

Article

Mitigating Imbalance of Land Cover Change Data for Deep Learning Models with Temporal and Spatiotemporal Sample Weighting Schemes

Alysha van Duynhoven  and Suzana Dragičević 

Spatial Analysis and Modeling Laboratory, Department of Geography, Simon Fraser University,
8888 University Drive, Burnaby, BC V5A1S6, Canada

* Correspondence: alyshav@sfu.ca; Tel.: +1-778-782-5628

Abstract: An open problem impeding the use of deep learning (DL) models for forecasting land cover (LC) changes is their bias toward persistent cells. By providing sample weights for model training, LC changes can be allocated greater influence in adjustments to model internal parameters. The main goal of this research study was to implement and evaluate temporal and spatiotemporal sample weighting schemes that manage the influence of persistent and formerly changed areas. The proposed sample weighting schemes allocate higher weights to more recently changed areas based on the inverse temporal and spatiotemporal distance from previous changes occurring at a location or within the location's neighborhood. Four spatiotemporal DL models (CNN-LSTM, CNN-GRU, CNN-TCN, and ConvLSTM) were used to compare the sample weighting schemes to forecast the LC changes of the Columbia-Shuswap Regional District in British Columbia, Canada, using data obtained from the MODIS annual LC dataset and other auxiliary spatial variables. The results indicate that the presented weighting schemes facilitated improvement over no sample weighting and the common inverse frequency weighting scheme for multi-year LC change forecasts, lowering errors due to quantity while reducing overall allocation error severity. This research study contributes to strategies for addressing the characteristic imbalances of multitemporal LC change datasets for DL modeling endeavors.

Keywords: land cover change; spatiotemporal deep learning; geospatial data imbalance; sample weights; inverse temporal distance weighting; spatiotemporal distance weighting; temporal convolutional networks; recurrent neural networks; convolutional neural networks; land cover data imbalance



Citation: van Duynhoven, A.; Dragičević, S. Mitigating Imbalance of Land Cover Change Data for Deep Learning Models with Temporal and Spatiotemporal Sample Weighting Schemes. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 587. <https://doi.org/10.3390/ijgi11120587>

Academic Editor: Wolfgang Kainz

Received: 18 September 2022

Accepted: 20 November 2022

Published: 23 November 2022

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



Copyright: © 2022 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/>).

1. Introduction

The abundance of openly available multitemporal remote sensing data continues to expand, accelerating studies of land change and pursuits of data-driven modeling techniques such as machine learning (ML) and deep learning (DL) [1]. These approaches have circumvented the need to encode nonlinear relationships between numerous variables in land change applications [2–4]. ML and DL methods are designed to learn patterns given training data, labels, and a loss function [5]. In particular, DL methodologies applied to multitemporal datasets have demonstrated favorable outcomes for LC classification and forecasting [5–7]. For example, recurrent neural networks (RNNs) are useful for extrapolating from timeseries data [5,6], and their combination with convolutional neural networks (CNNs) facilitates the extraction of spatial features from each timestep [7]. However, many DL models are sensitive to data imbalances, exhibiting biases toward samples belonging to majority groups characterizing the dataset [8–10]. Persistent areas dominate most land cover datasets [11], presenting a large challenge for the application of DL methodologies due to the relatively small amount of changed areas [12]. Therefore, mitigation of imbalance

between changed and persistent areas is required for DL methodologies to be effective for LC change forecasting.

Previous studies have applied sampling strategies and data augmentation techniques to address the imbalance of changed and unchanged areas. For example, balanced sampling schemes were implemented to provide ML models equal numbers of changed and persistent samples using random sampling [13,14] and an iterative bootstrap sampling approach [15]. However, DL models benefit from larger amounts of data [4], inspiring geographic dataset augmentation procedures including adaptations of the synthetic minority over-sample technique (SMOTE) [16], transformations of data samples such as rotations and flips [17], and synthetic sample generation with generative adversarial networks (GANs) [18]. Despite this, such approaches have not been adapted to the dynamic characteristics of land change timeseries data, nor do they maintain directional spatial relationships in the case of manual transformations or the real-world spatiotemporal context of geographic data samples. Other strategies aimed at mitigating data imbalance include cost-sensitive methods such as sample weighting [9]. By allocating increased penalties to significant or minority samples, the spatial phenomena of interest have greater influence on the learned parameters of DL models [19]. Using sample weighting schemes, scarcer or anomalous geographic phenomena can be allocated higher importance in model training procedures without data-level manipulations.

Sample weights were previously explored to improve forecasts of underrepresented peak air pollution concentrations [20], to increase the importance of nearby samples for location recommendations [21], and to boost multi-class LC classification accuracy through inverse frequency weighted categorical cross-entropy [22]. Temporal and spatiotemporal distance decay were also implemented for modeling house prices with local regression techniques [23] and video target tracking with spatiotemporal DL models [24] to diminish the influence of locations or features based on both temporal and spatial proximity. While variations of temporal and spatiotemporal sample weighting schemes were explored for other model types and applications, it remains unknown how temporally and spatiotemporally weighted samples would affect the capacity of spatiotemporal DL models to forecast LC changes versus the commonly used inverse frequency weighting scheme.

The main objective of this research study was to propose and evaluate the potential effect of temporally and spatiotemporally weighted samples for managing the imbalances that inhibit DL model capacity to forecast LC changes. While it is acknowledged that per-class disparities characterize a second dimension of imbalance impacting the capacity of DL models to forecast LC changes [12], this research study focuses on LC changes overall. The new sample weighting schemes are implemented with the intent of ensuring more recently changed areas and changed locations with transitioning neighborhoods are more impactful in model training. This is achieved through inverse distance weighting schemes which allocate changed samples variable importance based on temporal and spatiotemporal proximity. Meanwhile, all persistent LC samples are assigned lower importance. The experimental cases compare four types of spatiotemporal DL models trained using the proposed sample weighting schemes versus those trained with no sample weights and inverse frequency weighted samples. The assessment uses change-focused measures to highlight potential improvements associated with the proposed sample weighting schemes for 5-year LC change forecasts. The goal is to determine if trends observed across all model types are maintained for a multistep LC forecast.

2. Methodology

2.1. Study Area and Datasets

The study area selected for this research work was the Columbia-Shuswap Regional District in British Columbia, Canada (Figure 1a). In 2016, the population of this region was 51,366, and Salmon Arm was its largest city [25]. The MODIS dataset was utilized to provide annual LC data at 500 m spatial resolution, covering 2001 to 2020 [26]. The vast majority of the region was characterized by forests and shrublands, with these two

LC classes spanning 45% and 40% of the area in 2001, respectively. Because per-class LC changes are not the priority of this research study, eight aggregated LC classes were used to preserve the characteristics of the region. The net change that occurred between 2001 and 2020 was 21,202 km², with 2076 km² of average annual change (Figure 1b).

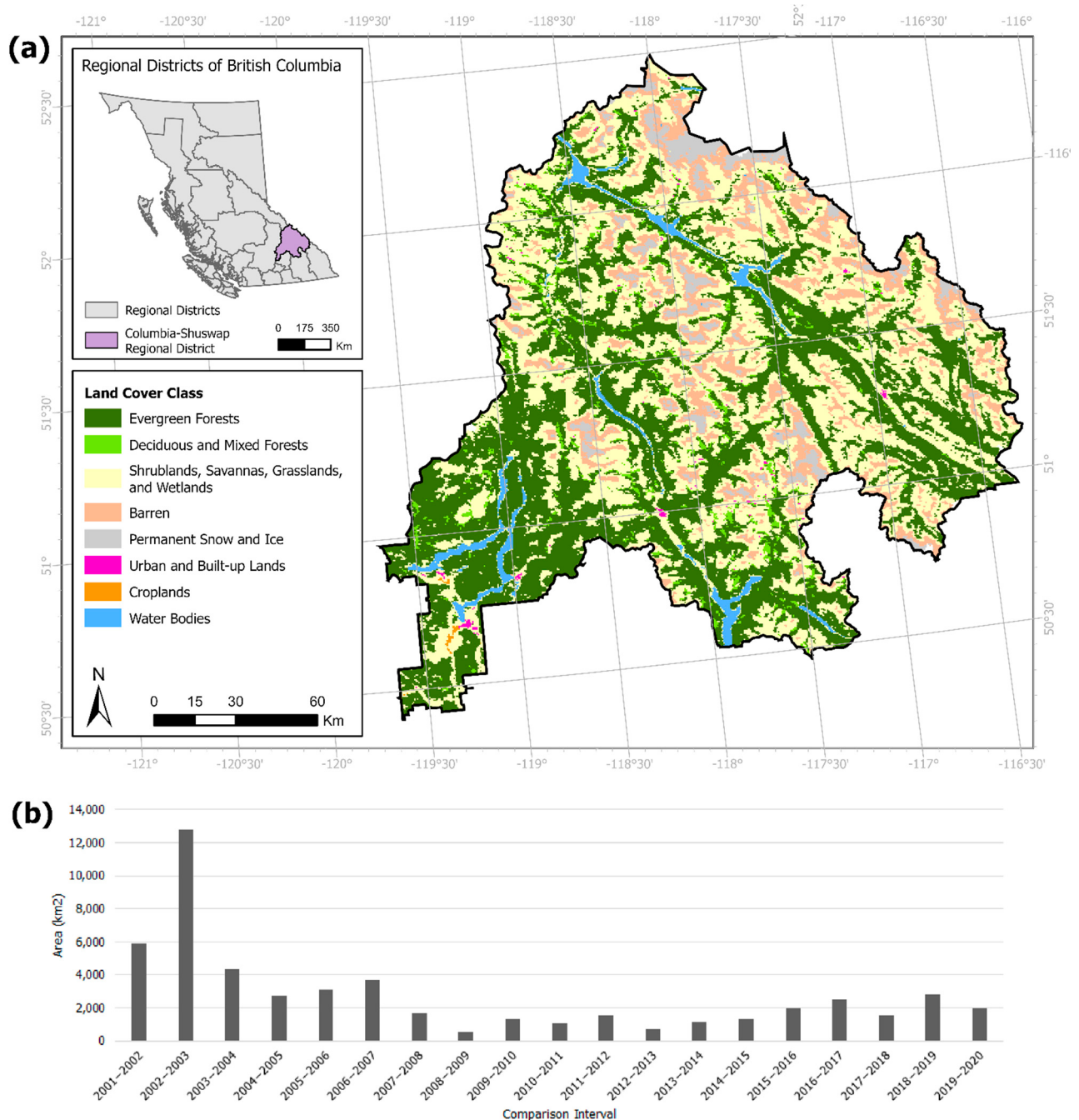


Figure 1. Study area of the Columbia-Shuswap Regional District, British Columbia, with (a) land cover for 2001 and (b) annual net land cover changes for the region from 2001 to 2020. Data are displayed with the NAD 1983 BC Environment Albers projected coordinate system.

In addition to multitemporal LC data, static spatial variables provide auxiliary information about drivers of land changes. Topographic variables provided are elevation and slope [27,28], alongside accessibility variables derived by calculating the Euclidean distances to population centers [6], roads [7], railways [3], protected areas [7,29], agricultural reserves [30], and lakes and rivers [29]. The ASTER global digital elevation model was

acquired for the topographic variables [31] and the proximity variables were computed with respect to data layers available from Statistics Canada [32,33]. All data layers were resampled to the MODIS dataset using nearest neighbor interpolation for categorical layers and bilinear interpolation for the continuous layers, since the LC dataset exhibits the coarsest spatial resolution. The LC data and auxiliary variables are reprojected to the NAD 1983 BC Environment Albers projected coordinate system, preserving planar area measurements. Data preprocessing procedures were completed in the ArcGIS Pro 2.9.1 software before deriving samples and sample weights in the next steps described.

2.2. Capturing Neighborhood Effects in Land Cover Data Samples

Before developing sample weights, the training data samples were extracted with respect to each cell of the study area. “Neighborhood effects” refer to the impact of changes and structure of phenomena in the vicinity of a sample location [34]. The LC at a sample location is highly dependent on the states and changes that occur around it, making neighborhood effects important to consider in LC change forecasting [4]. To integrate neighborhood effects in this research study, the LC composition surrounding each location at every timestep was included in each data sample. Considering the raster geospatial data layers described in Section 2.1, data samples were acquired for every cell comprising the study region using a Moore neighborhood configuration. A Moore neighborhood captures N cells provided a range parameter r , where $N = (2r + 1)^2$ cells [35]. Samples were obtained considering each cell as a central cell, with neighborhoods capturing cells within range r from the central location across all timesteps. The neighborhood size can also be referred to with respect to the areal dimensions captured by each sample, $M \times M$, where M refers to how many cells comprise the longest edge of the central cell’s neighborhood.

Data samples for the Columbia-Shuswap Regional District included both spatiotemporal LC and auxiliary spatial variables. The spatiotemporal LC obtained for each location was of size $T \times M \times M \times C$, where T signifies the number of timesteps, and C represents the number of LC classes. The spatial variables were provided in the form of $M \times M \times V$, where V denotes the number of variables. The neighborhood dimensions used in this research study were set to 9×9 (or $r = 4$ cells in every direction from the central cell), according to other studies employing DL models [36–38] and findings suggesting land changes are captured well when M is less than 5 km with 500 m spatial resolution data [39]. Therefore, the 9×9 neighborhood captured the LC and spatial variables for a 20.25 km² area around each cell in the study region.

2.3. Model Specifications

The models implemented in this research study included those adapted for the spatiotemporal samples provided alongside auxiliary spatial variables as described in Section 2.2. Specific DL models called recurrent neural networks (RNNs) are useful for extracting patterns from sequential or timeseries data [37]. RNN implementations exhibit different variations in architectures, including long short-term memory (LSTM) [40] and gated recurrent units (GRUs) [41]. Temporal convolutional networks (TCNs) provide an alternative to RNNs for sequence modeling, using convolutional neural networks (CNNs) that operate on the temporal dimension of a geospatial dataset [42]. However, on their own, these sequence modeling techniques do not extract spatial correlations within data samples. Instead, capturing neighborhood effects first with traditional CNNs [43] and providing the outputs to sequence models (i.e., LSTM, GRU, and TCN) implements hybrid spatiotemporal models, which have been used to capture spatial and temporal patterns of land change [7,44]. Despite demonstrated integrations, these hybrid models do not capture explicit spatial and temporal relationships simultaneously. As such, convolutional LSTM (ConvLSTM) was formulated to accommodate spatiotemporal relationships directly, providing benefits over CNN-LSTM implementations [45].

The four spatiotemporal DL model types implemented in this research study were CNN-LSTM, CNN-GRU, CNN-TCN, and ConvLSTM. Each model was equipped with two

input branches, accommodating the spatiotemporal LC sequences and the auxiliary spatial variables, respectively. This is similar to how 3D and 2D variables have been integrated in DL models in other geographic applications [46]. The general structure of the branched model implemented is shown in Figure 2.

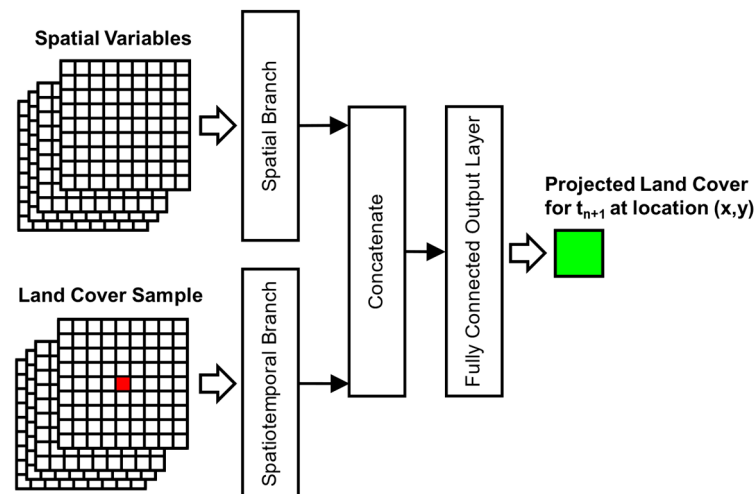


Figure 2. Overview of the basic branched model structure used to accommodate the 9×9 spatiotemporal land cover sample sequences and the auxiliary spatial variables. The “spatial branch” is implemented with CNN layers. The “spatiotemporal branch” implementation varies according to the model type, characterized by CNN-LSTM, CNN-GRU, CNN-TCN, or ConvLSTM layers. Location x, y in the land cover sample is denoted in red, indicating the central cell of the neighborhood.

For all model types, the spatial variable input branch (denoted as “spatial branch” in Figure 2) was implemented with two sets of two convolution layers with 32, 32, 64, and 64 filters, respectively. After each set of two CNN layers, a 2×2 max pooling operation was applied. Next, the spatiotemporal LC input branch (denoted as “spatiotemporal branch” in Figure 2) was implemented according to the model type. The implementations for CNN-LSTM, CNN-GRU, and CNN-TCN were adapted from previous studies [46,47]. To implement the spatiotemporal branch of the model, spatial relationships from each timestep of the LC data samples were first extracted using CNN. Two sets of two convolution layers were followed by 2×2 max pooling operations, where the CNN layers were parameterized by 32, 32, 64, and 64 filters, respectively, with the *ReLU* activation function. The outputs of the CNN operations were flattened and provided to the temporal model component characterized by LSTM, GRU, or TCN of the respective models. The CNN-LSTM and CNN-GRU models each featured two recurrent layers of 32 and 128 neurons using the *tanh* activation function, respectively, based on previous implementations [36]. The CNN-TCN model featured two layers with 32 and 128 filters, with the *ReLU* activation function. The ConvLSTM implementation had two ConvLSTM layers, featuring 32 and 128 filters, respectively, with the *ReLU* activation function. The kernel size for all CNN and ConvLSTM layers was set to 3×3 , as seen in previous research studies [36,48]. Before concatenating outputs of each model branch, a dropout factor of 10% was applied. Dropout regularization randomly drops a specified percentage of neurons with the intent of preventing the model from overfitting [49]. Lastly, the fully connected output layer featured nine neurons and the Softmax activation function, which outputted the probability of each LC class label [17]. For all models, the LC and spatial variable data samples were provided with neighborhood effects explained in Section 2.2. The LC data for each sample location were supplied directly to the spatiotemporal input branch in the form of $T \times M \times M \times C$, and auxiliary spatial variable data were provided in the form of $M \times M \times V$ to the spatial variable input branch.

2.4. Categorical Cross-Entropy Loss

Supervised DL models were trained by optimizing weights of a network with respect to some objective function or loss function [50]. Using the gradient descent (or stochastic gradient descent) algorithm, the aim is to minimize or maximize the amount of error between the real-world and projected value. Forecasting multi-class LC is formulated as a classification problem, requiring an objective function equipped for this probabilistic task. The categorical cross-entropy (CCE) is commonly used for classification problems, facilitating learning of multiple classes by backpropagating error with respect to forecasted values and real-world LC classes [51]. The CCE function computed with respect to sample i is expressed as follows as per previous studies [22]:

$$CCE_i = - \sum_{c=1}^C \hat{y}_c \log(y_c),$$

where LC class labels are one-hot encoded vectors, denoted by $\hat{y}_c$ and y_c , where $\hat{y}_c$ represents the real-world LC of class c , and y_c represents the forecasted LC for the location. C represents the number of LC classes.

In the computation of error with respect to the CCE function, every sample is given the same amount of influence on model updates. By adding a weight factor to the CCE function, samples can be given varying error penalties that consequently have differing impacts on model weight adjustments during model training procedures [52]. This is expressed as follows:

$$WCCE_i = - \sum_{c=1}^C \hat{y}_c \log(y_c) \cdot w_i,$$

where w_i represents the sample weight of sample i . Values for w_i assume the sample weights described in the next section.

2.5. Calculating Temporal and Spatiotemporal Sample Weights

This research study provides three new LC sample weighting schemes to improve model capacity to capture changes. The proposed temporal and spatiotemporal sample weights were implemented to bypass the need for any manual assignment of cost values to samples, which is challenging in many real-world applications [9]. The sample weighting schemes were based on temporal and spatiotemporal proximity to changes, achieved using the inverse distance weighting (IDW) approach [53]. For the task of LC change forecasting, the aim was to first identify whether a location underwent a change, then refine the sample weight based on whether the change happened recently or whether the location's neighborhood underwent a recent change. Accommodating temporal decay addresses temporal heterogeneity by giving higher importance to recently changed cells, thus decreasing the impact of changes that occurred long ago. For instance, notably larger quantities of changed areas were observed between 2002 and 2003 (Figure 1b). Under the proposed sample weighting schemes, locations that changed but remained persistent following 2003 were weighted less than those that underwent changes more recently. Additionally, even if the change occurred earlier, a sample of a formerly changed location with an actively changing neighborhood was weighted with higher importance.

The effects of the three new sample weighting schemes are compared against unweighted samples and the widely used inverse frequency weighting scheme. Therefore, the five sample weighting schemes implemented in this research study were as follows:

1. Unweighted (base case or "none"), where no sample weights were used;
2. Binary weights (BW), where a traditional inverse frequency weighting scheme used the inverse frequency of changed versus persistent sample counts to assign sample weights;
3. Temporal weighting scheme 1 (TW1), where the inverse temporal distance weight was computed with respect to the most recent change of the central cell;

4. Temporal weighting scheme 2 (TW2), where the inverse temporal distance weight was computed with respect to the most recent change of the cell's neighborhood;
5. Spatiotemporal weighting scheme (STW), where the inverse spatiotemporal distance weight was calculated with respect to the most recent change that occurred within the neighborhood of the central cell.

To compute the sample weights of the training data samples, changes were considered within the temporal extent of 2001 to 2014 (Figure 1b). The three steps to implement the weighting schemes are as follows:

Step 1. Identify whether a change has occurred at the sample location, or central cell of the neighborhood (Figure 2). If the number of change incidents was one or greater from 2001 to 2014, the cell was considered as changed.

Step 2. Compute the inverse frequency weight according to the counts of persistent and changed cells. Similar to another research study [54], the initial weights of the samples were calculated on the basis of their overall type. For LC data, the initial weight of the changed samples was $b_{c_i} = P / (P + C)$, and the weight of persistent samples was $b_{p_i} = C / (P + C)$, where the number of changed cells is denoted by C , and the number of persistent cells is denoted by P . Persistent samples were assigned non-zero weights because they were still important to the learned model structure. Therefore, persistent sample weights (w_{p_i}) assumed the values of b_{p_i} and required no further updates. This concludes the calculations required to implement the BW scheme, while temporal and spatiotemporal variation was added to changed sample weights in Step 3 for the TW1, TW2, and STW schemes.

Step 3. With the effect of persistent cells managed in Step 2, the sample weight calculations for the TW1, TW2, and STW schemes were applied to the changed sample weights (w_{c_i}) as a function of the temporal and spatial variation occurring at the central cell and within its neighborhood over time. To implement the temporal weighting schemes (TW1 and TW2), the temporal distance from the most recent year of the training sample was computed with respect to the most recent change at the central cell (d_{cc}) and to the year of the most recent change occurring in the neighborhood of the central cell (d_{cn}), respectively. For the spatiotemporal weighting scheme (STW), the spatiotemporal distance was computed from the location and year of the latest central cell in a sample and the location and year of the most recently changed cell in its neighborhood (d_{cn}^{ST}). Table 1 shows the formulations of TW1, TW2, and STW. By expanding the weighting schemes to consider changes taking place within a cell's neighborhood, the TW2 and STW schemes increased the weight of change samples with recent nearby transitions. This means that more dynamic change samples had greater influence on model parameter adjustments during the training procedure using the TW2 and STW schemes. The IDW power or exponent parameter was set to one because the spatial resolution of this research study was coarse and there were limited timesteps available, as well as to ensure weight values associated with of historical changes were non-zero. The resulting sample weights are presented in Figure 3. No manual adjustments or normalization techniques were applied to the weights to further influence sample importance. Therefore, the maximum possible weight values were constrained by Step 2, where the inverse proportion of changed and persistent cells was calculated. This led to the maximal weight value of 0.9 (Figure 3).

Table 1. Sample weight scheme applied to samples identified as changed.

Sample Weight Scheme for Changed Locations	Formula	Description
Binary weight (BW)	$w_{c_i} = b_{c_i} = P / (P + C)$	The inverse proportion of changed versus persistent samples
Cell-change temporal weight (TW1)	$w_{c_i} = b_{c_i} * \frac{1}{d_{cc}} = b_{c_i} * \frac{1}{t_n - t_{cc}}$	Temporal distance (d_{cc}) between most recent year (t_n) and the year of the most recent change event of the central cell (t_{cc})

Table 1. Cont.

Sample Weight Scheme for Changed Locations	Formula	Description
Neighborhood-change temporal weight (TW2)	$w_{c_i} = b_{c_i} * \frac{1}{d_{cn}} = b_{c_i} * \frac{1}{t_n - t_{cn}}$	Temporal distance (d_{cn}) from the most recent year (t_n) and the year of change event occurring in the neighborhood of the central cell (t_{cn})
Spatiotemporal weight (STW)	$w_{c_i} = b_{c_i} * \frac{1}{d_{cn}^{ST}}$ $= b_{c_i} * \frac{1}{\sqrt{(x - x_{cn})^2 + (y - y_{cn})^2 + (t_n - t_{cn})^2}}$	Spatiotemporal distance (d_{cn}^{ST}) from the central cell (x, y, t_n) to the nearest changed cell in its neighborhood (x_{cn}, y_{cn}, t_{cn})

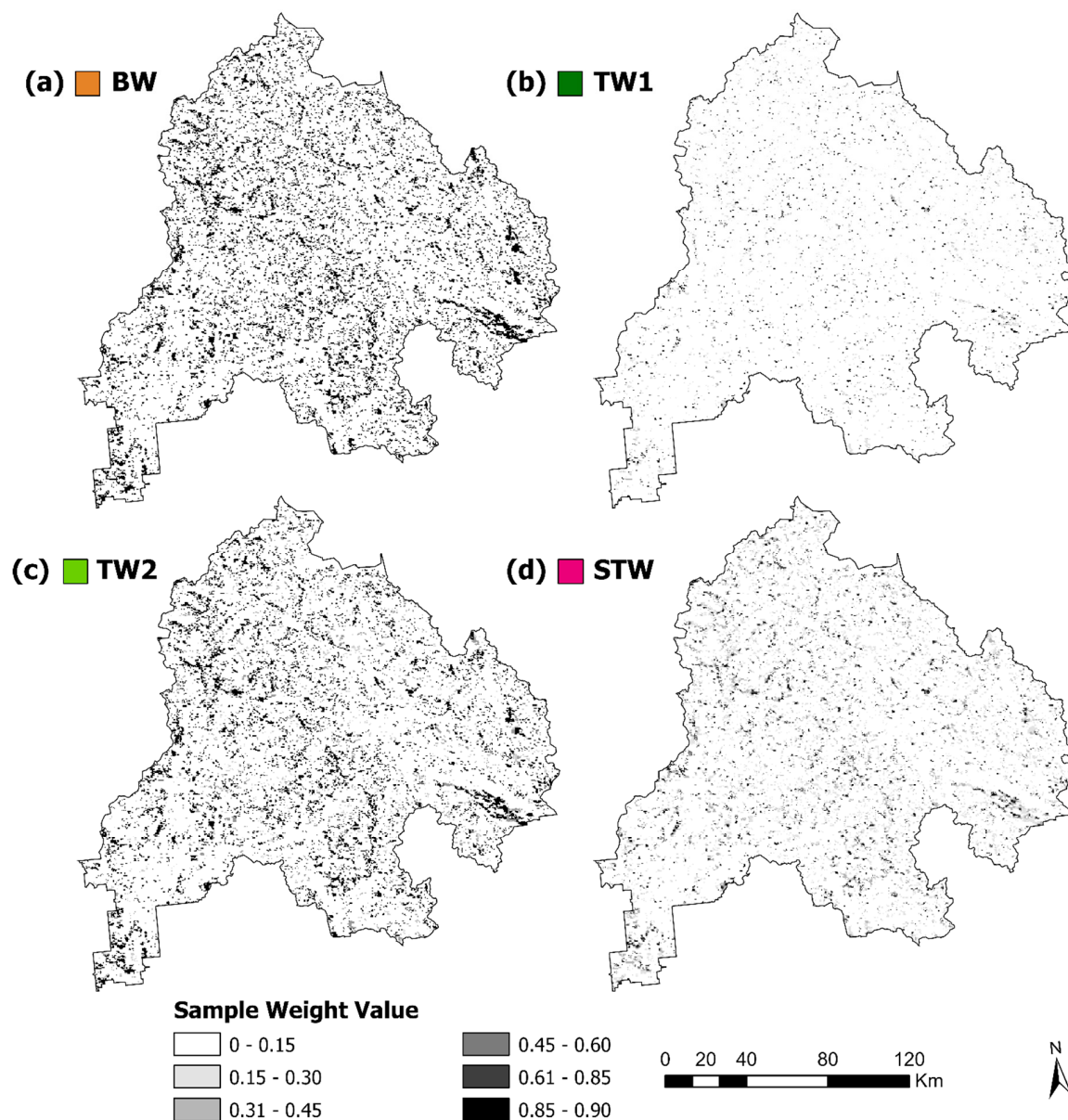


Figure 3. Maps depicting the variation of sample weights under each weighting scheme. The sample weights assigned to every location represent (a) binary weight (BW), (b) cell-change temporal weight (TW1), (c) neighborhood-change temporal weight (TW2), and (d) spatiotemporal weight (STW). Color swatches adjacent to each sample weight scheme correspond to those used in figures in subsequent sections.

2.6. Model Assessment

The measures selected for evaluating model performance in this study focused on LC changes instead of overall accuracy, since high accuracy values cannot be used to express the capacity of a model to forecast changes [55]. Instead, the assessment included figure of merit (FOM), producer's accuracy (PA), and user's accuracy (UA) [56], which were identified as more appropriate for evaluating LC change forecasts. FOM provides a measure of agreement between forecasted changes and real-world changes, calculated with respect to "hits", "misses", "wrong hits", and "false alarms". Computations for PA and UA involve subsets of the terms comprising the FOM measure. PA indicates the proportion of area that a model forecasts correctly as changed with respect to real-world changes, whereas UA suggests the amount of correctly forecasted changes versus all projected changes. The correct changes ("hits"), missed changes ("misses"), incorrect changes ("false alarms"), and correct persistence ("correct rejections") components of agreement and disagreement used to compute the FOM measure are also reported separately to showcase quantity and allocation of changes [57].

To conduct an error analysis, the error due quantity, error due to allocation, and allocation error distance [58] were calculated for each model and sample weight combination. Allocation error distance (AED) provides information about the severity of allocation errors, where the distance between the real-world locations and locations of erroneous forecasts were averaged with respect to each LC class [58]. Because LC class or category imbalance was not addressed in this research study, AED was computed with respect to all allocation errors (AED_{overall}), the largest classes (AED_{large}), the classes deemed "medium-sized" (AED_{medium}), and the smallest classes (AED_{small}). The purpose was to identify where the most severe AEs stemmed from. The large class size category encompassed evergreen forests, shrublands and savannas, and barren land, comprising 91.1% of the study area. The medium class size category captured permanent snow and ice, water bodies, and deciduous forests, covering 8.7% of the study area. The smallest classes were urban and built-up lands and croplands, occupying less than 1% of the study area.

Python 3.9.1, GDAL 3.3, and Rasterstats 0.17.0 were used to implement the allocation error distance approach conveyed in prior work [58] and the other change-focused measures. In this research study, the FOM measure was used to identify six top-performing models based on the 2016 LC forecast. The performance of these models was then compared across the multi-year LC change forecasts to determine if trends were maintained. The mapped FOM components also supported a visual assessment for the forecasts produced by the best-performing models identified with respect to their capacity to forecast changes.

2.7. Experiment Settings

The models were implemented to the specifications described in Section 2.3 with Python 3.9.1 [59], the Keras API [60], TensorFlow 2.5.0 [61], and an open-source implementation of TCN (Keras-TCN 3.4.0) [62]. The CCE loss function described in Section 2.4 was employed in all models. To train the models, the batch size was set to 128, and the Adam optimizer was used with an initial learning rate of 0.01 [7]. Early stopping and learning rate reductions (using the "ReduceLROnPlateau" function from TensorFlow) were utilized to ensure that model training ceased or that the learning rate was decreased when performance gains were negligible [63].

In this study, the multi-year forecasts were generated using a "rolling-window" strategy according to the implementation in previous work [64]. Instead, "sequence-to-sequence" forecasting was demonstrated to circumvent the propagation of error across projected timesteps [44], as this approach would reduce the already limited timesteps available and lead to short sequences of MODIS LC data, which impeded DL models for LC change forecasting in a prior study [65]. With the "rolling-window" strategy, a multi-year LC change projection was produced using the previous forecast as the next timestep of the testing sequence. The LC data spanning 2001–2014 were used to populate the training dataset, while 2015 was withheld for model validation and 2016–2020 was used for model

testing. Samples from the 2001–2014 training dataset were obtained using a rolling-window approach based on a previous application [66], where 10 timesteps comprised the training sequence, with the following timestep withheld as the training label. Therefore, the training data sequences spanned $t_n, t_{n+1}, \dots, t_{n+9}$, with t_{n+10} as the training label. Each data sample comprised spatiotemporal LC data and static spatial variables with the neighborhood specifications expressed in Section 2.2. Samples for every location are provided to train the model with the sample weights described in Section 2.5 and shown in Figure 3. The four model types were trained considering each sample weight scheme (none, BW, TW1, TW2, and STW), where “none” refers to the experimental combinations or base case in which no sample weighting scheme was applied.

3. Results

3.1. Multi-Year Change Assessment

The model assessment described in Section 2.6 was used to quantify the impact of the proposed sample weighting schemes across the four model types. With respect to the FOM measure obtained for 2016, the top six model and sample weight combinations were CNN-TCN_{STW}, CNN-GRU_{STW}, ConvLSTM_{TW2}, ConvLSTM_{STW}, CNN-LSTM_{TW1}, and ConvLSTM_{TW1} (Figure 4). Following the FOM values obtained by these model and sample weight combinations, there was a 19.7% difference between the next model and sample weight combinations (Figure 4) and a substantial drop in FOM values observed for the base case (denoted as “none”). All sample weighting schemes facilitated improved FOM measures over the base case for the 2016 LC change forecasts, regardless of model type. The BW scheme was associated with consistently improved FOM values compared to the base case, although the top combination using BW (CNN-GRU_{BW}) was 27.6–32.6% lower than the top six performers identified. In addition, the STW scheme was associated with higher FOM values for three of the four model types (CNN-GRU, CNN-TCN, and ConvLSTM) (Figure 4). FOM values obtained with the TW2 scheme also enabled improved performance versus the BW scheme for the same three models. The TW1 scheme worked well for CNN-LSTM and ConvLSTM, but reduced FOM values for CNN-GRU and CNN-TCN.

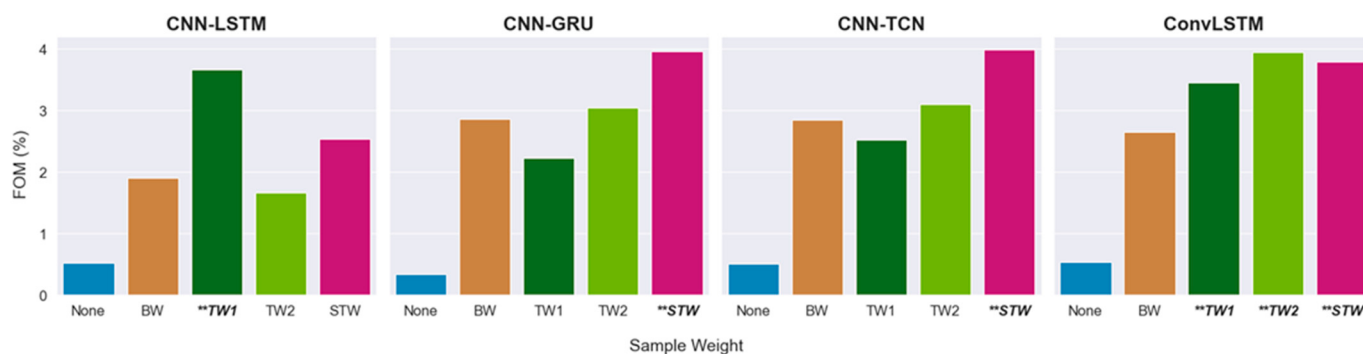


Figure 4. Figure of merit (FOM) values obtained for each model type and sample weight combination for the 2016 LC forecast. The model and sample weight combinations with the top six FOM values are denoted with the prefix **.

After identifying the top six models with respect to FOM values computed for the 2016 forecasts, it was observed that these models maintained the highest FOM measures over the 5-year projection (Figure 5a). Meanwhile, the FOM values remained low for the base case, showing that no sample weights produce underperforming LC change models, regardless of model type. The BW scheme maintained consistent effects on all model types, although CNN-TCN_{BW} and CNN-GRU_{BW} showed a slight increase over time. In contrast to the initial trends seen in Figure 4, ConvLSTM_{TW1} surpassed ConvLSTM_{STW} and ConvLSTM_{TW2} after the 2017 projection, while CNN-TCN_{STW}, CNN-GRU_{STW}, and CNN-LSTM_{TW1} continued to yield the highest FOM measures (Figure 5a). The PA measures

computed with respect to changed areas followed the same trends (Figure 5b), showing that the proportion of forecasted versus real-world changes was highest in forecasts obtained from the top six model and sample weight combinations. In contrast, the UA showed a trend dissimilar to those seen with FOM and PA measures (Figure 5c). The highest UA measures were obtained by CNN-LSTM_{None} and ConvLSTM_{None}, showing that the proportion of correctly simulated changes versus all projected change was high. The low quantities of changes projected with no sample weights boosted UA measures over time, as these measures were inflated by small amounts of projected LC change, as indicated by “hits”, “false alarms”, and “wrong changes” for both 2016 and 2020 (Figure 6a,b). Notably, ConvLSTM_{TW2} and ConvLSTM_{STW} exhibited higher UA measures than the other top six models. This indicates that these combinations forecasted higher amounts of correctly changed areas out of all forecasted changes and reduced incorrectly changed areas, despite not attaining the maximal FOM or amount of hits (Figures 5c and 6). Overall, CNN-TCN_{STW} yielded the highest amount of correctly changed area of all the 2020 LC change forecasts (Figure 6). The highest quantity of hits or correctly changed area for each model type was associated with TW1, TW2, and STW, except for CNN-GRU_{TW1}. Additionally, CNN-TCN_{STW} and ConvLSTM_{TW1} produced the highest number of false alarms of the top six models.

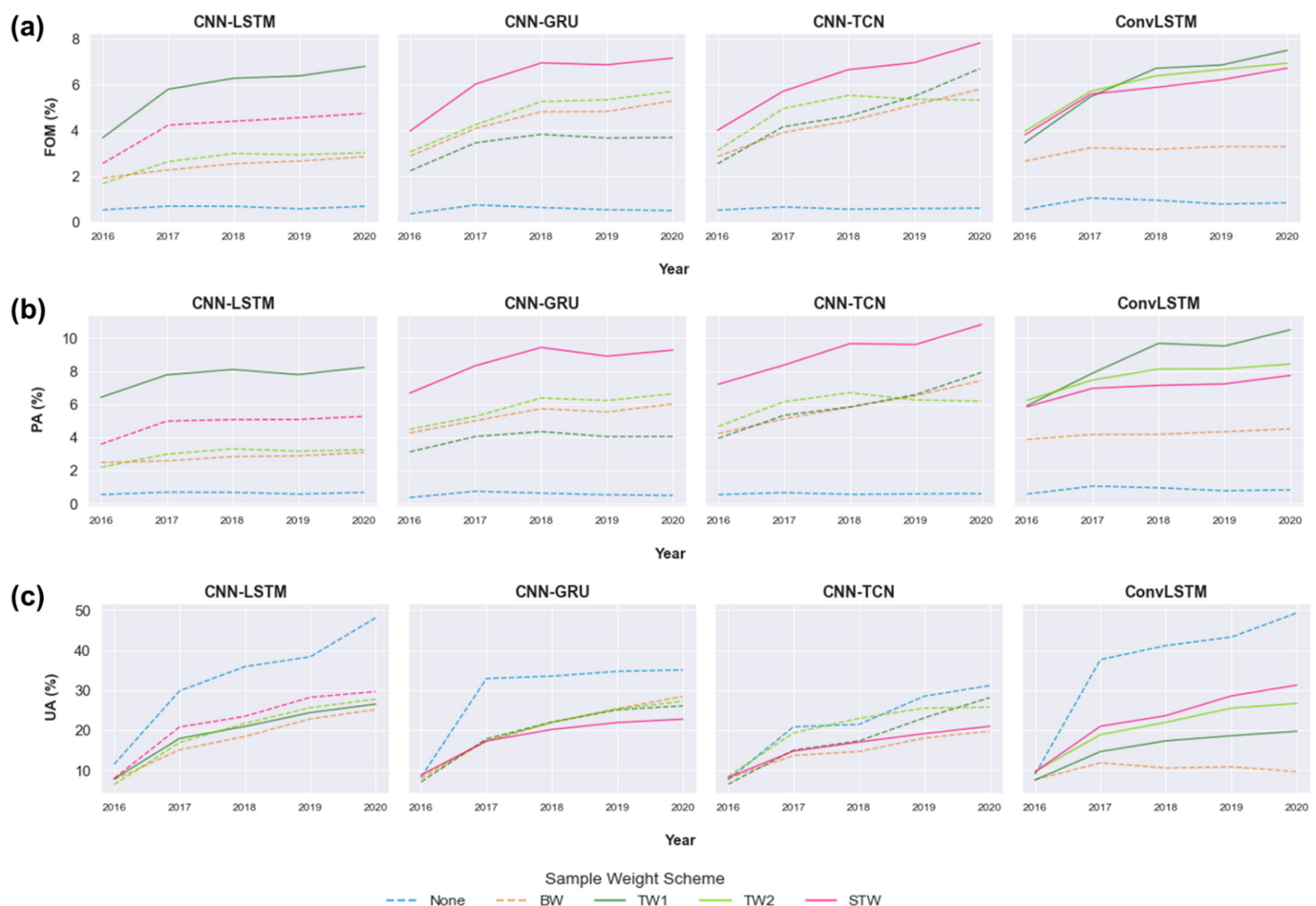


Figure 5. Values obtained for each model type and sample weight combination with respect to cumulative changes from 2016 to 2020 with measures: (a) figure of merit (FOM), (b) producer's accuracy (PA), and (c) user's accuracy (UA). The model and sample weight combinations with the top six FOM values are indicated by bold lines.

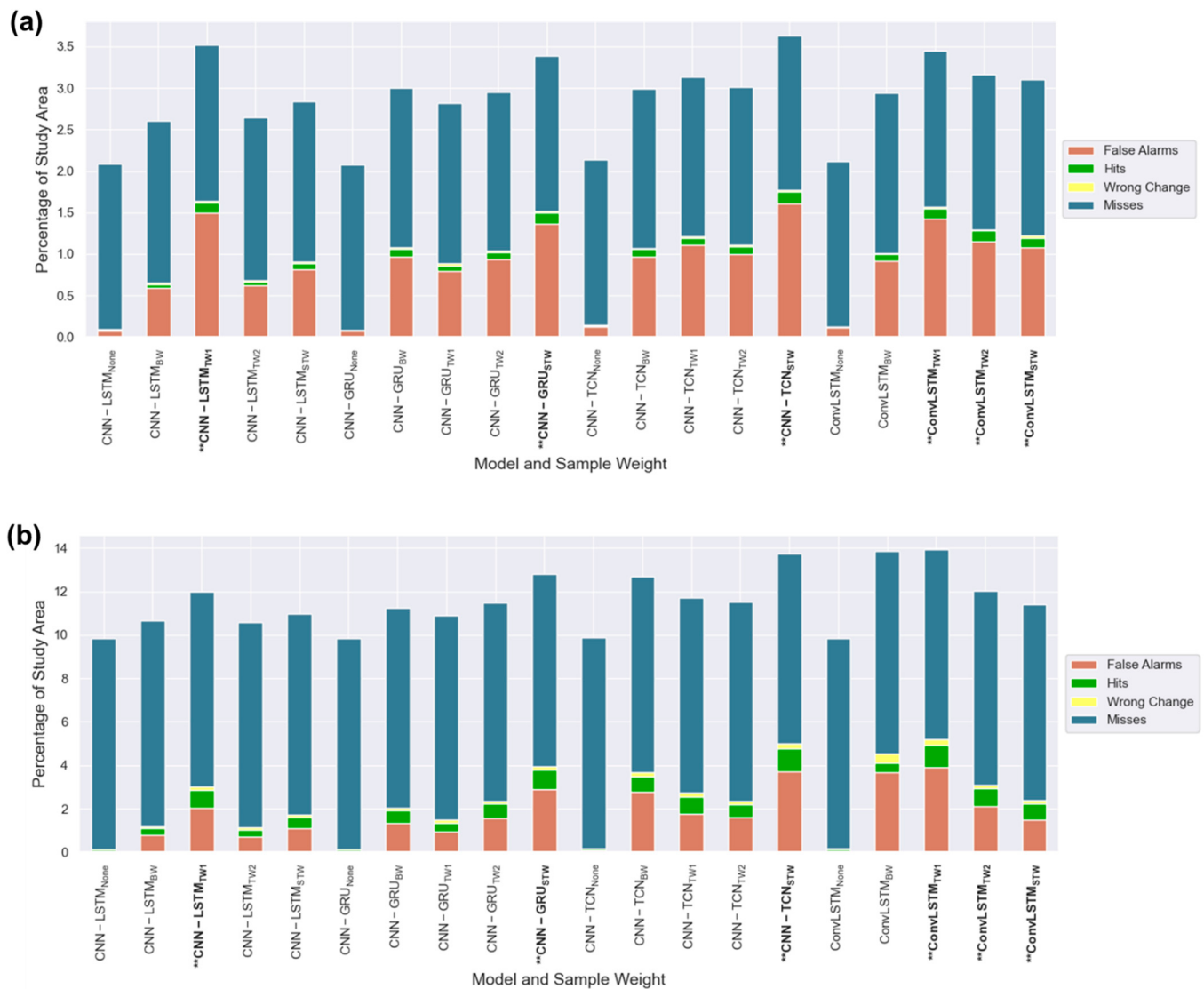


Figure 6. Components of agreement and disagreement as a percentage of the study area for (a) the 2016 forecast and (b) the 2020 forecast. The model and sample weight combinations with the top six FOM values (Figure 4) are denoted with the prefix **.

3.2. Multi-Year Error Analysis

The highest amount of EQ was observed for every model's base case, in which no sample weights were used (Figure 7a). Meanwhile, three of the top six models (CNN-LSTM_{TW1}, CNN-GRU_{STW}, and CNN-TCN_{STW}) provided the lowest EQ for their respective model types, showing that each projected more realistic quantities of changes. ConvLSTM_{TW1} preserved the lowest EQ for the ConvLSTM model type, while EQ values attributed to ConvLSTM_{TW2} and ConvLSTM_{STW} gradually exceed those of ConvLSTM_{BW}. Conversely, the base case attained the lowest EA measures, corresponding to the minimal quantities of false alarms and wrong changes observed (Figures 6 and 7b). Of the top six models, CNN-LSTM_{TW1}, CNN-GRU_{STW}, CNN-TCN_{STW}, and ConvLSTM_{TW1} forecasted the highest amounts of changed area allocated incorrectly, while ConvLSTM_{STW} projected the lowest EA for each step of the 5-year projection.

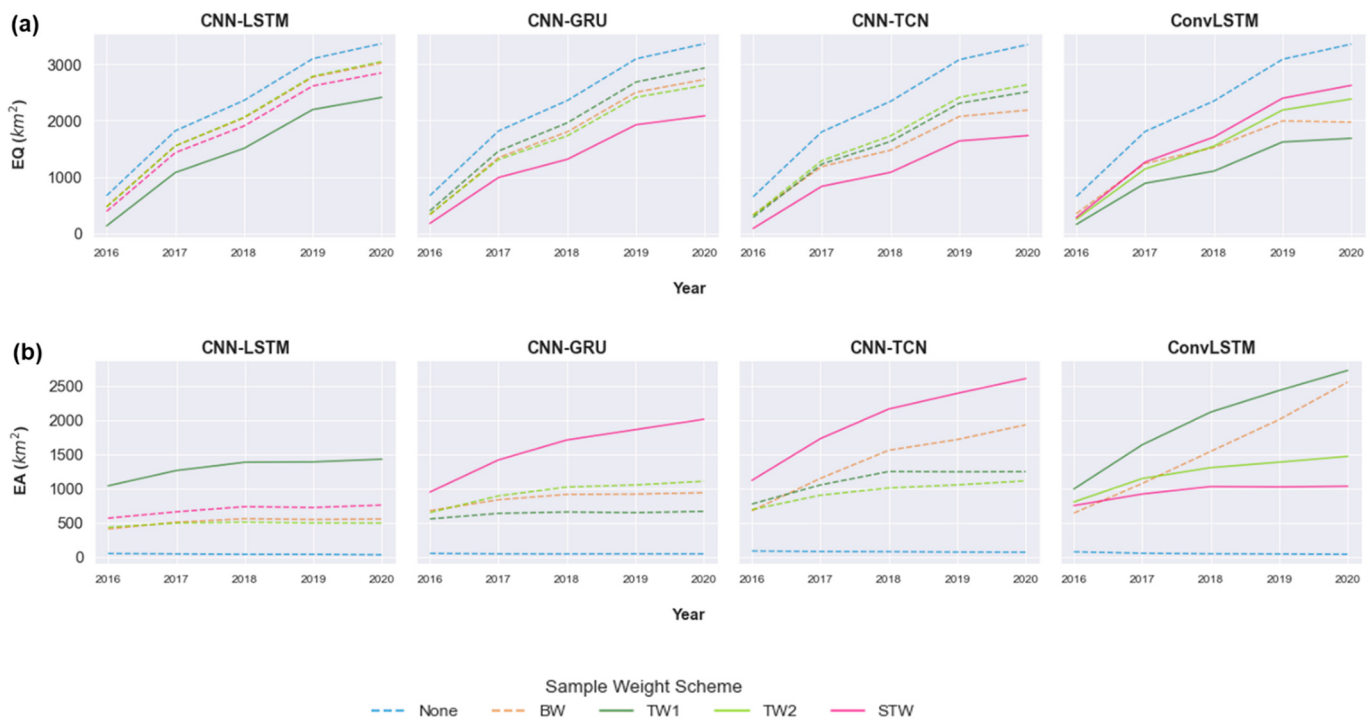


Figure 7. Cumulative errors from 2016 to 2020 for (a) error due to quantity (EQ) and (b) error due to allocation (EA). The model and sample weight combinations with the top six FOM values for the 2016 forecast are indicated by bold lines.

Considering the distance of erroneous allocations to real-world LC category locations, the $AED_{overall}$, AED_{large} , AED_{medium} , and AED_{small} maintained similar trends across the 5-year forecast (Figure 8). $AED_{overall}$ and AED_{large} values showing the smallest or nearest allocation errors were associated with the 2016 forecast by ConvLSTM_{TW2} (Figure 8a,b). For $AED_{overall}$, a notable deviation was observed for CNN-TCN_{BW}, where the overall allocation error severity exceeded the unweighted base case. The top six models produced overall allocation errors generally closer to the real-world areas than the unweighted base case in the projections for 2017–2020. The same trend was also noted for AED_{large} , except for the 2017 forecast produced by ConvLSTM_{TW1}. However, it was observed that the erroneous large class allocations were either corrected or were more agreeable with the real-world 2018 LC allocations, as the AED_{large} value decreased for the next timestep. It should be noted that the spread of $AED_{overall}$ and AED_{large} values was not substantial, indicating that overall allocation errors and allocation errors attributed to the largest LC classes were marginal with respect to the spatial resolution of the dataset. The AED_{medium} showed that the top six models produced allocation errors nearer to the real-world classes than the unweighted base case, except for the ConvLSTM_{STW} model (Figure 8c). While AED_{small} values computed from the 2016 forecast were zero for ConvLSTM_{STW} and ConvLSTM_{TW2}, larger deviations were observed between 2018 and 2020. This implies agricultural or built-up areas were forecasted far from their real-world allocations.

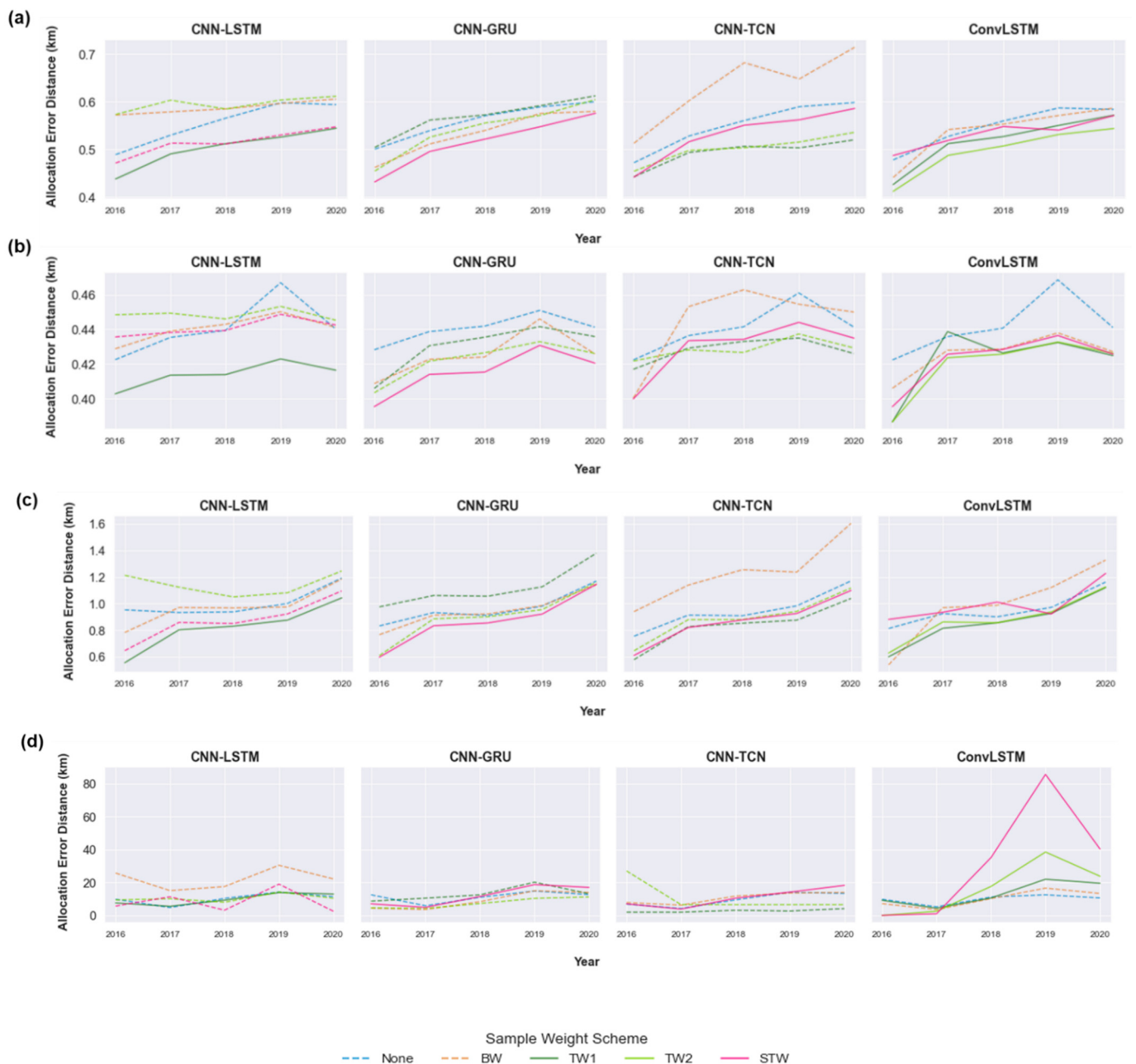


Figure 8. Average allocation error distance from real-world allocations in 2016–2020 considering (a) AED_{overall}, (b) AED_{large}, (c) AED_{medium}, and (d) AED_{small}. The model and sample weight combinations with the top six FOM values are indicated by bold lines.

3.3. Visual Assessment

For the visual assessment, CNN-TCN_{STW} and ConvLSTM_{STW} were considered because the first exhibited the highest FOM, while the latter forecasted the smallest number of false alarms among the top six models. For the 2016 forecasts, the false alarms and misses exhibited a “salt-and-pepper” appearance (Figure 9a,c). The CNN-TCN_{STW} exhibited a few small clusters of false alarms in the west and southwest of the study region. However, for the 2020 LC forecast, CNN-TCN_{STW} showed more distinctive clusters of false alarms surrounded by correct changes. Areas that were persistent that were forecasted incorrectly as changed appeared to be near to or at the locations that had higher sample weights (Figures 3d and 9b). The 2020 forecast from ConvLSTM_{STW} showed smaller and fewer clusters of false alarms. It was shown previously that misses contributed the most errors

(Figure 6), and missed changes were shown to be visually consistent across both 2020 projections (Figure 9b,d).

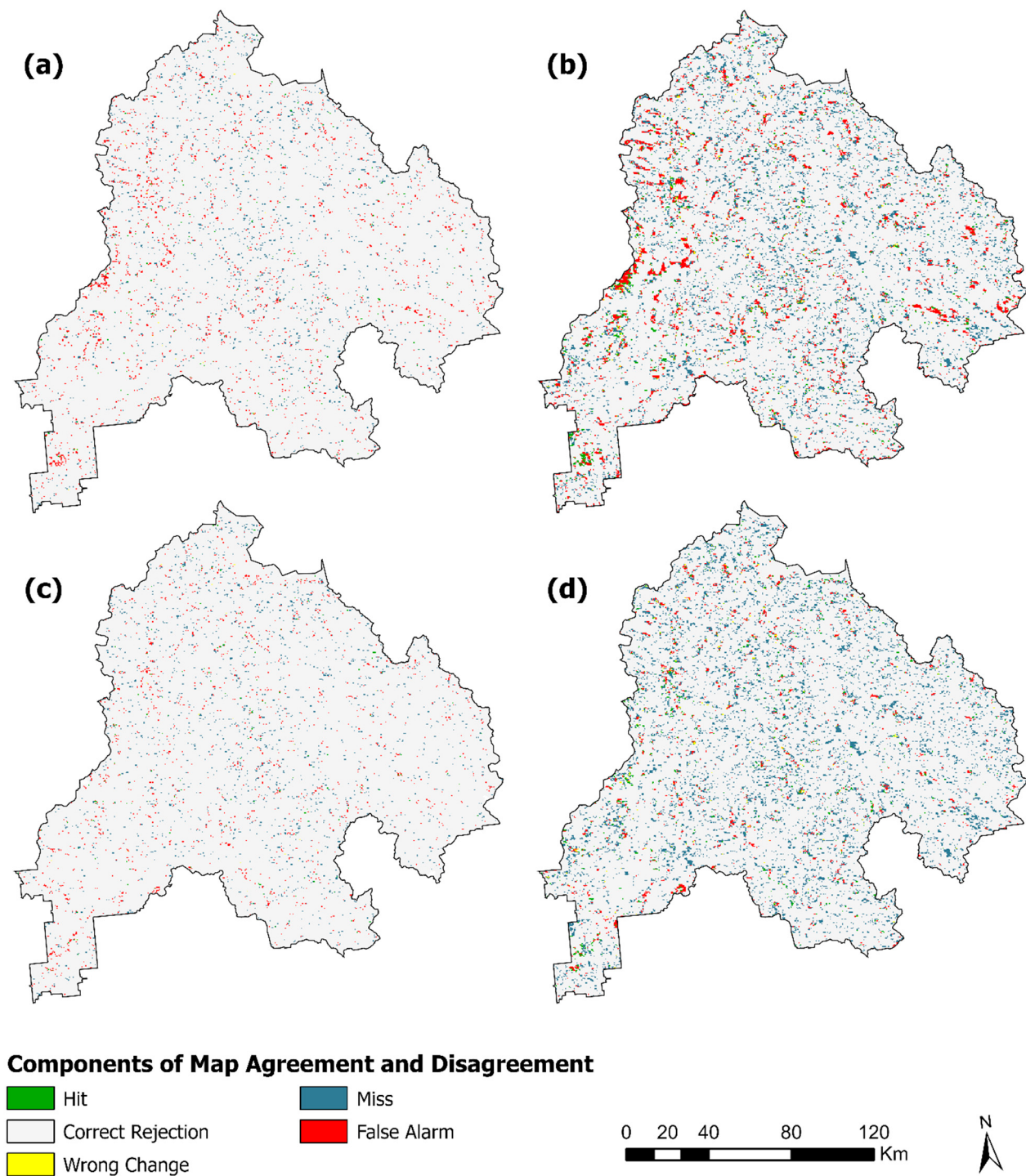


Figure 9. Maps representing the locations of obtained values for components of agreement and disagreement of forecasted LC with CNN-TCN_{STW} for (a) 2016 and (b) 2020 and with ConvLSTM_{STW} for (c) 2016 and (d) 2020.

4. Discussion

The results obtained in this research study demonstrated that, regardless of the DL model chosen, the proposed sample weighting schemes were beneficial for forecasting LC changes. The TW1, TW2, and STW schemes generally showed improvement over the traditional BW scheme. In particular, the STW scheme improved the FOM measures compared to the BW scheme across all model types explored. With respect to FOM values obtained for 2016 LC forecasts, six model and weight combinations were identified: CNN-TCN_{STW}, CNN-GRU_{STW}, ConvLSTM_{TW2}, ConvLSTM_{STW}, CNN-LSTM_{TW1}, and ConvLSTM_{TW1}. It was observed that the highest FOM measures associated with these six combinations were maintained for the 5-year projection. Furthermore, the top six models attained the highest PA measures. The UA measures highlighted some interesting trends among the weighting schemes, notably for ConvLSTM. The gradually increasing UA observed for ConvLSTM_{STW} with respect to cumulative forecasted changes indicated consistent increases in correct changes versus all projected changes, while forecasting fewer false alarms or incorrect LC transitions. Overall, STW was most beneficial for CNN-GRU and CNN-TCN across all experiments, while TW1, TW2, and STW similarly benefited ConvLSTM models. The TW1 scheme benefited CNN-LSTM most for all timesteps of the 5-year projection, and the FOM values associated with ConvLSTM_{TW1} also surpassed those of ConvLSTM_{TW2} and ConvLSTM_{STW} after 2017. This aligns with the observation that sample weight values and variations of TW1 and STW are more similar than those characterizing TW2 or BW schemes (Figure 3). However, the similarities did not expound the approximate 2% difference between FOM measures of CNN-LSTM and CNN-GRU with the TW1 and STW schemes (Figure 5a). This requires future investigation of model parameters, structure, and regularization techniques with respect to the weighting schemes. TW2 was associated with only one of the top six models (ConvLSTM_{TW2}) and did not facilitate similar performance to models trained with STW for the other model types. This may be because TW2 sample weight values were more like those of the traditional inverse frequency weighting (BW) scheme observed across the study area (Figure 3a,c).

Considering the types of errors associated with the top six models, CNN-TCN_{STW} forecasted the highest amount of correctly changed areas while forecasting the most false alarms for the 2016 projection. However, CNN-TCN_{STW} forecasted the second most false alarms for the 2020 projection, superseded by ConvLSTM_{TW1} projecting the most persistent area incorrectly as changed. If maximizing the agreement of changed areas for the 5-year forecast was the sole objective, CNN-TCN_{STW} would have been regarded as the “best” model and sample weight combination. Meanwhile, ConvLSTM_{STW} forecasted 86.3% fewer false alarms with less error due to quantity (EQ) after 5 years, which was apparent in the visual assessment (Figure 9). With respect to the AED measures, the expectation was that all sample weighting schemes would help mitigate allocation error severity overall and with respect to the largest classes, which was typically the case. The AED measures indicated that the worst allocation errors were generally associated with medium and small size classes, which makes sense because no techniques were used to address the LC class imbalance problem and since optimizing per-class change allocations was not the objective of this study. However, the AED_{medium} indicated that the top six models generally produced less severe allocation errors than the unweighted base case. The exception to this trend was ConvLSTM_{STW}, suggesting that one or more of the medium-sized classes were not well-allocated by this model. AED_{small} measures also indicated ConvLSTM_{STW} and ConvLSTM_{TW2} forecast built-up or agricultural areas far from their real-world allocations. This outcome was expected, as this second dimension of imbalance characterizing LC datasets adds challenges for multi-class change forecasting with DL models. As such, future research studies should investigate further combinations of sample weights with class weights or the focal loss function [67] to reduce the quantity and allocation error distance with respect to non-majority LC categories.

From the visual assessment, it was observed that the spatial variations of weights for the STW scheme (Figure 3d) were associated with the agreeing and disagreeing allocations

of LC changes forecasted with CNN-TCN_{STW} and ConvLSTM_{STW} (Figure 9). The CNN-TCN_{STW} combination appeared more sensitive to the higher sample weight values in some areas, which was more noticeable in the 2020 forecast (Figure 9b). For instance, the projected hits and false alarms were typically seen clustered around locations with larger sample weights (Figures 3d and 9b). This may suggest that persistent samples still require reduced weights or undersampling strategies to manage their influence on learned model parameters. This outcome may also hinge upon the model type, since the 2020 forecast produced by ConvLSTM_{STW} appeared less influenced by areas where the sample weight computed for the location was high. Conversely, in each resulting map, both sporadic and clustered areas of missed changes existed at similar spatial locations for areas with low weight values. However, given the appearance of more clustered hits and false alarms in the 2020 projections (Figure 9b,d), the STW scheme may be beneficial for future works seeking to manage properties like spatial variability [68].

The goal of this research study was to improve LC change forecasting with DL models by proposing and evaluating a sample weighting scheme that integrated temporal and spatiotemporal proximity from geographic timeseries data. The effect of sample weights derived from temporal and spatiotemporal distance from recent changes was unexplored for LC change forecasting with DL models, which are highly influenced by mostly persistent areas. It is also acknowledged LC change rates are typically small [69]. Missed changes were the most common error attributed to all model and sample weighting scheme combinations (Figure 6). Yet, previous modeling endeavors attained 3.38% correct changes using 15-year temporal resolution data [27]. Therefore, 1.05% of correctly changed areas achieved by the CNN-TCN_{STW} for the 5-year projection may have been somewhat comparable. Additionally, the FOM measure had a positive linear relationship with net observed changes [56]. For example, a land change model with 10-year temporal resolution obtained an FOM value of 9%, with less than 5% of the study area undergoing changes during that time period [28]. This corroborates the values obtained in this research study, in which CNN-TCN_{STW} attained an FOM value of 7.8% with only 3.8% of the region undergoing changes from 2016 to 2020. Future work should consider further optimization of models and sample weighting schemes for datasets with finer spatial resolutions, expanded neighborhoods, longer LC sequences, and noisy datasets. Climatic variables are also important drivers of LC change [70] and should also be integrated to further enhance model capacity to forecast changes alongside the sample weighting schemes. Additional adjustments to computed sample weights may also be beneficial, as previous work identified that low weights were negligible in their effect on model training procedures [20]. Lastly, combinations of data augmentation [16] and the removal or undersampling of persistent samples [14] may be beneficial alongside the proposed sample weighting schemes. Nevertheless, the risk of removing potentially important LC data samples remains an open problem.

5. Conclusions

This research study explored the potential of proposed inverse temporal distance and inverse spatiotemporal distance sample weighting schemes for LC change forecasting with spatiotemporal DL models and multitemporal LC data available for the Columbia-Shuswap Regional District of BC. The rationale for training sample weighting was to decrease the influence of samples that underwent changes long ago while increasing the influence of changed samples that underwent more recent changes at the central location or within its neighborhood. With temporally and spatiotemporally weighted LC change samples, model forecasts showed consistent improvements in FOM and FOM components of agreement and disagreement versus the unweighted base case (“none”) and the traditional inverse frequency weight (BW) schemes. While allocation errors remain an outstanding problem for the DL models, the proposed sample weighting schemes reduced the average distance of allocation errors to real-world LC categories overall and with respect to large and medium-sized LC classes. Based on the findings of this research study, the proposed sample weighting schemes enabled markedly better performance compared to using unweighted

samples for LC change forecasting with spatiotemporal DL models. Of all sample weighting schemes, STW was consistently associated with improved FOM and PA measures for all model types versus the traditional BW scheme. It is recommended to explore, analyze, and further adjust the TW1, TW2, and STW sample weighting schemes with respect to other datasets and spatiotemporal DL model configurations in future studies.

Given the typically slow and scarce nature of LC change events that impede direct applications of DL models, this research study contributes to advancing strategies used to mitigate data imbalance for LC change forecasting and other geographic phenomena. It was previously unknown how temporal and spatiotemporal sample weighting schemes contribute to DL model capacity to forecast LC changes. As such, this research study introduced simple sample weights based on temporal and spatiotemporal proximity to change events, demonstrating improvements in DL model capacity to forecast LC changes. This was accomplished without randomly discarding potentially useful examples or augmenting synthetic or transformed samples that defy real-world spatial context and relationships. This research study can benefit any DL modeling endeavor that deals with LC data or imbalanced geographic timeseries data. With increasingly available open-source data-driven modeling approaches featuring sample weight parameter options, cost-sensitive learning techniques can be achieved without complex programmatic modifications. The new sample weighting schemes can also contribute to improving non-timeseries DL models or ML models more generally by allocating less significance to older samples obtained for geographic applications such as projecting urban growth, forest change, or agricultural expansion.

Author Contributions: Conceptualization, formal analysis, investigation, methodology, writing—original draft, and writing—review and editing, Alysha van Duynhoven and Suzana Dragičević; funding acquisition and supervision, Suzana Dragičević; software, Alysha van Duynhoven. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada through the Postgraduate Scholarship-Doctoral Grant (PGS-D) and the Discovery Grant (RGPIN-2017-03939) awarded to the first and second authors, respectively.

Data Availability Statement: The publicly available datasets used in the current study include MODIS/Terra + Aqua Land Cover Global Land Cover <https://lpdaac.usgs.gov/products/mcd12c1v006> (accessed 20 July 2022), ASTER Digital Elevation Model <https://asterweb.jpl.nasa.gov/gdem.asp> (accessed 20 July 2022), Population Centers, Water, and Digital Boundary Files of Canadian Provinces and Territories <https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm> (accessed 21 July 2022), Canadian Census Road Network <https://open.canada.ca/data/en/dataset/57d5ffae-3048-4a19-9b4c-eab12f6322c5> (accessed 21 July 2022), Parks and Protected Areas <https://catalogue.data.gov.bc.ca/dataset/parks-and-protected-areas-regional-boundaries> (accessed 21 July 2022), and Agricultural Land Reserve (<https://catalogue.data.gov.bc.ca/dataset/alc-alr-polygons>) (accessed 21 July 2022).

Acknowledgments: The authors are grateful for the full support of this study by the Natural Sciences and Engineering Research Council (NSERC) of Canada through the Postgraduate Scholarship-Doctoral Grant (PGS-D) and the Discovery Grant (RGPIN-2017-03939) awarded to the first and second authors, respectively. The authors thank the three anonymous reviewers for the positive, constructive, and valuable comments and suggestions. The authors are also thankful to Compute Canada WestGrid Computing Facilities for enabling the experiments to be run for this study.

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

References

1. Shi, W.; Zhang, M.; Zhang, R.; Chen, S.; Zhan, Z. Change detection based on artificial intelligence: State-of-the-art and challenges. *Remote Sens.* **2020**, *12*, 1688. [\[CrossRef\]](#)
2. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Prabhat Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Aburas, M.M.; Ahamad, M.S.S.; Omar, N.Q. Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: A review. *Environ. Monit. Assess.* **2019**, *191*, 205. [\[CrossRef\]](#) [\[PubMed\]](#)

4. Wang, J.; Bretz, M.; Dewan, M.A.A.; Delavar, M.A. Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. *Sci. Total Environ.* **2022**, *822*, 153559. [CrossRef] [PubMed]
5. Rußwurm, M.; Körner, M. Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 129. [CrossRef]
6. Lee, C.; Lee, J.; Park, S. Forecasting the urbanization dynamics in the Seoul metropolitan area using a long short-term memory-based model. *Environ. Plan. B Urban Anal. City Sci.* **2022**, *59*, in press. [CrossRef]
7. Xiao, B.; Liu, J.; Jiao, J.; Li, Y.; Liu, X.; Zhu, W. Modeling dynamic land use changes in the eastern portion of the hexi corridor, China by cnn-gru hybrid model. *GIScience Remote Sens.* **2022**, *59*, 501–519. [CrossRef]
8. Kubat, M.; Matwin, S. Addressing the curse of imbalanced data sets: One-sided sampling. In Proceedings of the Fourteenth International Conference on Machine Learning, Nashville, TN, USA, 8–12 July 1997; pp. 179–186.
9. Krawczyk, B. Learning from imbalanced data: Open challenges and future directions. *Prog. Artif. Intell.* **2016**, *5*, 221–232. [CrossRef]
10. Buda, M.; Maki, A.; Mazurowski, M.A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.* **2018**, *106*, 249–259. [CrossRef] [PubMed]
11. Pontius, R.G.; Shusas, E.; McEachern, M. Detecting important categorical land changes while accounting for persistence. *Agric. Ecosyst. Environ.* **2004**, *101*, 251–268. [CrossRef]
12. Karpatne, A.; Jiang, Z.; Vatsavai, R.R.; Shekhar, S.; Kumar, V. Monitoring land-cover changes: A machine-learning perspective. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 8–21. [CrossRef]
13. Samardžić-Petrović, M.; Kovačević, M.; Bajat, B.; Dragičević, S. Machine Learning Techniques for Modelling Short Term Land-Use Change. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 387. [CrossRef]
14. Karimi, F.; Sultana, S.; Babakan, A.S.; Suthaharan, S. Urban expansion modeling using an enhanced decision tree algorithm. *Geoinformatica* **2021**, *25*, 715–731. [CrossRef]
15. Ahmadlou, M.; Karimi, M.; Pontius, R.G. A new framework to deal with the class imbalance problem in urban gain modeling based on clustering and ensemble models. *Geocarto Int.* **2022**, *37*, 5669–5692. [CrossRef]
16. Douzas, G.; Bacao, F.; Fonseca, J.; Khudinyan, M. Imbalanced learning in land cover classification: Improving minority classes' prediction accuracy using the geometric SMOTE algorithm. *Remote Sens.* **2019**, *11*, 3040. [CrossRef]
17. Yu, X.; Wu, X.; Luo, C.; Ren, P. Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework. *GIScience Remote Sens.* **2017**, *54*, 741–758. [CrossRef]
18. Kamel, A.; Issam, B. Data Augmentation for Land Cover Classification Using Generative Adversarial Networks. *Int. Geosci. Remote Sens. Symp.* **2021**, *2021*, 2309–2312. [CrossRef]
19. Lu, J.; Ren, K.; Li, X.; Zhao, Y.; Xu, Z.; Ren, X. From reanalysis to satellite observations: Gap-filling with imbalanced learning. *Geoinformatica* **2022**, *26*, 397–428. [CrossRef]
20. Ren, X.; Mi, Z.; Georgopoulos, P.G. Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: Modeling ozone concentrations across the contiguous United States. *Environ. Int.* **2020**, *142*, 105827. [CrossRef] [PubMed]
21. Lian, D.; Wu, Y.; Ge, Y.; Xie, X.; Chen, E. Geography-Aware Sequential Location Recommendation. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, San Diego, CA, USA, 23–27 August 2020; pp. 2009–2019.
22. Sun, P.; Lu, Y.; Zhai, J. Mapping land cover using a developed U-Net model with weighted cross entropy. *Geocarto Int.* **2021**, in press. [CrossRef]
23. Fotheringham, A.S.; Crespo, R.; Yao, J. Geographical and Temporal Weighted Regression (GTWR). *Geogr. Anal.* **2015**, *47*, 431–452. [CrossRef]
24. Li, D.; Wen, G.; Kuai, Y.; Wang, L. Spatio-temporally weighted multiple instance learning for visual tracking. *Optik* **2018**, *171*, 904–917. [CrossRef]
25. Statistics Canada “Population and Dwelling Counts: Canada and Census Subdivisions”. Available online: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=9810000201> (accessed on 1 June 2022).
26. Sulla-Menashe, D.; Friedl, M. The Terra and Aqua Combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) Version 6 Data Product. Available online: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1_v006 (accessed on 30 January 2022).
27. Wang, M.; Sun, X.; Fan, Z.; Yue, T. Investigation of Future Land Use Change and Implications for Cropland Quality: The Case of China. *Sustainability* **2019**, *11*, 3327. [CrossRef]
28. Singh, V.G.; Singh, S.K.; Kumar, N.; Singh, R.P. Simulation of land use/land cover change at a basin scale using satellite data and markov chain model. *Geocarto Int.* **2022**, in press. [CrossRef]
29. Van Berkel, D.; Shashidharan, A.; Mordecai, R.S.; Vatsavai, R.; Petrasova, A.; Petras, V.; Mitsova, H.; Vogler, J.B.; Meentemeyer, R.K. Projecting urbanization and landscape change at large scale using the FUTURES model. *Land* **2019**, *8*, 144. [CrossRef]
30. Stobbe, T.E.; Eagle, A.J.; Cottleer, G.; van Kooten, G.C. Farmland Preservation Verdicts-Rezoning Agricultural Land in British Columbia. *Can. J. Agric. Econ.* **2011**, *59*, 555–572. [CrossRef]

31. NASA/METI/AIST/Japan Spacesystems and U.S./Japan ASTER Science Team. "ASTER Global Digital Elevation Model V003". Available online: <https://lpdaac.usgs.gov/products/astgtmv003/> (accessed on 10 June 2022).
32. Statistics Canada. "2016 Census-Boundary Files". Available online: <https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm> (accessed on 10 May 2022).
33. Statistics Canada. "2016 Census Road Network File". Available online: <https://open.canada.ca/data/en/dataset/57d5ffae-3048-4a19-9b4c-eab12f6322c5> (accessed on 29 July 2022).
34. van Vliet, J.; Naus, N.; van Lammeren, R.J.A.; Bregt, A.K.; Hurkens, J.; van Delden, H. Measuring the neighbourhood effect to calibrate land use models. *Comput. Environ. Urban Syst.* **2013**, *41*, 55–64. [\[CrossRef\]](#)
35. Roodposhti, M.S.; Hewitt, R.J.; Bryan, B.A. Towards automatic calibration of neighbourhood influence in cellular automata land-use models. *Comput. Environ. Urban Syst.* **2020**, *79*, 101416. [\[CrossRef\]](#)
36. Masolele, R.N.; De Sy, V.; Herold, M.; Marcos Gonzalez, D.; Verbesselt, J.; Gieseke, F.; Mullissa, A.G.; Martius, C. Spatial and temporal deep learning methods for deriving land-use following deforestation: A pan-tropical case study using Landsat time series. *Remote Sens. Environ.* **2021**, *264*, 112600. [\[CrossRef\]](#)
37. Gray, P.C.; Chamorro, D.F.; Ridge, J.T.; Kerner, H.R.; Ury, E.A.; Johnston, D.W. Temporally Generalizable Land Cover Classification: A Recurrent Convolutional Neural Network Unveils Major Coastal Change through Time. *Remote Sens.* **2021**, *13*, 3953. [\[CrossRef\]](#)
38. van Duynhoven, A.; Dragičević, S. Assessing the Impact of Neighborhood Size on Temporal Convolutional Networks for Modeling Land Cover Change. *Remote Sens.* **2022**, *14*, 4957. [\[CrossRef\]](#)
39. Verburg, P.H.; de Nijs, T.C.M.; van Eck, J.R.; Visser, H.; de Jong, K. A method to analyse neighbourhood characteristics of land use patterns. *Comput. Environ. Urban Syst.* **2004**, *28*, 667–690. [\[CrossRef\]](#)
40. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#)
41. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 26–28 October 2014; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 1724–1734.
42. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271.
43. Yin, G.; Huang, Z.; Bao, Y.; Wang, H.; Li, L.; Ma, X.; Zhang, Y. ConvGCN-RF: A hybrid learning model for commuting flow prediction considering geographical semantics and neighborhood effects. *Geoinformatica* **2022**, *in press*. [\[CrossRef\]](#)
44. Yan, J.; Chen, X.; Chen, Y.; Liang, D. Multistep Prediction of Land Cover from Dense Time Series Remote Sensing Images with Temporal Convolutional Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5149–5161. [\[CrossRef\]](#)
45. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *2015*, 802–810.
46. Chen, R.; Wang, X.; Zhang, W.; Zhu, X.; Li, A.; Yang, C. A hybrid CNN-LSTM model for typhoon formation forecasting. *Geoinformatica* **2019**, *23*, 375–396. [\[CrossRef\]](#)
47. Huang, C.-J.; Kuo, P.-H. A Deep CNN-LSTM Model for Particulate Matter (PM_{2.5}) Forecasting in Smart Cities. *Sensors* **2018**, *18*, 2220. [\[CrossRef\]](#)
48. Seifin, O.; Riese, F.M.; Keller, S. Deep learning for land cover change detection. *Remote Sens.* **2021**, *13*, 78. [\[CrossRef\]](#)
49. Pham, V.; Bluche, T.; Kermorant, C.; Louradour, J. Dropout Improves Recurrent Neural Networks for Handwriting Recognition. In Proceedings of the 2014 14th International Conference on Frontiers in Handwriting Recognition, Crete, Greece, 1–4 September 2014; pp. 285–290.
50. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Gordon-Rodriguez, E.; Loaiza-Ganem, G.; Pleiss, G.; Cunningham, J.P. Uses and Abuses of the Cross-Entropy Loss: Case Studies in Modern Deep Learning. In Proceedings of the Proceedings on "I Can't Believe It's Not Better!" at NeurIPS Workshops, online, 12 December 2020; Volume 137, pp. 1–10.
52. Maretto, R.V.; Fonseca, L.M.G.; Jacobs, N.; Körting, T.S.; Bendini, H.N.; Parente, L.L. Spatio-Temporal Deep Learning Approach to Map Deforestation in Amazon Rainforest. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 771–775. [\[CrossRef\]](#)
53. Zimmerman, D.; Pavlik, C.; Ruggles, A.; Armstrong, M.P. An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Math. Geol.* **1999**, *31*, 375–390. [\[CrossRef\]](#)
54. Kang, M.; Liu, Y.; Wang, M.; Li, L.; Weng, M. A random forest classifier with cost-sensitive learning to extract urban landmarks from an imbalanced dataset. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 496–513. [\[CrossRef\]](#)
55. Tong, X.; Feng, Y. A review of assessment methods for cellular automata models of land-use change and urban growth. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 866–898. [\[CrossRef\]](#)
56. Pontius, R.G.; Boersma, W.; Castella, J.-C.C.; Clarke, K.; de Nijs, T.; Dietzel, C.; Duan, Z.; Fotsing, E.; Goldstein, N.; Kok, K.; et al. Comparing the input, output, and validation maps for several models of land change. *Ann. Reg. Sci.* **2008**, *42*, 11–37. [\[CrossRef\]](#)
57. Camacho Olmedo, M.T.; Pontius, R.G.; Paegelow, M.; Mas, J.F. Comparison of simulation models in terms of quantity and allocation of land change. *Environ. Model. Softw.* **2015**, *69*, 214–221. [\[CrossRef\]](#)
58. Paegelow, M.; Camacho Olmedo, M.T.; Mas, J.; Houet, T. Benchmarking of LUCC modelling tools by various validation techniques and error analysis. *Cybergeo Eur. J. Geogr.* **2014**, *701*. [\[CrossRef\]](#)
59. van Rossum, G. *Python Language Reference*; Python Software Foundation: Amsterdam, The Netherlands, 2009; ISBN 9780954161781.

60. Chollet, F. Keras: The Python Deep Learning Library. Available online: <https://keras.io/> (accessed on 26 May 2022).
61. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 265–283. [\[CrossRef\]](#)
62. Remy, P. Temporal Convolutional Networks for Keras. Available online: <https://github.com/philipperemy/keras-tcn> (accessed on 1 May 2022).
63. Naushad, R.; Kaur, T.; Ghaderpour, E. Deep transfer learning for land use and land cover classification: A comparative study. *Sensors* **2021**, *21*, 8083. [\[CrossRef\]](#)
64. Xiao, C.; Chen, N.; Hu, C.; Wang, K.; Xu, Z.; Cai, Y.; Xu, L.; Chen, Z.; Gong, J. A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environ. Model. Softw.* **2019**, *120*, 104502. [\[CrossRef\]](#)
65. van Duynhoven, A.; Dragičević, S. Exploring the sensitivity of recurrent neural network models for forecasting land cover change. *Land* **2021**, *10*, 282. [\[CrossRef\]](#)
66. Ma, J.; Ding, Y.; Gan, V.J.L.; Lin, C.; Wan, Z. Spatiotemporal Prediction of PM2.5 Concentrations at Different Time Granularities Using IDW-BLSTM. *IEEE Access* **2019**, *7*, 107897–107907. [\[CrossRef\]](#)
67. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Gupta, J.; Molnar, C.; Xie, Y.; Knight, J.; Shekhar, S. Spatial Variability Aware Deep Neural Networks (SVANN): A General Approach. *ACM Trans. Intell. Syst. Technol.* **2021**, *12*, 1–21. [\[CrossRef\]](#)
69. Costa, H.; Almeida, D.; Vala, F.; Marcelino, F.; Caetano, M. Land cover mapping from remotely sensed and auxiliary data for harmonized official statistics. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 157. [\[CrossRef\]](#)
70. da Silva, M.V.; Pandorfi, H.; de Oliveira-Júnior, J.F.; da Silva, J.L.B.; de Almeida, G.L.P.; de Assunção Montenegro, A.A.; Mesquita, M.; Ferreira, M.B.; Santana, T.C.; Marinho, G.T.B.; et al. Remote sensing techniques via Google Earth Engine for land degradation assessment in the Brazilian semiarid region, Brazil. *J. South Am. Earth Sci.* **2022**, *120*, 104061. [\[CrossRef\]](#)