Using Environmental Context to Synthesis Missing Pixels

Thaer F. Ali, Alan Woodley
School of Computer Science, Science and Engeering Faculty, Queensland Univeristy of Technology
ARC Centre of Excellence for Mathematical & Statistical Frontiers
Brisbane, Australia
{thaerfaragali.ali, a.woodley}@hdr.qut.edu.au

***Abstract*— Satellites have proven to be a technology that can help in a variety of environmental and human development contexts. However, at times some pixels in the satellite images are not captured. These uncaptured pixels are called missing pixels. Having these missing pixels means that important data for research and satellite imagery-based applications is lost. Therefore, people have developed pixel synthesis methods. This paper presents a new pixel synthesis method called the Iterative Self-Organizing Data Analysis Techniques Algorithm – Integration of Geostatistical and Temporal Missing Pixels' Properties (ISODATA-IGTMPP). The method is built upon the Integration of Geostatistical and Temporal Missing Pixels' Properties (IGTMPP) method and adds a seminal clustering technique called the Iterative Self-Organizing Data Analysis Techniques Algorithm (ISODATA). The clustering technique allows a new way of predicting the missing pixel from their environmental class with benefit of the spatial and temporal properties. Here, the ISODATA-IGTMPP method was tested on the Spatial-Temporal Change in the Environment Context (STCEC) dataset and was compared with results of four missing pixel predicting methods. The method shows the best performing results and preforms very well across different environment types.**

Keywords*— *Pixel Synthesis, Clustering, ISODATA-IGTMPP, ISODATA, IGTMPP

I. Introduction

The accuracy of satellites images is important since they provide environmental information to remote sensing-based applications [1]. Satellites sense the surface of the Earth and interpret the observation into images. However, sometimes satellites do not capture every pixel correctly due to atmospheric or electronic reasons [2]. In the literature these pixels are called missing pixels. Literature surrounding missing pixels can roughly be categorised into five classes: geostatistical methods, deterministic methods, time-series curve based methods, auxiliary image-based methods and hybrid methods.

Overall, hybrid methods have shown the best results [3]. However, each of the presented methods has limitations. This study demonstrates a new method with more emphases on the predicting the missing pixel from its environmental class and combines it with its spatial-temporal characters. The new algorithm is called the Iterative Self-Organizing Data Analysis Techniques Algorithm – Integration of Geostatistical and Temporal Missing Pixels' Properties (ISODATA-IGTMPP).

The algorithm is made up of two methods. First, is the ISODATA clustering method [4], a well-established clustering method. Here, it is used as pre-processing step to determine the environmental class of a pixel. The environmental class is then used in the second method, the IGTMPP method [5], which uses a temporal and spatial properties to predict missing pixels. Using the environmental class is a key advantage over the IGTMPP method, but its advantages would likely be repeated across similar methods. This work explores this and provides a deeper analysis of the previously published IGTMPP method

The Spatial-Temporal Change in the Environment Context (STCEC) [6] dataset was used to test the method. The root mean square deviation (RMSD) was used as the metric to compare the ISODATA-IGTMPP with the IGTMPP, and four baselines. A two-sample, unequal variance t-test was used. The results show that the ISODATA-IGTMPP method was the most accurate.

The rest of the paper is organized as follows. Section 2 presents a literature review of other clustering methods and the use of clustering in remote sensing. Section 3 outlines the ISODATA-IGTMPP method, presenting its underlying theory. Section 4 presents the results and compares them with previous algorithms.

II. Literature Review

A. Pixel Synthesis Methods

Earth observation satellites have been launched for decades, with the size of their image archives continuously increasing. However, sometimes images contain incorrectly captured spots or strips, which are called missing pixels, due to atmospheric or technical conditions. These missing pixels negatively affect the accuracy of the captured images [7] resulting in negative downstream effects, for example errors in ecological applications [8] and land-cover mapping [9]. Methods have been developed for predicting missing pixels but are still at insufficient level of accuracy. The missing pixel predicting methods are categorized into five classes which are geostatistical, deterministic, time-series curve based, axillary-based image, and hybrid methods.

Geostatistical methods apply statistics on the spatial information assuming that locations are dependent in a specific range [7], and, that changes in one location affect others. These methods predict missing pixels by measuring the relationship between pixels and their pattern of change [10]. The core of geostatistical missing pixel predicting method are built based on probability principles [11]; and often include a kriging model [12].

The second class of missing pixel predicting methods are deterministic methods. This class uses empirical mathematical relations for measuring the continuity of the missing pixel with other pixels. They can analyze their result through spatial or temporal properties [13].

Auxiliary image based methods are the third class of methods which merge together images from different sensors.

For example, merging together Landsat and MODSIS or merging together satellite images from different counties [14].

The fourth class of methods is time-series curve methods. These methods accept a large number of images and arrange them inside a data cube. There has been a large number of papers in recent years due to the widespread availability of a dense timeseries from government owned satellites [15].

The final class are hybrid methods. These combine one or more of the previous methods in a new method. The advantage of hybrid methods is that they are able to build upon the weakness of any single method [9].

B. Clustering Methods

Clustering is the process of grouping similar pixels together like objects which in the context of remote sensing this means similar pixels. Using the taxonomy of Fahad et al. [16], clustering techniques can be classified into partitioning-based algorithms, hierarchical algorithms, density-based algorithms, grid-based algorithms and model-based algorithms, all of which have been tested on datasets [17]. The majority of computational algorithms applied to remote sensing have been supervised methods, however, a number of clustering approaches have also been applied. The advantage of clustering compared to supervised approaches is that it does not require *a priori* knowledge. Approaches to use clustering have included: k-means [18, 19], k-means++ [20], k-tree [17, 21, 22] and ISODATA [23-25].

C. Integration of Geostatistic and Temporal Missing Pixels' Properties Method

Recently, the Integration of Geostatistical and Temporal Missing Pixels' Properties (IGTMPP) method was been developed for predicting pixels when the environment changes [5]. The IGTMPP method used the concept of estimating a proxy pixel (called the Optimal Pixel) that had similar spatio-temporal properties to the missing pixel (Missing Pixel). However, analysis showed an unsolved important issue, that is, predicting the right location to choose the proxy pixel. This step is vital since it is relied upon by the other steps in the IGTMPP method, and it relies on the Optimal Pixel"?

This is particularly important when the Optimal Pixel is selected from a different environmental class than the Missing Pixel. This can occur when an Optimal Pixel has a very similar spatio-temporal (band) value to the Missing Pixel, but in reality is an outliner. It has found that this was particularly prevalent when looking at heterogeneous examples that change over time. This paper presented a solution based upon clustering. It is based upon the theory that pixels in the same cluster will change in same way, for example, changing from green vegetation to brown vegetation.

The clustering algorithm used in the proposed method is called the Iterative Self Organising Data Analysis Technique Algorithm (ISODATA). ISODATA has been used previously in remote sensing, and it was used here for two main reasons:

1. ISODATA is an unsupervised clustering algorithm, which is useful since labelled data does not exist for the data and is a time-consuming task to generate manually.
2. ISODATA does not require the user to know the number of end clusters, which is useful when trying to cluster an unseen area.

The new method is called Iterative Self-Organizing Data Analysis Techniques Algorithm – Integration of Geostatistic and Temporal Missing Pixels' Properties (ISODATA-IGTMPP). The overall algorithm of ISODATA-IGTMPP is presented in Figure 1 and is detailled in the next section.

Algorithm 1 ISODATA-IGTMPP-methoed

```
procedure MAIN(Image1, Image2, Image3)
    Image1 ← READ-IMAGE(firstDateImage)
    procedure ISODATA(ISODATA-Algorithm)
        Clusters[ ] ← ISODATA-CODE(Image1, Thresholds)
    procedure CREATING-MAP-CODES(Clusters[], clusersNum)
        for I < numberClusters do
            while J < PixelNum do
            Map-Code[X][Y] ← CLUSTER[I][J].CODE()
    procedure WEIGHTING-SIMILARITIES(Image1, J , K)
        I = 0
        X ← J - 2
        Y ← K - 2
        for X → (J + 2)  do
            for Y → (K + 2) do
                localPixel ← PIXEL-LOCATION(Image1, X, Y)
                RMSD ← RMSD-FUNCTION(firstPixel, localPixel(X,Y))
                D ← DISTANCE(firstPixel, localPixel(X,Y))
                Similarities[I] ← RMSD * D
                I ← I + 1
        return Similarities
    procedure PREDICTING(Image2, Similarities[ ], firstPixel, k , J)
        code, code1
        for  Similarities do
            if (Similarities[ I ] > Similarities[ I + 1]) then
                code ← MAP-CODE(firsPixel, K, J)
                code1 ← MAP-CODE(optimalPixel, X, Y)
                optimalPixel ← Similarities[ I ]
        X ← XCOORDINATE(optimalPixel)
        Y ← YCOORDINATE(optimalPixle)
        secondPixel ← PIXEL-LOCATION(Image2, X, Y)
        fracChange ← FRACTION-CHANGE(optimalPixel, secondPixel)
        pmissingPixel ← firstPixel * fracChange
        return pmissingPixel , secondPixel, X , Y
    procedure ERROR-CORRECTION(Image3, pmissingPixel, secondPixel,
endPixel, X, Y)
        thirdPixel ← PIXEL-LOCATION(Image3, X, Y)
        fracChange ← FRACTION-CHANGE(secondPixel, thirdPixel)
        pendPixel ←  pmissingpixel * fracChange
        Diff ← endPixel - pendpixel
        if (Diff <> 0) then
            MissingPixel ←  pmissingPixel + (Diff / Frac-Change)
            return MissingPixel
        else
            return pmissingPixel
```

Fig. 1. The overall ISODATA-IGTMPP algorithm.

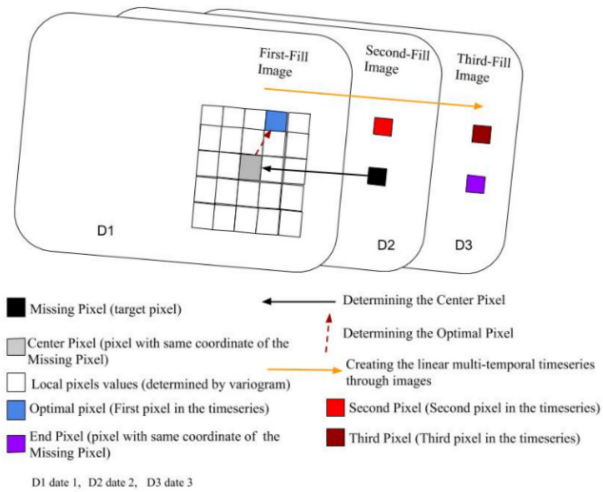

Fig. 2. The overall ISODATA-IGTMPP method. The figure outlines how the pixels are chosen from left to right.

III. THE ISODATA-IGTMPP METHOD

This section shows how the clustering, spatial, and temporal concepts are combined in a one algorithm called the ISODATA-IGTMPP method. Figure 2 presents the overall algorithm. The three images captured in different dates and are referred to as: the First-Fill image, the Second-Fill image, which is the image with missing pixels, and the Third-Fill image. The method consists of three main phases

1. Clustering phase
2. Predicting the missing pixel phase.
3. Adjusting the pixel deviation phase

A. *Clustering Phase*

The first phase of the ISODATA-IGTMPP method is to cluster pixels in the First-Fill image using the Iterative Self-Organizing Data algorithm (ISODATA). ISODATA produces a virtual raster that for each group of -pixel contains a cluster reference. Interpreted here to be a defined as an environmental reference [4]. A diagram explaining the clustering phase is presented in Figure 3.

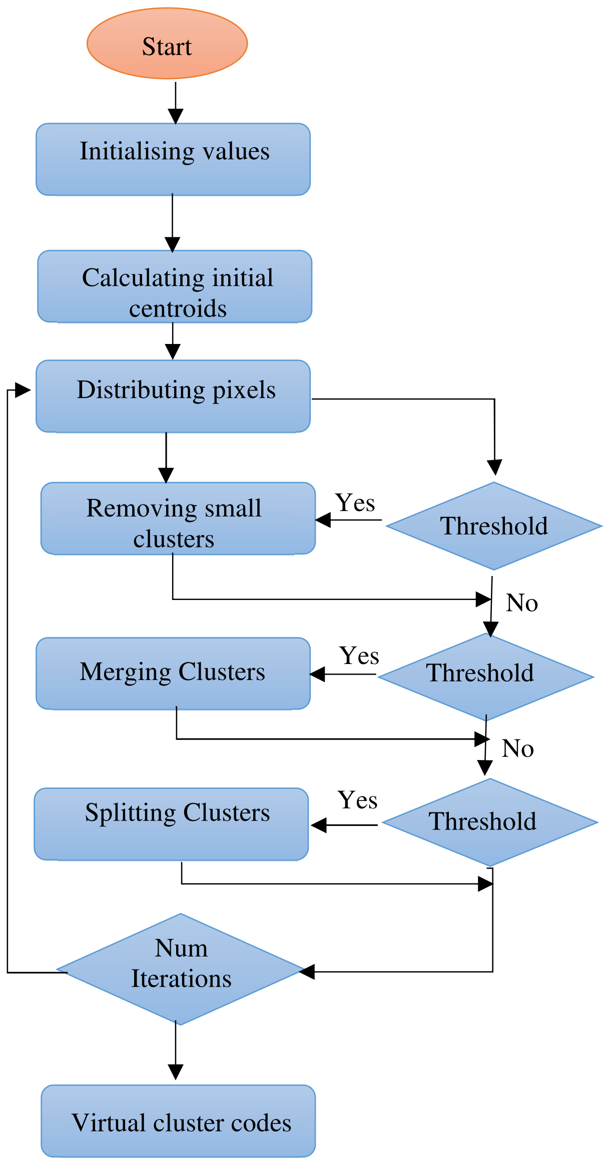

Fig. 3. The algorithm of the ISODATA method.

1) Initialising values: Firstly, the necessary variables are initialized:

- Maximum and minimum number of clusters (K_m and K_n respectively)
- Minimum required number of clusters (K_r)
- Cluster centroids (Z)
- Minimum distance between clusters (D)
- Number of iterations
- Maximum number of pixels in each cluster

2) Calculating initial centroids: The cluster centroids are initialized by reading in the first K_m values. A pixel was chosen as the centroid if it is not the same value of the previous centroid.

3) Distributing pixels: The algorithm assigns pixels to clusters with nearest centroid value.The centroid values must be re-measured each time a new pixel is added. This is shown in (1)

$$p \epsilon K_i \; if \; (|x - Z_i| \leq |x - Z_j|) \tag{1}$$

Where:
p is the pixel's cluster
K_i it the determined nearest cluster for pixel *i*
Z_i is the cluster's centroid
Z_j is the centroid of the next cluster
x is the pixel's band average

4) Removing small clusters: Clusters that contained a small number of pixels as specified by a threshold of $K_m/2$ are eliminated. All pixels in these clusters were redistributed to the cluster with the nearest centroid.

5) Merging clusters: Any two clusters that have close centroids are merged. Merging is based on a threshold, which is the distance between clusters, and the range of values within the cluster as show in (2).

$$K_m = (K_m - 1) \; if \; (|Z_i - Zj| \leq D) \tag{2}$$

Where:
K_m is the maximum number of clusters
D is the distance between clusters
Z_i is the centroid of the cluster
Z_j is the centroid of the next cluster

6) Spliting clusters: Clusters are split if they contain more than the maximum number of pixels.

$$K_n = (K_n + 1) if \; N(Z) \geq P \tag{3}$$

Where:

K_n is the number of clusters at this step
Z is the centroid of a cluster
N is the total number
P is the maximum number of pixels in the cluster

The process continues until the number of iterations is exhausted or the number of clusters is constant.

After the clustering is finished, a virtual raster is created containing the cluster values. This allows the algorithm to continue using the cluster values, a proxy for the pixel's environmental value. Figure 4 shows an example of the clustering phase using the clusters 0,1 and 2.

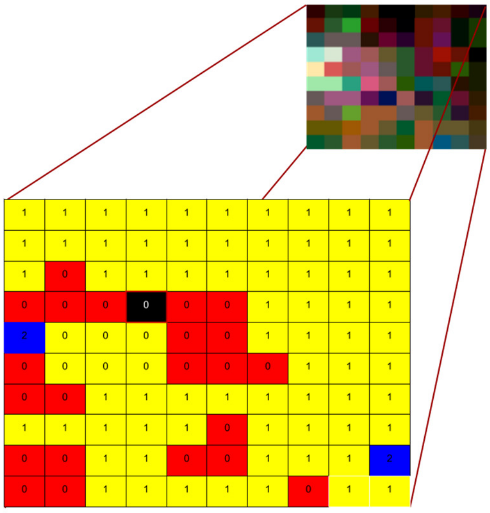

Fig 4. The output of the clustering algorithm.

B. *Predicting the Missing Pixel Phase*

This phase predicts the missing pixel. It uses a kriging model that selects the Optimal Pixel from the same environment class (for example vegetation) as the target pixel. This phase consists of four steps.

1) Estimating similarities: In the First-Fill image, a spatial model determines the distance around the Center Pixel. The Center Pixel is the pixel that located at the coordinate of the missing pixel in the First-Fill image.

2) Determining the Optimal Pixel: A kriging model is used to determine the Optimal Pixel from the same environment class of the Center Pixel. A 5 X 5 window is used to constrain the pixels around the Center Pixel. The technique measures similarities by the utilising of the root mean square deviation (RMSD) metric and the distance of the pixel, as (3) shows.

$$v_i = \sqrt{\frac{\sum_b^m (v_{s,b} - v_{i,b})^2}{m}} . \left(\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}\right) \quad (3)$$

Where:
vi is the similarity of pixel i from the Center Pixel
$v_{s,b}$ is the value of band b in the Center Pixel
$v_{i,b}$ is the value of band b in neighbor pixel i
m is the number of bands in the pixel
c is the Center Pixel
x and y are the pixels' coordinates

Equation 4 analyses pixel similarities and assigns the highest weight for the pixel with the lowest deviation from the Center Pixel. A default value (1/p) is recorded for pixels that have the same value of the Center Pixel, where p is the number of pixels that have the same value of the Center Pixel.

$$W_i = \frac{1}{d_i} * \left(\frac{1}{\sum_{i=1}^{n} \left(\frac{1}{d_i}\right)} \right) \quad (4)$$

Where:
W_i is the weight of the i^{th} neighbor pixel
d_i is the i^{th} neighbor pixel's distance

Next, the model compares the references of the Optimal Pixel and the Center Pixel, so that only pixels with the same environmental reference and highest similarity should be considered as the Optimal Pixel.

3) Building the linear multi-temporal relationship: At this step, the deterministic concept is used to build the liner multi-temporal relationship with input images. The temporal model links the three pixels in all images that are located in the same co-ordinates of the Optimal Pixel. It assumes that pixels from the same environmental class will change in the same way. Furthermore, the linkage guarantees that the pattern of change is measured from the same location.

4) Estimating the missing pixel: This step measures the pattern of change using pixels of the linear multi-temporal relationship and estimates the missing pixel value. The pattern of the change in the environment was obtained from subtracting the values in the First Pixel and the Second Pixel. This is shown in (5).

$$\forall b \epsilon B \ m_b = \left(\frac{v_{\beta,b} - v_{\alpha,b}}{v_{\alpha,b}} + 1\right) * v_{c,b} \quad (5)$$

Where:
m_b is the predicted value of the M*issing Pixel* in band *b,*
$v_{\alpha,b}$ is the value of band b in the First Pixel
$v_{\beta,b}$ is the value of band b in the Second Pixel
$v_{c,b}$ is the value of band *b* in the Center Pixel

C. *Adjusting the Deviation*

This step adjusts the errors in the predicted missing pixel by determining if the difference is linear between timesteps. If there is any difference, then the predicted value is required to be adjusted.

1) Estimating the End Pixel: A similar strategy for estimating the Missing Pixel is used to estimate the value of the End Pixel, as shown (6). Then, it is compared with the observed value. This allows us to determine the accuracy and to correct if necessary.

$$\forall b \epsilon B \ v_{eb} = \left(\left(\frac{v_{\gamma b} - v_{\beta b}}{v_{\beta b}}\right) + 1\right) * v_{m,b} \quad (6)$$

Where:
$v_{e,b}$ is the value of band b for the predicted Ending Pixel
$v_{\gamma,b}$ is the value of band b for the Third Pixel
$v_{\beta,b}$ is the value of band b for the Second Pixel
$v_{m,b}$ is the value of band b for the predicted Missing Pixel

2) Adjusting the Estimated Missing Pixel: Here, the difference between the estimated and real values of the End

Pixel was added to the fraction of change between the Second Pixel and Third Pixel. Then, the End Pixel's fraction of changes is used to adjust the Missing Pixel, as shown in (7).

$$\forall b \epsilon B \; v_{r,b} = v_{m,b} + \frac{(v_{e,b} - v_{p,b})}{c'_b} \qquad (7)$$

Where:
$v_{r,b}$ is the value of band b for the re-predicted Missing Pixel
$v_{m,b}$ is the value of band b for the predicted Missing Pixel
$v_{e,b}$ is the value of band b for the observed End Pixel
$v_{p,b}$ is the value of band b for the predicted End pixel
c'_b is the new change fraction of band b.

IV. RESULTS

A. *Spatial-Temporal Change in the Environment Context Dataset*

The Spatial-Temporal Change in the Environment Context (STCEC) dataset was used for testing [5]. The STCEC dataset is a standard dataset that tests remote sensing applications across different environmental contexts. The contexts are included in the STCEC dataset are the homogeneous with/without changes and heterogeneous with/without changes. The results w~~h~~ere also compared with IGMTPP methods and three baseline methods. These baseline methods were Average Between Two Pixels (ABTP), Weighted Liner Interpolation Method (WLIP) and Nearest Similar Pixel (NSSP) methods. More details about the baselines are available in [6].

1) Integration of Geostatistic and Temporal Missing Pixels' Properties Method (IGTMPP): The Integration of Geostatistic and Temporal Missing Pixels' Properties method combines together temporal and spatial properties. It determines a pixel based on the current location and band value. It then predicts the value of the pixel throughout time. Finally, it contains a correction method.

2) Average Between Two Pixels Method (ABTP): The average between two pixels method (ABTP) calculates the average between two pixels of the first and third image to find the value of the missing pixel in the second image.

3) Weighted Linear Interpolation Method (WLIP): WLIP is a linear function that assumes a straight-line trajectory between the pixels located in the same location in the first and third image. WLIP assumes that there is an intersection between the two points linearly weighted by the inverse temporal distance between the capture date of the two images.

4) Nearest Neighbour Similar Pixel Interpolation Method (NNSP) method: The NNSP method assigns values to missing pixels based on spectral values and geometric distance. For each missing pixel, it determines the most similar pixel in the corresponding co-ordinates from first image and combines this value with a weight inversely proportional to the its distance from the actual pixel. It then finds the co-ordinates of the most similar pixel in the second image and assigns those values to the missing pixel.

B. *Evaluation*

The methods predicted images that corresponded to the second dated images in the STCEC dataset. The metric used was the root mean square deviation. A t-test was used to determine which method produced the most accurate results. The most important comparison was with IGTMPP method since it already showed more accurate results than the baseline methods. The type of the t-test used was the two-sample that assumes unequal variances.

Table 1 shows the analysis in the dataset of the changed homogeneous samples. The test showed that the ISODATA-IGTMPP method produces significantly better results than the other methods. The t-test showed that the mean and variance values of ISODATA-IGTMPP (mean was 3.36 and variance was 1.95) method are lower than those for the IGTMPP method (mean was 4.76 and variance was 2.84). Furthermore, the t-stat is greater than the t-critical two-tail value.

TABLE 1. ROOT MEAN SQUARE DEVIATIONS OF EACH METHOD AND T-TEST VALUES FOR THE HOMOGENOUS ENVIRONMENT WITH CHANGE DATASET.

Method	**RMSD**	**σ^2**	**t-stat**	**p-value**	**t-critical**
ISODATA-IGTMPP	3.36	1.96			
IGTMPP	4.76	2.84	4.52	$1.80 * 10^{-5}$	1.99
ABTP	7.12	14.9	6.48	$1.72 * 10^{-8}$	2.00
WLIP	7.34	11.18	7.76	$7.13 * 10^{-11}$	2.00
NNSP	7.6	12.26	7.97	$3.79 * 10^{-11}$	2.00

An analysis on why the ISODATA-IGMTPP method produced the most accurate results for the changed homogeneous samples found the following reasons. Even though the pixels of the sub-dataset are from the same environment class, the method clustered the sample into two clusters. The clustering of pixels helped to minimize the range of the similarities between pixels in the kriging model which lead to a more accurate pixel being chosen as the Optimal Pixel. Choosing the right location for the Optimal Pixel was key for treating the errors that could happen by the irregular change of the data through time, as it changed in a similar way to the Missing Pixel. Likewise, the fraction of changes would be very close to the Missing Pixel, and so the prediction is very accurate.

Table 2 shows the results of the changed heterogeneous sub-dataset. The t-test confirmed that the ISODATA-IGTMPP method produced the most accurate results (mean was 3.26 and variance was 1.46) compared to the baselines and IGTMPP method (mean was 4.75 and the variance was 2.36). In addition, the t-stat values were always greater than the t-critical two-tail value.

TABLE 2. ROOT MEAN SQUARE DEVIATIONS OF EACH METHOD AND T-TEST VALUES FOR THE HETROGENOUS ENVIRONMENT WITH CHANGE DATASET.

Method	RMSD	σ^2	t-stat	p-value	t-critical
ISODATA-IGTMPP	3.36	1.96			
IGTMPP	4.75	2.36	5.90	$8.34 * 10^{-8}$	2.00
ABTP	7.32	10.89	8.40	$1.52 *10^{-11}$	2.01
WLIP	7.21	8.13	9.33	$3.27 *10^{-13}$	2.01
NNSP	7.45	7.96	10.1	$2.51 *10^{-14}$	2.01

The analysis of the ISODATA-IGTMPP method showed that the method had the advantage of choosing an Optimal Pixel that changed in the same way as the Target Pixel. This was particularly important for this dataset since the changed pixel likely changed in the same way to the Target Pixel.

Table 3 shows the performance of the ISODATA-IGTMPP method compared with the alternatives on the sub-dataset of homogeneous samples without changes. The t-test showed that the ISODATA-IGTMPP method (mean was 3.86 and variance 3.47) significantly outperformed the IGTMPP dataset (mean was 4.56 and variance 3.67) and the baseline methods. In addition, the t-stat values are always greater than the t-critical two-tail value.

TABLE 3. ROOT MEAN SQUARE DEVIATIONS OF EACH METHOD AND T-TEST VALUES FOR THE HOMOGENOUS ENVIRONMENT WITHOUT CHANGE DATASET.

Method	RMSD	σ^2	t-stat	p-value	t-critical
ISODATA-IGTMPP	3.86	3.47			
IGTMPP	4.56	3.76	1.84	$7.00 *10^{-2}$	1.99
ABTP	7.74	21.67	5.47	$8.10 *10^{-7}$	2.00
WLIP	7.72	15.15	6.33	$2.05 *10^{-8}$	2.00
NNSP	7.99	16.57	6.53	$9.63 *10^{-19}$	2.00

Analyzing the performance of the ISODATA-IGMPP method showed the advantage of using clustering and the correction model to help build the liner relationship between pixels over the time. The ISODATA-IGMTPP algorithm clusters pixels of the images into two clusters. While the correlations between pixels do not change that much through time, the clustering helps the correction model for measuring the compensation value from the cluster that shows the higher correlation with missing pixel.

Table 4 shows the results on the dataset of pixels from a heterogeneous environment without changes. It showed that the ISODATA-IGMTPP method produces superior results (mean 4.77 and variance 8.12) compared to the IGMTPP method (mean 5.20 and variance 8.15) and the other baseline methods. Likewise, the t-stat value was always greater than the t-critical two-tail value.

TABLE 4. ROOT MEAN SQUARE DEVIATIONS OF EACH METHOD AND T-TEST VALUES FOR THE HETROGENOUS ENVIRONMENT WITHOUT CHANGE DATASET.

Method	RMSD	σ^2	t-stat	p-value	t-critical
ISODATA-IGTMPP	4.77	8.12			
IGTMPP	5.20	8.15	0.76	$4.50*10^{-1}$	1.99
ABTP	7.70	12.08	4.63	$1.20 *10^{-5}$	1.99
WLIP	7.73	12.06	4.68	$9.89 *10^{-6}$	1.99
NNSP	7.92	14.61	4.69	$9.74 *10^{-6}$	1.99

The pixels in this sub-dataset have a large range of values but they do not change much through time. The clustering ensures that only pixels similar to the starting pixel are chosen as the optimal pixel. That helps the correction model to correct the errors that occur in the prediction process.

Figure 5 provides a summary of the performance of the all the methods. In particular, it shows the number of samples were one method outperformed the rest. It is clear that the ISODATA-IGTMPP method outperformed all the other methods with scores of 96%, 100%, 86% and 72%.

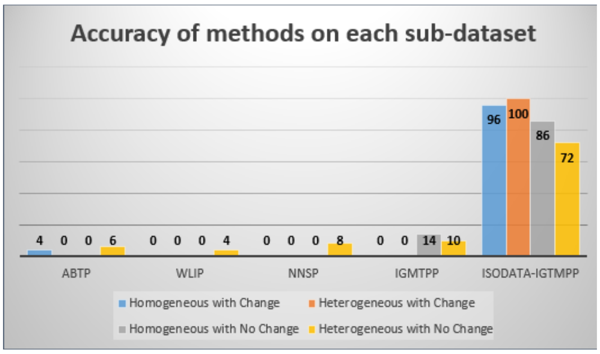

Fig 5. A comparison of between the ISODATA-IGTMPP compared to the other methods.

Table 5 presents results where IGTMPP method was outperformed by the baselines. In these examples it is shown that the ISODATA-IGMTPP achieved higher results. This demonstrates that the combination between geo-temporal and environmental class can help to enhance the performance of the pixel synthesis.

Overall, the results showed that the ISODATA-IGTMPP method produced the highest accuracy when compared to the IGMTPP and baseline methods. Predicting missing pixels from only their environmental class and using the combination between their geostatistical and temporal properties is advantageous.

TABLE 5. A COMPARISON OF THE ISODATA-IGTMPP METHOD IN AREAS WHERE THE BASELINE OUTPERFORMED THE IGTMPP METHOD.

Name	ABTP	WLIP	NNSP	IGMTPP	ISODATA-ISTMPP
Dunrobin1-hoc	5.80	5.98	5.17	6.16	5.17
Gardiner2-hoc	3.92	3.53	3.26	4.75	1.84
Gardiner3-hoc	2.64	1.89	1.47	3.41	1.45
Gardiner4-hoc	3.9	5.74	5.35	5.91	2.06
Gardiner5-hoc	3.22	5.68	5.40	4.21	2.04
Gardiner6-hoc	3.95	3.63	3.46	4.74	2.01
Boatman1-hoc	4.93	7.38	7.87	8.55	2.87
Boatman2-hoc	4.21	7.77	6.88	5.81	1.48
Boatman3-hoc	3.19	7.51	6.99	6.56	2.86
Boatman4-hoc	4.71	8.07	7.99	7.21	3.37
Boatman5-hoc	4.76	6.70	7.75	6.92	3.05
Boatman6-hoc	5.20	8.20	7.55	8.61	4.39
Dunrobin1-hec	4.24	6.00	6.34	4.58	4.11
Gardiner1-hec	2.90	2.61	2.74	3.75	1.58
Gardiner2-hec	4.75	3.55	3.06	4.90	2.09
Gardiner3-hec	5.99	4.33	4.84	5.56	2.60
Gardiner4-hec	4.88	5.34	5.46	7.63	2.50
Boatman1-hec	4.47	6.17	6.68	7.34	3.12
Boatman2-hec	4.17	7.70	7.09	4.43	1.80
Boatman3-hec	3.70	7.00	7.53	7.20	2.41
Boatman4-hec	4.85	7.71	7.56	7.12	2.68
Average	4.28	5.84	5.74	5.97	2.65

V. CONCLUSION

This paper showed that using a clustering algorithm can help to synthesis pixels. The output was the ISODATA-IGTMPP method, a combination between clustering and a pixel synthesis method. It showed that the ISODATA-IGTMPP algorithm performed well across all different environmental types. Future work will focus on enhancing the ISODATA-IGTMPP method in the clustering proportion.

REFERENCES

[1] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction* vol. 4th. Berlin: Springer, 2006.

[2] D. Fernández-Prieto and R. Sabia, *Remote Sensing Advances for Earth System Science*: Springer, 2013.

[3] H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, *et al.*, "Missing information reconstruction of remote sensing data: A technical review," *Geoscience and Remote Sensing Magazine, IEEE,* vol. 3, pp. 61-85, 2015.

[4] S.-W. Jang, G.-Y. Kim, and S. Byun, "Clustering-based pattern abnormality detection in distributed sensor networks," *International Journal of Distributed Sensor Networks,* vol. 10, pp. 438-468, 2014.

[5] T. F. Ali and A. Woodley, "IGTMPP: A hybrid method to predict missing pixels of remote sensing images using geo-temporal properties," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, Canberra, Australia, 2018, pp. 1-8.

[6] T. F. Ali and A. Woodley, "STCEC: A remote sensing dataset for identifying spatial-temporal change in homogeneous and heterogeneous environments," presented at the 2018 Digital Image Computing: Techniques and Applications (DICTA), Perth, Australia, 2019.

[7] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 56, pp. 4274-4288, 2018.

[8] W. B. Cohen and S. N. Goward, "Landsat's role in ecological applications of remote sensing," *BioScience,* vol. 54, pp. 535-545, 2004.

[9] H. Shen, A. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, *et al.*, "Missing information reconstruction of remote sensing data: A technical review," *IEEE Geoscience and Remote Sensing Magazine,* vol. 3, pp. 61-85, 2015.

[10] B. Gorte, F. van der Meer, A. Stein, and V. Springer, *Spatial Statistics for Remote Sensing* vol. 1. Dordrecht;Boston;: Kluwer Academic Publishers, 1999.

[11] R. Sunila and K. Kollo, "A comparison of geostatistics and fuzzy application for digital elevation model," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. 36, p. 2, 2007.

[12] Y. Chao, C. Liangfu, S. Lin, F. Meng, and L. Shenshen, "Kriging interpolation method and its application in retrieval of MODIS aerosol optical depth," in *19th International Conference on Geoinformatics 2011*, 2011, pp. 1-6.

[13] C. Feng, T. Lina, and Q. Quanyi, "Exploitation of CBERS-02B as auxiliary data in recovering the Landsat7 ETM+ SLC-off image," in *Geoinformatics, 2010 18th International Conference on*, 2010, pp. 1-6.

[14] M. Peralvo and D. Maidment, "Influence of DEM interpolation methods in drainage analysis," *Gis Hydro,* vol. 4, 2004.

[15] S. K. Maxwell, G. L. Schmidt, and J. C. Storey, "A multi-scale segmentation approach to filling gaps in Landsat ETM+ SLC-off images," *International Journal of Remote Sensing,* vol. 28, pp. 5339-5356, 2007.

[16] A. Fahad, N. Alshatri, Z. Tari, A. Alamri, I. Khalil, A. Zomaya, *et al.*, "A survey of clustering algorithms for big data: Taxonomy & empirical analysis," *IEEE Transactions on Emerging Topics in Computing,* vol. 2, 2014.

[17] A. Woodley, L.-X. Tang, S. Geva, R. Nayak, and T. Chappell, "Parallel K-Tree: A multicore, multinode solution to extreme clustering," *Future Generation Computer Systems,* vol. 99, pp. 333-345, 2019.

[18] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-means clustering," *IEEE Geoscience and Remote Sensing Letters,* vol. 6, pp. 772-776, 2009.

[19] A. Rekik, M. Zribi , M. Hamida, A. B. , and M. Benjelloun, "An optimal unsupervised satellite image segmentation approach based on pearson system and k-means clustering algorithm initialization," *Methods,* vol. 8, pp. 9-17, 2009.

[20] S. Owen, R. Anil, T. Dunning, and E. Friedman, *Mahout in Action*, 2011.

[21] A. Woodley, T. Chappell, S. Geva, and R. Nayak, "Efficient feature selection and nearest neighbour search for hyperspectral image classification," in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2016, pp. 1-8.

[22] A. Woodley, L. Tang, S. Geva, R. Nayak, and T. Chappell, "Using parallel hierarchical clustering to address spatial big data challenges," in *2016 IEEE International Conference on Big Data (Big Data)*, 2016, pp. 2692-2698.

[23] B. Li, H. Zhao, and Z. Lv, "Parallel ISODATA clustering of remote sensing images based on MapReduce," in *2010 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 2010, pp. 380-383.

[24] N. Memarsadeghi, D. M. Mount, N. S. Netanyahu, and J. Le Moigne, "A fast implementation of the ISODATA clustering algorithm," *International Journal of Computational Geometry & Applications,* vol. 17, pp. 71-103, 2007.

[25] Q. Wang, Q. Li, H. Liu, Y. Wang, and J. Zhu, "An improved ISODATA algorithm for hyperspectral image classification," presented at the 7th International Congress on Image and Signal Processing, 2014.