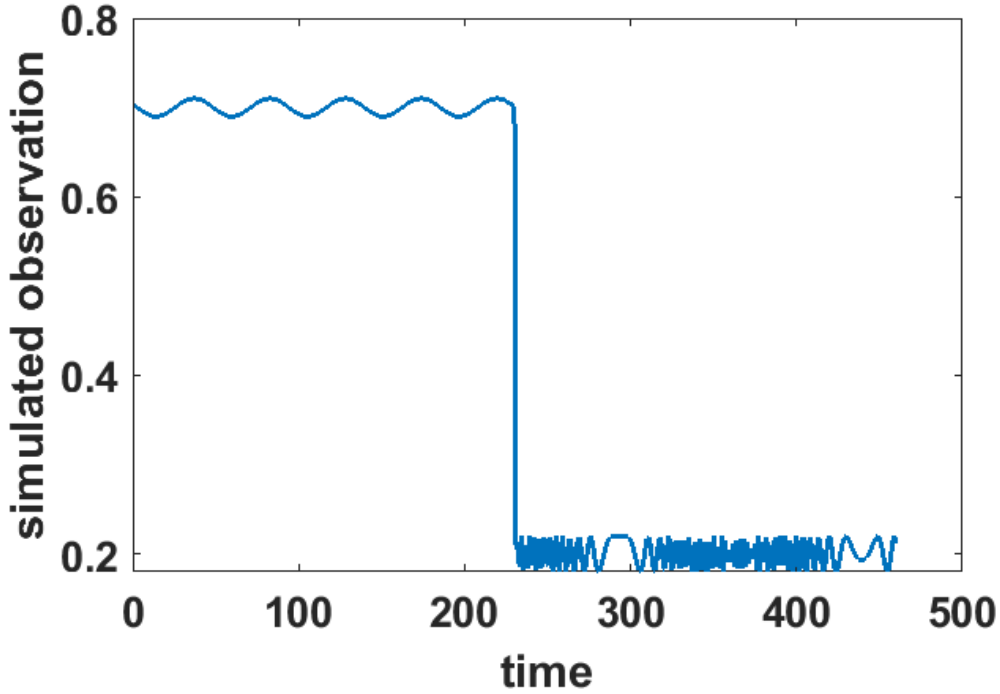
Figure 5.3: Case 2: Simulated Signal Formed from Two Models. Model 1: Constant Instantaneous Frequency, Model 2: Chirp Instantaneous Frequency.

$z_k = \mu_k + \alpha_k cos(2\pi b(T_s k)^2) + \eta_k$, whose IF matches III (described in Table 5.1) showing a linear increase. The state vector to be estimated for Model 2 is $\mathbf{x}_k = [\mu_k, \alpha_k, b]$, where b is an unknown constant. This is shown in Figure 5.3. The model selection (between the estimated model and predicted model) is shown in Figure 5.4. Here, n_k was simulated such that the SNR of the simulated data is 9 dB. The model switching is observed to happen with a delay at the change point which results in the relatively higher mean squared error immediately after change. Within a few observations the model switching occurs which then lowers the mean squared error as shown in Figure 5.4 and remains low subsequently. Thus, the correct instantaneous frequency model is selected and such variations in IF can be attributed to the regrowth phase with a gradual linear change in land cover frequency. On the other hand, the IF in IV can be related to the the sharp decrease in IF that can be expected immediately after forest fires.

76
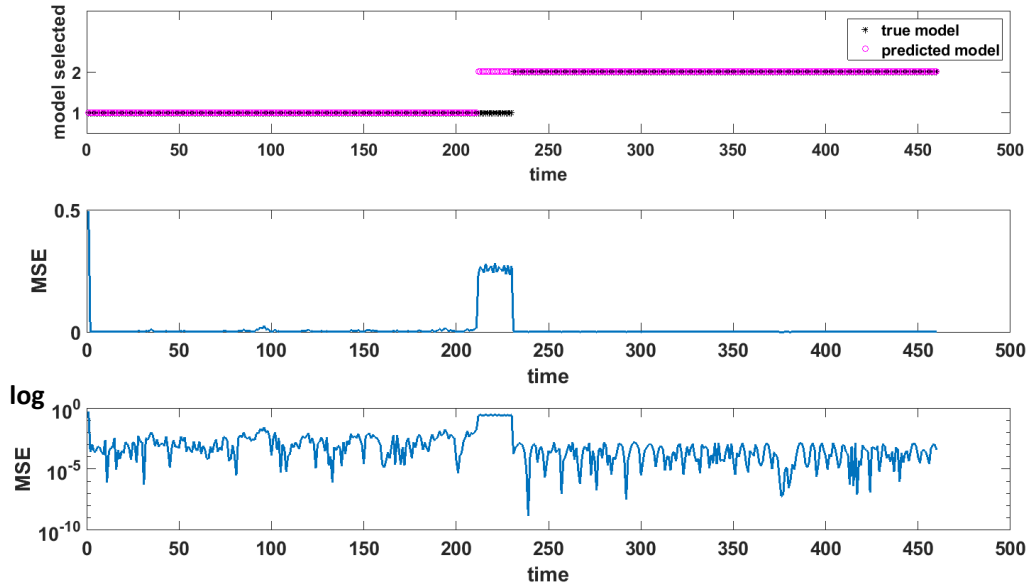
Figure 5.4: a) Model Switching on Simulated Signal (Case 2) at 9 dB, b) Mean Squared Error Between Estimated and Simulated Signal, c)Mean Squared Error in Logarithmic Scale

The delay in model switching can be explained due to the higher noise. In such cases the optimal τ_κ is critical. However, in this work, this threshold is determined empirically and is also a reason for the delay in model switching. However, in the presence of higher noise of SNR of 5 dB with a different η_k for the same data, the model switching and consequently the mean squared error is affected as shown in Figure 5.5 even after varying the KLD threshold. A very high threshold lowers the possibility of model switching even in the presence of change while a low threshold causes model switching even in the absence of change due to the high noise in the data.
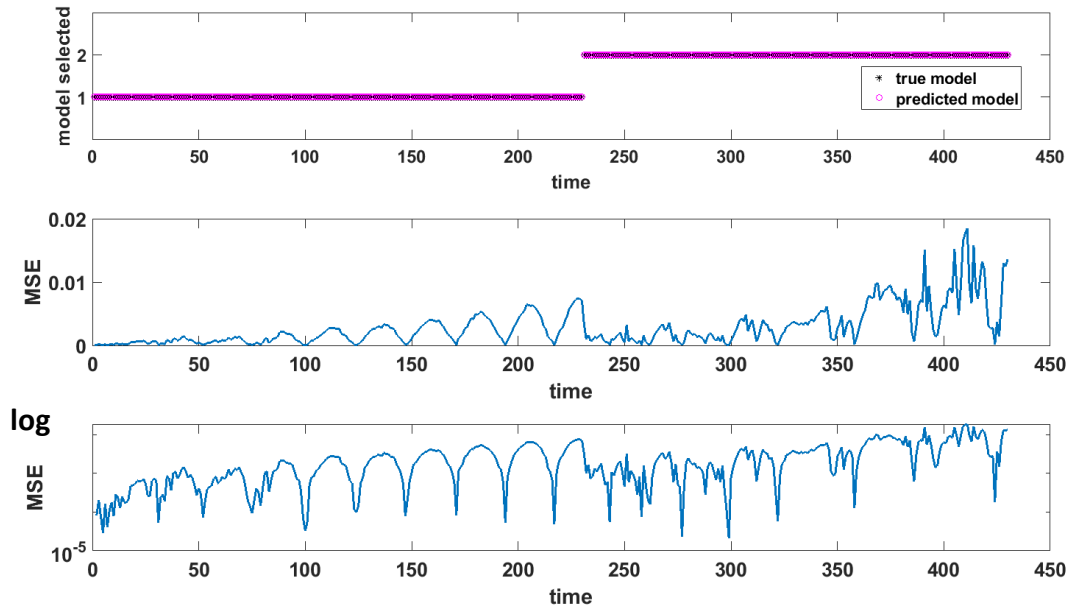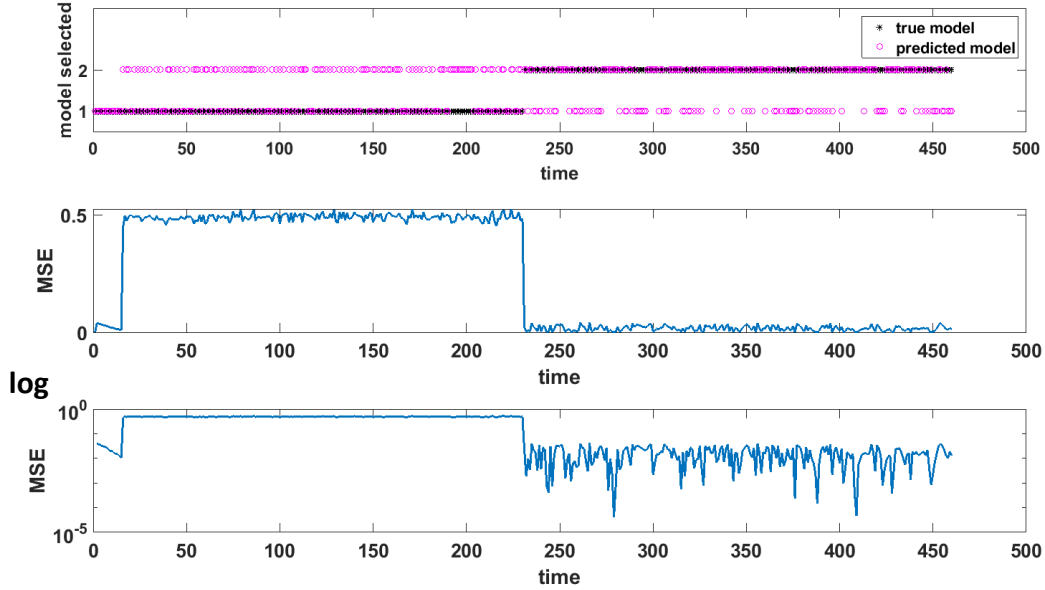
Figure 5.5: a) Model Switching on Simulated Signal (Case 2) at 5 dB, b) Mean Squared Error Between Estimated and Simulated Signal, c)Mean Squared Error in Logarithmic Scale

5.1.4   Real Data Analysis

In addition, this approach was also extended to real observations from CR-1. Model selection using KLD was tested with Model 1 as the CF model with IF defined in I (described in Table 5.1) and Model 2 having similar IF as defined by II (described in Table 5.1). The model switching results are shown in Figure 5.6, where the KLD threshold is also determined empirically and allows immediate model switching at the change point. However, the alternate model Model 2 can be improved to accurately model the post-fire behavior, which would allow the alternate model Model 2 to be selected in subsequent observations as well. The current alternate model defines the intra-annual behavior of crops and is hence a poor approximation of post-fire behavior.

Overall, the proposed approach is capable of performing model switching based on
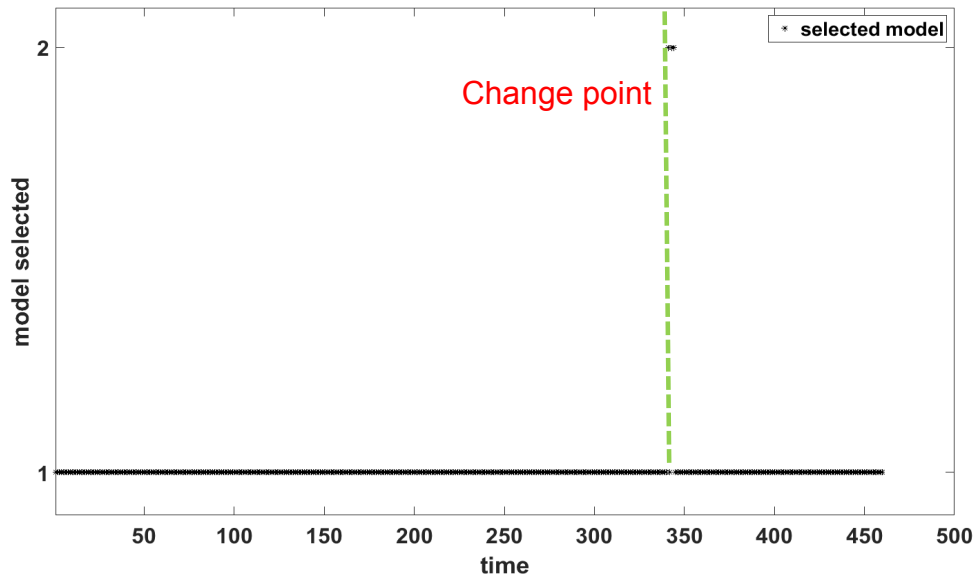
Figure 5.6: a) Model Switching on Real Forest Fire Time-Series with Constant Frequency (Model 1) and Sinusoidal Frequency (Model 2) Models.

KLD threshold to estimate instantaneous frequency of land surface time-series during change and post change periods. However, improved alternate models that govern change events more closely and optimal threshold selection are essential factors that would further improve the method.

Chapter 6

MULTISPECTRAL ANALYSIS OF LAND COVER TIME-SERIES FOR

CHANGE EVENT SEPARABILITY

Land cover time-series modeling and change detection is primarily addressed
by utilizing vegetation indices such as the Normalized Difference Vegetation In-
dex (NDVI) or Enhanced Vegetation Index (EVI). These indices relate to the vegeta-
tion condition or greenness of the underlying land cover and can capture degenerative
and regrowth trends in the vegetation of the area. Thus, land cover band ratio time-
series that are tailored towards highlighting vegetation condition of the underlying
land cover are well suited for detecting changes, both positive and negative trends.
However, the underlying signature of the change event is not captured by vegetation
indices, as they are only indicative of change. As a result, change detection approaches
based on NDVI or EVI cannot distinguish one change event from another. Change
signature of a land cover change event, is an information that is captured by the
spectral or band deviations at the change point. Therefore, analyzing the band devia-
tion combination at the change point reveals the spectral variation that is unique to
the change. It is also expected that similar change events should give rise to similar
change signatures and hence it should be possible to group land cover time-series
changes based on the type of change event. However, this maybe limited by pre-
change conditions due to geographic variability. This chapter introduces techniques
to model the expected spectral features of given region and examine the deviation at
change point and the feasibility of grouping events with similar change signature.

6.1   Modeling Multispectral Land Cover Time-series

Any changes in land cover over time causes it to transition from one stage to another. As the spectra of a given region captured by the satellites measures the energy absorbed by the surface due to its interactions with the incident light of a given wavelength, monitoring the spectra over time, sheds light on *how the surface has changed compositionally.* These change events however, are anomalous and temporally rare and the expected state of land cover is usually stable or unchanged. This behavior of multispectral land cover reflectance time-series is exploited to learn a model of the expected spectral characteristics in the absence of change, that govern most observations. The model describing the expected regional land cover spectra is then used to determine whether an incoming observation meets this expected multispectral characteristics or not. Observations that violate the expected trend are anomalous. This happens when the given observation is an out-of-distribution sample compared to the samples from which the model has been learned. The spectral deviations between the expected spectra and the true change point spectra are then analyzed to determine the feasibility of class separation using the spectral signatures. It is assumed that knowledge of all possible change events are not known a priori. This assumption is motivated by two scenarios - a) the knowledge of all possible change events may not be available to us a priori, and b) the global geographical variability of land cover might result in different change event signatures for the same change event (for example, flood signature in a coastal marsh may have a different spectral behavior than a flooding event in an urban area). Hence an unsupervised approach is adopted that can group unseen change events without assuming any prior knowledge about the kinds of change events.
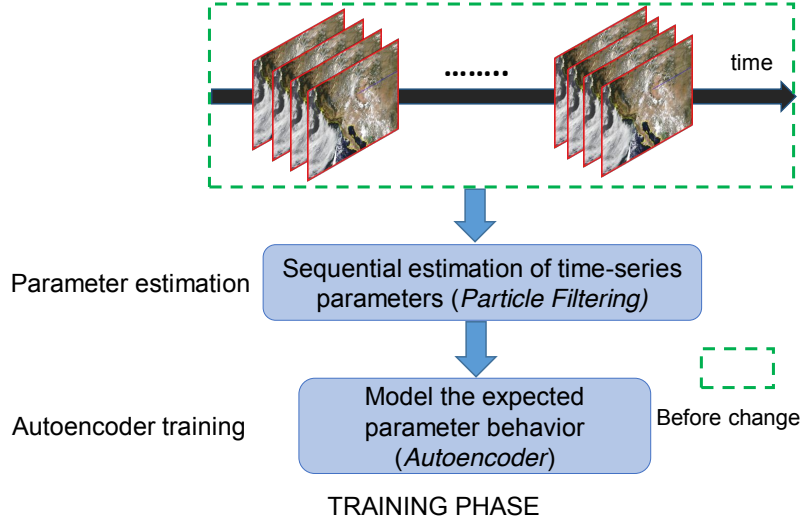
Figure 6.1: Learning Pre-Change Stable Phase Spectral Representation of Particle Filter Estimates Using an Autoencoder for a Given Region

At the change point, the deviation in the reflectance time-series across all bands captures the signature associated with the given type of change as each spectral band responds uniquely according to the nature of change. This signature is a difference between the change point spectra with that of the spectra prior to change to highlight the change in the reflectance (and hence the composition) time-series due to the change event. The expected spectral characteristics in the absence of change and other disturbances are modeled using a latent space representation of the estimated parameters $\hat{\mathbf{x}}_{k_\lambda}$ across multiple bands $\lambda = \{1, 2, \ldots, \Lambda\}$ prior to change. Latent space representations of the estimated parameter time-series across bands model the spectral characteristics of the land cover in its stable state. The model is formed by training on the parameter time-series of a set of observations during the stable stage for the given region. This training data $\mathcal{X} = \{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{tr}\}$ consists of tr samples or consecutive observations in time (during a stable period of no change) where each $\hat{X}_k$ consists of the state vector parameters (or its subset) $\hat{X}_k = \{\hat{\mathbf{x}}_{k,\lambda 1}, \ldots, \hat{\mathbf{x}}_{k,\lambda \Lambda}\}$

estimated by the particle filter across all Λ bands.

The autoencoder approach is adopted to learn the latent space representation from $\mathcal{X}$. Autoencoders are neural network architectures where the input is transformed to produce the output with minimum distortion. Here the input $\hat{X} \in \mathcal{R}^n$, is transformed to produce the output $\tilde{X} = \mathbb{D}(\mathbb{E}(\hat{X})) \in \mathcal{R}^n$, such that $\mathbb{E}$ is a class of functions forming the encoder that maps the input, which is the particle filter estimate of the state vector across all bands to its encoded representation $\mathbb{E}(\hat{X}) \in \mathcal{R}^m$, and $\mathbb{D}$ is a class of functions forming the decoder that maps the encoded input to its output representation $\mathbb{D}(\mathbb{E}(\hat{X})) \in \mathcal{R}^n$, where $m<n$. The corresponding output $\tilde{X}_k = \{\tilde{\mathbf{x}}_{k,\lambda 1}, \ldots, \tilde{\mathbf{x}}_{k,\lambda\Lambda}\}$, is essentially a reconstruction of the input (where $\tilde{(.)}$ represents a parameter reconstructed by the autoencoder). Thus, this form of the autoencoder (undercomplete) performs a compression, where it learns the most salient characteristics of the input data in the latent space and reconstructs it. The encoder and decoder models for a given region are built from a training set $\mathcal{X}$ during a stable period learned by minimizing the model prediction loss L

$$\min L(\mathbb{E}, \mathbb{D}) = \min_{\mathbb{E},\mathbb{D}} \sum_{\hat{X}_k \in \mathcal{X}} ||\hat{X}_k - \tilde{X}_k||, \tag{6.1}$$

over all training samples from that region as shown in Figure 6.1.

The autoencoder is therefore trained to reconstruct multispectral land cover parameters from the stable period and shows very low error in reconstruction. In the absence of change, the state vector distribution across all bands nearly remains the same, which the functions $\mathbb{E}$ and $\mathbb{D}$ learn to reconstruct. For incoming observations in the test phase, the magnitude of reconstruction error is directly related to the anomaly score $\mathbb{A}_k$ of that observation given as

$$\mathbb{A}_k = \sum_{\lambda=1}^{\Lambda} | \tilde{X}_{\lambda,k} - \hat{X}_{\lambda,k} | . \tag{6.2}$$

83
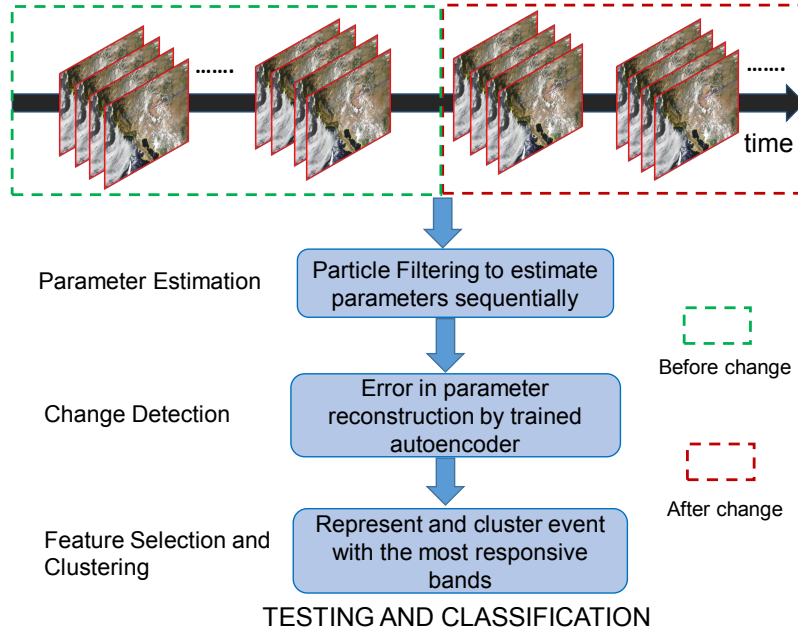
Figure 6.2: Test Phase to Monitor Land Cover Using Trained Autoencoder to Detect and Group Land Cover Change Events Based on Spectral Deviation

As change ensues, the parameters deviate from the expected distribution across all spectral channels. The trained autoencoder weights now fail at reconstructing the input as the trained model now poorly approximates the anomalous state parameters of the changed land cover class, resulting in high reconstruction errors and anomaly score. Here the state vector distributions varies (significantly or slightly depending on the change severity) from that of the stable phase, which the $\mathbb{E}$ and $\mathbb{D}$ functions reconstruct with very high error. Such a deviation in reconstruction error is then used as a detector of change points. This is shown in Figure 6.2. As an online change detection approach is adopted, the threshold for the change point is determined from the standard deviation in reconstruction error from the past data. When the reconstruction error for a given observation exceeds this threshold learned from the past data in the region, the observation is determined as anomalous. However, this threshold is multiplied with an additional factor that further increases or decreases the threshold. For example, a factor < 1 reduces the threshold and is used for detecting gradual changes

such as coastal land gain, droughts, while the factor > 1 increases the threshold and is used to detect more sharp and abrupt changes such as forest fires and floods. In addition, the threshold also controls the false positive and false negatives in the data.

Moreover, at the change point, $k = k_0$, the estimated parameters $\tilde{X}_{k_0}$ deviate from the representation over certain change-specific spectral channels. Each change event is associated with a deviation in the reflectance time-series in specific spectral bands that responds strongly to a given type of change event. This is because from the change point onward, the reflectance time-series exhibits spectral properties with respect to the new land cover class. As a result, it exhibits deviations in specific bands which distinguishes the new land cover class from the old. For changes due to natural events that has been considered in this work, all bands are observed to respond and deviate due to change, however, the magnitude of response varies across bands. It is hypothesized that capturing this trend in the multispectral reflectance time-series should enable change event separability. As the spectral properties of the post change class is dependent on the type of change event, analyzing the Λ dimensional spectral deviation vector Δ_{k_0} given by

$$\Delta_{k_0} = [\tilde{X}_{\lambda 1, k_0} - \hat{X}_{\lambda 1, k_0}, \ldots, \tilde{X}_{\lambda \Lambda, k_0} - \hat{X}_{\lambda \Lambda, k_0}], \tag{6.3}$$

of the multispectral time-series at k_0 can enable distinguishing one type of change event from the other. By inspecting the bands that respond to change and the nature of change in its trajectory, the type of land cover change can be identified. The bands that deviate strongly due to a given event are referred to as the most responsive bands with respect to that event. The onset of this deviation is at k_0 and lasts over the duration $\{k_0, \ldots, k_T\}$ of change. For events such as wildfire which leaves a greater impact on the ecosystem, this deviation lasts for a longer period before the regrowth phase sets in, while for short-duration events, this deviation is observed for a shorter

interval.

6.1.2 Interpretability Of Change Event Signatures

The spectral deviation vector Δ consists of information relating to both the magnitude or change severity and the direction indicative of positive or negative response due to change at each band. In this work, an approach is proposed that analyzes the spectral deviation vector to interpret the autoencoder detections of change. Here, interpretations refer to the explainability of autoencoder decisions in human or expert understandable form. These explanations offer insight into *why an observation was detected as anomalous* as well as examines the band-wise deviation to indicate *what features or bands resulted in the observation being detected as anomalous or a change point*. While the former can be directly addressed by noting the anomaly score of the observation, the latter can be assessed by noting the direction of spectral deviation as well as the bands that are most anomalous and are maximally responsible for the anomaly score of the observation.

The autoencoder reconstruction facilitates the interpretation of the change signatures. For an anomalous observation at k, the reconstruction by the model at k, $\tilde{\mathbf{x}}_{k,\lambda}$, is compared to the input parameter estimates (from the particle filter) at k given to the model, $\hat{\mathbf{x}}_{k,\lambda}$ for each band. *Here, the interpretability is maximal when both $\tilde{\mathbf{x}}_{k,\lambda}$ and $\hat{\mathbf{x}}_{k,\lambda}$ includes the estimated and reconstructed mean only.* This can be attributed to the fact that the mean captures the average land surface state and is therefore the most accurate representation of change response. The amplitude on the other hand represents the variation about the mean and does not carry information as to whether the surface reflectance increases or decreases. The change signature is then extracted using equation (6.3), where the input mean to the autoencoder across all bands $\hat{\mathbf{x}}_k$, is the true parameter estimate for that band, and the reconstructed spectral mean

$\tilde{\mathbf{x}}_k$ represents the model based reconstruction of the input using the weights learned from the stable or no change phase and hence represents a baseline. In addition, the Δ_k across each band can be interpreted in the following way:

1. if $\tilde{\mathbf{x}}_{k,\lambda} - \hat{\mathbf{x}}_{k,\lambda} > 0$

   then the corresponding change event caused the surface reflectance in λ to decrease.

2. if $\tilde{\mathbf{x}}_{k,\lambda} - \hat{\mathbf{x}}_{k,\lambda} < 0$

   then the corresponding change event caused the surface reflectance in λ to increase.
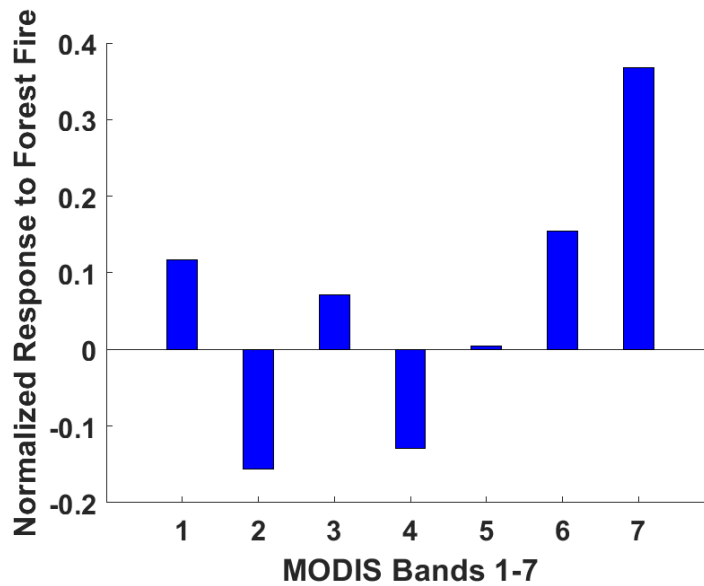
Figure 6.3: Normalized Spectral Response Across All MODIS Land Bands for Forest Fire (Horseshoe2 Fire, Southwest United States, CR-2).

Change signatures interpreted this way adds to the interpretability of autoencoder detections and is therefore a way to explain these decisions in a human interpretable form (based an spectral analysis of different compositions) [68]. The Δ_k for a given
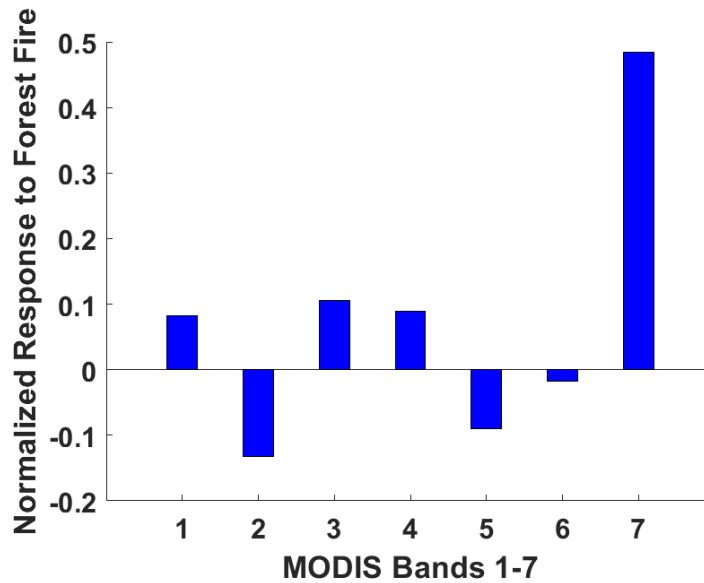
Figure 6.4: Normalized Spectral Response Across All MODIS Land Bands for Forest Fire (Wallow Fire, Southwest United States, CR-1).

region is calculated and summed across pixels from that region with a high anomaly score that are detected as change points using equation 6.2. The summing effect ensures that bands for which the reconstruction error is not consistent (same event exhibits both positive and negative deviations across the same band) across different pixels are given low priority in expressing the anomaly score. The spectral deviation due to forest fires are shown in Figures 6.3 and 6.4 which is a representation of autoencoder reconstructions. Here 0 represents the baseline error when the model perfectly reconstructs the input. These representations correspond with known band ratios such as NDVI, with the increase in Red band (1) and decrease in Near-Infrared band (2) reflectance due to vegetation loss. Similarly, the high increase in the Shortwave Infrared band (7) can be attributed to the fire.

On the other hand, Figure 6.5 shows that the expected decrease in the bands due to vegetation loss (decrease in band 2)/ water content increase (decrease in Visible
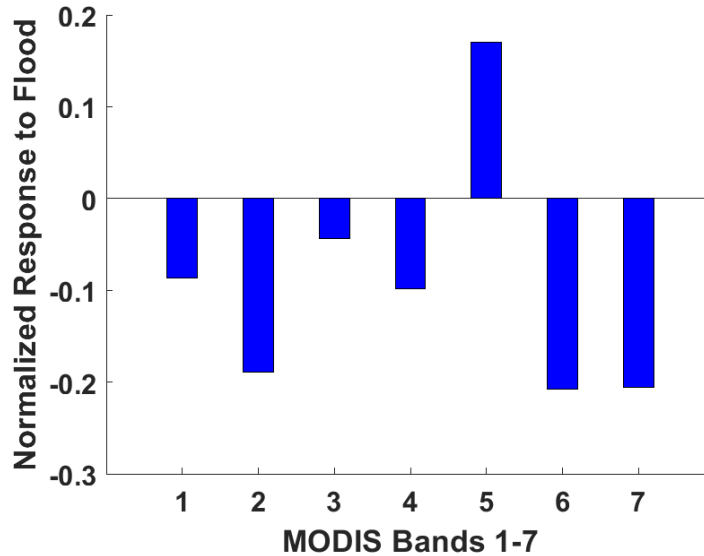
Figure 6.5: Normalized Spectral Response Across All MODIS Land Bands for Flood (Sava River Flooding, Croatia, CR-4).

bands 1, 3, 4 and Shortwave Infrared band 7 as water has high absorption in these wavelengths) due to flooding of an urban and agricultural region. This marked difference between signatures of different events is exploited to spectrally separate the change events. Moreover, such explanations or interpretations of change decisions demonstrates that anomaly scores can be expressed in a format that matches with domain knowledge. This is a key contribution compared to similar studies on multiple change detection in bitemporal images [69], [70], [71], [72], [73] that aim at ternary change detection.

However, as seen from the change signatures, the responses across multiple bands for similar events can often vary. To tackle this the top-r most responsive bands are selected for each event while the response across the remaining channels are set to zero and r is varied from $r = 1$ to $r = 7$. It is observed that the consistency (same direction of response) of the response across different channels is improved when $r = 2$. This

has also been reported in [72], [73], where increasing dimensionality of the data by considering all bands, increases complexity and noise in the feature space and can be less effective than single channel based change analysis. Here the original change response Δ_k is modified as $\Delta_{k,r}$ where the original reconstruction error is retained for the top r most responsive bands, while the the responses for the remaining channels are zero. In the subsequent steps all values of r are considered, however, the results discussed is for the case $r = 2$.

6.1.3 Spectral Separability Of Change Events

The separability of spectral deviation vector at the change point is analyzed to group different change events. Ideally no a priori assumption about the spectral properties associated with a given change event should be made as that i) allows identification of new change events with unknown spectral characteristics and ii) does not depend on the availability of labeled training data. Therefore, an unsupervised approach (clustering) is adopted that aims to group change events based on its spectral deviation. To further avoid the process of clustering across all bands, the angle $\mathcal{A}_{k,r,i}$ between the responses of each event $\Delta_{k,r,i}$ is compared with a reference vector v_{ref} and is given as follows:

$$\mathcal{A}_{k,r,i} = arccos(\sum_{\lambda=1}^{\Lambda}(\Delta_{k,r,i}v_{\text{ref}})/\sqrt{\Delta_{k,r,i}^2 v_{\text{ref}}^2}), \tag{6.4}$$

where i indicates that this angle is computed between the reference and each pixel i in the region, thereby forming the final feature which is to be grouped.

Number of change events are not assumed to be known a priori. The assumption of the number of components is restrictive as it requires the knowledge of the type of possible events. Moreover, this can further restrict the classification step where a new or previously unknown change event is also grouped into one of the known categories

90

without identifying it as a distinct event and grouping it as a new cluster. A Gaussian mixture modeling (GMM) approach is adopted to group the change events where the number of distinct spectral responses is indicative of the number of components C. The change events are assumed to be mixtures of normal distributions with mean ζ_c, covariance Σ_c and mixing coefficients π_c [74]. Each spectral deviation vector at a change point is grouped into its change category using the expectation-maximization algorithm as follows:

i) Initialize the cluster means, covariances and mixing coefficients for each C component.

ii) Expectation Step: This step calculates the responsibilities $\gamma_{i,c}$, where $\gamma_{i,c}$ denotes the posterior probability that the spectral deviation of the ith sample matches the change signature of the event described by cluster c as

$$\gamma_{i,c} = \frac{\pi_c \mathcal{N}(\mathcal{A}_{k,r,i}|\zeta_c, \Sigma_c)}{\sum_{j=1}^{C} \pi_j \mathcal{N}(\mathcal{A}_{k,r,i}|\zeta_j, \Sigma_j)}. \tag{6.5}$$

iii) Maximization Step: This is followed by re-estimating the new parameters of the each change event cluster using the responsibilities calculated in step i) as

$$\zeta_c^{new} = \frac{1}{N_c} \sum_{i=1}^{N} \gamma_{i,c} \mathcal{A}_{k,r,i}, \tag{6.6}$$

$$\Sigma_c^{new} = \frac{1}{N_c} \sum_{i=1}^{N} \gamma_{i,c} (\mathcal{A}_{k,r,i} - \zeta_c^{new})(\mathcal{A}_{k,r,i} - \zeta_c^{new})^T, \tag{6.7}$$

$$\pi_c^{new} = \frac{N_c}{N}, \tag{6.8}$$

where $N_c = \sum_{i=1}^{N} \gamma_{i,c}$ and N_c shows the total number of samples assigned to the $c - th$ cluster.

iv) Repeat steps ii) and iii) till the convergence criterion (change event assignment of spectral deviation vectors does not vary above a threshold between successive steps) is met.

However, GMMs assume that C or number of distinct clusters are known a priori. To tackle this problem, the number of clusters is inferred from the mixture modeling approach by iterating over a range of values of C (clusters) and determining the appropriate number of clusters from the smallest cluster number that significantly reduces the class assignment error.

The events that are under consideration include two forest fires (CR-1, CR-2), three floods (agricultural fields CR-4, CR-5, coastal marsh CR-6) and one drought (CR-8) event using all MODIS land cover bands. In Figure 6.6 the smallest cluster
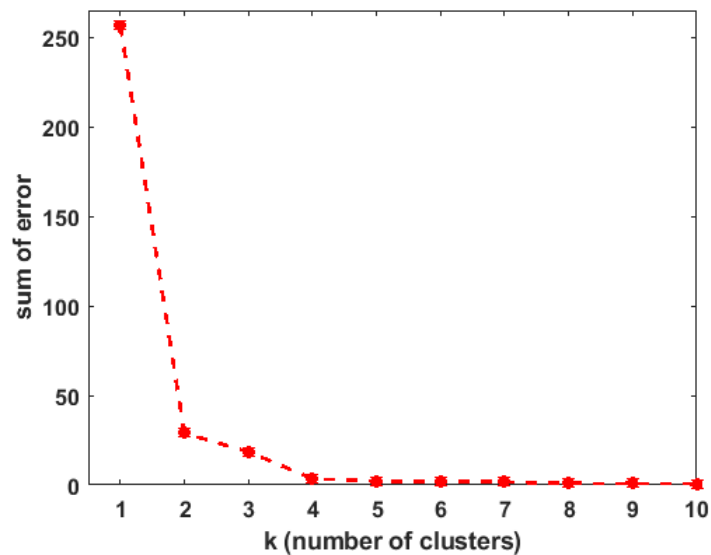
Figure 6.6: Sum of Error in Assignment Versus Number of Separable Change Events from Gaussian Mixture Modeling.

number when the cluster assignment error is lowered is when C changes from 3 to 4. Hence the number of clusters is 4. This gives a 98% accuracy in clustering.

However, in this case it is observed that the clusters are not grouped according to the type of event solely, but on the type of spectral deviation. Hence it can be concluded that the separability of land surface change events is feasible when the type of pre-change land cover class and the type of the change event are considered simultaneously to determine the spectral signature. This is also demonstrated by the polar coordinate representation of the $\mathcal{A}_{k,r,i}$ from the dataset under study as shown in Figure 6.7. The four classes identified from unique spectral angles are forest fire (both
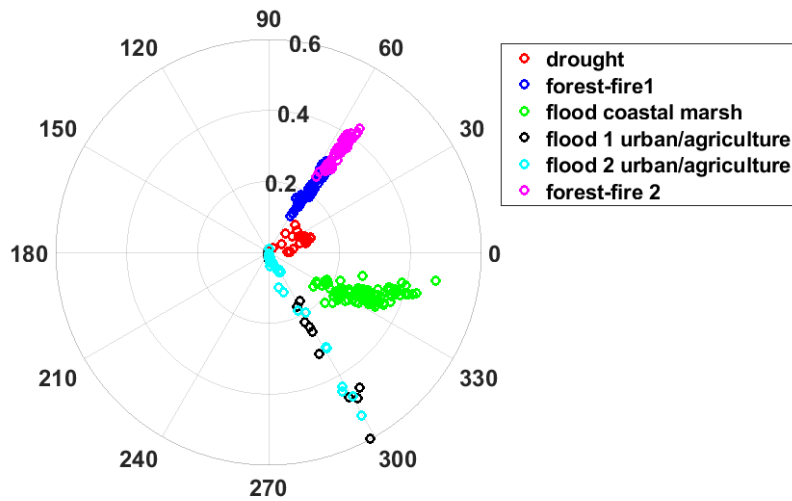
Figure 6.7: Polar Coordinate Representation of the Angle Between the Change Response Vector of Each Pixel with a Reference Vector.

examples from South Western United States), drought (Southwestern United States), flooding (example 1: Hurricane Katrina, coastal marsh, Louisiana), flooding (both examples: agricultural and urban area, Sava River, Croatia, Bosnia and Herzegovina). In addition, from the given dataset it can be concluded that the proposed separability approach is also scalable to different geographic regions. However, the performance of the approach can vary as the consistency in change signature and hence the spectral

angle varies (when coastal land gain signature from the Atchafalaya basin in Louisiana was considered). Overall, the proposed approach shows promise in identifying change event signatures as well in interpreting the change decisions from detections made from frequently acquired datasets.

Chapter 7

RETRIEVAL AND PRIORITIZATION OF NOVEL MULTISPECTRAL PLANETARY IMAGES

This chapter introduces a technique for context-specific analysis of planetary images by formulating a *representation* of geological landforms of interest. In addition to Earth observation, observations from remote sensing instruments also capture important planetary atmospheric and surfaces properties. However, the volume of the repository is so large that expert analysis of every incoming observation is infeasible. The objective of this work is to formulate methods that can analyze these observations with expert guidance by emulating expert-like decision making and hightlighting and prioritizing the subset of observations that are most interesting or scientifically relevant. As a result, the data volume to be analyzed by an expert is reduced, where only the most anomalous or relevant observations are to be examined, thereby making the analysis feasible.

These representations are important for retrieving instances that match a given criterion or are novel or unusual with respect to the expected trend from a the large repository. The representations are adaptive over over time (lifespan of an instrument) to newly discovered information from recently downlinked observations so that the updated representation is utilized for prioritization in the future. In order to support scientific discovery and improve the existing knowledge about a process of interest, emphasis is laid on designing the prioritization and retrieval methodology by assigning higher priority or relevance to observations with characteristics that are unusual with respect to the expected trend learned from the past data. To better match expert-like prioritization, the context of the observations are also interpreted by simultaneously

considering multiple features in the observations and not just the presence and absence of landforms. Here, the context is based on the presence of features of interest to human experts and are useful in post-downlink analysis such as retrieval and novelty detection for further analysis and subsequent planning.

In addition to retrieval tasks, the proposed prioritization approach can also impact prioritizing the downlinking operations onboard planetary orbiters. The limited bandwidth over which the data is transmitted from planetary orbiters is a limitation for transmitting the total possible data volume acquired. Onboard analysis to determine the context of the images and downlink the most relevant instances can potentially be useful to optimally utilize the bandwidth and enable the transmission of most significant observations.

In this chapter an approach is designed that analyzes the content of the image and assigns it a priority or rank by using expert guidance. This technique is useful for i) ground based retrieval of datasets with respect to a query, ii) ground based support in highlighting the most novel instances to help in targeting future observations, and iii) onboard ranking or prioritization of the dataset based on its relevance for downlinking over the limited bandwidth for optimally utilizing the resources.

We focus on formulating a generalized representation of an already observed landform and allow the prioritization or retrieval module to determine if a new instance with that instance is unusual. We combine strategies such as rule based knowledge repository construction and novelty detection to form a meaningful representation of interesting classes and identify outliers with respect to these representations to favor exploration. Additionally, what makes an instance *interesting* is subjective to expert interests. Our strategy therefore allows expert guidance to refine the rules with domain knowledge.

The objective of the proposed method is to emulate expert-like identification of

relevant datasets to determine the most scientifically interesting or important instances for ground based retrieval as well as to maximize the return of high value data over limited bandwidth. The focus is laid on geological features in planetary images *interesting* to experts. All observations with these expert identified interesting landforms (ILFs) have a high relevance. Additionally, relevance also depends on how novel or anomalous the instance is with respect to expert selected attributes. The more anomalous the instance is with respect to expert defined attributes, the higher is its value for analysis and is hence assigned a higher priority.

An expert guided rule formation methodology is introduced where the rules represent the expected characteristics of each ILF with respect to the attributes [22]. Deviation of an ILF bearing instance from its rule causes it to be more significant for further investigation. This this deviation is quantified as *relevance score* to characterize the degree of novelty of an ILF bearing instance. In order for the proposed approach to generate an expert acceptable priority order, the rules have to be aware of domain knowledge. This knowledge is elicited by seeking expert rating on observed relevant instances from which the rules are formed. The rating allows incorporation of prior domain knowledge to guide the rules in the decision making process. Along with deciding on the priority of an instance, the prioritization step also lists the attributes with respect to which the instance is novel, explaining *why* the decision was made to increase the expert's trust in the system. The proposed approach is evaluated on images from the early operation stage of the Thermal Emission Imaging System (THEMIS) instrument onboard the Mars Odyssey spacecraft [49]. The evaluation is repeated over successive batches of images where the rules are updated to account for changes and observe that the acceptability of the rule generated ranking to experts increases when domain knowledge has been incorporated.

7.1 Problem Definition

The key objective of this work is to prioritize or retrieve observations that are highly relevant (due to its context) to domain scientists for further analysis.

Relevance by context : An observation may be relevant for future investigation because its context may imply scientific novelty or it may be pertinent to answer a scientific question of interest. This is determined by the presence of *known, interesting* classes in that observation. A given observation is considered relevant due to the presence of

- *known interesting classes.* Interesting classes are the morphologies of interest to scientists. Thus if one or more of the interesting landforms chosen by experts is present in a given observation, that increases its relevance.

Identifying these observations from the large repository is pivotal to increase the scientific yield from an instrument as manually searching for interesting observations is likely to be infeasible. Formally, this problem can be defined as follows.

Let $O = \{o_1, \ldots, o_N\}$ be the set of N multispectral observations acquired and C denote the set of interesting classes chosen by experts. These C classes can be distinguished from the background class based on distinct texture or composition. Let C' denote the set of all other classes that belong to the background such that $C \cap C' = \emptyset$. The set of features that belong to the background class, are assumed to have lower significance to scientists. The relevance of a given observation o_j is determined based on whether it contains any feature that belongs to the set of interesting classes C. Additionally, each observation o_j can be summarized using a descriptor $d_{o_j} = < C_j, \ell_j, t_j >$ with respect to certain key attributes of the observation. Here, C_j denotes the set of classes present in the observation o_j and ℓ_j and t_j denotes its

spatial and temporal properties. The objective is to identify the relevant subset S, where $S = \{o_j \,|\, C_j \in C, S \subseteq O\}$.

The identified relevant subset S contains those observations that are expected to have greater significance for investigation. However, the relative relevance of each instance in S also varies based on the additional attributes such as ℓ_j and t_j. Thus, these instances in the relevant subset S are also to be evaluated based on these attributes to determine the ranking of order of importance. This ranking is based on the following:

- if the instance is novel either due to novelty (deviates from the expected properties of the class such that it is scientifically significant) or due to an unexpected combination of the attributes.

- if the combination of attributes in the instance is applicable to answer a scientific question of interest.

The overall priority is determined from the context of a given image determined by a simultaneously analyzing the attributes from the image descriptor. The weights or priority of the factors are determined from the expectation of scientists

7.2 Related Work

Relevant instance retrieval and *learning to rank* class of problems have been studied with applications in both document [75] and image retrieval [76], where conventionally features such as BM25 and Page Rank have been used to symbolize documents, while SIFT features, Fisher vectors have been utilized to represent images. Recent studies have utilized autoencoder features for document ranking given a query [77] and patch-based [78] or holistic neural codes extracted from the intermediate layers of neural networks for image retrieval. However, in the proposed work,

we emphasize on detecting novelty (by considering association, spatial and temporal attributes to interpret the scene content) instead of similar instances while allowing expert preference to guide the prioritization. Requirement based prioritization of test cases using k-means clustering ranks test cases that meet pre-specified requirements [79]. In contrast, this work has to effectively identify unseen ILF associations autonomously instead of assigning it to an already known cluster.

This work adopts a rule based approach to represent the classes of interest as it can extract meaningful relations, if any, even from a small amount of data. The proposed approach differs from existing rule based approaches [80] and semantics-enabled information mining [81] as the extracted rules favor discovering novel instances and incorporate domain specific preferences.

The proposed framework of rule based determination of scientific value of relevant instances is shown in Figure 7.1. In this study, the expert selected ILFs are dunes, dust devil tracks, channels and wind streaks and the rules are to be formed with respect to each ILF. The instances with ILFs are selected by experts from the existing observation repository. The expert also describes the selected instances with respect to the attributes to form the training set. The proposed approach then forms a rule based representation of each class from the training set. The rules are then updated using expert feedback to form the prioritization module, which is then tested to assign priority to a test set containing relevant instances from subsequent orbits. The instances from the test set are then used to update the rules, which are used to prioritize the next batch of instances.

7.3 Proposed Approach: Prioritization by Identifying Anomalous Instances

In this section, a rule based representation is proposed that defines the expected attributes of each ILF. Under the prioritization criterion chosen by experts, any in-
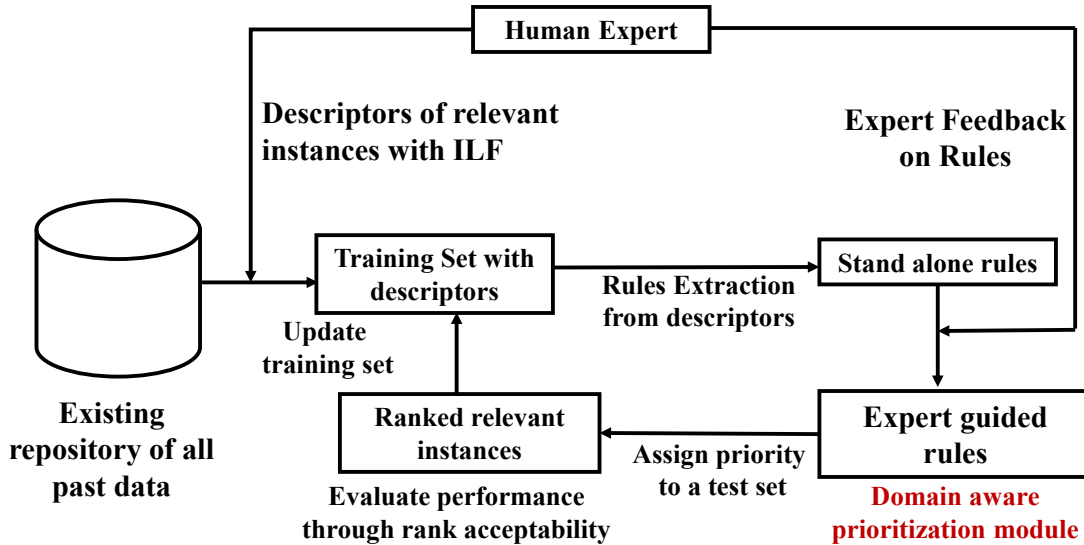
Figure 7.1: Overview of the Rule Based Prioritization Approach

stance containing one or more ILFs has a high priority for further analysis, expressed in terms of *relevance score*. However, the more unusual the instance is with respect to the rules, the higher its novelty and relevance score is for further investigation. Such unusual instances have higher priority than an ILF bearing instance that meets the expected rules. The novelty of a relevant instance with respect to the rules is determined based on expert selected attributes which are:

- **class relevance**: This denotes the relative relevance of the classes present in the instance for further analysis. Although all relevant instances contain one or more ILF, which increases its priority, the relevance of each class may vary from expert to expert. Presence of ILFs that are more relevant to the expert increases the relevance score of the instance.

- **spatial context or inter-class association**: The context of the images determined by the association of ILFs within an instance. For example, all past observations of a given ILF i may appear with another ILF j in the same instance. However, in a test image, if ILF i occurs by itself or is present with

101

another ILF l, it increases the relevance score of the instance as it implies aberration in the data generating process that is significant for further investigation.

- **spatial location**: Unusual spatial location of any ILF with respect to its previously observed spatial distribution. This increases the relevance score of the instance and assists experts to explore unresearched regions thereby enhancing existing surface maps.

- **seasonal information**: An ILF may be prevalent during a particular season. Any occurrence of that ILF in a different season is considered anomalous and increases the relevance score of the instance it appears in.

7.3.1 Rule Formation Of Relevant Classes

Rule based characterization of ILFs defines the expected attributes for each of the C ILFs from its observed instances. Presence of an ILF renders it important for further analysis however, the rules are used to identify the most unique instances and hence impact prioritization. The rules are learned iteratively from the observations acquired in each batch. In a given batch y, from the set $D_y = \{d_1, d_2, \ldots, d_N\}$ of all N observations acquired, the relevant subset $\Theta_y = \{\theta_1, \theta_2, \ldots, \theta_{N_C}\}$ containing N_C instances with ILFs are identified by experts. For all N_C instances, experts also provide the instance descriptor $\theta \equiv\ <labels, latitude, longitude, season>$ for ground based rule generation with respect to the attributes. These instance descriptors form the training dataset for rule generation. Here, the *labels* are to be obtained from the convolutional neural network based ILF detection technique described in Section 2.4. An overview of this process is shown in Figure 7.1.

Strength or frequency of any combination of inter-class association between ILFs $(i, j, \ldots, n)$ is the joint probability of its occurrence together within an instance and

is computed empirically from the training dataset. Strength indicates the likelihood of co-occurrence of two or more ILFs in one instance and does not necessarily imply causality. If the ILFs $(i, j, \ldots, n)$ co-occur in θ_s instances in the set Θ_y and $|\theta_s|$ denotes the number of such instances, then the strength of association $s_{i,j,\ldots,n}$, between the given combinations of ILFs is computed using

$$s_{i,j,\ldots,n} = \frac{|\theta_s|}{N_c}, \tag{7.1}$$

where $\theta_s \subseteq \Theta_y$. While a high s indicates that the association between the ILFs is frequently observed, low s implies that the association is rarely observed, which increases its relevance for further analysis. For unseen associations, $s = 0$.

In addition to inter-class associations, the rules also incorporate the set of expected latitude and longitude center for each ILF i to determine spatial anomalies. If a given ILF is detected in a spatial location that is unusual with respect to all previously observed instances of that ILF, such a test instance has a high priority as it prompts experts towards new and previously unknown locations of that ILF, thereby aiding in exploration and enhancing existing surface maps. If X_i and Y_i are the vector of latitudes and longitudes of all instances of ILF i in the training data, the spatial clusters of the given class are identified by grouping these location parameters of the given class within a given distance which acts as the threshold. In subsequent steps, all instances of ILF$_i$ are then assigned to the closest cluster center $(X_{\omega_i}, Y_{\omega_i})$, where ω is the number of clusters of the ILF i determined by clustering. A given instance in the training dataset is also used to create a new cluster if cannot be assigned to any existing cluster, thereby allowing the expected spatial location of the class to be updated. Following cluster assignment, the rules for each ILF is updated with the cluster centers and spatial variances $(\sigma_{lat_{\omega_i}}, \sigma_{long_{\omega_i}})$ of each cluster of that ILF, where the variances are determined from clustering the data or from the default distance

(threshold) parameters. The predominant spatial locations of each ILF is the set $\Omega_i = \{(X_{\omega_i} \pm \sigma_{lat_{\omega_i}}, Y_{\omega_i} \pm \sigma_{long_{\omega_i}})\}$.

Season of prevalence learned from past data of each ILF is used to characterize its seasonality. Season of occurrence of each observed instance of every ILF i is expressed as a binary vector $\mathbf{t}$, whose length corresponds to the number of seasons ϕ on Mars [82] or the planetary surface under study. For every past instance θ of ILF i, if the associated season is φ, then $t_\theta[k] = 1$, for $k = \varphi$ and $t_\theta[k] = 0$ otherwise, where $k = 1, \ldots, \phi$. The expected season of prevalence μ_{k_i}, of each ILF i can be expressed as $\mu_{k_i} = E[k_i] = \sum_{k=1}^{\phi} k \cdot f(k)_i$, while the seasonal variance is given by $\sigma(k_i) = \sum_{k=1}^{\phi}(k - E[k_i])^2 \cdot f(k)_i$. Here $f(k)_i = \frac{1}{N_{c_i}} \sum_{I \in \text{ILF}_{c_i}} t_\theta[k]$, represents the prevalence of ILF i in season k learned from N_{c_i} instances of N where i appears.

7.3.2 Incorporating Expert Knowledge

Rules extracted for every relevant class is uninformed of expert preferences and domain knowledge. This may hamper the prioritization module's effectiveness in generating expert acceptable ranking. Here a technique is described for enriching the extracted rules by incorporating it with expert knowledge. This is essential to determine expert preference of relative importance of each class. Additionally, expert preferences are crucial for refining association strength. This is important when the rules are formulated from local knowledge on a small training dataset. For example, in the early stages of operation when only a small fraction of the surface has been observed, expert knowledge is useful to incorporate any known prior global context, which reduces any bias that may arise due to limited imaging of the surface. Moreover, this knowledge is utilized to override any data driven rule and guide it with expert knowledge when the relevance of that rule cannot be directly inferred from the data. For example, a frequent association may still bear high significance due to a domain

specific reason. Seeking expert feedback ensures that the rules are aware of this context.

However, expecting accurate quantification of this knowledge even by experts is unrealistic and remains a challenge [83]. This knowledge is elicited from experts by recording their rating of images from the training data of batch y. For every expert identified relevant instance, the preferences are determined by recording the *factor* (ILF relevance, interesting inter-class association) that caused the instance to be relevant on a scale of 1 to 5, indicating the *strength of preference.*

These ratings are used to guide the rules with expert opinion in the form of weights governing the class importance w_i and weights for all combinations of observed inter-class associations $w_{i,j,\ldots,n}$. Here w_i represents the relative importance of ILF i to the experts, whereas $w_{i,j,\ldots,n}$ represents the expert rating on the novelty or importance of association between ILFs $i, j, \ldots, n$. These weights are initially set to unity and then updated to the average normalized expert rating of every instance bearing the ILF i or observed inter-class association $i, j, \ldots, n$ respectively. The rules formed for each ILF after expert feedback can be expressed as $< w_i, s_i, \Omega_i, (\mu_{k_i}, \sigma_{k_i}) >$, where s_i is a vector indicating association strength of i with all other class combinations. A feedback form from the experts were utilized that displays an image of each ILF category and each observed ILF association to record the corresponding expert weight. These weights are used to determine the class relevance and update the inter-class association strength, while the expected spatial location and season is learned in a data driven manner.

7.3.3 Inference Rules For Prioritization

The rules extracted define the expected attributes with respect to each relevant class. Here, an inference methodology is proposed to determine the relevance score

of the test instances in batch $y+1$ with respect to each attribute of the rules. A higher relevance score is indicative of greater significance of the given instance for further analysis because it is more likely to be relevant to experts. The novelty of each instance with the ILFs are determined using the attributes defined above, namely class weight, association strengths of multiple classes, spatial and temporal anomaly. For an ILF bearing instance τ in the test set, that can be described as $< \{i, \ldots, n\}, (X_\tau, Y_\tau), t_\tau >$, where $\{i, \ldots, n\}$ denotes ILFs present in τ, (X_τ, Y_τ), t_τ indicates the latitude/longitude and season of τ respectively, the inference rules to compute the relevance scores of τ with respect to each attribute is determined by:

- overall relevance of each ILF i in the instance, computed from expert weight $\sum_{\forall i \in \tau} w_i$,

- novelty of association $a_{i,j,\ldots,n}$ between all ILFs in the given instance, which is given as

$$a_{i,j,\ldots,n} = \frac{w_{(i,j,\ldots,n)}}{\left(s_{(i,j,\ldots,n)} + \epsilon\right)}, \forall (i, j, \ldots, n) \in \tau \tag{7.2}$$

where s is the association strength of all ILFs in τ computed using equation (7.1) and ϵ is a constant ($\epsilon << s$), representing the strength of previously unseen associations. For unobserved associations, $w_{(i,j,\ldots,n)} = 1$. Here, the inverse of strength indicates how rare the association is.

- occurrence of ILFs in a previously unknown location. This is indicated when (X_τ, Y_τ) is further away than the expected variance from its nearest cluster center. Spatial novelty is computed using

$$\sum_{\forall i \in \tau} \delta_i - \sigma_{lat_{\omega_i}}, \text{if} \quad \delta_i > \sigma_{lat_{\omega_i}}, \tag{7.3}$$

where $\delta_i = min(|X_\tau - X_{\omega_i}|)$ for all spatial clusters of ILF i. Anomalous longitude occurrence also contributes to relevance score similarly. This step is similar to

assigning an instance to a previously known cluster.

- occurrence of i in its non-prevalent season, indicating temporal anomaly. For a given instance τ, the temporal novelty score is computed using

$$\sum_{\forall i \in \tau} \delta_t - \sigma_{k_i}, \text{if} \quad \delta_t > \sigma_{k_i}, \text{where} (\delta_t = |t_\tau - \mu_{k_i}|). \tag{7.4}$$

Thus, the overall relevance score of an instance is the sum of the ILF relevance score, inter-class association, spatial and temporal relevance scores. However, when two or more instances have equal relevance scores, higher priority is assigned to the instance that has ILFs with higher w_i and $a_{i,j,...,n}$. A second tie-breaker is added that assigns higher priority to the instances with higher types of novelties. For example, an instance with inter-class association, spatial and temporal novelties is assigned higher priority than an instance with inter-class association and spatial novelties only.

7.3.4 Rule Update

The relevant instances from batch $y+1$ are added to the training dataset to update the previously known rules using the proposed rule formation approach. Although the expert knowledge elicited from the previous training dataset is retained to guide the updated rules, an expert can update their ratings at this phase to appropriately direct the rules to reflect any modifications to the geological classes and its attributes of interest. However, feedback during update phases are expected to be infrequent.These updated rules are in turn utilized for prioritizing the test data in the following batch.

7.4 Evaluation Metrics

This approach ranks all instances in a test set batch based on how interesting its contents are for experts to analyze. Its effectiveness is determined by its ability to generate expert acceptable ranking in a given test batch. Hence, for every test batch,

the similarity in order of ranking $\mathbf{r} = [r_{\tau_1}, \ldots, r_{\tau_B}]$ generated from the rules for each test set batch with that of the expert order of ranking $\mathbf{e} = [e_{\tau_1}, \ldots, e_{\tau_B}]$ of the same set is computed, where B denotes the number of instances in each batch and r_{τ_b} and e_{τ_b} denotes the rule generated rank and expert rank of instance τ_b. The metrics used to quantify this similarity in ranking are

- Mean Average Precision (MAP): For a given test set batch q, we compute the average precision (AP) at every position b where the rule generated rank r_{τ_b} of the instance τ_b matches its expert rank e_{τ_b}. This is expressed as

$$\text{AP(b)}_q = \frac{1}{B} \sum_{b=1}^{B} \gamma_b, \tag{7.5}$$

  where

$$\gamma_b = \begin{cases} p_b, & \text{if } r_{\tau_b} = e_{\tau_b} \\ 0, & \text{otherwise.} \end{cases}$$

  Here p_b denotes the precision in detecting the correct ranking of the relevant instances up to position b. MAP is then computed as the mean AP over all Q test set batches and is given by $\text{MAP} = \frac{1}{Q} \sum_{q=1}^{Q} \text{AP(b)}_q$.

- Spearman Rank Correlation (SRC): This computes the correlation in the rule generated rank with the expert ranking of each dataset.

These metrics quantify the goodness of ranking and hence the prioritization order. A high value (≈ 1) denotes acceptability of the rule generated ranking to experts and implies effectiveness of the rule based methodology in capturing expert preferences by identifying the most relevant instances.

ILF A – Dune; ILF B - Dust devil; ILF C – Channel; ILF D – Wind streak

Iteration 1
Highest priority instance:
<wind streak, channel>

Iteration 2
Highest priority instance:
<channel>

Iteration 3
Highest priority instance:
<dune, channel>
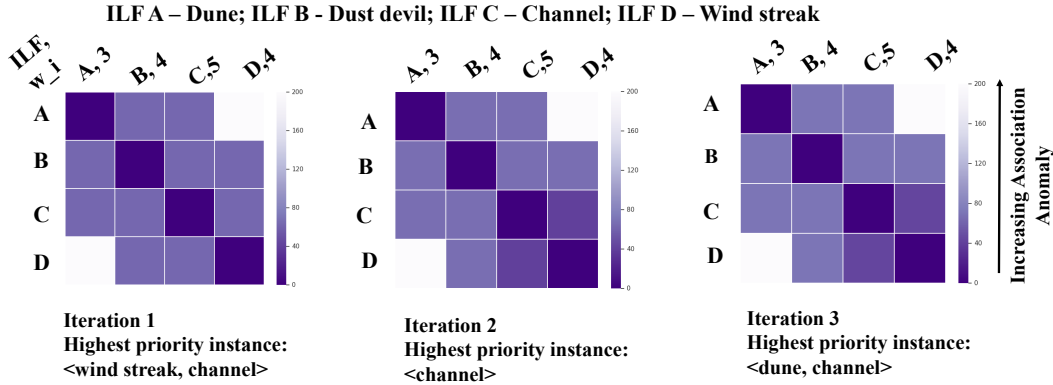
Figure 7.2: Inter-Class Association Anomaly Score (Shown Between Two Classes Only) Update Over Iterations for Identification of Novel Associations.

7.5 Dataset And Results

7.5.1 Dataset Description

The visible imager datasets from THEMIS has been used in this study (`http://viewer.mars.asu.edu`). This dataset consists of all 1600 (single channel band 3) datasets collected from orbits 81 to 239, out of which 300 instances were identified by experts for its relevance due to presence of ILFs. Along with the spatial and seasonal attributes of these instances, experts annotated each instance to identify the interesting classes that appeared in it. The annotated images were divided into training and test sets. The training set consists of 99 relevant images from the first 71 orbits. This training dataset is used in extracting the rules which are then guided with expert feedback. The remaining 201 relevant images from the 87 subsequent orbits form the test set batches on which the rules are applied to rank it according to its scientific value. Each test set batch consists of eight images which are ranked and validated with expert acceptance of the ranking to evaluate the effectiveness of our approach. All instances from the test batch are used to update the rules, which is then used to rank the next batch.

EQ- Equatorial; SP- South Polar; NP – North Polar

Iteration 1:
• Known spatial locations of dunes

Iteration 2:
• Known spatial locations of dunes modified
• Dune observed at an unusual location

Iteration 3:
• Known spatial locations of dunes modified to include north pole
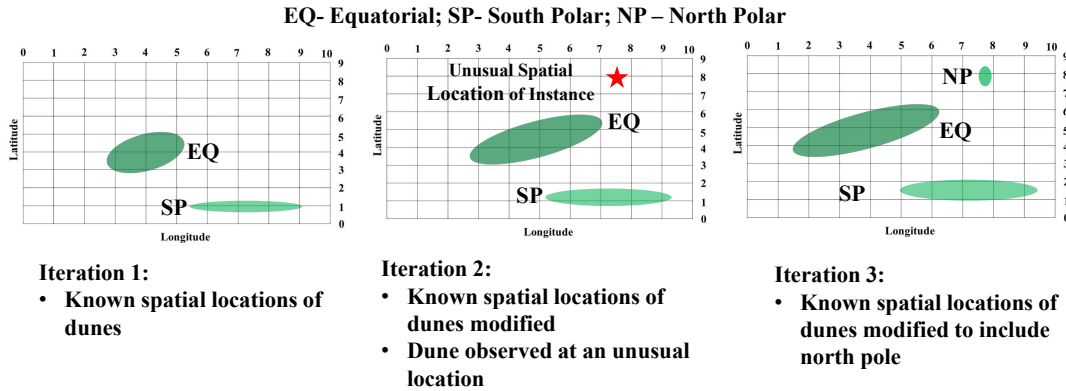
Figure 7.3: Learning Season of Prevalence of a Given ILF (Wind Streak) Through Rule Update.
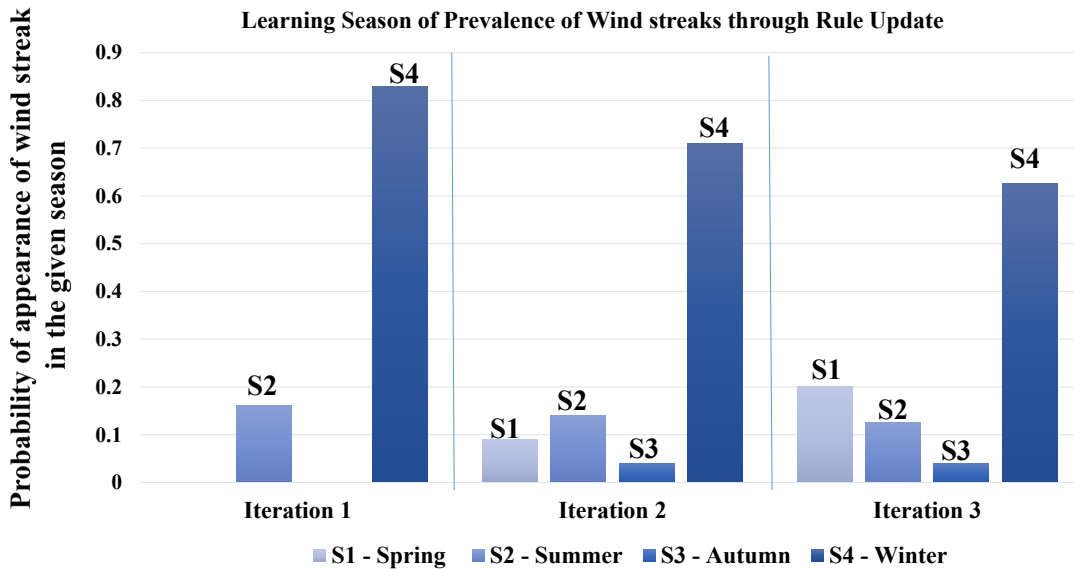
Figure 7.4: Learning Season of Prevalence of a Given ILF (Wind Streak) Through rule update.

7.5.2   Results

In this section the results obtained to evaluate the effectiveness of our proposed rule based methodology in identifying unusual instances and generating an expert acceptable ranking based on its relevance are discussed. Here, the test sets contain only expert identified instances with the ILFs from each orbit. The ranking from both stand-alone rules that are solely extracted from the data and the expert guided

Table 7.1: Acceptability of Rule Generated Priority to Experts Using MAP and SRC Over All Test Set Batches.

Rule Forming Methodology	MAP	SRC
Standalone Rules	0.1864	0.5482
Expert Guided Rules	**0.765**	**0.984**

rules with that of expert ranking (from one expert) of the test set is compared using the evaluation metrics discussed before. The expert ranking of the test set is only utilized to validate our approach. Table 7.1 shows the average acceptability of the prioritization order to an expert over all test set batches using all attributes of the inference rules. The stand-alone rules score poorly using both metrics as these are uninformed of expert preferences. These rules give equal weight to all ILFs and inter-class associations. However, both MAP and SRC demonstrate the improvement in generating expert acceptable rank after incorporating expert knowledge. This is expected as the rules are aware of expert preferences. We also explain our decisions to increase interpretability by listing all attributes with respect to which an instance was novel by identifying the attribute which increased the relevance score of the observation.

The detection of novel instances is demonstrated by highlighting some examples of high priority instances identified in the test set batches. In Figure 7.2, the grids denote the association anomaly score of each ILF with every other ILF (only two-way associations are shown here). The corresponding expert weights of each ILF w_i, are shown as (ILF$_i$, w_i). Based on these association anomaly scores in iteration 1, the highest priority instance identified contains wind streak and channel, that have high class relevance and association anomaly scores. In the subsequent iterations, the relevance of this association therefore decreases, while the association relevance scores of unobserved associations such as dune and wind streak remains high. The

proposed method also favors instances with channels as it has highest relevance to the expert. Figure 7.3 shows the discovery of dunes at a location that was unknown based on previous observations. This discovery of north polar dunes then updates the existing planetary map for subsequent iterations. Figure 7.4 shows an example of the seasonal rule update for wind streak. In iteration 1, the knowledge of predominant season is winter, which makes wind streak instances during autumn and summer in the next orbit/batch very unusual and hence is assigned high downlink priority. This is used to update the knowledge of predominant season which results in the updated probabilities in iteration 2. By discovering different types of novel instances, the rules are updated such that highest preference is always assigned to instances that are most novel with respect to the attributes or are previously unobserved. This refinement of the existing knowledge base allows learning the ILF characteristics over time while informing experts about such updates.

This chapter presented an expert guided rule based methodology for identifying the most relevant images with interesting landforms and rank them in order of their scientific context. The rules thus formed represent the expected attributes of the interesting landforms and creates a knowledge repository from planetary images. A priority inference approach is proposed where the relevance of an image is determined by its deviation from the formulated rules. The effectiveness of this methodology is evaluated by determining how acceptable the prioritization order is to experts. As expected, the prioritization acceptability increases when the rules are guided by expert feedback as rules formulated solely from the data is uninformed of domain specific preferences. Here the results of rule based prioritization are presented where the images have been hand labeled. In order to operate with minimal human intervention, the labels of the images are obtained using the convolutional neural network approach and its other variants such as convolutional autoencoder as discussed in Chapter 2.

This has applications in both assisting domain scientists by highlighting anomalous or novel observations, aiding in generating feature maps and improving the efficiency of the data processing and analysis pipeline. In addition, the promising results in performing such tasks also offer further justification in enabling onboard systems to have such capabilities. However, for these methods to successfully operate in onboard environments, computational resources (processing, power as well as storage) require significant enhancement. This would not only allow assigning low priority to observations that are of poor quality due to noise, but also assigning higher priority to novel observations and facilitate planetary exploration with limited downlinking resources.

Chapter 8

CONCLUSION AND FUTURE WORK

8.1 Conclusion

This work focused on developing advanced processing techniques for time-varying representations for the characterization and interpretation of remote sensing observations for the purpose of change and novelty detection.

For frequently acquired image time-series, this representation is in the form of time-varying models that capture the region specific seasonality to estimate model parameters at different phases of the land surface state, including a time-varying frequency parameter that signifies signal regrowth rate. From time-frequency representations the varying growth rate or change rate of land surface reflectance time-series is observed. A multiple model selection approach is also proposed that can select best models for a given state based on instantaneous phase functions that relates to the instantaneous frequency or land surface variation rate at different stages. The parameters estimated from these models are further utilized for unsupervised change detection as well for spectral separability analysis of different change events. Overall, this results in a framework for multitemporal analysis where each observation is monitored and change decisions are made sequentially. The change points are then monitored for change rate variation using model switching to allow temporal adaptability as well as change signature analysis. The proposed approaches are particularly suitable for detecting long-term changes in land cover. Short-term events that occur and normalize at at rate faster than the sampling rate cannot be detected as anomalous samples are not observed. Atmospheric effects and interference due to cloud cover

can also affect the detection of short-term events. However, the long-term effects of such changes can still be detected. This not only provides improved insight into land surface processes but also reduces the data volume to be analyzed human experts by highlighting anomalies and supporting further analysis of those observations. In addition, the proposed techniques have also been examined for spatial adaptability. By considering datasets from geographically heterogeneous areas, it is observed that these techniques can adjust to the spatial variability and can therefore be extended both spatially and temporally to different change events.

On the other hand, for planetary images, the time-varying representations are formed with respect to geological landforms of interest chosen by domain scientists which defines the expected attributes of each landform. These representations are refined with every batch of available downlinked data to update the knowledge of planetary landforms of interest and aids in the process of exploration and mapping. This aligns closely with retrieval techniques for selecting the most anomalous observations, but has a significant difference as the objective here is to retrieve the most anomalous observation. Highlighting these observations to domain scientists assists in pointing out the most interesting observations for further analysis and in pointing the most scientifically relevant observations to aid in planning subsequent observations. In addition, with the improvement in computational resources onboard, expert-like decision making ability can also be imparted for prioritization and maximizing scientifically valuable (novel and high quality) data downlinking capabilities.

8.2   Future Work

8.2.1   Modeling Frequently Acquired MODIS Time-Series

Several extensions of this work is possible and these are summarized below:

1. Improved seasonality models that accounts for higher order periodicities in Earth surface time-series: In various cases, Earth system components have several periodic components. For example, several classes of land cover have biannual components, annual land surface temperature time-series exhibit diurnal as well as seasonal cycles. An improved model that accounts for these variations will enhance our interpretation of the time-varying processes.

2. Improved post-change models: Although the proposed model selection technique allows switching between land cover models, better post-change functions are to be designed for accurately estimating the instantaneous frequency that matches post-change behavior and growth rate trajectories and to allow correct model selection.

3. Adaptive estimation of change rate function: A primary observation which this work has been centered around is the time-varying frequency of land cover, specially during change and regrowth. This variation in frequency is proposed to be tackled by allowing model selection based on the similarity between the observed change rate and the change rate modeled by pre-determined functions of time-varying frequency. Further studies in this area would examine different classes of change rate function or allow adaptive estimation of the function to model previously unobserved change and regrowth examples.

4. Modeling frequency variation due to multiple and simultaneous change events: This work addressed modeling change rate variation in the event of a primary change event. An improvement possible in this direction would model multiple frequency components due to simultaneous events or when a given type of change induces another event, for example gradual degenerative trend of droughts can often make the vegetation in the area susceptible to forest fires.

5. Modeling seasonality-aware expected spectral behavior: The proposed autoen-

coding technique to learn a representation of the expected spectral signature can be improved further by ensuring that the learned representation is aware of the seasonal variation. This is essential to extract a better representation of the expected signature and in turn extract better change signatures. This is crucial for detecting change (specially of small magnitude) in regions with a high seasonal variation and distinguish between change and seasonality. Additionally, certain regions can have different land covers and hence different surface reflectance and spectra at different times of the year. Incorporating this knowledge will further enhance the autoencoder based models. For example, urban or forested regions can have snow cover in winter months. Awareness of this change in at specific seasons, will improve the regional spectral representation. This improvement will also be of use for detecting temporal deviations in expected events, for example early or late onset of snow cover or phenological stages.

6. Fusing complementary sensor observations: While the temporal resolution of MODIS forms an excellent repository for monitoring changes, the spatial resolution can be enhanced by utilizing complementary observations from higher spatial resolution data from instruments such as Landsat, ASTER, Sentinel; higher spectral resolution airborne data and observations from other modalities such as LiDAR and SAR imagery.

7. Improving evaluation techniques: Additional insight can be gained by analyzing the receiver operator characteristics for the change detection approaches to examine the variation of the probability of detection with probability of false alarms at different SNR. Additionally, shifting from the single pixel approach to a neighborhood based approach will provide additional spatial context of change and improve the SNR in the datasets. Moreover, considering multiple pixels in a neighborhood can also improve decision making where the particle filter obtains estimates of all pixels within a spatial

117

window to improve detection decisions.

8.2.2 *Landform Representation for Prioritization and Retrieval of Planetary Images*

Improvements and future research possible in this area would include:

1. Feasibility of long-term prioritization: An important continuation of this work is to analyze the feasibility of deploying such an approach throughout the life cycle of instrument as it would require fully autonomous functionality. A key source of error in prioritization is incorrect label prediction by visual feature recognition approaches as mislabeling an instance would invoke the incorrect comparison of class attributes leading to incorrect prioritization and retrieval. Furthermore, for long-term deployment, such a system should correctly generate the image labels and update the class rules independently. Incorrect labels will also hamper rule update and consequently prioritization and retrieval in the future. Thus, in addition to studying the feasibility of this approach, enhancing landform detection should also be explored.

2. Improved decision making by considering multi channel images: Retrieval and prioritization techniques proposed in this work introduces a novel system for analyzing planetary images and incorporating expert feedback with applications in both ground based and onboard analysis of data. Further improvement in this work is possible by modifying the decision making process to include both the infrared and visible images for enhanced context.

3. Integrating multi sensor observations for enhanced definition of context: The definition of context considered here is spatial association with other landforms of interest, spatial and seasonal attributes. This can be further enhanced by considering observations from multiple instruments at varying spatial and spectral data, as well as data from altimeters, atmospheric observations and rovers for an improved definition of context that can potentially allow more precise retrieval. Ultimately, such

techniques could lead to a comprehensive knowledge base of planetary surfaces and can be extended to any planetary surface of interest for efficiently allowing combined analysis from all observations of the given region.

In addition, the developed time-series approaches are also applicable (with domain specific modifications) in different areas such as seismic, meteorological, financial and biomedical time-series data, while the prioritization approach can be used to formulate knowledge bases for spatio-temporal datasets and facilitate expert-guided data retrieval and collaborative tasks.

REFERENCES

[1] G. R. Van der Werf, D. C. Morton, R. S. DeFries, J. G. Olivier, P. S. Kasibhatla, R. B. Jackson, G. J. Collatz, and J. T. Randerson, "CO_2 emissions from forest loss," *Nature Geoscience*, vol. 2, no. 11, p. 737, 2009.

[2] R. E. Dickinson and P. Kennedy, "Impacts on regional climate of Amazon deforestation," *Geophysical Research Letters*, vol. 19, no. 19, pp. 1947–1950, 1992.

[3] A. Henderson-Sellers, R. E. Dickinson, T. Durbidge, P. Kennedy, K. McGuffie, and A. Pitman, "Tropical deforestation: Modeling local-to regional-scale climate change," *Journal of Geophysical Research: Atmospheres*, vol. 98, no. D4, pp. 7289–7315, 1993.

[4] D. Schimel and D. Baker, "Carbon cycle: The wildfire factor," *Nature*, vol. 420, no. 6911, p. 29, 2002.

[5] Z. Liu, A. P. Ballantyne, and L. A. Cooper, "Biophysical feedback of global forest fires on surface temperature," *Nature Communications*, vol. 10, no. 1, p. 214, 2019.

[6] Y. Liu, S. Goodrick, and W. Heilman, "Wildland fire emissions, carbon, and climate: Wildfire–climate interactions," *Forest Ecology and Management*, vol. 317, pp. 80–96, 2014.

[7] Y. Shu-ren, "Effects of fire disturbance on forest hydrology," *Journal of forestry research*, vol. 14, no. 4, pp. 331–334, 2003.

[8] S. V. Ollinger, A. D. Richardson, M. E. Martin, D. Y. Hollinger, S. E. Frolking, P. B. Reich, L. C. Plourde, G. G. Katul, J. W. Munger, R. Oren *et al.*, "Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks," *Proceedings of the National Academy of Sciences*, vol. 105, no. 49, pp. 19 336–19 341, 2008.

[9] N. Ramankutty, H. K. Gibbs, F. Achard, R. Defries, J. A. Foley, and R. Houghton, "Challenges to estimating carbon emissions from tropical deforestation," *Global Change Biology*, vol. 13, no. 1, pp. 51–66, 2007.

[10] S. Boriah, "Time Series Change Detection: Algorithms for Land Cover Change," Ph.D. dissertation, Univ. of Minnesota, 2010.

[11] T. L. Grobler, E. R. Ackermann, A. J. van Zyl, J. C. Olivier, W. Kleynhans, and B. P. Salmon, "Using Page's Cumulative Sum Test on MODIS Time Series to Detect Land-cover Changes," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 332–336, 2013.

[12] W. Kleynhans, B. P. Salmon, J. C. Olivier, F. Van den Bergh, K. J. Wessels, T. L. Grobler, and K. C. Steenkamp, "Land cover change detection using autocorrelation analysis on modis time-series data: Detection of new human settlements in the gauteng province of south africa," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 3, pp. 777–783, 2012.

[13] W. C. Olding, J. C. Olivier, B. P. Salmon, and W. Kleynhans, "A forecasting approach to online change detection in land cover time series," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019.

[14] W. Kleynhans, J. C. Olivier, K. J. Wessels *et al.*, "Improving land cover class separation using an extended Kalman filter on MODIS NDVI time-series data," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, pp. 381–385, 2010.

[15] ——, "Detecting land cover change using an extended Kalman filter on MODIS NDVI, time-series data," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, pp. 507–511, 2011.

[16] H. Carrão, P. Gonçalves, and M. Caetano, "A nonlinear harmonic model for fitting satellite image time series: Analysis and prediction of land cover dynamics," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, pp. 1919–1930, 2010.

[17] A. Anees and J. Aryal, "A Statistical Framework for Near-Real Time Detection of Beetle Infestation in Pine Forests Using MODIS Data," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1717–1721, 2014.

[18] ——, "Near-real time detection of beetle infestation in pine forests using MODIS data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 3713–3723, 2014.

[19] A. Anees, J. Aryal, M. M. OReilly, and T. J. Gale, "A relative density ratio-based framework for detection of land cover changes in MODIS NDVI time series," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 8, pp. 3359–3371, 2015.

[20] S. Chakraborty, A. Banerjee, S. Gupta, A. Papandreou-Suppappola, and P. Christensen, "Estimation of dynamic parameters of MODIS NDVI time series nonlinear model using particle filtering," in *IEEE Int. Geosci Remote Sens. Symp.*, 2017, pp. 1091–1094.

[21] S. Chakraborty, A. Banerjee, S. K. Gupta, P. R. Christensen, and A. Papandreou-Suppappola, "Time-varying modeling of land cover change dynamics due to forest fires," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 6, pp. 1769–1776, 2018.

[22] S. Chakraborty, S. Das, A. Banerjee, S. K. Gupta, and P. R. Christensen, "Expert guided rule based prioritization of scientifically relevant images for downlinking over limited bandwidth from planetary orbiters," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 9440–9445.

[23] L. Bandeira, J. S. Marques, J. Saraiva, and P. Pina, "Automated detection of martian dune fields," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 626–630, 2011.

[24] L. F. Palafox, C. W. Hamilton, S. P. Scheidt, and A. M. Alvarez, "Automated detection of geological landforms on mars using convolutional neural networks," *Computers & Geosciences*, vol. 101, pp. 48–56, 2017.

[25] K. L. Wagstaff, Y. Lu, A. Stanboli, K. Grimes, T. Gowda, and J. Padams, "Deep Mars: CNN classification of mars imagery for the PDS imaging atlas," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[26] R. Castano, K. L. Wagstaff, S. Chien, T. M. Stough, and B. Tang, "On-board Analysis of Uncalibrated Data for a Spacecraft at Mars," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 2007, pp. 922–930.

[27] A. Castano, A. Fukunaga, J. Biesiadecki, L. Neakrase, P. Whelley, R. Greeley, M. Lemmon, R. Castano, and S. Chien, "Automatic Detection of Dust Devils and Clouds on Mars," *Machine Vision and Applications*, vol. 19, no. 5-6, pp. 467–482, 2008.

[28] T. A. Estlin, B. J. Bornstein, D. M. Gaines, R. C. Anderson, D. R. Thompson, M. Burl, R. Castaño, and M. Judd, "AEGIS Automated Science Targeting for the MER Opportunity Rover," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 3, no. 3, p. 50, 2012.

[29] D. R. Thompson, A. Candela, D. S. Wettergreen, E. N. Dobrea, G. A. Swayze, R. N. Clark, and R. Greenberger, "Spatial Spectroscopic Models for Remote Exploration," *Astrobiology*, vol. 18, no. 7, pp. 934–954, 2018.

[30] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, pp. 174–188, 2002.

[31] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373).* Ieee, 2000, pp. 153–158.

[32] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and bayesian missing data problems," *Journal of the American Statistical Association*, vol. 89, no. 425, pp. 278–288, 1994.

[33] S. Lhermitte, J. Verbesselt, I. Jonckheere *et al.*, "Hierarchical image segmentation based on similarity of NDVI time series," *Remote Sens. Environ.*, vol. 112, pp. 506–521, 2008.

[34] M. E. Jakubauskas, D. R. Legates *et al.*, "Harmonic analysis of time-series AVHRR NDVI data," *Photogrammetric Engineering & Remote Sensing*, vol. 67, pp. 461–470, 2001.

[35] M. E. Jakubauskas, D. R. Legates, and J. H. Kastens, "Crop identification using harmonic analysis of time-series AVHRR NDVI data," *Computers and Electronics in Agriculture*, vol. 37, pp. 127–139, 2002.

[36] L. Cohen, *Time-frequency analysis.* Prentice hall, 1995, vol. 778.

[37] P. Flandrin, F. Auger, E. Chassande-Mottin *et al.*, "Time-frequency reassignment: From principles to algorithms," *Applications in Time-Frequency Signal Processing*, vol. 5, no. 179-203, p. 102, 2003.

[38] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H.-T. Wu, "Time-frequency reassignment and synchrosqueezing: An overview," *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 32–41, 2013.

[39] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, 1992.

[40] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT press, 2016.

[41] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 2012, pp. 37–49.

[42] C. Elachi and J. J. Van Zyl, *Introduction to the physics and techniques of remote sensing*. John Wiley & Sons, 2006, vol. 28.

[43] M. Minomura, H. KUzE, and N. Takeuchi, "Adjacency effect in the atmospheric correction of satellite remote sensing data: Evaluation of the influence of aerosol extinction profiles," *Optical Review*, vol. 8, no. 2, pp. 133–141, 2001.

[44] E. Vermote and A. Vermeulen, "Atmospheric correction algorithm: spectral reflectances (mod09)," *ATBD version*, vol. 4, pp. 1–107, 1999.

[45] G. W. Petty, *A first course in atmospheric radiation*. Sundog Pub, 2006.

[46] C. B. Schaaf, F. Gao, A. H. Strahler *et al.*, "First operational BRDF, albedo nadir reflectance products from MODIS," *Remote Sens. Environ.*, vol. 83, pp. 135–148, 2002.

[47] L. Eklundh and P. Jonsson, "Timesat 3. 2 with Parallel Processing–Software Manual," *Lund University*, 2012.

[48] J. Eidenshink, B. Schwind, K. Brewer, Z. Zhu *et al.*, "A project for monitoring trends in burn severity," *Fire Ecology*, vol. 3, pp. 3–21, 2007.

[49] P. R. Christensen, B. M. Jakosky, H. H. Kieffer, M. C. Malin, H. Y. McSween, K. Nealson, G. L. Mehall, S. H. Silverman, S. Ferry, M. Caplinger *et al.*, "The Thermal Emission Imaging System (THEMIS) for the Mars 2001 Odyssey Mission," *Space Science Reviews*, vol. 110, no. 1-2, pp. 85–130, 2004.

[50] A. Anees, J. Aryal, M. M. O'Reilly, and T. J. Gale, "A relative density ratio-based framework for detection of land cover changes in MODIS NDVI time-series," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, pp. 3359–3371, 2016.

[51] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, 1992.

[52] B. P. Salmon, W. Kleynhans, F. van den Bergh, J. C. Olivier *et al.*, "Meta-optimization of the extended Kalman filters parameters through the use of the bias variance equilibrium point criterion," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, pp. 5072–5087, 2014.

[53] B. P. Salmon, W. Kleynhans, F. van den Bergh, J. C. Olivier *et al.*, "Land cover change detection using the internal covariance matrix of the extended Kalman filter over multiple spectral bands," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1079–1085, 2013.

[54] P. S. Beck, C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore, "Improved monitoring of vegetation dynamics at very high latitudes: A new method using modis ndvi," *Remote Sensing of Environment*, vol. 100, no. 3, pp. 321–334, 2006.

[55] C. Bacour, F.-M. Bréon, and F. Maignan, "Normalization of the directional effects in noaa–avhrr reflectance measurements for an improved monitoring of vegetation cycles," *Remote Sensing of Environment*, vol. 102, no. 3-4, pp. 402–413, 2006.

[56] R. S. Lunetta, J. F. Knight, J. Ediriwickrema *et al.*, "Land-cover change detection using multi-temporal MODIS NDVI data," *Remote Sensing of Environment*, vol. 105, pp. 142–154, 2006.

[57] S. Lhermitte, J. Verbesselt, W. W. Verstraeten, and P. Coppin, "A comparison of time series similarity measures for classification and change detection of ecosystem dynamics," *Remote Sensing of Environment*, vol. 115, pp. 3129–3152, 2011.

[58] F. Provost and T. Fawcett, "Robust classification for imprecise environments," *Machine Learning*, vol. 42, pp. 203–231, 2001.

[59] C. Chen, A. Liaw, and L. Breiman, "Using random forest to learn imbalanced data," Univ. of California, Berkeley, Tech. Rep. 666, 2004.

[60] C. Elkan, "The foundations of cost-sensitive learning," in *International Joint Conference on Artificial Intelligence*, vol. 17, 2001, pp. 973–978.

[61] M. Kubat, R. Holte, and S. Matwin, "Learning When Negative Examples Abound," in *European Conference on Machine Learning*, 1997, pp. 146–153.

[62] S. P. Parambath, N. Usunier, and Y. Grandvalet, "Optimizing F-measures by Cost-Sensitive Classification," in *Advances in Neural Information Processing Systems*, 2014, pp. 2123–2131.

[63] C. J. Rijsbergen, *Information Retrieval*, 2nd ed. Butterworth-Heinemann, 1979.

[64] J. E. Keeley, "Fire intensity, fire severity and burn severity: A brief review and suggested usage," *International Journal of Wildland Fire*, vol. 18, pp. 116–126, 2009.

[65] C. J. Chafer, M. Noonan, and E. Macnaught, "The post-fire measurement of fire severity and intensity in the Christmas 2001 Sydney wildfires," *International Journal of Wildland Fire*, vol. 13, pp. 227–240, 2004.

[66] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *ACM-SIAM Symp. Discrete Alg.*, 2007, pp. 1027–1035.

[67] H. Carrão, P. Gonçalves, and M. Caetano, "A Nonlinear Harmonic Model for Fitting Satellite Image Time Series: Analysis and Prediction of Land Cover Dynamics," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1919–1930, 2010.

[68] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.

[69] T. Zhan, M. Gong, J. Liu, and P. Zhang, "Iterative feature mapping network for detecting multiple changes in multi-source remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 146, pp. 38–51, 2018.

[70] M. Gong, H. Yang, and P. Zhang, "Feature learning and change feature classification based on deep learning for ternary change detection in sar images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 129, pp. 212–225, 2017.

[71] H. Zhang, M. Gong, P. Zhang, L. Su, and J. Shi, "Feature-level change detection using deep representation and feature change analysis for multispectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 11, pp. 1666–1670, 2016.

[72] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 1, pp. 218–236, 2006.

[73] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2196–2212, 2011.

[74] C. M. Bishop, *Pattern recognition and machine learning.* Springer Science+ Business Media, 2006.

[75] R. Silva, M. A. Gonçalves, and A. Veloso, "Rule-based Active Sampling for Learning to Rank," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases.* Springer, 2011, pp. 240–255.

[76] W. Zhou, H. Li, and Q. Tian, "Recent Advance in Content-based Image Retrieval: A Literature Survey," *arXiv preprint arXiv:1706.06064*, 2017.

[77] A. Albuquerque, T. Amador, R. Ferreira, A. Veloso, and N. Ziviani, "Learning to Rank with Deep Autoencoder Features," in *2018 International Joint Conference on Neural Networks (IJCNN)*.   IEEE, 2018, pp. 1–8.

[78] M. Paulin, M. Douze, Z. Harchaoui, J. Mairal, F. Perronin, and C. Schmid, "Local Convolutional Features with Unsupervised Training for Image Retrieval," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 91–99.

[79] M. J. Arafeen and H. Do, "Test Case Prioritization Using Requirements-Based Clustering," in *IEEE Sixth International Conference on Software Testing, Verification and Validation (ICST), 2013*.   IEEE, 2013, pp. 312–321.

[80] A. J. Lee, R.-W. Hong, W.-M. Ko, W.-K. Tsao, and H.-H. Lin, "Mining Spatial Association Rules in Image Databases," *Information Sciences*, vol. 177, no. 7, pp. 1593–1608, 2007.

[81] K. R. Kurte, S. S. Durbha, R. L. King, N. H. Younan, and R. Vatsavai, "Semantics-enabled framework for spatial image information mining of linked earth observation data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 29–44, 2017.

[82] B. A. Cantor, P. B. James, and W. M. Calvin, "MARCI and MOC Observations of the Atmosphere and Surface Cap in the North Polar Region of Mars," *Icarus*, vol. 208, no. 1, pp. 61–81, 2010.

[83] R. Castano, K. Wagstaff, L. Song, and R. Anderson, "Validating Rover Image Prioritizations," *The Interplanetary Network Progress Report*, vol. 42, p. 160, 2005.