Doctoral Thesis: Graph based Image fusion and features extraction for remote sensing applications.

David Alejandro Jimenez Sierra, M.Sc

Advisor: Hernán Darío Benítez Resrepo, Ph.D

Co-Advisor: Hernán Darío Vargas Cardona, Ph.D

Pontificia Universidad Javeriana Cali
Engineering Faculty - Ph.D in Engineering and Applied Science
Cali, Valle del Cauca, Colombia
2024- II

Contents

Executive Summary

Recent advances in sensor technology have lead to the increased availability of hyper-spectral, multi-spectral (MS), and synthetic aperture radar (SAR) images (at very high spatial and spectral resolutions), which describe an object or phenomenon. Each sensor captures different information that explains physical features. For example, a SAR sensor captures information about the physical characteristics of a surface (such as roughness, geometric structure, and orientation), and an MS sensor captures reflectances at different wavelengths from objects. Nonetheless, the ground information obtained from a single sensor is limited to draw reliable conclusions about phenomena such as changes in land cover and biomass growth. In contrast, image fusion techniques integrate spectral, spatial, and temporal information from several sensors to obtain appropriate information and construct images that are more suitable for human and machine perception. Image fusion is a process of combining two or more images into a single image which is more informative, hence useful and helpful in different remote sensing applications (i.e geology, agriculture, military, etc). Therefore, it is generally desirable to use data captured from different sensors.

Even though data fusion contributes to better performance in classification and detection in remote sensing, it is a complex task. For example, the different resolutions, units, dimensions, and formats are challenges imposed by raw data. Furthermore, homogeneous data (i.e. data acquired from the same sensor) presents small intra-class variability and brightness distortions (artifacts), in the case of heterogeneous data (i.e. data acquired from different sensors) the pixels have different signatures, and therefore follow distinct statistical behaviors that makes difficult the extraction of relevant information from the fused data. Consequently, pre-processing and post-processing stages might be needed. Despite the numerous data fusion proposed methods in the past few decades that focuses on approaches such as feature extraction, space embedding, data modeling, domain adaptation, data transformation, transfer learning, and image-to-image translation, structural analysis induced by graphs have not been widely explored. More precisely, the current graph-based fusion algorithms have shown their ability to cope with the variability of data formats and provide a flexible way of representing the relationship between data entities. Nonetheless, the graph-based fusion methods do not exploit prior information embedded in the data (i.e. graph signal processing). In addition, they are highly impacted by the image representation (i.e. pixels as a node, super-pixels as a node, or patches as nodes), and the fusion rule commonly relies more in the weighted adjacency matrix rather than the spectral basis.

As consequence, four approaches were proposed to address the previous problems in graph-based

data fusion. First, this thesis proposes a graph-based framework of data fusion for remotely sensed images applied to (i) data-driven semi-unsupervised change detection (CD) and (ii) biomass estimation in rice crops. The main contribution of this study was a "data-driven" framework used to capture unique information from multi-temporal, and multi-modal/heterogeneous images in a fused graph. The following stages implement this process: 1) Construct an approximate local graph related to remotely sensed images (such as an MS image captured by Landsat/UAV) by using the Nyström extension; 2) perform data fusion over the local graphs by minimizing the similarity between the connections of the graphs to capture relevant information embedded in the case studies; 3) extract different relationships given in the fused data by computing the eigenvectors/eigenvalues; 4) evaluate the performance of the proposed graph-based data fusion in the applications of change detection and biomass estimation. The performance in the CD application of the proposed method was compared against four state-of-the-art models and validated in fourteen datasets, where our proposed method outperformed the competing approaches in eight out of fourteen cases. For the biomass estimation, we were able to generated a unique model over the full growth stage of rice crops and outperformed the prediction of biomass in rice crops with respect to a recent state-of-the-art method.

Secondly, we extended the graph-based data fusion framework, by adding a new sampling technique known as blue-noise sampling. This is combined with a graph signal processing (GSP) approach to enhance the estimation given by the Nyström extension of the graph, where, the eigenvector from the graphs are used as the features extracted from the multi-spectral images. The proposed methods has 4 stages: 1) We down-sampled the image obtained by the segmentation algorithm GFKuts (i.e. which contains probabilities of the pixels that belong or do not belong to the crop), the down-sampling is conducted by dividing the images into uniform square patches; 2) we learn the graph by using the prior of smoothness, where the nodes are the vectorized version of each patch in the image; 3) we apply the blue-noise sampling over the smoothed graph and inject these samples to the graph based fusion framework; 4) use as feature the eigenvector with the highest eigenvalue of the fused-graph; 5) apply the t distributed stochastic neighbor embedding (t-SNE) to reduce the dimensionality of the feature. As a result, the proposed method to select relevant samples by applying GSP techniques improved the estimation of the above-ground biomass compared with two state-of-the-art methods.

Next, we propose a graph-based framework for CD. Similar to the previous approach the graph learning is driven by a GSP approach based on prior signal smoothness. However, we use the smoothed graph directly for the CD rather than using it to extract samples. In addition, we

proposed a false RGB image to generate super-pixels to be used as nodes, and a new optimization problem to find the change map that relies only on the normalized graph Laplacian of the fused graph. Unlike previously proposed methods, to detect the change map we do not require spectral decomposition, instead we use all the spectral basis of the fused graph. We compared against eight state-of-the-art methods and our method outperformed in ten out of the fourteen datasets used to test the performance of the models. These results support that the proposed method is more efficient and stable than the comparison methods.

Finally, a non local block-matching (NL-BM) to generate super-pixels as nodes and a novel optimization problem for subspace fusion from graphs to detect the change map were proposed. The proposed method has two main stages: 1) NL-BM, to generate super-pixels iteratively based on the cosine distance from uniformly distributed patches over the image; 2) an optimization problem, that does not rely on a prior knowledge directly, instead it uses a penalty term over the sparsity of the desired change map, and fused the subspace by the graphs. We tested the performance of the model in twenty cases against eleven state-of-the-art methods and outperformed them in ten out of twenty datasets.

Overall, limited research has been conducted in graph-based data fusion applied to remote sensing tasks. This thesis deepens in the learning of GSP to enhance the learned graph from remote sensing images, optimization problem to address the fusion rule, and a proper representation of a remote sensing image as a graph by using super-pixels as nodes. The results demonstrate the importance and the impact of using approaches based on structural analysis (graphs) for remote sensing applications such as CD and biomass estimation. Hence, this thesis advances further research on graph-based data fusion. In addition two databases [1] (see Section 3.4 and Section 5.1.4) with diversity of cases were generated to evaluate the performance of models in change detection.

Keywords

Biomass estimation, block matching, change detection, graphs, graph fusion, graph signal processing, optimization.

[1] Available at: https://github.com/DavidJimenezS

Resumen

Recientemente los avances en la tecnología de sensores han conllevado a un incremento en la disponibilidad de imágenes (a una alta resolución espacial y espectral) hiper-espectrales, multi-espectrales (ME), y de radar de apertura sintética (SAR), las cuales permiten describir un objeto o un fenómeno. Cada sensor captura diferente información que explica distintas características físicas. Por ejemplo, un sensor SAR captura información relacionada a características de la superficie (como la aspereza, estructura geométrica, y orientación), y un sensor ME captura la reflectancia de los objetos a diferentes longitudes de onda. Sin embargo, la información obtenida en tierra por un solo sensor es limitada para sacar conclusiones confiables sobre algún fenómeno como la detección de cambios en la cobertura del suelo y el crecimiento de biomasa. Por el contrario, las técnicas de fusión de imágenes integran información espectral, espacial y temporal de diferentes sensores con el fin de obtener información apropiada y generar imágenes adecuadas para la percepción humana y de maquina. La fusión de imágenes es el proceso de combinar dos o más imágenes en una sola, la cual debe de ser más informativa y por lo tanto útil en diferentes aplicaciones de sensado remoto (i.e. geología, agricultura, miliar, etc). Por ende, generalmente es deseado el uso de datos capturados por diferentes sensores.

Aunque la fusión de datos contribuye a mejorar el desempeño en tareas de clasificación y detección en sensado remoto, es una tarea que es compleja. Por ejemplo, las diferentes resoluciones, unidades, dimensiones, y formatos son retos impuestos por los datos sin procesamiento alguno. Además, los datos homogéneos (i.e. datos captados por el mismo sensor) presentan pequeñas variaciones intra-clase y distorsiones por brillo (artefactos), para el caso de datos heterogéneos (i.e. datos captados por diferentes sensores) los pixeles poseen diferentes firmas y por lo tanto siguen un comportamiento estadístico diferente lo cual dificulta la extracción de información relevante de los datos fusionados. En consecuencia, puede ser necesario el uso de pre-procesamiento y post-procesamiento. A pesar de que numerosos métodos propuestos en las últimas décadas para la fusión de datos que se enfocan en la extracción de características, embebimiento de espacios, modelamiento de datos, adaptación de dominio, transformación de datos, aprendizaje por transferencia, y traducción de imagen a imagen, el análisis estructural inducido por los grafos no ha sido ampliamente explorado. Más precisamente, los actuales algoritmos de fusión basados en grafos han mostrado su habilidad para lidiar con la variabilidad que presentan el formato de los datos y han permitido de una manera flexible representar la relación entre entidades de datos. No obstante, los métodos de fusión de datos basados en grafos no explotan la información prior embebida en los datos (i.e. procesamiento de

señales en grafos), son altamente impactados por la forma de representar una imagen (i.e. pixeles como nodos, super-píxeles como nodos, y parches como nodos), y la regla de fusión que se utiliza usualmente depende más de la matriz de pesos que de las bases espectrales.

Cómo consecuencia, se propusieron cuatro métodos para abordar los problemas descritos anteriormente en la fusion de datos basada en grafos. Primero, se propuso una metodología basada en grafos para fusionar datos provenientges de imágenes de sensado remoto y se aplico a (i) la detección de cambios (DC) semi-no-supervisada orientada por datos y (ii) la estimación de biomasa en cultivos de arroz. La contribución principal de este estudio es el uso de una metodología "orientada por datos" para capturar información única en un grafo fusionado de imágenes multi-temporales, y multi-modal/heterogeneas. El proceso se lleva acabo por las siguientes etapas: 1) La construcción de un grafo local aproximado relacionado a las imágenes de sensado remoto (como lo es una imagen MS capturada por el sensor Landsat o UAV) a través del uso de la extension de Nyström; 2) realizar la fusión sobre los grafos locales a través de la minimización de las similitudes entre las conexiones de los grafos con elf in de capturar la información relevante que está embebida en los casos de estudio; 3) extraer diferentes relaciones dadas en el grafo fusionado a través del calculo de los eigenvectores/eigenvalores; 4) evaluar el desempeño de la metodología propuesta para la fusion de datos basada en grafos en las aplicaciones de detección de cambios y estimación de bioamsa. El desempeño de la metodología propuesta en la DC se comparo contra cuatro métodos del estado-del-arte y se valido en catorce casos de estudio, de los cuales nuestro método propuesto obtuvo el mejor desempeño en ocho de los catorce casos de estudio. Para la estimación de biomasa, se genero un único modelo relacionado al crecimiento de la biomasa en todas sus etapas en el arroz y se supero la predicción con respecto al método más reciente del estado-del-arte.

En segundo lugar, extendimos la metodología de fusión de datos basada en grafos añadiendo una nueva técnica de muestreo conocida como muestreo de ruido-azul. Esto se combino con un método de procesamiento de señales en grafos (PSG) para mejorar la estimación del grafo generada por la extensión de Nyström, donde, los eigenvectores del grafo son utilizados como las características extraídas de las imágenes multi-espectrales. El método propuesto consta de 4 etapas: 1) Se submuestra la imagen proveniente de la segmentación de GFKuts (i.e. la cual contiene las probabilidades de los pixeles de pertenecer o no al cultivo), el submuestreo se realiza a través de la división de la imagen en parches cuadrados uniformes; 2) aprendemos el grafo a partir del uso del prior de suavidad, donde los nodos son la versión vectorizada de los parches en la imagen; 3) se aplica el muestreo de ruido-azul sobre el grafo suavizado y se inyectan las muestras a la metodología de fusión de datos basada en grafos; 4) se utilizan como características el eigenvector con el mayor valor

propio asociado del grafo fusionado; 5) se aplica la distribución estocástica t para embeber vecindarios (t-SNE) con el fin de reducir la dimensión de las características. Como resultado, el método propuesto selecciona muestras relevantes a partir de la aplicación de técnicas de PSG mejorando la estimación de biomasa sobre-tierra comparada con respecto a dos métodos del estado-del-arte.

Posteriormente, se propuso una metodología para la fusión de datos basada en grafos para la DC. Donde, de forma similar como las anteriores aproximaciones el aprendizaje del grafo es guiado por un método de PSG basado en el suavizado de la señal prior. Sin embargo, hacemos uso del grafo suavizado directamente en la DC en lugar de utilizarlo para extraer muestras. Adicionalmente, se propuso una imagen RGB falsa para generar super-pixeles, para posteriormente ser utilizados como nodos, y también se propuso un nuevo problema de optimización para encontrar el mapa de cambio que se encuentra en el Laplaciano normalizado del grafo fusionado. A diferencia de los métodos propuestos anteriormente, para detectar el mapa de cambio no requerimos de la descomposición espectral del grafo, por el contrario, hacemos uso de todas las bases espectrales del grafo fusionado. Nos comparamos contra ocho métodos del estado-del-arte y nuestro método supero a los demás en diez de los catorce casos de estudio utilizados para evaluar el desempeño de los modelos. Los resultados soportan que nuestro método propuesto es más eficiente y estable que los métodos utilizados como comparación.

Por último, se propuso un emparejamiento de bloques no-local para generar nodos como super-pixeles y un novedoso problema de optimización para detectar el mapa de cambio a partir de la fusión de los subespacios generados por los grafos. La metodología propuesta tiene dos etapas principales: 1) Emparejamiento de bloques, para generar super-pixeles de forma iterativa basado en la distancia del coseno entre parches cuadrados uniformemente distribuidos en la imagen; 2) un problema de optimización, que no depende directamente de un conocimiento a priori, por el contrario, hace uso de un término de penalización sobre la dispersión del mapa de cambio deseado, y fusiona los subespacios generados por los grafos. Evaluamos el desempeño del modelo propuesto en catorce casos de estudio contra once métodos del estado-del-arte y los superamos en siete de los catorce casos de estudio.

En general, se ha investigado de manera limitada modelos de fusión basada en grafos aplicados a tareas de sensado remoto. Esta tesis, profundiza en el uso de GSP para mejorar el aprendizaje de grafos de imágenes de sensado remoto, en la regla de fusión de datos a través de un problema de optimización, y la apropiada representación de imágenes a través de super-pixeles como nodos del grafo. Los resultados demuestran la importancia y el impacto del uso de aproximaciones basadas en el análisis de estructuras (grafos) para aplicaciones de sensado remoto como DC y estimación

de biomasa. Por lo tanto, esta tesis avanza en la investigación de fusión de datos basada en grafos. Adicionalmente, se generaron dos bases de datos [2] (ver Sección 3.4 y Sección 5.1.4) con diversos escenarios para evaluar modelos para la detección de cambios.

Palabras clave

Detección de cambios, emparejamiento de bloques, estimación de biomasa, fusión de grafios, grafos, optimización, procesamiento de señales en grafos.

[2] Available at: https://github.com/DavidJimenezS

1 Introduction

1.1 Problem Statement

Recent advances in sensors technology have lead to an increased availability of hyper-spectral, multi-spectral and synthetic aperture radar (SAR) images at very high spatial and spectral resolutions [Iyer et al., 2017]. These sensors capture information from the earth by sensing energy in different portions of the electromagnetic spectrum [Bai et al., 2015]. Nonetheless, the ground information obtained from a single sensor is limited to draw reliable conclusions about phenomena such as changes in land cover and biomass growth. An exciting direction of research is the image fusion or in general data fusion, which is a process of combining two or more images (i.e. integrate spectral, spatial, and temporal information from several sensors) into a single image that is more suitable for human and machine perception [Ghassemian, 2016]. Therefore, image fusion is useful and helpful for different remote sensing applications (i.e geology, agriculture, military, etc) [Kulkarni and Rege, 2020, Chawla et al., 2020, Karthikeyan et al., 2020, Lechner et al., 2020]. However, depending on the nature of the data, the fusion process presents some issues. For instance, in homogeneous data present small intra-class variability and brightness distortions (artifacts), in the case of heterogeneous data pixels have different signatures, and therefore follows different statistical behaviors that difficult the extraction of relevant information from the fused data, consequently pre-processing and post-processing stages might be needed [Luppino et al., 2017, Liu et al., 2017b].

In recent years, graph-based fusion algorithms have shown in the last years their ability to cope with the variability of data formats and provide a flexible way of representing the relationship between data entities [Dong et al., 2019]. Unsurprisingly, graph-based approaches are also applied to the task of data fusion [An et al., 2017, Tong et al., 2017, Amiri and Jamzad, 2018, Kang et al., 2020]. For instances, in [Debes et al., 2014] the authors proposed a graph-based data fusion (GDF) method to couple data and dimension reduction for classification of multi-sensor imagery. The GDF [Debes et al., 2014] combined multiple feature sources through a fused graph. However, it requires large storage capacity and computational load to process large training data sets (i.e. for an image with size of 3000×2000 it is necessary 67 GB of RAM available). As a reasonable solution to this issue the authors of [Liao et al., 2015] proposed a local approach by using a sliding window. In [Iyer et al., 2017], the authors proposed an approximated global graph using Nyström extension to generate a graph for each modality in order to fuse RGB and LiDAR images. The authors segment the urban areas by applying K-means on the eigenvectors of the fused graph to classify the areas of interest. Furthermore, Graph signal processing (GSP) has encouraged the analysis of

signals that live naturally on irregular data kernels described by graphs such as multispectral (MS) images. Though a MS image contains pixels that reside on a regularly sampled $2-D$ grid, if one can design an appropriate underlying graph connecting pixels with weights that reflect the image structure at each band, then one can interpret the image (or image patches) as a signal on a graph, and apply GSP tools for processing [Cheung et al., 2018].

The most recent contributions in graph based image fusion [Debes et al., 2014, Liao et al., 2015, Luppino et al., 2019, Iyer et al., 2020] that rely on a common and simple approach for graph construction that is based on a Gaussian radius basis function kernel to quantify the similarity between data samples. Even though, this methods has achieved good performances at classification and fusion in remote sensing images, they are based purely on observations without any explicit prior in the model of the data, hence they may be sensitive to noise and have difficulty in tuning the hyper-parameters. A meaningful data model or accurate prior may, however, guide the graph-inference process and lead to a graph topology that reveals the intrinsic relationship among the pixels in the different modalities [Cheung et al., 2018, Kalofolias and Perraudin, 2019, Dong et al., 2019, Shekkizhar and Ortega, 2020] improving the data fusion process. Therefore, a main challenge is to define a model for the generative process or function F, so that it captures the relationship between the observed data X and the learned graph topology G (i.e. weighted, adjacency, or Laplacian matrices). Typically, such models often correspond to specific criteria to describe or estimate structures among the image pixels, e.g., models that put a smoothness assumption on the data, or represent features in spectral domain by using graph signal processing based on Fourier transforms in graphs [Perraudin and Vandergheynst, 2017, Ricaud et al., 2019].

These issues raise a series of questions such as:

1. How to define a generative process or function F such that it captures the relationship between the observed remote sensing data and the learned graph topology G (i.e. weighted, adjacency, or Laplacian matrices) to fuse multi-modal/heterogeneous images?
2. How to extract relevant features from a graph signal processing based image fusion framework?
3. How to validate or measure the performance of the proposed model?

These are the open questions that the development of this doctoral thesis intends to respond.

1.2 State of the art

This section aims at presenting the state-of-the-art of image fusion focusing in remote sensing images (multi-spectral, multi-temporal and multi-modal data).

1.2.1 Image Fusion

Image fusion can be classified according to the type of data acquisition and the fusion level. For data acquisition we have multi-view, multi-model, multi-temporal, and multi-focus modalities [Sroubek et al., 2010]. Fusion algorithms also operate at pixel, feature, and decision levels [Kulkarni and Rege, 2020]. Next, we describe in detail these fusion approaches.

1.2.1.1 Pixel level

Pixel level fusion techniques work directly on the pixel data of the image, by combining two or more geometrically registered images of the same scene into one more representative image [Ghassemian, 2016, Kulkarni and Rege, 2020]. It is useful for improvement in further processing [Piella, 2003], better visual appearance [Zhang, 2004], or enhancement in visual interpretation of data [Lewis et al., 2004]. In general, this fusion level requires the following stages: forward transform, coefficient fusion and inverse transform.

Literature surveys [Ghassemian, 2016, Kulkarni and Rege, 2020], classify pixel level fusion techniques such as: component substitution methods (CS), multiscale decomposition methods (MSD), hybrid methods and model-based methods.

Component substitution methods (CS): The CS techniques are based on the projection (forward transformation) of multi-spectral data into another space by a transformation that allows to separate spatial from spectral data. Then, the spatial data is replaced (coefficient fusion) by a high resolution spacial PAN/SAR data and project the data back to the original space (inverse transformation) [Ghassemian, 2016, Kulkarni and Rege, 2020]. Popular techniques based on CS are Principal Component Analysis (PCA) [Pohl, 1998], Gram-Schmidt (GS) Orthogonalization [Laben and Brower, 2000], Intensity-Hue-Saturation (IHS) transform [CARPER et al., 1990], Brovery Transform (BT), High Pass Filtering (HPF) [Palubinskas, 2013] and Ehlers fusion [Ehlers et al., 2010]. Nevertheless, the methods based on CS are highly dependent on the correlation between images and the spectral mismatch in the channels of different types of sensors, causing local differences between the images to be fused resulting in spectral distortion in the fused product [Ghassemian, 2016, Kulkarni and Rege, 2020].

Multi-resolution analysis methods (MRA): In these methods, source images to be fused are first decomposed at different scales and then fusion rules are applied to each of the decomposed image sub-bands at all levels and the fused sub-bands are inverted to generate a

fused image. Most common methods are average filter [Schowengerdt, 2006], multiresolution decompositions of the image based on Laplacian pyramids [Chandrakanth et al., 2014], or wavelet/contourlet [Pandit and Bhiwani, 2015]. Nonetheless, MRA may impair spatial fidelity of the details, by producing, in some cases, spatial artifacts or blurred areas. [Dalla Mura et al., 2015].

Hybrid methods: Hybrid methods combine component substitution and multi scale decomposition techniques to exploit their advantages and improve the spatial and spectral resolutions. Some proposed methods combine IHS/DWT [Huang et al., 2005], PCA/IHS and wavelet transform [Liu et al., 2015], and PCA/DWT [Kulkarni et al., 2019]. These methods achieve a better performance than the CS and MRA separately. It is important to keep the correlation between the levels (scale) of the MRA and their components, in order to improve the performance in the final fusion [Zhang et al., 2018, Kulkarni and Rege, 2020].

Model based methods: These models can be separated in variational models [Zhang and Yu, 2010] and sparse-representation based models [Zhang et al., 2018]. The model based methods are robust to misregistration problems, but present a high computational complexity [Kulkarni and Rege, 2020].

1.2.1.2 Feature level

Fusion at feature level carries out image processing, such as image segmentation, at a higher abstraction than pixel-based approaches. Some feature level fusion methods combine texture, shape, and spectral attributes [Mirzapour and Ghassemian, 2015]. Other examples include hierarchical segmentation results into markov random fields (MRF), learning based super resolution fusion [Golipour et al., 2015], and softmax regression-based feature fusion [Bai et al., 2014].

1.2.1.3 Decision level

Decision level fusion methods combine a set of decision images or label maps. The decision combine extracted information to reinforce common interpretation and resolve differences to ease a better understanding of the observed objects [Ghassemian, 2016, Imani and Ghassemian, 2020]. Recent methods have been proposed at this level such as Bayesian inference [Mahmoudi et al., 2014], adaptive decision fusion based on the local scale of the structure [Luo et al., 2012], and discriminant analysis in morphological feature space [Imani and Ghassemian, 2018]. Nonetheless, the selection of features sets or choice of classifiers containing complementary information and the high computation time due to implementation of multiple classification processes are difficulties of the decision fusion methods [Imani and Ghassemian, 2020].

1.2.1.4 Based on data acquisition

Multi-angular Image Fusion: Here the problem is how to fuse images with a variety of acquisition angles from a scene. This process implies significant differences in acquisition geometries, radiometry, atmosphere, and illumination. Two approaches have been proposed to solve this issue: one is to consider the images as different views of the same object (exploit difference in the sequence as features) [Longbotham et al., 2011]; another is to consider the images as a series of observations with spectral distortion to be removed [Pan and Yang, 2009].

Multi-temporal Image Fusion: This type of fusion is based on the classification of pixels whose intensities evolve over time. Nonetheless, this approach has issues such as the curse of dimensionality [Liu et al., 2005], misclassification leading to wrong assessments (especially when context or textural information is not considered) [Gamba et al., 2006], and the lack of representation capability to account for nonlinear relationship among pixels, and features at different times [Melgani and Serpico, 2003].

Multi-sensor Image Fusion: These algorithms fuse different sensor sources such as LIDAR and optical images into a common representation space [Ma and Fu, 2011].

1.2.2 Deep learning techniques

Since 2012, with the breakthrough of deep learning in image classification [Krizhevsky et al., 2012], image super-resolution and remote sensing image fusion based DL methods have been proposed [Liu et al., 2018, Shi et al., 2020a]. These approaches outperform traditional methods because they are able to characterize better the complex relationship between the inputs and the target images [Liu et al., 2018, Ma et al., 2019]. The pioneer approaches are focused on PAN-Sharpening (MS and PAN fusion) such as stacked auto-encoders (SAE) model to generate a direct mapping among the low and high resolution MS image patches [Huang et al., 2015], a deep metric learning approach to learn the refined geometric multi-manifold embedding via multiple stacked sparse auto-encoders [Xing et al., 2018], and a multiscale and multidepth CNN architecture [Shao and Cai, 2018]. Other approaches focus in Hyper-spectral (HS) and Multi-spectral (MS) fusion, in [Palsson et al., 2017] the authors proposed a DL-based HS and PAN image fusion method through a 3D-CNN, later a two-branch convolutional network is proposed [Yang et al., 2018], and [Dian et al., 2018] combine a deep convolutional network and a model-based approach. In general, DL approaches present issues such as the over-fitting of data when the training dataset is small, the large dataset size required to train properly the network, and the optimization of hyper-parameters [Ma et al., 2019].

1.2.3 Graph based methods

Since graph captures structural information related to an interaction between nodes, they had been widely used in computer vision and pattern recognition [Kandel et al., 2007]. More precisely, the graph theory for remote sensing has focused on image classification [Sawant and Prabukumar, 2018, Hassanzadeh et al., 2018]. Nonetheless, graph-based methods have also contributed to image fusion. For instance, in [Debes et al., 2014] the authors proposed a graph-based data fusion (GDF) method to combine data and dimension reduction for classification of multisensor imagery. The GDF [Debes et al., 2014] combined multiple feature sources through a fused graph. However, the GDF may cause some problems on storage resources and computational load especially when using large training data sets. Therefore, as a solution of this problem in [Liao et al., 2015] the authors proposed a local approach by using a sliding window. In [Iyer et al., 2020] was proposed an approximated global graph using Nyström extension to generate a graph for each sensor to fuse RGB and LiDAR images. The authors perform a segmentation of urban areas by using K-means on the eigenvectors of the fused graph and then classify the areas. The most recent approach presented in [Luppino et al., 2019], selects samples based on local graph analysis for each sensor and fuses these outcomes by a Frobenius norm to train a regression model and detect changes in heterogeneous data.

1.2.4 Change detection

Change detection (CD), in the context of remote sensing, involves analyzing two images of the same land area taken at different points in time, which are often referred to as pre-event and post-event satellite images. The aim is to detect patches where the land-cover type has changed between acquisitions [Mura et al., 2015] [ZhiYong et al., 2021]. Modern satellites are able to capture multispectral (MS) images (images at different frequency bands), for which the red and near-infrared (NIR) bands are normally the most effective at identifying events such as fires and floods [Liu et al., 2017a]. Satellite images may be affected by noise, variations in luminosity, and misalignment. These images thus need to be coregistered before carrying out a comparison between them [Ding et al., 2010, Vakalopoulou et al., 2015]. CD algorithms can demand high computational cost due to the size of the image.Therefore, it is important to develop algorithms that operate efficiently and quickly in order to assist with prevention and evaluation of the impact of catastrophes.

Over the years, numerous methods for CD have been developed, from the most traditional methods based on hypothesis testing (e.g., **KI** [Kittler and Illingworth, 1986], **rR** [Zanetti et al., 2015], **rrR-EM** [Zanetti and Bruzzone, 2018]) to the most recent, based on machine and deep image

translation [Shi et al., 2020b] (e.g., **CAN** [Niu et al., 2018], **U-HPT** [Luppino et al., 2019], **X-Net** [Luppino et al., 2021]). Probabilistic methods often require the definition of a suitable null-hypothesis and a threshold to detect whether a pixel has changed or not between acquisitions. Their accuracy is thus determined by the truthfulness of the hypothesis, and the estimation of the threshold. In practice however, obtaining a desirable level of accuracy may be difficult in heterogeneous settings [Luppino et al., 2019, Luppino et al., 2021], where both the pre- and post-event images come from different sensing mechanisms and are thus affected by different sources of uncertainty (e.g., speckle noise in SAR images, and brightness distortion in MS images). Although recent learning-based methods may be trained to handle heterogeneous settings, such methods can be limited in their ability to generalize over different datasets and sensors (i.e. due to the limited number of training data available). Furthermore, there are also deep learning (DL) approaches ([Lei et al., 2019b, Zhang et al., 2020a, Zhang and Shi, 2020]) based on convolutional neural networks (CNNs) that have demonstrated an excellent power of feature extraction from remote sensing images. For instance, in [Lei et al., 2019b] authors extract features at different levels from a pyramid pooling setup to enhance the landslides detection. In [Zhang et al., 2020a] a difference discrimination network (DDN) is proposed to fuse features from pre- and post-events to detect the change map between the images. Here the authors extract features from images with a shared structure parameter for a fair comparison in a common space. In addition, authors in [Zhang and Shi, 2020] proposed a feature difference network (FD-Net) guided by a change magnitude as a loss function. The output of the FD-Net is a change magnitude map where a threshold is applied to get the final change map. In general, the DL approaches are highly computational demanding and at some point, the network is a black box where the uncertainty of what is happening is high [Zhang and Shi, 2020].

Recently, a new class of change detection methods based on graphs have been proposed. Unlike the aforementioned approaches, graph-based change detection methods (e.g., **GBF-CD** [Jimenez-Sierra et al., 2020], **INLPG** [Sun et al., 2021d], **NPSG** [Sun et al., 2021c], and **IRGM** [Sun et al., 2021a]) require little supervision and have proven to be robust in handling data coming from heterogeneous image sources [Iyer et al., 2020, Sun et al., 2021d, Sun et al., 2021a]. Graph-based approaches often find a graph representation for both the pre- and post-event images. Next, the graphs generated are fused (or combined) into a single graph, which encodes the most relevant information from the input images [Iyer et al., 2020, Jimenez-Sierra et al., 2020, Sun et al., 2021d, Sun et al., 2021c, Sun et al., 2021a, Jimenez-Sierra et al., 2021a]. Using this type of graph, it is possible to discover the intricate clustering structure of the change map, by examining the eigenvectors of the kernel matrix associated with the graph. Although not explicitly stated by Jimenez-Sierra *et*

al. [Jimenez-Sierra et al., 2020], the accuracy of their approach appears to lie in the fact that the change map belongs within the span of the most significant eigenvectors, i.e., the ones associated with the largest eigenvalues of the fused kernel matrix. Despite the undoubted benefits of graph-based change detection, current methods [Iyer et al., 2020, Jimenez-Sierra et al., 2020, Sun et al., 2021d, Sun et al., 2021c, Sun et al., 2021a, Jimenez-Sierra et al., 2021a] find the graph representation of the input images by means of Gaussian kernel matrices. However, this does not necessarily provide the best representation of the images as stated by [Dong et al., 2019] [Mateos et al., 2019], because the Gaussian kernel does not exploit priors or models of the data, which makes it sensitive to noise and hyper-parameter tuning. It is therefore reasonable to suggest that if graph representation is enhanced, then the performance of the method would improve.

1.2.5 Biomass estimation

Monitoring biomass at larger crop scales requires remote sensing approaches for precision sampling, mostly based on unmanned aerial vehicle (UAV-driven) multispectral imagery [Yue et al., 2018a, Yue et al., 2018b, Jimenez-Sierra et al., 2020, Colorado et al., 2020]. Furthermore, the low cost and flexibility of UAVs have created in the last years new opportunities in precision agriculture and phenotyping. UAVs imagery provides relevant relationships between vegetation indices, photosynthetic activity and canopy structural properties [Harrell et al., 2011, Bendig et al., 2014, Honrado et al., 2017, Devia et al., 2019], creating advantages with respect to alternative method for gathering data such as manual collection or satellite image analysis [Honrado et al., 2017, Devia et al., 2019] in agriculture applications. For this purpose, Above-Ground Biomass Estimation (AGBE) methods have recently gained significant traction [Cheng et al., 2017, Yang et al., 2019, Xiao et al., 2020, Li et al., 2021a], since machine learning models can be trained by features extracted according to different plant reflectance wavelengths, namely vegetation indices (VIs). These VIs, are computed from multispectral imagery captured by UAVs or satellites. Although the combination of several VIs tends to avoid saturation issues with higher values of accumulated biomass [Gitelson et al., 2002], in some cases the estimation of the biomass (based on VI extracted features) requires independent models that must be characterized and calibrated dependent on the crop growth stages. Furthermore, VI features are highly affected by the genotype (rice variety) and external abiotic conditions [Arroyo et al., 2017, Yue et al., 2018a, Yue et al., 2018b, Ndikumana et al., 2018, Viljanen et al., 2018, Devia et al., 2019, Lin et al., 2020, Harrell et al., 2011, Campos et al., 2021].

Most of the AGBE methods reported in the literature require image segmentation algorithms to extract the VI formulas from the corresponding canopy [Yue et al., 2018a, Yue et al., 2018b, Devia et al., 2019, Jimenez-Sierra et al., 2020, Colorado et al., 2020]. Therefore, the accuracy of the

extracted VI-features strictly depends on the quality of the segmentation, where several factors such as image resolution, saturation and radiation noise, play an important role in achieving an accurate correlation between the segmented pixels (VI-features) with the biomass.

1.3 Research Hypothesis

Learning graphs (i.e. weighted, adjacency, or Laplacian matrices) from graph signal processing approaches to represent, fuse, and extract features from remote sensing images increase the predictive performance in biomass estimation and change detection.

1.4 Research Scope

This thesis aims to propose new methodologies based on graph signal processing to represent and fuse remote sensing images for the applications of biomass estimation in rice crops and change detection on the ground.

1.5 Document Outline

In detail, the rest of the thesis is structured as follows:

- Chapter 1 provides a brief introduction to the presented problem. We briefly presented the state-of-the-art in image fusion, graph-based fusion, change detection and biomass estimation.
- Chapter 2 presents the general and specific objectives accomplished with this research.
- Chapter 3 presents a proposed graph-based data fusion applied to CD and biomass estimation, compared to three and one state-of-the-art methods respectively. This chapter was published in [Jimenez-Sierra et al., 2020].
- Chapter 4 introduces GSP approaches to enhance the proposal of Chapter 3. This Chapter mainly describes the use of GSP for sampling, to learn a graph (i.e. weighted, adjacency, or Laplacian matrices) and a new optimization problem for CD. This chapter was published in [Jimenez-Sierra et al., 2021b, Jimenez-Sierra et al., 2022].
- Chapter 5 describes a novel way to generate super-pixels in a non-local way and also a new optimization problem that addresses both the graph fusion and the CD. In addition, we also applied the graph-subspace fusion to biomass estimation.
- Chapter 6 provides conclusions of the results obtained from the proposed approaches.
- Chapter 7 presents recommendations for further work, based on the findings and results.
- Chapter 8 summarizes the contributions made by this doctoral thesis.
- Chapter 9 presents the publications produced by the development of this thesis.

2 Objectives

2.1 General objective

To develop a graph signal processing based image fusion method for remote sensing applications.

2.2 Specific objectives

SO.1

To define a function F to capture the relationship between the observed remote sensing data and the learned graph topology G (i.e. weighted, adjacency, or Laplacian matrices) to fuse multi-modal images.

SO.2

To analyze and select relevant features from graph signal processing approaches for the graph based image fusion framework.

SO.3

To validate the effectiveness of features extracted from graph based image fusion in remote sensing applications such as change detection and biomass estimation in terms of performance metrics such as: accuracy, overall error, root mean square error (RMSE), precision, and recall.

3 Graph-based data fusion framework for remote sensing images

Since graphs explain the structural relationships between nodes (such as image pixels) and also capture local information related to data (such as radiometric similarities), this chapter shows the proposed graph-based data fusion approach that aims to:

- Construct an approximate local graph related to remotely sensed images such as an MS image captured by Landsat and UAV sensors by using the Nyström extension.
- Perform data fusion over the local graphs by minimizing the similarity between the connections of the graphs to capture relevant information (i.e changes in the landcover or features of rice crops) embedded in the case studies.
- Extract different relationships (i.e. different connections in the graph) given in the fused data by computing the eigenvectors/eigenvalues of the fused graphs.
- Evaluate the performance of the proposed graph-based data fusion in the applications of change detection and biomass estimation.

3.1 Graph-Based Data Fusion

MS images contain pixels that reside on a regularly sampled 2D grid. Thus, it is possible to interpret them as a signal on a graph with edges that connect each pixel in each band to its neighborhood of pixels. A graph is a nonlinear structural representation of data, defined by $G = (V, E)$, where G is the graph, V is a set of nodes, and E refers to the arcs or edges that explain the directed or undirected relationship between nodes. The edges have an associated weight of $w_{i,j}$, which quantifies how strong the relationship is between the nodes. The common measure used for each weight is a Gaussian kernel [Alfredo Caicedo Barrero, 2010]:

$$w_{i,j} = \exp\left(-\frac{d(V_i, V_j)^2}{\sigma^2}\right), \tag{1}$$

where $d(V_i, V_j)$ is the distance between two nodes and σ is the standard deviation of all $d(V_i, V_j)$. A common application for graphs is the embedding of G based on the Laplacian ($\mathbf{L}$) matrix into a space $\mathbb{R}^m$. That keeps the graph nodes as close as they were in the input space. In short, the embedding of a graph is given by the eigen problem $\mathbf{Ly} = \boldsymbol{\lambda}\mathbf{Dy}$ [Belkin and Niyogi, 2003], where $\mathbf{L} = \mathbf{D} - \mathbf{W}$, $\mathbf{W}$ is known as the adjacency matrix, or weights of the graph (each component is given by Equation (1)), and $\mathbf{D}$ is a diagonal matrix, its components being the degree of the node ($di = \sum_j w_{i,j}$).

As there is such a high number of pixels in an MS image, the computational cost of calculating the

full matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$ is extremely high (an image with a resolution of 1280×960 is equivalent to $N = 1{,}228{,}800$). To solve this problem, we apply the Nyström extension [Fowlkes et al., 2004] to find an approximation of $\mathbf{W}$ in significantly reduced time. To select points uniformly distributed across the image, n_s samples are selected by spatial grid sampling and re-indexing the matrix $\mathbf{W}$ as:

$$\mathbf{W} = \kappa_G \left(\begin{bmatrix} \mathbf{d_{AA}} & \mathbf{d_{AB}} \\ \mathbf{d_{AB}}^\top & \mathbf{C} \end{bmatrix} \right), \tag{2}$$

where κ_G is a Gaussian kernel, $\mathbf{d_{AA}} \in \mathbb{R}^{n_s\times n_s}$ represents the graph distances within the n_s sample nodes, $\mathbf{d_{AB}} \in \mathbb{R}^{n_s\times(N-n_s)}$ are the distances between the n_s sample nodes and the remaining $N - n_s$ nodes, and $\mathbf{C} \in \mathbb{R}^{(N-n_s)\times(N-n_s)}$ are the distances within the unsampled nodes. This method approximates $\mathbf{C}$ by choosing n_s samples uniformly distributed across the image from the dataset of size N ($n_s \ll N$), hence:

$$\mathbf{W} \approx \widehat{\mathbf{W}} = \kappa_G \left(\begin{bmatrix} \mathbf{d_{AA}} \\ \mathbf{d_{AB}} \end{bmatrix} \right). \tag{3}$$

Thus, the eigenvectors of the matrix $\widehat{\mathbf{W}}$ can be spanned by the eigenvalues and eigenvectors of $\kappa_G(\mathbf{d_{AA}})$. Solving the diagonalization of $\kappa_G(\mathbf{d_{AA}})$ (eigenvalues λ and eigenvectors $\mathbf{U}$: $\kappa_G(\mathbf{d_{AA}}) = \mathbf{U}^\top\mathbf{\Lambda}\mathbf{U}$), the eigenvectors of $\widehat{\mathbf{W}}$ can be spanned by:

$$\hat{\mathbf{U}} = \begin{bmatrix} \mathbf{U} \\ \kappa_G(\mathbf{d_{AB}})^\top\mathbf{U}\mathbf{\Lambda}^{-1} \end{bmatrix}. \tag{4}$$

Since the approximated eigenvectors $\hat{\mathbf{U}}$ are not orthogonal, as explained in [Fowlkes et al., 2004], to obtain orthogonal eigenvectors, we use $\mathbf{S} = \kappa_G(\mathbf{d_{AA}}) + \kappa_G(\mathbf{d_{AA}})^{-\frac{1}{2}}\kappa_G(\mathbf{d_{AB}})\kappa_G(\mathbf{d_{AB}})^\top\kappa_G(\mathbf{d_{AA}})^{-\frac{1}{2}}$. Then, by diagonalization of $\mathbf{S}$ ($\mathbf{S} = \mathbf{U_s}\mathbf{\Lambda_s}\mathbf{U_s}$), the final approximated eigenvectors of $\mathbf{W}$ are given by:

$$\hat{\mathbf{U}} = \begin{bmatrix} \kappa_G(\mathbf{d_{AA}}) \\ \kappa_G(\mathbf{d_{AB}})^\top\kappa_G(\mathbf{d_{AA}})^{-\frac{1}{2}} \end{bmatrix} \mathbf{U_s}\mathbf{\Lambda_s}^{-\frac{1}{2}}. \tag{5}$$

Fusion Stage

In this section, we present the fusion stage to integrate the multi-temporal data. We model each pixel as a node in the graph and assume that pre-event and post-event images contain the same number of elements and that they are co-registered. Figure 1 presents a diagram of the method explained in Algorithm 1, which processes each instance of band b and time k ($\mathbf{X}^{b,k}$) and the number of samples n_s as inputs. The output of Algorithm 1 for one instance of time of a selected band or bands $X^{b,k}$ corresponds to the approximate normalized adjacency matrix ($\widehat{\mathbf{W}}_N^{b,k}$) [Fowlkes

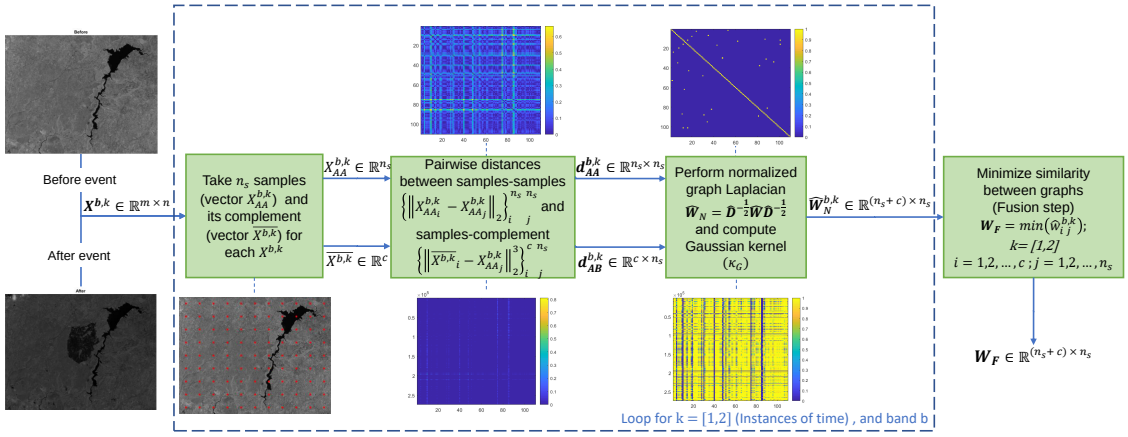


Figure 1: Graph-based fusion, where k is the time of Events 1 (pre) and 2 (post), b refers to the band, $X^{b,k}$ is an image that represents an event, $X_{AA}^{b,k}$ represents the samples from $X^{b,k}$, $\overline{X}^{b,k}$ is the complement, $\mathbf{d_{AA}}^{b,k}$ is the pairwise distance between the samples in $X_{AA}^{b,k}$, $\mathbf{d_{AB}}^{b,k}$ is the pairwise distance between $X_{AA}^{b,k}$ and $\overline{X}^{b,k}$, $\widehat{\mathbf{D}} = Diag(d_1, d_2, \ldots, d_{n_s})$ with $d_i = \sum_j^{n_s} \hat{w}_{ij}^{b,k}$ is the approximated degree matrix, and $\widehat{\mathbf{W}}_N^{b,k}$ is the normalized Laplacian calculated by using the Nyström approximation.

et al., 2004]. Consequently, the fusion step consists of capturing the unique information given by each graph ($\widehat{\mathbf{W}}_N^{b,k}$) into one fused graph ($\mathbf{W_F}$). In order to achieve this fusion, we maximize the distance (or minimize the similarity) among the pixels:

$$\mathbf{W_F} = min(\hat{w}_{N_{ij}}^{b,k}), \text{with } k = [1, 2],$$

where $w_{i,j}$ represents the weight of the node for each instance of time ($i = 1, 2, \ldots, c; j = 1, 2, \ldots, n_s$). This learning approach is data driven (uses a few uniformly distributed n_s samples across the image). It is restarted for each dataset. The graph $\mathbf{W_F}$ represents the relationship in terms of dissimilarity between the pre-event and post-event images.

3.2 Change Detection Scheme Based on the Multi-Temporal Graph (GFB-CD)

The change detection scheme presented in Figure 2 uses the approximated eigenvectors and eigenvalues found by Nyström's extension from $\mathbf{W_F}$, as features to represent the change between the pre-event and post-event images. The number of eigenvectors is equal to the number of samples (n_s) taken from an instance of time k. To build the change map, we select the scaled eigenvector ($\mathbf{I}_{u_i}$) that maximizes the mutual information (MI) [Cover and Thomas, 2012] of this eigenvector with a binarized prior signal. The prior signal comes from the normalized differences between pre-event and post-event images.

Algorithm 1: GBF for temporal data.

Input: Temporal images from band b or set of bands $\mathbf{X}^{b,k} \in \mathbb{R}^{m \times n}$, number of samples n_s
Output: Fused graph $\mathbf{W_F} \in \mathbb{R}^{(n_s+c)\times n_s}$
Initialize: $k = 1$, $N = m \times n$
while $k \leq 2$ **do**

(1) Scale the data by $\mathbf{X}^{b,k} = \frac{\mathbf{X}^{b,k}}{max(\mathbf{X}^{b,k})}$

(2) Take n_s samples uniformly distributed across $\mathbf{X}^{b,k}$ by spatial grid sampling.
$X_{AA}^{b,k} = \text{sampler}(\mathbf{X}^{b,k}, n_s), \in \mathbb{R}^{n_s}$

(3) Find the complement $\overline{X}^{b,k} \in \mathbb{R}^{c}$ of $X_{AA}^{b,k}$ in $\mathbf{X}^{b,k}$.

(4) For each set $X_{AA}^{b,k}$ and $\overline{X}^{b,k}$, perform the pairwise distance between samples-samples ($\mathbf{d_{AA}}^{b,k} \in \mathbb{R}^{n_s \times n_s}$) and samples-complement ($\mathbf{d_{AB}}^{b,k} \in \mathbb{R}^{c \times n_s}$).

$$\mathbf{d_{AA}}^{b,k} = \left\{ \left\| x_{AA_i}^{b,k} - x_{AA_j}^{b,k} \right\|_2 \right\}_{i \;\; j}^{n_s n_s}, \forall i \neq j$$

$$\mathbf{d_{AB}}^{b,k} = \left\{ \left\| \overline{x}_{i}^{b,k} - x_{AA_j}^{b,k} \right\|_2^3 \right\}_{i \;\; j}^{c \;\; n_s}, \forall i \neq j$$

(5) Apply the normalized graph Laplacian ($\hat{\mathbf{D}}^{-\frac{1}{2}} \widehat{\mathbf{W}} \hat{\mathbf{D}}^{-\frac{1}{2}}$) by using the code in [Fowlkes et al., 2004].

(6) Apply a Gaussian kernel ($\kappa_G(.)$) with $\sigma = mean(\mathbf{d_{AB}})$ on the normalized distances, and build the approximated normalized Laplacian matrix based on the Nyström approximation.

$$\widehat{\mathbf{W}}_N^{b,k} = \left[\kappa_G(\mathbf{d_{AA}}^{b,k}); \kappa_G(\mathbf{d_{AB}}^{b,k})\right]^\top$$

(7) $k = k + 1$

end

$\mathbf{W_F} = min(\widehat{w}_{N_{ij}}^{b,k})$, with $i = 1, .., c; j = 1, .., n_s$.

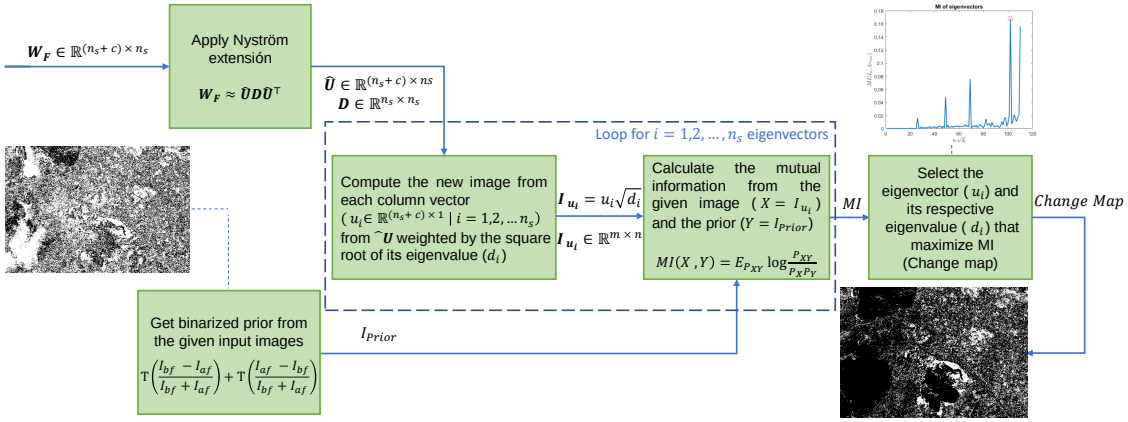


Figure 2: Change detection, where $\widehat{\mathbf{W}}_\mathbf{F}$ is the fused graph, $\widehat{\mathbf{U}}$ is the approximated eigenvectors, $\mathbf{D}$ is the eigenvalues, and T in the prior stands for a binarization operator.

3.3 Graph-Based Fusion Regression for Estimating Biomass in Rice Crops

In terms of image processing, the analysis of images related to crops implies important challenges. Weather conditions can affect the quality of the data (sunny or cloudy). Another important factor

is the appearance (architectural morphology) of the plant as it grows. The development of tillers occurs in the vegetative stage; the number of leaves increases, as well as the height of the plant. In this stage, the color green is predominant. In the reproductive stage, panicle formation starts, and thus, yellow features appear in the images. In the final ripening stage, the flowering, the grain filling, and the maturation of the plant occur, while the leaves begin to senesce. In this stage, the color yellow is predominant, and the plot can barely be distinguished from panicles, while grains and senescent leaves predominate. In conclusion, it is possible to observe (see Figure 5) that during a plant's growth, it becomes more difficult to separate plots and distinguish between plants and background, using RGB images. Therefore, general assumptions about the color, the size of the plant, and the color of the soil will not always be correct [Devia et al., 2019]. Considering these limitations, we believe that the graph-based fusion of MS bands provides a flexible way of representing useful combinations of surface reflectance, to produce features at two or more wavelengths that predict biomass in rice crops at different stages of growth. We developed our method inspired by the work in [Devia et al., 2019], in which the authors estimated rice biomass as a function of one of the growth stages (vegetative, reproductive, and ripening). They proposed three models of linear regressions, one for each stage of the crop. Those models have inputs that are features extracted from VIs. A comprehensive survey of the specialized literature was carried out, in order to identify which vegetation indices are suitable for estimating rice biomass as a function of the growth stage of the crop [Prabhakara et al., 2015, Gnyp et al., 2014, Kanke et al., 2016, Arroyo et al., 2017]. The results of this survey are summarized in Table 1:

Table 1: VIs for biomass estimation.

Name	**Equation**
Ratio Vegetation Index (**RVI**) [Kanke et al., 2016]	$\frac{NIR}{RED}$
Difference Vegetation Index (**DVI**) [Naito et al., 2017]	$NIR - RED$
Normalized DVI (**NDVI**) [Kanke et al., 2016]	$\frac{NIR-RED}{NIR+RED}$
Green NDVI (**GNDVI**) [Prabhakara et al., 2015]	$\frac{NIR-GREEN}{NIR+GREEN}$
Corrected Transformed Vegetation Index (**CTVI**) [Naito et al., 2017]	$\frac{NDVI+0.5}{\lvert NDVI+0.5\rvert}\sqrt{\lvert NDVI+0.5\rvert}$
Soil-Adjusted Vegetation Index (**SAVI**) [Prabhakara et al., 2015]	$(1+L)\left(\frac{NIR-RED}{NIR+RED+L}\right)$, with $L=0.5$
Modified SAVI (**MSAVI**) [Gnyp et al., 2014]	$\frac{1}{2}(2NIR)+1-\sqrt{(2NIR+1)^2-8(NIR-RED)}$

The following is a brief explanation of the procedure used by the authors in [Devia et al., 2019]: (i) segment the area covered by the crop from the soil by using k-means clustering ($K = 2$); (ii)

extract VIs (features) from the crop pixels; (iii) train a linear regression model for each stage of the crop.

Firstly, the bands ($\mathbf{X}^{b,k}$) that are to be fused are red (R), green (G), and near-infrared (NIR) ($b = [1, 2, 3]$). Secondly, there are n_s eigenvectors for each fused graph ($\mathbf{W_F}$). For each graph, we took the eigenvector with the associated highest eigenvalue, as it provides the strongest contribution to the Laplacian reconstruction. Thirdly, we stacked all these features as row vectors from each image into a matrix $\mathbf{F} \in \mathbb{R}^{q\times(n_s+c)}$. Fourthly, since there are $q = 489$ images with a size of 1280×960, the dimensionality of the features is high (≈ 1.2 million dimensions). Consequently, we reduced the number of features to z dimensions by applying two well-known techniques: principal component analysis (PCA) [Jolliffe and Cadima, 2016] and t distributed stochastic neighbor embedding (t-SNE) [Maaten and Hinton, 2008]. Lastly, we trained a support vector machine (SVM) regressor [Awad and Khanna, 2015] with a Gaussian kernel to predict the biomass over all growth.

Figure 3 illustrates our proposed method for estimating biomass in rice crops based on Algorithm 1 (setting $k = 1$ and $b = [1, 2, 3]$) and the graph-based fusion methodology shown in Figure 1. The procedure used to train the models is given in Algorithm 2 below. The outputs of Algorithm 2 are

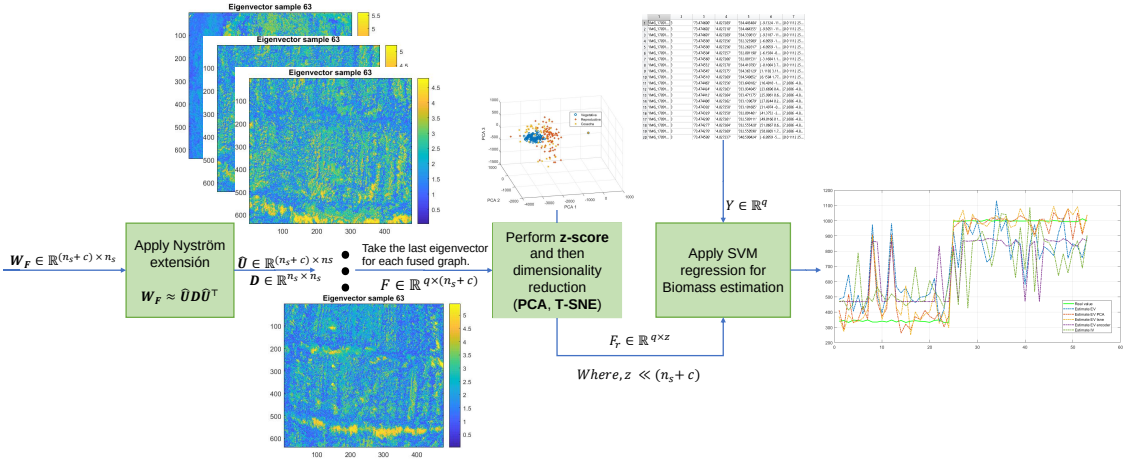


Figure 3: The proposed methodology based on graph fusion for estimating biomass in rice crops, from q images.

two regression models that predict the biomass related to an image of rice crops. These models use the reduced dimensions as inputs (such as PCA or t-SNE) of the eigenvectors from the fused graph with respect to the red, green, and near-infrared bands. The reason for applying the z-score is to avoid the high variability of features given for the entire growth stage of the rice crops and to decrease unstable biomass estimations.

Algorithm 2: SVM-regression models, trained on features extracted from the fused graph.

Input: Images of rice crops to train $\mathbf{I_i} \in \mathbb{R}^{m \times n}$, number of images q
, biomass related to the image $Y \in \mathbb{R}^q$
Output: Regression models: $\mathbf{M_{TSNE}}$ and $\mathbf{M_{PCA}}$
Initialize: $N = m \times n$, $i = 1$, $z = 16$
while $i \leq q$ **do**
  (1) Execute Algorithm 1 with $X^{b,k} = I_i$, with $b = [1, 2, 3]$, $k = 1$, and $n_s = 100$.
  (Output $\mathbf{W_F}^i$)
  (2) Take the eigenvector with the highest associated eigenvalue of the fused graph:
  $v = \hat{\mathbf{U}}^i(:, n_s) \in \mathbb{R}^N$.
  (3) Stack the vector as features for the regression $F(:, i) = v^\top \in \mathbb{R}^{1 \times N}$
  $i = i + 1$
end
(4) Apply the z-score to the features.
(5) Perform dimensionality reduction of features ($\mathbf{F} \in \mathbb{R}^{q \times N}$):
$\mathbf{F_{TSNE}} = tsne(\mathbf{F}, z) \in \mathbb{R}^{q \times z}$, $\mathbf{F_{PCA}} = pca(F, z) \in \mathbb{R}^{q \times z}$
(6) Train the models with SVM-regression:
$\mathbf{M_{TSNE}} = fitrsvm(\mathbf{F_{TSNE}}, Y)$, $\mathbf{M_{PCA}} = fitrsvm(\mathbf{F_{PCA}}, Y)$

Finally, the time complexity ($\mathcal{O}$) for both Algorithms 1 and 2 taking into account the upper-bound is given by the distance between samples and the complement (d_{AB}) and it is $\mathcal{O}(n_s^2(nm))$.

3.4 Datasets' Description

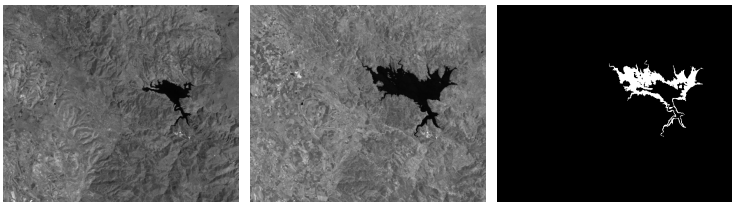
Here, we describe the datasets used to measure the performance of the proposed graph-based data fusion method. For the change detection application, we considered fourteen real scenes captured by MS and SAR sensors (as shown in Table 2 and Figure 4), which include events such as: earthquakes, floods, wildfires, melted ice, farming, and building. In addition, for the biomass estimation task, we used 560 UAV images with their corresponding value of biomass measured by the destructive method.

The authors of [Devia et al., 2019] provided a dataset that contains 321, 96, and 72 images, as well as biomass measurements for vegetation, reproductive, and ripening stages, respectively (see Figure 5). The biomass (ground truth) associated with each image is defined as follows: For each plot of the crop, one linear meter of the plant was cut from the ground. Plants were sampled and weighed (fresh weight), then put in an oven at 65 °C for four days, or until a constant weight was reached. Later, the vegetative biomass (leaves and stems) was separated from the reproductive biomass (panicles and grains). Both vegetative and reproductive biomass were then weighed (dry weight). The images were taken by a UAV equipped with the Tetracam ADC-lite multispectral

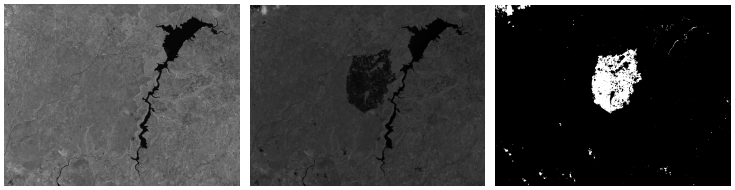
[3] Available at: SF_dataset

Table 2: Datasets to evaluate CD performance. The date format is dd/mm/yyyy.

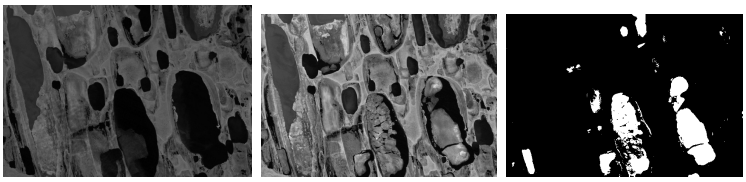
Place	Event	Pre-event Date	Post-event Date	Size	Pre-Band	Post-Band	Pre-Sensor	Post-Sensor
Italy, Mulargia lake	Flood	03/09/1995	03/07/1996	479 × 573	Nir	Nir	Landsat-5 TM	Landsat-5 TM
Italy, Omodeo lake	Fire	25/07/2013	10/08/2013	742 × 965	Red	Red	Landsat-8	Landsat-8
U.S., Alaska	Melting Ice	24/07/1985	13/07/2005	443 × 642	Nir	Nir	Landsat-5 TM	Landsat-5 TM
Brasil, Madeirinha	Farming/ Construction	15/07/2000	16/07/2006	364 × 527	Red	Red	Landsat-5 TM	Landsat-5 TM
Canada, Prince George	Fire	05/07/2017	22/08/2017	2479 × 1905	Nir	Nir	Landsat-8	Landsat-8
U.K., Gloucester-1	Flood	05/09/1999	17/11/2000	4220 × 2320	Nir	Nir	SPOT	SPOT
Colombia, Katios National Park	Fire	10/03/2019	27/04/2019	879 × 1319	SAR	SAR	Sentinel 1 A	Sentinel 1 A
Colombia, Atlantico	Flood (dam)	28/04/2010	16/03/2011	729 × 1056	SAR	SAR	ALOS/ PALSAR	ALOS/ PALSAR
U.S., San Francisco [3]	Flood	10/08/2003	16/05/2004	275 × 400	SAR	SAR	ERS-2 SAR	ERS-2 SAR
China, WenChuan	Earthquake	03/03/2008	16/06/2008	301 × 442	SAR	SAR	ESA/ ASAR	ESA/ ASAR
France, Toulouse [4]	Building	10/02/2009	15/07/2013	2604 × 4404	SAR	Nir	TerraSAR-X	Pleiades
U.S., California [5]	Flood	11/01/2017	26/02/2017	3500 × 2000	Nir	SAR	Landsat-8	Sentinel 1A
U.S., Bastrop [6]	Fire	08/09/2011	22/10/2011	1534 × 808	Nir	Nir	Landsat-5 TM	EO-1 ALI
U.K., Gloucester-2 [4]	Flood	14/06/2006	25/07/2007	4220 × 2320	Nir	SAR	Quickbird 02	TerraSAR-X



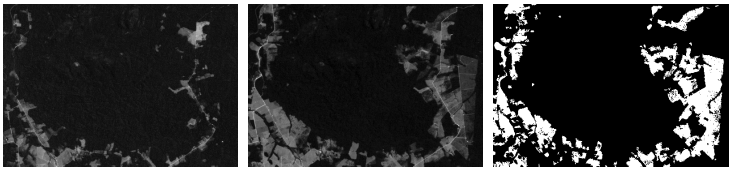
(**a**) Mulargia



(**b**) Omodeo



(**c**) Alaska

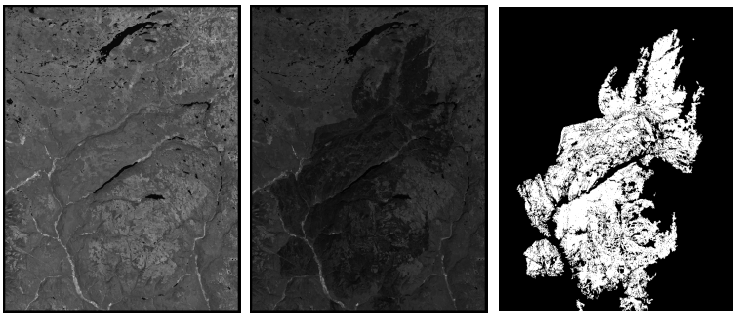


(**d**) Madeirinha

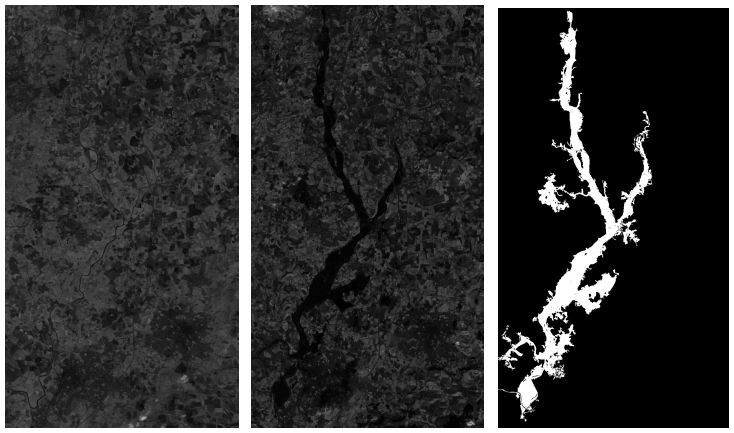
Figure 4: *Continued.*

camera capable of capturing images up to 72.26 mm/pixel in resolution flying at an altitude of 122 m. In our study, the UAV took images of the rice crops, flying over them at a steady altitude of 12 m above ground level (5.93 mm/pixel of resolution). The images (resolution of 1280 × 960) were

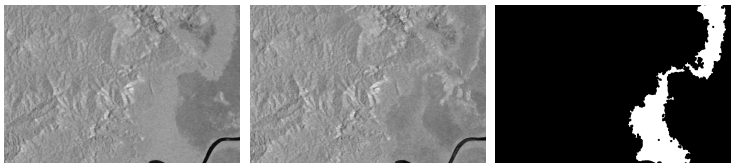
[4] Available at : M3CD
[5] Available at: california_dataset
[6] Available at: Bastrop_dataset



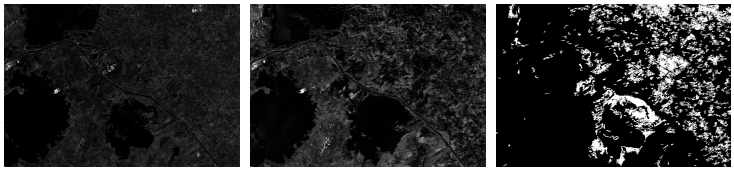
(**j**) Prince George



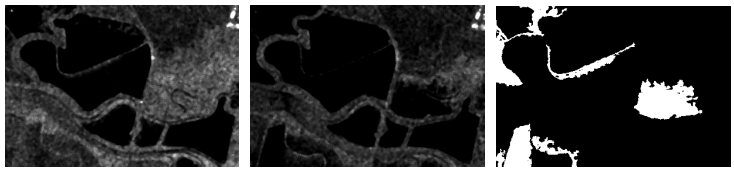
(**l**) Gloucester-1



(**e**) Katios National Park



(**f**) Atlantico



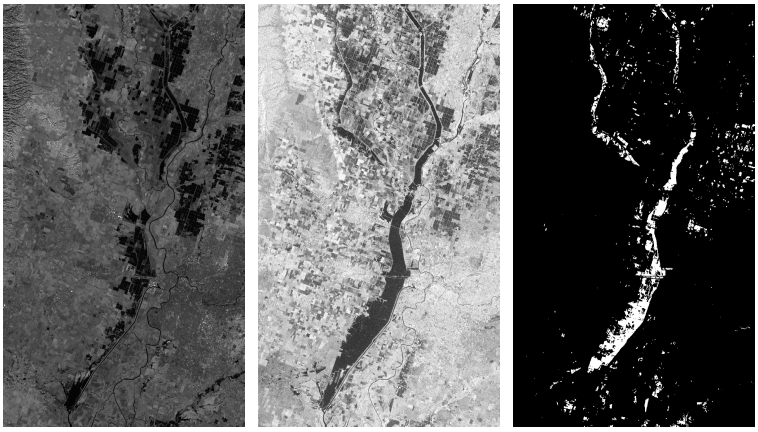
(**g**) San Francisco

Figure 4: *Continued.*

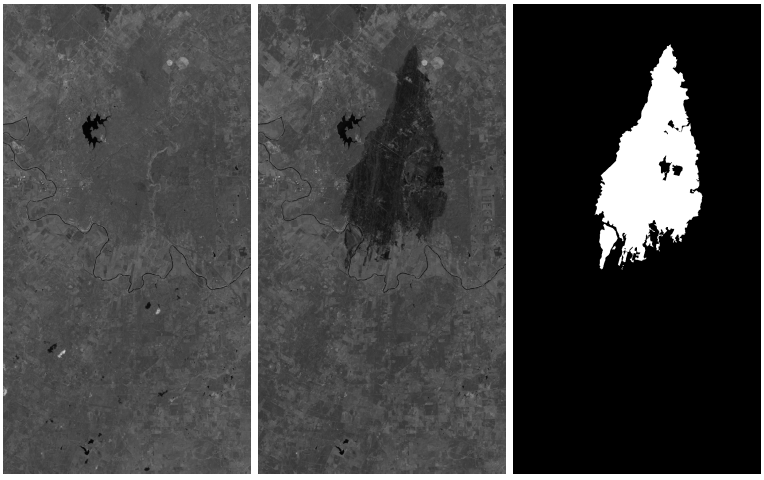
co-registered, and the bands used to extract the information from the crops were NIR, red, and green.

(**h**) WenChuan

(**i**) Toulouse

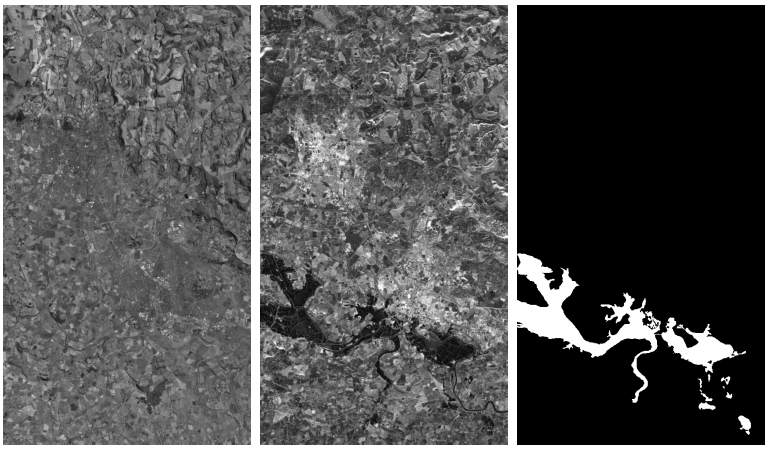


(**k**) California



(**m**) Bastrop

Figure 4: *Continued.*



(**n**) Gloucester-2

Figure 4: Datasets used in this study to test the proposed framework for change detection. From left to right: pre-event, post-event, and the reference change map.

Figure 5: Images from left to right represent the stages of the crop: vegetative, reproductive, and ripening, respectively for the genotype Tropical Japonica sub-species.

3.5 Experimental Setup

All the codes [7] ran in a server with two processors, Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz, with a total of 24 physical cores, 48 threads of processes, and 252 GB of memory RAM @2400 MHz.

3.5.1 Change Detection

The proposed graph-based fusion (GBF)-CD was compared with the classical Kittler–Illingworth (KI) [Kittler and Illingworth, 1986] and state-of-the-art techniques: Rayleigh-Rice (rR) [Zanetti et al., 2015], Rayleigh-Rayleigh-Rice (rrR) [Zanetti and Bruzzone, 2017], and unsupervised change detection using the regression homogeneous pixel transformation (U-CD-HPT) (code available at `https://github.com/llu025/Heterogeneous_CD`) [Luppino et al., 2019]. We evaluated each change map generated by all the methods with respect to the ground truth by using relevant metrics in

[7] to ensure the reproducibility of the proposed method, the code and all datasets are publicly available at: `https://github.com/DavidJimenezS/GBF-CD`

change detection such as: false negatives (FNs), false positives (FPs), precision (P, Equation (8)), recall (R, Equation (9)), Cohen's kappa (κ, Equation (6)), and overall error (OE), where the metrics FN, FP, and OE are measured in percentage with respect to the number of real change pixels, real non-change pixels, and all the pixels in the image, respectively. The method U-CD-HPT is the only one that requires a post-processing stage by filtering and thresholding to build the change map. The selected parameters for this post-processing stage were set according to the values presented by the authors in [Luppino et al., 2019].

The metrics are expressed as follows:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{6}$$

where p_o is the observed agreement between predictions and labels (the overall accuracy), while p_e is the probability of random agreement, which is estimated from the observed true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) as:

$$p_e = \left(\frac{TP + FP}{N}\frac{FN + TN}{N}\right) + \left(\frac{TP + FN}{N}\frac{FP + TN}{N}\right). \tag{7}$$

Precision and recall measure the agreement between the predicted and the real changed pixels as:

$$P = \frac{TP}{TP + FP}, \tag{8}$$

$$R = \frac{TP}{TP + FN}. \tag{9}$$

The number of samples (n_s) taken by spatial grid sampling and the standard deviations (σ) for the kernels were set through exhaustive grid-search using $MATLAB^{®}2017a$. Table 3 shows the parameter values of the proposed method for each database. According to the exhaustive grid-search show in Table 3, we recommend to use the number of samples in the range of 50 to 100.

3.5.2 Estimating Biomass in Rice Crops

The number of samples (n_s) was set to 100, and they were selected using a grid mesh on the image. We used cosine distance for t-SNE. For both t-SNE and PCA, the dimension z and the standard deviations (σ) for the kernels were set through exhaustive grid-search using MATLAB$^{®}2017a$, which gave us the dimensions $z = 16$. Table 4 shows the mean results of σ parameters for each stage of the crop.

In order to evaluate the performance of the proposed features for biomass estimation, we used cross-validation splitting the data into training (70%) and testing (30%) datasets. The model considers the whole growth stage of rice crops (vegetative, reproductive, and ripening). To measure the accuracy of the proposed features and the commonly used vegetation indices for biomass estimation,

Table 3: Parameters used for evaluation of the datasets. The superscripts 1 and 2 stand for pre- and post-event, respectively.

Database	n_s	σ^{1}	σ^{2}
Mulargia	93	2.5299×10^{-10}	1.5561×10^{-10}
Omodeo	93	2.7930×10^{-11}	1.6533×10^{-10}
Alaska	2	1.3720×10^{-9}	-6.7521×10^{-10}
Madeirinha	9	1.3841×10^{-5}	7.5380×10^{-9}
Katios National Park	60	1.0319×10^{-13}	-3.2947×10^{-15}
Atlantico	240	0.0012	-2.6971×10^{-6}
San Francisco [3]	4	8.3849×10^{-9}	7.5754×10^{-7}
WenChuan	39	-5.6319×10^{-8}	7.6359×10^{-7}
Toulouse	96	-8.9790×10^{-15}	-1.4351×10^{-14}
Prince George	110	-1.9516×10^{-12}	2.6925×10^{-9}
California [4]	270	-4.7062×10^{-14}	1.9471×10^{-16}
Gloucester-1	12	-3.5108×10^{-11}	-1.0611×10^{-10}
Bastrop	96	-1.2140×10^{-9}	-3.6741×10^{-11}
Gloucester-2	76	-7.7131×10^{-13}	-1.6947×10^{-14}

[3] Available at http://earth.esa.int/ers/ers_action/SanFrancisco_SAR_IM_Orbit_47426_20040516.html.
[4] Available at https://sites.google.com/view/luppino/data.

Table 4: Parameters used to evaluate the datasets. The superscripts 1, 2, and 3 stand for bands R, G, and NIR, respectively.

Stage	$\overline{\sigma}^{1}$	$\overline{\sigma}^{2}$	$\overline{\sigma}^{3}$
Vegetative	1.0490×10^{-14}	0.9850×10^{-14}	1.2650×10^{-14}
Reproductive	1.0290×10^{-14}	0.7080×10^{-14}	1.1260×10^{-14}
Ripening	1.1090×10^{-14}	0.7840×10^{-14}	1.4260×10^{-14}

we calculated the root mean squared error (Equation (10)):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}, \tag{10}$$

where y_i are for the real values of the biomass and $\hat{y}_i$ are the estimations of the model.

3.6 Results and Discussion

3.6.1 Change Detection

The visual comparison of the estimated change maps and the corresponding ground truths provide a qualitative assessment of the performance for each method.

Figure 38 in the supplementary material illustrates the resulting change maps for the same geographical area, in which each row represents a dataset and each column is one of these methods: KI [Kittler and Illingworth, 1986], rR-EM [Zanetti et al., 2015], rrR-EM [Zanetti and Bruzzone, 2017], U-CD-HPT [Luppino et al., 2019], and the proposed GBF-CD, respectively. The change maps that were obtained for all the methods show that the most challenging datasets were the

Katios National Park and Atlantico (see the fifth and sixth row in Figure 38). The images corresponding to pre- and post-events have similar variability in their pixel intensities. Therefore, the assumption of the probabilistic approaches [Zanetti et al., 2015, Zanetti and Bruzzone, 2017] (that the data follow a certain distribution for non-change and change pixels) does not hold. For both the thresholding algorithm (KI) [Kittler and Illingworth, 1986] and the unsupervised method based on image-to-image translation (U-CD-HPT) [Luppino et al., 2019], the estimated thresholding and the Frobenius distance between affinity matrices were unable to detect real change. This is because of the similarity between the distributions of change and no-change pixel intensities. In contrast, in the proposed GBF-CD method, the results came from building a fused graph (that minimized the similarities between the pixel intensities in the pre-event and post-event images) and from selecting an approximated eigenvector. This methodology maximizes the mutual information with a prior change map and yields change maps with lower false negative rates.

In terms of false negatives (FNs) and false positives (FPs), the probabilistic method (rR) [Zanetti et al., 2015] provided the worst performance. This was because the assumption of a large difference between pre-event and post-event images was not true in some of the test scenarios. The KI [Kittler and Illingworth, 1986] and rrR-EM [Zanetti and Bruzzone, 2017] algorithms classified all pixels in the San Francisco and California scenarios as belonging to the change category, producing zero FN and very high FP rates. In summary, the U-CD-HPT [Luppino et al., 2019] and GBF-CD methods provide a reasonable compromise between the correctly detected change pixels, FNs, and FP rates (see Tables 5–18, where the best performance with respect to the metrics is written in bold type).

Table 5: Performance of the models for the Mulargia dataset. OE, overall error.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	10.24	1.04	72.30	89.76	79.41	1.32	1.467
rR-EM [Zanetti et al., 2015]	**5.72**	4.01	41.73	**94.28**	56.05	4.06	9.881
rrR-EM [Zanetti and Bruzzone, 2017]	10.14	1.06	72.04	89.86	79.29	1.33	13.895
U-CD-HPT [Luppino et al., 2019]	9.03	2.00	58.12	90.96	69.84	2.20	107.978
GBF-CD	12.33	**0.17**	**93.96**	87.67	**90.43**	**0.53**	19.208

Table 6: Performance of the models for the Omodeo dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	**0.00**	3.42	59.04	**1.00**	72.62	3.26	4.850
rR-EM [Zanetti et al., 2015]	0.01	3.73	56.93	**1.00**	70.80	3.56	14.489
rrR-EM [Zanetti and Bruzzone, 2017]	0.01	2.14	69.73	**1.00**	**81.12**	**2.04**	9.928
U-CD-HPT [Luppino et al., 2019]	45.88	**0.55**	**82.90**	54.11	64.14	2.68	294.320
GBF-CD	77.00	1.26	47.26	22.99	28.73	4.83	91.624

Table 7: Performance of the models for the Alaska dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	14.13	3.57	74.23	85.86	76.98	4.70	1.424
rR-EM [Zanetti et al., 2015]	**8.07**	10.91	50.24	**91.92**	59.34	10.60	7.638
rrR-EM [Zanetti and Bruzzone, 2017]	12.52	4.81	68.51	87.48	73.68	5.64	8.322
U-CD-HPT [Luppino et al., 2019]	22.01	**0.15**	**98.38**	77.98	85.65	2.49	123.214
GBF-CD	11.66	0.87	92.36	88.34	**89.17**	**2.02**	3.623

Table 8: Performance of the models for the Madeirinha dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	**0.01**	10.44	69.47	**99.99**	76.70	8.44	1.347
rR-EM [Zanetti et al., 2015]	**0.01**	10.19	69.98	**99.99**	77.18	8.23	6.171
rrR-EM [Zanetti and Bruzzone, 2017]	40.31	1.32	91.45	59.69	67.27	8.81	16.320
U-CD-HPT [Luppino et al., 2019]	61.05	**0.11**	**98.78**	38.94	50.48	11.81	77.366
GBF-CD	24.44	1.13	94.06	75.56	**80.46**	**5.60**	4.100

Table 9: Performance of the models for the Prince George dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	**0.60**	16.15	70.23	**99.39**	**73.79**	**11.84**	2.575
rR-EM [Zanetti et al., 2015]	100.00	**0.00**	–	0.00	0.00	27.71	4.764
rrR-EM [Zanetti and Bruzzone, 2017]	54.01	1.13	93.93	45.99	53.22	15.79	723.778
U-CD-HPT [Luppino et al., 2019]	61.23	0.20	**98.61**	38.76	47.42	17.12	2075.130
GBF-CD	54.10	0.38	97.86	45.90	54.42	15.27	240.742

Table 10: Performance of the models for the Gloucester-1 dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	43.16	2.33	**69.62**	56.83	**59.44**	**5.85**	2.933
rR-EM [Zanetti et al., 2015]	99.99	**0.04**	0.03	0.00	-0.07	8.64	8.202
rrR-EM [Zanetti and Bruzzone, 2017]	**2.35**	44.06	17.27	**97.65**	17.24	40.47	24.540
U-CD-HPT [Luppino et al., 2019]	44.60	2.41	68.31	55.39	57.94	6.05	3808.564
GBF-CD	23.80	26.57	21.26	76.19	22.86	26.33	96.464

Table 11: Performance of the models for the Katios dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	67.88	5.87	39.20	32.12	28.51	12.42	1.769
rR-EM [Zanetti et al., 2015]	99.84	**1.18**	1.49	0.15	-1.72	11.60	4.013
rrR-EM [Zanetti and Bruzzone, 2017]	99.79	1.29	1.85	0.21	-1.79	11.69	4.083
U-CD-HPT [Luppino et al., 2019]	73.00	3.58	**47.03**	26.99	28.82	**10.90**	457.230
GBF-CD	**52.05**	10.63	34.74	**47.95**	**31.96**	15.00	34.481

Table 12: Performance of the models for the Atlantico dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	98.34	3.00	9.12	1.65	-2.03	17.72	1.652
rR-EM [Zanetti et al., 2015]	99.69	0.29	15.70	0.30	0.01	15.63	5.099
rrR-EM [Zanetti and Bruzzone, 2017]	99.93	**0.08**	11.62	0.06	-0.04	**15.49**	–
U-CD-HPT [Luppino et al., 2019]	99.13	0.28	36.01	0.86	0.97	15.53	333.742
GBF-CD	**30.42**	13.69	**48.11**	**69.57**	**47.26**	16.27	103.911

Table 13: Performance of the models for the San Francisco dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	**1.08**	63.16	18.55	**98.92**	12.54	55.28	1.315
rR-EM [Zanetti et al., 2015]	92.75	**0.59**	64.05	7.24	10.71	12.29	3.282
rrR-EM [Zanetti and Bruzzone, 2017]	2.19	61.23	18.85	97.80	13.11	53.73	3.813
U-CD-HPT [Luppino et al., 2019]	75.81	1.52	**69.62**	24.19	31.43	**10.92**	64.899
GBF-CD	48.82	7.64	49.34	51.17	**42.85**	12.87	3.213

Table 14: Performance of the models for the WenChuan dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	93.29	22.11	5.94	6.70	-14.67	34.38	1.380
rR-EM [Zanetti et al., 2015]	99.79	**1.07**	3.72	0.20	-1.41	**18.10**	3.318
rrR-EM [Zanetti and Bruzzone, 2017]	41.61	53.95	18.40	58.39	2.38	51.83	3.678
U-CD-HPT [Luppino et al., 2019]	99.69	2.06	3.00	0.30	-2.73	18.88	65.025
GBF-CD	**35.82**	22.52	**37.25**	**64.17**	**32.39**	24.81	6.235

Table 15: Performance of the models for the Toulouse dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	74.42	8.33	20.97	25.57	15.66	13.59	1.380
rR-EM [Zanetti et al., 2015]	74.94	8.11	**21.07**	25.05	15.59	13.43	3.318
rrR-EM [Zanetti and Bruzzone, 2017]	**52.32**	22.07	15.74	**47.67**	13.29	24.47	3.678
U-CD-HPT [Luppino et al., 2019]	98.30	**0.97**	13.11	1.69	1.20	**8.72**	4449.601
GBF-CD	54.27	17.33	18.57	45.72	**17.02**	20.27	839.940

Table 16: Performance of the models for the California dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	**0.17**	99.97	4.36	**99.83**	-0.01	95.61	2.910
rR-EM [Zanetti et al., 2015]	97.74	31.71	0.32	2.26	-7.66	34.60	9.989
rrR-EM [Zanetti and Bruzzone, 2017]	18.01	97.85	3.69	81.98	-1.43	94.36	25.521
U-CD-HPT [Luppino et al., 2019]	58.21	**2.79**	**40.59**	41.79	**38.45**	**5.21**	2955.937
GBF-CD	11.93	11.79	25.44	88.06	35.07	11.80	921.624

Table 17: Performance of the models for the Bastrop dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	73.30	99.16	3.10	26.69	-16.67	96.41	1.380
rR-EM [Zanetti et al., 2015]	100.00	**0.00**	–	0.00	0.00	10.63	3.318
rrR-EM [Zanetti and Bruzzone, 2017]	100.00	**0.00**	–	0.00	0.00	10.63	3.678
U-CD-HPT [Luppino et al., 2019]	**15.50**	0.39	96.17	**84.49**	**88.84**	2.00	365.296
GBF-CD	16.83	0.23	**97.71**	83.16	88.75	**1.99**	109.347

Table 18: Performance of the models for the Gloucester-2 dataset.

Model	FN (%)	FP (%)	Recall (%)	Precision (%)	κ (%)	OE (%)	Time (s)
KI [Kittler and Illingworth, 1986]	90.34	4.25	13.46	9.65	6.21	9.78	1.380
rR-EM [Zanetti et al., 2015]	96.29	2.33	9.80	3.70	1.92	8.36	3.318
rrR-EM [Zanetti and Bruzzone, 2017]	44.12	19.72	**16.26**	55.87	**16.93**	21.29	3.678
U-CD-HPT [Luppino et al., 2019]	98.36	**1.57**	1.63	6.63	0.08	**7.78**	3767.047
GBF-CD	**29.39**	27.71	14.86	**70.60**	15.62	27.82	543.650

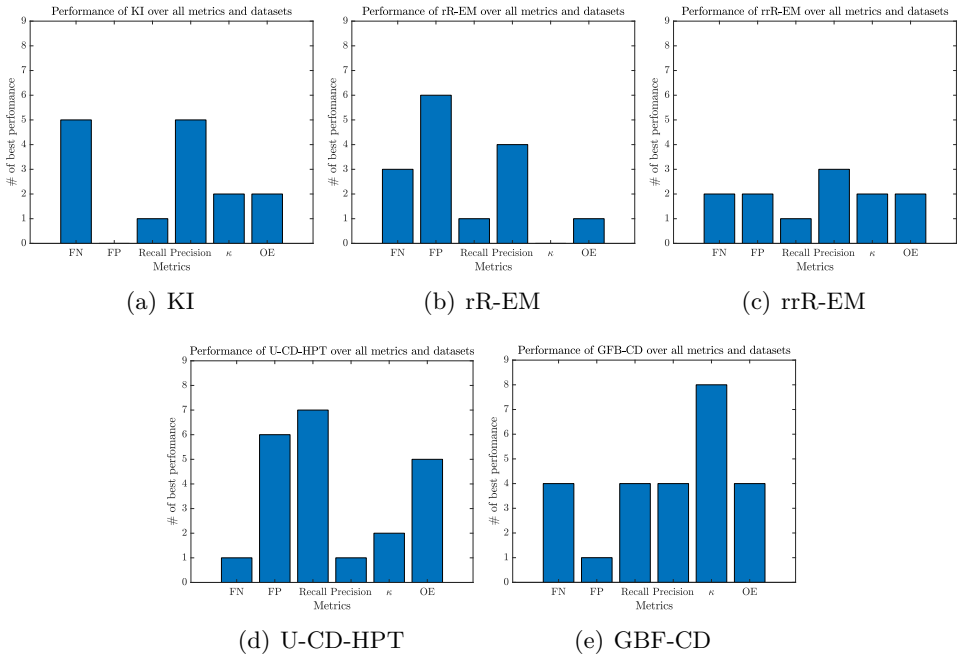
The Toulouse, California, Bastrop, and Gloucester-2 test scenarios are represented by NIR and SAR images. With regard to the Toulouse and Gloucester-2 datasets, the rrR-EM and GBF-CD methods yielded change maps with high TPs, low FNs, and high FP rates. In contrast, KI, rR-EM, and U-CD-HPT provided low TPs and high FN rates. In the case of the California dataset, the KI, rR-EM, and rrR-EM algorithms provided inaccurate change maps due to the fact that these methods were devised for processing homogeneous (one modality) input data. Despite the data heterogeneity, the U-CD-HPT [Luppino et al., 2019] and GBF-CD algorithms provided better performance in terms of FNs, FPs, and κ. Unlike the KI, rR-EM, and rrR-EM methods, the algorithms U-CD-HPT and GBF-CD used for the Bastrop dataset yielded an accurate change map.

To illustrate the relative performance of each CD method in all the challenging test scenarios, we counted the number of times a given CD method outperformed the competing algorithms in a specific performance metric (see Figure 6). We observed that the proposed GBF-CD method outperformed (in terms of κ) the competing algorithms in eight (Mulargia, Alaska, Madeirinha, Katios, Atlantico, San Francisco, WenChuan, and Toulouse) of the fourteen datasets. Moreover, the GBF-CD algorithm achieved the best performance metrics (FN, recall, precision, and OE) in four (Katios, Atlantico, WenChuan, and Gloucester-2) of the test scenarios. It also showed the lowest FP rate in one scenario (Mulargia). Overall, the proposed GBF-CD algorithm outperformed the comparison methods in at least one performance metric. In contrast, the KI, rR-EM, rrR-EM, and U-CD-HPT algorithms did not surpass other competing methodologies in at least one performance metric.

3.6.2 Biomass Estimation

Figure 7 illustrates the comparison of the biomass prediction results. This was achieved by applying the dimensionality reduction techniques t-SNE and PCA to the features extracted from the proposed graph-based fusion approach, in addition to the biomass estimation yielded by using the traditional VIs. These results show that the VI does not capture the biomass features during the growth of rice crops. In contrast, the regressor that was trained with the features obtained after applying the dimensionality reduction techniques provided better prediction results and lower estimation errors (as shown in Table 19).

Even though the proposed graph-based fusion features outperformed the traditional VIs for biomass estimation, there is still a need for further work; for instance, to decrease the computation time, as it currently takes approximately three hours to extract the features and train the models. It would also be advantageous to reduce the dependency of the performance on the number of selected



(a) KI (b) rR-EM (c) rrR-EM

(d) U-CD-HPT (e) GBF-CD

Figure 6: Bar charts that evaluate the performance of each method over all metrics and datasets. The count for each method in one of the six possible metrics means that in one dataset, the model outperformed all the competing methods in that metric.

samples n_s and the standard deviation σ and explore parameter tuning methods beyond exhaustive grid search. However, one regression model based on the proposed features predicted the biomass well, despite its variability during different growth stages of rice crops. This is a result of the fact that the graph-based features capture both radiometric and structurally useful information from the MS bands. In contrast, the VIs are not able to capture the biomass variability for rice crop growth, requiring three separate regression models [Devia et al., 2019].

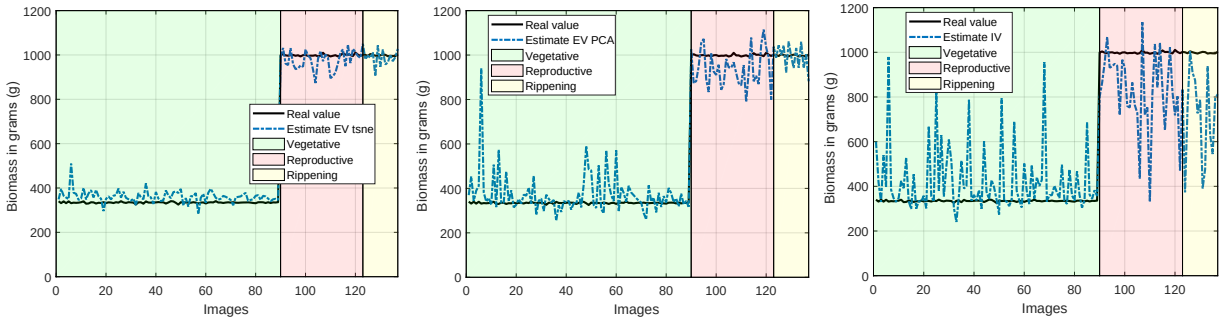


Figure 7: Regression performance by one model for all rice crop growth stages. From left to right, the models are: t-SNE, PCA, and vegetation indices (VIs).

Table 19: Performance of each model for biomass estimation in terms of RMSE.

	VI	PCA	t-SNE
RMSE	213.290	95.795	**40.273**

4 Graph signal processing for Graph-Based data fusion

In this section is describe an improvement of the previous proposed method explained in Section 3. In which, a smoothness prior for the graph learning is applied and then a sampling method in the graph space is made to generate more representative samples.

4.1 Feature extraction based on Graph-fusion with GSP for Biomass estimation

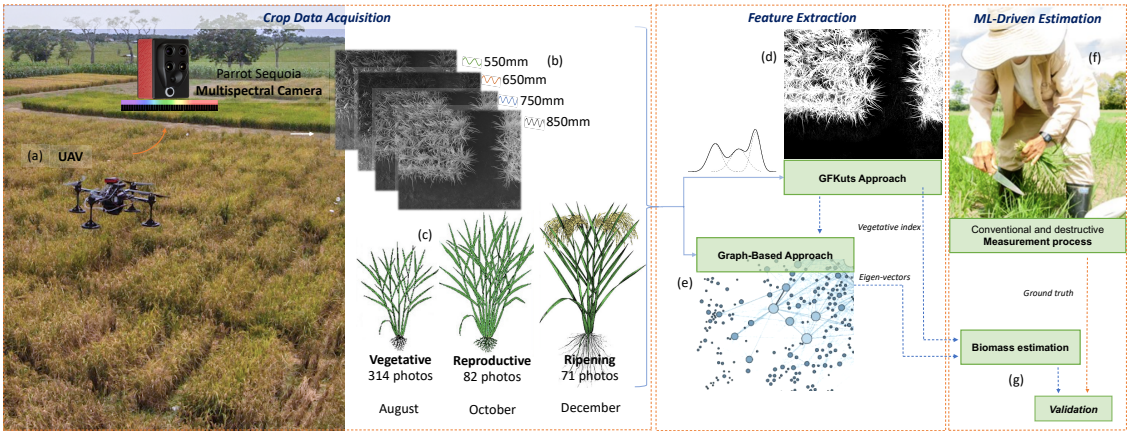


Figure 8: (a) UAV-driven remote sensing of above-ground biomass in rice crops. (b) Multispectral imagery. (c) Dataset amount & crop stages (d) First methodology - GFKuts. (e) Second methodology - Graph based. (f) Destructive biomass sampling. (g) Validation and correlation stage

The general architecture for UAV-driven remote sensing of above-ground biomass in rice crops is presented in Figure 8. An UAV is equipped with a multispectral camera onboard that captures the reflectance of the light spectrum in four different bands (Green, Red, Red-Edge, and NIR). A set of 1868 (i.e. one image per band) images was acquired for the crop's three growth stages: vegetative, reproductive, and ripening. Further details on the dataset acquisition and crop characteristics can be found in [Colorado et al., 2020].

As mentioned, the proposed method presents two approaches for feature extraction. The first approach is GFKuts, which is based on the image processing area. This approach is an entirely automatic proposal that integrates a binary classification technique with an optimization approach based on a Gaussian mixture model, followed by a filtering stage, to extract a mask related to pixels

that belong to the crop canopy. The second approach is a graph-based framework that aims to fuse data by taking advantage of the structural relationship between the data captured by independent graphs (i.e., the structural relationship of the pixels of an image with one graph per image channel). The method generates a resultant fused graph with the most relevant information. Unlike GFkuts, this method uses eigenvectors of the fused-graph as features rather than vegetation indices. Lastly, the extracted features obtained from both methods are the inputs to a regression model to estimate the above-ground biomass. The next subsections explain the features extraction approaches and the regressors.

4.1.1 GFKuts

Figure 9 details the proposed GFKuts approach. The first contribution of GFKuts is to apply a random sampling approach based on Monte Carlo method and integrate these samples into a binary classification method [Shapiro, 2003]. This random arrangement and the binary classification algorithm solve a time-demanding computational iteration problem as they provide an efficient stochastic numerical method based on a photogrammetric image technique used in precision agriculture.

An array $\mathbf{I} = (i_1, ..., i_n)$ of size N, denotes the input image in the standard RGB color space (sRGB). The binary non-supervised classification method implemented in this work is the k-means clustering algorithm that cluster the set of N samples into K groups. The semantic labeling of the groups is random according to the nature of the algorithm. Thus, there is no control of the true class of the label assigned. Given that the results are highly repeatable, the identification of each group is carried out through the characterization of the centroid. This characterization runs only once and operates as long as the conditions of the images do not change (e.g., the type of crop, growing, or environmental conditions).

The Montecarlo sample K-means classification strategy generates a selection of pixels with a uniform random distribution of the spatial coordinates $n_{(x,y)}$ of the image. The result is a subset of sRGB values for each selected pixel N. These values come from the refractive bands of the light spectrum captured by the sensor. The implementation of this first stage is shown in algorithm 3. The result is a pair of masks that allow us to initialize regions of the image to implement the optimization process.

The image segmentation consists of inferring the unknown opacity variables, denoted as α, from the given image data I and the model θ [Rother et al., 2004]. The opacity term $\alpha = (\alpha_1, ..., \alpha_N)$ is the image segmentation weighed by each pixel with $0 \leq \alpha_N \leq 1$ for soft segmentation and

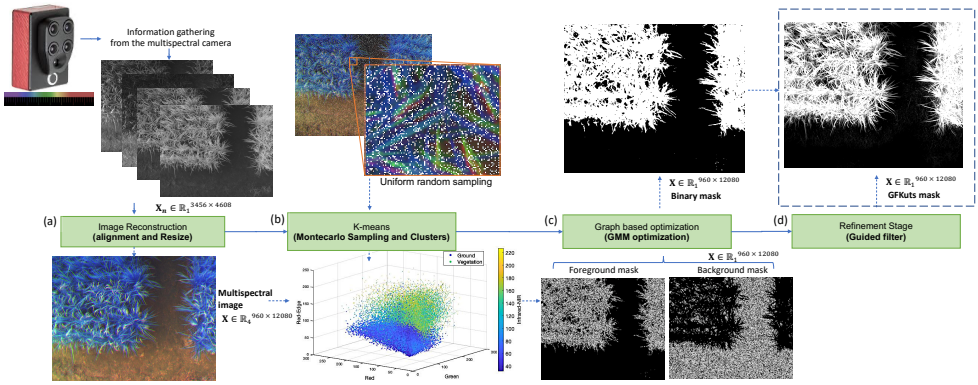


Figure 9: GFKuts approach. (a) Preprocessing stage. (b) Binary classification approach. (c) GMM modeling & optimization. (d) Filter stage.

Algorithm 3: Montecarlo Sampled K-means.
Inputs: the image **I**, the number of samples N, and two heuristic values associated with the mean expected radiance of the canopy: C_1, and the ground: C_2.

for Each pixel in range (1 ... n) **do**
  Random (x,y) pixel selection from **I** to P
  Store sRGB value from P to $\mathbf{Feature}_n$
  Store pixel coordinates $\mathbf{I}_{x,y}$
end for
Run K-means over $Feature$ to get labeling $\mathbf{I}_{n,k=1}$ and $\mathbf{I}_{n,k=2}$
calculate the Euclidean distance between C1 and the centroid of the group K = 1 to $m1$;
calculate the Euclidean distance between C1 and the centroid of the group K = 2 to $m2$;
if $m1 < m2$ **then**
  group K = 1 to *canopy*
  group K = 2 to *ground*
else
  group K = 1 to *ground*
  group K = 2 to *canopy*
end if
Create a mask T_F and set the coordinates in $\mathbf{I}_{x,y}$ of each pixel in $\mathbf{I}_{n,canopy}$ as the foreground.
Create a mask T_B and set the coordinates in $\mathbf{I}_{x,y}$ of each pixel in $\mathbf{I}_{n,ground}$ as the background.

$\alpha_N \in (0,1)$ for hard segmentation, with 1 denoted as foreground and 0 for background. The parameters θ describe image foreground and background distributions modeled by GMMs values, that introduces the covariance parameter $\mathbf{k} = (k_1, ..., k_N)$ with $k_n \in (1, ..., K)$ assigned for each pixel. The "Gibbs" energy function of Eq.11 denoted as **E**, models a trend of solidity of the objects through the opacity parameter, whose minimum corresponds to an optimal segmentation.

$$\mathbf{E}(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{I}) = U(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{I}) + V(\underline{\alpha}, \mathbf{I}). \tag{11}$$

The function $U()$ in Eq.11 is explained in Eq. 12; this function estimate the fitness of the opacity distribution α of the image, given the model θ, in which $p()$ is a Gaussian probability distribution and $\pi()$ are mixed weights coefficients.

$$U(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{I}) = \sum_n -logp(z_n|\alpha_n, k_n, \theta) - log\pi(\alpha_n, k_n) \tag{12}$$

The remaining function $V()$ shown in Eq. 13 is called the smoothness term. The constant γ is defined empirically as 50 by [Rother et al., 2004], given images in a 0 - 255 range. β is a constant that regulates the smoothness term.

$$V(\underline{\alpha}, \mathbf{z}) = \gamma \sum_{(m,n)} dis(m,n)^{-1}[\alpha_m \neq \alpha_n]\beta * exp(Z_m - Z_n)^2 \tag{13}$$

The segmentation can be estimated as a global minimum of α parameter over the energy model, as seen in Eq. 14

$$\alpha' = \arg\min_{\alpha} E(\underline{\alpha}, \theta) \tag{14}$$

After the optimization process, an image filtering stage is introduced. Linear and time-invariant filters are usually used. A powerful approach, is to optimize a quadratic function that directly imposes constraints on the unknown output. The solution is obtained by solving a dispersion matrix encoded with the information from the guide image [Petschnigg et al., 2004] [Correa et al., 2020]. Processing is done around each pixel using a weighted average of the nearby pixels. The processing information is based on the color and intensity of the guide image. [Petschnigg et al., 2004]. This proposal is developed with the bilateral filter, starting with the guided filter. The filter can preserve the edges and smooth out small fluctuations. [He et al., 2010]. This approach is based on explicitly constructing the cores using an orientation image. The output is a linear transform of the guide image. This filter has the edge-preserving and anti-aliasing property such as a bilateral filter, but does not suffer from gradient inversion problems [He et al., 2010] [Correa et al., 2020].

The filter stage has two inputs, a guide image 'I' and an input image 'p'. The output is expressed as a weighted average 'q'. The parameters i and j are the pixel indices, as seen in Eq.15

$$q_i = \sum_j W_{ij}(I)p_j \tag{15}$$

The guide image 'I' is the reference of the filter kernel W_{ij} independent of 'p'. The kernel weights can be explicitly expressed by Eq. 16

$$W_{ij}^{GF}(I) = \frac{1}{|\omega|^2} \sum_{k;(i,j)\in\omega_k} \left(1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma_k^2 + \epsilon}\right) \tag{16}$$

The parameters μ_k and σ_k^2 are the mean and variance of w_k in image 'I' respectively, ϵ is a regularization parameter and $\|w\|$ is the number of pixels in w_k [Correa et al., 2020].

4.1.2 Enhanced Graph-based data fusion with GSP

Multi-channel input images denoted as $\mathbf{X}_n^c \in \mathbb{R}^{m\times n}$, with $c = 1, \ldots, C_h$ and $n = 1, \ldots, N$, where C_h is the number of channels of each image and N is the total number of images. Here, the goal is to extract relevant features given by each channel of the images, in order to train a model that

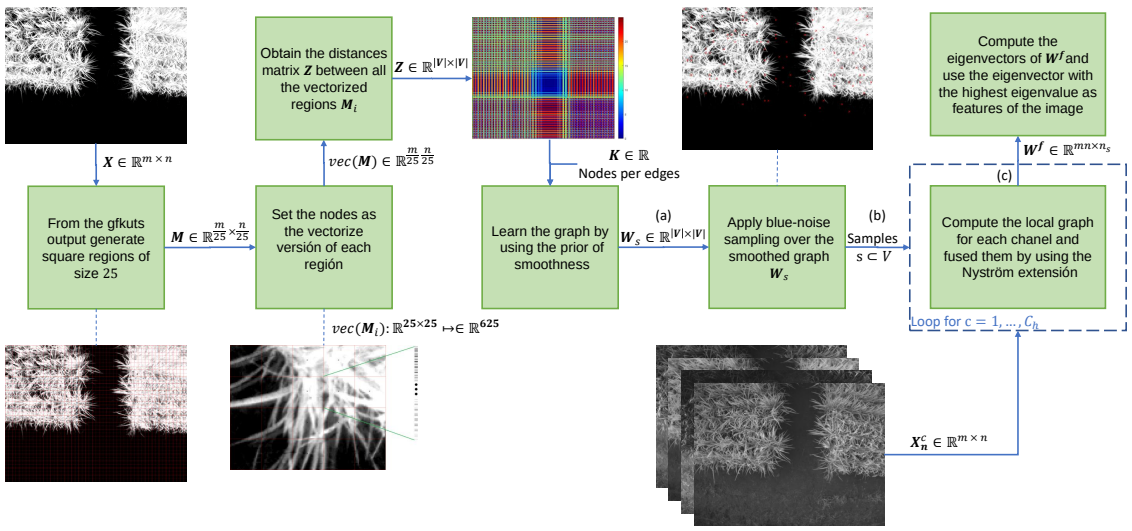


Figure 10: Proposed method based on the three stages: (a) Graph learning with prior smoothness [Kalofolias and Perraudin, 2019], (b) blue-noise sampling [Parada-Mayorga et al., 2019] to inject the samples to Nyström extension in [Jimenez-Sierra et al., 2020] and (c) the fusion of the the multispectral images to extract features of the crop.

captures the biomass behavior across the crop stages. As explained in [Jimenez-Sierra et al., 2020], each channel of images are represented by a corresponding graph, with the purpose of obtaining a single fused-graph from all the channels. The fused-graph is then used as an embedded space of images to extract relevant features.

The graphs used in the approach in [Jimenez-Sierra et al., 2020] are undirected, and they are denoted as a triplet $\mathbf{G} = (\mathbf{V}, \mathbf{E}, w)$, consisting of vertexes $\mathbf{V}$, and edges $\mathbf{E} \subset \mathbf{V} \times \mathbf{V}$, and a non-negative weight function $w : \mathbf{V} \times \mathbf{V} \mapsto [0, \infty)$. Therefore, a multichannel image $\mathbf{X}_n^c$ can be represented as graph $\mathbf{G}$ by defining a vertex set $\mathbf{V}$, an edge set $\mathbf{E} \subset \mathbf{V} \times \mathbf{V}$, where each edge represents the relationship between two pixels, and a weight function w that measures the strength of that relationship. Typically w is defined by:

$$w(v_i, v_j) = \exp\left(\frac{-\text{dist}\,(v_i, v_j)^2}{\sigma^2}\right), \tag{17}$$

where $\text{dist}(v_i, v_j)$ is the Euclidean distance between the nodes v_i and v_j associated to the pixels $\in \mathbf{X}_n^c$, and $\sigma > 0$ is a scaling parameter.

Given the graphs for all channels in the image ($\mathbf{X}_n^c$) $\mathbf{G}^c = (\mathbf{V}, \mathbf{E}^c, w^c) \in \mathbb{R}^{mn\times mn}$, a fused graph $\mathbf{G}^f = (\mathbf{V}, \mathbf{E}^f, w^f) \in \mathbb{R}^{mn\times mn}$ can be defined by combining the weight functions of G^c as follows:

$$w^f(v_i, v_j) = \min(w^1(v_i, v_j), w^2(v_i, v_j), \ldots, w^{C_h}(v_i, v_j)), \tag{18}$$

Therefore, the eigenvectors ($\mathbf{u}$) and eigenvalues ($\boldsymbol{\lambda}$) are computed from $\mathbf{W}^f \in \mathbb{R}^{mn\times mn}$ by solving

the eigen problem of the normalized Laplacian, defined as:

$$\mathbf{Lu} = \boldsymbol{\lambda}\mathbf{Du}, \tag{19}$$

where the Laplacian $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}^f\mathbf{D}^{-1/2}$ and $\mathbf{D}$ is the degree matrix that is diagonal has elements $d_j = \sum_i w^f(v_i, v_j))$.

Graph Signal Processing: Smoothness prior

The problem of learning a graph with a prior of smoothness, is to learn the Laplacian matrix ($\mathbf{L}$), so that the signal variation on the resulting graph ($\mathcal{Q}(\mathbf{L})$), is small. Here a small value of $\mathcal{Q}(\mathbf{L})$, means that the signal on the graph takes similar values to its neighbors, resulting in edge disconnections from the graph [Dong et al., 2019]. The measure of smoothness of a signal x on a graph is given by:

$$\mathcal{Q}(\mathbf{L}) = \frac{1}{2}\sum_{i,j} w_{ij}(x(i) - x(j))^2,$$

where w_{ij} is the ijth entry of matrix $\mathbf{W}$. To do so, we use the recent approach in [Kalofolias and Perraudin, 2019] for large scale graphs with prior smoothness, where the authors leverage the desired graph sparsity to reduce computational cost. This model needs as inputs the number of neighbors K that are related to the edges per node, and the parameters α and β. Therefore, the minimization problem of [Kalofolias and Perraudin, 2019] can be computed as:

$$\begin{aligned} \min_{\mathbf{W}\in\mathbb{R}^{P\times P}} \; & \operatorname{tr}(\mathbf{W}^T\mathbf{Z}) - \alpha\sum_i \log(\sum_j w_{ij}) \\ & + \frac{\beta}{2}\left\|\mathbf{W}\right\|_F^2 + \frac{c}{2}\left\|\mathbf{W} - \mathbf{W}_0\right\|_F^2 \\ \text{s.t. } & w_{ij} = w_{ji} \geq 0,\ i \neq j,\ w_{ij} = 0,\ i = j \end{aligned} \tag{20}$$

where $\mathbf{Z}$ is a pairwise distances matrix, α is a log prior constant ($> \alpha \rightarrow >$ weights in $\mathbf{W}$), β is a $\|\mathbf{W}\|_F^2$ prior constant ($> \beta \rightarrow$ less sparsity in $\mathbf{W}$), and the parameter c encourages adjacency matrices close to a pre-specified adjacency matrix $\mathbf{W}_0$. To reduce the complexity of Eq. (20) the authors in [Kalofolias and Perraudin, 2019] fix the parameters $\alpha = \beta = 1$ and multiply the pairwise distances matrix ($\mathbf{Z}$) by a parameter θ to make the sparsity analysis simpler. The parameter θ is computed by each column $\mathbf{b}$ of $\mathbf{Z}$, and it has upper and lower bounds, as described by Eq. (21).

$$\begin{aligned} \theta_{lower} &= \sum_{j=1}^{n} \frac{1}{n\sqrt{Kz_{K+1,j}^2 - b_{k,j}z_{K+1,j}}} \\ \theta_{upper} &= \sum_{j=1}^{n} \frac{1}{n\sqrt{Kz_{K,j}^2 - b_{k,j}z_{k,j}}} \end{aligned} \tag{21}$$

The process is summarized in Algorithm 4.

Algorithm 4: Graph learning with prior smoothness [Kalofolias and Perraudin, 2019].

Input: Matrix of distances Z, K edges per node (sparsity level)
Output: Graph learned with prior smoothness $\mathbf{W}_s$

Step to compute θ

Compute bounds of θ with Eq. (21).

Compute θ as a geometric mean between the bounds θ_{lower} and θ_{upper}.

Step to compute adjacency matrix $\mathbf{W_s}$.

Compute $\mathbf{W}_s$ from Eq. (20) with $\mathbf{Z} = \theta\mathbf{Z}$

Blue-Noise sampling

Many methods to perform a sampling over graphs can be found in the literature [Tanaka et al., 2020] in order to find a suitable sampling set $\mathbf{S} \subset \mathbf{V} \in \mathbf{G}$. However, as explained in [Iyer et al., 2020], is desirable to get samples from different structures or objects that are present in the image, that from the graph perspective this structures/objects can be capture by sub-graphs. Therefore, we will focus on the sampling method known as *blue-noise sampling* since it focus on extracting samples related to sub-graphs. This kind of sampling has been widely used in digital half-toning [Lau and Arce, 2018], but has recently been extended to graphs [Parada-Mayorga et al., 2019].

The blue-noise sampling gives as a result a subset $\mathbf{S}$ of vertices $\mathbf{V}$ that are as far as possible from each other in terms of the geodesic distances on $\mathbf{G}$. The work in [Parada-Mayorga et al., 2019] demonstrated that the final subset given by blue-noise sampling leads to an accurate reconstruction of signals that is energy is mostly concentrated at the lowest eigenvectors of the graph Laplacian of $\mathbf{G}$.

In order to assemble all the aforementioned methods into a single approach, we need a graph that contains the most relevant samples (i.e pixels). To do this, we use a down-sampled version of the image taken from the GFKuts method [Colorado et al., 2020] which contains probabilities of the pixels that belong or do not belong to the crop (see Figure 10). The down-sampling is conducted to avoid computer memory saturation, by applying a square grid $\mathbf{M} \in \mathbb{R}^{\frac{m}{25} \times \frac{n}{25}}$. Secondly, to learn a graph by using the prior of smoothness, we use, as nodes, the vectorized version ($\mathbf{M}_i \in \mathbb{R}^{625}$) of each square in the image, computing the distances between the nodes ($\mathbf{Z} \in \mathbb{R}^{|\mathbf{V}| \times |\mathbf{V}|}$) and apply the prior of smoothness [Kalofolias and Perraudin, 2019]. Lastly, we apply the blue-noise sampling over the smoothed graph $\mathbf{W}_s$ and inject these samples to the graph based fusion algorithm proposed in [Jimenez-Sierra et al., 2020], in order to obtain the eigenvector with the highest eigenvalue of the fused-graph $\mathbf{W}^f$. Since the eigenvectors are embedded in a high-dimensional space equal to the resolution of the images (i.e 960×1280), we use the same dimensionality reduction technique

used in [Jimenez-Sierra et al., 2020], based on the t distributed stochastic neighbor embedding (t-SNE) [Van der Maaten and Hinton, 2008]. The time complexity of this approach depends on the graph learning phase and it is $\mathcal{O}(V^2K)$, where V are the nodes that are represented as patches $\mathbf{M}$ and K is the sparsity level. Figure 10 summarizes the proposed method.

4.1.3 Non-linear regression models

Support Vector Machine Regression (SVM-R)

The purpose of a support vector machine regression is to solve a linear regression problem by mapping the input data into a high dimensional space (feature space). The following model is considered:

$$f(w_e, x) = (w_e * \phi(x)) + b, \tag{22}$$

where w_e are the regression coefficients (weights), $\phi(x) : \mathbb{R}^{in} \mapsto \mathbb{R}^{out}$ with $out \gg in$, and b is the bias term. To obtain the optimal values of $\mathbf{W}_e$, a loss function can be defined such as the Laplacian, Huber's, Gaussian or ϵ-intensive. We will use the most common loss function that is the ϵ-intensive with a regularization parameter C:

$$\min_{\mathbf{W}_e,\xi,b} \frac{1}{2} \|\mathbf{W}_e\| + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{23}$$

$$\text{s.t.} \begin{cases} y_i - f(w_e, x) - b \leq \epsilon + \xi_i^* \\ f(w_e, x) + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i^*, \xi_i \geq 0 \end{cases},$$

where ξ_i^* and ξ_i are slack variables with measure deviations larger than ϵ. By transforming the optimization problem in Eq. (23) into a dual problem, the final solution is obtained by using Lagrange, the Karush–Kuhn–Tucker conditions, and the kernel trick [Laref et al., 2019], giving:

$$f(\mathbf{x}) = (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\, K(\mathbf{x}) + \mathbf{b}, \tag{24}$$

where $\boldsymbol{\alpha}, \boldsymbol{\alpha}^*$ are the Lagrange multipliers and $K(\mathbf{x})$ is a Kernel function. In this case we use the Radial Basis Function (RBF) Kernel (similar to Eq. (17)).

A Nonlinear Autoregressive Exogenous (NARX)

An exogenous autoregressive model (NARX) aimed at the identification of non-linear systems involves both current and past values of the impulsive series that models the dynamics of the system. NARMAX (Nonlinear Autoregressive Moving Average with Exogenous Inputs) methods provide models that are transparent, and easily solve many problems such as contour errors in

CNC machines, forecasting, network traffic, and prediction of daily solar radiation [Huo and Poo, 2013, Men et al., 2014, Alfred et al., 2015, Boussaada et al., 2018]. The NARMAX model is defined in Eq. (25). Where y(k) is the output system, u(k) is the input system and e(k) is the noise sequence. $\mathbf{n}_y$, $\mathbf{n}_u$ y $\mathbf{n}_e$ are the maximum lags for the system output, input and noise respectively; $\mathbf{F}[.]$ refers to the function and d is a time delay usually set to $\mathbf{d} = \mathbf{1}$.

$$Y(\mathbf{k}) = F[y(\mathbf{k}-\mathbf{1}), .., y(\mathbf{k}-\mathbf{n_y}), u(\mathbf{k}-\mathbf{d}), .., u(\mathbf{k}-\mathbf{d}-\mathbf{n_u}), e(\mathbf{k}-\mathbf{1}), .., e(\mathbf{k}-\mathbf{n_e})] + e(\mathbf{k})] \quad (25)$$

In this work, the function $\mathbf{F}[.]$ has been implemented using artificial neural networks (ANN) with two hidden layers and one output layer.

4.1.4 Experimental Setup

In order to compare the effectiveness of the proposed feature extraction based on graphs, the results were compared with two approaches: (i) the former graph method introduced in [Jimenez-Sierra et al., 2020], namely GBF, and (ii) the GFKuts approach [Colorado et al., 2020]. We used the datasets and Ground-Truth reported in [Jimenez-Sierra et al., 2020, Colorado et al., 2020], containing 314 images for the vegetation stage, 82 for the reproductive, and 71 for the ripening. The captured images have a resolution of 960×1280 pixels, geo-referenced with the corresponding biomass measurements in grams (g) from the Ground-Truth[8], which was assembled as follows: 1 linear meter of the plants were cut from each plot of the crop and weighted to obtain the fresh biomass. Subsequently, the samples are placed inside an oven at 65 degrees Celsius for 4 days or until a constant weight is reached. This is known as the dry biomass.

Having the imagery dataset and the Ground-Truth, we trained two estimation models: classical SVM-R regression and a robust nonlinear autoregressive network with exogenous inputs (NARX), both accounting for 70% of the dataset for the training, and the remaining 30% for testing and validation. Lastly, the performance of the models was measured in terms of the root mean squared error ($RMSE$), the linear correlation (r), and the coefficient of determination (R^2). Additionally, we conducted several experiments with the SVM regressor considering different kernel functions (i.e. linear, polynomial, and RBF), while the best performance was achieved by the RBF Kernel (similar to the work in [Jimenez-Sierra et al., 2020]).

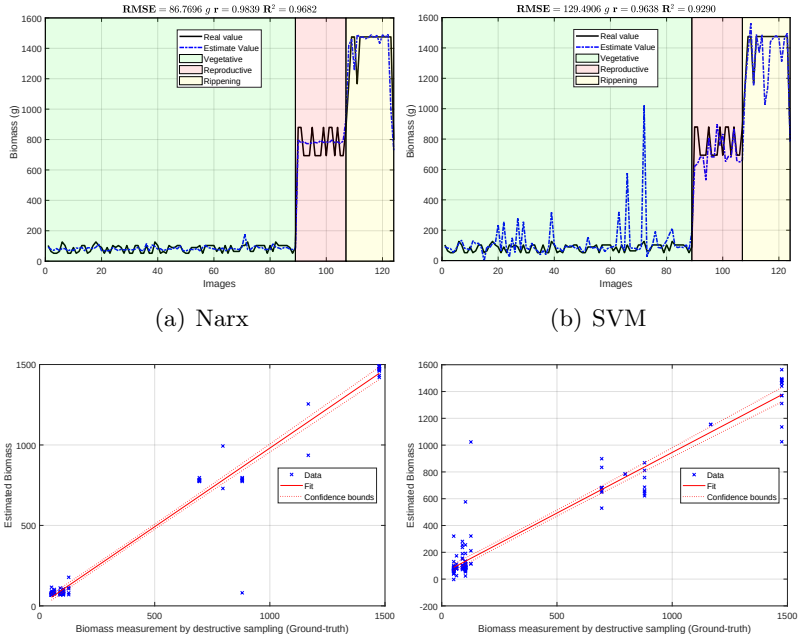
The algorithms were tested in a server with two Intel(R) Xeon(R) CPUs E5-2650 v4 @ 2.20GHz, with 24 physical cores, 48 threads of processes, and 252GB of RAM. Parameters are detailed in Table 20.

[8]Further information of the experimental protocol can be found at https://www.protocols.io/view/protocol-bjxskpne

Table 20: Parameters used for the models GFKuts [Colorado et al., 2020], GBF [Jimenez-Sierra et al., 2020], and our proposal GBF-Sm-Bs model.

Model	**Parameters**			
GFKuts [Colorado et al., 2020]	k-neighbors = 2	window radius = 60	$\epsilon = \mathbf{10^{-6}}$	-
GBF [Jimenez-Sierra et al., 2020]	-	-	t-SNE$_{dim} = 16$	$n_s = 100$
GBF-Sm-Bs	edges/node = $\frac{\lvert\mathbf{V}\rvert}{2}$	window size = 25	t-SNE$_{dim} = 89$	$n_s = 100$

4.1.5 Results and Discussion



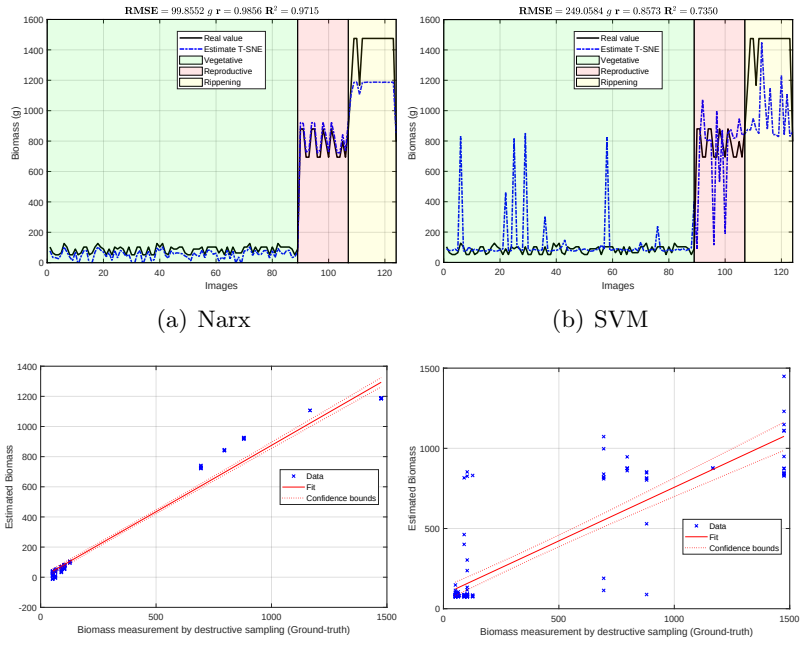
(a) Narx (b) SVM

(c) Lineal fit of Narx (d) Lineal fit of SVM

Figure 11: Biomass estimation using the GFKuts approach [Colorado et al., 2020].

Figures 11 to 13 show the experimental results obtained from the 3 methods in the following order: (i) GFKuts, (ii) former graph method (GBF), and (iii) proposed graph method with blue-noise and prior smoothness (GBF-Sm-Bs). Numerical data containing the above-ground biomass estimation results for each model are presented in Table 21.

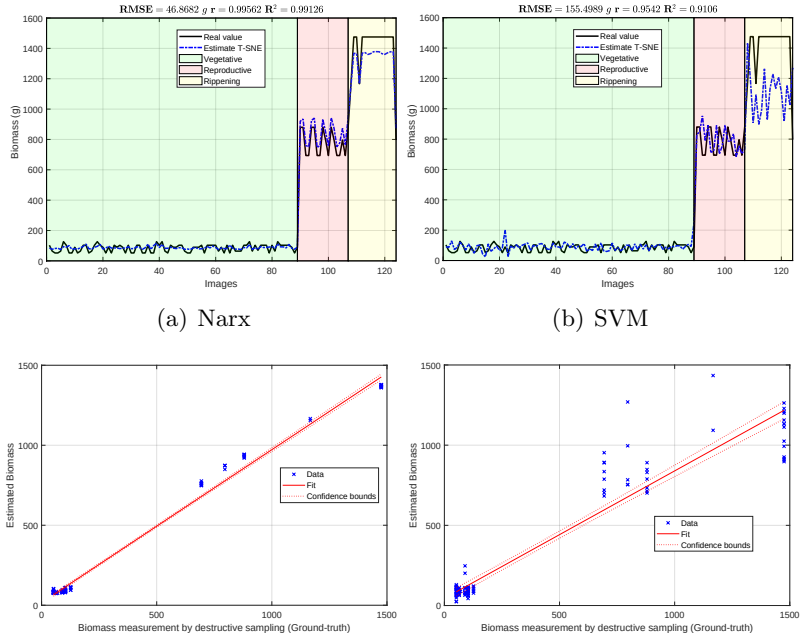
According to Table 21, the GKFkuts approach achieved the lowest $RMSE = 129.490$ g , followed by the proposed graph-based model with $RMSE = 155.498$ g. However, by applying the robust Narx regressor, the best performance was achieved by our proposal, with a $RMSE = 45.358$ g. Results were also correlated to the linear fit, presented from figures 11 to 13. There we can observe



(a) Narx (b) SVM

(c) Lineal fit of Narx (d) Lineal fit of SVM

Figure 12: Biomass estimations using the graph-based approach with uniform sampling features (GBF). [Jimenez-Sierra et al., 2020].



(a) Narx (b) SVM

(c) Lineal fit of Narx (d) Lineal fit of SVM

Figure 13: Biomass estimations using the proposed graph-based approach with blue-noise and prior smoothness features (GBF-Sm-Bs).

that the Narx regressor improves the fitting to a linear model, in which the proposed method achieved a linear correlation value of $r = 0.995$.

Table 21: Performance in terms of $RMSE$, r and R^2 for the models: GFKuts [Colorado et al., 2020], GBF [Jimenez-Sierra et al., 2020], and our proposed GBF-Sm-Bs model with both Narx and SVM regressors.

Model/Regressor	**RMSE** (in grams [g])	**r**	$\mathbf{R^2}$
GFKuts [Colorado et al., 2020]/SVM	**129.490**	**0.963**	**0.929**
GBF [Jimenez-Sierra et al., 2020]/SVM	249.058	0.857	0.735
GBF-Sm-Bs/SVM	155.498	0.954	0.910
GFKuts [Colorado et al., 2020]/Narx	86.769	0.983	0.968
GBF [Jimenez-Sierra et al., 2020]/Narx	99.855	0.9856	0.971
GBF-Sm-Bs/Narx	**45.358**	**0.995**	**0.991**

Contrasting with the previous work from [Colorado et al., 2020], the Narx model enabled a slight improvement of about 3% over the former GFKuts method, due mainly to the configuration of the auto regressive inputs, and the use of vegetative indices with the soft mask generated by the GFKuts approach. Regarding the former graph-based method from [Jimenez-Sierra et al., 2020] (GBF), results were improved by around 62.43% (155.498/249.058) and 45.42% (45.358/99.855) for the SVM and Narx regressors in terms of the $RMSE$ metric, respectively. For the SVM regressor, the GBF and the GBF-Sm-Bs methods achieved a $RMSE = 259.058$ g and $RMSE = 155.498$ g respectively, as observed in Table 21, while for the Narx regressor, the GBF and the GBF-Sm-Bs methods achieved a $RMSE = 99.855$ g and $RMSE = 45.358$ g respectively, as presented in Table 21. This demonstrates the effectiveness of using a structured sampling method combined with graph learning, based on a prior of smoothness, for the extraction of relevant samples to be apply into Nyström extension.

Overall, the proposed GBF-Sm-Bs approach using Narx achieved an improvement in the $RMSE$ metric of about 50%, while also improving the biomass correlations reported from previous works. Part of the improvement achieved with the GBF-Sm-Bs method comes from extracting more relevant features than GFKuts, precisely 89 features. While GFKuts works with only 7 VI-features (per image) highly sensitive to biomass variations, the graph-based method is characterized by the eigenvectors with the highest eigenvalue associated with the fused-graph $\mathbf{W}^f$, yielding 89 features per image.

4.2 Graph-based data fusion with GSP learning from superpixels for CD

In this section we propose a graph-based framework for CD. Where, the graph learning is driven by a GSP approach based on prior signal smoothness and it is used directly for the CD rather than using it to extract samples. In addition, we proposed a false RGB image to generate a fixed

number of super-pixels (i.e. the graph nodes), and a new optimization problem to find the change map that uses all the spectral basis of the fused graph.

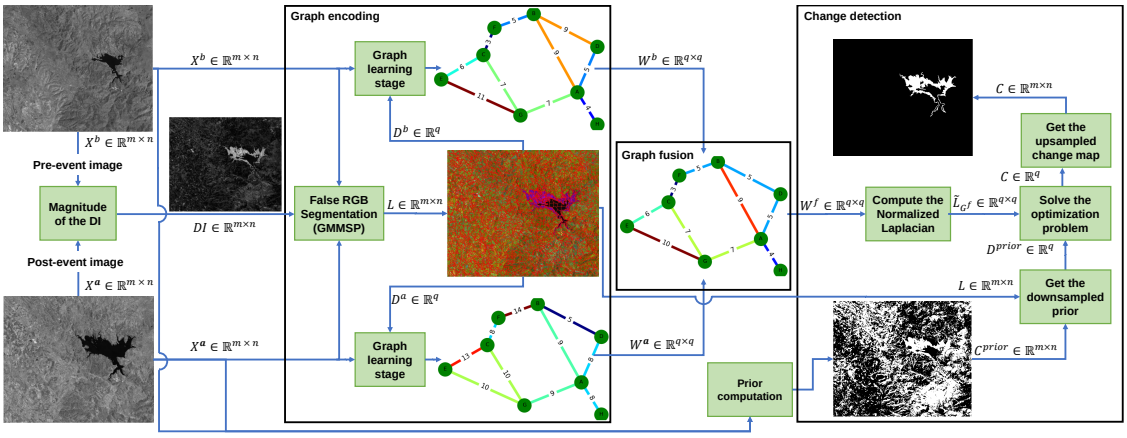


Figure 14: Framework of the proposed method for homogeneous and heterogeneous change detection.

4.2.1 The Proposed Method

This section explains the theory behind the proposed approach. First, in Section V-A, we explain the GMMSP framework to downsample the images into regions, including the representation that we use for the downsampled (D) and upsampled images (C). Finally, in section V-B we explain the steps involved in the algorithm 5.

4.2.1.1 Gaussian mixture models for superpixel segmentation

In order to select a proper superpixel segmentation algorithm, we conducted two experiments to compare the elapsed time and performance with respect to the state-of-the art approaches SLIC [Achanta et al., 2012] and AMR [Lei et al., 2019a]. In the first experiment, we measure the time elapsed for each algorithm to generate the superpixels (see Table 34 in the supplementary material). The fastest approach is GMMSP [Ban et al., 2018]. In the second experiment, we measure the performance of each superpixel method in the proposed framework in Figure 14, the results demonstrate that the best performance in terms of κ was achieved by the GMMSP (see Table 35 in the supplementary material). Therefore, we select GMMSP [Ban et al., 2018] as the superpixel segmentation algorithm.

The GMMSP [Ban et al., 2018] approach associates each superpixel with width v_x and height v_y to a Gaussian distribution. Then each pixel is considered as a random variable described by a weighted sum of several Gaussian functions (Gaussian Mixture Model (GMM)). To reduce the computational complexity of knowing all the parameters of the Gaussian models, Ban *et al.* model

each pixel as a pixel-related GMM. Therefore, a subset of Gaussian functions is built related to the spatial position of that pixel. GMMSP solves the Expectation-Maximization (EM) problem to tune the parameters of the Gaussian functions.

Since we need to create a graph for each image to later perform the fusion step, GMMSP algorithm should have the same regions when applied to the pre- and post-event images. To achieve this, we generated what we call a false RGB image, where the channels are the pre-event image, post-event image, and DI magnitude. Next, GMMSP segments the false RGB image into q regions, and outputs the label matrix L of size $m \times n$ (same size as the original images). L represents the label for each pixel related to its corresponding region. Figure 15 shows the results of applying GMMSP on the Sardania dataset.

Using the label matrix L that's entries take values in $\{1, \ldots, q\}$, we represent an image X of size $m \times n$ with a vector $D \in \mathbb{R}^q$ with entries D_i given by:

$$D_i = \frac{\sum_{r=1}^{m}\sum_{s=1}^{n} \mathbf{1}_{B_i}(r,s)X(r,s)}{\sum_{r=1}^{m}\sum_{s=1}^{n} \mathbf{1}_{B_i}(r,s)}, \ i = 1, \ldots, q, \tag{26}$$

where $1_{B_i}(,)$ denotes the indicator function of the set B_i given by $\{(r,s) : L(r,s) = i\}$. That is, if $(r,s) \in B_i$, $1_{B_i}(r,s) = 1$, but if not, $1_{B_i}(r,s) = 0$. The vector D can be considered a downsampled version of X, but due to the irregular nature of the downsampling regions, D does not necessarily form an image.

We also define vector representations $D^b, D^a, D^{\text{prior}}$ of the pre-event, post-event, and prior images $X^{b,p}$, $X^{a,p}$, C^{prior} as presented in (26).

4.2.1.2 Change Map Estimation

Provided $D^b, D^a,$ and D^{prior}, we now describe the procedure to estimate the change map. We first find the graph representations G^b, and G^a from D^b, D^a by using the gsp method of prior smoothness (see Section 4.1.2), respectively, and obtain the fused graph G^f as:

$$w^f(v_i, v_j) = \min(w^b(v_i, v_j), w^a(v_i, v_j)), \tag{27}$$

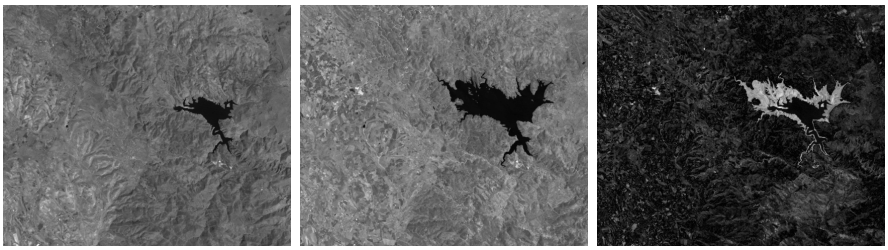
for $(v_i, v_j) \in E^b \cap E^a$. To illustrate the above, consider the following adjacency matrices W^b and W^a and their fused matrix W^f obtained using (2). Where, $W^b(v_1, v_2) = 0.68$, and $W^a(v_1, v_2) = 0.2$. Hence, $W^f(v_1, v_2) = 0.2$. Then compute the normalized graph Laplacian $\tilde{L}_{G^f}$ from G^f as follows:

$$\tilde{L}_{G^f} = D_{G^f}^{-1/2} L_{G^f} D_{G^f}^{-1/2}, \tag{28}$$

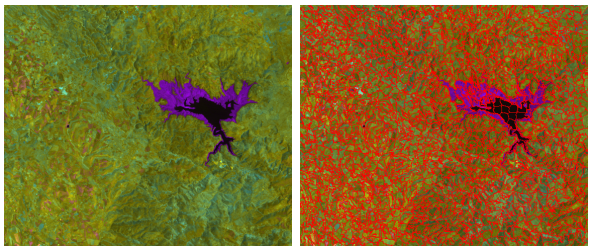
where D_{G^f} and L_{G^f} denote the degree and the graph Laplacian matrices of G^f.

The change map $c \in \mathbb{R}^q$ is obtained by denoising the prior change map with respect to $\tilde{L}_{G^f}$ as follows:

$$c = \arg\min_{x \in \mathbb{R}^q} \ x^\top \tilde{L}_{G^f} x + \alpha \left\| x - D^{\text{prior}} \right\|_2^2, \tag{29}$$



(a) Pre-event (b) Post-event (c) DI Magnitude



(d) False RGB (e) Downsampled image

Figure 15: GMMSP applied to Sardania dataset transform into the false RGB image (d). The channels are: pre-event image (a), post-event image (b), and the magnitude of the DI (c). The downsampled output image (e) contains around 2000 regions.

where $\alpha > 0$ is a regularization parameter, controlling the degree to which the solution of (29) is close to the prior D^{prior} and simultaneously smooth on the fused graph G^f. Here, the smaller α is set, the smoother the solution c on G^f will be. We note that unlike the previous approach, in which the change map corresponded to an eigenvector of $\mathbf{W}^f$, the new approach depends on the whole spectral basis of $\tilde{L}_{G^f}$. To clarify, note that the solution to problem (29) satisfies:

$$c = (\tilde{L}_{G^f} + \alpha I)^{-1} \alpha D^{\text{prior}}. \tag{30}$$

As a consequence, c is not exactly formed by an eigenvector of $\tilde{L}_{G^f}$. As will be explained shortly, this leads to more robust and stable change detection. Since $c = (c_1, \ldots, c_q)^T$ is a q dimensional vector, we transform it into a change detection image C of size $n \times m$ with entries $C(r, s)$ given by

$$C(r, s) = \sum_{i=1}^{q} c_i \mathbf{1}_{B_i}(r, s), \tag{31}$$

for $(r, s) \in \{1, \ldots, m\} \times \{1, \ldots, n\}$. Finally, we apply the Otsu's threhold to binarize the change map C as follows:

$$C(r, s) = \begin{cases} 0 & C(r, s) < T \\ 1 & C(r, s) \geq T \end{cases}, \tag{32}$$

where T is the global threshold that minimize the intra-class variance:

$$\sigma_b^2 = \omega_0(T)\omega_1(T) \left[\mu_0(T) - \mu_1(T)\right], \tag{33}$$

where ω_0 and ω_1 are the probabilities of the two classes separated by a threshold T, and μ_0 and μ_1 are the class mean (for further details see [Otsu, 1979]).

Algorithm 5: Graph based fusion for change detection using the signal smoothness representation approach combined with GMMSP.

Input: Images pre- and post-event $X^b, X^a \in \mathbb{R}^{m\times n}$, v_x.
GL-SMO: number of edges per node on average $K > 0$.
Output: Change Map $C \in \mathbb{R}^{m\times n}$

Find the Change Map prior $\Rightarrow C^{\text{prior}}$ by using $C^{\text{prior}} = \mathcal{T}\left(\frac{X^{b,p}-X^{a,p}}{X^{b,p}+X^{a,p}}\right) + \mathcal{T}\left(\frac{X^{a,p}-X^{b,p}}{X^{b,p}+X^{a,p}}\right)$, where $\mathcal{T}$ is a binarization process, implemented with Otsu's thresholding method [Otsu, 1979].

GMMSP [Ban et al., 2018] applied to the false RGB image $\Rightarrow D^{\text{prior}}, D^b, D^a$

Fixing $\boldsymbol{v_x}$ to group C^{prior} in $\boldsymbol{q}$ regions (each pixel value now has a region number) $\Rightarrow L$
Using L, compute $D^b, D^a, D^{\text{prior}}$ as in (26), explained in Sec. 4.2.1.1.

Graph Learning (GL):
1. Learn graph representations W^b, W^a from the pre- and post-event vector representations D^a and D^b by using **GL-SMO**.
2. Construct a fused adjacency matrix W^f of size $q \times q$ such that $W^f(i,j) = \min(W^b(i,j), W^a(i,j))$, with $(i,j) \in \{1,..,q\} \times \{1,..,q\}$.

Change Detection:
$\tilde{L}_{G^f}$ by using (28).
Find the downsampled CD vector C by using (30).

Image representation for change detection vector:
By using the indicator function $1_{B_i}(,)$, obtain the upsampled change detection map C with (31).
Apply the Otsu's threhold for image binarization as show in (32).

The time complexity for Algorithm 5 taking into account the upper-bound is given by the graph learning and it is $\mathcal{O}(q^2K)$, where q are the number of regions/super-pixels generated and K is the sparsity level.

4.2.1.3 Description

Here, we present our proposal for integrating graph learning based on the signal smoothness representation approach, the GMMSP as a downsampling module into the graph based fusion change detection (**GBF-CD**) model [Jimenez-Sierra et al., 2020], and the estimation of the change map by using Equation (30). The algorithm 5 and Figure 14 show the proposed approach. First, we apply the GMMSP method [Ban et al., 2018], which segments a false RGB image as described in subsection 4.2.1.1. Then, we use the given local homogeneous regions from the downsampling process for the pre- and post-event images, as well as the prior C^{prior} to compute the vectorized version of the data by using equation (26). By doing this, we obtain two vectors that represent

the pre- and post-event images (D^b and D^a) and another that represents the prior (D^{prior}). These vectors have q elements, where q is the number of regions. It is important to highlight that we also applied the GMMSP to the prior C^{prior} because we need this region-based version to solve the proposed optimization problem in (29).

After downsampling the images (i.e. by using the GMMSP), we continue with the GL-stage, using the smoothness approach, which is the **GL-SMO** [Kalofolias and Perraudin, 2019]. The integration of the downsampling module, the GL stage, and the change map estimation results in the following approach, referred to as **G-SMO-CD**. As a result, we obtain the two graphs that represent the pre- and post-event images as adjacency matrices with sizes $q \times q$. Next, we apply the fusion stage, by using the minimization of the similarity as in the **GBF-CD** method [Jimenez-Sierra et al., 2020] (see Equation (27)). Then, from the fused graph W^f we computed the normalized Laplacian matrix by using equation (28) and obtain the change map by solving the optimization problem as shown in Equation (30). Afterwards, we perform the Otsu's thresholding technique [Otsu, 1979] ($\mathcal{T}(C) \in \mathbb{R}^q$) to obtain the binary change map. Lastly, to obtain the original size of the change map, we make use of the label matrix L and the indicator function $1_{B_i}(,)$ in equation (31) as explained in 4.2.1.1.

4.2.2 Results and Discussion

Databases: We tested our approach on 14 real change detection scenarios captured by MS and SAR sensors presented in [Jimenez-Sierra et al., 2020]. In particular, the datasets were acquired by different satellites such as Landsat, ALOS/PALSAR, ERS-2 SAR, Sentinel 1A, and ESA/ASAR (see Table 2 in Section 3). The image resolutions range from 275×400 to 4220×2320. These datasets include events such as earthquakes, floods, wildfires, melted ice, farming, and building of a construction site. These datasets incorporate 6 MS homogeneous cases, 4 SAR homogeneous cases, and 4 heterogeneous cases (Toulouse, California, Bastrop, Gloucester-2) which combine SAR/MS images (i.e. respectively positioned in Table 2). The metadata for the datasets is detailed in Table 2 and the images are shown in Figure 4 (see Section 3). Furthermore, to the best of our knowledge, we consider that our datasets are large and diverse for CD benchmark, compared to recent CD datesets. For instance, previous works [Niu et al., 2018, Zanetti and Bruzzone, 2018, Luppino et al., 2019, Luppino et al., 2021, Sun et al., 2021c, Sun et al., 2021a] have shown the application of their approaches in experimental setups that contain between two and six datasets, and the approach in [Sun et al., 2021d] uses twelve datasets. By contrast, our approaches are tested on 14 datasets. Recently, in [Chen and Shi, 2020] the authors proposed a CD dataset that contains 637 images. However, these images are RGB (not MS), with a resolution of 1024×1024, that mainly contain

building CD scenes.

Experimental Set-Up: We ran all the codes in MATLAB®2018a using a server with two processors, Intel(R) Xeon(R) CPU E5-2650 v4 @2.20GHz, with a total of 24 physical cores, 48 threads, and 252 GB of memory RAM @2400MHz.

Parameter settings: Firstly, MS images were normalized according to their maximum value and SAR images were normalized in logarithmic scale since their scale is defined in dB. Secondly, for the GMMSP[9] based downsampling process, we set $v_x = v_y$ for the superpixels and then we selected the values of v_x to obtain the desired number of regions (i.e. from 500 to 2500) over the images. Afterwards, we carried out a grid-search for the parameter K, for the GL based on smoothness, with a fixed value of $\alpha = 0.5$. We performed a grid search over K from $K = 2$ to the maximum number of regions in steps of one. Lastly, once we find the optimal parameter K, we fine-tune α with a grid-search from 0.001 to 1 with a step size of 0.001.

To summarize, our experiments consist of: i) Normalization of the data, ii) Downsampling of the data, iii) Tuning of the GL parameter, iv) Fine tuning of the penalization parameter, and v) Change map estimation. In order to evaluate the influence of the number of regions in the performance of the proposed framework, we ran 5 experiments from 500 to 2500 regions with a step of 500. With respect to the competing methods used for comparison (the **rrR-EM** [10] [Zanetti and Bruzzone, 2018], the **U-HPT**[11] [Luppino et al., 2019], the **CAN**[12] [Niu et al., 2018], the **X-Net**[13] [Luppino et al., 2021], the **GBF-CD**[14] [Jimenez-Sierra et al., 2020], the **INLPG**[15] [Sun et al., 2021d], the **NPSG**[16] [Sun et al., 2021c], and the **IRGM**[17] [Sun et al., 2021a]), we used the parameters recommended by these authors.

In the tables 22 to 26 we refer to our method as **G-SMO-CD**. Additionally, for **CAN** and **X-Net** since their training varies in each run (i.e. randomly), we performed 20 runs and reported the mean value and the standard deviation for the κ metric.

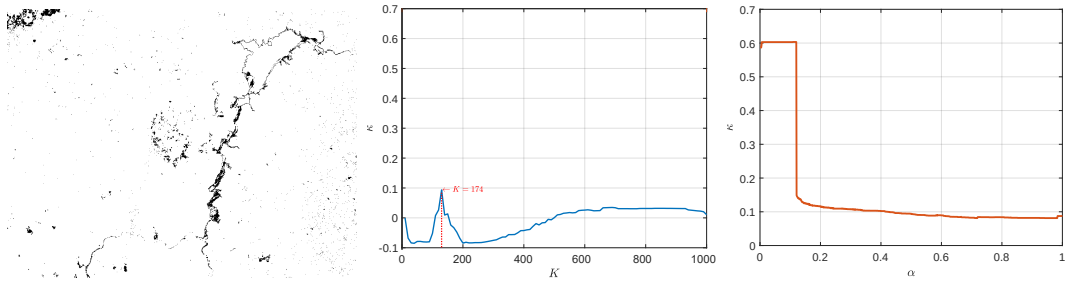
Results: First of all, we show the influence of the regularization parameter α in (29), when we have a noisy prior. Secondly, we summarized the sensitivity of κ with respect to parameter K, for the graph learning based on smoothness. Then, we show the impact of the number of regions in the downsampling process with respect to the performance of the change map detection. Afterwards,

[9] Available at github.com/ahban/GMMSP-superpixel
[10] Reproduced by authors in [Jimenez-Sierra et al., 2020] at: github.com/DavidJimenezS/rrR_EM
[11] Available at: github.com/llu025/Image_Regression
[12] Reproduce by authors in [Luppino et al., 2021] at: github.com/llu025/CAN
[13] Available at: github.com/llu025/Deep_Image_Translation
[14] Available at: github.com/DavidJimenezS/GBF-CD
[15] Available at: github.com/yulisun/INLPG
[16] Available at: github.com/yulisun/NPSG
[17] Available at: github.com/yulisun/IRG-McS

Table 22: Parameters used for evaluation of the datasets.

DATASETS	$\mathbf{v_x} = \mathbf{v_y}$	REGIONS	K	$\boldsymbol{\alpha}$
Mulargia	11	2658	1464	0.1
Omodeo	20	1928	174	0.109
Alaska	11	2613	1479	0.013
Madeirinha	10	1953	410	0.2
Prince George	44	2526	2400	0.1
Gloucester-1	144	516	171	0.1
Katios	23	2087	102	0.738
Atlantico	17	2581	2029	0.103
San Francisco	16	510	260	0.215
Wenchuan	8	2309	1740	0.5
Toulouse	86	2017	192	0.002
California	73	1493	291	0.1
Bastrop	23	2435	393	0.1
Gloucester-2	150	518	444	0.042

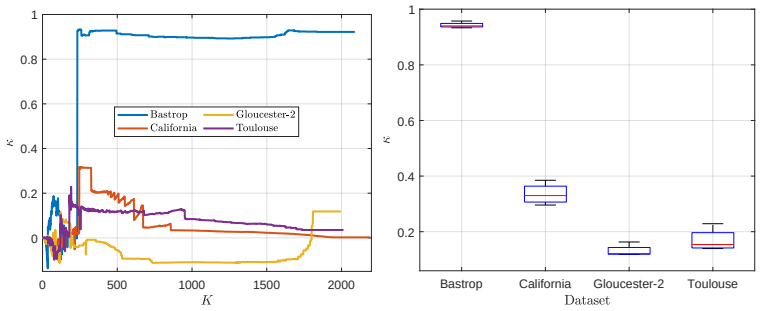
we analyzed three scenarios: MS homogeneous, SAR homogeneous and MS/SAR heterogeneous. We measure performance based on the metrics: false negative (FN), false positive (FP), precision (P), recall (R), Cohen's kappa coefficient ($\boldsymbol{\kappa}$), overall error (OE), and execution time (ET), in Tables 23 to 25 (see Tables 36 to 39 in the supplementary material for FN, FP, R, and P metrics). Additionally, to provide a visual comparison between the methods, we present the change maps for both homogeneous (Figure 39 to Figure 48) and heterogeneous (Figure 49 to Figure 52) cases in the supplementary material section. The maps show false negatives (FN), false positives (FP) and correct pixels (C). Finally, we show and discuss the ablation results. In the ablation process, we remove the learning stage, the optimization problem, and estimate the change map, as authors in [Jimenez-Sierra et al., 2020] (explained in Section 3).



(a) Omodeo C^{prior} (b) GL tuning with $\alpha = 0.5$ (c) α tuning with $K = 174$

Figure 16: Parameter α penalization over a noisy prior map (C^{prior}).

Discussion: On the one hand, Figure 16 (b) in blue shows the tuning of GL parameter K (i.e. the number of neighbors or edges per node) by using grid search. The parameter K ranges from 2



(a) Sensitivity of the GL stage (b) Sensitivity of the number of regions

Figure 17: Sensitivity test in the proposed framework for the GL stage and the downsampling module. For the Heterogeneous cases.

Table 23: Models performance in terms of $\boldsymbol{\kappa}$.

Dataset	rrR-EM	U-HPT	CAN	X-Net	GBF-CD	INLPG	NPSG	IRGM	G-SMO-CD
Mulargia	0.7929	0.6984	0.1564 ± 0.0124	0.7395 ± 0.0161	0.9043	0.8906	0.6431	0.8431	**0.9165**
Omodeo	**0.8112**	0.6414	0.7388 ± 0.0029	0.7219 ± 0.0068	0.2873	0.1058	0.1386	0.1661	0.6032
Alaska	0.7368	0.8565	0.8815 ± 0.0037	0.8707 ± 0.0036	0.8917	0.6963	0.5320	0.8388	**0.9106**
Madeirinha	0.6727	0.5048	0.1122 ± 0.0287	0.3900 ± 0.0218	0.8046	0.4138	0.1440	0.4446	**0.8303**
Prince George	0.5322	0.4742	0.2434 ± 0.0116	0.3020 ± 0.0171	0.5442	0.2906	0.2921	0.2018	**0.8319**
Gloucester-1	0.1724	0.5794	0.1534 ± 0.0043	0.5102 ± 0.0123	0.2286	0.3717	0.1283	0.5259	**0.7033**
Katios	-0.1790	0.2882	0.1503 ± 0.0973	0.1045 ± 0.0001	0.3196	0	0.0794	0.4208	**0.7103**
Atlantico	-0.0040	0.0970	0.2315 ± 0.0142	0.1861 ± 0.0202	**0.4726**	0.2759	0.4474	0.4496	0.2901
San Franciso	0.1311	0.3143	0.4775 ± 0.0089	0.4107 ± 0.0142	0.4285	0.6767	-0.0246	-0.0235	**0.6947**
Wenchuan	0.2380	-0.2730	-0.0790 ± 0.0028	-0.0693 ± 0.0014	0.3239	-0.2134	-0.1789	-0.0463	**0.7991**
Toulouse	0.1329	0.1200	-0.0266 ± 0.0151	0.2085 ± 0.0205	0.1702	**0.3118**	0.0029	0.2688	0.2287
California	-0.1430	0.3845	0.3646 ± 0.0282	0.3764 ± 0.0046	0.3507	-0.0617	0.1015	0.3288	**0.3850**
Bastrop	0.000	0.8884	0.1392 ± 0.0115	0.8541 ± 0.0127	0.8875	0.7618	0.2727	0.0791	**0.9566**
Gloucester-2	0.1693	0.0080	0.0041 ± 0.0058	0.1972 ± 0.0390	0.1562	**0.3436**	-0.0176	0.1629	0.1685

to the maximum number of regions. For this GL tuning we fixed the regularization parameter α to 0.5. On the other hand, Figure 16 (c) shows the fine-tuning of the α parameter that ranges from 0.001 to 1 with a step size of 0.001. For the tuning of α, the learning parameter for the graph is set as the one that achieves the best κ during the learning stage (i.e. in this case $K = 174$ shown in red in 16 (b)). Then, if the C^{prior} is noisy (see Figure 16 (a)) then a higher penalization of the C^{prior} (i.e. a low value of α) avoids a misleading result given by such a noisy prior (C^{prior}) in the final change map. In consequence, the proposed optimization approach presented in (29) proves to be robust against noisy priors when the penalization parameter α has low values ($\alpha \leq 0.15$).

We carried out a sensitivity analysis of the GL process by analyzing the impact on the change map estimation caused by the tuning of the parameter K. For instance, Figure 17(a) shows the

Table 24: Models performance in terms of OE (units in %).

Dataset	rrR-EM	U-HPT	CAN	X-Net	GBF-CD	INLPG	NPSG	IRGM	G-SMO-CD
Mulargia	1.33	2.20	5.05	1.78	0.53	0.61	2.98	0.90	**0.49**
Omodeo	**2.04**	2.68	2.32	2.33	4.83	4.47	4.53	3.72	2.83
Alaska	5.64	2.49	2.17	2.30	2.02	6.90	11.36	2.95	**1.77**
Madeirinha	8.81	11.81	21.72	14.76	5.60	16.90	16.32	12.95	**5.25**
Prince George	15.79	17.12	26.22	23.20	15.27	24.82	24.40	24.30	**7.13**
Gloucester-1	40.47	6.05	13.96	7.87	26.33	16.53	22.89	6.95	**4.97**
Katios	11.69	10.90	15.77	43.17	15.00	89.44	11.90	11.93	**4.48**
Atlantico	15.49	15.53	15.89	15.97	26.17	15.63	31.97	**14.47**	23.11
San Franciso	53.73	10.92	9.68	10.84	12.87	**5.94**	23.91	9.57	7.87
Wenchuan	51.83	18.88	22.74	22.58	24.81	48.73	37.55	21.60	**5.89**
Toulouse	24.47	8.72	12.84	14.11	20.27	9.51	90.33	9.92	**7.67**
California	94.36	5.21	7.90	11.85	11.80	73.71	36.63	8.51	**4.35**
Bastrop	10.63	2.00	12.37	2.92	1.99	4.14	11.48	10.62	**0.83**
Gloucester-2	21.29	**7.78**	15.38	12.64	27.82	13.16	77.3	9.15	31.61

Table 25: Models performance in terms of ET (units in s).

Dataset	rrR-EM	U-HPT	CAN	X-Net	GBF-CD	INLPG	NPSG	IRGM	G-SMO-CD
Mulargia	**13.90**	107.98	941.70	611.71	19.21	65.45	112.92	45.89	127.54
Omodeo	**9.93**	294.32	2330.25	1623.01	91.62	412.88	394.22	50.30	95.03
Alaska	8.32	123.21	911.47	602.99	**3.62**	67.89	117.14	45.44	133.57
Madeirinha	16.32	77.37	584.55	474.47	**4.10**	15.57	42.02	46.32	61.07
Prince George	723.78	2045.13	3861.93	3619.16	240.74	1831.71	2454.10	**61.54**	133.91
Gloucester-1	**24.54**	3808.46	8666.61	14105.61	94.46	8448.30	5234.0	71.63	31.69
Katios	**4.08**	427.53	3871.94	3620.19	34.48	365.19	304.18	48.0	131.36
Atlantico	**5.49**	333.74	2472.46	1926.65	103.91	38.70	77.08	46.34	136.82
San Franciso	3.81	64.90	322.49	399.45	**3.21**	949.19	500.89	51.49	**3.21**
Wenchuan	**3.68**	65.03	441.33	419.70	6.24	23.86	53.0	53.94	81.35
Toulouse	**3.68**	4449.60	4434.27	15250.43	839.94	11958.0	5940.10	74.05	100.26
California	**25.52**	2955.94	5865.05	7378.58	921.62	3344.30	3713.20	64.90	59.49
Bastrop	**3.68**	365.30	4111.59	4088.67	109.35	1074.30	561.69	51.93	117.40
Gloucester-2	**3.68**	3767.05	9172.24	13079.53	543.65	8328.90	5084.30	72.94	27.59

sensitivity for the GL process for the heterogeneous datasets (please refer to the supplementary material to see the Figures 53 to 55 related to the sensitivity in the homogeneous cases).

Figure 17(a) shows that for low values of K, the model presents a moderate variability in the $\boldsymbol{\kappa}$ metric. This is because, the lower the value of K, the denser, or less sparse the graph is. Hence, the disconnections of the regions related to the real change are more likely to be affected by regions that are not related to the change. Moreover, as shown in Figure 17(a) with a value of $K > 10\%$ of the total regions (around of 2000 regions in these cases), the performance of the change map detected is less sensitive, or shows low variability to the changes presented in the parameter K.

In general, the proposed framework in the learning stage presents low variability with respect to changes in the parameters. Nevertheless, tuning is needed in the learning stage, as shown in Figure 17(a), to achieve adequate performance in the change map detection.

Figure 17(b) shows the box plots of the region sensitivity analysis for the heterogeneous cases (please refer to the supplementary material to see the Figures 53 to 55 related to the homogeneous cases). The regions range from 500 to 2500 with a step of 500 for a total of 5 scenarios with a different number of regions. In Figure 17(b), we can see that the proposed G-SMO-CD method presents the lowest deviation of 0.025 with a mean value of 0.93 for the Bastrop dataset, and the highest deviation of 0.05 with a mean value of 0.18. These results show that the G-SMO-CD is robust to the number of regions. In addition, the learning stage and the optimization process are the greatest influence in the change map detection.

With regard to MS homogeneous cases given in Figures 39 to 44 (see the supplementary material), our proposed G-SMO-CD method surpasses the 8 comparison methods in 5 out of the 6 datasets (see Table 23). G-SMO-CD achieves the highest performance in the Mulargia dataset with a $\kappa = 0.9165$, and the lowest in the Omodeo dataset with a $\kappa = 0.6032$. In the Omodeo dataset, the rrR-EM method achieved the best performance with $\kappa = 0.8112$. We believe, that the performance of the Omodeo dataset could improve with a better prior (see Figure 16 (a)). Additionally, the G-SMO-CD demonstrates a balance between the FN and FP (see Tables 36 and 37 in the supplementary material respectively). The G-SMO-CD has the lowest FN and highest P in Alaska, Madeirinha, and Prince George datasets (see Table 36 and 38 in the supplementary material). G-SMO-CD presents the lowest OE in 5 out of the 6 datasets (see Table 24).

For the second scenario of SAR homogeneous cases presented in Figures 45 to 48 (see supplementary material). The G-SMO-CD method achieved the best overall performance in three out of the four datasets (see Table 23). The highest value was achieved in the Wenchuan dataset with a $\kappa = 0.7991$ and the lowest in Atlantico dataset with a $\kappa = 0.2901$. One of the most challenging datasets presented in the three scenarios is the Wenchuan dataset. In that dataset, six of the comparison methods achieved a negative κ (meaning that the performance of the method is even worse than the one obtained by random selection) and only the rrR-EM and GBF-CD achieved a positive value of $\kappa = 0.2380$ and $\kappa = 0.3239$ respectively. Furthermore, the G-SMO-CD obtained the highest R of 98.63% and 80.66% (see Table 39 in the supplementary material), the lowest OE 4.48 and 5.89 (see Table 24) for Katios and the Wenchuan datasets respectively, and the highest P of 86.53 in Wenchuan dataset. The G-SMO-CD also presented the lowest FP and FN rate of 0.095 in the Katios and 13.46 in the Wenchuan datasets respectively.

Lastly for the third scenario of the heterogeneous cases, the G-SMO-CD outperformed all the competing methods in 2 out of the 4 datasets. The G-SMO-CD achieved the highest κ of all the comparison methods in the Bastrop dataset with a value of $\kappa = 0.9566$. Here, the G-SMO-CD obtained the lowest FN and the highest P in Bastrop and Gloucester-2 datasets (see Tables 36 and 38 respectively in the supplementary material).

With respect to the execution times, presented in Table 25, the methods rrR-EM and IRGM are the fastest in the change map estimation. However, the best overall results are achieved by the proposed method G-SMO-CD, that's execution times for large datasets (i.e. the homogeneous cases) does not take more than approximately two minutes to compute. Meanwhile, the competing methods (with the exception of the rrR-EM and IRGM) take more than ten times longer than our proposal, to achieve results that in general are not better than ours (see Tables 23 and 25).

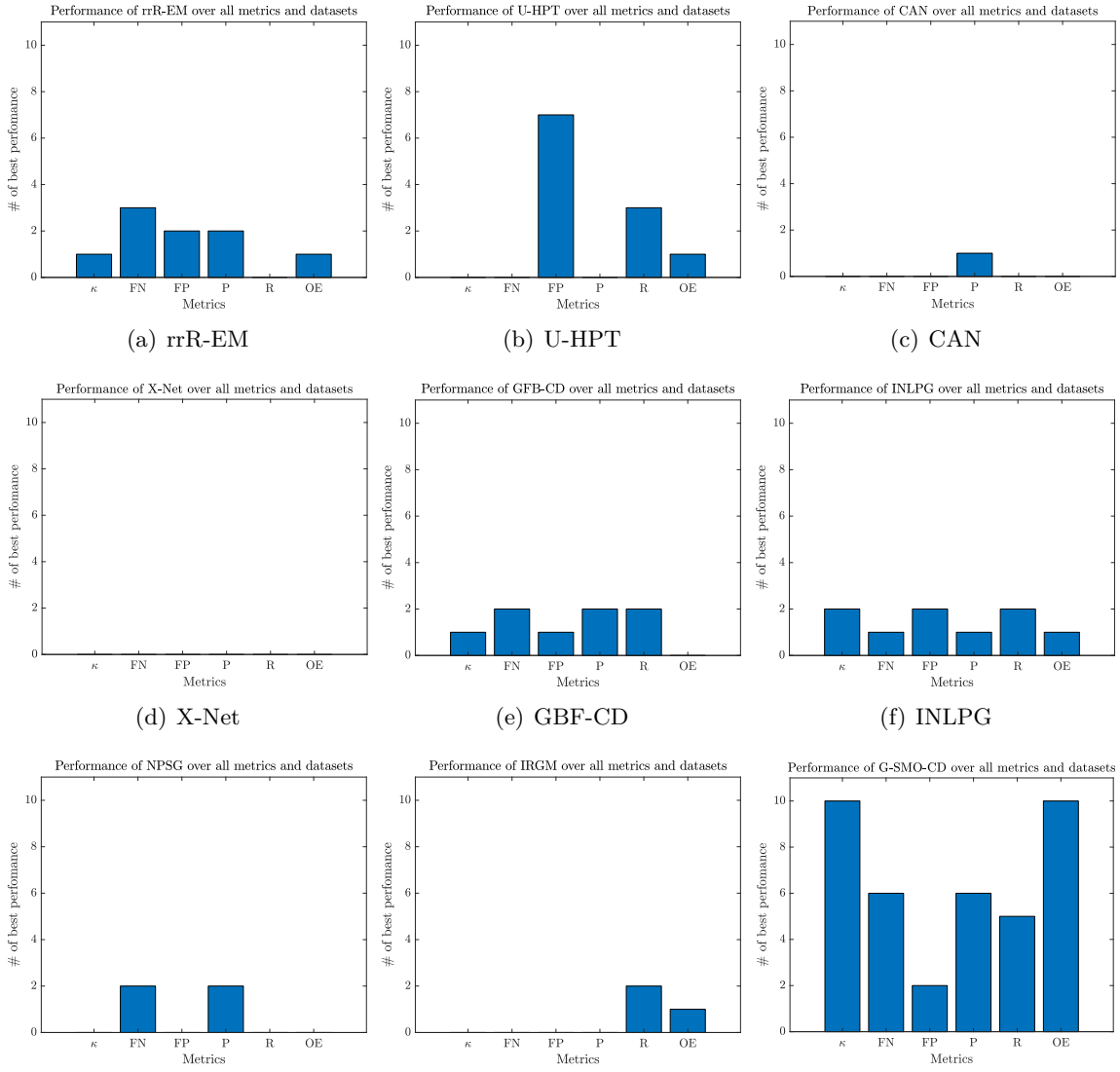
In addition, Figure 18 shows that the **G-SMO-CD** method outperformed the competing approaches in all of the metrics except for false positive (FP). On the one hand, the smoothness approach (**G-SMO-CD**) outperformed the competing models (**rrR-EM** [Zanetti and Bruzzone, 2018], **U-HPT** [Luppino et al., 2019], **CAN** [Niu et al., 2018], **X-Net** [Luppino et al., 2021], **GBF-CD** [Jimenez-Sierra et al., 2020], **INLPG** [Sun et al., 2021d], **NPSG** [Sun et al., 2021c], and **IRGM** [Sun et al., 2021a]) in terms of Cohen's Kappa coefficient ($\boldsymbol{\kappa}$) and overall error (OE) in 10 out of the 14 datasets. For the metrics false negatives (FN) and precision (P), G-SMO-CD surpasses the comparison models in 6 out of 14 datasets. In the case of the Recall (R) metric, G-SMO-CD achieved the best value in 5 out of 14 datasets. And for the false positive metric (FP), G-SMO-CD achieved the best values for 2 out of 14 datasets. It is important to point out, that the **rrR-EM**, **GBF-CD**, and **INLPG** approaches achieved the best $\boldsymbol{\kappa}$ in the remaining 4 datasets for a total of 14, which includes the Omodeo, Atlantico, and Toulouse/Gloucester-2 datasets respectively. The machine learning method (**U-HPT**), deep learning approaches (**CAN** and **X-Net**), and the graph-based methods (**NPSG** and **IRGM**) did not outperform the competing methods in any of the 14 datasets with respect to the κ metric. **X-Net** is the method that showed the poorest performance, which did not surpass in any metric the competing methods for the 14 datasets. In this benchmark, **CAN** came after **X-Net** in terms of performance. **CAN** only did better, with respect to the P metric in the Omodeo dataset (see Table 38 in the supplementary material). Furthermore, we note that even though the learning-based models such as **U-HTP** are trained on samples labeled using the ground truth, such models seem unable to generalize well over the entire dataset. This can be a consequence of not using a large enough set of training examples, which is unrealistic in practice because the ground truth is unknown. Unlike this approach, our method produces notably better results even though there is no training with clean labels except for

the selection of a regularization parameter that controls the amount of filtering applied to a noisy prior change map on the proposed fused graph. Similar to our approach, **X-Net** based on **CAN** also relies on a noisy prior change map, which is used to generate noisy training examples for a change detection framework based on image translation. This approach, **X-Net**, however, exhibits poorer performance than our graph-based method as shown by our results. The poor performance of **X-Net** appears to be due to the image translation process that fails to transform the pre- and post-event images into the same feature domain, particularly in the case of heterogeneous datasets such as Toulouse, California, Bastrop, and Gloucester-2.

Lastly, in Table 26 we present the performance of our proposed method with respect to the ablation approach. In the ablation process, we remove the GL based on GSP from the proposed framework and the proposed optimization based approach to find the change map. Instead, we use the classical approach explained in Section 3, which learns the graph by using the Gaussian kernel with a a value of $\sigma = \text{mean}(\|v_i - v_j\|_2)$. The results shown in Table 26, confirm that the GL and the optimization stages improve the results for the proposed graph-based approach. In addition, supported by the results we demonstrate that our proposal is useful for creating generic models of signal structure that enhance change detection even for heterogeneous datasets.

Table 26: Ablation test in terms of $\boldsymbol{\kappa}$.

Dataset	**G-SMO-CD**	**Ablation**
Mulargia	**0.9165**	−0.0182
Omodeo	**0.6029**	−0.0343
Alaska	**0.9106**	0.4822
Madeirinha	**0.8303**	0.1115
Prince George	**0.8319**	−0.1293
Gloucester-1	**0.7033**	−0.0388
Katios	**0.7103**	0.0316
Atlantico	**0.2901**	0.0465
San Franciso	**0.6947**	0.0810
Wenchuan	**0.7991**	−0.2010
Toulouse	**0.2287**	−0.0570
California	**0.3850**	−0.0208
Bastrop	**0.9566**	−0.0929
Gloucester-2	**0.1685**	−0.0422



(a) rrR-EM (b) U-HPT (c) CAN

(d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD

Figure 18: Overall performance of the models. Bar charts that evaluate the performance of each method over all metrics and datasets. The count for each method in one of the six possible metrics means that in one dataset, the model outperformed all the competing methods in that metric.

5 Graph Subspace Fusion and Non-local Block Matching

This chapter describes an improvement of the previous proposed graph-based approaches (i.e. applied to change detection and biomass estimation) explained before. Where we focus on a novel way to fuse the graphs by addressing the fusion as a subspace fusion by using total variation (TV) and l_0 norm to extract relevant features for biomass estimation and change detection respectively. Furthermore, we propose a non-local block matching (Nl-BM) as a solution to the pixel neighborhood restriction that classic super-pixel based approaches imposes.

5.1 Non-local Block Matching and l_0 norm for CD

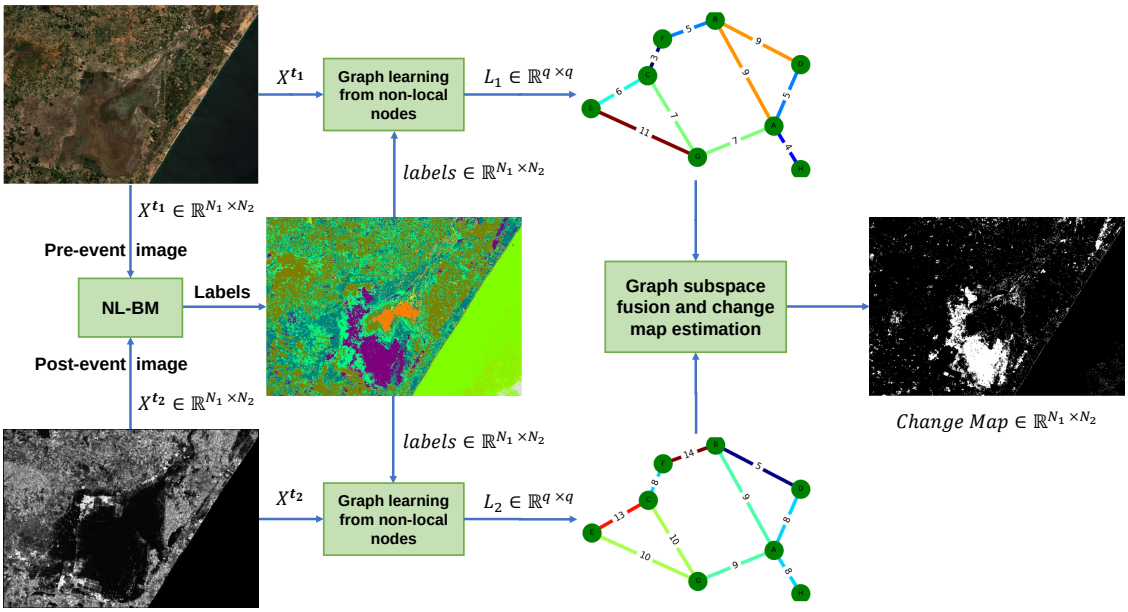


Figure 19: Proposed framework for graph subspace fusion and change map detection. Our proposal has three main stages: (a) The number q of non-local stacked patches (regions) generated by the proposed NL-BM from the input image $\mathbf{X}_\mathbf{i}^\mathbf{t}$, (b) graph learning by using smoothness prior, and (c) graph subspace fusion and change map estimation.

In this section, we propose a framework to perform graph fusion over subspaces with l_0 norm regularization and present a non-local relationship of similar patches to be used as graph nodes. In short, first we perform non-local block matching to generate the nodes, then we learn the graphs by using prior smoothness (see Section 4.1.2), and finally we apply the propose graph subspace fusion and change map estimation explained below (See Figure 19).

The classical block matching (BM) in [Dabov et al., 2007], was used alongside with a collaborative filtering. For our proposal, we will just discuss the BM part that it is of our interest. In [Dabov

et al., 2007] the BM procedure aims to find signal fragments similar to a given reference as follows:

- Divide the image into windows with size $l_w \times l_w$.
- Divide each local window in small patches with size $\ll l_w \times l_w$ and with a given stride for both rows and columns (i.e. overlapping the patches).
- For each iteration of the patches within the local window, compare the current patch similarity against the overlapping patches.
- Stack in a $3D$ array the patches that are the most similar to the reference.

In short, after grouping the patches the authors in [Dabov et al., 2007] apply a collaborative filtering to denoise an image.

Inspire by the previous mentioned method for BM, we extend this notion to a non-local BM (NL-BM) to group similar structures in a image. In contrast with the classical BM, on the one hand we do not use overlapping patches since we are interested in grouping similar patches to generate regions that share similar features in remote sensing images (i.e. MS, SAR, and SAR/MS). On the other hand, since we need to combine temporal information we propose the use of a compositive image and an automatic tolerance selection.

5.1.1 Proposed NL-BM

Consider an image $X^{t_i} \in \mathbb{R}^{N_1 \times N_2 \times c}$, where t_i indicates image acquisition time (i.e. t_1 for pre-event and t_2 for post-event), and c is the number of spectral bands. Then, since change detection involves two images captured in times t_1 and t_2, the NL-BM process must contain information of the pre- and post- events images. To do so, we generate what we call a composite image whose first two bands are always the gray scale images for RGB bands (i.e. by using a weighted sum of the of the R, G, and B components) of the pre- and post- events respectively if available. For the rest of the bands, we apply the following procedure depending on the sensor that captures the images:

- **Homogeneous multi-spectral cases:** we use as input for the NL-BM algorithm the magnitude of difference image (MDI) $MDI = \left\| X^{t_1} - X^{t_2} \right\|_2$.
- **Homogeneous SAR cases:** we use as input for the NL-BM algorithm a false RGB image. The band one is related to the pre-event image, band two is the post-event image and the band three is the MDI.
- **Heterogeneous (multi-spectral/SAR) cases:** here, we use as input the MDI. But, since one of the images has one band related to the SAR sensor, we replicate this band for the same number of bands in the MS image to perform the MDI.

In order to perform the NL-BM over the composite image, first of all we generate patches P_i with size $B_s \times B_s$ uniformly distributed over the image. Furthermore, since the patch must contain all the spectral information within the image (i.e. for the MS cases), we extend the patch along the third dimension as $P_i \in \mathbb{R}^{B_s \times B_s \times c}$, where i indicates the i-th patch and c is the number of the spectral bands. Then, the main idea of the proposed NL-BM is to iteratively match the patches available to stack as similar defined by the measure of a similarity metric $\mathcal{M}$ and a tolerance ϵ. The similarity metric uses as input the vectorized version of the patch as $p_i \in \mathbb{R}^{(B_s \times B_s \times c)}$ as shown in Figure 20). Once we have the patches of the composite image, we now must select a reference patch

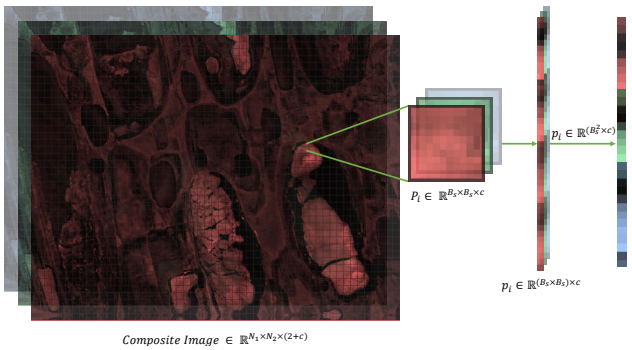


Figure 20: Patches generation from the MDI and its vectorization.

to measure the similarity with respect to the other patches available to stack them if the similarity metric is below the tolerance ϵ and update the available patches (i.e. once a set of patches are stacked can not be considered in posterior iterations). This process continues iteratively until there are no more patches available as shown in Algorithm 1.

To select the reference patch as the one that contains most of the information among all the patches, we select the reference as the patch with the highest variance level (i.e. radiometric values).

We find out that the Spectral Angle Mapper (SAM) in classification and pansharpening applications is a reliable metric to compare spectral information because it is not highly affect by the intensity or spectral content [Vivone et al., 2014, Shivakumar and Rajashekararadhya, 2017].Therefore, we use the cosine distance $(1 - Cosine(v_1, v_2))$ as the similarity metric to select whether a patch is similar to others or not.

The time complexity of Algorithm 6 taking into account the upper-bound is $\mathcal{O}(B_s^2 c(\#Patches)i)$, where B_s is the dimension of the patches, c are the channels and i are the number of iterations.

Algorithm 6: NL-BM algorithm.

Input: Composite image $\in \mathbb{R}^{N_1 \times N_2 \times (2+C)}$, B_s, ϵ
Output: Labeled map $\in \mathbb{R}^{N_1 \times N_2}$ with q labels
Initialize: $P = mat2tiles(\text{Composite image}, B_s)$, Reference patch $p_{\text{ref}} = \{max(var(p_i)) \forall i \subset P\}$, Available Patches $\in \{0,1\}^{\#Patches}$, Match = true, Labeled map $= zeros(N_1, N_2)$, Label = 1
while *Match* **do**
  (1) Measure the similarities among the reference patch and the ones available:
  Similarities $= \{\mathcal{M}_{dcos}(p_{\text{ref}}, p_i), \forall i | \text{Available Patches}(i) = 1\}$
  (2) Select as match the patches that satisfies the rule:
  Matches = (Similarities $< \epsilon$)
  (3) Update the available patches:
  Available Patches$(i) = \{0, \forall i | Matches(i) = 1\}$
  (4) Indicate the similar patches by using a label in the input space ($\mathbb{R}^{N_1 \times N_2}$):
  Labeled map$(r,s) = \{\text{label }, \forall P_i \in (r,s) \mid Matches(i) = 1, i \in \{1, 2, \ldots, \#\text{Patches}\}\}$
  (5) Update the new reference patch:
  $p_{\text{ref}} = \{max(var(p_i)) \forall i \subset P | \text{Available Patches}(i) = 1\}$
  (6) Label = Label + 1
  if *Available Patches*$(i) = \{0, \forall i \in 1, 2, \ldots, \#Patches\}$ **then**
    Match = false
  end
end

5.1.2 Automatic tolerance selection

For the automatic selection of the tolerance ϵ, we make use of a well known practice applied in different fields which is the use of the histogram of an image to select a proper threshold either to segment the image or to make a selection of the data for posterior processes [Li et al., 2020, Mohamed Shakeel et al., 2020, Li et al., 2021b]. Therefore, we make use of the histogram thresholding but instead of applying it to the image pixels, we apply it over the similarity metric in the first iteration of the NL-BM (i.e. since we have all the patches available to obtain the global tolerance). From the histogram, since we are measuring the similarity as the cosine distance most of the data will be accumulated in the first bins as the closest patches to the reference (i.e. due to the low variance between the values presented in remote sensing images [Yavariabdi and Kusetogullari, 2017, Li et al., 2022] which makes the CD application a complex task to address). In consequence, we select the center of the first bin as the tolerance ϵ as shown in Figure 21.

5.1.3 Graph Subspace Fusion and Change map detection

The proposed graph-based approaches [Jimenez-Sierra et al., 2020, Jimenez-Sierra et al., 2021a, Jimenez-Sierra et al., 2022] as described in previous sections carries out the fusion of the graphs

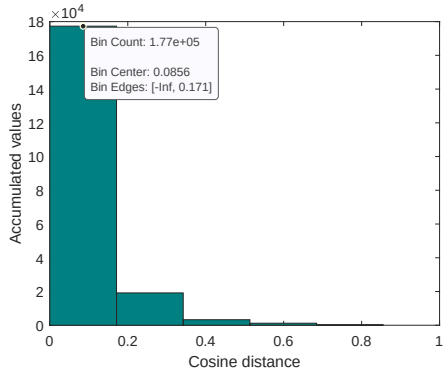


Figure 21: Selection of ϵ from the histogram of the cosine distances in the first iteration of the NL-BM algorithm. For 5 bins we take the center of the first bin of the cosine distance in the x axis as the value of $\epsilon = 0.0856$. This example is for the Mulargia dataset.

and the change detection map as separated stages. Here, inspired by the work in [Rasti and Ghamisi, 2020] we perform the fusion by merging the subspaces generated by the Laplacian matrices of the graphs and also consider the change map detection in the same optimization problem as follows:

$$\begin{aligned} c \quad = \quad & \underset{F,V_1,V_2,x\in\mathbb{R}^q}{\arg\min} \ \tfrac{\lambda_1}{2}\left\|L_1 - FV_1^\top\right\|_{\mathcal{F}}^2 + \tfrac{\lambda_2}{2}\left\|L_2 - FV_2^\top\right\|_{\mathcal{F}}^2 \\ & + x^\top Fx + \lambda_3 \left\|x\right\|_0 \\ \text{S.T} \quad & V_1^\top V_1 = I \text{ and } V_2^\top V_2 = I, \end{aligned} \tag{34}$$

where $\|.\|_{\mathcal{F}}$ is the frobenius norm, L_1 and L_2 are the Laplacian matrices (i.e. learned by using prior smoothness, see Section 4.1.2) related to the pre-event and post-event respectively, V_i are the subspace basis generated by L_i, F is the common subspace generated by the Laplacian matrices, and x is the desired change map.

With this novel way to fuse the graphs and detect the change map we aim to: i) *Remove* the dependency of a prior knowledge in the change map detection (i.e. some methods make use of a prior image either to train a model or to lead the change map detection [Jimenez-Sierra et al., 2020, Jimenez-Sierra et al., 2021a, Luppino et al., 2021, Jimenez-Sierra et al., 2022]), ii) Apply a *sparsity penalty* since the change maps are sparse which is an intrinsic prior knowledge, iii) Represent the fused graph with a few learnable fused features (F) to estimate the change map. The fused features are *co-living in the subspaces of each graph.*

In order to solve the optimization problem in (34) we apply variable splitting [Afonso et al., 2010]:

$$\begin{aligned} c \quad = \quad & \underset{F,V_1,V_2,S,x\in\mathbb{R}^q}{\arg\min} \ \tfrac{\lambda_1}{2}\left\|L_1 - FV_1^\top\right\|_{\mathcal{F}}^2 + \tfrac{\lambda_2}{2}\left\|L_2 - FV_2^\top\right\|_{\mathcal{F}}^2 \\ & + x^\top Fx + \lambda_3 \left\|S\right\|_0 \\ & \text{S.T} \quad V_1^\top V_1 = I \text{ and } V_2^\top V_2 = I, \text{ and } x = S, \end{aligned} \tag{35}$$

Then to remove the constraint on x in (35), we add it as a penalty term:

$$\begin{aligned} c \quad = \quad & \underset{F,V_1,V_2,S,x\in\mathbb{R}^q}{\arg\min} \ \tfrac{\lambda_1}{2}\left\|L_1 - FV_1^\top\right\|_{\mathcal{F}}^2 + \tfrac{\lambda_2}{2}\left\|L_2 - FV_2^\top\right\|_{\mathcal{F}}^2 \\ & + x^\top Fx + \lambda_3 \left\|S\right\|_0 + \tfrac{\mu}{2}\left\|S - x - L_m\right\|_2^2 \\ & \text{S.T} \quad V_1^\top V_1 = I,\ V_2^\top V_2 = I, \end{aligned} \tag{36}$$

where L_m is a Lagrangian multiplier. Finally, we solve the optimization problem in (36) by using the cyclic descent algorithm [Marjanovic and Solo, 2014] (i.e. solve the problem with respect to one of the unknown variables while the other ones are assumed to be fixed) to obtain an iterative solution with respect to F, V_1, V_2, x, and S as follows:

1. By fixing V_1, V_2, and x, problem in (36) turns to the following subproblem with respect to F:

$$\begin{aligned} \underset{F}{\arg\min} \quad & \tfrac{\lambda_1}{2}\left\|L_1 - FV_1^\top\right\|_{\mathcal{F}}^2 + \tfrac{\lambda_2}{2}\left\|L_2 - FV_2^\top\right\|_{\mathcal{F}}^2 \\ & + x^\top Fx, \end{aligned} \tag{37}$$

   this can be solved by taking the derivative with respect to F and set it to zero, giving as a result:

$$F = (\lambda_1 L_1 V_1 + \lambda_2 L_2 V_2 - xx^\top)/(\lambda_1 + \lambda_2). \tag{38}$$

2. By fixing the V_i variables (36) it is reduced to the following low-rank orthogonal Procrustes problem:

$$\begin{aligned} \underset{V_i}{\arg\min} \quad & \tfrac{\lambda_i}{2}\left\|L_i - FV_i^\top\right\|_{\mathcal{F}}^2 \\ & \text{S.T} \quad V_i^\top V_i = I, \end{aligned} \tag{39}$$

   then the solution is given by:

$$V_i = A_i C_i^\top, \text{where } SVD(L_i^\top F) = A_i \Sigma_i C_i^\top. \tag{40}$$

3. By fixing F, S, and L_m the subproblem w.r.t. to x is given by:

$$\underset{x}{\arg\min} \quad x^\top Fx + \tfrac{\mu}{2}\left\|S - x - L_m\right\|_2^2, \tag{41}$$

in consequence, the solution of the change map is given by:

$$x = (F + F^\top + \mu I)^{-1}\mu(S - L_m). \tag{42}$$

4. The subproblem w.r.t. S is the following:

$$\arg\min_S \quad \lambda_3 \|S\|_0 + \tfrac{\mu}{2}\|S - x - L_m\|_2^2, \tag{43}$$

then the solution of the previous problem is given a hardthresholding as shown below:

$$S = (x + L_m)\mathbf{1}_{|x+L_m|>\frac{\mu}{\lambda_3}}. \tag{44}$$

5. Lastly, we need to update the Lagrangian multiplier:

$$L_m = L_m + x - S. \tag{45}$$

The time complexity of solving this optimization problem taking into account the upper-bound is given by the SVD algorithm and it is $\mathcal{O}(q^3)$, where q is the number of regions/super-pixels generated.

5.1.4 Result and Discussion

Dataset's Description: We tested our approach on 20 real change detection scenarios captured by the MS and SAR sensors presented in Figure 22 and Table 27. In particular, the datasets were acquired by different satellites such as Landsat, ALOS/PALSAR, ERS-2 SAR, Sentinel 1A, Sentinel 2A, and ESA/ASAR (see Table 27). The datasets contains low resolution images such as 275×400 and high resolution images of 4220×2320. These datasets include diversity of events, such as earthquakes, floods, wildfires, melted ice, tsunamis, volcano eruption, farming, and construction (building bridges, etc.). These datasets incorporate 6 MS homogeneous cases, 4 SAR homogeneous cases, and 10 heterogeneous cases which combine SAR/MS or MS/SAR images. All the details of the datasets are presented in Table 27 and the images are shown in Figure 22.

Experimental set-up: We ran all the codes in MATLAB®2021b using a server with two processors, Intel(R) Xeon(R) CPU E5-2650 v4 @2.20GHz, with a total of 24 physical cores, 48 threads, and 252 GB of memory RAM @2400MHz. Firstly, MS images were normalized according to their maximum value and SAR images were normalized in logarithmic scale since they are in dB. Secondly, we fixed the contribution of the subspaces generated by the Laplacian matrices to be equal (i.e. $\lambda_1 = \lambda_2 = 1$) and for the μ parameter we performed a test for values from 0.1 to 1 with steps of 0.1 to ensure

[18] Available at: SF_dataset
[19] Available at : Mignotte
[20] Available at: california_dataset
[21] Available at: Bastrop_dataset
[22] Available at: Shuguang

Table 27: Databases used to evaluate the performance of the proposed method. The date format is dd/mm/yyyy.

Place	Event	Pre-event Date	Post-event Date	Size	Bands(Pre/Post)	Pre-Sensor	Post-Sensor
Italy, Mulargia lake	Flood	03/09/1995	03/07/1996	479×573	6/6	Landsat-5 TM	Landsat-5 TM
Italy, Omodeo lake	Fire	25/07/2013	10/08/2013	664×966	8/8	Landsat-8	Landsat-8
U.S., Alaska	Melting Ice	24/07/1985	13/07/2005	443×642	6/6	Landsat-5 TM	Landsat-5 TM
Brasil, Machadinho d'Oeste	Farming/ Construction	15/07/2000	16/07/2006	363×528	6/6	Landsat-5 TM	Landsat-5 TM
Canada, Prince George	Fire	05/07/2017	22/08/2017	2507×1904	6/6	Landsat-8	Landsat-8
U.K., Gloucester-1	Flood	05/09/1999	17/11/2000	4220×2320	3/3	SPOT	SPOT
Colombia, Katios National Park	Fire	10/03/2019	27/04/2019	879×1319	1/1	Sentinel 1A	Sentinel 1A
Colombia, Atlantico	Flood (dam)	28/04/2010	16/03/2011	729×1056	1/1	ALOS/PALSAR	ALOS/PALSAR
U.S., San Francisco [18]	Flood	10/08/2003	16/05/2004	275×400	1/1	ERS-2 SAR	ERS-2 SAR
China, WenChuan	Earthquake	03/03/2008	16/06/2008	301×442	1/1	ESA/ASAR	ESA/ASAR
France, Toulouse-1 [19]	Building	10/02/2009	15/07/2013	2604×4404	1/1	TerraSAR-X	Pleiades
U.S., California [20]	Flood	11/01/2017	26/02/2017	3500×2000	6/1	Landsat-8	Sentinel 1A
U.S., Bastrop [21]	Fire	08/09/2011	22/10/2011	1534×808	9/6	Landsat-5 TM	EO-1 ALI
U.K., Gloucester-2 [19]	Flood	14/06/2006	25/07/2007	4220×2320	3/1	Quickbird 02	TerraSAR-X
Spain, Palma	Volcano	08/02/2011	30/09/2021	696×1155	1/10	ALOS/PALSAR	Sentinel 2A
China, Shuguang [22]	Building	18/07/2006	20/09/2012	342×419	1/3	RadarSat-2	Google Earth
India, Kaliveli Lake	Flood	30/05/2021	16/11/2021	1400×2000	10/1	Sentinel 2A	Sentinel 1A
Hong-Kong, Town Island [19]	Building	06/06/2008	21/06/2013	403×415	1/3	RadarSat-2	Google Earth
Japan, Watari Iwanuma	Tsunami	24/08/2010	13/03/2011	3050×2255	6/1	Landsat-5 TM	ALOS/PALSAR
Japan, Higashimatsushima	Tsunami	19/03/2011	01/08/2017	969×969	1/10	TerraSAR-X	Sentinel 2A

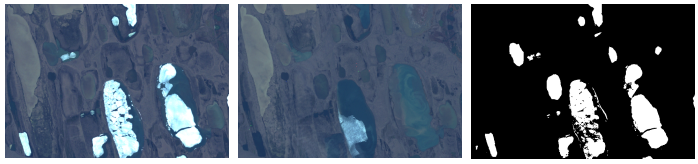
the convergence of the solution in 400 iteration, hence, we fix $\mu = 0.8$. Afterwards, we carried out a grid-search for the parameters K for the GL based on smoothness and for the parameter λ_3 that regularized the sparsity level of the change map. The grid search over K starts from $K = 4$ to the maximum number of regions in steps of one, and for μ stars from 0.005 to 1 with steps of 0.005. To summarize, our experiments consist of: i) Normalization of the data, ii) NL-BM, iii) Tuning of the parameters, and iv) Fusion and Change map estimation. With respect to the eleven state-of-the-art methods used for comparison **SCCN** [23] [Liu et al., 2016], **CAN**[24] [Niu et al., 2018], **X-Net**[25] [Luppino et al., 2021], **CA-Ae**[26] [Luppino et al., 2022], **INLPG**[27] [Sun et al., 2021d], **NPSG**[28] [Sun et al., 2021c], **IRGM**[29] [Sun et al., 2021a] [Sun et al., 2021a], **G-SMO**[30] [Jimenez-Sierra et al., 2022], **SCASC**[31] [Sun et al., 2021b], **GIR-MRF**[32] [Sun et al., 2022b], and **IST-CRF**[33] [Sun et al., 2022a], we used the parameters recommended by these authors. We refer to our proposed approach as **NL-BM-GSF** in the tables 29 to 32.

Parameter Settings: An important parameter that impacts directly in the final label map given by the proposed NL-BM, is the number of the bins to generate the histogram of the similarity measures. Therefore, we conducted an experiment by changing the number of the bins from 4 to

[23] Reproduced by authors in [Luppino et al., 2022] at: github.com/llu025/SCCN
[24] Reproduced by authors in [Luppino et al., 2021] at: github.com/llu025/CAN
[25] Available at: github.com/llu025/X-Net
[26] Available at: github.com/llu025/CA-Ae
[27] Available at: github.com/yulisun/INLPG
[28] Available at: github.com/yulisun/NPSG
[29] Available at: github.com/yulisun/IRG-McS
[30] Available at: github.com/DavidJimenezS/G-SMO-CD
[31] Available at: github.com/yulisun/SCASC
[32] Available at: github.com/yulisun/GIR-MRF
[33] Available at: github.com/yulisun/IST-CRF

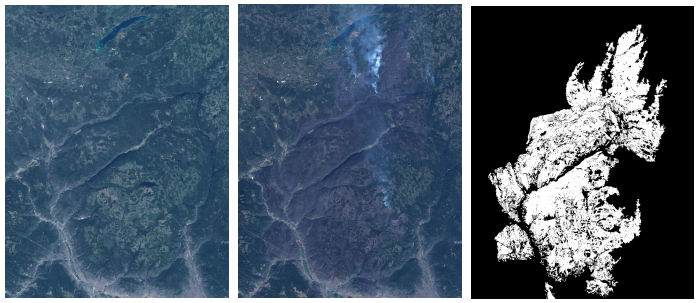
(**a**) Mulargia

(**b**) Omodeo



(**c**) Alaska

(**d**) Madeirinha

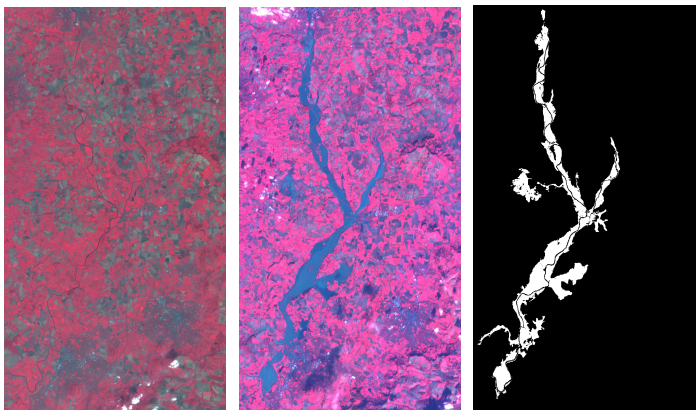


(**j**) Prince George

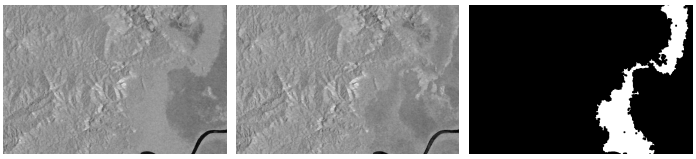
Figure 22: *Continued.*

10 (see Figure 23) and the best results were achieved for 4 bins for homogeneous cases and 5 bins for heterogeneous cases. In addition, we performed a experiment with different patch sizes from 2 to 5, and the structures were well preserved with a patch size of 2.

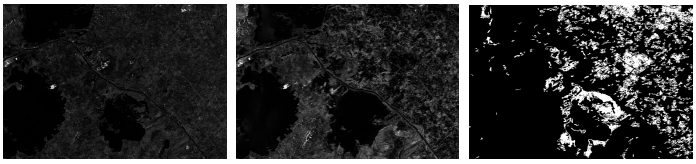
In Figure 23, we find out that the higher the amount of bins we use the higher the number of regions will be generated. Additionally, more redundancy may be presented since the number of bins define the groups according to the cosine distances, therefore the more bins are used the longer distance is considered that could lead that two similar structures may be seem as different ones. Nonetheless, most of the changes presented in the datasets are represented by large changes and



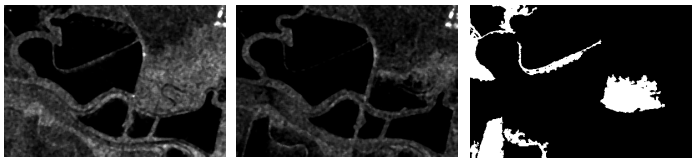
(**l**) Gloucester-1



(**e**) Katios National Park



(**f**) Atlantico



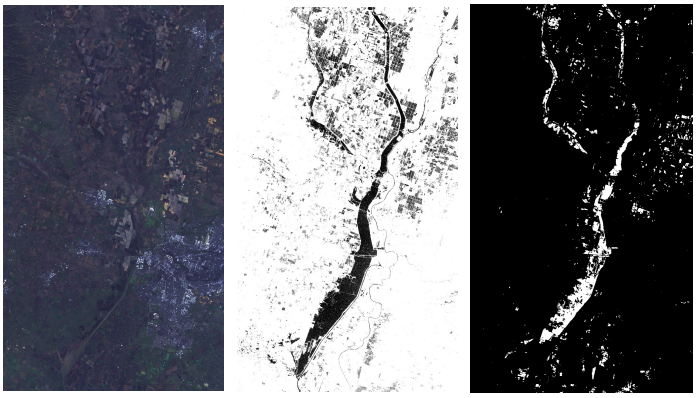
(**g**) San Francisco

(**h**) WenChuan

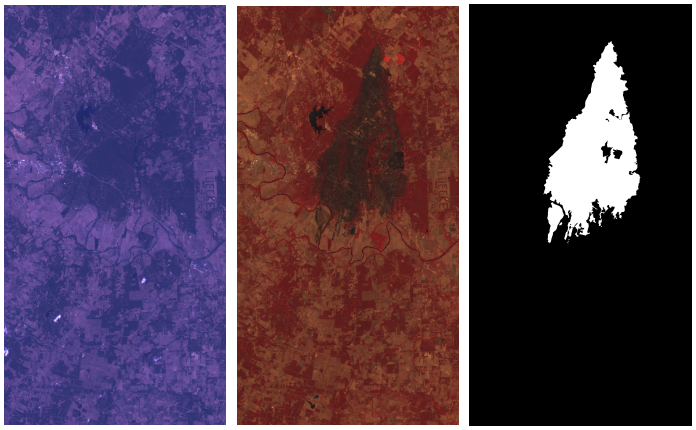
(**i**) Toulouse

Figure 22: *Continued.*

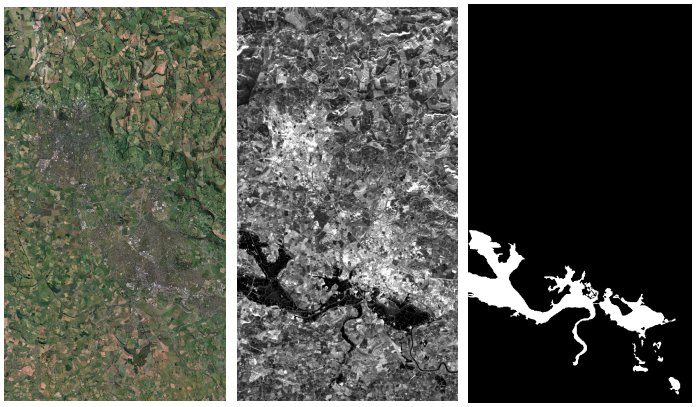
the details are preserved by using a small patch such it is 2×2 over the images. In consequence, the best parameters for K and μ were found through grid search and are tabulated in Table 28.



(**k**) California



(**m**) Bastrop

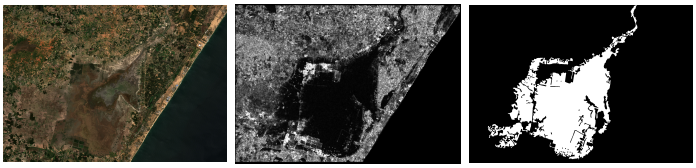


(**n**) Gloucester-2

(**o**) Palma

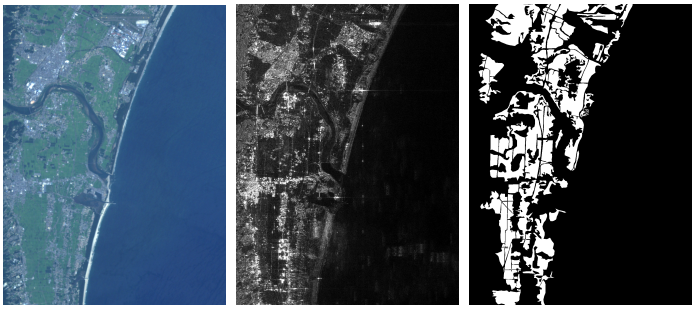
Figure 22: *Continued.*

(**p**) Shuguang



(**q**) Kaliveli Lake

(**r**) Town Island

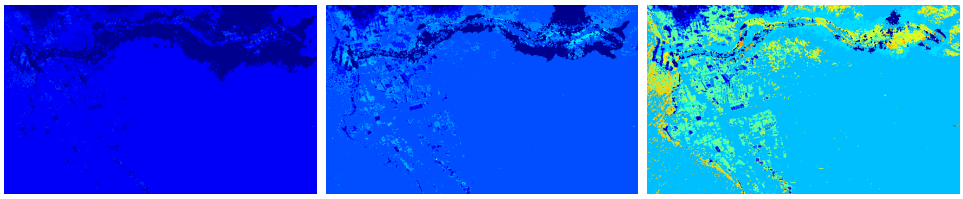


(**s**) Watari Iwanuma

(**t**) Higashimatsushima

Figure 22: Datasets to test the proposed framework for CD. From left to right: pre-event, post-event, and the reference change map. The RGB images shown does not implies the non existence of more channels (see Table 27).

Results: First of all, we show some results of the proposed NL-BM. Secondly, we analyzed three scenarios: MS homogeneous, SAR homogeneous and MS/SAR or SAR/MS heterogeneous. We measure performance based on the metrics: Cohen's kappa coefficient ($\boldsymbol{\kappa}$), overall error (OE), and execution time (ET), in Tables 29 to 31. Additionally, to provide a visual comparison between the



(a) 5 bins (b) 7 bins (c) 10 bins

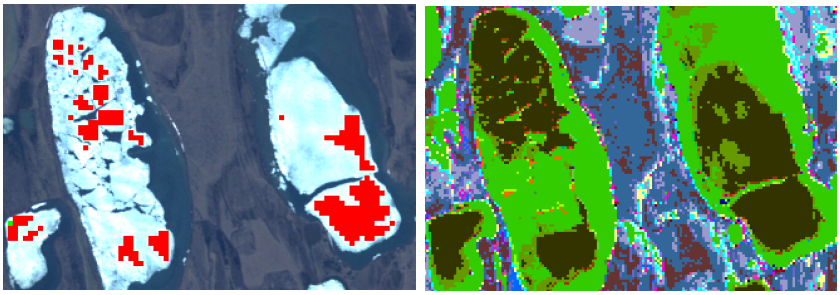
Figure 23: Final label map in function of the number of bins for Palma dataset. With 5 bins the final label map has 36 regions, with 7 there are 69 regions and with 10 a total of 141 regions.

Table 28: Parameters used for evaluation of the datasets.

Homogeneous Cases				**Heterogeneous Cases**			
Dataset	**# of Nodes**	**K**	$\boldsymbol{\lambda_3}$	**Dataset**	**# of Nodes**	**K**	$\boldsymbol{\lambda_3}$
Mulargia	76	38	0.160	Toulouse-1	26	9	0.880
Omodeo	140	21	0.470	California	693	192	0.120
Alaska	72	18	0.340	Bastrop	143	22	0.230
Machadinho d'Oeste	88	27	0.215	Gloucester-2	106	67	0.0350
Prince George	57	14	0.675	Palma	36	18	0.030
Gloucester-1	22	16	0.610	Shuguang	287	101	0.020
Katios	11	9	0.975	Kaliveli Lake	55	14	0.560
Atlantico	13	7	1	Town Island	175	37	0.040
San Franciso	86	83	0.030	Watari Iwanuma	56	23	0.595
Wenchuan	54	19	0.305	Higashima- tsushima	173	95	0.040

methods, we present the change maps for the heterogeneous cases from Figure 25 to Figure 34 (for the homogeneous cases see the Figure 57 to Figure 66 in the supplementary material). The maps show false negatives (FN), false positives (FP) and correct pixels (C). Then we perform an analysis by using the two-way analysis of variance (ANOVA) [Yigit and Mendes, 2018, Wen et al., 2022] to show if the proposed method performance across the datasets is remarkably different from the comparison methods (see Figure 56). Finally, Table 32 shows the ablation results. In the ablation process, we remove the proposed optimization problem that involves both the graph fusion and the change map estimation, instead we use the proposed fusion rule and optimization problem to obtain the change map used by authors in [Jimenez-Sierra et al., 2020] (explained in Section 3.1). Figure 24 (a) is shows that the proposed NL-BM is able to find relationships between patches that are not close in the image domain but indeed they refer to a similar structure or object in the images. In Figure 24 (b), the final results exhibit regions that are not constrained to local neighbors and therefore we induce non-locality knowledge in the segmentation.

Concerning to MS homogeneous cases given in Figures 57 to 62 (see the supplementary material), our proposed NL-BM-GSF method surpasses the 11 comparison methods in 4 out of the 6 datasets (see Table 29), and achieves the second best in one (i.e. Omodeo) of the two datasets (i.e. Omodeo and Canada) where our model does not outperform the comparison methods. The NL-BM-GSF



(a) 1st Iteration (b) Final result

Figure 24: NL-BM first iteration results. Where in (a) the green patch is the reference patch in the first iteration and the red patches are the ones stack as similar to the reference, and in (b) is shown the regions generated by the NL-BM.

Table 29: Models performance in terms of $\boldsymbol{\kappa}$.

Dataset	SCCN	CAN	X-Net	CA-Ae	INLPG	NPSG	IRGM	G-SMO	SCASC	GIR-MRF	IST-CRF	NL-BM-GSF
Mulargia	0.3095	0.1347	0.3580	0.2694	0.4631	0.6589	0.8339	0.5746	0.7509	0.7005	0.8248	**0.8921**
Omodeo	−0.0572	0.6438	0.8828	0.8675	0.7757	0.0343	**0.8967**	0.6759	0.8602	0.8814	0.8734	0.8286
Alaska	0.7771	0.2630	0.4947	0.5035	0.6211	0.4064	0.1608	0.2972	0.6787	0.6351	0.7681	**0.8164**
Machadinho d'Oeste	0.2475	0.1357	0.4171	0.6091	0.4870	0.4901	0.4228	0.8513	0.3805	0.4554	0.4097	**0.8548**
Prince George	−0.1101	0.2464	0.3383	0.2333	0.2515	0.0401	0.1382	−0.1402	0.2606	0.2761	**0.6059**	0.5330
Gloucester-1	0.0738	0.0300	0.5385	0.4795	0.1801	0.0068	0.5413	−0.0733	0.5066	0.5848	0.5828	**0.6543**
Katios	−0.1210	0.1503	0.1045	−0.0046	0	−0.0246	0.0235	0.7103	0.6070	−0.0655	−0.0603	**0.7208**
Atlantico	−0.1385	0.2315	0.1861	0.5517	0.2759	0.1386	0.1661	0.2901	0.0760	0.0115	0.3627	**0.7411**
San Franciso	−0.1657	0.4775	0.4107	–	0.6767	0.1440	0.4446	0.6486	0.0109	0.4311	**0.7502**	0.5638
Wenchuan	−0.1827	−0.0790	−0.0693	−0.0740	−0.2134	−0.1789	−0.0463	**0.7758**	−0.0111	−0.1386	−0.0576	0.5151
Toulouse-1	0.2136	−0.0266	0.2085	0.2082	0.3118	0.0029	0.2688	0.0322	0.3359	**0.4439**	0.3402	0.1693
California	0.3730	0.2797	−0.0579	0.2575	0.1825	0.0208	**0.4660**	0.0048	0.3406	0.3216	0.4347	0.3136
Bastrop	−0.1055	0.4249	0.6848	0.7696	0.3721	0.0357	0.1607	**0.8928**	0.5501	0.8314	0.2080	0.8788
Gloucester-2	0.2666	0.0006	0.1427	−0.0592	0.0417	−0.0198	0.1035	0.1928	0.4950	0.5072	0.1514	**0.5518**
Palma	0.5549	−0.0332	0.2501	0.1797	0.4809	0.4758	0.4092	−0.2254	0.4086	0.5374	0.2273	**0.6365**
Shuguang	0.1947	0.4306	0.6376	0.6927	0.5072	0.2765	0.7016	0.0559	0.7061	**0.7145**	0.6449	0.4408
Kaliveli Lake	**0.4040**	−0.0737	−0.0545	−0.1058	0.1758	0.1101	−0.6719	0.1101	−0.0046	0.0181	−0.0253	0.3813
Town Island	−0.0548	0.2008	0.1222	0.0342	0.0881	0.0022	0.0989	0.0881	0.2635	**0.2643**	0.0138	0.2150
Watari Iwanuma	0.4250	−0.0554	−0.0763	−0.0889	−0.0051	0.0098	−0.0133	−0.2716	−0.0151	−0.0407	0.1033	**0.6092**
Higashima- tsushima	−0.1508	0.1850	0.0474	−0.1321	0.3200	−0.0789	−0.0040	0.3477	−0.0583	–	0.3890	**0.4235**

Table 30: Models performance in terms of OE (units in %).

Dataset	SCCN	CAN	X-Net	CA-Ae	INLPG	NPSG	IRGM	G-SMO	SCASC	GIR-MRF	IST-CRF	NL-BM-GSF
Mulargia	11.32	7.68	7.54	9.54	6.17	2.04	0.94	3.79	1.48	2.14	0.93	**0.60**
Omodeo	22.32	3.30	1.23	1.44	2.65	5.01	**1.02**	4.09	1.39	1.13	1.23	1.53
Alaska	4.67	13.40	9.97	11.50	9.61	11.72	12.66	27.29	6.13	12.81	4.60	**3.58**
Machadinho d'Oeste	20.64	22.86	15.29	15.24	16.76	16.39	14.98	4.89	15.92	13.60	15.38	**4.80**
Prince George	43.91	26.09	24.14	28.80	26.12	28.24	26.41	63.58	23.28	23.74	**14.50**	18.91
Gloucester-1	27.07	14.58	7.71	8.86	32.47	8.47	4.92	46.15	4.73	7.25	5.84	**3.59**
Katios	28.51	15.77	43.17	10.8	13.44	11.91	11.93	10.72	6.76	6.82	14.89	**5.45**
Atlantico	27.19	15.89	15.87	10.54	15.63	31.97	14.48	29.76	15.23	16.18	14.18	**7.38**
San Franciso	32.32	9.68	10.84	–	5.94	23.92	9.58	9.35	13.46	21.51	**5.29**	13.09
Wenchuan	32.88	22.74	22.58	22.65	82.76	39.05	21.61	**6.73**	17.88	17.46	29.80	16.14
Toulouse-1	18.96	13.72	12.40	15.33	9.51	90.34	9.93	41.36	**8.31**	12.25	8.99	19.34
California	13.72	7.54	16.92	9.26	25.31	67.09	**4.88**	77.78	5.46	14.54	5.89	7.88
Bastrop	27.95	8.43	4.78	4.04	8.61	10.80	9.93	**2.07**	7.27	5.32	9.45	2.21
Gloucester-2	18.05	15.19	14.11	18.77	28.43	80.81	10.16	25.77	6.57	15.61	10.30	**5.67**
Palma	10.48	24.62	22.69	18.40	11.57	18.68	13.00	58.54	12.58	11.82	15.30	**10.33**
Shuguang	14.31	5.54	4.46	3.80	8.55	22.16	2.82	42.67	2.88	**2.71**	3.19	5.29
Kaliveli Lake	28.40	26.47	28.46	26.69	60.93	39.54	22.46	59.60	20.57	23.55	29.06	**16.00**
Town Island	19.42	8.43	17.58	15.53	32.00	93.44	**4.24**	31.14	7.76	6.96	4.39	6.71
Watari Iwanuma	28.22	25.96	29.32	29.73	37.65	32.82	22.27	58.25	19.93	22.98	22.99	**13.83**
Higashima- tsushima	29.24	16.85	24.56	27.16	22.19	34.13	16.77	28.91	21.67	**13.11**	–	21.32

achieves the highest performance in the Mulargia dataset with a $\kappa = 0.8921$, and the lowest in the Canada dataset with a $\kappa = 0.5330$. In the Omodeo dataset, the IRGM approach achieved the

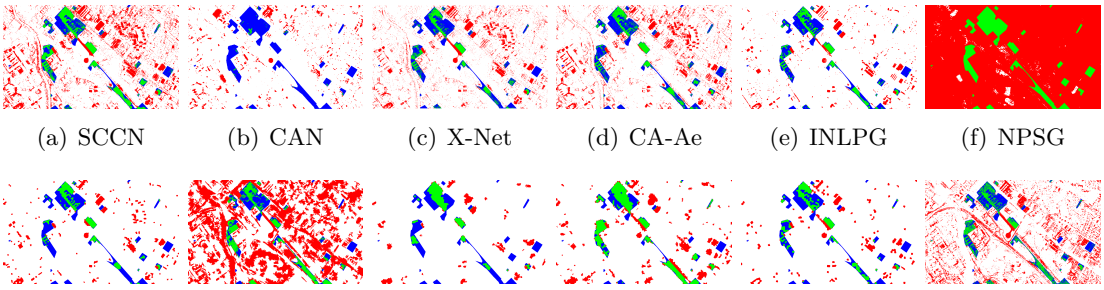
Table 31: Models computing time in seconds s.

Dataset	SCCN	CAN	X-Net	CA-Ae	INLPG	NPSG	IRGM	G-SMO	SCASC	GIR-MRF	IST-CRF	NL-BM-GSF
Mulargia	281.59	925.09	617.64	356.92	269.93	119.73	38.40	157.49	**5.51**	118.86	11.72	36.77
Omodeo	546.72	2270.26	1491.69	571.60	1966.29	316.57	43.47	60.86	**6.33**	90.74	14.03	21.74
Alaska	357.11	913.01	644.28	313.78	283.72	123.94	37.64	62.15	4.86	96.86	10.50	**4.68**
Machadinho d'Oeste	335.96	614.18	504.43	270.75	143.50	82.35	38.20	35.32	4.73	142.44	9.54	**4.67**
Prince George	909.89	4090.01	3758.42	722.08	4564.03	3458.32	56.58	146.07	**19.60**	152.53	27.46	50.80
Gloucester-1	1530.08	3470.04	3099.27	1281.39	2196.71	5485.75	53.07	17.10	**16.62**	145.83	29.74	44.51
Katios	877.44	3871.94	3620.19	256.59	993.61	400.02	36.59	60.16	5.03	93.09	8.61	**3.95**
Atlantico	771.17	2472.46	1926.65	455.16	436.50	256.00	35.47	110.86	3.75	64.22	7.25	**2.81**
San Franciso	511.69	322.49	399.45	–	15.73	39.06	32.48	1.44	**3.57**	87.19	5.94	4.03
Wenchuan	563.57	441.33	419.70	252.98	5.26	49.51	38.07	47.63	3.02	66.38	6.16	**1.95**
Toulouse-1	1935.90	3984.19	4042.90	860.63	1157.04	4504.95	51.36	96.67	13.83	134.26	24.95	45.48
California	1522.23	2427.74	1886.05	618.31	1409.73	4016.63	55.05	51.10	19.34	125.47	26.95	610.32
Bastrop	1376.66	4543.93	3929.28	985.25	4693.14	644.98	43.45	92.30	7.08	118.18	14.83	27.37
Gloucester-2	1913.62	3385.65	3074.81	748.59	1638.56	4778.59	51.79	16.54	**15.30**	126.76	28.26	89.33
Palma	1091.97	2884.19	2022.41	371.50	323.73	535.25	39.73	28.65	**5.06**	126.87	11.08	9.17
Shuguang	838.46	426.46	432.41	217.64	7.80	64.54	36.97	56.16	**4.03**	133.73	6.90	33.36
Kaliveli Lake	2431.49	986.86	675.12	1266.81	3417.83	1842.72	46.53	79.14	**10.40**	109.04	16.58	31.32
Town Island	955.80	469.81	450.63	392.11	9.72	70.82	35.64	60.39	**4.56**	132.05	6.89	12.09
Watari Iwanuma	1794.76	2311.85	1803.23	679.09	1370.60	3953.33	52.00	95.30	**17.30**	118.67	26.14	66.15
Higashima- tsushima	1649.70	3229.26	2627.52	432.98	436.58	662.58	42.26	62.09	**5.97**	–	11.94	35.90

best performance with $\kappa = 0.8967$. We believe, that the performance in the Canada dataset could improve by using more regions (i.e. more bins). Additionally, the NL-BM-GSF presents the lowest OE in 4 out of the 6 datasets (see Table 30). In general, the proposed method in MS homogeneous cases outperforms the comparison models with the exception of two datasets where the results of our proposed approach still achieve a competitive performance.

For the second scenario of SAR homogeneous cases presented in Figures 63 to 66 (see the supplementary material). The NL-BM-GSF method achieved the best overall performance in two out of the four datasets, also achieved the second best in Wenchuan dataset, and the four best in San Francisco dataset where other methods outperforms all the comparison approaches (see Table 29). The highest value was achieved in the Atlantico dataset with a $\kappa = 0.7411$ and the lowest in Wenchuan dataset with a $\kappa = 0.5151$. It is worth to remark that one of the most challenging datasets presented in the three scenarios is the Wenchuan dataset. In which, ten of the comparison methods achieved a negative κ value (meaning a poorly performance of the method which is even worse than the one obtained by random selection) and only the G-SMO and our proposal NL-BM-GSF achieved a positive value of $\kappa = 0.7758$ and $\kappa = 0.5151$ respectively. Furthermore, the NL-BM-GSF obtained the lowest OE 4.48 and 5.89 (see Table 30) for Katios and the Atlantico datasets respectively. The NL-BM-GSF performance for SAR homogeneous cases achieved a mean value in the κ metric of 0.6352 meanwhile the G-SMO (second best) method in this settings achieve a value of 0.6032. Showing, that even without outperforming in all datasets the proposed approach NL-BM-GSF has a better performance than the comparison methods.

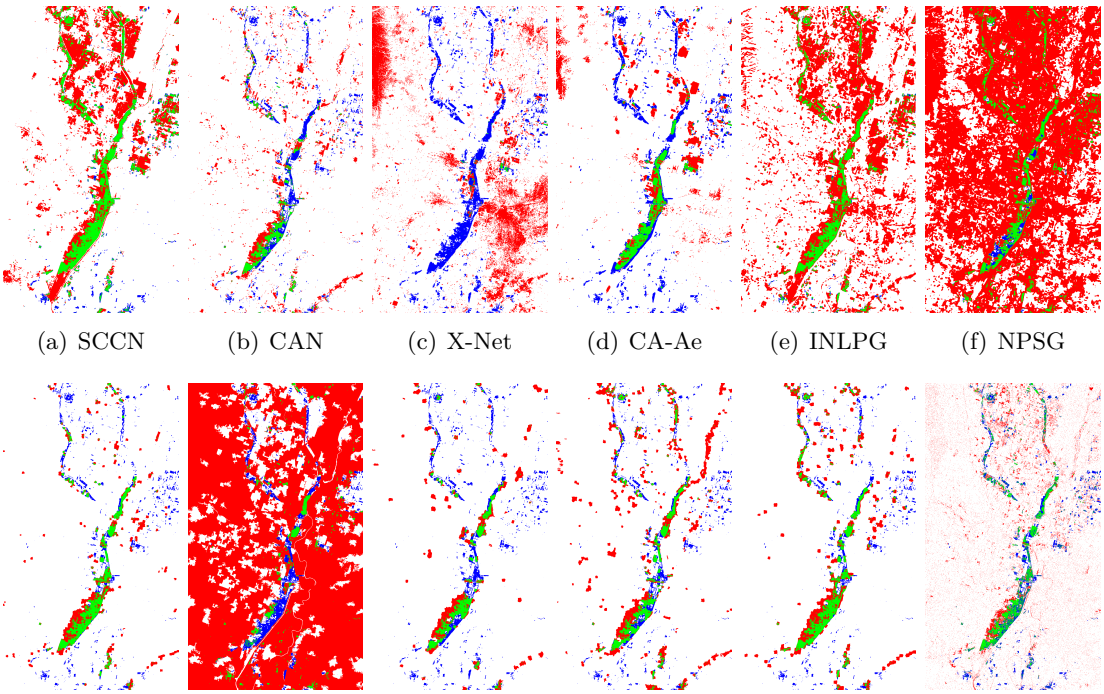
Lastly for the third scenario of the heterogeneous cases and the largest one with ten cases, the NL-BM-GSF outperformed all the competing methods in 4 out of the 10 datasets and achieved the



(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 25: Change maps for Toulouse.



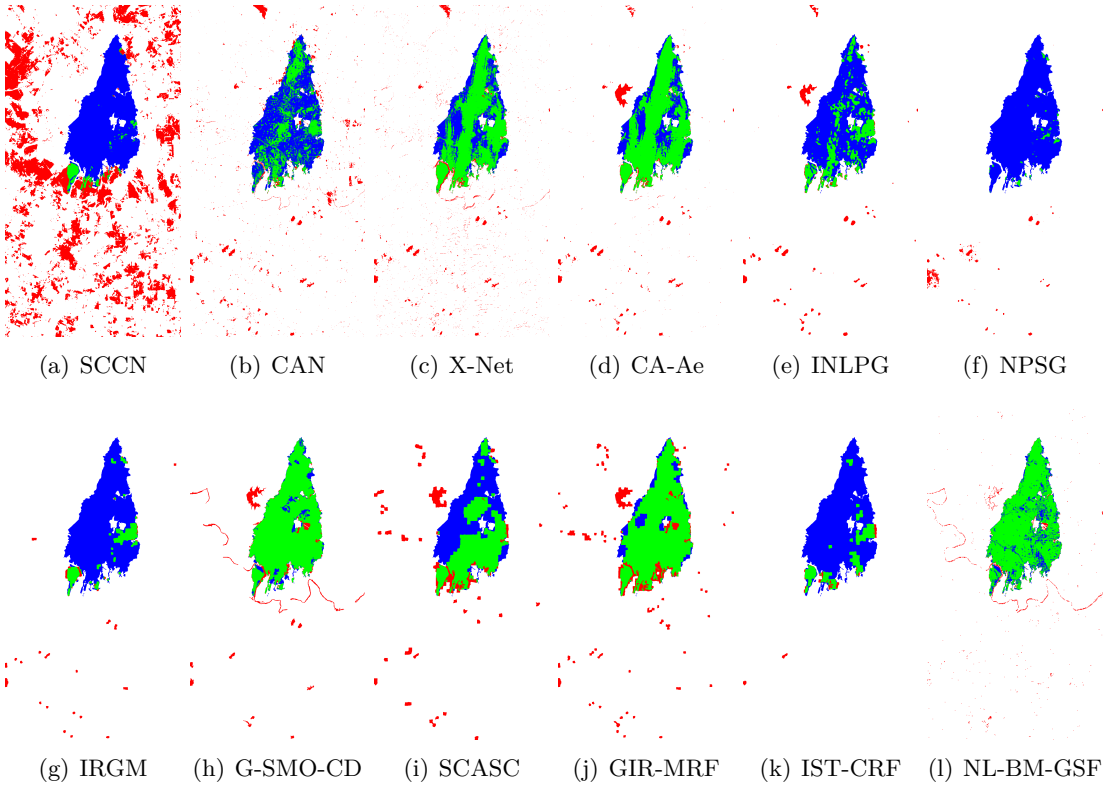
(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 26: Change maps for California.

second best in two (i.e. Bastrop and Kaliveli Lake) of the remaining six datasets (i.e. Toulouse, California, Shuguang, and Town Island). The NL-BM-GSF achieved the highest κ of all the comparison methods in the Palma dataset with a value of $\kappa = 0.6365$ where the second best is SCCN with $\kappa = 0.5549$. Here, the NL-BM-GSF obtained the lowest OE in Gloucester-2, Palma, Kaliveli Lake and Watari Iwanuma datasets (see Table 30). Additionally, for the challenging datasets such as Kaliveli Lake and Watari Iwanuma where most of the comparison methods score negative values in terms of the κ metric, the proposed NL-BM-GSF achieved the second best $\kappa = 0.3813$ and the best $\kappa = 0.6092$ performance respectively.

With respect to the execution times, presented in Table 31, the graph-based methods from INLPG



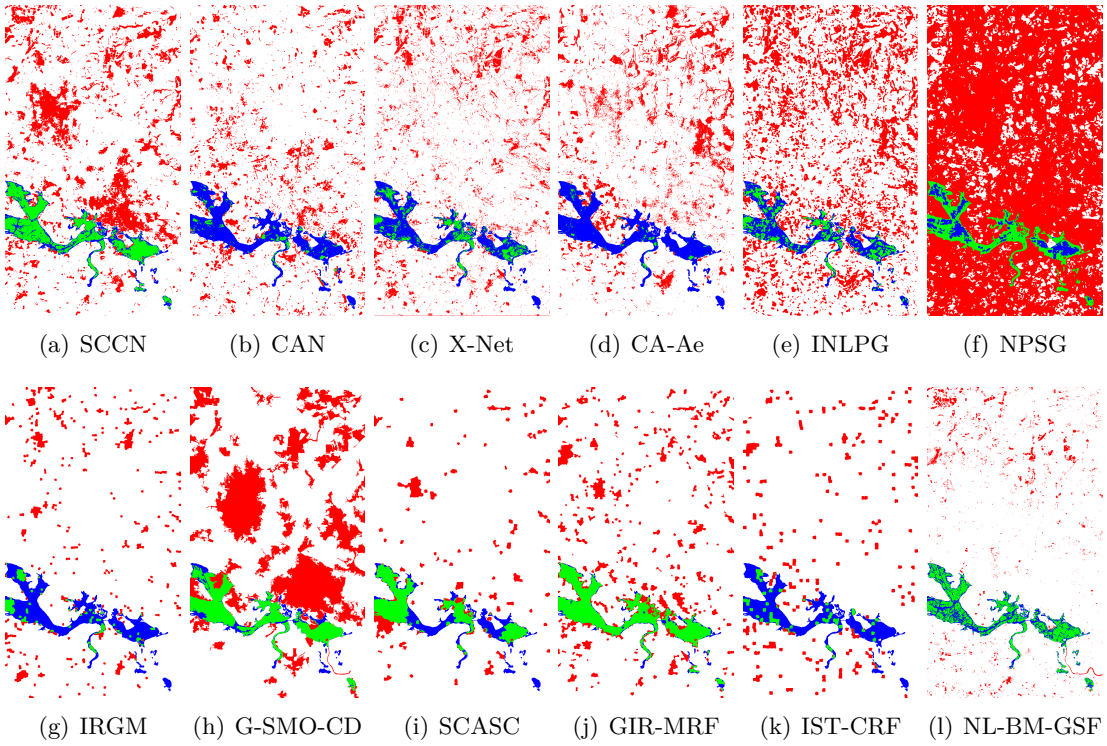
(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 27: Change maps for Bastrop.

to our proposal NL-BM-GSF are the fastest ones in general. However, the best overall results are achieved by the proposed method NL-BM-GSF, that's execution times for large datasets (i.e. the homogeneous cases) does not take more than approximately fifteen seconds to compute. Meanwhile, the competing methods take around twenty seconds, to achieve results that in general are not better than ours (see Tables 29 and 31).

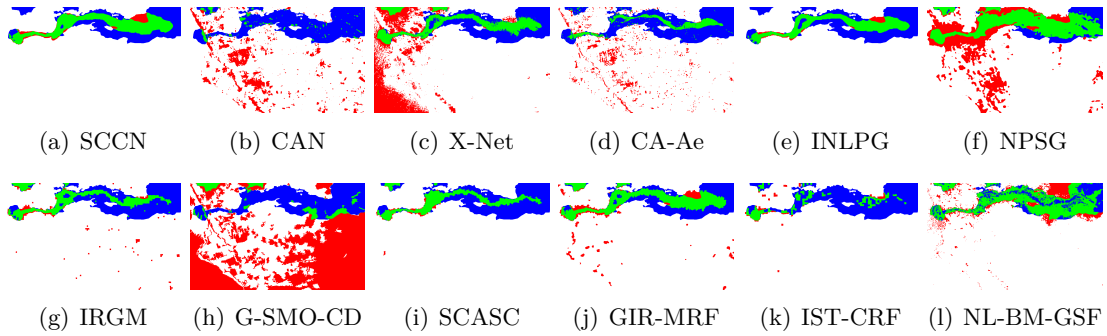
In addition, Figures 56 (a) shows that the **NL-BM-GSF** method is significant different from 9 out of the 11 comparison methods according to the two-way ANOVA test, and in 56 (b) is denoted that in general the mean performance of the proposed model is higher than the others and show a low deviation compare to the other approaches in terms of the κ metric. On the one hand, the non-local graph subspace approach (**NL-BM-GSF**) outperformed the competing models (**SCCN** [Liu et al., 2016], **CAN** [Niu et al., 2018], **X-Net** [Luppino et al., 2021], **CA-Ae** [Luppino et al., 2022], **INLPG** [Sun et al., 2021d], **NPSG** [Sun et al., 2021c], **IRGM** [Sun et al., 2021a], **G-SMO** [Jimenez-Sierra et al., 2022], **SCASC** [Sun et al., 2021b], **GIR-MRF** [Sun et al., 2022b], and **IST-CRF** [Sun et al., 2022a]) in terms of Cohen's Kappa coefficient ($\boldsymbol{\kappa}$) and overall error (OE) in 10 out of the 20 datasets, and achieved the second best in 4 of the remaining 10 datasets. It is important to point out, that the **SCCN**, **IRGM**, **G-SMO**, **GIR-MRF**, and **IST-CRF** approaches



(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 28: Change maps for Gloucester-2.

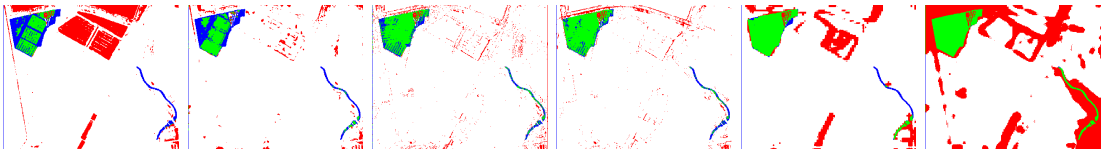


(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

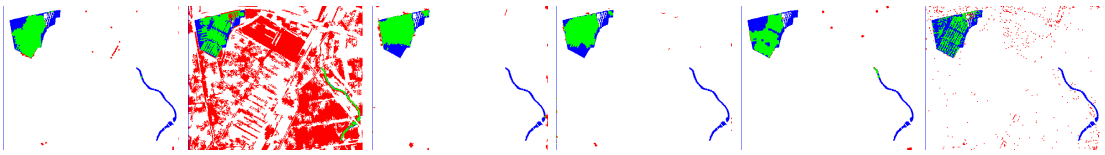
(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 29: Change maps for Palma.

achieved the best $\boldsymbol{\kappa}$ in the remaining 10 datasets outperming the other methods in 1, 2, 2, 3, and 2 datasets respectively. The deep learning approaches (**CAN**, **X-Net**, and **CA-Ae**), and the graph-based methods (**INLPG**, **NPSG**, and **SCASC**) did not outperform the competing methods in any of the 20 datasets with respect to the κ metric. **SCCN** is the method that showed the poorest performance, which did not surpass in any metric the competing methods for the 14 datasets. In this benchmark, **CAN** came after **SCCN** in terms of performance. Furthermore, we note that even though the graph-translation based models such as **INLPG** [Sun et al., 2021d], **NPSG** [Sun et al., 2021c], **IRGM** [Sun et al., 2021a], **SCASC** [Sun et al., 2021b], **GIR-MRF** [Sun et al., 2022b], and **IST-CRF** [Sun et al., 2022a] focus into finding the DI trough the mapping from one graph to

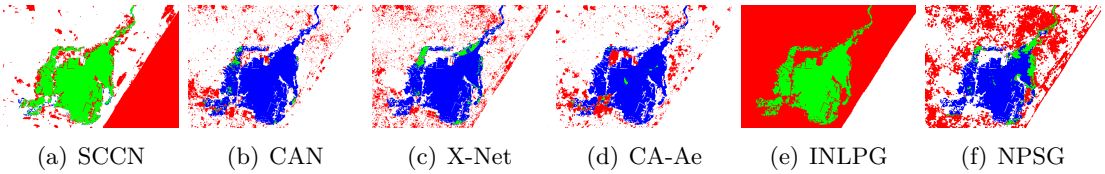


(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

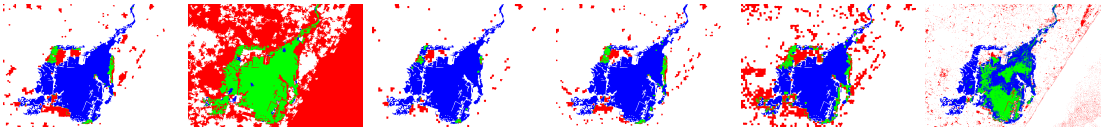


(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 30: Change maps for Shuguang.

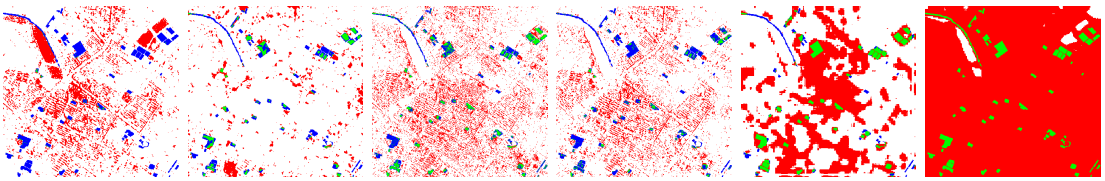


(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

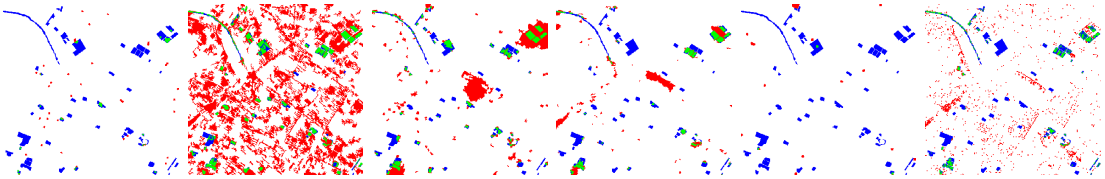


(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 31: Change maps for Kaliveli lake.



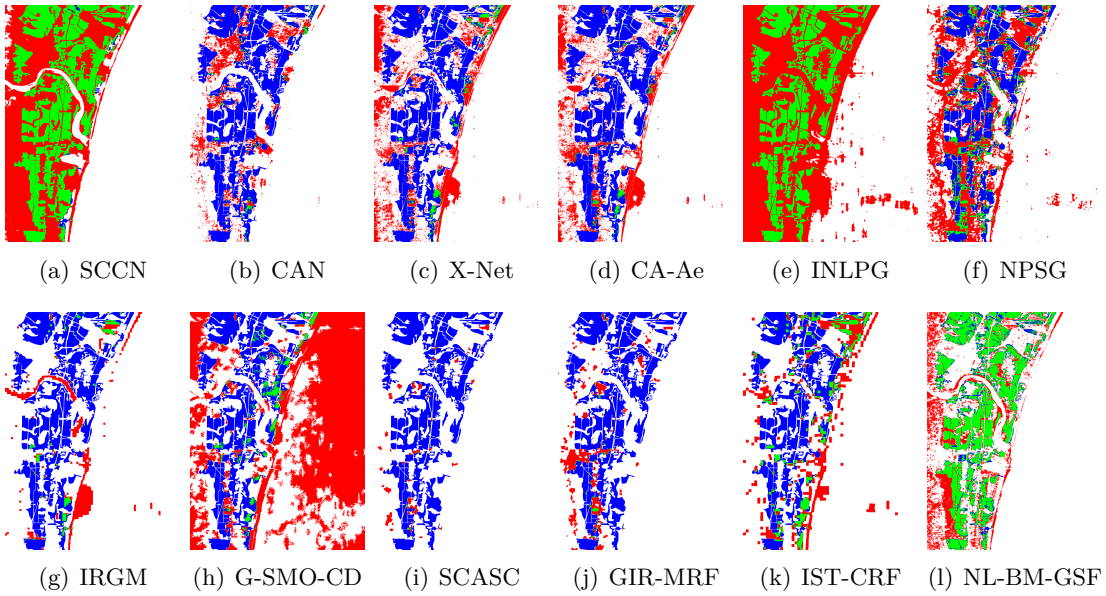
(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG



(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 32: Change maps for Town Island.

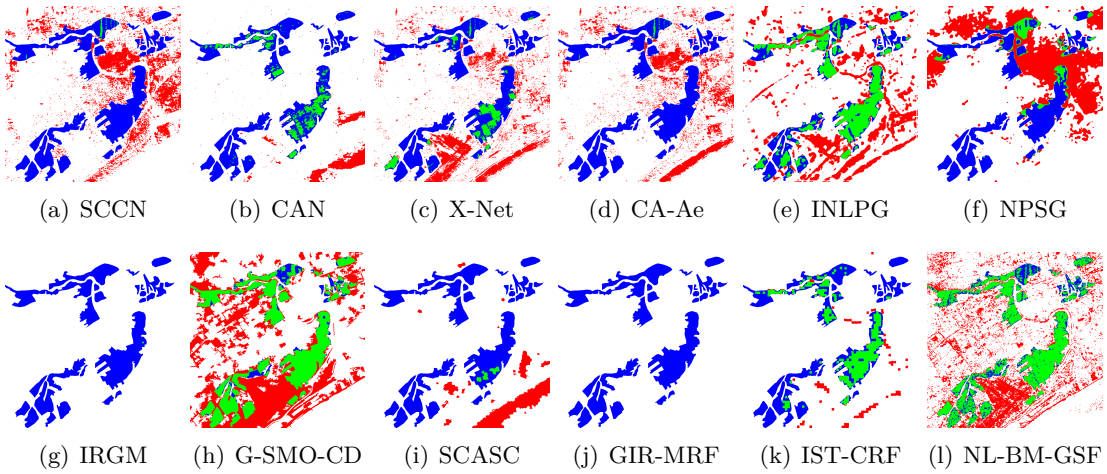
another and vice versa, these models are unable to generalize well over the entire datasets tested in this work. Unlike these approaches, our method produces notably better results even though it is limited by the super-pixel segmentation since those are the nodes of the graphs. Nonetheless,



(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 33: Change maps for Watari Iwanuma.



(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 34: Change maps for Higashimatsushima.

the proposed NL-BM as shown in Figure 24 preserves the correlation of the structures and add non-locality to the model. Similar to our approach, **G-SMO** also relies on a prior knowledge (i.e. difference image), which is used in the optimization problem to obtain the final change map, but in our proposal we use that the change maps are sparse as a prior knowledge in a implicit way in the l_0 norm. Meanwhile, in the approach **G-SMO** the prior is used explicitly as the difference image.

Finally, in Table 32 we present the performance of our proposed method with respect to the ablation approach. In the ablation process, we remove the proposed fusion rule and change map estimation. Instead, we use the approach explained in Section 4.1.2, which separates the fusion and

Table 32: Ablation test for NL-BM with the fusion rule and the optimization problem proposed in [Jimenez-Sierra et al., 2022] (NL-BM-Prev), in terms of $\boldsymbol{\kappa}$.

Homogeneous Cases			Heterogeneous Cases		
Dataset	**NL-BM-Prev**	**NL-BM-GSF**	**Dataset**	**NL-BM-Prev**	**NL-BM-GSF**
Mulargia	0.7610	**0.8921**	Toulouse-1	0.0558	**0.1693**
Omodeo	0.6758	**0.8286**	California	0.1786	**0.3136**
Alaska	0.4539	**0.8164**	Bastrop	0.8739	**0.8788**
Machadinho d'Oeste	0.8495	**0.8548**	Gloucester-2	0.1242	**0.5518**
Prince George	−0.3562	**0.5330**	Palma	0.0655	**0.6365**
Gloucester-1	−0.0175	**0.6543**	Shuguang	0.1445	**0.4408**
Katios	0.6205	**0.7208**	Kaliveli Lake	0.1217	**0.3813**
Atlantico	0.7410	**0.7411**	Town Island	0.0829	**0.2150**
San Franciso	0.5370	**0.5638**	Watari Iwanuma	0.0382	**0.6092**
Wenchuan	0.4407	**0.5151**	Higashima- tsushima	0.2337	**0.4235**

the change map estimation. The results shown in Table 32, confirm that the new fusion rule based on subspace fusion and the integration with the change map estimation in the same optimization problem improve the results for the proposed graph-based approach.

These results, demonstrate that our proposal is useful for creating generic models of signal structure that enhance change detection even for variety of heterogeneous datasets.

5.2 Graph Subspace Fusion and TV for Biomass estimation

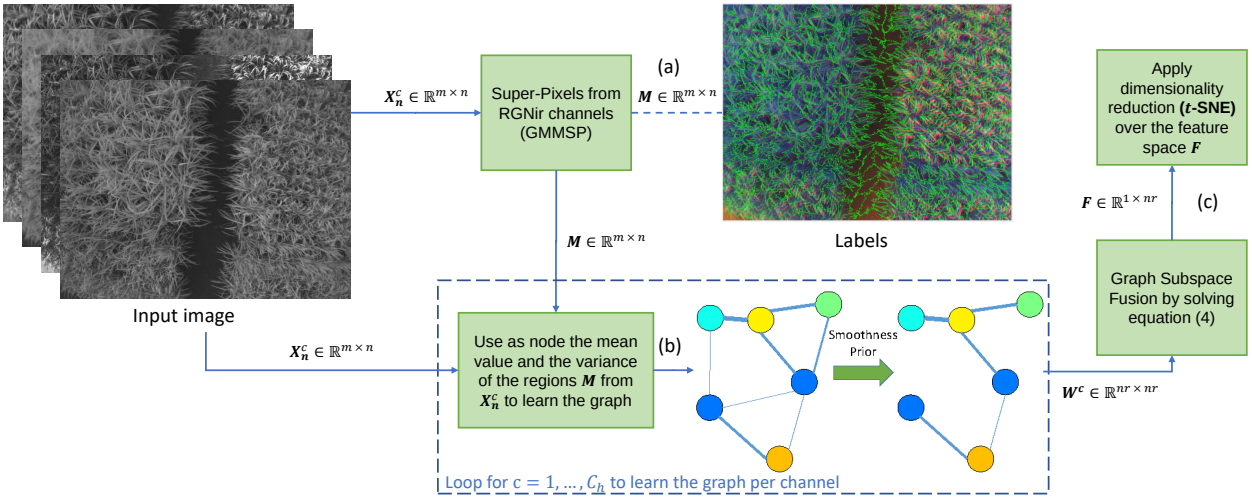


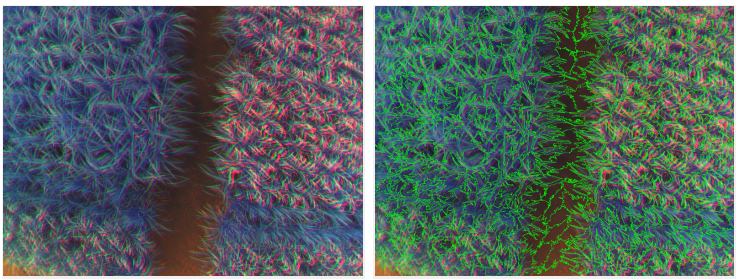
Figure 35: Proposed framework for graph subspace fusion. Our proposal has three main stages: (a) The number nr of super-pixels $\mathbf{M}$ generated by GMMSP from the R-G-Nir channels c of the n-th input image $\mathbf{X_n^c}$, (b) learning a graph per channel related to the super-pixels W^c, and (c) the subspace fusion F of the graphs. Where, n indicates the current image in the dataset, c is the number of channels, and F are the features from the subspace fusion.

Inspire by the previous described approach in Section 5.1, we proposed a framework to perform fusion but with total variation regularization to smooth the extracted features. In short, first we perform super-pixel segmentation by using GMMSP (see Section 4.2.1). Then, as shown in Figure

35 we apply the proposed graph subspace fusion explained below to extract graph-based features and train a machine learning algorithm for biomass estimation.

5.2.1 Nodes as Super-Pixels

The learning of a graph from the pixels in an image demands a high computational memory capacity (i.e. RAM) [Jimenez-Sierra et al., 2020, Jimenez-Sierra et al., 2021b]. Therefore, we use the Gaussian Mixture Model for Super-Pixels (GMMSP) explained in Section 4.2.1.
Since we need to create a graph for each image to later perform the fusion step, GMMSP algorithm should have the same regions when applied to the images. Therefore, we use a false RGB image, where the channels are Red, Green, and NIR that in previous works have shown to be relevant for crop phenotyping. Next, GMMSP segments the false RGB image into nr regions, and outputs the label matrix M of size $m \times n$ (same size as the input images). M represents the label for each pixel related to its corresponding region. Finally, we define the nodes as 2-D points for the graph learning stage (explained in Section 4.1.2). The node values (i.e. descriptors of the super-pixels) are defined by the mean value and variance of the related super-pixel to the node. Figure 36 shows the results of applying GMMSP on one image in the dataset. In previous works based on



(a) RGNir image (b) Super-Pixels

Figure 36: GMMSP applied to an image transform into the false RGB image (a). The generated super-pixels (b) contains around 500 regions.

graphs to extract feature from crops for above-ground biomass dynamics (ABGD) [Jimenez-Sierra et al., 2020, Colorado et al., 2020, Jimenez-Sierra et al., 2021b], the authors focused in the crop segmentation or in selecting the most relevant spectral basis given by the graph fusion. However, the fusion step relies in the combination of the graph weights functions given by the minimal value (i.e. denoting a low similarity).
In [Rasti and Ghamisi, 2020], the authors fuse a hyper-spectral image with a multi-spectral image by finding a common subspace. Motivated by this fusion principle, we proposed to fuse the subspaces generated by the Laplacian of the graphs that in our application are related to the channels of an

image $\mathbf{X_n^c} \in \mathbb{R}^{\mathbf{m}\times\mathbf{n}}$ as follows:

$$\begin{aligned} F &= \arg\min_F \tfrac{\lambda_1}{2}\left\|W_1 - FV_1^\top\right\|_{\mathcal{F}}^2 + \tfrac{\lambda_2}{2}\left\|W_2 - FV_2^\top\right\|_{\mathcal{F}}^2 + \\ &\tfrac{\lambda_3}{2}\left\|W_3 - FV_3^\top\right\|_{\mathcal{F}}^2 + \tfrac{\lambda_4}{2}\left\|W_4 - FV_4^\top\right\|_{\mathcal{F}}^2 + \lambda_5\left\|F\right\|_{\mathrm{TV}}, \end{aligned} \tag{46}$$

where $L_1, \ldots, L_4$ are the Laplacian matrices related to the channels Red, Green, Nir, and Red-Edge respectively, $V_1, \ldots, V_4$ are the basis generated by the subspace of the graphs (i.e. weighted matrix), $\|.\|_{\mathcal{F}}$ represents de Frobenius norm, and $\|F\|_{\mathrm{TV}} = \left\|\sqrt{(D_hF)^2 + (D_vF)^2F}\right\|$ is the total variation norm. In order to solve the problem in (46), we apply variable splitting with $F = S$ and add this constrain as a penalty term. Then, the minimization problem to solve is:

$$\begin{aligned} F &= \arg\min_{F,S} \tfrac{\lambda_1}{2}\left\|W_1 - FV_1^\top\right\|_{\mathcal{F}}^2 + \tfrac{\lambda_2}{2}\left\|W_2 - FV_2^\top\right\|_{\mathcal{F}}^2 \\ &+\tfrac{\lambda_3}{2}\left\|W_3 - FV_3^\top\right\|_{\mathcal{F}}^2 + \tfrac{\lambda_4}{2}\left\|W_4 - FV_4^\top\right\|_{\mathcal{F}}^2 \\ &+\lambda_5\left\|S\right\|_{\mathrm{TV}} + \tfrac{\mu}{2}\left\|S - F - L_m\right\|_{\mathcal{F}}^2, \end{aligned} \tag{47}$$

where L_m is a Lagrangian multiplier. Finally, the optimization problem in (47) is solved cyclic descent algorithm (i.e. compute the derivatives with respect to the minimization arguments) as follows:

$$F = \frac{\lambda_1 W_1 V_1 + \cdots + \lambda_4 W_4 V_4 + \mu(S - L_m)}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \mu}, \tag{48}$$

to solve with respect to S we use the implementation of the method SplitBregman [Goldstein and Osher, 2009] as in [Rasti and Ghamisi, 2020]:

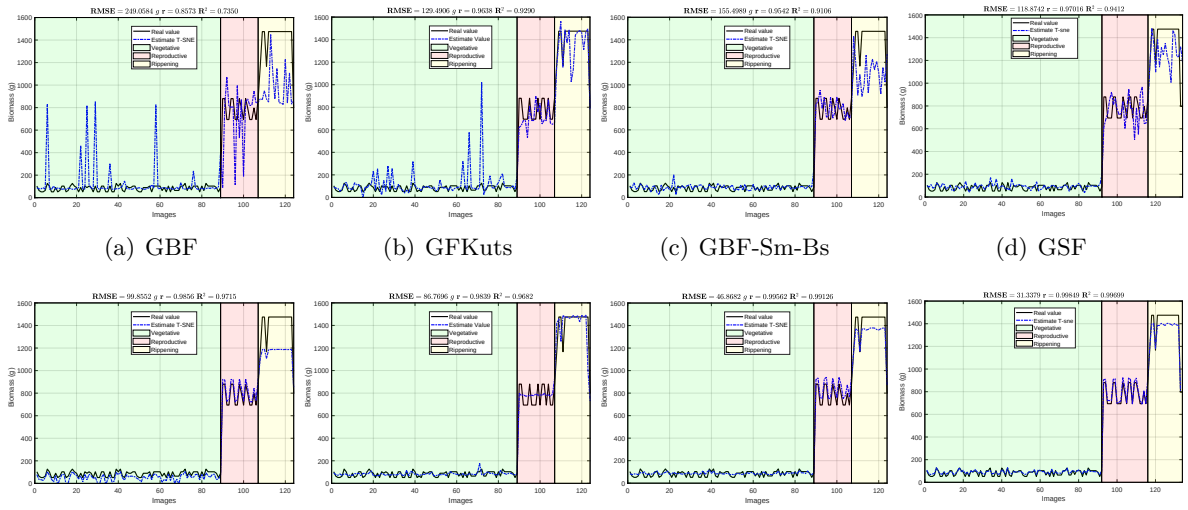
$$S = \mathrm{SplitBregman}(L_m + F, \mu/\lambda_5). \tag{49}$$

Lastly, we update the Lagrangian multiplier $L_m = L_m + F - S$.

The proposed framework is summarized in figure 35. Where, we have three main stages: (a) The super-pixels generation $M \in \mathbb{R}^{m\times n}$, (b) the graph learning W^c from the super-pixels, and (c) the graph subspace fusion F.

5.2.2 Experimental results and discussion

Experimental set-up: In order to compare the effectiveness of the features extracted from the proposed graph subsapce fusion, the results were compared with three approaches: (i) the former graph method introduced in [Jimenez-Sierra et al., 2020], namely GBF, (ii) the GFKuts approach [Colorado et al., 2020], and (ii) the GBF-Sm-Bs in [Jimenez-Sierra et al., 2021b]. We used the datasets and Ground-Truth reported in [Jimenez-Sierra et al., 2020, Colorado et al., 2020], containing 314 images for the vegetation stage, 82 for the reproductive, and 71 for the ripening. The captured images have a resolution of 960 × 1280 pixels, geo-referenced with the corresponding



(a) GBF (b) GFKuts (c) GBF-Sm-Bs (d) GSF

(e) GBF (f) GFKuts (g) GBF-Sm-Bs (h) GSF

Figure 37: Biomass estimation curve for rice crop related to the whole phenological cycle (i.e. Vegetation in green, reproductive in red, and ripening in yellow). The columns shows the results by method: GBF, GFKuts, GBF-Sm-Bs, and GSF. The rows indicate the regressor used: first row is related to the SVM and second row to the NARX regressors respectively.

biomass measurements in grams (g) from the Ground-Truth, which was assembled as follows: 1 linear meter of the plants were cut from each plot of the crop and weighted to obtain the fresh biomass. Subsequently, the samples are placed inside an oven at 65 degrees Celsius for 4 days or until a constant weight is reached. This is known as the dry biomass.

Having the imagery dataset and the Ground-Truth, we trained two estimation models: classical SVM-R regression and a robust nonlinear autoregressive network with exogenous inputs (NARX) [Boussaada et al., 2018], both accounting for 70% of the dataset for the training, and the remaining 30% for testing and validation. Lastly, the performance of the models was measured in terms of the root mean squared error ($RMSE$), the linear correlation (r), and the coefficient of determination (R^2). Furthermore, we use the RBF Kernel for the SVM regressor since in previous works it has better performance than the linear or polynomial kernels [Jimenez-Sierra et al., 2020, Jimenez-Sierra et al., 2021b].

The algorithms were tested in a server with two Intel(R) Xeon(R) CPUs E5-2650 v4 @ 2.20 GHz, with 24 physical cores, 48 threads of processes, and 252 GB of RAM. We use the parameters recommended by each author for the GBF [Jimenez-Sierra et al., 2020], GFKuts [Colorado et al., 2020], and GBF-Sm-Bs [Jimenez-Sierra et al., 2021b]. For our proposal, we fixed the parameters $v_x = v_y = 35$ for the super-pixels, the contribution of each weight matrix and the total variation norm of the optimization problem in (47) were set to be equal (i.e. $\lambda_1 = \cdots = \lambda_5 = 1$), the value of K for the graph learning was fix to $\frac{nr}{2}$, and the value related to the convergence of the optimization

problem was set $\mu = 0.01$ according to [Rasti and Ghamisi, 2020]. In addition, we performed a grid search over the number of dimensions to reduce the feature space and the results suggest to use 10 dimensions. For the cyclic descent algorithm we use 30 iteration, S, and L_m are initialized as matrices with ones.

Results and Discussion: Figure 37 shows in the columns the predicted biomass (i.e. qualitative results) obtained from the four methods in the following order: GBF [Jimenez-Sierra et al., 2020], GFKuts [Colorado et al., 2020], GBF-Sm-Bs [Jimenez-Sierra et al., 2021b], and our proposal GSF. The performance of the models is measured by the accuracy in the biomass prediction according to the metrics presented in table 33. According to Table 33, the proposed approach GSF achieved

Table 33: Performance in terms of $RMSE$, r, and R^2 for the models: GFKuts [Colorado et al., 2020], GBF [Jimenez-Sierra et al., 2020], GBF-Sm-Bs [Jimenez-Sierra et al., 2021b], and ours GSF.

Model/Regressor	RMSE (in grams [g])	r	R^2
GFKuts [Colorado et al., 2020]/SVM	129.490	0.963	0.929
GBF [Jimenez-Sierra et al., 2020]/SVM	249.058	0.857	0.735
GBF-Sm-Bs [Jimenez-Sierra et al., 2021b]/SVM	155.498	0.954	0.910
GSF/SVM	**118.8742**	**0.970**	**0.9412**
GFKuts [Colorado et al., 2020]/Narx	86.769	0.983	0.968
GBF [Jimenez-Sierra et al., 2020]/Narx	99.855	0.9856	0.971
GBF-Sm-Bs [Jimenez-Sierra et al., 2021b]/Narx	45.358	0.995	0.991
GSF/Narx	**31.337**	**0.998**	**0.996**

the lowest $RMSE = 118.8742$ g and $RMSE = 31.337$ g fpr the SVM and NARX regressors respectively. Furthermore, the Narx regressor improves the fitting to a linear model (i.e. see r and R^2 coefficients), in which the proposed method achieved a linear correlation value of $r = 0.998$. Contrasting with the previous graph-based method [Jimenez-Sierra et al., 2020] (GBF) which uses an approximated graph based on the Nyström extension, the results were improved by decreasing the $RMSE$ metric by 130.1838 g and 68.5180 g for the SVM and Narx regressors respectively. Regarding the GFKuts method that relies on VIs as features over a segmentation generated by graphs, our proposal obtains a slight improvement for the SVM regressor by decreasing the $RMSE$ value by 10.6158 g, however with the Narx regressor the improvement regarding to the proposed GSF was more notorious by a decrease of the $RMSE$ value in 55.4320 g. Finally, with respect to the most recent graph-based method GBF-Sm-Bs which improves the samples used to approximate the graph in the Nyström extension, the GSF decreases the $RMSE$ value in 36.6238 g and 14.0210 g for SVM and Narx regressors respectively.

Overall, the results presented in Table 33 demonstrates the effectiveness of the extracted features related to the estimation of biomass in rice crops by using the proposed subspace fusion of the graphs based on the generated super-pixels.

6 Conclusions

We introduced a graph-based data fusion methodology for remote sensing images and tested it in two applications: change detection and biomass estimation in rice crops. The main contribution of this thesis is a "data-driven" framework used to capture unique information for multi-temporal, and multi-modal/heterogeneous (Toulouse, California, Bastrop, and Gloucester-2 datasets) images in a fused graph. The fused graph stage captures information in one graph from a small set of samples (i.e. ranges between 50 and 150 samples that is less than 10% of the total pixels within the input image) for each dataset (in different times or bands for homogeneous or heterogeneous data). For the change detection application, we utilized a mutual information criterion to select from a prior and an eigen-image to build the final change map. In this case, our method is parametric since it depends on a number of samples and the prior information (difference images). Thereby, from the results for all datasets, we observed that our model obtained coherent change maps and outperformed state-of-the-art methods [Kittler and Illingworth, 1986, Zanetti et al., 2015, Zanetti and Bruzzone, 2017]. The method proposed in this study performed well with respect to the metrics TP and FN in multi-sensor datasets such as: Toulouse, California, Bastrop, and Gloucester-2. In addition, the model developed in this paper does not require a post-processing stage, such as that needed by the U-CD-HPT method. In biomass estimation, the model showed that the features extracted from the fused graph with a dimensionality reduction technique (i.e., PCA or t-SNE) capture the variability of biomass in rice crops. This model makes possible to predict the biomass features throughout the growth stages in rice crops, by using one regression model. These outcomes are more comprehensive than those reported by the authors in [Devia et al., 2019], in which three separate regression models estimated the biomass at each stage of the rice crop, based on VI features.

Furthermore, we proposed a method for above-ground biomass estimation, that not only enabled the precise characterization of the biomass behavior of the rice crops through the entire phenological cycle, but also worked as both image segmentation and feature extraction techniques, by associating relevant features from the canopy. It is worth mentioning that most of the existing body of work in remote sensing methods for high-throughput biomass estimations based on multispectral imagery, requires dedicated photogrammetry methods for image correction, segmentation and feature extraction. In this study, the GBF-Sm-Bs methods solved those stages by combining them into one single approach: First we apply GFKuts over the input image, secondly we use a down-sampled version of the image by applying a square grid over the GFKuts output, then the graph is learnt by using the prior of smoothness (i.e. where the nodes are the vectorized version

of each square in the image), next we apply the blue-noise sampling over the smoothed graph, and finally inject these samples to the graph based fusion algorithm proposed in [Jimenez-Sierra et al., 2020]. According to the findings reported in Table 21, the proposed GBF-Sm-Bs approach obtained an average biomass correlation of 0.995 with $R^2 = 0.991$ and $RMSE = 45.358$ g, increasing the precision in the estimation by around 45.42% (45.358/99.855), compared to the GBF method, and about 52.27% (45.358/86.769) compared to the GFKuts approach for the NARX regressor. This is a promising result towards GWAS gene characterization (Genome-wide Association Study), that requires larger amounts of precise and accurate phenotyping data for the association of gene functions with a specific trait. In this regard, future work is aimed at the inclusion of clustering approaches within the proposed algorithms, to enable the extraction of features according to several plant varieties (genotypes).

Moreover, we have proposed a new graph-based change detection (CD) framework from superpixels, which builds on state-of-the-art graph learning methods (GL) based on signal smoothness priors. Unlike some early works on graph-based change detection, our method does not rely on eigendecomposition for change detection. Since the GL problem may demand high computational complexity, a downsampling module based on GMMSP was leveraged to alleviate such requirements. The proposed optimization based approach reduces the execution time necessary to build the change map. As shown by our results, the optimization based approach helps to cope with noisy prior change maps, preventing errors in the final change map detection. Particularly, when the prior change map is very noisy, the value of the regularization parameter α needs to be set small enough, so that we mostly rely on the fused graph for change detection, leading to accurate change map estimates, which are highly smooth on the graph. Here an important question to ask is: under what conditions on input images we can assume that the original change map is sufficiently smooth on the fused graph?. This is however a more fundamental issue that is currently under investigation. Our experimental results show that the smoothness assumption on the fused graph leads to remarkable results in change detection. We showed that our approach outperforms existing state-of-the-art methods such as probabilistic thresholding [Zanetti and Bruzzone, 2018], machine learning [Luppino et al., 2019], deep learning approaches [Niu et al., 2018,Luppino et al., 2021], and other graph-based frameworks [Jimenez-Sierra et al., 2020, Sun et al., 2021d, Sun et al., 2021c, Sun et al., 2021a] in 10 out of 14 datasets. Lastly, our findings prove the effectiveness of incorporating GSP models for graph learning in the context of change detection from remote sensing images. In addition, the proposed optimization approach improved change detection performance over different cases (i.e. MS homogeneous, SAR homogeneous, and MS/SAR heterogeneous cases).

Finally, we introduced a novel way to represent the graphs by inducing non-locality by using the proposed NL-BM. The experimental results, shows that the NL-BM is able to correlate areas in the images that are not close in location but that indeed share or define the same structure (see Figure 24). Furthermore, we have proposed a different way of fusing the graphs by addressing the problem as subspace fusion, where the fused features are co-living in the subspaces of each graphs. Here, we presented two version of the subspace fusion problem. On the one hand, the first optimization problem (see Equation (34)) takes into account the change map and uses the l_0 norm to regularized the solution since we know beforehand that the change map is sparse. The performance of the NL-BM alongside with the graph-subspace fusion in twenty datasets shows that by improving the fusion and the node representation we can have a general model applicable in diversity of cases (i.e. with different settings, complexity of the images, coping with different noises, low resolution, high resolution, etc) that performs better in 10 out of the 20 datasets against recent state-of-the-art approaches. Moreover, from the results we can tell that our proposal is competitive across all the datasets and promising towards a general model for change detection in homogeneous and heterogeneous cases. On the other hand, we use a graph-subspace fusion with total variation regularization to smooth the fused features and used them as descriptors of rice crops in all the growth stage. The results tabulated in Table 33 reflect that the fusion rule proposed is able to outperform the previously proposed approaches considerably.

Overall the approaches and results presented in this thesis represent a important progress in the state-of-the-art of graph-based data fusion methods for remote sensing applications such as change detection and above ground biomass estimation.

7 Future work

In this section, based on the findings reported, we propose the following directions for further research:

7.1 Integration of graph-fusion and learning stages

Recently, it has been proposed graph-translation approaches that address the change detection problem similar to the image-translation problem. However, instead of learning how to map one image to another, the graph-translation learns how to map from one structure to another (i.e., map from one graph to another). For instance, **INLPG** [Sun et al., 2021d], **NPSG** [Sun et al., 2021c], and **IRGM** [Sun et al., 2021a] consider overlapped patches as nodes over the images [Sun et al., 2021d, Sun et al., 2021c] or super-pixels [Sun et al., 2021a] and learn the graph using a K-nearest neighbors [Sun et al., 2021d, Sun et al., 2021a] or a Gaussian kernel [Sun et al., 2021c]. Then, a function f_i is defined to map from G_1 to G_2 and vice versa to seek changes in the learned graphs that capture the structure for pre- and post-events. Once the mapping functions have computed the approaches [Sun et al., 2021d, Sun et al., 2021c, Sun et al., 2021a] apply the forward and backward difference image (DI) over the mapped graphs and the final change map is detected by using a threshold algorithm (i.e. Otsu's [Otsu, 1979], PCA with k-means [Celik, 2009], or Markov Random field [Szeliski et al., 2008]). The differences among these approaches rely on: (i) The definition of nodes with respect to the spatial relationships in the images, (ii) the graph learning approaches, and (iii) the functions f_i that performs the graph-translation.

More recently, the approaches **SCASC** [Sun et al., 2021b] and **GIR-MRF** [Sun et al., 2022b] use the graph-translation problem to solve the image-translation problem. In short, the authors learn a graph by using features (i.e., mean, median, and variance) extracted from super-pixels, perform the graph-translation to detect a change feature matrix to map the pre-event image in to the post-event image, and then combine the change information with spatial data to detect the final change map. For instance, in the method **SCASC** [Sun et al., 2021b], an adaptive probabilistic graph is learned to perform the graph-translation as an optimization problem to find a regression feature R. With R, the regression image is computed and apply a super-pixel label problem based on MRF to detect the final change map. A similar procedure is followed by the approach **GIR-MRF** [Sun et al., 2022b]. However, here the authors learn the graph by combining local and global structural information. Moreover, the learned graph is a hypergraph that takes into account structure consistency imposed by a regularization based on the hypergraph Laplacian, and the final

change map involves the use of graph-cut problem.

In general current state-of-the-art related to graphs in remote sensing and our proposal in this thesis carries out the fusion and the learning as separated stages as described before (i.e. the graph is learned in Section 3 by using a Gaussian Kernel, in Sections 4 and 5 by applying smoothness prior). However, it would be very interesting to be able to learn a graph that involves the fusion. For instances, Kalofolias *et al.* [Kalofolias and Perraudin, 2019] find W by solving the following optimization problem:

$$\begin{aligned} \min_{W\in\mathbb{R}^{d\times d}} \; & \operatorname{tr}(W^T Z) - \alpha \sum_i \log(\sum_j W_{ij}) \\ & + \frac{\beta}{2}\|W\|_{\mathcal{F}}^2 + \frac{c}{2}\|W - W_0\|_{\mathcal{F}}^2 \\ \text{s.t. } & W_{ij} = W_{ji} \geq 0,\ i \neq j,\ W_{ij} = 0,\ i = j \\ & W_{i,j} = 0,\ (i,j) \notin E^{\text{allowed}} \subset V \times V \end{aligned} \tag{50}$$

where α is a log prior constant (the larger α, the larger the weights in W), β is a regularity constant that controls the sparisty level of W (the larger β, the denser W), the parameter c encourages adjacency matrices close to a pre-specified adjacency matrix W_0, and the edge set E^{allowed} is a pre-specified set of edges. Although E^{allowed} can be arbitrarily chosen, Kalofolias propose using the edge set of a KNN graph, constructed using approximate nearest neighbors (A-NN) algorithms. This in turn allows us to reduce the time complexity of solving the problem (50) as demonstrated in [Kalofolias and Perraudin, 2019].

Inspire by the the Kalofolias proposal we think that will be possible to adress both the learning and the fusion as follows:

$$\begin{aligned} \min_{W\in\mathbb{R}^{d\times d}} \; & \operatorname{tr}(W_F^T Z_1) + \lambda_1 \operatorname{tr}(W_F^T Z_2) - \alpha \sum_i \log(\sum_j W_{F_{ij}}) \\ & + \frac{\beta}{2}\|W_F\|_{\mathcal{F}}^2 + \frac{c}{2}\|W_F - W_0\|_{\mathcal{F}}^2 \\ \text{s.t. } & W_{ij} = W_{F_{ji}} \geq 0,\ i \neq j,\ W_{F_{ij}} = 0,\ i = j \\ & W_{F_{i,j}} = 0,\ (i,j) \notin E^{\text{allowed}} \subset V \times V \end{aligned} \tag{51}$$

where W_F will be learnt from the distance matrices related Z_1 and Z_2 related to time 1 and time 2, an initial guest for the fuse graph W_0 could be the rule proposed in Section 3.1. Nonetheless, more robust constrains should be considered. For instance, the total variation to smooth the graph, or structure consistency as the one used in [Sun et al., 2021d, Sun et al., 2022b].

7.2 Graph neural networks

Another future direction of this work is to explore the graph convolutional network (GCN), which combines the structural information captured by the graphs with the high level of abstraction given by neural networks. For instances, there are two recent approaches [Gitelson et al., 2002, Chen et al., 2022] that exploits the level of abstraction of GCN to learn local an non-local structure relationship from graphs.

There is a variant of GCN in [Welling and Kipf, 2016] that focuses on the spectral convolution on graphs. In a nutshell, given a signal $s \in \mathbb{R}^N$ that relates a scalar value to the nodes/vertex and a filter $g\theta = diag(\theta)$ with parameters $\theta \in \mathbb{R}^N$. Then, the spectral convolution of s and $g\theta$ is perform by applying the Fourier transform on S and multiplying each frequency by $g\theta$ as:

$$g\theta \star s = U g\theta U^\top s, \tag{52}$$

where U are the eigenvectors of the Laplacian matrix $L = D - A = U\Lambda U^\top$, and Λ is a diagonal matrix of with the eigenvalues of L. Finally, $U^\top s$ is the graph Fourier transform of s and $g\theta$ can be understood as a function of the eigenvalues of L. As demonstrated in the approaches proposed in this thesis, for the change detection application there is a signal s that is spread over the the spectral basis of the fused graph L_F and is able to capture the change map. Therefore, it will be interesting to bring the proposed framework to deep learning field by using GCN. For instance, to use a GCN framework to learn the optimal parameters related to the GSP smoothness prior learning or other graph learning (i.e. based on semi-supervised learning [He et al., 2019], learning graphs from sets [Serviansky et al., 2020], adaptive graph learning [Zhang et al., 2020b], or discriminative learning [Saboksayr et al., 2021]) and the sparsity penalty over signal s to represent properly the change map.

Another direction by using GCN inspired by the works in [Sun et al., 2021a, Sun et al., 2022a] will be to find a graph-translation, but instead of using the vertex information use the GFT to find the spectral basis that is able to translate the graphs and therefore find in that learned base the change map.

7.3 Segmentation and Classification

Given the results of the NL-BM to correlate in a consistent way structures that share common features and are not constrain to be local neighbors, it will be an interesting direction of work a framework that is able to classify given the information of NL-BM and also to detect changes over

the time. For instance inspire in recent works in object detectors [Ladickỳ et al., 2010, Bi et al., 2022], we think that is feasible to train a model that combines the NL-BM information (i.e. about the structure, material, location, etc) and the change detection capture by the fused graph. To bring a result that is able to classify the changes according to the matches find by NL-BM and also count them.

Furthermore, the proposed work in this thesis goes beyond remote sensing applications. For example with the NL-BM will be interesting to check how well segments tissues or cerebral structures in a dMRI and MRI respectively. In addition, to capture descriptors in the evolution of a patient over the time by applying the graph-based framework either to detect changes or to perform follow up given the variation in the structure captured by graphs.

8 Contributions

1. Two open-source databases for change detection[34]: 1-) Fourteen cases for single channel images, and 2-) Twenty cases for multi-channel images with variety of Heterogeneous cases. These datasets are real change detection scenarios captured by MS and SAR sensors presented in [Jimenez-Sierra et al., 2020, Jimenez-Sierra et al., 2021a, Jimenez-Sierra et al., 2022]. In particular, the datasets were acquired by different satellites such as Landsat, ALOS/PALSAR, ERS-2 SAR, Sentinel 1A, and ESA/ASAR. The image resolutions range from 275×400 to 4220×2320. These datasets include events such as earthquakes, floods, wildfires, melted ice, farming, building of a construction site, tsunamis, and volcano eruption. The first datasets incorporate 6 MS single-channel homogeneous cases, 4 SAR homogeneous cases, and 4 single-channel heterogeneous cases which combine SAR/MS images; the second dataset is the augmented version of the previous one that includes multi-channel for the MS images and increases the heterogeneous cases from 4 to 10.

2. A graph-based framework for data fusion applied to change detection and biomass estimation. The main contribution of this study was a "data-driven" framework used to capture unique information from multi-temporal, and multi-modal/heterogeneous images in a fused graph. The fused graph stage captures information in one graph from a small set of samples (less than 10% of the total pixels in the images) for each dataset (in different times or bands for homogeneous or heterogeneous cases). We tested the framework in both change detection and biomass estimation applications. In this experimental set-up, the performance of the proposed method compared with four state-of-the-art models outperformed the competing approaches in eight out of fourteen cases for change detection and compared to a recent state-of-the-art model outperformed the prediction of biomass in rice crops.

3. An adaptation of blue-noise sampling for remote sensing images in the Graph-based framework for data fusion, and the use of smoothness prior in the graph learning. The proposed method for above-ground biomass estimation, enabled the precise characterization of the biomass behavior in rice crops through the entire phenological cycle by extracting relevant feature capture by graphs. We compared the proposed framework with two state-of-the-art methods by using a base line regressor as it is the SVM and a robust regressor base on NN as it is the Narx. By using the SVM our method achieved the second best performance

[34] Available at: https://github.com/DavidJimenezS

with a difference in performance of a 14%, and with the Narx our method achieved the best performance with a difference near to the 50% against the second best method. Hence, the extracted feature from the proposed framework are more relevant and does not highly depend on the regressor for the prediction of biomass in rice crops.

4. A graph-based framework from super-pixels for data fusion that uses GSP smoothness prior in the graph learning phase, and the use of the whole spectral basis for the change map detection. The proposed optimization based approach helps to cope with noisy prior change maps preventing errors, reduces the execution time, and unlike previous methods uses all the spectral basis generated by the graph Laplacian to detect the final change map. We compared the proposed framework against eight state-of-the-art methods such as thresholding, machine learning, deep learning, and other graph-based frameworks. The proposed framework outperformed the comparison methods in 10 out of 14 datasets. In addition, this model is more general since it is able to cope with multi-channel images.

5. A non local block-matching (NL-BM) based approach to generate super-pixels as nodes, along with an optimization problem for graph-subspace fusion applied to CD and biomass estimation. The proposed NL-BM, generates super-pixels iteratively based on the cosine distance from uniformly distributed patches over the image. The super-pixels generated by the NL-BM shows consistency with respect to the different structures and land cover over the images, reducing the amount of nodes needed in previous methods considerably to generate a graph, and therefore the execution time to learn the graph is reduced. The proposed optimization problem, compared to previous state-of-the-art graph-based approaches for change detection does not rely on prior knowledge directly, instead it uses a penalty term over the sparsity of the desired change map, and fuses the data as a subspace fusion generated by the graphs. Furthermore, the variation of the graph-subspace fusion with total variation is able to extract relevant features from rice crops to improve the above ground biomass estimation.

6. The code developed [35] in this thesis is publicly available.

[35] Available at: https://github.com/DavidJimenezS

9 Publications

9.1 Journal Articles

- D. A. Jimenez-Sierra, B. Rasti, H. D. Benítez-Restrepo, and J. Chanussot, "Non-Local Block Matching and Graph Subspace Fusion for Homogeneous/Heterogeneous Change Detection", IEEE Transactions on Geoscience and Remote Sensing, Dic. 2022. (**To be submitted**)
- D. A. Jimenez-Sierra, D. A. Quintero-Olaya, J. C. Alvear-Muñoz, H. D. Benítez-Restrepo, J. F. Florez-Ospina, and J. Chanussot, "Graph Learning Based on Signal Smoothness Representation for Homogeneous and Heterogeneous Change Detection", IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-16, Apr. 2022, doi: 10.1109/TGRS.2022.3168126.
- D. A. Jimenez-Sierra, E. S. Correa, H. D. Benítez-Restrepo, F. C. Calderon, I. F. Mondragon, and J. D. Colorado, "Novel feature-extraction methods for the estimation of above-ground biomass in rice crops", Sensors, vol. 21, no. 13, p. 4369, Jun. 2021, doi: 10.3390/s21134369.
- D. A. Jimenez-Sierra, H. D. Benítez-Restrepo, H. D. Vargas-Cardona, and J. Chanussot, "Graph-based data fusion applied to: Change detection and biomass estimation in Rice crops," Remote Sensing, vol. 12, no. 17, p. 2683, Aug. 2020, doi: 10.3390/rs12172683.

9.2 Proceedings Articles

- D. A. Jimenez-Sierra, B. Rasti, H. D. Benítez-Restrepo, J. Chanussot, "Graph Subspace Fusion from Super-Pixels for Biomass Estimation in Rice Crops", In 2022 Virtual Symposium in Plant Omics Sciences (OMICAS), IEEE, Nov. 2022. (**Accepted**)
- D. A. Jimenez-Sierra, H. D. Benítez-Restrepo, G. Arce, and J. F. Florez-Ospina, "Blue noise sampling and Nyström extension for graph based change detection", 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 2895-2898, Jul. 2021, doi: 10.1109/IGARSS47720.2021.9555107.

10 Acknowledgments

This thesis was made possible due to the support received from numerous people. First and foremost, I want to extend my deepest acknowledgments to God for allowing me to be enrolled in my Doctoral studies at Pontificia Universidad Javeriana Cali. Similarly, my advisor Dr. Hernán Darío Benítez Restrepo inspired me with his brilliant research abilities and his strive for excellence. Without his intellect, guidance, and discipline the results made in this Doctoral thesis wouldn't be possible. Many thanks, Dr. Hernán for all your efforts made during my research to achieve this important goal in my professional career.

I also would like to thank my co-advisor, Dr. Hernán Darío Dario Vargas Cardona, for his valuable guidance during this doctoral thesis. His contributions, corrections, ideas, and suggestions made possible the success presented in this work.

Similarly, I want to thank Dr. Jocelyn Chanussot from the Grenoble Institute of Technology (INP) and Dr. Behnood Rasti from Helmholtz-Zentrum Dresden-Rossendorf (HZDR). They were extremely welcoming and friendly during my internship at INRIA in Grenoble, France. Furthermore, their advice was valuable in finding new research paths and methods.

I also want to thank Dr. Julián David Colorado Montaño and Dr. Francisco Carlos Calderón Bocanegra, who support me by granting me access to their server Cratos. In addition, many thanks to all involved in the phenomics project (P4).

I also want to thank the Pontificia Universidad Javeriana Cali for supporting all the publication processes. Similarly, thanks to the Doctorate in Engineering and Applied Sciences program, especially its Director Dr. Andrés Jaramillo Botero and Assistant Andrea Ramirez Fernandez, for all the support given in the administrative processes required and financial support in the INRIA six-month internship. Furthermore, many special thanks to Leidi Johana Rojas Arroyave and all the Ómicas program administrative committee.

We acknowledge the funding provided by the OMICAS program: "Optimización Multiescala Insilico de Cultivos Agrícolas Sostenibles (Infraestructura y validación en Arroz y Caña de Azúcar)", anchored at the Pontificia Universidad Javeriana in Cali and funded within the Colombian Scientific Ecosystem by The World Bank, the Colombian Ministry of Science, Technology and Innovation, the Colombian Ministry of Education, the Colombian Ministry of Industry and Tourism, and ICETEX, under grant ID: FP44842-217-2018 and OMICAS Award ID: 792-61187.

This research also received funding from PROGRAMA ECOS-NORD INTERCAMBIO DE INVESTIGADORES COLOMBIA – FRANCIA 2021, supported by MINCIENCIAS, The Ministère de l'Europe et des Affaires étrangères (MEAE), and Le Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI).

Thanks to all my friends and colleagues at Pontificia Universidad Javeriana Cali. It was a privilege to get to know each one of them. Mainly, I would like to thank Miguel Romero, Camila Riccio, Jenny Gallo, Nicolas Lopez, Chrystian Sosa, Gustavo Lara, Carolina Saavedra, and Sammy Perdomo for their friendship and support during this process.

Last, but not least, I want to thank my family for their love and support. Special thanks to Mabelis Sierra, Gilberto Mejia, Fernando Jimenez, and Jessica Valencia for their never-ending love, support, and patience.

Dedicated to my parents

Mabelis Sierra, Gilberto Mejia, and Fernando Jimenez

11 References

[Achanta et al., 2012] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282.

[Afonso et al., 2010] Afonso, M. V., Bioucas-Dias, J. M., and Figueiredo, M. A. (2010). Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(9):2345–2356.

[Alfred et al., 2015] Alfred, R. et al. (2015). Performance of modeling time series using nonlinear autoregressive with exogenous input (narx) in the network traffic forecasting. In *2015 International Conference on Science in Information Technology (ICSITech)*, pages 164–168. IEEE.

[Alfredo Caicedo Barrero, 2010] Alfredo Caicedo Barrero, Graciela Wagner de García, R. M. M. P. (2010). Introducción a la teoría de grafos. volume 1, pages 1–12. Elizcom s.a.s.

[Amiri and Jamzad, 2018] Amiri, S. H. and Jamzad, M. (2018). Leveraging multi-modal fusion for graph-based image annotation. *Journal of Visual Communication and Image Representation*, 55:816–828.

[An et al., 2017] An, L., Chen, X., and Yang, S. (2017). Multi-graph feature level fusion for person re-identification. *Neurocomputing*, 259:39–45.

[Arroyo et al., 2017] Arroyo, J. A., Gomez-Castaneda, C., Ruiz, E., de Cote, E. M., Gavi, F., and Sucar, L. E. (2017). Uav technology and machine learning techniques applied to the yield improvement in precision agriculture. In *2017 IEEE Mexican Humanitarian Technology Conference (MHTC)*, pages 137–143. IEEE.

[Awad and Khanna, 2015] Awad, M. and Khanna, R. (2015). Support vector regression. In *Efficient Learning Machines*, pages 67–80. Springer.

[Bai et al., 2015] Bai, L., Xu, C., and Wang, C. (2015). A review of fusion methods of multi-spectral image. *Optik*, 126(24):4804–4807.

[Bai et al., 2014] Bai, X., Liu, C., Ren, P., Zhou, J., Zhao, H., and Su, Y. (2014). Object classification via feature fusion based marginalized kernels. *IEEE Geoscience and Remote Sensing Letters*, 12(1):8–12.

[Ban et al., 2018] Ban, Z., Liu, J., and Cao, L. (2018). Superpixel segmentation using gaussian mixture model. *IEEE Transactions on Image Processing*, 27(8):4105–4117.

[Belkin and Niyogi, 2003] Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396.

[Bendig et al., 2014] Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., and Bareth, G. (2014). Estimating biomass of barley using crop surface models (csms) derived from uav-based rgb imaging. *Remote Sensing*, 6(11):10395–10412.

[Bi et al., 2022] Bi, H., Xu, C., Shi, C., Liu, G., Li, Y., Zhang, H., and Qu, J. (2022). Srrv: A novel document object detector based on spatial-related relation and vision. *IEEE Transactions on Multimedia*.

[Boussaada et al., 2018] Boussaada, Z., Curea, O., Remaci, A., Camblong, H., and Mrabet Bellaaj, N. (2018). A nonlinear autoregressive exogenous (narx) neural network model for the prediction of the daily direct solar radiation. *Energies*, 11(3):620.

[Campos et al., 2021] Campos, J., García-Ruíz, F., and Gil, E. (2021). Assessment of vineyard canopy characteristics from vigour maps obtained using uav and satellite imagery. *Sensors*, 21(7):2363.

[CARPER et al., 1990] CARPER, W., LILLESAND, T., and KIEFER, R. (1990). The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data. *Photogrammetric Engineering and remote sensing*, 56(4):459–467.

[Celik, 2009] Celik, T. (2009). Unsupervised change detection in satellite images using principal component analysis and k-means clustering. *IEEE Geoscience and Remote Sensing Letters*, 6(4):772–776.

[Chandrakanth et al., 2014] Chandrakanth, R., Saibaba, J., Varadan, G., and Raj, P. A. (2014). A novel image fusion system for multisensor and multiband remote sensing data. *IETE Journal of Research*, 60(2):168–182.

[Chawla et al., 2020] Chawla, I., Karthikeyan, L., and Mishra, A. K. (2020). A review of remote sensing applications for water security: Quantity, quality, and extremes. *Journal of Hydrology*, page 124826.

[Chen and Shi, 2020] Chen, H. and Shi, Z. (2020). A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12(10):1662.

[Chen et al., 2022] Chen, H., Yokoya, N., Wu, C., and Du, B. (2022). Unsupervised multimodal change detection based on structural relationship graph representation learning. *arXiv preprint arXiv:2210.00941*.

[Cheng et al., 2017] Cheng, T., Song, R., Li, D., Zhou, K., Zheng, H., Yao, X., Tian, Y., Cao, W., and Zhu, Y. (2017). Spectroscopic estimation of biomass in canopy components of paddy rice using dry matter and chlorophyll indices. *Remote Sensing*, 9(4):319.

[Cheung et al., 2018] Cheung, G., Magli, E., Tanaka, Y., and Ng, M. K. (2018). Graph spectral image processing. *Proceedings of the IEEE*, 106(5):907–930.

[Colorado et al., 2020] Colorado, J. D., Calderon, F., Mendez, D., Petro, E., Rojas, J. P., Correa, E. S., Mondragon, I. F., Rebolledo, M. C., and Jaramillo-Botero, A. (2020). A novel nir-image segmentation method for the precise estimation of above-ground biomass in rice crops. *PloS one*, 15(10):e0239591.

[Correa et al., 2020] Correa, E. S., Calderon, F., and Colorado, J. D. (2020). Gfkuts: a novel multispectral image segmentation method applied to precision agriculture. In *2020 Virtual Symposium in Plant Omics Sciences (OMICAS)*, pages 1–6. IEEE.

[Cover and Thomas, 2012] Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

[Dabov et al., 2007] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095.

[Dalla Mura et al., 2015] Dalla Mura, M., Prasad, S., Pacifici, F., Gamba, P., Chanussot, J., and Benediktsson, J. A. (2015). Challenges and opportunities of multimodality and data fusion in remote sensing. *Proceedings of the IEEE*, 103(9):1585–1601.

[Debes et al., 2014] Debes, C., Merentitis, A., Heremans, R., Hahn, J., Frangiadakis, N., van Kasteren, T., Liao, W., Bellens, R., Pižurica, A., Gautama, S., et al. (2014). Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2405–2418.

[Devia et al., 2019] Devia, C. A., Rojas, J. P., Petro, E., Martinez, C., Mondragon, I. F., Patino, D., Rebolledo, M. C., and Colorado, J. (2019). High-throughput biomass estimation in rice crops using uav multispectral imagery. *Journal of Intelligent & Robotic Systems*, 96(3-4):573–589.

[Dian et al., 2018] Dian, R., Li, S., Guo, A., and Fang, L. (2018). Deep hyperspectral image sharpening. *IEEE transactions on neural networks and learning systems*, (99):1–11.

[Ding et al., 2010] Ding, M., Tian, Z., Jin, Z., Xu, M., and Cao, C. (2010). Registration using robust kernel principal component for object-based change detection. *IEEE Geoscience and Remote Sensing Letters*, 7(4):761–765.

[Dong et al., 2019] Dong, X., Thanou, D., Rabbat, M., and Frossard, P. (2019). Learning graphs from data: A signal representation perspective. *IEEE Signal Processing Magazine*, 36(3):44–63.

[Ehlers et al., 2010] Ehlers, M., Klonus, S., Johan Åstrand, P., and Rosso, P. (2010). Multi-sensor image fusion for pansharpening in remote sensing. *International Journal of Image and Data Fusion*, 1(1):25–45.

[Fowlkes et al., 2004] Fowlkes, C., Belongie, S., Chung, F., and Malik, J. (2004). Spectral grouping using the nystrom method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):214–225.

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

[Ghassemian, 2016] Ghassemian, H. (2016). A review of remote sensing image fusion methods. *Information Fusion*, 32:75–89.

[Gitelson et al., 2002] Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote sensing of Environment*, 80(1):76–87.

[Gnyp et al., 2014] Gnyp, M. L., Miao, Y., Yuan, F., Ustin, S. L., Yu, K., Yao, Y., Huang, S., and Bareth, G. (2014). Hyperspectral canopy sensing of paddy rice aboveground biomass at different growth stages. *Field Crops Research*, 155:42–55.

[Goldstein and Osher, 2009] Goldstein, T. and Osher, S. (2009). The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343.

[Golipour et al., 2015] Golipour, M., Ghassemian, H., and Mirzapour, F. (2015). Integrating hierarchical segmentation maps with mrf prior for classification of hyperspectral images in a bayesian framework. *IEEE Transactions on Geoscience and remote Sensing*, 54(2):805–816.

[Harrell et al., 2011] Harrell, D., Tubana, B., Walker, T., and Phillips, S. (2011). Estimating rice grain yield potential using normalized difference vegetation index. *Agronomy Journal*, 103(6):1717–1723.

[Hassanzadeh et al., 2018] Hassanzadeh, A., Kaarna, A., and Kauranne, T. (2018). Sequential spectral clustering of hyperspectral remote sensing image over bipartite graph. *Applied Soft Computing*, 73:727–734.

[He et al., 2019] He, F., Nie, F., Wang, R., Li, X., and Jia, W. (2019). Fast semisupervised learning with bipartite graph for large-scale data. *IEEE Transactions on Neural Networks and Learning Systems*, 31(2):626–638.

[He et al., 2010] He, K., Sun, J., and Tang, X. (2010). Guided image filtering. In *European conference on computer vision*, pages 1–14. Springer.

[Honrado et al., 2017] Honrado, J., Solpico, D. B., Favila, C., Tongson, E., Tangonan, G. L., and Libatique, N. J. (2017). Uav imaging with low-cost multispectral imaging system for precision agriculture applications. In *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, pages 1–7. IEEE.

[Huang et al., 2015] Huang, W., Xiao, L., Wei, Z., Liu, H., and Tang, S. (2015). A new pan-sharpening method with deep neural networks. *IEEE Geoscience and Remote Sensing Letters*, 12(5):1037–1041.

[Huang et al., 2005] Huang, Y., Liao, J., Guo, H., and Zhong, X. (2005). The fusion of multispectral and sar images based wavelet transformation over urban area. In *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05.*, volume 6, pages 3942–3944. Ieee.

[Huo and Poo, 2013] Huo, F. and Poo, A.-N. (2013). Nonlinear autoregressive network with exogenous inputs based contour error reduction in cnc machines. *International Journal of Machine Tools and Manufacture*, 67:45–52.

[Imani and Ghassemian, 2018] Imani, M. and Ghassemian, H. (2018). Discriminant analysis in morphological feature space for high-dimensional image spatial–spectral classification. *Journal of Applied Remote Sensing*, 12(1):016024.

[Imani and Ghassemian, 2020] Imani, M. and Ghassemian, H. (2020). An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges. *Information fusion*, 59:59–83.

[Iyer et al., 2017] Iyer, G., Chanussot, J., and Bertozzi, A. L. (2017). A graph-based approach for feature extraction and segmentation of multimodal images. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3320–3324. IEEE.

[Iyer et al., 2020] Iyer, G., Chanussot, J., and Bertozzi, A. L. (2020). A graph-based approach for data fusion and segmentation of multimodal images. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–11.

[Jimenez-Sierra et al., 2021a] Jimenez-Sierra, D. A., Benítez-Restrepo, H. D., Arce, G. R., and Florez-Ospina, J. F. (2021a). Blue noise sampling and nystrom extension for graph based change detection. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 2895–2898. IEEE.

[Jimenez-Sierra et al., 2020] Jimenez-Sierra, D. A., Benítez-Restrepo, H. D., Vargas-Cardona, H. D., and Chanussot, J. (2020). Graph-based data fusion applied to: Change detection and biomass estimation in rice crops. *Remote Sensing*, 12(17):2683.

[Jimenez-Sierra et al., 2021b] Jimenez-Sierra, D. A., Correa, E. S., Benítez-Restrepo, H. D., Calderon, F. C., Mondragon, I. F., and Colorado, J. D. (2021b). Novel feature-extraction methods for the estimation of above-ground biomass in rice crops. *Sensors*, 21(13):4369.

[Jimenez-Sierra et al., 2022] Jimenez-Sierra, D. A., Quintero-Olaya, D. A., Alvear-Muñoz, J. C., Benítez-Restrepo, H. D., Florez-Ospina, J. F., and Chanussot, J. (2022). Graph learning based on signal smoothness representation for homogeneous and heterogeneous change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16.

[Jolliffe and Cadima, 2016] Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202.

[Kalofolias and Perraudin, 2019] Kalofolias, V. and Perraudin, N. (2019). Large scale graph learning from smooth signals. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*.

[Kandel et al., 2007] Kandel, A., Bunke, H., and Last, M. (2007). *Applied graph theory in computer vision and pattern recognition*. Springer Science & Business Media.

[Kang et al., 2020] Kang, Z., Shi, G., Huang, S., Chen, W., Pu, X., Zhou, J. T., and Xu, Z. (2020). Multi-graph fusion for multi-view spectral clustering. *Knowledge-Based Systems*, 189:105102.

[Kanke et al., 2016] Kanke, Y., Tubana, B., Dalen, M., and Harrell, D. (2016). Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields. *Precision agriculture*, 17(5):507–530.

[Karthikeyan et al., 2020] Karthikeyan, L., Chawla, I., and Mishra, A. K. (2020). A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *Journal of Hydrology*, page 124905.

[Kittler and Illingworth, 1986] Kittler, J. and Illingworth, J. (1986). Minimum error thresholding. *Pattern recognition*, 19(1):41–47.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

[Kulkarni and Rege, 2020] Kulkarni, S. C. and Rege, P. P. (2020). Pixel level fusion techniques for sar and optical images: A review. *Information Fusion*.

[Kulkarni et al., 2019] Kulkarni, S. C., Rege, P. P., and Parishwad, O. (2019). Hybrid fusion approach for synthetic aperture radar and multispectral imagery for improvement in land use land cover classification. *Journal of Applied Remote Sensing*, 13(3):034516.

[Laben and Brower, 2000] Laben, C. A. and Brower, B. V. (2000). Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. US Patent 6,011,875.

[Ladickỳ et al., 2010] Ladickỳ, L., Sturgess, P., Alahari, K., Russell, C., and Torr, P. H. (2010). What, where and how many? combining object detectors and crfs. In *European Conference on Computer Vision*, pages 424–437. Springer.

[Laref et al., 2019] Laref, R., Losson, E., Sava, A., and Siadat, M. (2019). On the optimization of the support vector machine regression hyperparameters setting for gas sensors array applications. *Chemometrics and Intelligent Laboratory Systems*, 184:22–27.

[Lau and Arce, 2018] Lau, D. L. and Arce, G. R. (2018). *Modern digital halftoning*, volume 1. CRC Press.

[Lechner et al., 2020] Lechner, A. M., Foody, G. M., and Boyd, D. S. (2020). Applications in remote sensing to forest ecology and management. *One Earth*, 2(5):405–412.

[Lei et al., 2019a] Lei, T., Jia, X., Liu, T., Liu, S., Meng, H., and Nandi, A. K. (2019a). Adaptive morphological reconstruction for seeded image segmentation. *IEEE Transactions on Image Processing*, 28(11):5510–5523.

[Lei et al., 2019b] Lei, T., Zhang, Y., Lv, Z., Li, S., Liu, S., and Nandi, A. K. (2019b). Landslide inventory mapping from bitemporal images using deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 16(6):982–986.

[Lewis et al., 2004] Lewis, J., O'callaghan, R., Nikolov, S., Bull, D., and Canagarajah, C. (2004). Region-based image fusion using complex wavelets. In *Seventh International Conference on Information Fusion (FUSION)*, volume 1, pages 555–562. Citeseer.

[Li et al., 2022] Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B., and Chanussot, J. (2022). Deep learning in multimodal remote sensing data fusion: A comprehensive review. *International Journal of Applied Earth Observation and Geoinformation*, 112:102926.

[Li et al., 2021a] Li, J., Wu, Z., Hu, Z., Li, Z., Wang, Y., and Molinier, M. (2021a). Deep learning based thin cloud removal fusing vegetation red edge and short wave infrared spectral information for sentinel-2a imagery. *Remote Sensing*, 13(1):157.

[Li et al., 2020] Li, M., Wang, L., Deng, S., and Zhou, C. (2020). Color image segmentation using adaptive hierarchical-histogram thresholding. *PloS one*, 15(1):e0226345.

[Li et al., 2021b] Li, Y., Xu, W., Chen, H., Jiang, J., and Li, X. (2021b). A novel framework based on mask r-cnn and histogram thresholding for scalable segmentation of new and old rural buildings. *Remote Sensing*, 13(6):1070.

[Liao et al., 2015] Liao, W., Dalla Mura, M., Chanussot, J., and Pižurica, A. (2015). Fusion of spectral and spatial information for classification of hyperspectral remote-sensed imagery by local graph. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(2):583–594.

[Lin et al., 2020] Lin, F., Guo, S., Tan, C., Zhou, X., and Zhang, D. (2020). Identification of rice sheath blight through spectral responses using hyperspectral images. *Sensors*, 20(21):6243.

[Liu et al., 2005] Liu, D., Kelly, M., and Gong, P. (2005). Classifying multi-temporal tm imagery using markov random fields and support vector machines. In *International Workshop on the Analysis of Multi-Temporal Remote Sensing Images, 2005.*, pages 225–228. IEEE.

[Liu et al., 2016] Liu, J., Gong, M., Qin, K., and Zhang, P. (2016). A deep convolutional coupling network for change detection based on heterogeneous optical and radar images. *IEEE Transactions on Neural Networks and Learning Systems*, 29(3):545–559.

[Liu et al., 2015] Liu, M., Dai, Y., Zhang, J., Zhang, X., Meng, J., and Xie, Q. (2015). Pca-based sea-ice image fusion of optical data by his transform and sar data by wavelet transform. *Acta Oceanologica Sinica*, 34(3):59–67.

[Liu et al., 2017a] Liu, S., Du, Q., and Tong, X. (2017a). Band selection for change detection from hyperspectral images. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIII*, volume 10198, page 101980T. International Society for Optics and Photonics.

[Liu et al., 2018] Liu, Y., Chen, X., Wang, Z., Wang, Z. J., Ward, R. K., and Wang, X. (2018). Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion*, 42:158–173.

[Liu et al., 2017b] Liu, Z., Li, G., Mercier, G., He, Y., and Pan, Q. (2017b). Change detection in heterogenous remote sensing images via homogeneous pixel transformation. *IEEE Transactions on Image Processing*, 27(4):1822–1834.

[Longbotham et al., 2011] Longbotham, N., Chaapel, C., Bleiler, L., Padwick, C., Emery, W. J., and Pacifici, F. (2011). Very high resolution multiangle urban classification analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4):1155–1170.

[Luo et al., 2012] Luo, B., Khan, M. M., Bienvenu, T., Chanussot, J., and Zhang, L. (2012). Decision-based fusion for pansharpening of remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 10(1):19–23.

[Luppino et al., 2017] Luppino, L. T., Anfinsen, S. N., Moser, G., Jenssen, R., Bianchi, F. M., Serpico, S., and Mercier, G. (2017). A clustering approach to heterogeneous change detection. In *Scandinavian Conference on Image Analysis*, pages 181–192. Springer.

[Luppino et al., 2019] Luppino, L. T., Bianchi, F. M., Moser, G., and Anfinsen, S. N. (2019). Unsupervised image regression for heterogeneous change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):9960–9975.

[Luppino et al., 2022] Luppino, L. T., Hansen, M. A., Kampffmeyer, M., Bianchi, F. M., Moser, G., Jenssen, R., and Anfinsen, S. N. (2022). Code-aligned autoencoders for unsupervised change

detection in multimodal remote sensing images. *IEEE Transactions on Neural Networks and Learning Systems*.

[Luppino et al., 2021] Luppino, L. T., Kampffmeyer, M., Bianchi, F. M., Moser, G., Serpico, S. B., Jenssen, R., and Anfinsen, S. N. (2021). Deep image translation with an affinity-based change prior for unsupervised multimodal change detection. *IEEE Transactions on Geoscience and Remote Sensing*.

[Ma et al., 2019] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152:166–177.

[Ma and Fu, 2011] Ma, Y. and Fu, Y. (2011). *Manifold learning theory and applications*. CRC press.

[Maaten and Hinton, 2008] Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605.

[Mahmoudi et al., 2014] Mahmoudi, F. T., Samadzadegan, F., and Reinartz, P. (2014). Object recognition based on the context aware decision-level fusion in multiviews imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(1):12–22.

[Marjanovic and Solo, 2014] Marjanovic, G. and Solo, V. (2014). $l\_\{q\}$ sparsity penalized linear regression with cyclic descent. *IEEE Transactions on Signal Processing*, 62(6):1464–1475.

[Mateos et al., 2019] Mateos, G., Segarra, S., Marques, A. G., and Ribeiro, A. (2019). Connecting the dots: Identifying network structure via graph signal processing. *IEEE Signal Processing Magazine*, 36(3):16–43.

[Melgani and Serpico, 2003] Melgani, F. and Serpico, S. B. (2003). A markov random field approach to spatio-temporal contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 41(11):2478–2487.

[Men et al., 2014] Men, Z., Yee, E., Lien, F.-S., Yang, Z., and Liu, Y. (2014). Ensemble nonlinear autoregressive exogenous artificial neural networks for short-term wind speed and power forecasting. *International Scholarly Research Notices*, 2014.

[Mirzapour and Ghassemian, 2015] Mirzapour, F. and Ghassemian, H. (2015). Improving hyperspectral image classification by combining spectral, texture, and shape features. *International Journal of Remote Sensing*, 36(4):1070–1096.

[Mohamed Shakeel et al., 2020] Mohamed Shakeel, P., Desa, M. I., and Burhanuddin, M. (2020). Improved watershed histogram thresholding with probabilistic neural networks for lung cancer diagnosis for cbmir systems. *Multimedia Tools and Applications*, 79(23):17115–17133.

[Mura et al., 2015] Mura, M. D., Prasad, S., Pacifici, F., Gamba, P., Chanussot, J., and Benediktsson, J. A. (2015). Challenges and opportunities of multimodality and data fusion in remote sensing. *Proceedings of the IEEE*, 103(9):1585–1601.

[Naito et al., 2017] Naito, H., Ogawa, S., Valencia, M. O., Mohri, H., Urano, Y., Hosoi, F., Shimizu, Y., Chavez, A. L., Ishitani, M., Selvaraj, M. G., et al. (2017). Estimating rice yield related traits and quantitative trait loci analysis under different nitrogen treatments using a simple tower-based field phenotyping system with modified single-lens reflex cameras. *ISPRS Journal of Photogrammetry and Remote Sensing*, 125:50–62.

[Ndikumana et al., 2018] Ndikumana, E., Minh, D. H. T., Thu, D. N. H., Baghdadi, N., Courault, D., Hossard, L., and El Moussawi, I. (2018). Rice height and biomass estimations using multi-temporal sar sentinel-1: Camargue case study. In *Remote Sensing for Agriculture, Ecosystems, and Hydrology XX*, volume 10783, page 107830U. International Society for Optics and Photonics.

[Niu et al., 2018] Niu, X., Gong, M., Zhan, T., and Yang, Y. (2018). A conditional adversarial network for change detection in heterogeneous images. *IEEE Geoscience and Remote Sensing Letters*, 16(1):45–49.

[Otsu, 1979] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66.

[Palsson et al., 2017] Palsson, F., Sveinsson, J. R., and Ulfarsson, M. O. (2017). Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network. *IEEE Geoscience and Remote Sensing Letters*, 14(5):639–643.

[Palubinskas, 2013] Palubinskas, G. (2013). Fast, simple, and good pan-sharpening method. *Journal of Applied Remote Sensing*, 7(1):073526.

[Pan and Yang, 2009] Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

[Pandit and Bhiwani, 2015] Pandit, V. R. and Bhiwani, R. (2015). Image fusion in remote sensing applications: A review. *International journal of computer applications*, 120(10).

[Parada-Mayorga et al., 2019] Parada-Mayorga, A., Lau, D. L., Giraldo, J. H., and Arce, G. R. (2019). Blue-noise sampling on graphs. *IEEE Transactions on Signal and Information Processing over Networks*, 5(3):554–569.

[Perraudin and Vandergheynst, 2017] Perraudin, N. and Vandergheynst, P. (2017). Stationary signal processing on graphs. *IEEE Transactions on Signal Processing*, 65(13):3462–3477.

[Petschnigg et al., 2004] Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., and Toyama, K. (2004). Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics (TOG)*, 23(3):664–672.

[Piella, 2003] Piella, G. (2003). A general framework for multiresolution image fusion: from pixels to regions. *Information fusion*, 4(4):259–280.

[Pohl, 1998] Pohl, C. (1998). Multisensor image fusion in remote sensing, review article. *Int. J. Remote Sensing*, 19(5):823–854.

[Prabhakara et al., 2015] Prabhakara, K., Hively, W. D., and McCarty, G. W. (2015). Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in maryland, united states. *International journal of applied earth observation and geoinformation*, 39:88–102.

[Rasti and Ghamisi, 2020] Rasti, B. and Ghamisi, P. (2020). Remote sensing image classification using subspace sensor fusion. *Information Fusion*, 64:121–130.

[Ricaud et al., 2019] Ricaud, B., Borgnat, P., Tremblay, N., Gonçalves, P., and Vandergheynst, P. (2019). Fourier could be a data scientist: From graph fourier transform to signal processing on graphs. *Comptes Rendus Physique*, 20(5):474–488.

[Rother et al., 2004] Rother, C., Kolmogorov, V., and Blake, A. (2004). " grabcut" interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG)*, 23(3):309–314.

[Saboksayr et al., 2021] Saboksayr, S. S., Mateos, G., and Cetin, M. (2021). Online discriminative graph learning from multi-class smooth signals. *Signal Processing*, 186:108101.

[Sawant and Prabukumar, 2018] Sawant, S. S. and Prabukumar, M. (2018). A review on graph-based semi-supervised learning methods for hyperspectral image classification. *The Egyptian Journal of Remote Sensing and Space Science*.

[Schowengerdt, 2006] Schowengerdt, R. A. (2006). *Remote sensing: models and methods for image processing*. Elsevier.

[Serviansky et al., 2020] Serviansky, H., Segol, N., Shlomi, J., Cranmer, K., Gross, E., Maron, H., and Lipman, Y. (2020). Set2graph: Learning graphs from sets. *Advances in Neural Information Processing Systems*, 33.

[Shao and Cai, 2018] Shao, Z. and Cai, J. (2018). Remote sensing image fusion with deep convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5):1656–1669.

[Shapiro, 2003] Shapiro, A. (2003). Monte carlo sampling methods. *Handbooks in Operations Research and Management Science*, 10:353–425.

[Shekkizhar and Ortega, 2020] Shekkizhar, S. and Ortega, A. (2020). Efficient graph construction for image representation. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1956–1960. IEEE.

[Shi et al., 2020a] Shi, W., Zhang, M., Zhang, R., Chen, S., and Zhan, Z. (2020a). Change detection based on artificial intelligence: State-of-the-art and challenges. *Remote Sensing*, 12(10):1688.

[Shi et al., 2020b] Shi, W., Zhang, M., Zhang, R., Chen, S., and Zhan, Z. (2020b). Change detection based on artificial intelligence: State-of-the-art and challenges. *Remote Sensing*, 12(10).

[Shivakumar and Rajashekararadhya, 2017] Shivakumar, B. and Rajashekararadhya, S. (2017). Performance evaluation of spectral angle mapper and spectral correlation mapper classifiers over multiple remote sensor data. In *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pages 1–6. IEEE.

[Sroubek et al., 2010] Sroubek, F., Flusser, J., and Zitova, B. (2010). Image fusion: Principles, methods, and applications. In *International Conference on Computer Graphics and Imaging*, pages 800–870.

[Sun et al., 2021a] Sun, Y., Lei, L., Guan, D., and Kuang, G. (2021a). Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images. *IEEE Transactions on Image Processing*, 30:6277–6291.

[Sun et al., 2021b] Sun, Y., Lei, L., Guan, D., Li, M., and Kuang, G. (2021b). Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote-sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14.

[Sun et al., 2022a] Sun, Y., Lei, L., Guan, D., Wu, J., and Kuang, G. (2022a). Iterative structure transformation and conditional random field based method for unsupervised multimodal change detection. *Pattern Recognition*, 131:108845.

[Sun et al., 2021c] Sun, Y., Lei, L., Li, X., Sun, H., and Kuang, G. (2021c). Nonlocal patch similarity based heterogeneous remote sensing change detection. *Pattern Recognition*, 109:107598.

[Sun et al., 2021d] Sun, Y., Lei, L., Li, X., Tan, X., and Kuang, G. (2021d). Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*.

[Sun et al., 2022b] Sun, Y., Lei, L., Tan, X., Guan, D., Wu, J., and Kuang, G. (2022b). Structured graph based image regression for unsupervised multimodal change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 185:16–31.

[Szeliski et al., 2008] Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C. (2008). A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080.

[Tanaka et al., 2020] Tanaka, Y., Eldar, Y. C., Ortega, A., and Cheung, G. (2020). Sampling signals on graphs: From theory to applications. *IEEE Signal Processing Magazine*, 37(6):14–30.

[Tong et al., 2017] Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D., Initiative, A. D. N., et al. (2017). Multi-modal classification of alzheimer's disease using nonlinear graph fusion. *Pattern recognition*, 63:171–181.

[Vakalopoulou et al., 2015] Vakalopoulou, M., Karantzalos, K., Komodakis, N., and Paragios, N. (2015). Simultaneous registration and change detection in multitemporal, very high resolution remote sensing data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 61–69.

[Van der Maaten and Hinton, 2008] Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11).

[Viljanen et al., 2018] Viljanen, N., Honkavaara, E., Näsi, R., Hakala, T., Niemeläinen, O., and Kaivosoja, J. (2018). A novel machine learning method for estimating biomass of grass swards using a photogrammetric canopy height model, images and vegetation indices captured by a drone. *Agriculture*, 8(5):70.

[Vivone et al., 2014] Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A., Restaino, R., and Wald, L. (2014). A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2565–2586.

[Welling and Kipf, 2016] Welling, M. and Kipf, T. N. (2016). Semi-supervised classification with graph convolutional networks. In *J. International Conference on Learning Representations (ICLR 2017)*.

[Wen et al., 2022] Wen, W., Timmermans, J., Chen, Q., and van Bodegom, P. M. (2022). Monitoring the combined effects of drought and salinity stress on crops using remote sensing in the netherlands. *Hydrology and Earth System Sciences*, 26(17):4537–4552.

[Xiao et al., 2020] Xiao, Z., Yun-xuan, B., Lin, W., Zheng-zhen, D., Qian, T., and Can, C. (2020). Hyperspectral features of rice canopy and spad values estimation under the stress of rice leaf folder. *Chinese Journal of Agrometeorology*, 41(03):173.

[Xing et al., 2018] Xing, Y., Wang, M., Yang, S., and Jiao, L. (2018). Pan-sharpening via deep metric learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:165–183.

[Yang et al., 2018] Yang, J., Zhao, Y.-Q., and Chan, J. C.-W. (2018). Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network. *Remote Sensing*, 10(5):800.

[Yang et al., 2019] Yang, X., Jia, Z., Yang, J., and Kasabov, N. (2019). Change detection of optical remote sensing image disturbed by thin cloud using wavelet coefficient substitution algorithm. *Sensors*, 19(9):1972.

[Yavariabdi and Kusetogullari, 2017] Yavariabdi, A. and Kusetogullari, H. (2017). Change detection in multispectral landsat images using multiobjective evolutionary algorithm. *IEEE Geoscience and Remote Sensing Letters*, 14(3):414–418.

[Yigit and Mendes, 2018] Yigit, S. and Mendes, M. (2018). Which effect size measure is appropriate for one-way and two-way anova models?: A monte carlo simulation study. *REVSTAT-Statistical Journal*, 16(3):295–313.

[Yue et al., 2018a] Yue, J., Feng, H., Jin, X., Yuan, H., Li, Z., Zhou, C., Yang, G., and Tian, Q. (2018a). A comparison of crop parameters estimation using images from uav-mounted snapshot hyperspectral sensor and high-definition digital camera. *Remote Sensing*, 10(7):1138.

[Yue et al., 2018b] Yue, J., Feng, H., Yang, G., and Li, Z. (2018b). A comparison of regression techniques for estimation of above-ground winter wheat biomass using near-surface spectroscopy. *Remote Sensing*, 10(1):66.

[Zanetti et al., 2015] Zanetti, M., Bovolo, F., and Bruzzone, L. (2015). Rayleigh-rice mixture parameter estimation via em algorithm for change detection in multispectral images. *IEEE Transactions on Image Processing*, 24(12):5004–5016.

[Zanetti and Bruzzone, 2017] Zanetti, M. and Bruzzone, L. (2017). A theoretical framework for change detection based on a compound multiclass statistical model of the difference image. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):1129–1143.

[Zanetti and Bruzzone, 2018] Zanetti, M. and Bruzzone, L. (2018). A theoretical framework for change detection based on a compound multiclass statistical model of the difference image. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):1129–1143.

[Zhang et al., 2020a] Zhang, C., Yue, P., Tapete, D., Jiang, L., Shangguan, B., Huang, L., and Liu, G. (2020a). A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:183–200.

[Zhang and Shi, 2020] Zhang, M. and Shi, W. (2020). A feature difference convolutional neural network-based change detection method. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10):7232–7246.

[Zhang et al., 2018] Zhang, Q., Liu, Y., Blum, R. S., Han, J., and Tao, D. (2018). Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Information Fusion*, 40:57–75.

[Zhang et al., 2020b] Zhang, R., Zhang, Y., and Li, X. (2020b). Unsupervised feature selection via adaptive graph learning and constraint. *IEEE Transactions on Neural Networks and Learning Systems*.

[Zhang and Yu, 2010] Zhang, W. and Yu, L. (2010). Sar and landsat etm+ image fusion using variational model. In *2010 International Conference on Computer and Communication Technologies in Agriculture Engineering*, volume 3, pages 205–207. IEEE.

[Zhang, 2004] Zhang, Y. (2004). Understanding image fusion. *Photogrammetric engineering and remote sensing*, 70(6):657–661.

[ZhiYong et al., 2021] ZhiYong, L., Liu, T., Benediktsson, J. A., and Falco, N. (2021). Land cover change detection techniques: Very-high-resolution optical images: A review. *IEEE Geoscience and Remote Sensing Magazine*.

Annexes

A Abbreviations

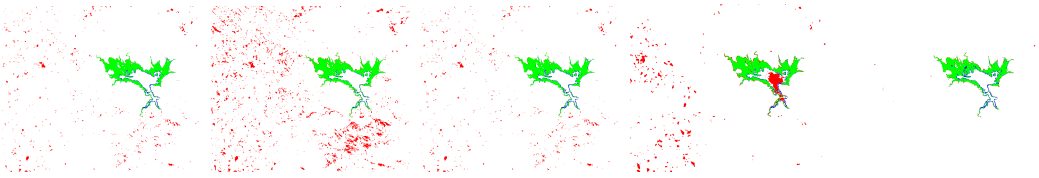
- **AEs:** Autoencoders.
- **BT:** Brovery Transform.
- **CNN:** Convolutional Neural Network.
- **CS:** Component Substitution Methods.
- **DBNs:** Deep Belief Networks.
- **DL:** Deep Learning.
- **DWT:** Discrete Wavelet Transform.
- **FA:** False Alarms.
- **GANs:** Generative Adversarial Networks.
- **GBF:** Graph-Based Fusion.
- **GBF-CD:** Graph-Based Fusion for Change Detection.
- **GDF:** Graph-based Data Fusion.
- **GS:** Gram-Schmidt.
- **HS:** Hyper-spectral.
- **HPF:** High Pass Filtering.
- **K:** Cohen's kappa.
- **LiDAR:** Light Detection and Ranging.
- **IHS:** Intensity-Hue-Saturation.
- **MA:** Missed Alarms.
- **MRA:** Multi-Resolution Analysis methods
- **MRF:** Markov Random Fields.
- **MS:** Multi-Spectral.
- **OE:** Overall Error.
- **P:** Precision.

- **PAN:** Panchromatic.
- **PCA:** Principal Component Analysis.
- **PCNNs:** Pulse Couple Neural Networks.
- **RGB:** Red-Green-Blue (color).
- **RNNs:** Recurrent Neural Networks.
- **R:** Recall.
- **RQ:** Research Questions.
- **rR:** Rayleigh-Rice.
- **rrR:** Rayleigh-Rayleigh- Rice.
- **SO:** Specific Objective.
- **SAE:** Stacked Auto-Encoders.
- **SAR:** Synthetic Aperture Radar.
- **U-CD-HPT:** Unsupervised Change Detection using Homogeneous Pixel Transformation regression.

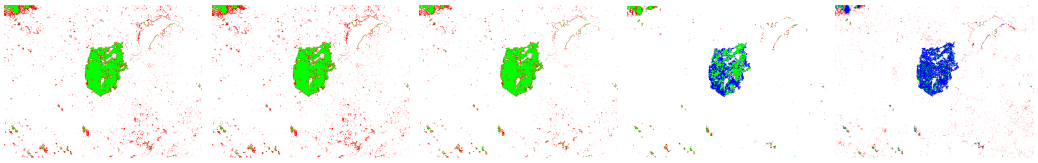
B Supplementary material

Table 34: Elapsed time of algorithms to generate superpixels in seconds.

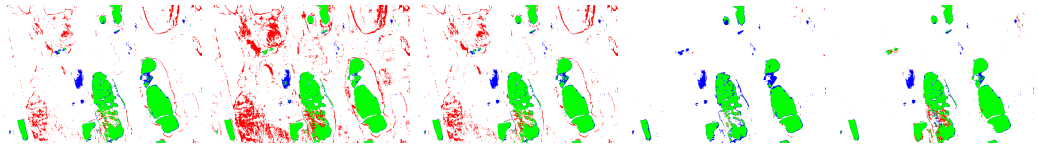
DATASETS	GMMSP	SLIC	AMR
Mulargia	**0.1569**	0.2043	0.6833
Omodeo	**0.2952**	0.3443	2.8955
Alaska	**0.1575**	0.1873	0.8754
Madeirinha	**0.0971**	0.1720	0.8380
Prince George	1.7699	**1.5796**	18.8972
Gloucester-1	3.3665	**3.1336**	50.6731
Katios	**0.3978**	0.4612	3.9608
Atlantico	**0.2691**	0.3331	2.6071
San Francisco	**0.0591**	0.1301	0.4178
Wenchuan	**0.0591**	0.1516	0.7310
Toulouse	4.5445	**3.6434**	43.7703
California	2.8939	**2.3647**	41.1072
Bastrop	**0.4834**	0.4971	4.1190
Gloucester-2	3.8686	**3.0163**	37.5695



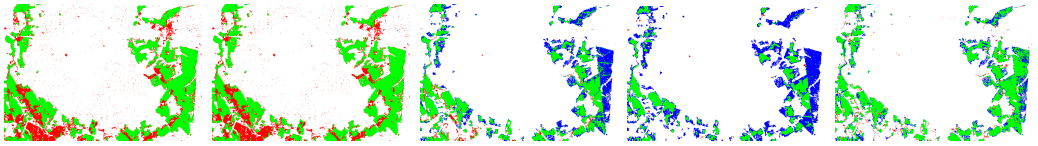
(**a**) Mulargia



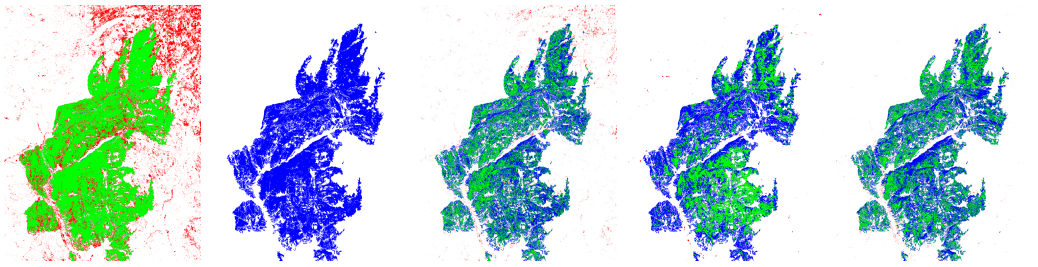
(**b**) Omodeo



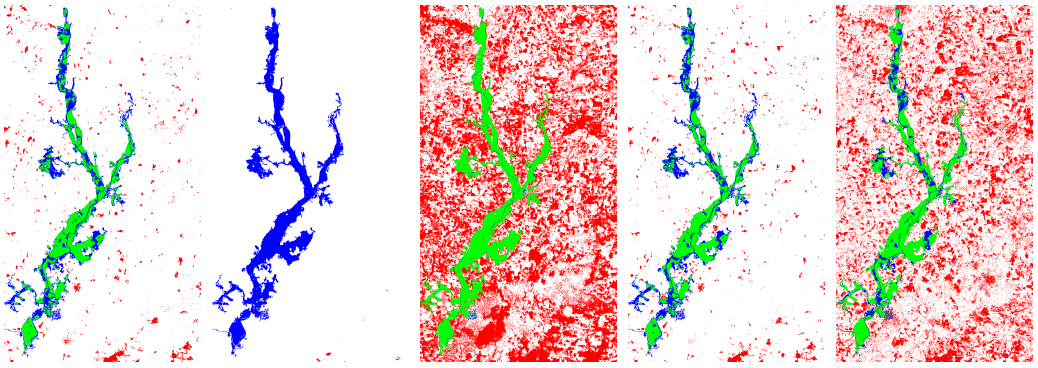
(**c**) Alaska



(**d**) Madeirinha

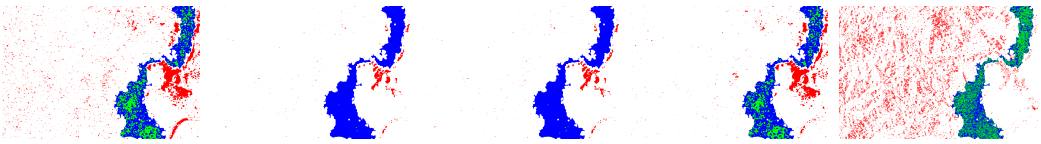


(**e**) Prince George

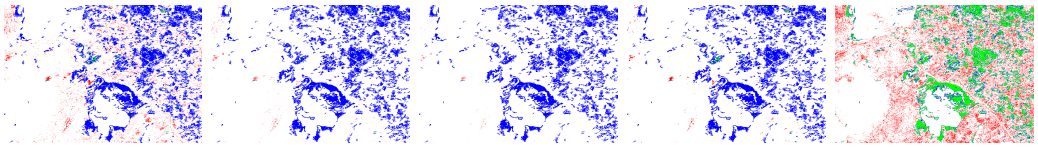


(**f**) Gloucester-1

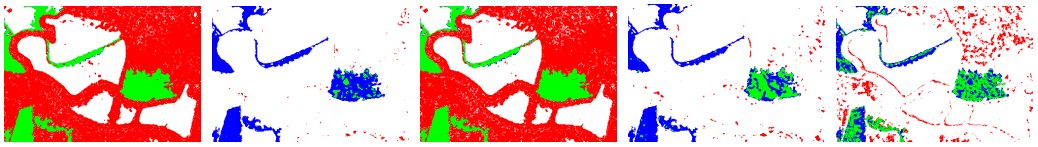
Figure 38: *Cont.*



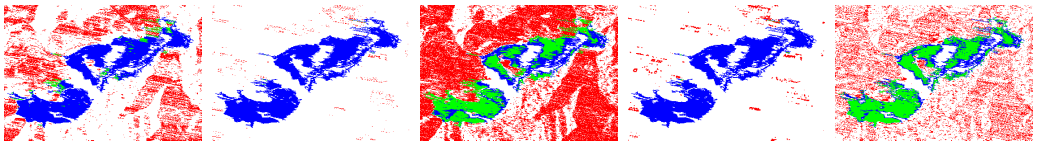
(**g**) Katios National Park



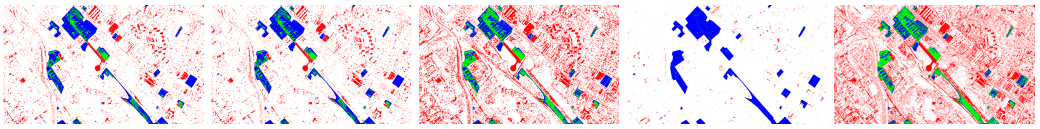
(**h**) Atlantico



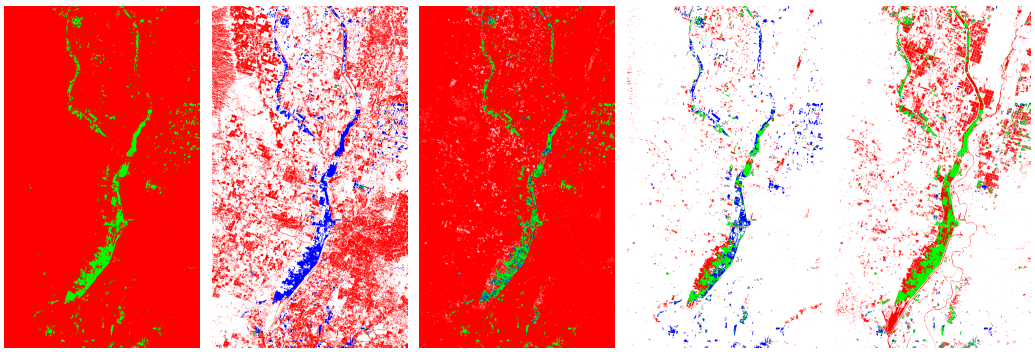
(**i**) San Francisco



(**j**) WenChuan

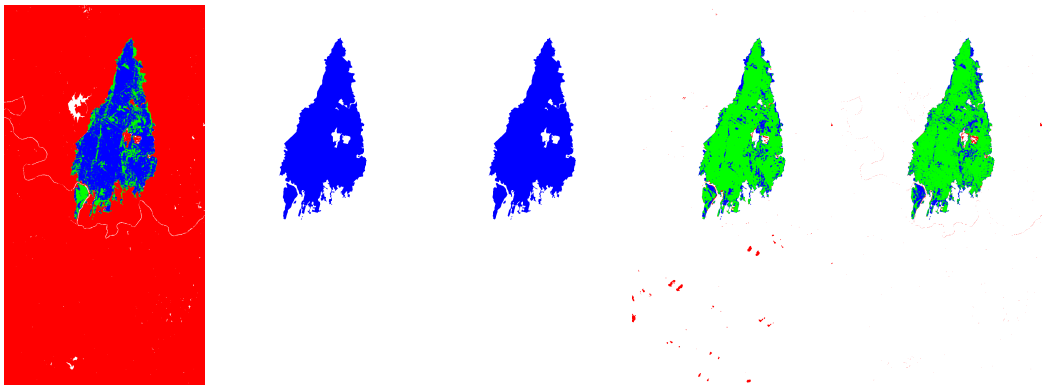


(**k**) Toulouse

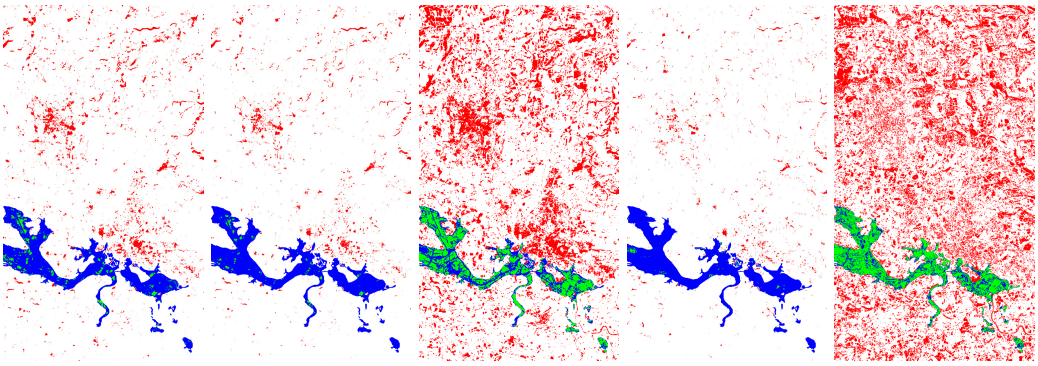


(**l**) California

Figure 38: *Cont.*

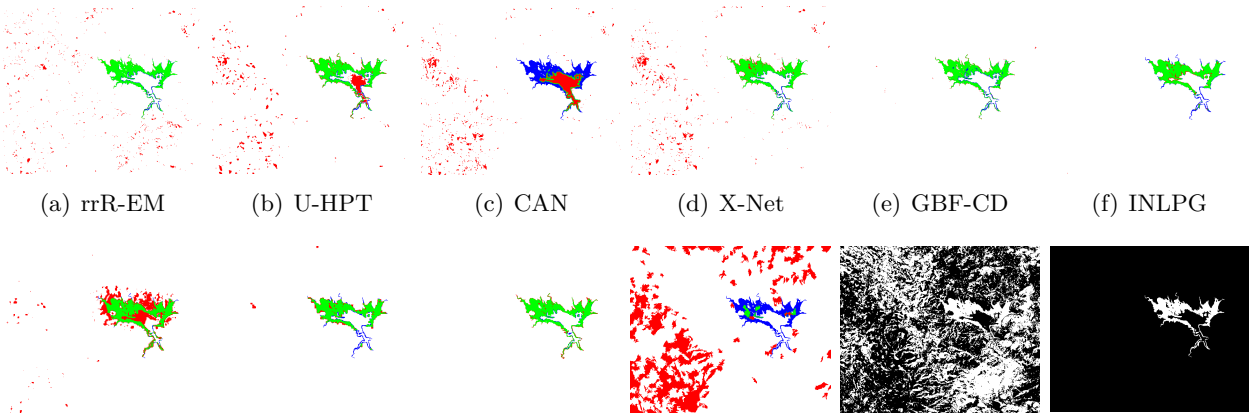


(**m**) Bastrop



(**n**) Gloucester-2

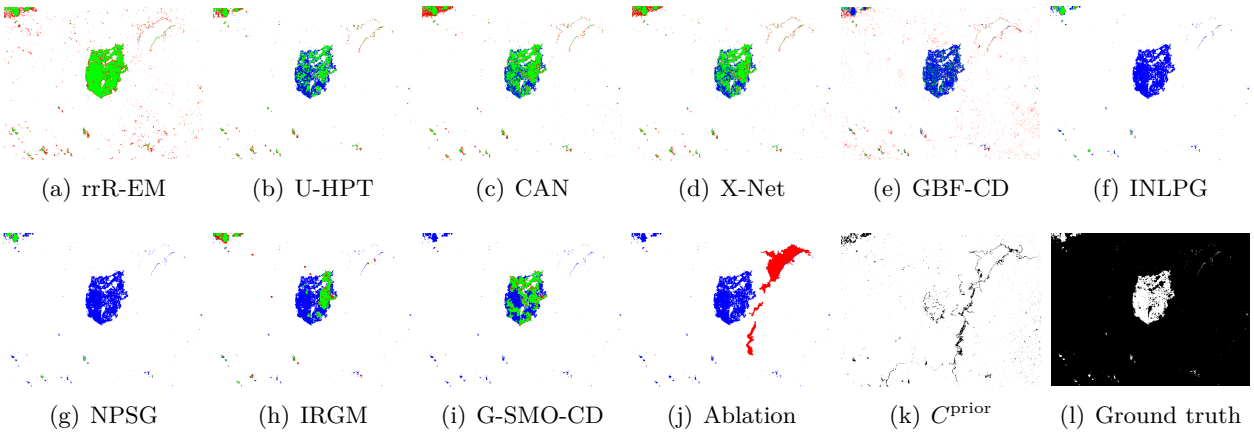
Figure 38: Change detection maps highlighting the false negatives (FNs), false positives (FPs), and correct changed pixels (Cs). Each row corresponds to a dataset and each column to a method: KI, rR-EM, rrR-EM, U-CD-HPT, and GBF-CD.



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

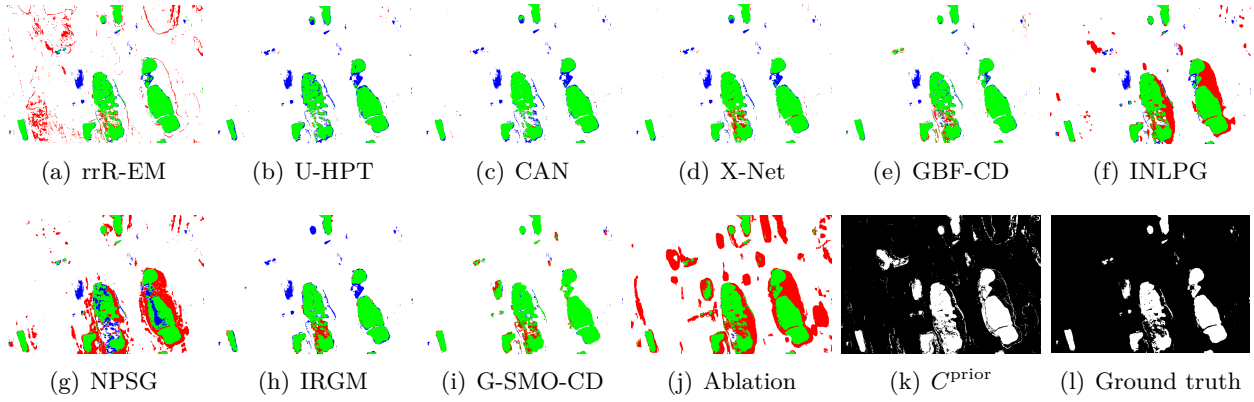
Figure 39: Change maps for Mulargia.



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

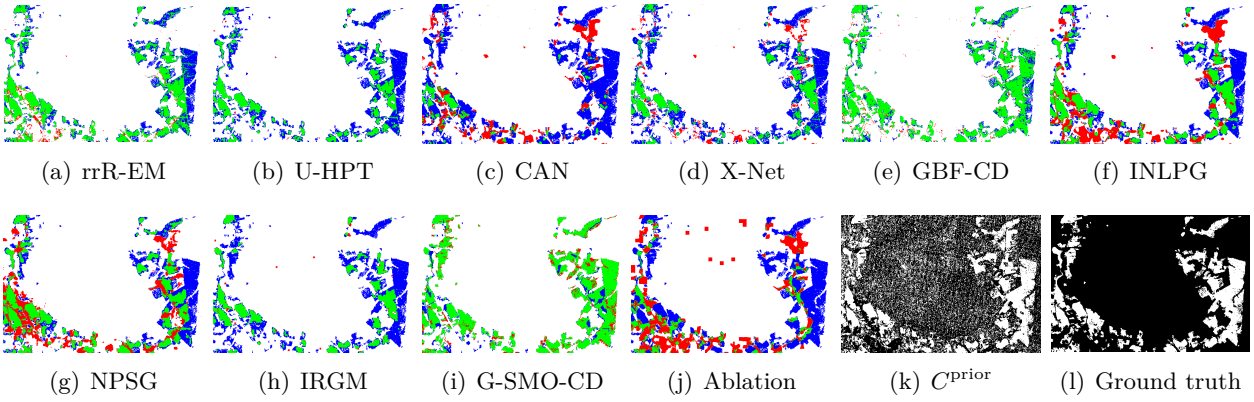
Figure 40: Change maps for Omodeo.



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

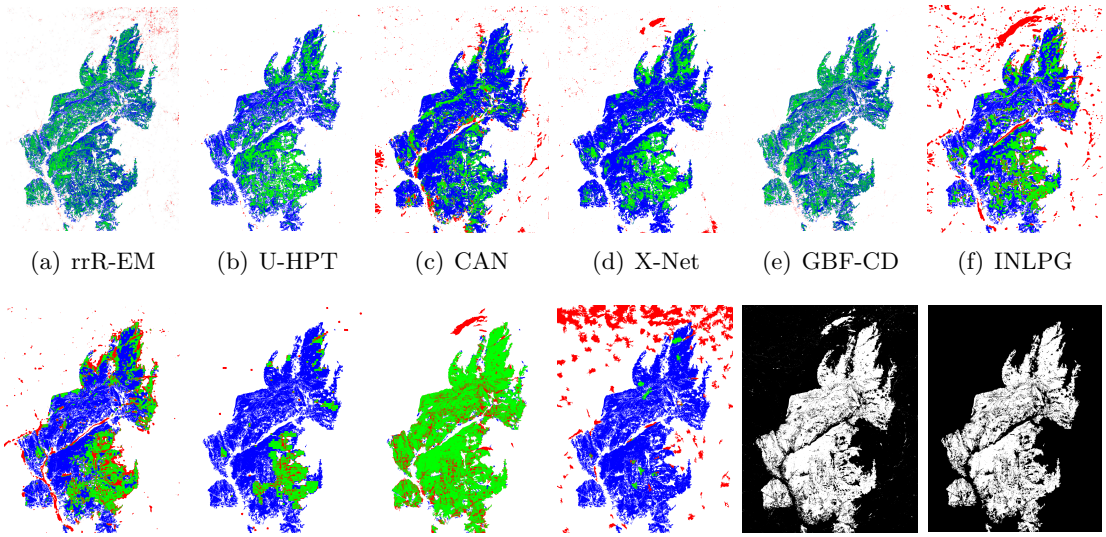
Figure 41: Change maps for Alaska.



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

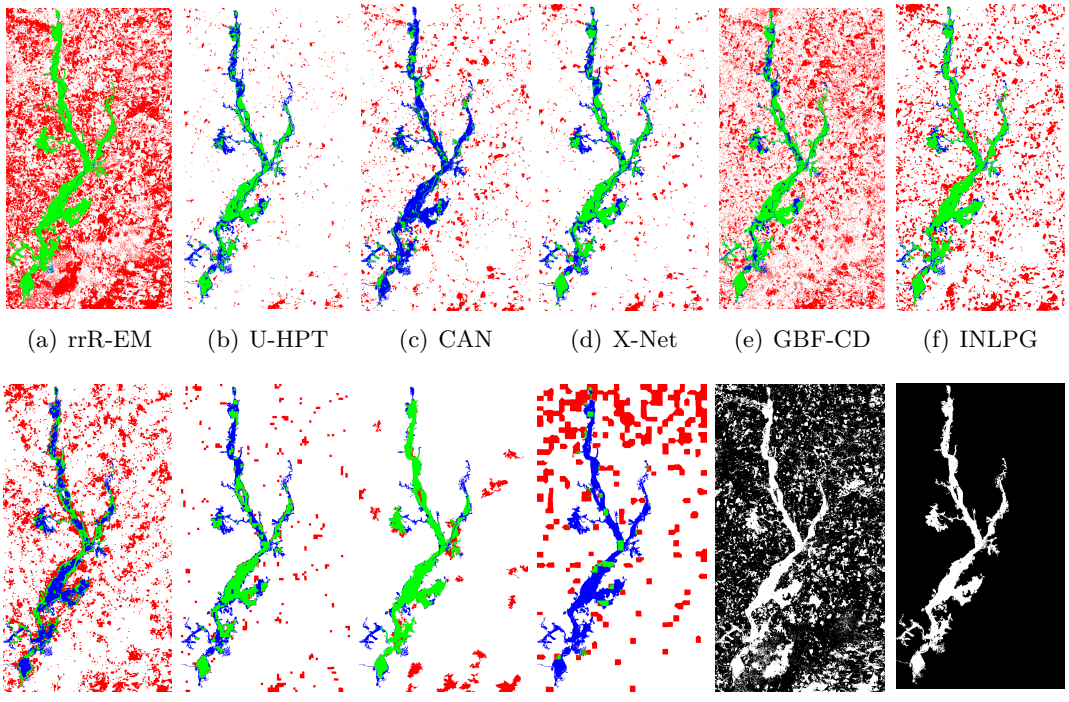
Figure 42: Change maps for Madeirinha.



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

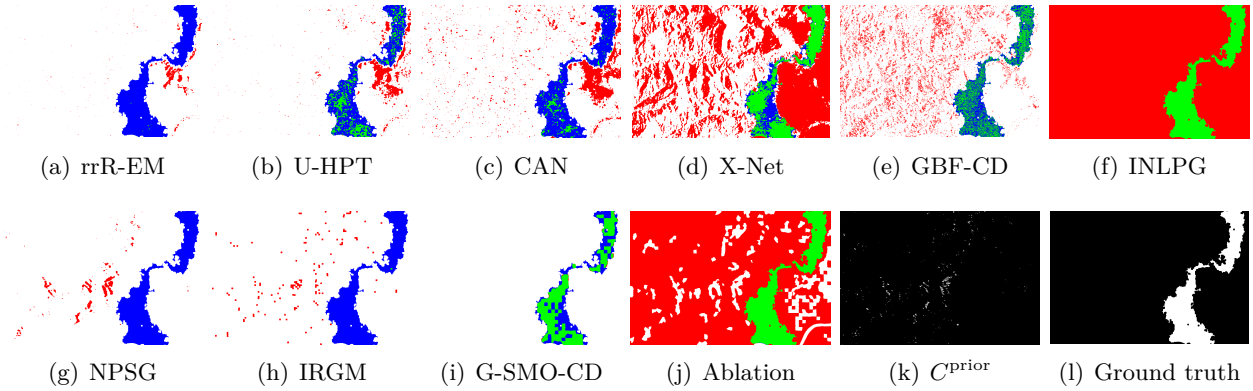
Figure 43: Change maps for Canada.



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

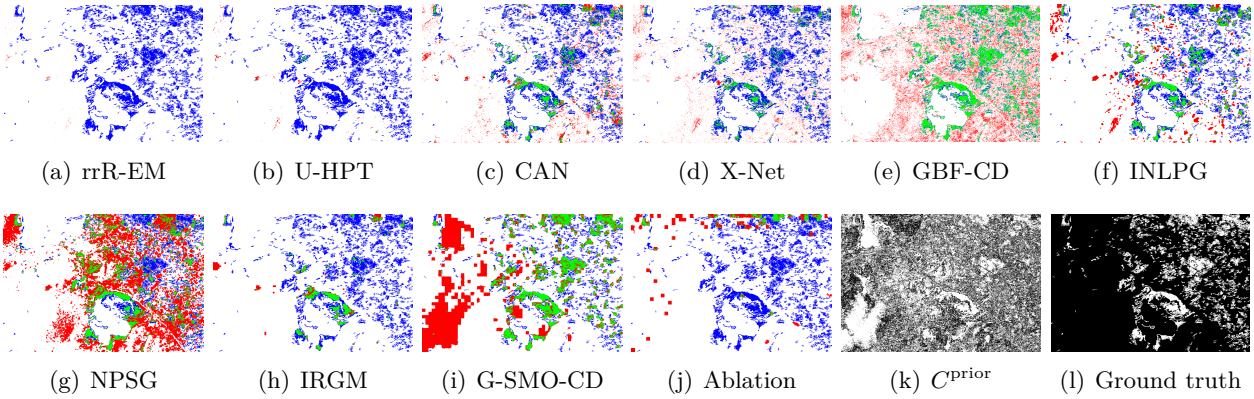
Figure 44: Change maps for Gloucester-1.



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

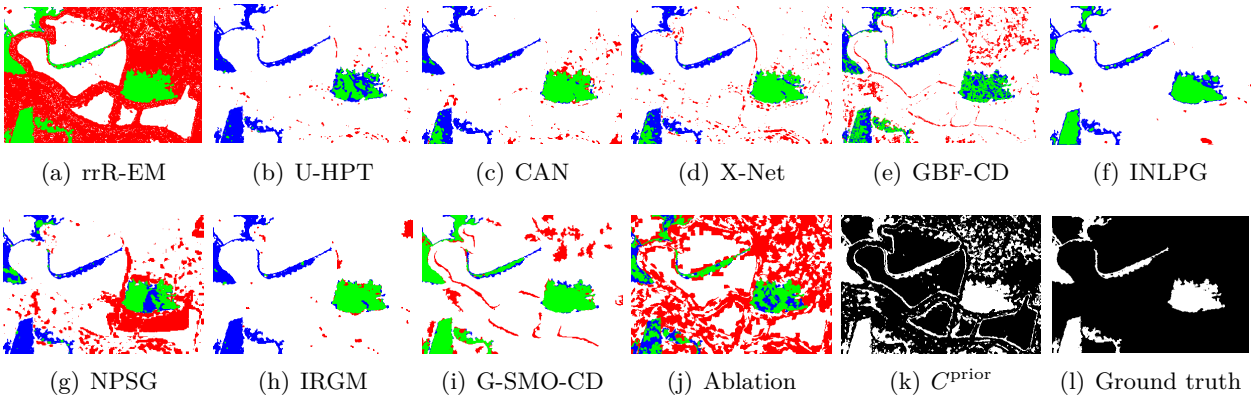
Figure 45: Change maps for Katios.



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

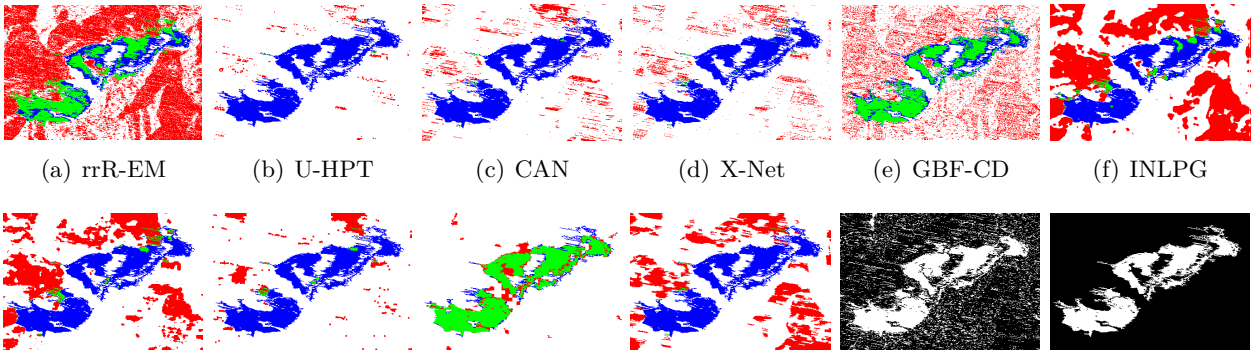
Figure 46: Change maps for Atlantico.



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

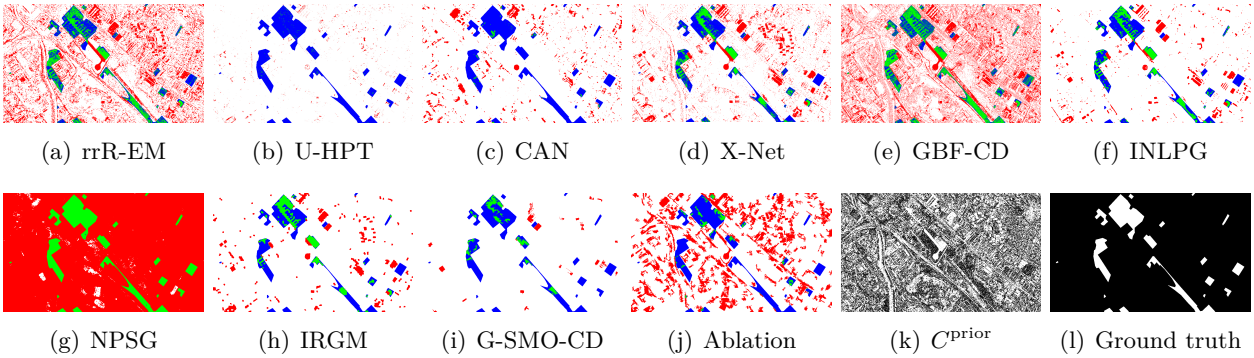
Figure 47: Change maps for San Francisco.



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

Figure 48: Change maps for Wenchuan.



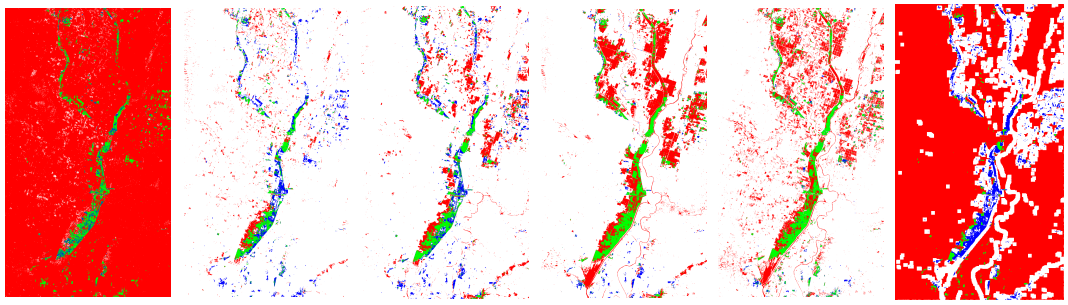
(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

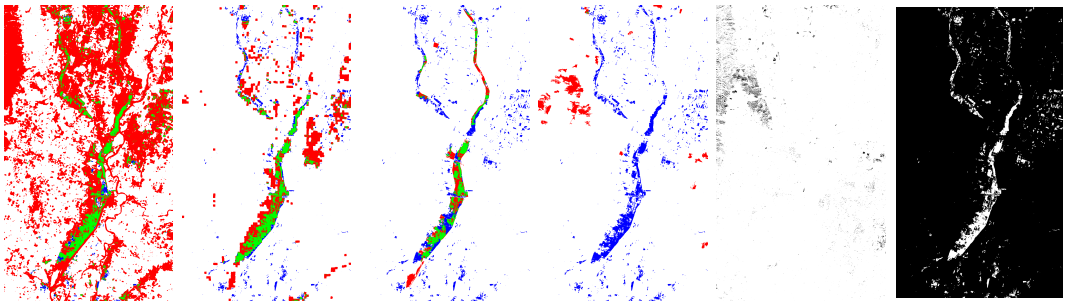
Figure 49: Change maps for Toulouse.

Table 35: Performance of superpixel algorithms in the proposed framework for CD in terms of $\boldsymbol{\kappa}$.

DATASETS	**GMMSP**	**SLIC**	**AMR**
Mulargia	**0.9165**	0.9114	0.3217
Omodeo	**0.6032**	0.0198	0.0815
Alaska	**0.9106**	0.7836	0.7699
Madeirinha	**0.8303**	0.8205	0.5821
Prince George	**0.8319**	0.8037	0.7665
Gloucester-1	**0.7033**	0.7029	0.3675
Katios	**0.7103**	−0.2012	0.2584
Atlantico	**0.2901**	0.1580	0.1935
San Francisco	**0.6947**	0.6262	0.6847
Wenchuan	**0.7991**	0.7689	0.0623
Toulouse	0.2287	−0.0243	**0.2515**
California	**0.3850**	0.1277	0.1785
Bastrop	**0.9566**	0.9397	0.9141
Gloucester-2	0.1685	0.1996	**0.5223**



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

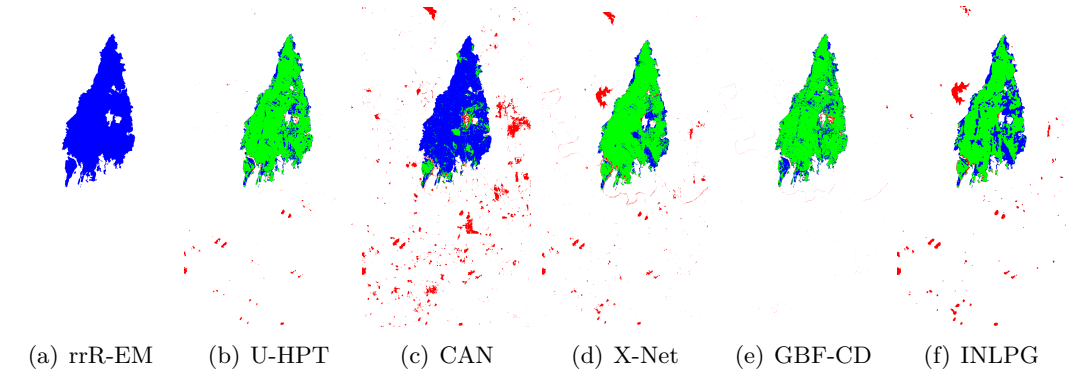


(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

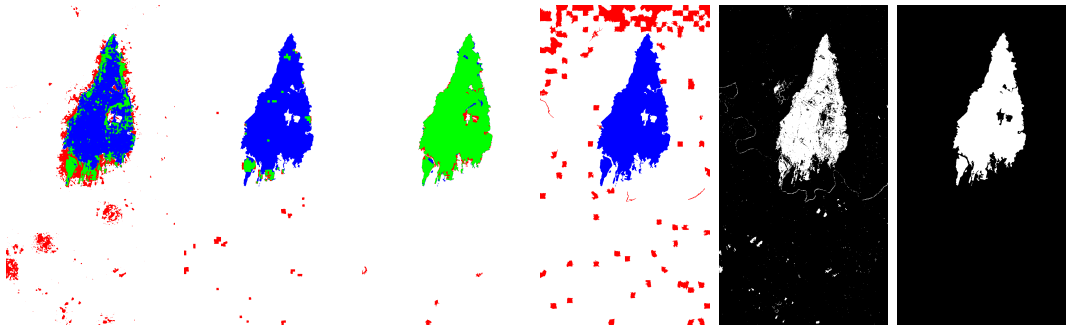
Figure 50: Change maps for California.

Table 36: Models performance in terms of FN (units in %).

Dataset	**rrR-EM**	**U-HPT**	**CAN**	**X-Net**	**GBF-CD**	**INLPG**	**NPSG**	**IRGM**	**G-SMO-CD**
Mulargia	10.14	9.03	82.48	8.91	12.33	13.22	**3.39**	15.30	4.64
Omodeo	**0.01**	45.88	26.71	29.06	77.00	94.05	95.60	69.35	51.49
Alaska	12.52	22.01	18.91	14.80	11.66	10.70	22.37	18.45	**3.72**
Madeirinha	40.31	61.05	87.18	66.09	24.44	53.26	48.79	65.39	**13.88**
Prince George	54.01	61.23	73.26	71.79	54.10	65.31	66.65	82.77	**3.13**
Gloucester-1	**2.35**	44.60	76.85	43.09	23.80	21.80	59.10	47.85	21.57
Katios	99.79	73.00	88.45	28.40	52.05	**0**	100	99.89	41.68
Atlantico	99.93	99.13	78.13	82.9	**30.42**	71.79	51.97	87.87	44.64
San Franciso	**2.19**	75.81	57.37	57.72	45.82	43.41	63.72	63.72	11.94
Wenchuan	41.61	99.69	99.27	98.43	35.82	89.86	94.22	97.57	**13.46**
Toulouse	52.32	98.30	97.19	66.95	54.27	65.96	**0.07**	70.41	83.35
California	18.01	58.21	50.69	18.18	**11.93**	77.74	12.74	43.19	65.88
Bastrop	100.00	15.50	86.41	15.36	16.83	29.01	73.22	94.57	**2.17**
Gloucester-2	44.12	98.36	87.83	64.81	29.39	30.67	30.61	80.94	**16.80**



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG

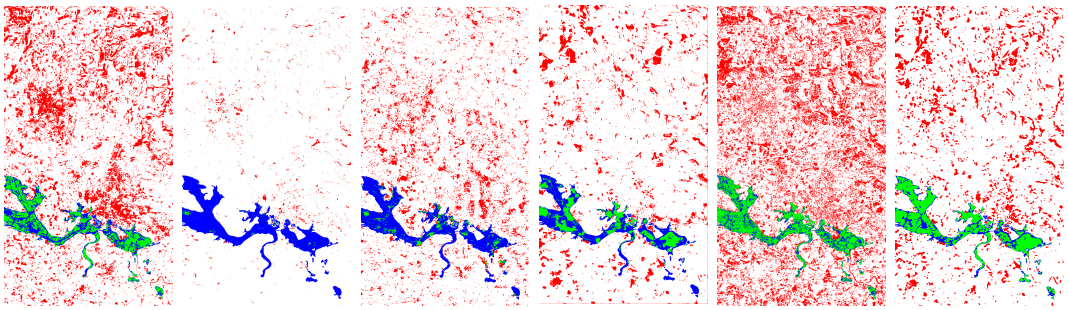


(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

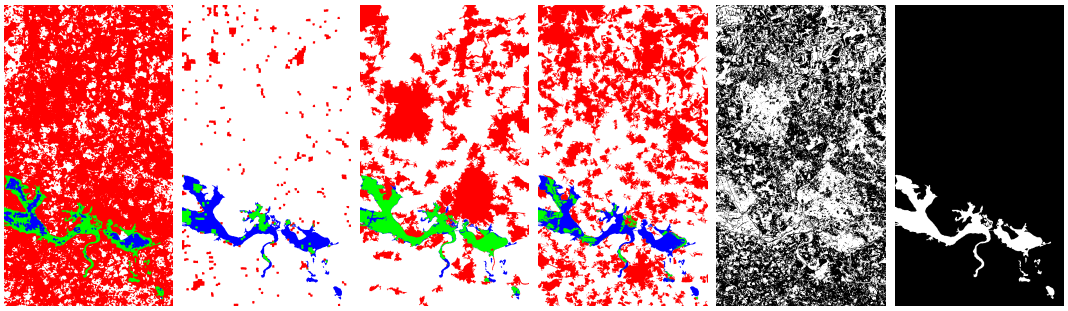
Figure 51: Change maps for Bastrop.

Table 37: Models performance in terms of FP (units in %).

Dataset	**rrR-EM**	**U-HPT**	**CAN**	**X-Net**	**GBF-CD**	**INLPG**	**NPSG**	**IRGM**	**G-SMO-CD**
Mulargia	1.06	2.00	2.68	1.56	**0.17**	0.22	2.97	0.46	0.37
Omodeo	2.14	0.55	1.12	1.01	1.26	**0.04**	0.03	0.48	0.43
Alaska	4.81	**0.15**	0.16	0.80	0.87	6.44	10.04	1.09	1.54
Madeirinha	1.32	**0.11**	6.17	2.56	1.13	8.26	8.60	0.49	3.20
Prince George	1.13	**0.20**	5.10	1.43	0.38	9.30	8.20	1.88	8.67
Gloucester-1	44.06	**2.41**	7.77	4.40	26.57	16.03	19.47	3.10	3.41
Katios	1.29	3.58	7.19	44.91	10.63	100	1.51	1.55	**0.095**
Atlantico	**0.08**	0.28	4.53	3.75	13.69	5.38	28.32	1.08	19.18
San Franciso	61.23	1.52	2.74	4.02	7.64	**0.49**	18.12	1.69	7.21
Wenchuan	53.95	**2.06**	6.80	6.77	22.52	40.16	25.74	5.78	4.32
Toulouse	22.07	**0.97**	5.33	9.40	17.33	4.62	98.14	4.69	1.12
California	97.85	2.79	5.67	11.52	11.79	73.53	37.72	6.92	**1.53**
Bastrop	**0.00**	0.39	3.56	1.44	0.23	1.18	4.12	0.62	0.67
Gloucester-2	19.72	**1.57**	10.32	9.00	27.71	11.96	80.50	4.22	31.55



(a) rrR-EM (b) U-HPT (c) CAN (d) X-Net (e) GBF-CD (f) INLPG



(g) NPSG (h) IRGM (i) G-SMO-CD (j) Ablation (k) C^{prior} (l) Ground truth

Figure 52: Change maps for Gloucester-2.

Table 38: Models performance in terms of P (units in %).

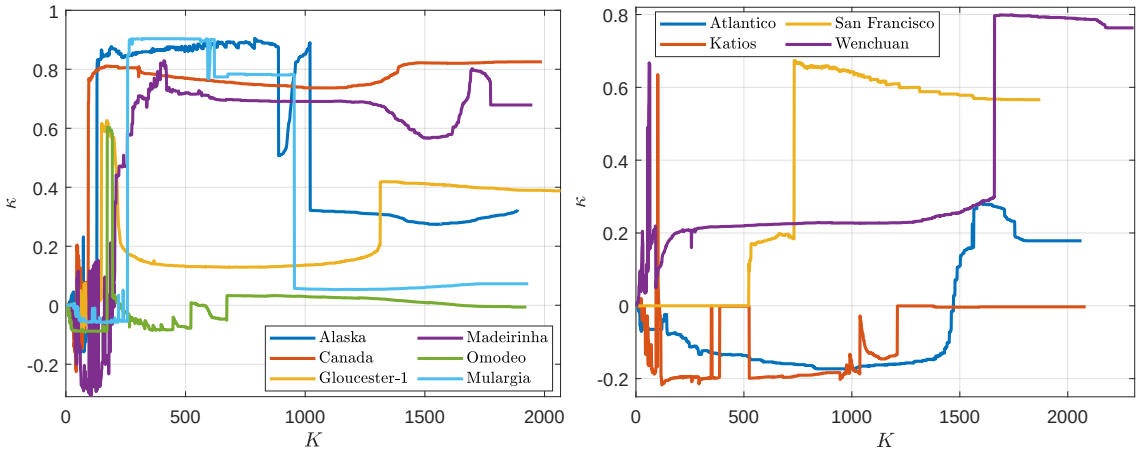
Dataset	**rrR-EM**	**U-HPT**	**CAN**	**X-Net**	**GBF-CD**	**INLPG**	**NPSG**	**IRGM**	**G-SMO-CD**
Mulargia	89.86	90.96	17.51	91.08	87.67	86.77	**96.60**	84.69	95.35
Omodeo	1.00	54.11	**73.28**	70.94	22.99	5.94	4.39	30.64	48.50
Alaska	87.48	77.98	81.09	85.19	88.34	89.29	77.62	81.55	**96.27**
Madeirinha	59.69	38.94	12.81	33.90	75.56	46.73	51.20	34.60	**86.12**
Prince George	45.99	38.76	26.73	28.20	45.90	34.68	33.34	17.22	**96.86**
Gloucester-1	**97.65**	55.39	23.15	56.90	76.19	78.19	40.89	52.14	78.42
Katios	0.21	26.99	11.54	71.59	47.95	**100**	0	10.23	58.31
Atlantico	0.06	0.86	21.86	17.10	**69.57**	28.20	48.02	12.13	55.35
San Franciso	**97.80**	24.19	42.63	42.27	51.17	56.59	36.28	36.27	88.05
Wenchuan	58.39	0.30	0.72	1.56	64.17	10.13	5.77	2.42	**86.53**
Toulouse	47.67	1.69	2.80	33.04	45.72	34.03	**99.92**	29.58	16.64
California	81.98	41.79	49.30	81.81	**88.06**	22.25	87.25	56.81	34.11
Bastrop	0.00	84.49	13.58	84.63	83.16	70.98	26.77	5.42	**97.82**
Gloucester-2	55.87	6.63	12.16	35.18	70.60	69.32	69.38	19.05	**83.19**

Table 39: Models performance in terms of R (units in %).

Dataset	**rrR-EM**	**U-HPT**	**CAN**	**X-Net**	**GBF-CD**	**INLPG**	**NPSG**	**IRGM**	**G-SMO-CD**
Mulargia	72.04	58.12	16.57	63.90	**93.96**	92.13	49.78	84.84	88.69
Omodeo	69.73	82.90	76.33	77.57	47.26	**86.26**	86.15	75.82	84.71
Alaska	68.51	**98.38**	98.30	92.70	92.36	62.39	48.08	89.89	88.20
Madeirinha	91.45	**98.78**	33.03	75.83	94.06	57.31	58.56	94.31	86.44
Prince George	93.93	**98.61**	70.16	89.81	97.86	58.83	60.89	77.75	81.06
Gloucester-1	17.27	68.31	22.67	55.97	27.26	31.47	16.50	61.26	**68.37**
Katios	1.85	47.03	15.93	15.83	34.74	10.55	0	76.03	**98.63**
Atlantico	11.62	36.01	46.80	45.39	48.11	48.88	23.63	**67.18**	34.50
San Franciso	18.85	69.62	69.31	60.40	49.44	94.35	22.54	**75.64**	63.76
Wenchuan	18.40	3.00	2.18	4.60	37.25	4.99	4.46	8.04	**80.66**
Toulouse	15.74	13.11	4.46	23.80	18.57	38.87	8.09	35.28	**53.13**
California	3.69	40.59	31.16	27.01	25.44	1.36	9.55	27.26	**50.33**
Bastrop	0.00	96.17	31.23	87.43	**97.71**	87.65	43.56	50.81	94.49
Gloucester-2	16.26	1.63	7.60	21.43	14.86	**28.43**	5.57	23.60	15.30

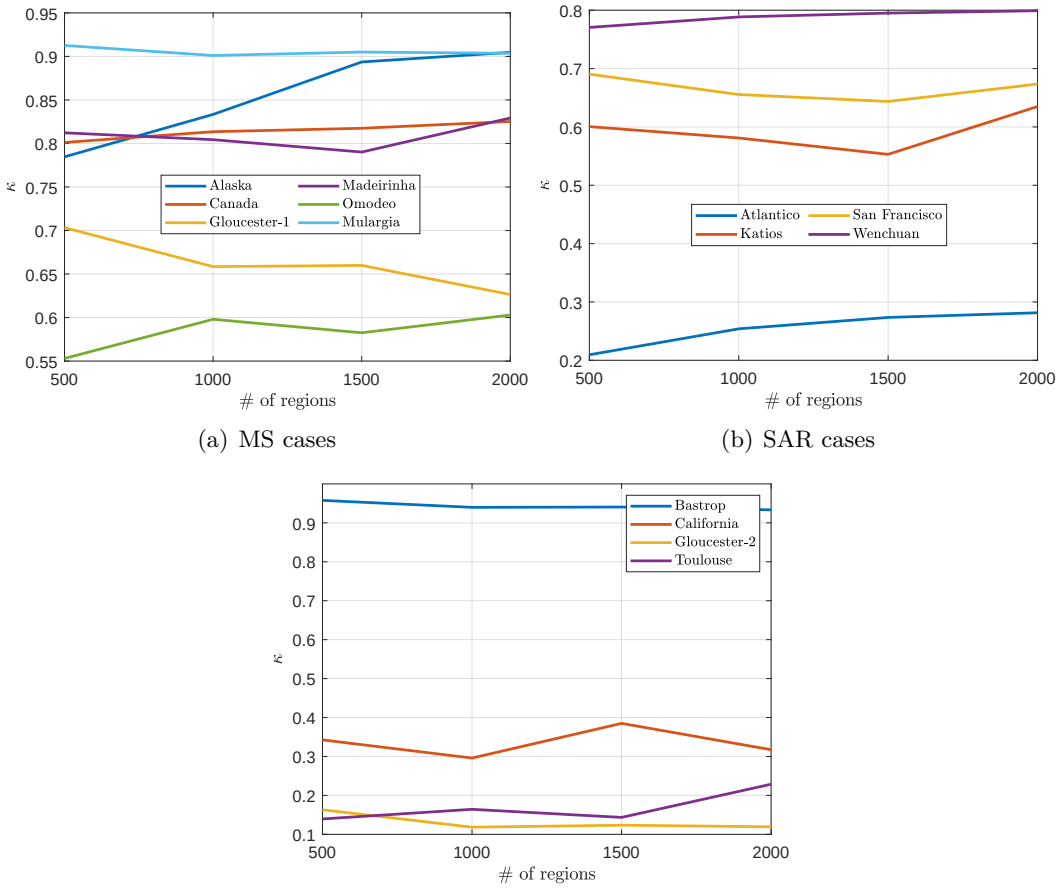
Table 40: Number of the training samples of the learning-based methods.

DATASETS	**U-HPT**	**X-Net**
Mulargia	10000	26880
Omodeo	10000	64193
Alaska	10000	95265
Madeirinha	10000	131557
Prince George	10000	480812
Gloucester-1	100000	539684
Katios	10000	618451
Atlantico	10000	525540
San Francisco	10000	72905
Wenchuan	10000	69286
Toulouse	100000	2194380
California	100000	1572357
Bastrop	100000	451108
Gloucester-2	100000	2027475



(a) MS cases (b) SAR cases

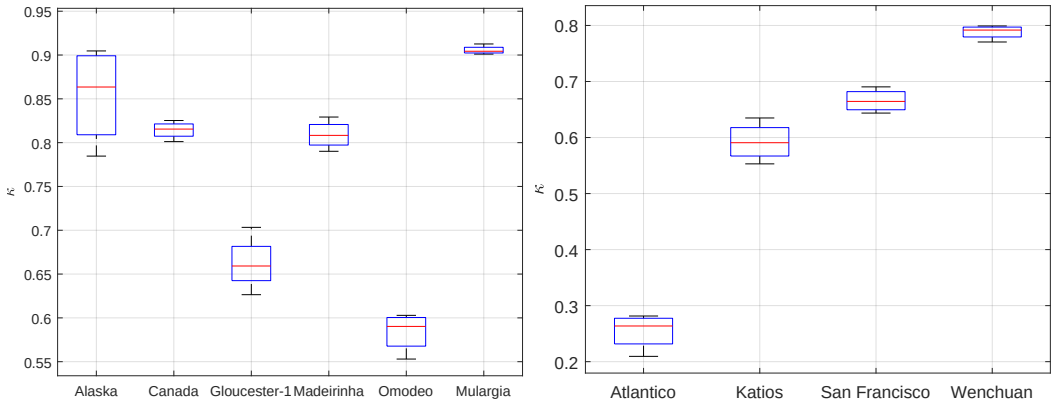
Figure 53: Parameter sensitivity analysis in the graph learning process for MS and SAR homogeneous cases.



(a) MS cases (b) SAR cases

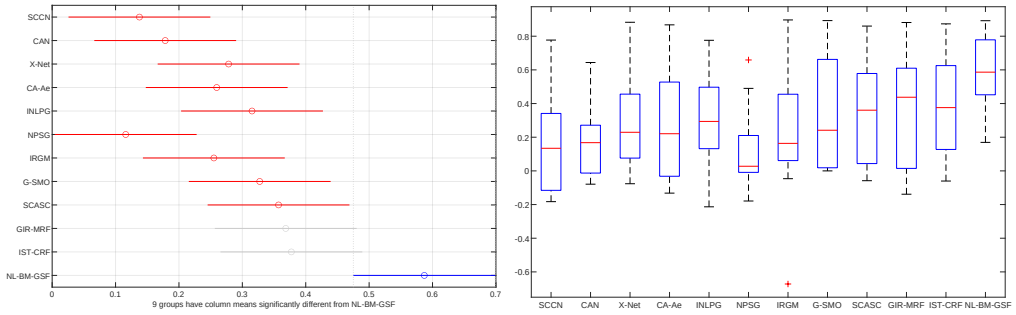
(c) Heterogenous cases

Figure 54: Region sensitivity analysis for MS and SAR homogeneous, and heterogeneous cases.



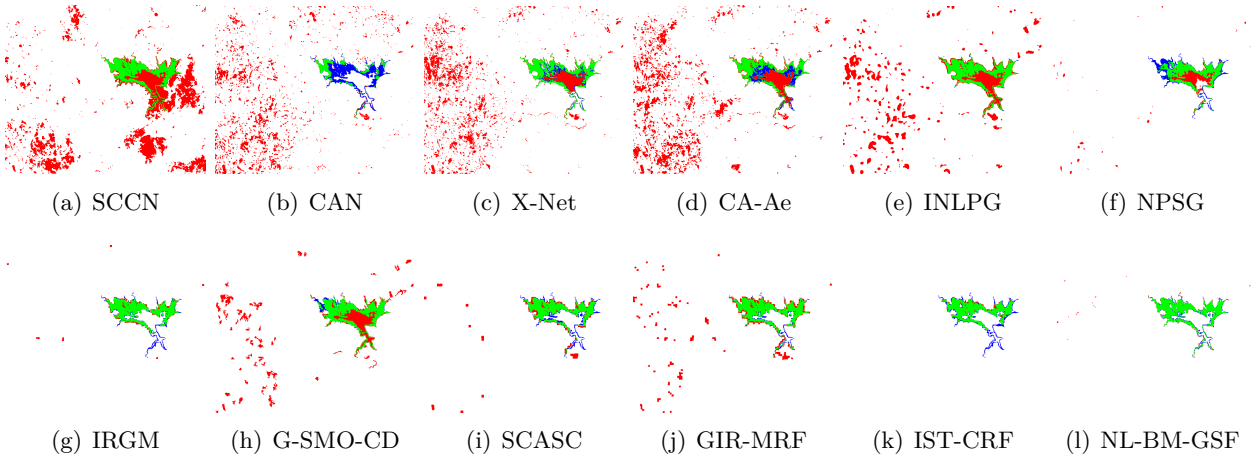
(a) MS cases (b) SAR cases

Figure 55: Region sensitivity analysis box-plot for MS and SAR homogeneous cases.



(a) Two way ANOVA (b) Box-Plot

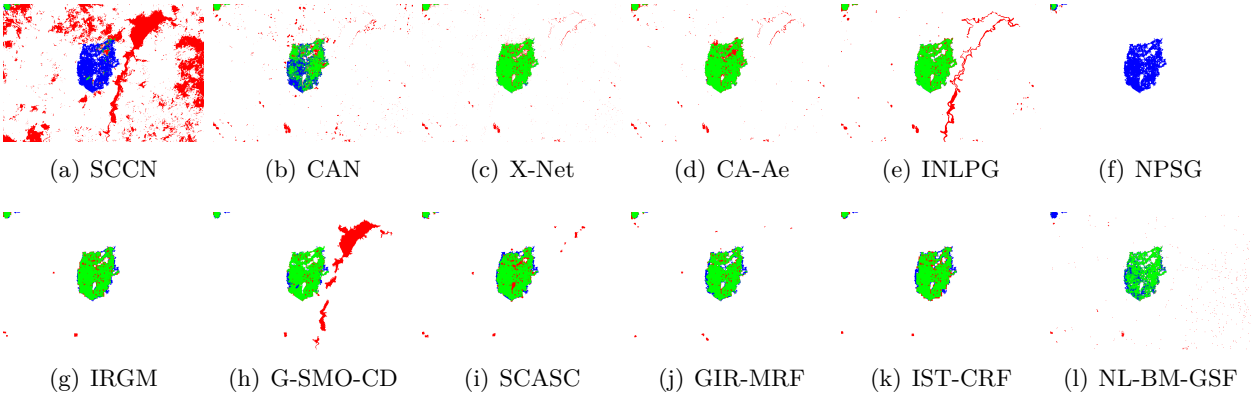
Figure 56: In this figure, we show the difference with respect to the metric κ (see table 29) for all the approachs tested. Wherem (a) is the Two-way analysis of variance that shows the significance difference between the models, and (b) is the box-plots of the models.



(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

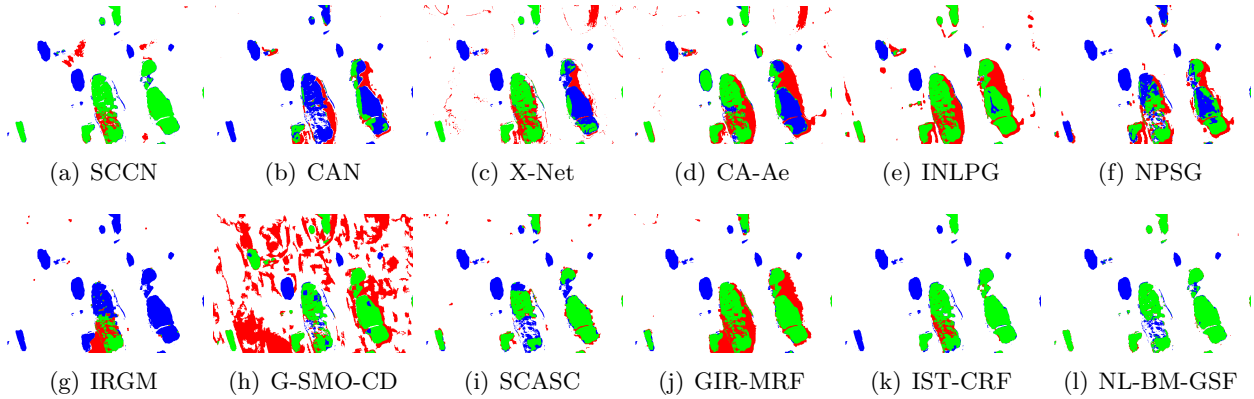
Figure 57: Change maps for Mulargia.



(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

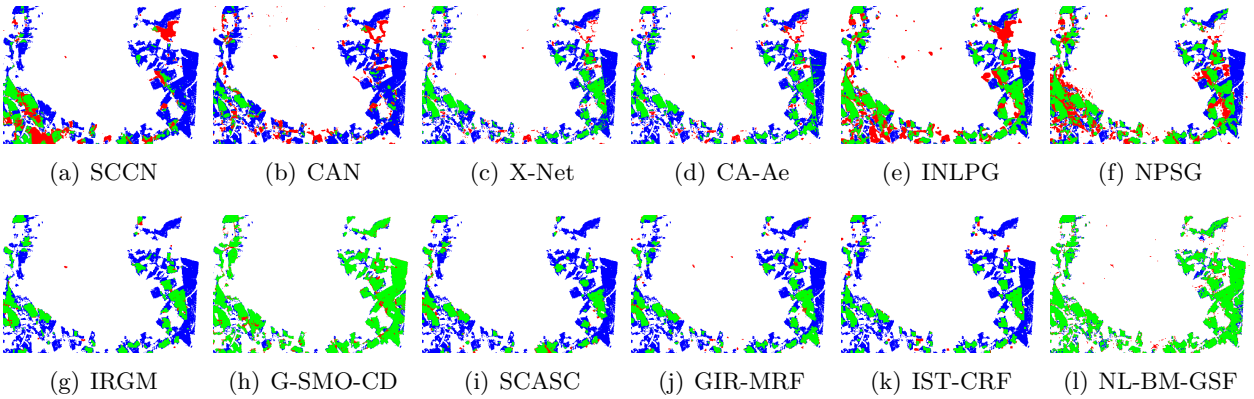
Figure 58: Change maps for Omodeo.



(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

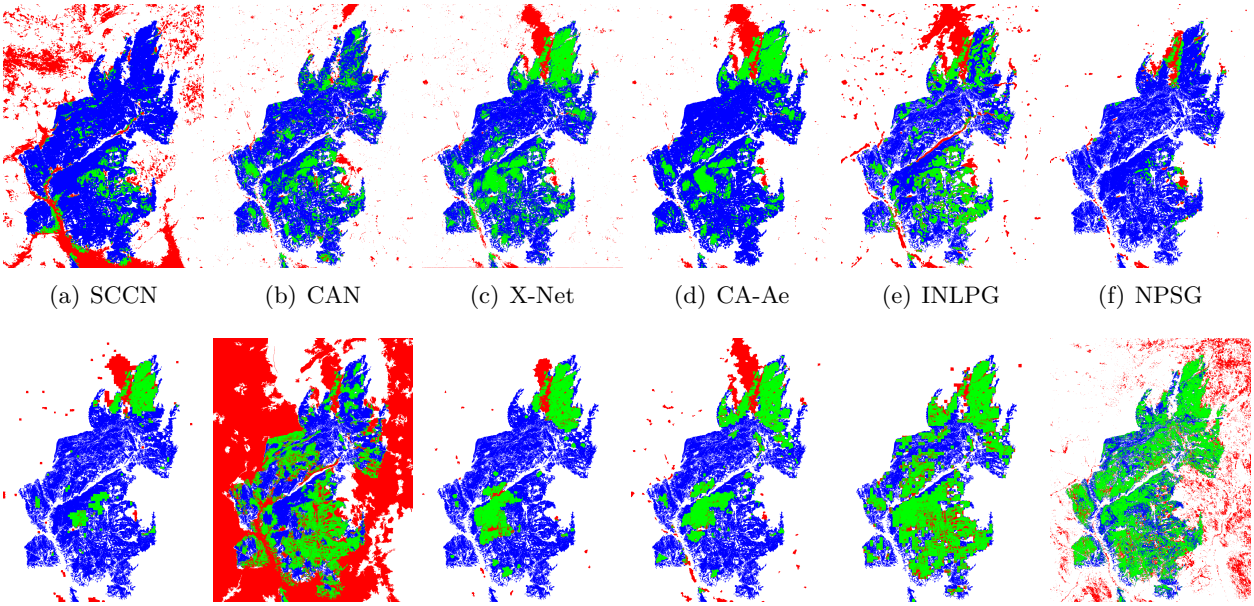
Figure 59: Change maps for Alaska.



(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

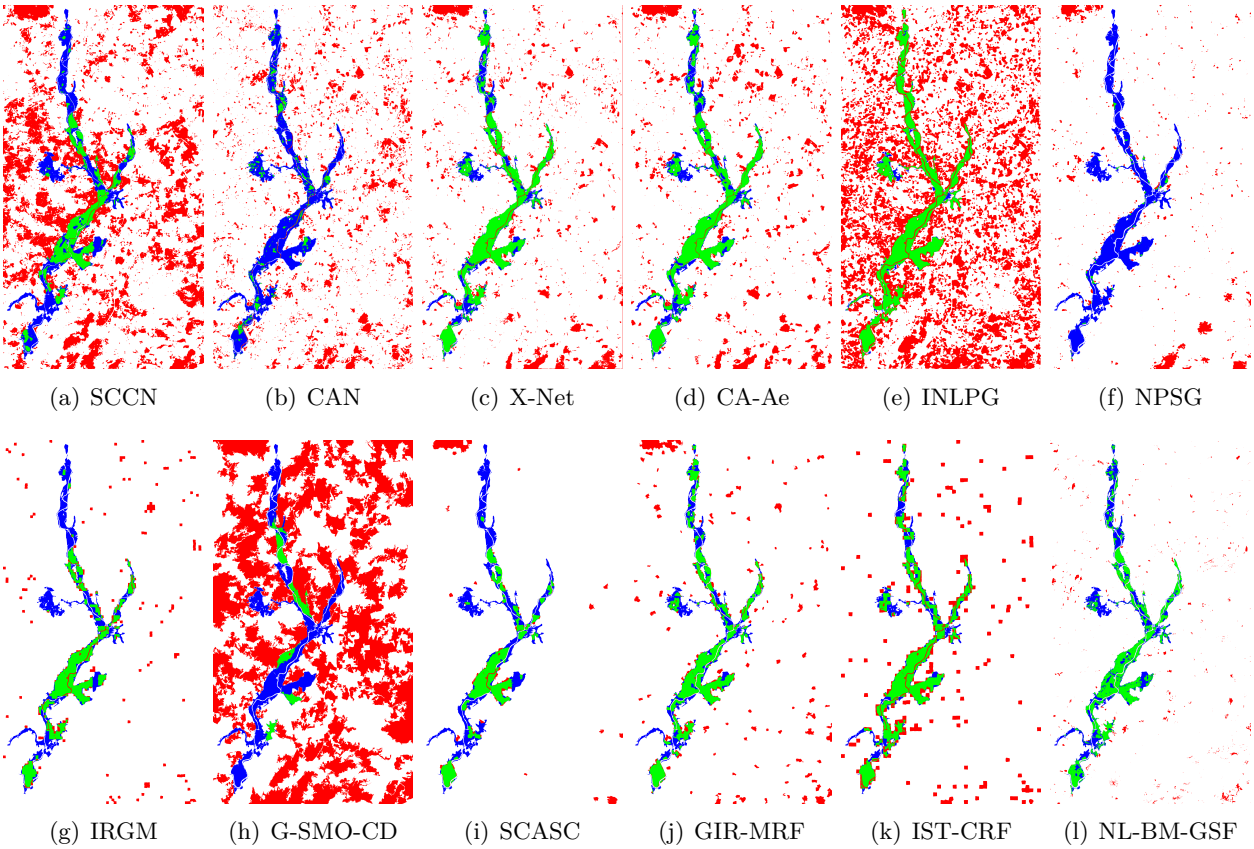
Figure 60: Change maps for Madeirinha.



(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

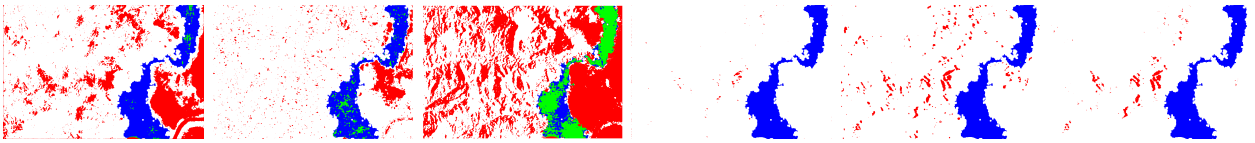
Figure 61: Change maps for Canada.



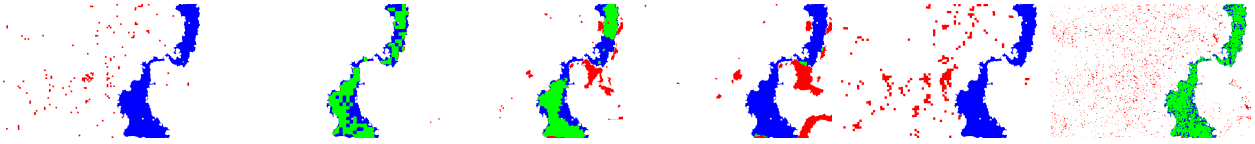
(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 62: Change maps for Gloucester-1.

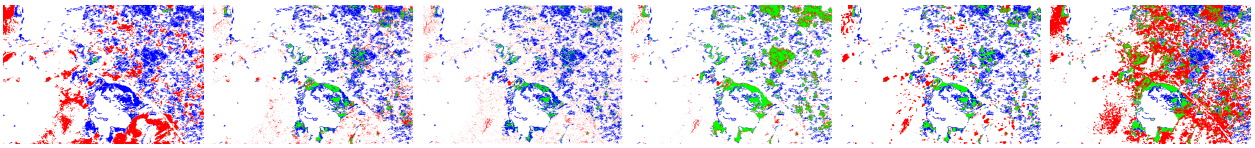


(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

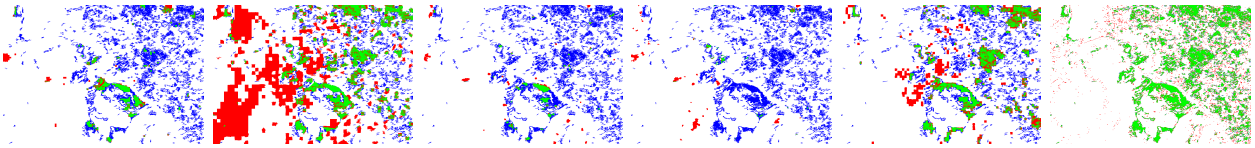


(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 63: Change maps for Katios.

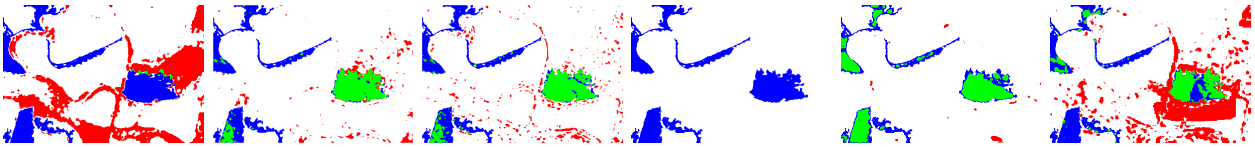


(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

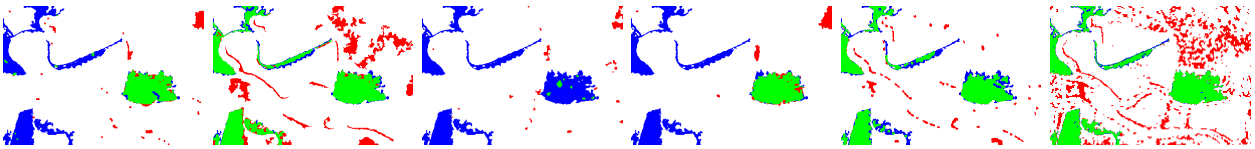


(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 64: Change maps for Atlantico.

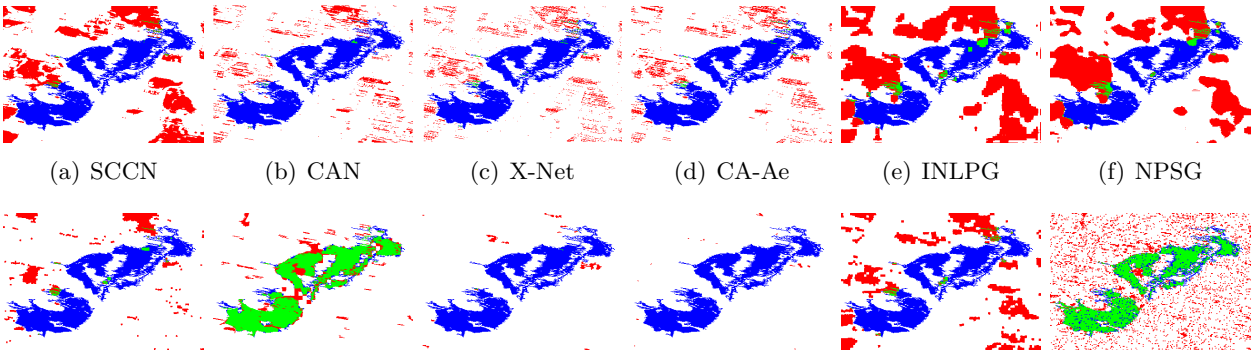


(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG



(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 65: Change maps for San Francisco.



(a) SCCN (b) CAN (c) X-Net (d) CA-Ae (e) INLPG (f) NPSG

(g) IRGM (h) G-SMO-CD (i) SCASC (j) GIR-MRF (k) IST-CRF (l) NL-BM-GSF

Figure 66: Change maps for Wenchuan.