



Article

High-Dimensional Satellite Image Compositing and Statistics for Enhanced Irrigated Crop Mapping

Michael J. Wellington *  and Luigi J. Renzullo

Fenner School of Environment and Society, The Australian National University, Canberra, ACT 2601, Australia; luigi.renzullo@anu.edu.au

* Correspondence: michael.wellington@anu.edu.au

Abstract: Accurate irrigated area maps remain difficult to generate, as smallholder irrigation schemes often escape detection. Efforts to map smallholder irrigation have often relied on complex classification models fitted to temporal image stacks. The use of high-dimensional geometric median composites (geomedians) and high-dimensional statistics of time-series may simplify classification models and enhance accuracy. High-dimensional statistics for temporal variation, such as the spectral median absolute deviation, indicate spectral variability within a period contributing to a geomedian. The Ord River Irrigation Area was used to validate Digital Earth Australia's annual geomedian and temporal variation products. Geomedian composites and the spectral median absolute deviation were then calculated on Sentinel-2 images for three smallholder irrigation schemes in Matabeleland, Zimbabwe, none of which were classified as areas equipped for irrigation in AQUASTAT's Global Map of Irrigated Areas. Supervised random forest classification was applied to all sites. For the three Matabeleland sites, the average Kappa coefficient was 0.87 and overall accuracy was 95.9% on validation data. This compared with 0.12 and 77.2%, respectively, for the Food and Agriculture Organisation's Water Productivity through Open access of Remotely sensed derived data (WaPOR) land use classification map. The spectral median absolute deviation was ranked among the most important variables across all models based on mean decrease in accuracy. Change detection capacity also means the spectral median absolute deviation has some advantages for cropland mapping over indices such as the Normalized Difference Vegetation Index. The method demonstrated shows potential to be deployed across countries and regions where smallholder irrigation schemes account for large proportions of irrigated area.



Citation: Wellington, M.J.; Renzullo, L.J. High-Dimensional Satellite Image Compositing and Statistics for Enhanced Irrigated Crop Mapping. *Remote Sens.* **2021**, *13*, 1300. <https://doi.org/10.3390/rs13071300>

Academic Editor: Mehdi Hosseini

Received: 26 February 2021

Accepted: 25 March 2021

Published: 29 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: geomedian; smallholder; irrigation; random forest; high-dimensional

1. Introduction

Mapping and quantifying irrigated areas are critical to local, national, and international organizations that aim to understand and govern land and water resources for food security and sustainable development. National estimates are often based on non-exhaustive on-ground surveys, or low-resolution remote sensing estimates for continental scale applications. This can result in a broad range of estimations for a given area. For example, Vogels et al. [1] recently found that irrigated area across the Horn of Africa was on the order of two to four times greater than several official estimates.

Irrigated farming in Zimbabwe occurs along a spectrum of scales, from small informal plots to large commercial operations. Based on government records, Landsat, and MODerate resolution Imaging Spectroradiometer (MODIS) imagery, current estimates of irrigated area range from 123,900 to 202,600 ha [2,3]. Most of the area included in these estimates is in Mashonaland where large-scale irrigation schemes are prevalent [4]. Irrigated area estimates for Matabeleland, where smallholder schemes predominate, are much lower [4]. However, these estimates overlook some irrigation schemes and smallholder activity. It is hypothesized that official figures underestimate irrigated area, as Vogels et al. [1] observed

for the Horn of Africa. Agencies in Zimbabwe are interested in detailed irrigated cropland mapping for the region of Matabeleland, as this would inform natural resource policies and agricultural research, development, and extension priorities [5]. This region continues to be affected by drought, poverty, and food and water scarcity, so sustainable irrigation is critical to development efforts [6,7]. A reliable method of irrigated cropland classification that accounts for smallholder activity in Matabeleland is therefore required.

Smallholder irrigation activity in tropical environments remains more difficult to accurately map than large-scale commercial irrigation. Small and irregular field size and shape, synchrony of green vegetation phenology in the wet season, cloud cover during crop seasons, and in-field heterogeneity are major challenges for remote sensing of smallholder irrigation [8,9]. The Sentinel-2 satellite mission offers increased sensor resolution capability over the Landsat and MODIS satellites which have been used extensively for land use classification, including irrigated area, over recent decades. Recent work towards improved mapping of small-scale irrigation has largely focused on the application of Sentinel-2 imagery [10].

Several classification techniques have been tested for efficacy on small-scale irrigated areas. Vogels et al. [1] applied a supervised, object-based approach on dry-season mosaic images, with object symmetry and roundness variables included in the random forest classification model. Bousbih et al. [11] explored the fusion of Sentinel-2 optical imagery with a soil moisture product, although the relatively coarse resolution of satellite soil moisture data limits their application in smallholder contexts. Hollander [12] collected training polygons in Mozambique to apply a Sentinel-2 supervised learning model, although found that the ground-collected training data did not include sufficient samples of non-irrigated areas that shared similar spectral characteristics, such as light seasonal vegetation, to fit a sound classification model. It is likely that Vogels et al. [1] approach of collecting training data from visual interpretation produces a better representation of landscape variability, and, therefore, a more robust supervised classification model.

Combining finer resolution satellite imagery with novel approaches to land use classification problems offers a path towards more reliable irrigated area detection. Traditionally, land use classification methods have tended to use a selection of clean images, or simple composite mosaics of clean images [1,13]. The high-dimensional geometric median, hereafter geomedian, has been proposed as a way of constructing high-quality, cloud-free composites whilst preserving high-dimensional relationships between spectral bands [14]. It was developed with the aim of replacing the need for temporal stacks of poorer quality images, which is a popular means of training complex classification models [1,13–17]. Additionally, high-dimensional statistics of temporal variation [18] may also be important predictors of irrigation and other agricultural activity due to the spectral variability associated with cultivation, crop growth, and harvest activities. Therefore, augmenting geomedian images with high-dimensional statistics of time-series may enhance the accuracy of land use classification models for irrigated cropland.

Digital Earth Australia, an Open Data Cube (odc) initiative, offers several high-dimensional statistical products of value for land use classification. For example, an annual geomedian product derived from Landsat-8 is available for the entire Australian continent from 2013 to 2018 [14,19]. Additionally, a triple median absolute deviation [20] product is available for the same period, also derived from Landsat-8 [18,20]. The rationale for this product is change detection and machine learning for land use classification, especially over areas that undergo large changes in cover within a year, such as irrigated croplands [18]. The triple median absolute deviation product has three measures of temporal variation: the Euclidean median absolute deviation, spectral (cosine) median absolute deviation (SMAD), and Bray Curtis dissimilarity. Of these, the SMAD is most appropriate for highlighting areas of change within the period contributing to a given geomedian [18]. Therefore, the SMAD may be an important and useful variable in machine learning models for cropland mapping.

The analysis-ready annual products available from Digital Earth Australia make data acquisition simple, although recalculation of high-dimensional statistics is necessary for applications beyond the Australian continent. For example, high-dimensional products are not yet available in Digital Earth Africa, a parallel odc initiative to Digital Earth Australia. Application of high-dimensional compositing across satellites and continents therefore requires a methodology that calculates necessary images and statistics from available reflectance data.

This research paper aims to demonstrate the improvements in irrigation area mapping resulting from the use of high-dimensional statistics and supervised satellite image classification using odc infrastructure. The paper first demonstrates the use of existing geomedian and SMAD products over a much-studied irrigation scheme in Australia. Secondly, geomedian and SMAD are derived from Sentinel-2 imagery through Digital Earth Africa. Finally, the performance of the high-dimensional dataset approach to classification of smallholder irrigation schemes is evaluated with reference to existing mapping products.

2. Materials and Methods

2.1. Site Selection and Information

The Ord River Irrigation Area (ORIA) surrounding the town of Kununurra in Western Australia was chosen for validating the Digital Earth Australia products (Table 1). This was because it is a well-studied, irrigated area, and shares similar biophysical and climatic characteristics to irrigation schemes in southern Africa, with frequent cloud cover in the monsoonal wet season [21]. Furthermore, the ORIA supports numerous annual crops in addition to perennial tree crops, despite having less within-field heterogeneity and much larger field size than smallholder irrigation schemes [22].

In total, three sites in Zimbabwe were selected from active irrigation schemes covered by phase 2 of the Transforming Irrigation in Southern Africa (TISA) project, funded by the Australian Centre for International Agricultural Research (ACIAR). The three sites: Silalatshani, Nabusenga, and Lungwalala, are geographically disparate within Matabeleland and vary in scale (Table 1). Furthermore, they are surrounded by various other land types including dryland (rainfed) cultivation, water bodies, and natural vegetation. Locations were confirmed by TISA project leaders [23].

Table 1. Irrigation schemes used for irrigated land use classification, their location, coordinates, and approximate area.

Irrigation Scheme	Location	Coordinates	Area Equipped for Irrigation (ha)
Ord River Irrigation Area	Kununurra, Western Australia	−15.601, 128.762	14,000 [24]
Silalatshani	Matabeleland South Province, Zimbabwe	−20.799, 29.296	442 [25]
Nabusenga	Matabeleland North Province, Zimbabwe	−17.462, 28.063	19 (measured)
Lungwalala	Matabeleland North Province, Zimbabwe	−17.938, 27.561	132 (measured)

2.2. Data Collection and Preprocessing-Digital Earth Australia

Data for the ORIA were collected within the Digital Earth Australia ‘sandbox’, which provides access to odc products in a Jupyter Notebook environment. The Landsat-8 geomedian product (ls8_nbart_geomedian_annual) was generated at 25-m resolution for the year 2017. The triple median absolute deviation product (ls8_nbart_tmad_annual) was derived with the same resolution, extents, and period. Then, three indices were calculated on the geomedian product: Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Bare Soil Index (BSI) using the Digital Earth Australia indices package [26]. The geomedian and triple median absolute deviation datasets were

then merged to form a 12 variable classification dataset comprising six spectral bands, three indices, and three measures of temporal variation (Table 2).

Table 2. Variables used for irrigated land use classification for the Ord River Irrigation Area site.

Group	Variable	Band or Source
Spectral bands	Blue	B1
	Green	B2
	Red	B3
	Near Infrared	B4
	Shortwave Infrared 1	B5
	Shortwave Infrared 2	B6
Indices	Normalized Difference Vegetation Index (NDVI)	[27]
	Normalized Difference Water Index (NDWI)	[28]
	Bare Soil Index (BSI)	[29]
Temporal variation	Spectral median absolute deviation (SMAD)	[18]
	Euclidean median absolute deviation (EMAD)	[18]
	Bray-Curtis Dissimilarity (bcdev)	[18]

2.3. Data Collection and Preprocessing-Digital Earth Africa

Data for the three Matabeleland sites were generated and collected from the Digital Earth Africa ‘sandbox’, a parallel initiative to Digital Earth Australia. All available cloud optimized Sentinel-2 (s2_l2a) images were collected in 10-m resolution over each site for the year 2019 [30]. This means that cloudy pixels were attributed as missing values and thus ignored in the calculation of high-dimensional composites and statistics.

The two classification datasets were retrieved for each Matabeleland site: an annual geomedian composite, and a stack of four (January–March, April–June, July–September, October–December) geomedian composites. Cloud effects in wet season months meant that complete geomedian composites could not be generated for each month, so the 3-monthly (quarterly) approach was taken. Annual and quarterly geomedian images were calculated with the `odc` package using the command ‘`xr_geomedian`’ which computes geomedians from a defined image stack [31]. The geomedian (g) was calculated on a collection of images (x, \dots, x_n), based on Roberts et al. [14], as:

$$g = \underset{x}{\operatorname{argmin}} \sum_{i=1}^n \|x - x_i\| \quad (1)$$

where *argmin* is the “argument of the minima” [14,18].

The SMAD was calculated on the six spectral bands for Sentinel-2 (Table 3). The SMAD was defined, based on Roberts et al. [18] as:

$$\text{SMAD} = \text{median}(\text{cosdist}(x^{(t)}, g), t = 1, \dots, n) \quad (2)$$

where x is a temporal stack dataset of images over a given period (t) contributing to the geomedian (g). The cosine distance was calculated as:

$$\text{cosdist}(x, g) = 1 - \frac{x^T g}{(\|x\| \cdot \|g\|)} \quad (3)$$

where the numerator on the righthand side of Equation (3) is the dot product of spectral data for vectors x and g and $\|\cdot\|$ is the product of the L-2 norms of each vector.

The SMAD calculation used to collect data for Matabeleland sites was validated against the existing Digital Earth Australia SMAD product over the ORIA before application in Digital Earth Africa. SMAD was chosen over other high-dimensional deviation statistics due to its relative capacity to highlight change within periods of interest [18]. The NDVI, NDWI, and BSI were calculated for each geomedian image with the Digital Earth Africa indices package [32].

The collection process resulted in an annual dataset of 10 variables comprising six spectral bands, three indices, and SMAD for each site (Table 3). Consequently, the stacked classification dataset of four quarterly geomedians comprised 40 (4 quarters per year \times 10 variables) classification variables.

Table 3. Variables used for irrigated land use classification for three irrigation schemes in Matabeleland.

Group	Variable	Formula	Band or Source
Spectral bands	Green		B1
	Red		B2
	Blue		B3
	Near Infrared (NIR)		B4
	Shortwave Infrared 1 (SWIR1)		B5
	Shortwave Infrared 2 (SWIR2)		B6
Indices	Normalised Difference Vegetation Index (NDVI)	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$	[27]
	Normalised Difference Water Index (NDWI)	$(\text{NIR} - \text{SWIR}) / (\text{NIR} + \text{SWIR})$	[28]
	Bare Soil Index (BSI)	$((\text{Red} + \text{SWIR}) - (\text{NIR} + \text{Blue})) / ((\text{Red} + \text{SWIR}) + (\text{NIR} + \text{Blue}))$	[29]
Temporal variation	Spectral median absolute deviation (SMAD)	Equation (2)	[14]

2.4. Data Sampling and Classification

Approximately 100 polygons were drawn over each image and labeled as either ‘irrigated’ or ‘other’ based on visual interpretation as shown in Figure 1. The rule for labeling a field as ‘irrigated’ was the appearance of bright green vegetative crop in at least one of the single-time images within the year of interest. For the ORIA, this was generally in the dry-season when water from channels is applied to crops [22]. For the Matabeleland schemes, each site has been studied as part of the TISA project and are known to be irrigated from channels, especially in the late dry-season. However, visual inspection of the image time-series ensured only actively irrigated fields were included. The polygon shapefiles and geomedian images were then read into R and 80% of polygons were sampled as training polygons, with the remaining 20% retained as validation polygons. The *sp* package was used to randomly sample 80,000 pixels from within the training polygons and 20,000 pixels from the validation polygons [33]. The *caret* package was then used to partition the sample into 80% training data and 20% validation data [34]. The classification was therefore pixel-based, although training and validation data were sampled from separate polygons.

The *randomForest* package was used to train the classification model with 500 trees. Variable importance for the random forest model was reported using mean decrease in accuracy. The *caret* package was used to generate confusion matrices, overall accuracies, and Kappa coefficients for model performance on the validation data [34,35]. Finally, the relevant classification model was applied to the entire extent for each site to classify pixels not included within either training or validation datasets.

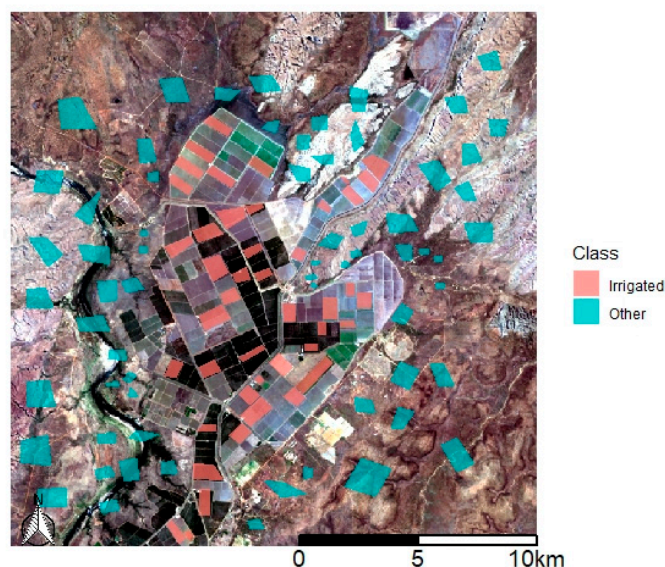


Figure 1. Annual geomedian image of the Ord River Irrigation Area overlaid with hand-drawn polygons labeled as either 'Irrigated' or 'Other' by visual interpretation, from which 80,000 training and 20,000 validation pixels were randomly sampled for classification.

2.5. Comparison to Existing Products

The Global Map of Irrigated Areas was inspected on the AQUAMAPS web application [36–38]. As none of the Matabeleland sites were classified as equipped for irrigation in the Global Map of Irrigated Areas, no further comparisons were made. For the ORIA, the AQUAMAPS product 'percent of area equipped for irrigation' was downloaded for comparison to classification results.

The FAO portal for monitoring Water Productivity through Open access of Remotely sensed derived data (WaPOR) [39] was used as comparison for the Matabeleland sites. The continental scale, 250-m resolution, WaPOR land use classification product was downloaded from the WaPOR database [40] and cropped to the extent of each site. As the product comprises 24 land use classes, all classes except 'cropland, irrigated' were combined to form an 'other' class for ease of comparison. Confusion matrices were then generated for the WaPOR classification against the validation (20,000 pixels) dataset using the caret package [34]. Accuracy statistics for the WaPOR classification were compared with those for the high-dimensional classification method.

3. Results

3.1. Calculation of Geomedian and Spectral Median Absolute Deviation

Geomedians and SMAD were derived as existing Landsat-8 based datasets from Digital Earth Australia and recalculated from Sentinel-2 images in the Digital Earth Africa platform. Visual inspection of SMAD plotted as a single band image demonstrates its potential for use in classifying cropland areas, along with other established predictors such as NDVI (Figure 2).

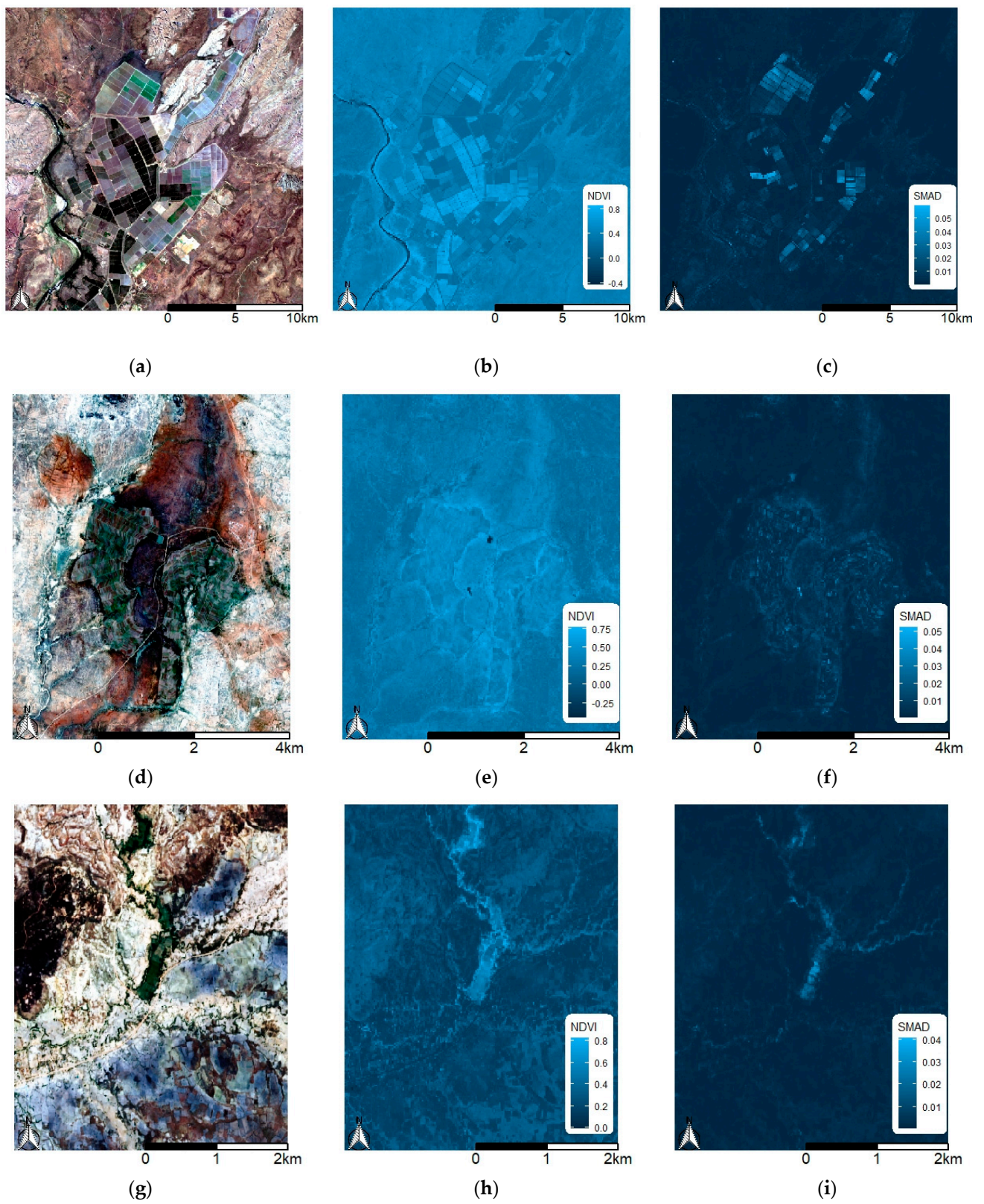


Figure 2. Cont.

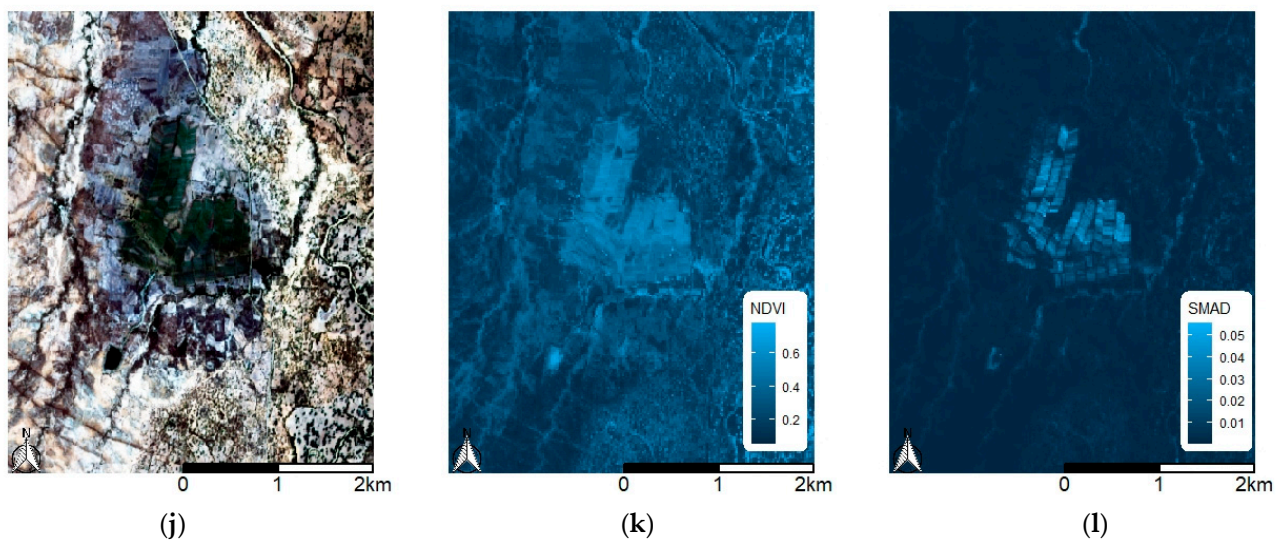


Figure 2. Images of the (a–c) Ord River, (d–f) Silalatshani, (g–i) Nabusenga, and (j–l) Lungwalala irrigation schemes showing (a,d,g,j) the calculated geomedian true color composite image, (b,e,h,k) a single band image for Normalized Difference Vegetation Index (NDVI) calculated on the geomedian, and (c,f,i,l) a single band image for the calculated spectral median absolute deviation (SMAD).

3.2. Irrigated Area Classification

Confusion matrices and prediction statistics for each classification model show model performance on validation data (Table 4). All models accurately classified irrigated areas for all sites, with each model giving overall accuracy levels greater than 84% and Kappa coefficients greater than 0.6. The stacked quarter datasets gave better accuracy results than the annual datasets.

Table 4. Confusion matrices, overall accuracy, and Kappa coefficients for classification model performance on validation data (20,000 pixels) for all datasets used.

Ord River Irrigation Area		Observed	
Predicted		Irrigated	Other
Irrigated	8382	8382	351
Other	2820	2820	8447
Overall accuracy (%)	84.1		
Kappa coefficient	0.69		
Silalatshani			
(A) Annual dataset		Observed	
Predicted		Irrigated	Other
Irrigated	2583	2583	226
Other	1554	1554	15,637
Overall accuracy (%)	91.1		
Kappa coefficient	0.69		
(B) Stacked quarter dataset		Observed	
Predicted		Irrigated	Other
Irrigated	3498	3498	0
Other	627	627	15,860
Overall accuracy (%)	96.8		
Kappa coefficient	0.90		

Table 4. Cont.

Nabusenga			
(A) Annual dataset	Observed		
Predicted	Irrigated		Other
Irrigated	4284		0
Other	583		15,133
Overall accuracy (%)	97.1		
Kappa coefficient	0.92		
(B) Stacked quarter dataset	Observed		
Predicted	Irrigated		Other
Irrigated	4789		0
Other	161		15,050
Overall accuracy (%)	99.2		
Kappa coefficient	0.98		
Lungwalala			
(A) Annual dataset	Observed		
Predicted	Irrigated		Other
Irrigated	4411		46
Other	41		15,502
Overall accuracy (%)	99.6		
Kappa coefficient	0.99		
(B) Stacked quarter dataset	Observed		
Predicted	Irrigated		Other
Irrigated	4452		0
Other	0		15,548
Overall accuracy (%)	1		
Kappa coefficient	1		

Visual inspection of classification maps (Figure 3) for the annual datasets supports the high accuracy statistics. These plots depict the probability of a given pixel being classified as irrigated, calculated as the proportion of 500 trees in the random forest giving an ‘irrigated’ vote. Pixels with a probability greater than 0.5 are classified as ‘irrigated’. Brighter areas are classified as ‘irrigated’ with higher model confidence. While pixels with values less than 0.5 are classified as ‘other’, brightness levels identify some landscape features, such as riparian vegetation, which are prone to misclassification.

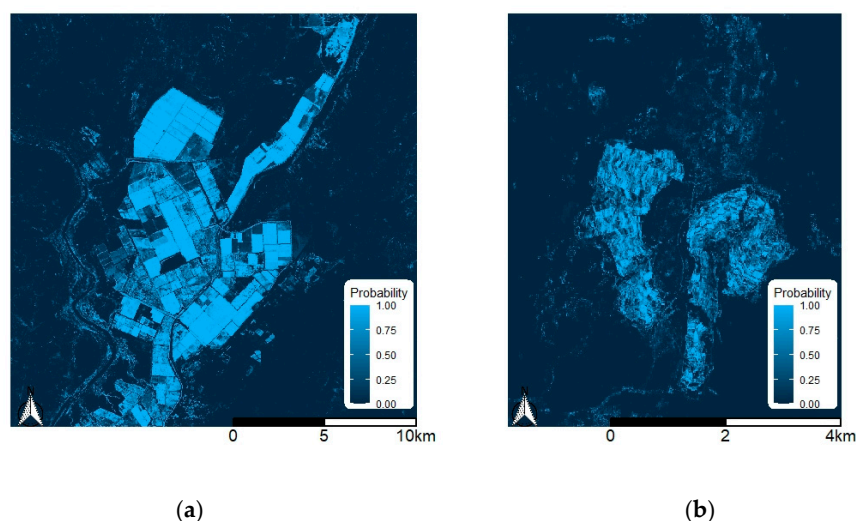


Figure 3. Cont.

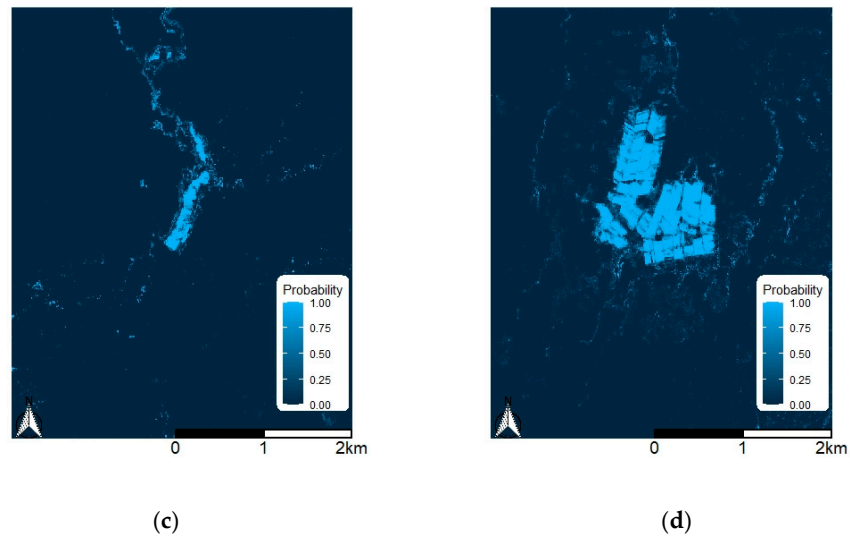


Figure 3. Probability of pixels being classified as irrigated, defined as the proportion of 500 decision trees in the random forest model, for annual datasets over (a) Ord River Irrigation Area, (b) Silalatshani, (c) Nabusenga, and (d) Lungwalala.

3.3. Variable Importance for Irrigated Area Classification

Variable importance analyses for the classification results shown in Table 4 and Figure 3 showed that SMAD was the most important variable for all classification models, except Silalatshani, applied to annual datasets based on mean decrease in accuracy (Figure 4). Higher values indicate a greater decrease in model accuracy if that variable is omitted from the model. The mean decrease in accuracy can therefore be interpreted as a test or summary statistic for variable importance in a random forest classification model.

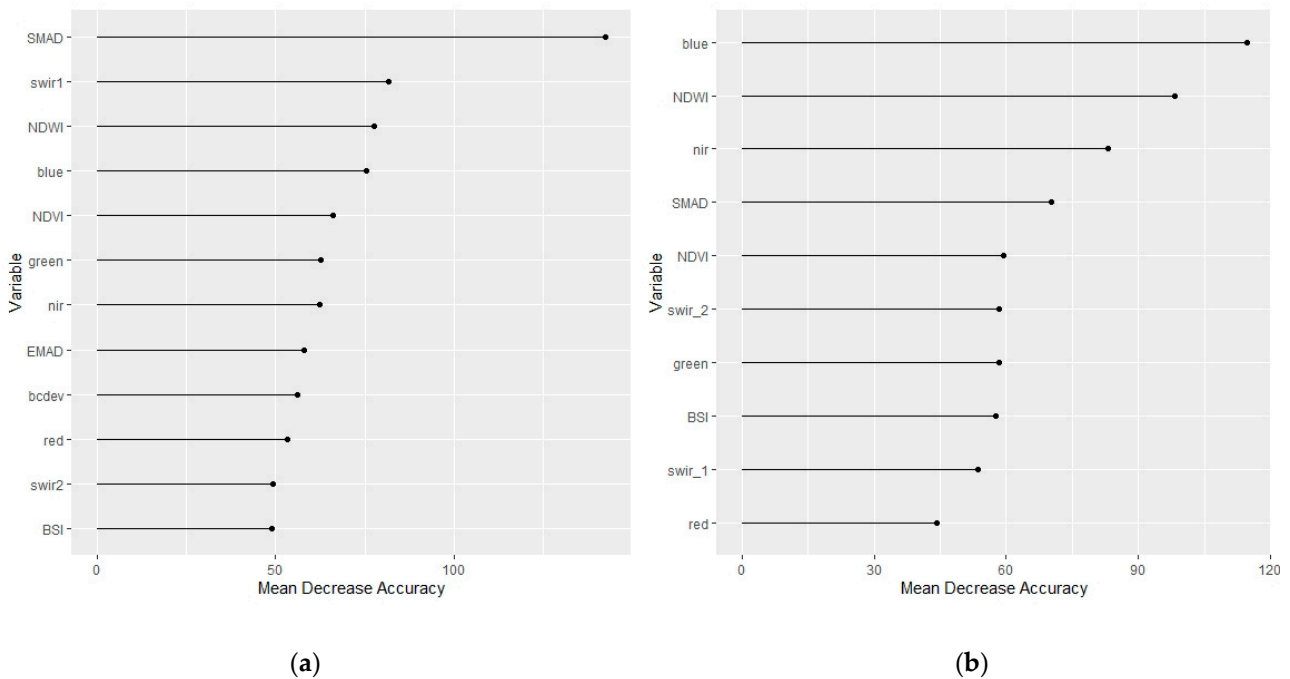


Figure 4. Cont.

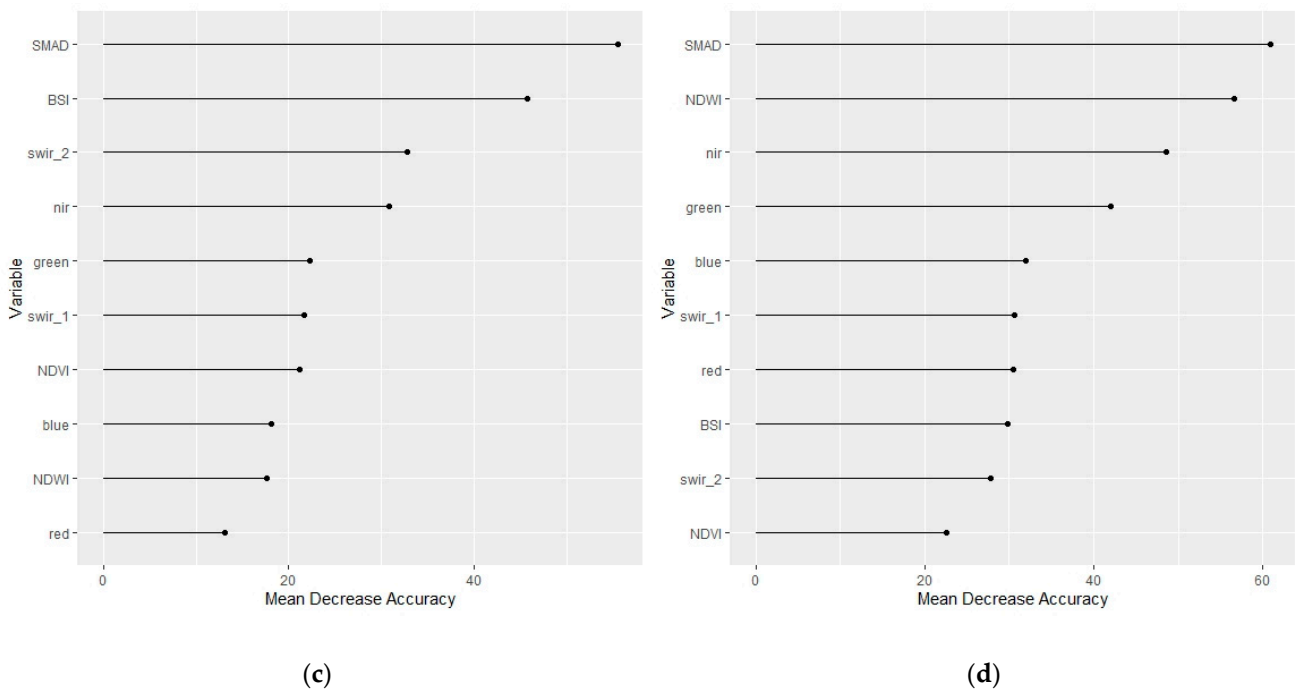


Figure 4. Variable importance expressed as mean decrease in accuracy for the random forest classification models on annual datasets, shown in Figure 3, for (a) Ord River Irrigation Area, (b) Silalatshani, (c) Nabusenga, and (d) Lungwalala.

Figure 5 shows the variable importance plots for the stacked quarterly (4 geomedian images x 10 predictors described in Table 2 = 40 predictors) classification datasets. The SMAD variable for each quarter appeared in the top 15 most important variables of 40 for each site. Like the annual datasets, the importance of indices and spectral bands varied between sites. There was no discernible trend for the importance of specific quarters in any classification model (Figure 5).

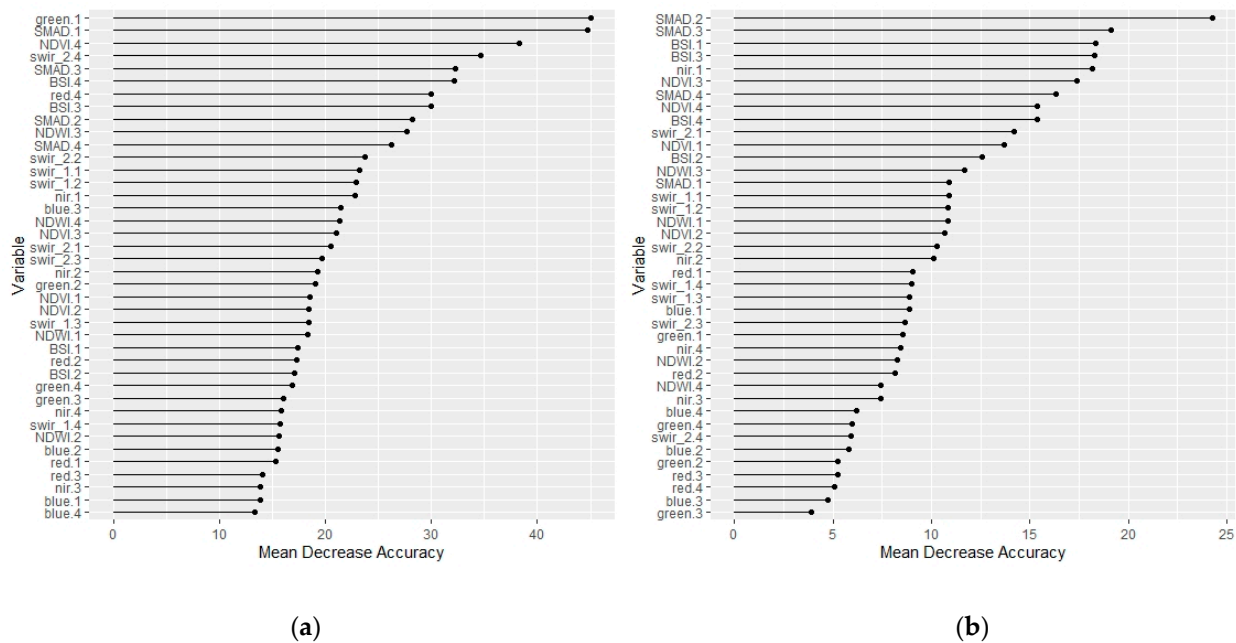
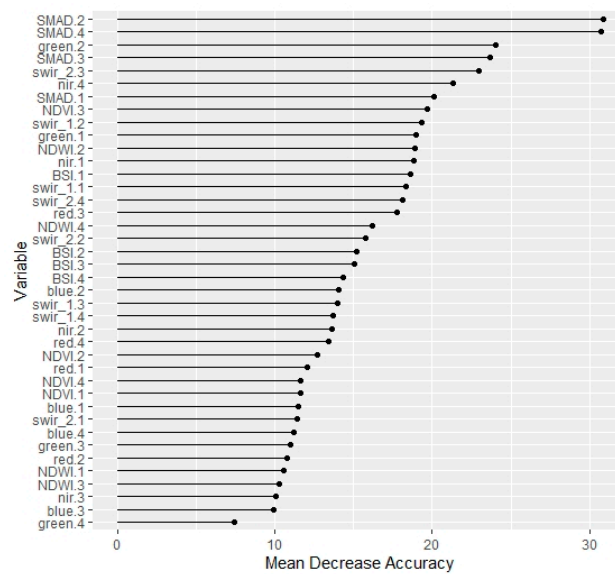


Figure 5. Cont.



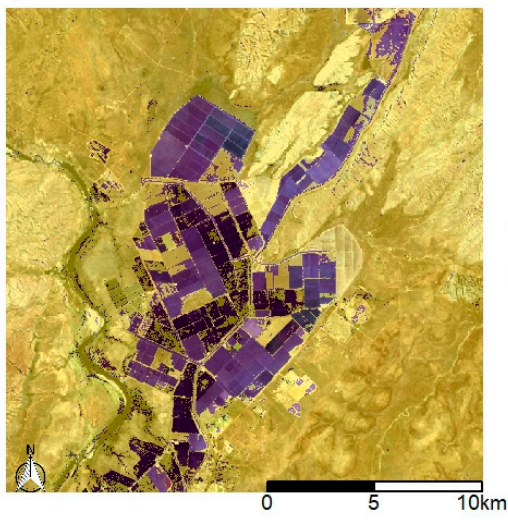
(c)

Figure 5. Variable importance expressed as mean decrease in accuracy for the stacked quarter datasets where the digit following each variable represents the quarter (1 = January–March, 2 = April–June, 3 = July–September, 4 = October–December) for (a) Silalatshani, (b) Nabusenga, and (c) Lungwalala.

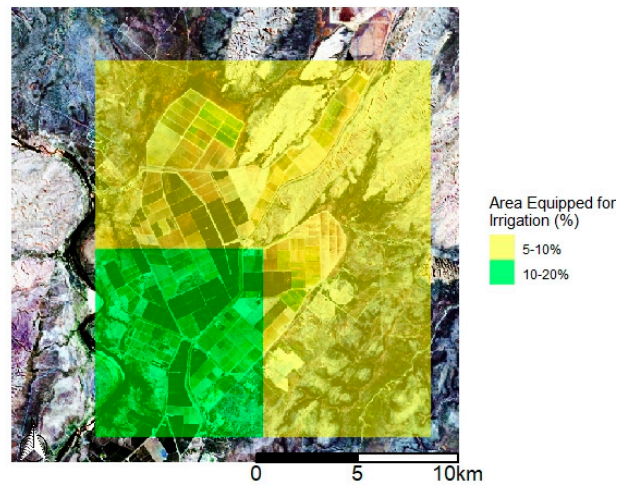
3.4. Comparison to Existing Products

Comparing the classification results detailed in Table 4 and Figure 3 with existing irrigated land maps illustrates differences in accuracy and resolution. Figure 6 compares the classified image of ORIA predicted on the random forest classification model for the annual Digital Earth Australia products with the Global Map of Irrigated Areas. Notably, this map product classified none of the Matabeleland sites as irrigated. Instead, the Matabeleland sites are compared with the WaPOR land use classification product for the African continent (Figure 6).

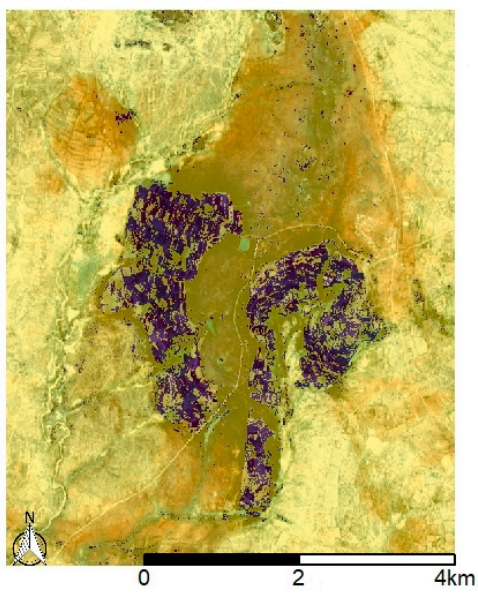
The confusion matrices for the WaPOR classification against training and validation data (Table 5) can be compared with those in Table 4. The high-dimensional method on annual datasets produced overall accuracy of 95.9% on average across the three sites, while the average for WaPOR was 77.2%. The average Kappa coefficient of 0.87 was also higher than the WaPOR average of 0.12. The discrepancy between 250-m resolution for the continental WaPOR product and 10-m resolution for the Sentinel-2 dataset contributes to the accuracy results, as illustrated in Figure 6.



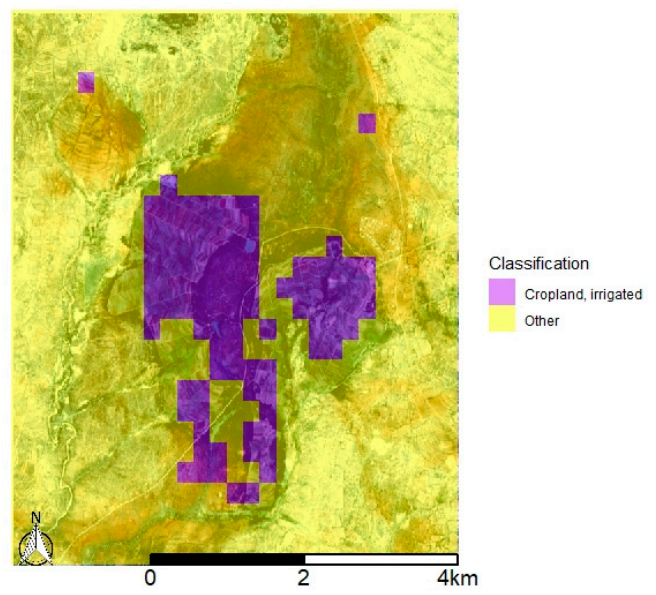
(a)



(b)



(c)



(d)

Figure 6. Cont.

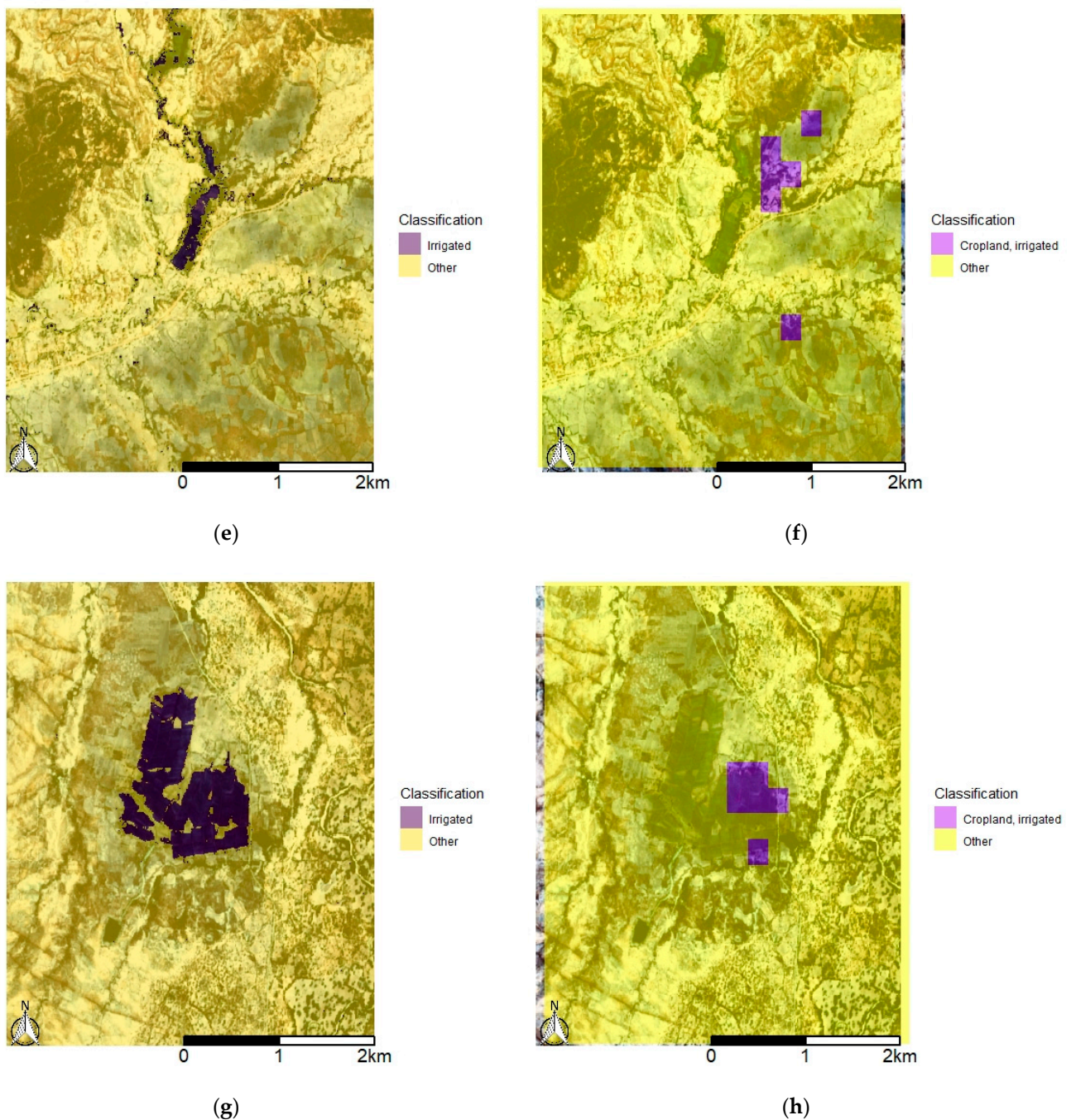


Figure 6. True color geomedian composites overlaid with a transparent layer showing classification results for the random forest models trained on annual datasets for (a) the Ord River Irrigation Area, compared with (b) AQUASTAT's Global Map of Irrigated Areas [37]. Transparent classification layers over true color geomedian composites derived from Sentinel-2 data for (c) Silalatshani, (e) Nabusenga, and (g) Lungwalala are compared with Water Productivity through Open access of Remotely sensed derived data (WaPOR) [40] classification (d,f,h).

Table 5. Confusion matrices, overall accuracy, and Kappa coefficients for the Water Productivity through Open access of Remotely sensed derived data (WaPOR) classification product [40] performance on combined training and validation data (20,000 pixels).

Silalatshani		
	Observed	
Predicted (WaPOR)	Irrigated	Other
Irrigated	489	1919
Other	3644	13,948
Overall accuracy (%)	72.2	
Kappa coefficient	−0.003	
Nabusenga		
	Observed	
Predicted (WaPOR)	Irrigated	Other
Irrigated	0	0
Other	4867	15,133
Overall accuracy (%)	75.7	
Kappa coefficient	0	
Lungwalala		
	Observed	
Predicted (WaPOR)	Irrigated	Other
Irrigated	1174	0
Other	3278	15,548
Overall accuracy (%)	83.6	
Kappa coefficient	0.36	

4. Discussion

4.1. High-Dimensional Geomedians and Statistics for Irrigated Cropland Mapping

The collection of geomedian images overcomes the need for temporal stacking of poorer quality, cloud-contaminated images for land use classification. Furthermore, annual geomedian composites are sufficient for cropland mapping as results show negligible improvement in accuracy compared to using a stack of seasonally derived geomedian composites for the given year (Table 4). Marginal improvement in accuracy is likely to be outweighed by the several-fold increase in the number of predictor variables which may lead to model overfitting. Therefore, the existing annual geomedian and SMAD products reduce the dimensionality of classification problem and represent useful data sources for cropland mapping.

The SMAD, as a high-dimensional temporal variation statistic, is critical to the application of annual geomedian composites to cropland mapping. SMAD featured as a key variable to the accuracy of all classification models tested (Figures 4 and 5). Additionally, visual inspection of true color images against the single band image for SMAD in Figure 2 demonstrates that SMAD corresponds strongly to irrigated croplands. Importantly, SMAD also shows greater deviation from surrounding landscapes than the NDVI. Synchrony of crop phenology with surrounding grasslands in the semi-arid tropics is a key limitation to accurate cropland mapping [9,15]. This is especially evident for the Silalatshani site where leakage from irrigation dams and channels [25], and seepage from irrigation plots appears to cause greenness and high NDVI values for surrounding vegetation (Figure 2). The inability to distinguish between rainfed cropland, irrigated cropland, and other green vegetation has been recognized as a limitation of the NDVI [11]. Therefore, the SMAD has desirable properties for irrigated area classification which overcomes some limitations of the NDVI.

Observation of the true color, SMAD, and NDVI plots for the ORIA reveals some further properties of the SMAD. Fields which appear dark green in the true color image show very high NDVI values but are not discernible from the surrounding landscape in the

SMAD plot (Figure 2). Conversely, fields on the north-west corner of the ORIA show low to moderate NDVI values but very high SMAD values. It was hypothesized that high NDVI values and low SMAD values corresponded to fields with perennial tree crops, and that annual cropping was conducted on fields that showed low to moderate NDVI values and high SMAD values. This observation was confirmed by a farmer in the ORIA [41]. Cyclical land cover changes associated with annual crop production give high SMAD values, and fallow periods mean that perennial vegetation may have higher NDVI values in annual geomedian composites than seasonally or annually cropped fields. This means that in addition to cropland mapping, the simultaneous use of SMAD and NDVI may be useful for crop type classification within irrigation schemes.

Beyond SMAD, differences in variable importance between schemes show that indices vary spatiotemporally in their contribution to distinguishing irrigated cropland (Figures 4 and 5). Figure 4 shows that NDWI was an important variable in classifying the ORIA, Silalatshani, and Lungwalala sites but was relatively unimportant for the Nabusenga site. This may be because periodically flood-irrigated fields in the former sites sit within a relatively dry landscape, whereas Nabusenga sits among a wetter landscape meaning NDWI is not an important distinguisher [42]. Additionally, BSI features as an important variable in the Nabusenga classification model, meaning fallow periods in the irrigation scheme may contribute to distinguishing fields from surrounding green vegetation. The modeling results show that indices vary in their relevance to cropland mapping, even within regions and timeframes. Therefore, indices should be selected for irrigated cropland classification with consideration for their relevance to the area of interest.

While all models tested gave very high accuracy statistics, some areas remain prone to confusion and misclassification. The probability of being classified as irrigated maps in Figure 3 show that pixels in the riparian zones of watercourses are subject to misclassification as irrigated. Vegetation in these locations exhibit similar spectral characteristics to irrigated vegetation, given that seasonal waterlogging is likely to occur due to wet and dry extremes of the semi-arid tropics. Noise reduction based on pixel neighborhood information could remove this misclassification.

4.2. Application to Irrigated Area Mapping

Generating geomedian composites and high-dimensional statistics shows potential for mapping smallholder irrigation schemes in southern Africa. However, the small scale of these schemes still limits the accuracy of mapping. Figure 6 shows a cleaner classified image for the larger scale ORIA than for any of the Matabeleland schemes, despite the higher resolution of the Matabeleland images. Furthermore, indistinct field boundaries and in-field heterogeneity limit image cleanness for Matabeleland sites [15,25]. Within schemes, there may also be a proportion of fields that are inactive and abandoned at any given time; this proportion ranged from 40 to 60% at Silalatshani from 2013–2018 [43]. These factors contribute to scattering in the classified images shown and continue to limit the accuracy of irrigated cropland mapping across southern Africa. Despite the limitations, the method demonstrated is a substantial advancement on current official irrigated area estimates.

The misclassification of all three Matabeleland sites in the Global Map of Irrigated Areas demonstrates the likely underestimation of official irrigated area statistics in Zimbabwe. Sound classification of the ORIA demonstrates that this product is effective over large-scale schemes in the tropics, but the nature of smallholder Matabeleland schemes means other methods are necessary. Importantly, the AQUASTAT product developers acknowledge this in rating the ‘area equipped for irrigation’ map quality as ‘good’ for Australia, and ‘poor’ for Zimbabwe [38].

The continental-scale WaPOR land use classification product performs more accurately than the Global Map of Irrigated Areas, though may still underestimate irrigated area. However, AQUASTAT is generally referred to for official irrigated area statistics over WaPOR [38]. Our results show WaPOR would provide a more accurate estimate for

Zimbabwe, and likely the African continent, and the demonstrated method would be preferable for regional and country scale mapping.

The demonstrated method has potential to be deployed across the semi-arid tropic areas of sub-Saharan Africa, and other global areas where smallholder irrigation schemes predominate. Geomedians and high-dimensional statistics such as the SMAD can be calculated in Digital Earth Africa or other platforms, although additional computational resources may be required for larger-scale applications. Obtaining training and validation data from additional sites would enhance robustness, as would ground-collected data from within irrigation schemes. However, training data for all other land uses may need to be collected from visual inspection of satellite images due to difficulties obtaining non-biased training data for entire regions using ground surveys [12].

While this method accurately maps smallholder irrigation schemes at known locations, an important limitation is that its ability to detect farmer-led, informal irrigation is unquantified. This form of irrigation generally occurs on dambo landforms; wetlands at the headwaters of river systems where water tables are easily accessible [44]. Field sizes are likely to be smaller than in smallholder schemes and fields are unlikely to be contiguous. Cropping is also likely to be integrated with livestock production and husbandry [45]. These factors combine to make detection of farmer-led irrigation difficult [12]. An extensive ground-collected training dataset for informal irrigation would be required to quantify performance on these areas and further develop the method.

5. Conclusions

Annual geomedian composite images combined with high-dimensional statistics of time-series are useful products for irrigated cropland mapping. Digital Earth Australia's geomedian and SMAD products were validated for irrigated cropland classification over the ORIA in north-west Australia. Recalculating annual geomedian composites and the SMAD on Sentinel-2 imagery in Digital Earth Africa generated useful datasets for cropland mapping over Matabeleland, Zimbabwe. Supervised classification using random forest for three pilot sites confirmed that SMAD is a critical variable for irrigated area detection and has advantages over traditionally used vegetation indices such as the NDVI. It may also be useful for differentiating between annual and perennial crops, and detecting cropping activity within a year, season, or other period.

While the Digital Earth Australia analysis-ready products were useful and reduced computation time in this instance, geomedian and high-dimensional statistic calculation packages which allow data collection across continents and satellites may be more valuable. This would negate the need for manual calculation of the SMAD.

The method piloted in this study can be deployed across the entirety of Matabeleland with additional training and validation data. It may also be useful for regional and national mapping in other areas where smallholder irrigation schemes comprise a large portion of irrigated area. Inherent characteristics of smallholder irrigated farming in Zimbabwe continue to limit the accuracy of cropland mapping. However, the application of this method at the regional scale would be an advancement on existing maps and information and has the capacity to reveal previously unrecorded areas of irrigation activity.

Author Contributions: Conceptualization and design of the methodology was jointly developed by both authors. L.J.R. conducted supervision and critical review of the formal analysis and writing, and also contributed interpretation of results. M.J.W. conducted data curation, formal analysis, and original draft preparation. All authors have read and agreed to the published version of the manuscript.

Funding: The research in this paper was associated with the project 'Transforming Irrigation in Southern Africa' largely funded the Australian Centre for International Agricultural Research under grant number LWR-2016-137.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used in this paper is available through Digital Earth Australia, Digital Earth Africa, FAO’s WaPOR, and AQUASTAT, code is available at <https://github.com/mickwelli/Mapping-smallholder-irrigation.git> (accessed on 29 March 2021).

Acknowledgments: We acknowledge the generous assistance of Jamie Pittock, Petra Kuhnert, Roger Lawes, Karthikeyan Matheswaran, and Peter Ramshaw in preparing this work. This research was undertaken while supported by the Australian National University (ANU) University Research Scholarship and a Commonwealth Scientific and Industrial Research Organisation (CSIRO) and ANU Digital Agriculture Supplementary Scholarship through the Centre for Entrepreneurial Agri-Technology.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Vogels, M.F.; De Jong, S.M.; Sterk, G.; Douma, H.; Addink, E.A. Spatio-Temporal Patterns of Smallholder Irrigated Agriculture in the Horn of Africa Using GEOBIA and Sentinel-2 Imagery. *Remote Sens.* **2019**, *11*, 143. [CrossRef]
- Zawe, C.; Madyiwa, S.; Matete, M. *Trends and Outlook: Agricultural Water Management in Southern Africa-Country Report Zimbabwe*; International Water Management Institute: Colombo, Sri Lanka, 2015.
- AQUASTAT. Agricultural Water Use Statistics: Zimbabwe. Available online: <http://www.fao.org/nr/water/aquastat/data/query/index.html?lang=en> (accessed on 1 September 2020).
- Food and Agriculture Organisation. Zimbabwe. Available online: <http://www.fao.org/aquastat/en/geospatial-information/global-maps-irrigated-areas/irrigation-by-country/country/ZWE> (accessed on 10 December 2020).
- Pittock, J.; Ramshaw, P.; Bjornlund, H.; Kimaro, E.; Mdemu, M.V.; Moyo, M.; Ndema, S.; van Rooyen, A.; Stirzaker, R.; de Sousa, W. *Transforming Smallholder Irrigation Schemes in Africa. A Guide to Help Farmers Become More Profitable and Sustainable*; Australian Centre for International Agricultural Research, Australian National University: Canberra, Australia, 2018; Volume 202, p. 64.
- Bjornlund, V.; Bjornlund, H.; Van Rooyen, A.F. Why agricultural production in sub-Saharan Africa remains low compared to the rest of the world—A historical perspective. *Int. J. Water Resour. Dev.* **2020**, *36*, S20–S53. [CrossRef]
- Mwamakamba, S.N.; Sibanda, L.M.; Pittock, J.; Stirzaker, R.; Bjornlund, H.; Van Rooyen, A.; Munguambe, P.; Mdemu, M.V.; Kashaigili, J.J. Irrigating Africa: Policy barriers and opportunities for enhanced productivity of smallholder farmers. *Int. J. Water Resour. Dev.* **2017**, *33*, 824–838. [CrossRef]
- Ozdogan, M.; Yang, Y.; Allez, G.; Cervantes, C. Remote Sensing of Irrigated Agriculture: Opportunities and Challenges. *Remote Sens.* **2010**, *2*, 2274–2304. [CrossRef]
- Lebourgeois, V.; Dupuy, S.; Vintrou, É.; Ameline, M.; Butler, S.; Bégué, A. A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM). *Remote Sens.* **2017**, *9*, 259. [CrossRef]
- Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
- Bousbih, S.; Zribi, M.; El Hajj, M.; Baghdadi, N.; Lili-Chabaane, Z.; Gao, Q.; Fanise, P. Soil Moisture and Irrigation Mapping in A Semi-Arid Region, Based on the Synergetic Use of Sentinel-1 and Sentinel-2 Data. *Remote Sens.* **2018**, *10*, 1953. [CrossRef]
- Hollander, V.R. *Mapping of Farmer-Led Irrigated Agriculture with Remote Sensing: A Case Study in Central Mozambique*; Delft University of Technology: Delft, The Netherlands, 2018.
- Ouattara, B.; Forkuor, G.; Zoungrana, B.J.B.; Dimobe, K.; Danumah, J.; Saley, B.; Tondoh, J.E. Crops monitoring and yield estimation using sentinel products in semi-arid smallholder irrigation schemes. *Int. J. Remote Sens.* **2020**, *41*, 6527–6549. [CrossRef]
- Roberts, D.; Mueller, N.; Mcintyre, A. High-Dimensional Pixel Composites from Earth Observation Time Series. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6254–6264. [CrossRef]
- Landmann, T.; Eidmann, D.; Cornish, N.; Franke, J.; Siebert, S. Optimizing harmonics from Landsat time series data: The case of mapping rainfed and irrigated agriculture in Zimbabwe. *Remote Sens. Lett.* **2019**, *10*, 1038–1046. [CrossRef]
- Belgiu, M.; Csillik, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* **2018**, *204*, 509–523. [CrossRef]
- Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.-T. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 122–130. [CrossRef]
- Roberts, D.; Dunn, B.; Mueller, N. Open Data Cube Products Using High-Dimensional Statistics of Time Series. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2018; pp. 8647–8650.
- Geoscience Australia. DEA Surface Reflectance Geomedian (Landsat). Available online: <https://cmi.ga.gov.au/data-products/dea/140/dea-surface-reflectance-geomedian-landsat> (accessed on 1 September 2020).
- Geoscience Australia. DEA Surface Reflectance Median Absolute Deviation (Landsat). Available online: <https://cmi.ga.gov.au/data-products/dea/346/dea-surface-reflectance-median-absolute-deviation-landsat> (accessed on 1 September 2020).

21. Davies, A.; Strickland, G.; Moulden, J.; Yeates, S. NORpak Ord River Irrigation Area Cotton Production and Management Guidelines for the Ord River Irrigation Area (ORIA) 2007. Available online: <http://www.insidecotton.com/xmlui/handle/1/204> (accessed on 1 September 2020).
22. Ash, A.; Gleeson, T.; Hall, M.; Higgins, A.; Hopwood, G.; MacLeod, N.; Paini, D.; Poulton, P.; Prestwidge, D.; Webster, T.; et al. Irrigated agricultural development in northern Australia: Value-chain challenges and opportunities. *Agric. Syst.* **2017**, *155*, 116–125. [CrossRef]
23. Ramshaw, P. *Locations of TISA Irrigation Schemes in Zimbabwe*; Wellington, M., Ed.; Australian National University: Canberra, Australia, 2020.
24. DAFWA. Ord River Development and Irrigated Agriculture. Available online: <https://www.agric.wa.gov.au/assessment-agricultural-expansion/ord-river-development-and-irrigated-agriculture> (accessed on 12 February 2021).
25. Moyo, M.; Van Rooyen, A.; Chivenge, P.; Bjornlund, H. Irrigation development in Zimbabwe: Understanding productivity barriers and opportunities at Mkoba and Silalatshani irrigation schemes. *Int. J. Water Resour. Dev.* **2017**, *33*, 740–754. [CrossRef]
26. Digital Earth Australia. Calculating Band Indices. Available online: https://docs.dea.ga.gov.au/notebooks/Frequently_used_code/Calculating_band_indices.html (accessed on 1 September 2020).
27. Rouse, J.; Haas, R.; Deering, D.; Schell, J.A.; Harlan, J. Monitoring vegetation systems in the Great Plains with ERTS. In Proceedings of the 3rd ERTS Symposium, Washington, DC, USA, 10–14 December 1973; pp. 309–317.
28. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
29. Rikimaru, A.; Roy, P.; Miyatake, S. Tropical forest cover density mapping. *Trop. Ecol.* **2002**, *43*, 39–47.
30. Digital Earth Africa. s2_l2a. Available online: https://explorer.digitalearth.africa/products/s2_l2a (accessed on 1 September 2020).
31. Kirill Kouzoubov. `_geomedian.py`. Available online: https://github.com/opendatacube/odc-tools/blob/develop/libs/algos/odc/algos/_geomedian.py (accessed on 1 September 2020).
32. Digital Earth Africa. Band Indices. Available online: https://training.digitalearthafrica.org/en/latest/session_4/01_band_indices.html (accessed on 1 September 2020).
33. Pebesma, E.J.; Bivand, R.S. Classes and methods for spatial data in R. *R News*, 1 February 2005; Volume 5, 1–21.
34. Kuhn, M. caret: Classification and regression training. R package version 6.0-86. *R News*, 20 March 2020; 1–223.
35. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News*, 30 November 2001; Volume 2, 1–6.
36. Food and Agriculture Organisation. AQUASTAT Maps. Available online: <https://data.apps.fao.org/aquamaps/> (accessed on 15 January 2020).
37. Food and Agriculture Organisation. Global Map of Irrigation Areas (GMIA). Available online: <http://www.fao.org/aquastat/en/geospatial-information/global-maps-irrigated-areas/> (accessed on 14 February 2021).
38. Siebert, S.; Henrich, V.; Frenken, K.; Burke, J. *Global Map of Irrigation Areas Version 5*; Rheinische Friedrich-Wilhelms-University: Bonn, Germany, 2013.
39. Food and Agriculture Organisation. *WaPOR Database Methodology: Level 1*; Remote Sensing for Water Productivity Technical Report; Food and Agriculture Organisation: Rome, Italy, 2018.
40. Food and Agriculture Organisation. WaPOR 2.1. Available online: https://wapor.apps.fao.org/home/WAPOR_2/1 (accessed on 12 February 2021).
41. Bolten, R. *Distribution of Crop Types in the Ord River Irrigation Area*; Wellington, M., Ed.; Australian National University: Canberra, Australia, 2021.
42. Savva, A.P.; Frenken, K. *Planning, Development, Monitoring and Evaluation of Irrigated Agriculture with Farmer Participation*; Food and Agriculture Organisation: Harare, Zimbabwe, 2002.
43. Quirke, G. *Assessing Smallholder Irrigated Plot-Use in Zimbabwe*; Australian National University: Canberra, Australia, 2019.
44. Bell, M.; Faulkner, R.; Hotchkiss, P.; Lambert, R.; Roberts, N.; Windram, A. *The Use of Dambos in Rural Development, with Reference to Zimbabwe*; Loughborough University, University of Zimbabwe: Leicestershire, UK, 1987.
45. Nyamadzawo, G.; Wuta, M.; Nyamangara, J.; Nyamugafata, P.; Chirinda, N. Optimizing dambo (seasonal wetland) cultivation for climate change adaptation and sustainable crop production in the smallholder farming areas of Zimbabwe. *Int. J. Agric. Sustain.* **2014**, *13*, 23–39. [CrossRef]