



## Article

# Assessing the Impact of Neighborhood Size on Temporal Convolutional Networks for Modeling Land Cover Change

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:** Land cover change (LCC) studies are increasingly using deep learning (DL) modeling techniques. Past studies have leveraged temporal or spatiotemporal sequences of historical LC data to forecast changes with DL models. However, these studies do not adequately assess the association between neighborhood size and DL model capability to forecast LCCs, where neighborhood size refers to the spatial extent captured by each data sample. The objectives of this research study were to: (1) evaluate the effect of neighborhood size on the capacity of DL models to forecast LCCs, specifically Temporal Convolutional Networks (TCN) and Convolutional Neural Networks (CNN-TCN), and (2) assess the effect of auxiliary spatial variables on model capacity to forecast LCCs. First, each model type and neighborhood setting configuration was assessed using data derived from multitemporal MODIS LC for the Regional District of Bulkley-Nechako, Canada, comparing subareas exhibiting different amounts of LCCs with trends obtained for the full region. Next, outcomes were compared with three other study regions. The modeling results were evaluated with three-map comparison measures, where the real-world LC for the next timestep, the real-world LC for the previous timestep, and the forecasted LC for the next year were used to calculate correctly transitioned areas. Across all regions explored, it was observed that increasing neighborhood sizes improved the DL model's capabilities to forecast short-term LCCs. CNN-TCN models forecasted the most correct LCCs for several regions while reducing error due to quantity when provided additional spatial variables. This study contributes to the systematic exploration of neighborhood sizes on selected spatiotemporal DL techniques for geographic applications.

**Keywords:** temporal convolutional networks; convolutional neural networks; long short-term memory; temporal deep learning; spatiotemporal deep learning; CNN-TCN; CNN-LSTM; land cover change; neighborhood size effects



**Citation:** 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. <https://doi.org/10.3390/rs14194957>

Academic Editors: Dino Ienco, Fernando Bação and Eric Vaz

Received: 15 July 2022

Accepted: 30 September 2022

Published: 4 October 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

Land cover changes (LCCs) arise from human activities and environmental processes that alter the condition of the Earth's surface [1], contributing directly and indirectly to environmental changes in regional and global systems [2]. The interconnection of LC with local, regional, and global systems makes studying and forecasting LCCs a vital component in fields such as geography [3] and climatology [4]. The proliferation of open LC datasets has resulted in the increasing use of computational approaches such as deep learning (DL) methods for classification and modeling tasks [5]. Neural networks of expanding breadth and depth underpin DL algorithms, facilitating complex pattern recognition in remote sensing data sources [6]. To forecast LC given a remotely sensed image timeseries, temporal DL models called Recurrent Neural Networks (RNNs) and Temporal Convolutional Networks (TCNs) can be used to obtain patterns from sequences.

Many research studies use timeseries data extracted from each cell or pixel comprising historical raster GIS data layers to classify LC and to forecast LCCs presuming enough data is available [7,8]. For example, TCNs were used to project vegetation cover with

multitemporal remotely sensed data [8]. However, these research studies extracted temporal sequences from each cell comprising the raster data layers, excluding consideration of proximal LC dynamics. While temporal sequences provide historical information of individual cells comprising a raster, influences occurring within the neighborhood of each cell are also important for LCC modeling [9]. In order to extract spatial features from a neighborhood in DL models, Convolutional Neural Networks (CNNs) are typically used to process neighborhoods provided as a grid of cells at each timestep before extracting temporal patterns [5]. Models containing CNN and temporal DL techniques are often referred to as “spatiotemporal” models and have been leveraged in past studies on land change applications. For example, a model composed of LSTM layers is considered a temporal model. Therefore, CNN and LSTM layers can be integrated to implement spatiotemporal models, such as CNN–LSTM.

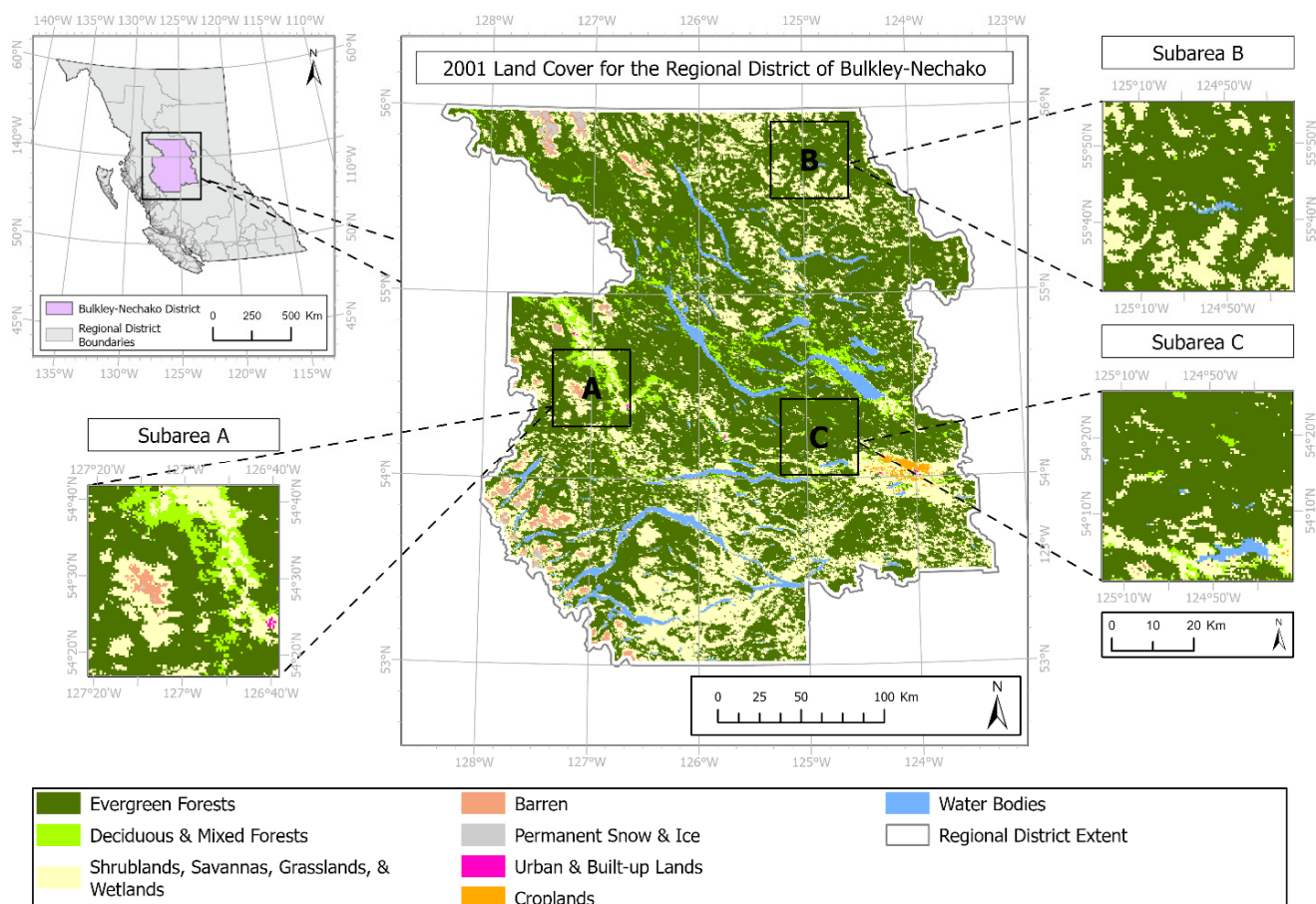
Constructing a dataset from a multi-temporal geographic data source for spatiotemporal DL models requires a neighborhood to be specified. Previous research studies have used a variety of neighborhood settings for land change modeling and classification. For instance, land change classification and forecasting applications involving datasets with spatial resolutions of 30 meters used Moore neighborhoods spanning  $3 \times 3$ ,  $5 \times 5$ , and  $9 \times 9$  cells [10–12]. With finer spatial resolution training data samples,  $9 \times 9$  neighborhoods were specified to classify LC with DL models [13]. With coarser spatial resolution data,  $11 \times 11$  neighborhoods and below were found to better capture nearby land change interactions [14]. Neighborhood dimensions not following the Moore neighborhood convention also involved  $10 \times 10$  [15],  $32 \times 32$  [5], and  $64 \times 64$  neighborhood settings [16] with 10 and 20 meter spatial resolution datasets. However, despite the range of neighborhood dimensions investigated in previous studies, the impact of neighborhood dimensions on DL models for the task of LCC forecasting is undetermined. While it is acknowledged that neighborhoods encompassing hundreds or thousands of pixels from fine spatial resolution data are commonly used in other applications such as change detection, the relationship between neighborhood size and a model’s capacity to forecast LCCs is unknown.

Only a few studies have explored the influence of neighborhood sizes on the outcomes of DL methods used for LC classification and modeling and were primarily selected for reasons such as to operate within computational constraints [16], to maintain available dataset qualities or formats [5], or to optimize overall accuracy measures [12]. They may also be specified arbitrarily [11,17]. It becomes difficult to determine optimal neighborhood settings from the current literature for the task of modeling LCC with DL. Thus, further systematic exploration of the relationship between neighborhood size and capacity of DL models to forecast LCCs is needed. Additionally, TCN and CNN–TCN models are unexplored with neighborhood settings beyond cell-level sequences for LCC forecasting or with other spatial drivers of LCC. Therefore, the main objectives this research study were to: (1) evaluate the effect of neighborhood size on the capacity of DL models to forecast LCCs, specifically TCN and CNN–TCN, and (2) assess the effect of auxiliary spatial variables on model capacity to forecast LCCs. Preliminary outcomes were explored with respect to datasets for the Regional District of Bulkley–Nechako in the province of British Columbia (BC), Canada, where trends were first explored for subareas and the entire region. Next, the findings obtained for the full region were compared with those obtained for three separate regions in the province. Lastly, the influence of spatial variables was compared across model and neighborhood configurations applied to all four regions. Experiments conducted on the respective datasets used two temporal DL models (LSTM and TCN) and two spatiotemporal models (CNN–LSTM and CNN–TCN). The purpose of using the LSTM and CNN–LSTM models was to provide a baseline for comparing the TCN and CNN–TCN model performance, because they are more commonly used. Outcomes were examined using change-focused metrics. The specific details of the research work are outlined in the next sections.

## 2. Methodology

### 2.1. Study Area and Datasets

The primary experimental study area used was the Regional District of Bulkley-Nechako, located in northern BC (Figure 1). LCCs over the past two decades have been characterized by a notable loss of forested areas in this region (Table 1). The spatiotemporal LC dataset was obtained from the Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) dataset [18] featuring yearly temporal resolution data from 2001 to 2019. The Land Cover Type 1: Annual International Geosphere–Biosphere Programme (IGBP) classification data layer featuring 17 LC classes was used for this study. In particular, the IGBP raster product was provided at a  $0.05^\circ$  spatial resolution, which is approximately 463.31 m spatial resolution in the Bulkley-Nechako region [19]. LC classes were aggregated based on previous research studies [20] to produce eight classes including: (a) evergreen forests, (b) deciduous and mixed forests, (c) shrublands, savannas, grasslands, and wetlands, (d) barren land, (e) permanent snow and ice, (f) urban and built-up lands, (g) croplands, and (h) water bodies. To support the exploration of trends across experiments, three alternative study areas were also considered (Figure 2). These include the Regional District of Central Kootenay, the Northern Rockies Regional Municipality, and the Cariboo Regional District, where the amount of change between 2001 and 2019 differs from the primary experimental study area (Table 1).



**Figure 1.** 2001 land cover of the primary study area of the Regional District of Bulkley-Nechako, BC. Data are displayed with the NAD 1983 BC Environment Albers projected coordinate system. Subareas A, B, and C are subsets of the Regional District of Bulkley-Nechako involved in exploring and identifying preliminary experimental trends in subsequent sections.

**Table 1.** Area covered by each land cover class in 2001 and 2019 for (R1) the Regional District of Bulkley-Nechako, (R2) the Regional District of Central Kootenay, (R3) the Northern Rockies Regional Municipality, and (R4) the Cariboo Regional District.

Study Region	Year	Evergreen Forests	Deciduous and Mixed Forests	Shrublands, Savannas, Grasslands, and Wetlands	Barren	Permanent Snow and Ice	Urban and Built-Up Lands	Croplands	Water Bodies
(R1) Bulkley-Nechako	2001	46,710.59	2383.57	20,373.90	802.61	314.69	16.74	230.54	2949.41
	2019	33,772.89	1692.15	33,669.00	665.01	167.86	16.74	211.87	3586.52
	% Change	−27.70%	−29.01%	65.26%	−17.14%	−46.66%	0%	−8.10%	21.60%
(R2) Central Kootenay	2001	11,194.45	227.97	7759.48	525.48	71.48	23.61	108.62	698.07
	2019	9250.07	227.75	9586.01	501.23	116.99	23.61	69.55	833.95
	% Change	−17.37%	−0.09%	23.54%	−4.62%	63.66%	0%	−35.97%	19.46%
(R3) Northern Rockies	2001	26,927.21	14,911.48	39,286.83	1503.04	119.14	0.86	0.21	196.20
	2019	32,664.82	12,936.62	35,361.15	1403.22	137.60	0.86	0.21	440.48
	% Change	21.31%	−13.24%	−9.99%	−6.64%	15.50%	0%	0%	124.51%
(R4) Cariboo	2001	30,322.90	1603.93	43,009.01	2240.82	466.88	24.90	23.83	844.25
	2019	21,650.69	1601.35	51,251.26	2116.32	539.65	24.90	17.60	1334.75
	% Change	−28.60%	−0.16%	19.16%	−5.56%	15.59%	0%	−26.13%	58.10%

In addition to spatiotemporal LC data, seven static spatial variables were included to capture drivers of LCCs. The selected variables capture a single snapshot of factors related to accessibility and topography. Accessibility factors used in prior LCC applications and adopted in this research study include the proximity to various features, including proximity to population centers [21], roads [22], railways [23], lakes and rivers [24], and coastal waters [25]. These were derived by computing the Euclidean distance from each feature using the point and vector data layers available from Statistics Canada [26,27]. Topographic variables involved in prior LCC studies include elevation and slope [28]. The ASTER Global digital elevation model was used to source the elevation data and derive the slope data layer for this research study [29].

All datasets are reprojected to the NAD 1983 BC Environment Albers projected coordinate system where planar area measurements are preserved. Spatial data layers are resampled to or computed relative to the MODIS LC data, which has the coarsest spatial resolution. Bilinear interpolation was used to resample the DEM. Edge effects were minimized by eliminating the influence of cells outside of the study region, with a 2.78 km buffer from the edges of the study area [30].

## 2.2. Overview of Deep Learning Models

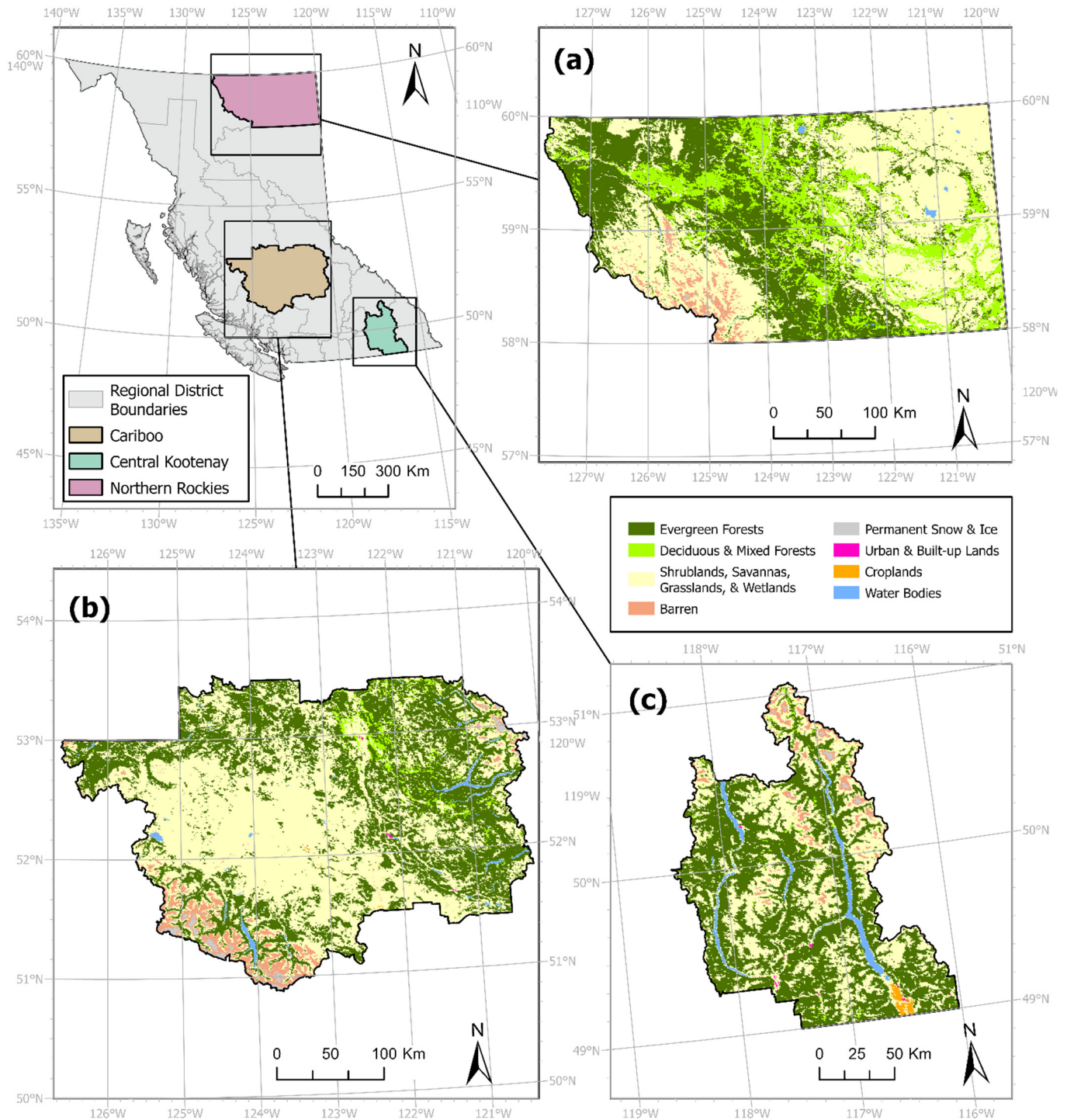
Four models were implemented in this research study, including two temporal DL models (LSTM and TCN) and two spatiotemporal models (CNN-LSTM and CNN-TCN). All models were implemented with Python 3.9.1 [31], the Keras API [32], and TensorFlow 2.5.0 [33]. The open-source KerasTCN API was used to implement the TCN and CNN-TCN models [34].

### 2.2.1. Temporal Models (LSTM and TCN)

LSTM models are extensively used to model temporal patterns [35]. The development of LSTM models occurred in response to the vanishing gradient problem that affected traditional RNN approaches for long timeseries data [36,37]. The unique features of an LSTM unit include “gating” functions and internal memory cells which maintain, drop, or inhibit the flow of information through an LSTM unit [38]. The main equations of an LSTM are expressed in previous studies [38]. The LSTM models specified in this research study feature two LSTM layers with 32 and 128 units, respectively, that use the default *tanh* activation function [11]. The layers are followed by a dropout regularization factor of 10% informed by previous studies [11,39]. Dropout regularization inhibits networks from overfitting by randomly dropping a specified percentage of neurons [40]. The fully-connected output



layer features 9 neurons to forecast the likelihood of each of the 8 LC classes. The activation of the output layer is the Softmax function which is useful for multi-class forecasting and classification problems [41]. Neighborhood inputs are provided as one-dimensional arrays for each timestep, described in detail in Section 2.2.3.



**Figure 2.** 2001 land cover of the additional study areas of (a) the Northern Rockies Regional Municipality, (b) the Cariboo Regional District, and (c) the Regional District of Central Kootenay. Data are displayed with the NAD 1983 BC Environment Albers projected coordinate system.

TCNs are also used to obtain long-term patterns from sequential data [42]. While TCNs are used for extracting proximal temporal features, they also benefit from the highly

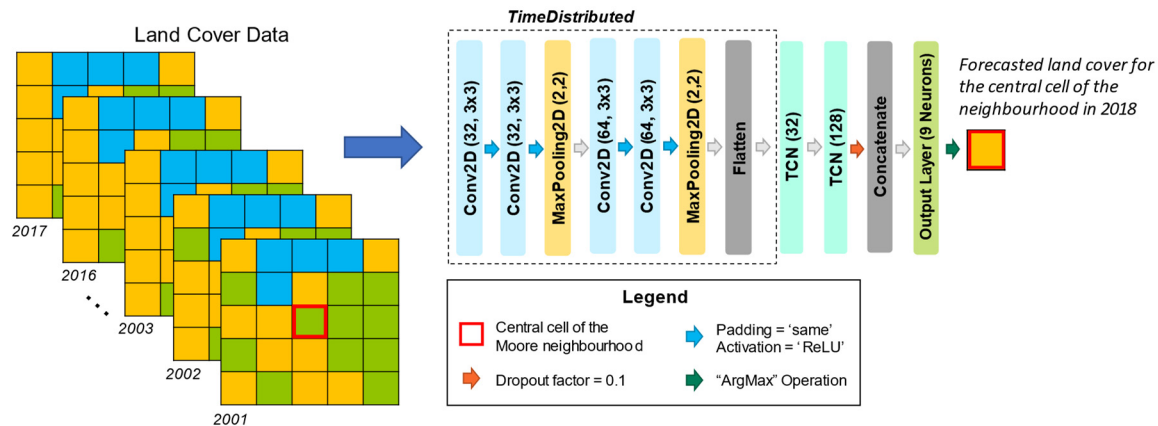
parallelizable structure of CNNs, providing a less computationally intensive alternative to LSTM models. The primary components of a TCN model include a one-dimensional, fully convolutional network to extract temporal patterns and dilated causal convolutions, which prevent the purging of historical information [43]. Dilations enable the receptive field to expand exponentially, thus preserving longer-term information from sequential datasets. The formulation of TCNs has been discussed in detail in prior research studies [42]. The implemented TCN models are characterized by two layers featuring 32 and 128 filters, respectively, using the ReLU activation function. Filters characterizing a TCN refer to a vector of learnable weights. The dilations parameter setting is 1, 2, 4, and 8 [44], and the kernel size is set to 2 [45]. The kernel refers to a matrix of weights that is multiplied with each subset of the input sequence. To match the LSTM model implementation, a dropout factor of 10% is applied before the output layer featuring the Softmax activation function. Neighborhood inputs are provided as one-dimensional arrays for each timestep, described in Section 2.2.3.

### 2.2.2. Spatiotemporal Models (CNN–LSTM and CNN–TCN)

In combinations of spatial and sequential DL techniques, spatial features are often first extracted using Convolutional Neural Networks (CNNs). CNNs can be used to obtain abstract spatial features from remotely sensed imagery [6]. The formulas describing CNNs have been disseminated in previous research studies [12]. Through a combination of convolutional layers and pooling layers, spatial patterns from a two-dimensional raster or image can be extracted [46]. This makes a CNN a useful construct for extending the spatial coverage of sequential DL models beyond per-cell timeseries to capture the states of neighboring cells [9]. As such, CNNs facilitate the extraction of spatial relationships present within a neighborhood at each timestep, while temporal relationships are captured using a sequential DL approach. Such models are commonly used for video recognition tasks (i.e., CNN–LSTM [47]) that require spatial dependencies to be first extracted before obtaining temporal dependencies from samples of independent videos. However, in geographic applications employing raster data layers, training datasets are populated by extracting spatiotemporal neighborhoods or “patches” across the study area, where the spatial coverage of each sample is specified by the neighborhood dimensions.

In this research study, CNN–LSTM and CNN–TCN were implemented to extract spatial relationships within neighborhoods of each cell for each timestep. These spatiotemporal DL models were implemented first with a single-branch input structure that accepts the LC data sequences only, adapted from prior model implementations [48,49]. LC sequences were provided as input to a branch consisting of four CNN layers and two  $2 \times 2$  max-pooling operations (Figure 3). The CNN layers feature 32, 32, 64, and 64 filters, respectively, each followed with a ReLU activation function. Using multiple CNN layers and expanded filter sizes facilitates the extraction of increasingly complex spatial features [50]. The CNN layers were each parameterized with a  $3 \times 3$  kernel, zero-padding, and a stride of one [11]. A  $3 \times 3$  kernel size was specified given the sample input dimensions [51,52] and to obtain local relationships of LC from sample neighborhoods [53]. Likewise, while increasing kernel sizes may be beneficial for finer-scale spatial resolution data for the purpose of extracting larger scale features [54], the coarser spatial resolution of the data and sample input dimensions used in this research study are better suited to a smaller kernel size for considering the local context of LC. The CNN operations were followed by a dropout factor of 10%. The TimeDistributed functionality afforded by the Keras API enables these operations to be applied to each timestep in the sequence of LC data provided [11]. This facilitates the extraction of spatial features from each timestep. Outputs of this branch were flattened, then provided as inputs to the sequential layers. For example, TCN layers are shown in Figure 3 following the concatenation operation. In the CNN–LSTM approach, this was replaced with LSTM layers with 32 and 128 units, respectively. The general structure of the CNN–TCN model is inspired by a previous approach demonstrating how outputted feature vectors from CNNs were provided as inputs to TCNs for a video classification

task [55]. However, the CNN–TCN model structure in this research study does not require the advanced CNN component for the MODIS LC data. The CNN–TCN model was implemented with TensorFlow and KerasTCN, with TCN layers parameterized the same as in Section 2.2.1.



**Figure 3.** Configuration for training a single-branch CNN–TCN model accepting LC data sequences with a  $5 \times 5$  Moore neighborhood.

### 2.2.3. Neighborhood Effects in Deep Learning Models

For each cell, the influence of neighborhood composition and changes within it are referred to as the “neighborhood effect” [56]. In this research study, DL models were used to obtain the strength of neighborhood effects and influential spatiotemporal patterns from samples without user intervention. The spatial coverage of neighborhood effects is controlled by specifying a neighborhood size, which is a highly influential parameter in land change modeling endeavors [56–58]. It should be noted that the term “neighborhood” is synonymous to spatiotemporal data “tiles” (i.e., [5]) or “patches” (i.e., [17,52]) prepared for DL model training datasets described in some research studies. The terms “tiles” or “patches” were previously used to refer to some  $M \times M$  area obtained from an overall study region, where  $M$  refers to the number of cells along the longest edge of the cell neighborhood. This research study follows the Moore neighborhood convention, where a target cell refers to the cell located at the center of the neighborhood. For example, a  $3 \times 3$  Moore neighborhood contains a target central cell and the eight surrounding cells [58]. Moore neighborhoods contain  $N$  cells given a distance range parameter  $r$ , where  $N = (2r + 1)^2$ . The neighborhood sizes explored in this study and their respective surface areas they cover given the spatial resolution of the dataset are presented in Table 2. The neighborhood dimensions selected to explore with the DL models are based on previous findings that showed that including cells 2.5 km to 3.5 km from the central cell captured the characteristics of neighborhoods with 500 m spatial resolution data [14].

In this research study, the task of LCC forecasting was framed as a spatiotemporal DL problem with respect to the CNN–LSTM and CNN–TCN modeling techniques. The intended effect of including neighborhoods in DL models is to capture a range of influence surrounding each cell by integrating information about state changes occurring near a central cell. For CNN–LSTM and CNN–TCN models, neighborhoods surrounding each target cell at each timestep are provided explicitly as the model’s input. For example, these models accept spatiotemporal inputs in the form of  $T \times M \times M \times C$ , where  $T$  denotes the number of timesteps,  $M$  denotes the neighborhood dimension, and  $C$  denotes the number of LC classes. For the temporal models, LSTM and TCN, the neighborhood structure is not inherently preserved. In previous studies, the states of each neighboring cell were explicitly provided as variables to LSTM models in the form of a one-dimensional array for each timestep [11,17]. To provide neighborhoods of size  $3 \times 3$  and larger to the temporal models, the process of “flattening” neighborhood information to compatible one-dimensional sequences was carried out. This means that the  $T \times M \times M \times C$  spatiotemporal

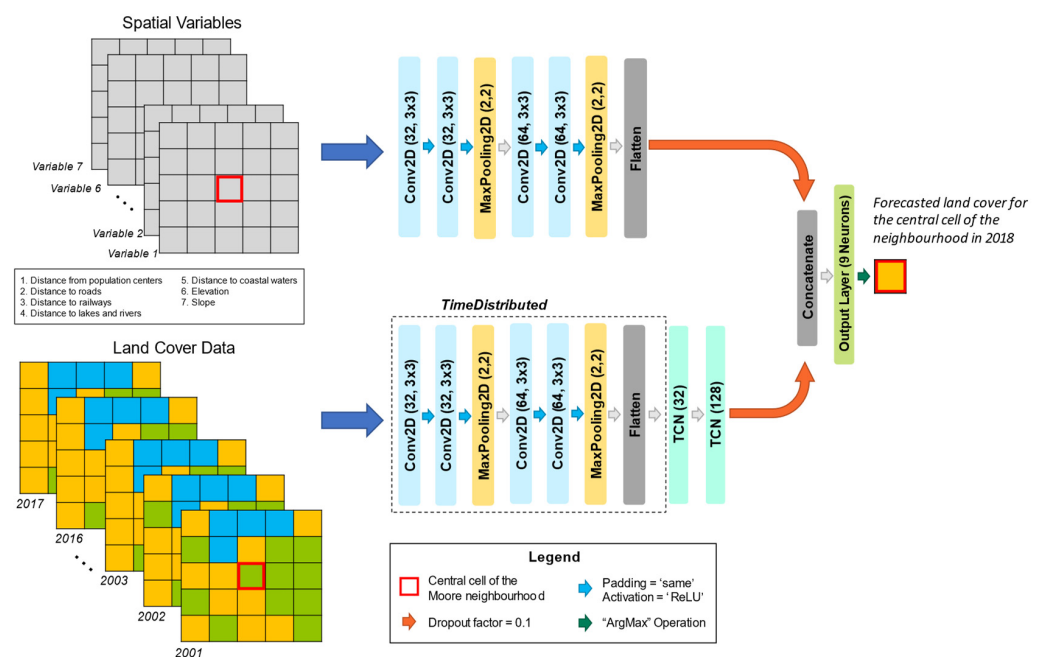
input sequence for each cell was transformed to a one-dimensional array with dimensions  $T \times M \times M \times C$ , where  $M \times M \times C$  is the length of the input vector at each timestep.

**Table 2.** Surface area covered by a training sample with each neighborhood configuration. A Moore neighborhood encompasses  $N$  cells given a range parameter  $r$  specifying the distance from the central cell, where  $N = (2r + 1)^2$ .

Range from Central Cell ( $r$ )	Neighborhood Size ( $M \times M$ )	# of Cells or Land Cover States Contributing to Neighborhood Effects ( $N$ )	Spatial Coverage Considered for Neighborhood Effects ( $\text{km}^2$ )
0	$1 \times 1$	1	0.21
1	$3 \times 3$	9	1.93
2	$5 \times 5$	25	5.37
3	$7 \times 7$	49	10.52
4	$9 \times 9$	81	17.39
5	$11 \times 11$	121	25.97

#### 2.2.4. Adding Spatial Variables

To accommodate both spatiotemporal LC data and static spatial variables, the previous single-branch model was extended for the spatial variable input (Figure 4). This structure was expanded from previous studies [48,49]. For the multi-branch model implementation used to add the spatial variables for the Case 3 experiments, the LC input branch implementation is the same. The second branch of the model uses convolutional layers to extract spatial relationships from the seven spatial variables of each training sample. Six convolution layers,  $2 \times 2$  max-pooling operations, and dropout regularization were used to process the auxiliary spatial features (Figure 4). The number of filters parameterizing each CNN layer of this branch was 32, 32, 64, and 64, respectively. Outputs of the spatial input branch were flattened and combined with outputs from the LC data branch using the concatenation operation. The activation function of all CNN layers is ReLU and the output layer with 9 neurons uses the Softmax function.



**Figure 4.** Configuration for training a multi-branch CNN-TCN model accepting the spatiotemporal LC data sequences and static spatial variables with a  $5 \times 5$  Moore neighborhood.



### 2.3. Overview of Experiments

The LSTM, TCN, CNN–LSTM, and CNN–TCN models were explored in three experimental cases. Real-world multi-temporal MODIS LC data and static spatial variables were used to forecast one year of LCCs for the selected study regions. LSTM and CNN–LSTM models were used as a baseline to compare TCN and CNN–TCN models, as the former are more common in geographic applications. Case 1 considered the LC data of the Regional District of Bulkley–Nechako (R1) to explore whether trends in the model’s capacity to forecast LCCs for three subareas is similar to that observed for the full region from which they were extracted from (Figure 1). Next, Case 2 explored the respective models on separate study regions, including the Regional District of Central Kootenay (R2), the Northern Rockies Regional Municipality (R3), and the Cariboo Regional District (R4) (Figure 2). Outcomes from Case 1 and the additional regions were compared to identify recurring trends. Lastly, Case 3 explored the effect of the spatial variables branch across model types and neighborhood sizes.

The temporal resolution of the models was one year. A “rolling-window” forecasting scheme used a specified interval to train the model, while the interval is advanced by one year to test the model by evaluating its capacity to project the next timestep [59]. The LCC modeling task was framed as a multi-class forecasting task, where each cell was assigned the most likely LC type according to the per-class probabilities produced by the model. Training datasets are formed in the same way as previous work [20]. The training datasets for each study region and subarea were comprised of 17 timesteps of annual LC data, where  $(t_0, t_1, \dots, t_{16})$  are used for the training data sequences and  $t_{17}$  providing the training label, with 20% of the training dataset withheld for validation purposes. The test dataset input sequences span  $(t_1, t_2, \dots, t_{17})$ , with  $t_{18}$  being the forecasted LC data. The first timestep  $t_0$  was supplied with LC data for 2001. Therefore, the training datasets include LC data from 2001 to 2018, where 2001 to 2017 were used as input for model training and the 2018 LC data layer providing the training label. Sequences from 2002 to 2018 were then used to forecast the 2019 LC of BC. Training data samples of size  $T \times M \times M \times C$  were obtained for every cell in the study area, excluding those within the specified distance from study region boundaries to manage edge effects (Section 2.1). This means that every cell with its surrounding neighborhood provides a sample within the training dataset. Each model type and neighborhood configuration were provided the same training datasets for the respective regions.

In all experimental cases, the LSTM, TCN, CNN–LSTM, and CNN–TCN models were run with neighborhood dimension settings of  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ , and  $11 \times 11$ . In Cases 1 and 2, the model configurations trained with LC data only (Figure 3) are referred to as LSTM $_{M \times M}$ , TCN $_{M \times M}$ , CNN–LSTM $_{M \times M}$ , and CNN–TCN $_{M \times M}$  where  $M$  refers to the neighborhood size setting (Table 2). For Case 3, models trained with LC data and the static spatial variables are referred to as CNN–TCN $_{M \times M (LC+SVs)}$  (Figure 4).

### 2.4. Model Assessment

Change-focused measures provide a means to compare the model output with the real-world data for areas that have undergone LCCs. These were selected because measures including overall accuracy and traditional Kappa statistics are impacted by the prevalence of non-changing areas, along with other issues that conceal the true nature of map agreement [60,61]. Therefore, three-map comparison measures have been used to differentiate changed and persistent areas by comparing an initial map representing the real-world LC states at time  $t_0$ , a reference map representing the real-world LC at time  $t_n$ , and the forecasted LC map for time  $t_n$  [60]. Measures that involve three-map comparison include Figure of Merit (FOM), Producer’s Accuracy (PA), and User’s Accuracy (UA), as expressed in previous work [62–65] and in Appendix A. FOM indicates the ratio of correctly forecasted changes versus the total amount of real-world and projected changes. Derived from components of FOM (Table A1), PA expresses the area forecasted correctly as changed versus the real-world changed areas, while UA expresses the amount of correctly changed

area versus the real-world changed areas. FOM, PA, and UA measures were selected for the purpose of determining the model's capacity to forecast LCCs, providing the primary means of quantifying model performance for full region experiments in Cases 1 and 2.

To investigate the types of spatial errors associated with each forecast, the total error or disagreement between real-world change and forecasted change can be explored with respect to error due to quantity (EQ) and error due to allocation (EA), as expressed in prior works [64–66] and in Appendix A. EQ indicates the difference in forecasted and real-world changed area, while EA indicates the amount of area allocated incorrectly as changed or to the wrong LC class. Both are expressed in terms of disagreeing area. The EQ and EA measures provide a disaggregated view of projected changes compared to FOM, PA, and UA, highlighting instead the disagreement in quantity and spatial disagreement of observed versus forecasted changes. Further descriptions and interpretation of the selected measures are included in Table A1. The EQ and EA measures were used to assess the impact of spatial variables added to models in Case 3. Additionally, simulation maps exhibiting correctly changed or persistent areas along with different types of forecasting errors were used to support the visual assessment of outcomes associated with the best performing models for each full study region, selected according to the highest FOM values found across Cases 2 and 3.

The initial reference year selected for a three-map comparison determines areas flagged as changed. The measures were computed using the 2018 LC data as the initial map, the 2019 LC data as the reference map, and the forecasted 2019 LC data in the three-map comparison evaluation measures. The goal was to show the model's capacity to forecast changes overall for the one-year interval. This means that though the models output the probability of each LC class for each cell in a study region, the correctly forecasted changes for each class were summed to attain the overall change measures used in this research study. Thus, the change from each LC class was treated the same way. The surface area that underwent transitions between 2018 and 2019 is shown in Table 3.

**Table 3.** Summary of the total changed area across all LC classes for each experimental study region between 2001–2019, and 2018–2019. The change summaries are a summation of net change area for each LC class from the real-world LC data between 2001 and 2019, and 2018 and 2019.

Case	Study Region	Changed Area between 2001 and 2019 (km <sup>2</sup> )	Changed Area between 2018 and 2019 (km <sup>2</sup> )
1	Subarea A <sup>1</sup>	365.35 (17.02%)	66.33 (3.09%)
	Subarea B <sup>1</sup>	661.15 (30.80%)	138.88 (6.47%)
	Subarea C <sup>1</sup>	1150.36 (53.59%)	187.61 (8.74%)
	Full Regional District of Bulkley-Nechako	23,452.10 (31.79%)	4317.00 (5.85%)
2	Regional District of Central Kootenay	4015.83 (19.49%)	310.40 (1.51%)
	Northern Rockies Regional Municipality	12,000.71 (14.47%)	1940.51 (2.34%)
	Cariboo Regional District	17,611.03 (22.42%)	2932.67 (3.73%)

<sup>1</sup> Subareas A, B, and C are subsets of the Full Regional District of Bulkley-Nechako (Figure 1).

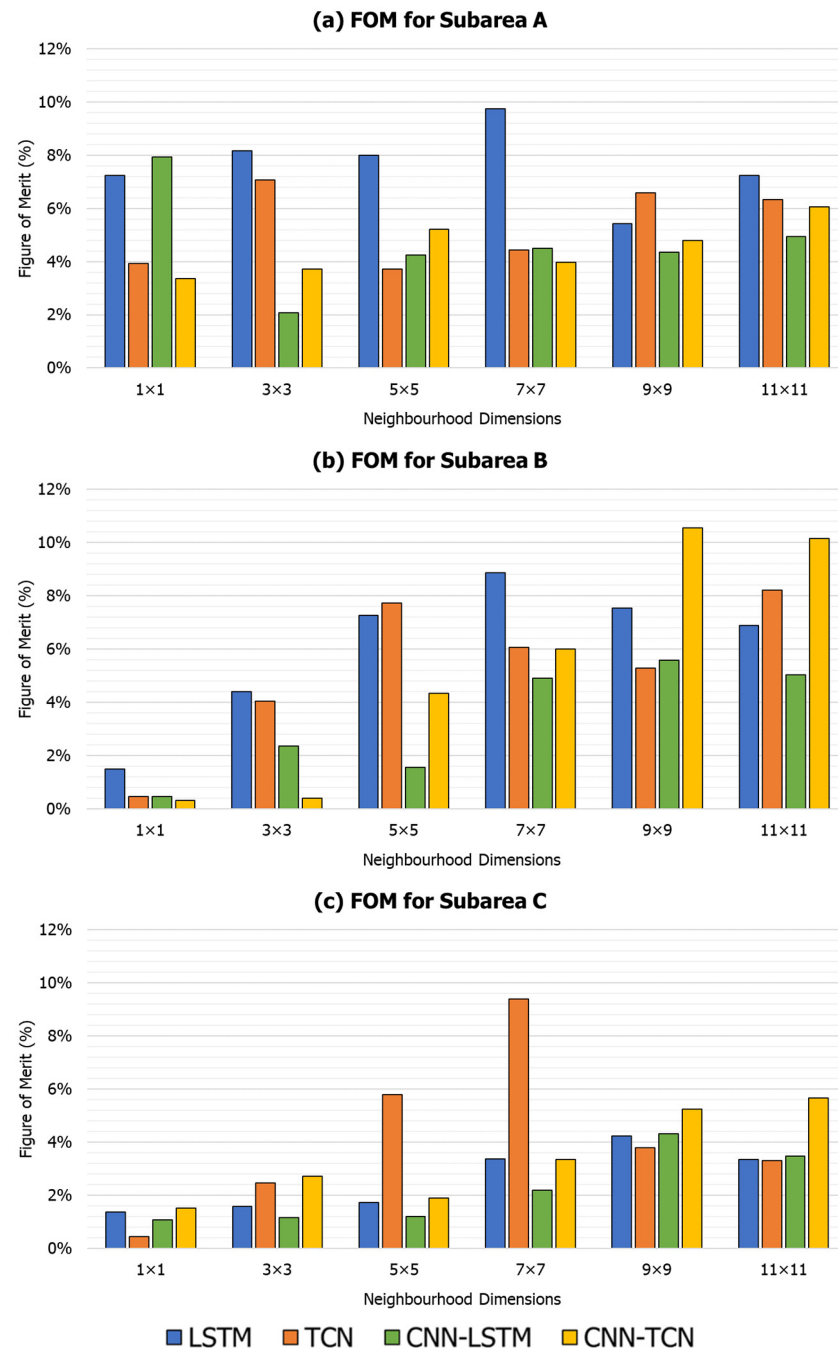
### 3. Results

#### 3.1. Case 1: Regional District of Bulkley-Nechako Experiment Results

##### 3.1.1. Subarea Experiment Results

Trends across neighborhood sizes and DL models were explored for three subareas within the Regional District of Bulkley-Nechako (R1). Subarea A underwent the least amount of change, while Subarea C underwent the most amount of change (Table 3). For Subarea A, the highest FOM measure was 9.7%, associated with forecasts from LSTM<sub>7×7</sub> (Figure 5a). For this subarea, there was no distinctive trend in FOM values versus neighborhood size increments. For Subarea B, the top FOM value (10.6%) was associated with the CNN-TCN<sub>9×9</sub> model (Figure 5b). For outcomes obtained for Subarea B, the overall trend observed was that the top FOM values increase until the 9 × 9 neighborhood specification. For Subarea C, the highest FOM value (9.4%) was associated with the TCN<sub>7×7</sub> model

(Figure 5c). The FOM values generally increase as neighborhood dimensions increase until the  $9 \times 9$  neighborhood dimensions. The exceptions to this trend are the  $\text{TCN}_{5 \times 5}$  and  $\text{TCN}_{7 \times 7}$  models, where FOM values are distinctively higher at 5.8% and 9.4%, respectively (Figure 5c).



**Figure 5.** FOM measures computed for (a) Subarea A, (b) Subarea B, and (c) Subarea C in the Regional District of Bulkley-Nechako.

The LSTM, TCN, CNN-LSTM, and CNN-TCN models also showed different trends with respect to each subarea. With Subarea A, FOM measures indicated inconsistent model performance. The exception was the LSTM models, which produced the highest FOM values for many of the neighborhood sizes (Figure 5a). The  $\text{CNN-LSTM}_{1 \times 1}$  and  $\text{TCN}_{9 \times 9}$  models produced the only FOM values that surpassed LSTM model performance for Subarea A. Considering Subareas B and C, the highest performing models varied

between LSTM, TCN, and CNN-TCN. For Subareas A and B, which exhibited low to moderate amounts of LCCs, the LSTM models attained the highest FOM values with  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$  neighborhood dimensions. For Subareas B and C, which exhibited moderate to large amounts of LCCs, the CNN-TCN models obtained the highest FOM values with  $9 \times 9$  and  $11 \times 11$  neighborhoods.

### 3.1.2. Entire Regional District of Bulkley-Nechako Experiment Results

Following the application of the respective models to various subareas of the Regional District of Bulkley-Nechako, new models were trained using the LC for the full study region to determine whether subarea trends continued for the larger extent. For the full study region, 31.8% of the LC differed between 2001 and 2019, and 5.6% differed between 2018 and 2019 (Table 3). Therefore, the percentage of changed area characterizing the full study region is most similar to that observed for Subarea B. The largest FOM and PA values were 3.0% and 3.2%, respectively, obtained with CNN-TCN $_{9 \times 9}$  (Figure 6a,b). Meanwhile, the highest UA was 41.7%, obtained by CNN-LSTM $_{3 \times 3}$ , indicating that more of its forecasted changes were at correct real-world locations (Figure 6c).

### 3.2. Case 2: Comparison with Alternative Regions

To explore the transferability of outcomes attained in Case 1 for the Regional District of Bulkley-Nechako (R1), three other regions were considered. These include the Regional District of Central Kootenay (R2), the Northern Rockies Regional Municipality (R3), and the Cariboo Regional District (R4). FOM values obtained with respect to models trained with each dataset generally increased as neighborhood size expanded (Table 4). Many of the highest amounts of correctly forecasted changes were associated with neighborhood specifications of  $7 \times 7$ ,  $9 \times 9$ , and  $11 \times 11$ , as indicated by FOM and PA (Tables 4 and 5). This follows a similar trend to those observed in Case 1. However, the same drop-off of FOM and PA values after CNN-TCN $_{9 \times 9}$  was not observed for the alternative regions (R2-R4). Instead, CNN-TCN $_{11 \times 11}$  attained higher FOM and PA measures, which were also some of the highest for each of the alternative regions (Tables 4 and 5). For the temporal models, a similar decrease in FOM values obtained after LSTM $_{9 \times 9}$  followed the observed trend in Figures 5b,c and 6a. Trends across UA values were also similar to those observed in Case 1, where smaller neighborhoods of  $1 \times 1$ ,  $3 \times 3$ , and  $5 \times 5$ , were associated with the highest UA values in R4, R2, and R3, respectively (Table 6). While the UA measures indicated that a higher proportion of forecasted changed areas intersected with real-world changed areas with several configurations using smaller neighborhoods, this measure is also swayed by correct forecasts of persistence (Table A1). Meanwhile, the PA measures indicate that more correctly changed area with less persistent area forecasted incorrectly was attained with the larger neighborhood settings of  $7 \times 7$ ,  $9 \times 9$ , and  $11 \times 11$  for all regions.

**Table 4.** Figure of merit (FOM) values obtained for (R1) the Regional District of Bulkley-Nechako, (R2) the Regional District of Central Kootenay, (R3) the Northern Rockies Regional Municipality, and (R4) the Cariboo Regional District.

Region	$M \times M$	Figure of Merit (%)			
		LSTM	TCN	CNN-LSTM	CNN-TCN
R1	$1 \times 1$	0.25	0.54	0.25	0.51
	$3 \times 3$	1.39	1.39	1.00	1.04
	$5 \times 5$	1.83	1.58	1.25	2.14
	$7 \times 7$	2.04	1.90	1.67	1.53
	$9 \times 9$	2.44	1.46	1.92	<b>3.01</b>
	$11 \times 11$	1.95	2.27	2.44	1.72

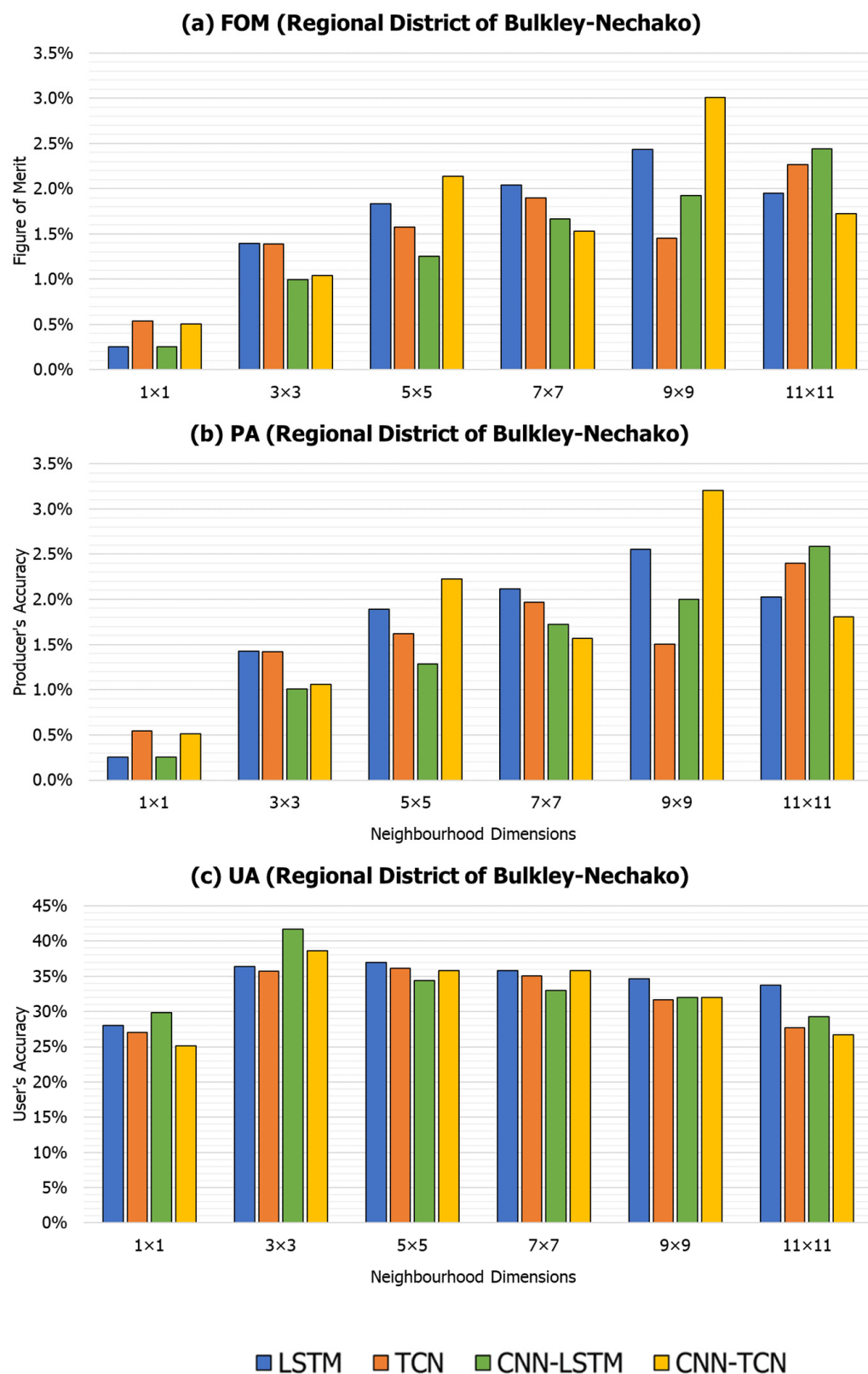


Table 4. Cont.

Region	$M \times M$	Figure of Merit (%)			
		LSTM	TCN	CNN-LSTM	CNN-TCN
R2	$1 \times 1$	1.82	1.65	1.14	1.03
	$3 \times 3$	3.36	3.00	2.09	2.89
	$5 \times 5$	4.31	4.15	2.65	3.56
	$7 \times 7$	4.33	4.61	2.53	3.28
	$9 \times 9$	<b>5.11</b>	3.79	4.73	3.79
	$11 \times 11$	4.52	3.86	4.14	4.24
R3	$1 \times 1$	0.05	0.42	0.08	0.22
	$3 \times 3$	1.62	2.18	1.00	1.40
	$5 \times 5$	2.32	2.23	1.38	1.37
	$7 \times 7$	3.19	3.25	1.53	1.92
	$9 \times 9$	3.09	2.88	2.10	2.47
	$11 \times 11$	2.55	2.79	2.31	<b>3.65</b>
R4	$1 \times 1$	0.60	0.65	0.49	0.59
	$3 \times 3$	3.73	3.52	2.98	3.33
	$5 \times 5$	4.60	4.28	5.76	5.40
	$7 \times 7$	5.93	5.89	6.66	6.11
	$9 \times 9$	7.29	6.86	5.97	6.50
	$11 \times 11$	6.87	<b>7.37</b>	6.43	7.02

**Table 5.** Producer's accuracy (PA) values obtained for (R1) the Regional District of Bulkley-Nechako, (R2) the Regional District of Central Kootenay, (R3) the Northern Rockies Regional Municipality, and (R4) the Cariboo Regional District.

Region	$M \times M$	Producer's Accuracy (%)			
		LSTM	TCN	CNN-LSTM	CNN-TCN
R1	$1 \times 1$	0.25	0.55	0.25	0.51
	$3 \times 3$	1.43	1.42	1.01	1.06
	$5 \times 5$	1.89	1.62	1.28	2.22
	$7 \times 7$	2.11	1.97	1.73	1.57
	$9 \times 9$	2.55	1.50	2.00	3.20
	$11 \times 11$	2.02	2.40	2.59	1.80
R2	$1 \times 1$	1.91	1.74	1.19	1.07
	$3 \times 3$	3.62	3.17	2.19	3.07
	$5 \times 5$	4.86	4.55	2.84	3.96
	$7 \times 7$	4.93	5.22	2.79	3.74
	$9 \times 9$	5.79	4.29	5.53	4.57
	$11 \times 11$	5.08	4.43	5.10	5.55
R3	$1 \times 1$	0.05	0.45	0.08	0.22
	$3 \times 3$	1.66	2.27	1.02	1.44
	$5 \times 5$	2.40	2.30	1.41	1.40
	$7 \times 7$	3.37	3.41	1.58	1.98
	$9 \times 9$	3.25	3.01	2.19	2.59
	$11 \times 11$	2.66	2.93	2.48	4.06
R4	$1 \times 1$	0.60	0.66	0.49	0.59
	$3 \times 3$	3.91	3.70	3.11	3.52
	$5 \times 5$	4.89	4.51	6.37	5.96
	$7 \times 7$	6.47	6.48	7.43	6.80
	$9 \times 9$	8.21	7.67	6.65	7.43
	$11 \times 11$	7.56	8.43	7.30	8.44



**Figure 6.** Measures of (a) Figure of Merit (FOM), (b) Producer's Accuracy (PA), and (c) User's Accuracy (UA) obtained for the full Regional District of Bulkley-Nechako.

**Table 6.** User’s accuracy (UA) values obtained for (R1) the Regional District of Bulkley-Nechako, (R2) the Regional District of Central Kootenay, (R3) the Northern Rockies Regional Municipality, and (R4) the Cariboo Regional District.

Region	$M \times M$	User’s Accuracy (%)			
		LSTM	TCN	CNN-LSTM	CNN-TCN
R1	$1 \times 1$	28.02	27.03	29.82	25.12
	$3 \times 3$	36.42	35.75	41.68	38.66
	$5 \times 5$	36.96	36.10	34.40	35.82
	$7 \times 7$	35.77	35.08	32.95	35.79
	$9 \times 9$	34.69	31.69	32.03	31.96
	$11 \times 11$	33.78	27.71	29.30	26.67
R2	$1 \times 1$	28.99	22.81	20.75	19.23
	$3 \times 3$	30.58	35.00	28.40	31.70
	$5 \times 5$	26.88	31.26	27.36	25.58
	$7 \times 7$	25.56	27.86	20.21	20.15
	$9 \times 9$	29.67	23.75	23.92	17.57
	$11 \times 11$	28.10	22.52	17.20	14.72
R3	$1 \times 1$	30.77	6.27	16.28	21.02
	$3 \times 3$	38.76	35.29	35.27	33.15
	$5 \times 5$	40.26	40.51	38.78	40.38
	$7 \times 7$	36.65	38.67	32.35	34.20
	$9 \times 9$	37.23	37.64	32.65	33.88
	$11 \times 11$	36.07	35.11	24.67	25.44
R4	$1 \times 1$	56.25	40.83	56.41	38.77
	$3 \times 3$	43.75	41.80	42.09	37.48
	$5 \times 5$	43.24	44.15	37.43	35.92
	$7 \times 7$	40.77	39.00	38.74	36.89
	$9 \times 9$	39.06	38.99	36.17	33.76
	$11 \times 11$	42.62	36.38	34.64	28.93

### 3.3. Case 3: Spatial Variables Experiment Results

In Case 3, the model configurations trained with LC data only (Figure 3) from Case 2 were compared with models trained with the addition of spatial variables (Figure 4) for the four separate study regions (R1–R4). For R1, the highest FOM was 4.1%, obtained by CNN-TCN $_{9 \times 9(LC+SVs)}$  (Table 7). This exceeded the largest FOM obtained in Case 2 (3%), which was obtained with CNN-TCN $_{9 \times 9}$  (Figure 6a, Table 4). R2 was associated with several anomalous outcomes. For instance, while many model and neighborhood size configurations for R2 benefitted from the addition of spatial variables, the highest FOM (5.1%) obtained by LSTM $_{9 \times 9}$  in Case 2 was not surpassed. Likewise, this region showed the singular instance (CNN-LSTM $_{9 \times 9}$ ) where CNN-LSTM did not attain an improved FOM value given the spatial variable input. Meanwhile, TCN $_{9 \times 9(LC+SVs)}$  and CNN-TCN $_{9 \times 9(LC+SVs)}$  attained the highest FOM values for R3 and R4, respectively. Across all four study regions and all model types, the overall trends of FOM values increasing with neighborhood size observed in Case 2 continued, with the auxiliary input of static spatial variables providing common drivers of LCC.

Considering the error types associated with the respective model and neighborhood configurations, EQ generally decreased with neighborhood size across models explored in both Case 2 and 3 (Table 8). However, the addition of spatial variables was observed to reduce the EQ for many model types and neighborhood size combinations (Table 8). This means that the amount of real-world changed area and amount of forecasted changed area is typically closer than many of the shown forecasts that were produced with models trained using only the LC data. Meanwhile, EA increased overall as neighborhoods expanded for each model and study region (Table 9), with less distinct patterns between models trained with and without the auxiliary spatial variables. This suggests that while quantity disagreement is reduced with neighborhood size increments and the static spatial variables,

the precise allocations of the greater quantities of changed areas are not necessarily being allocated to the correct real-world locations. This can be observed in the simulation maps where areas are forecasted incorrectly as changed or as the wrong change (Figure 7).

**Table 7.** Figure of Merit (FOM) values obtained for (R1) the Regional District of Bulkley-Nechako, (R2) the Regional District of Central Kootenay, (R3) the Northern Rockies Regional Municipality, and (R4) the Cariboo Regional District. “LC+SV” refers to the model configuration provided the added spatial variables input (Figure 3), with differences between FOM values in Case 2 shown under the respective “Diff” columns for each model type. Values in bold and with \* indicate the FOM is highest for the region across Cases 2 and 3.

Effect of Adding Spatial Variables on Figure of Merit (%)									
Region	$M \times M$	LSTM		TCN		CNN-LSTM		CNN-TCN	
		LC+SV	Diff.	LC+SV	Diff.	LC+SV	Diff.	LC+SV	Diff.
R1	$1 \times 1$	0.29	+0.04	0.64	+0.10	0.86	+0.61	1.02	+0.51
	$3 \times 3$	1.29	−0.10	1.41	+0.02	1.70	+0.70	2.38	+1.34
	$5 \times 5$	1.86	+0.03	1.33	−0.25	3.51	+2.26	3.28	+1.14
	$7 \times 7$	2.26	+0.22	1.76	−0.14	3.44	+1.77	3.40	+1.87
	$9 \times 9$	2.44	+0.00	2.40	+0.94	3.87	+1.95	<b>4.10 *</b>	+1.09
	$11 \times 11$	1.7	−0.25	1.74	−0.53	3.28	+0.84	3.56	+1.84
R2	$1 \times 1$	1.96	+0.14	3.00	+1.35	1.49	+0.35	2.50	+1.47
	$3 \times 3$	3.73	+0.37	3.63	+0.63	3.00	+0.91	2.90	+0.01
	$5 \times 5$	4.15	−0.16	2.33	−1.82	3.53	+0.88	3.33	−0.23
	$7 \times 7$	4.08	−0.25	4.74	+0.13	3.15	+0.62	3.12	−0.16
	$9 \times 9$	5.03	−0.08	3.90	+0.11	4.23	−0.50	3.92	+0.13
	$11 \times 11$	4.9	+0.38	3.59	−0.27	4.25	+0.11	4.64	+0.40
R3	$1 \times 1$	0.24	+0.19	0.73	+0.31	0.87	+0.79	1.37	+1.15
	$3 \times 3$	1.71	+0.09	1.90	−0.28	1.56	+0.56	2.33	+0.93
	$5 \times 5$	2.29	−0.03	1.79	−0.44	1.84	+0.46	2.20	+0.83
	$7 \times 7$	2.29	−0.90	2.59	−0.66	2.06	+0.53	2.68	+0.76
	$9 \times 9$	3.08	−0.01	<b>3.51 *</b>	+0.63	2.71	+0.61	3.07	+0.60
	$11 \times 11$	2.72	+0.17	3.11	+0.32	3.12	+0.81	3.39	−0.26
R4	$1 \times 1$	0.6	0.00	0.95	+0.30	1.14	+0.65	1.79	+1.20
	$3 \times 3$	3.9	+0.17	3.45	−0.07	4.92	+1.94	4.86	+1.53
	$5 \times 5$	5.64	+1.04	5.28	+1.00	6.06	+0.30	5.40	+0.00
	$7 \times 7$	6.81	+0.88	6.69	+0.80	7.03	+0.37	5.24	−0.87
	$9 \times 9$	6.83	−0.46	6.86	0.00	7.32	+1.35	<b>7.40 *</b>	+0.90
	$11 \times 11$	6.72	−0.15	6.87	−0.50	7.04	+0.61	7.17	+0.15

**Table 8.** Area associated with error due to quantity (EQ) values obtained for (R1) the Regional District of Bulkley-Nechako, (R2) the Regional District of Central Kootenay, (R3) the Northern Rockies Regional Municipality, and (R4) the Cariboo Regional District. “LC Only” refers to models considering the LC data as the only input (Figure 2), while “LC+SV” refers to models provided the added spatial variables input (Figure 3). The lowest EQ values for each model type and region are indicated in bold and with \*. ± indicates if LC+SV added or reduced EQ.

Effect of Adding Spatial Variables on Error due to Quantity (km <sup>2</sup> )									
	$M \times M$	LSTM		TCN		CNN-LSTM		CNN-TCN	
		LC Only	LC+SV	LC Only	LC+SV	LC Only	LC+SV	LC Only	LC+SV
R1	$1 \times 1$	4280	4279 (−)	4235	4221 (−)	4283	4214 (−)	4236	4175 (−)
	$3 \times 3$	4153	4160 (+)	4150	4156 (+)	4214	4114 (−)	4203	4025 (−)
	$5 \times 5$	4102	4089 (−)	4130	4166 (+)	4162	3872 (−)	4056	3894 (−)
	$7 \times 7$	4070	4041 (−)	4079	4105 (+)	4098	3865 (−)	4135	3898 (−)
	$9 \times 9$	4006	4001 (−)	4119	3994 (−)	4057	<b>3830 *</b> (−)	3896	<b>3758 *</b> (−)
	$11 \times 11$	4065	<b>3950 *</b> (−)	3953	<b>3891 *</b> (−)	3947	3839 (−)	4035	3801 (−)

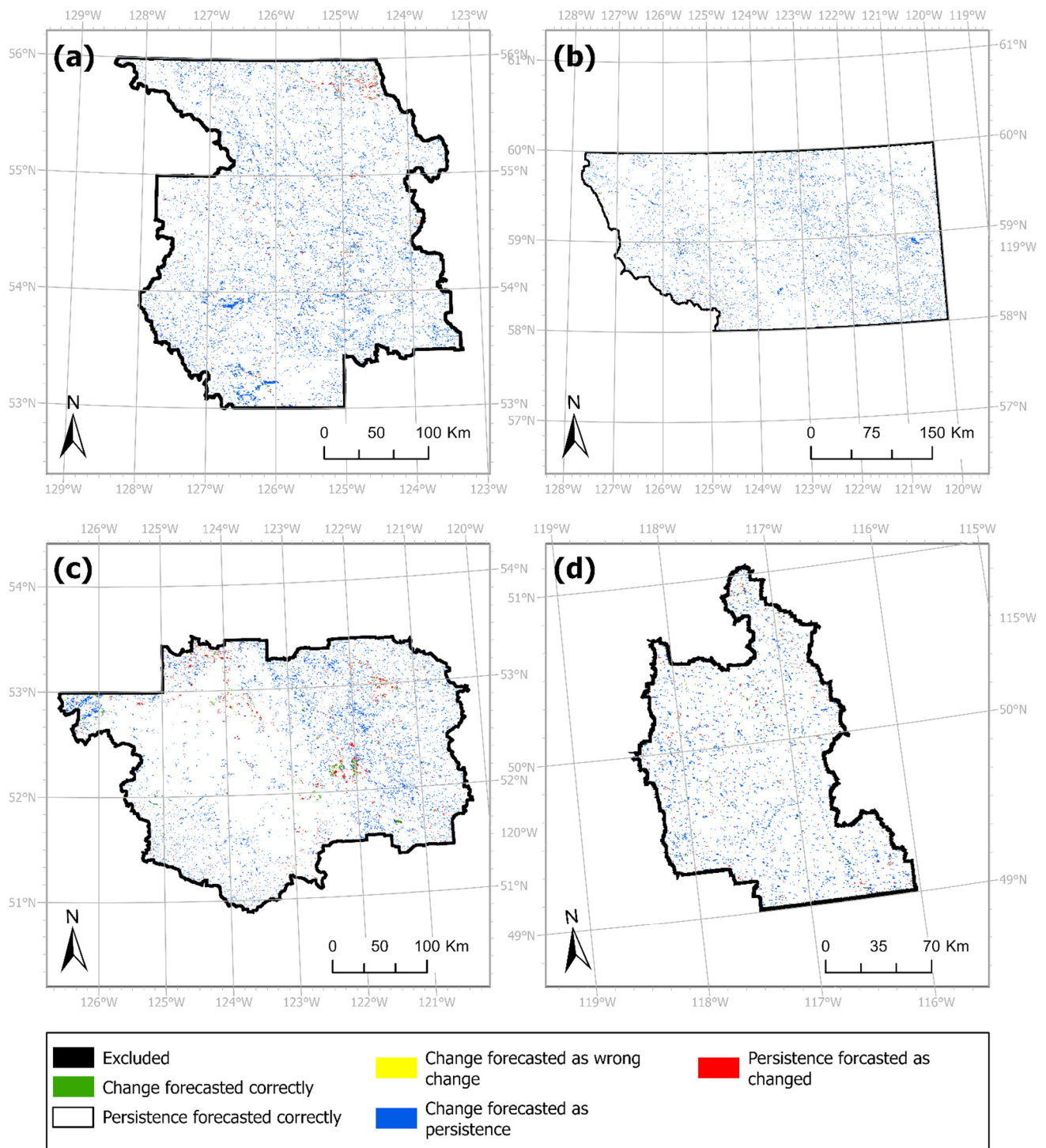


Table 8. Cont.

Effect of Adding Spatial Variables on Error due to Quantity (km <sup>2</sup> )									
	$M \times M$	LSTM		TCN		CNN-LSTM		CNN-TCN	
		LC Only	LC+SV	LC Only	LC+SV	LC Only	LC+SV	LC Only	LC+SV
R2	1 × 1	842	832 (−)	837	785 (−)	853	850 (−)	855	816 (−)
	3 × 3	798	776 (−)	822	773 (−)	835	813 (−)	815	804 (−)
	5 × 5	742	758 (+)	773	805 (+)	812	776 (−)	764	777 (+)
	7 × 7	731	<b>712</b> * (−)	736	729 (−)	783	763 (−)	739	759 (+)
	9 × 9	729	691 (−)	743	726 (−)	700	717 (+)	674	676 (+)
	11 × 11	744	639 (−)	729	<b>649</b> * (−)	645	<b>597</b> * (−)	573	<b>549</b> * (−)
R3	1 × 1	3567	3549 (−)	3318	3371 (+)	3554	3469 (−)	3535	3390 (−)
	3 × 3	3424	3405 (−)	3346	3385 (+)	3471	3420 (−)	3421	3312 (−)
	5 × 5	3364	3362 (−)	3376	3413 (+)	3445	3390 (−)	3452	3343 (−)
	7 × 7	3256	3349 (+)	3265	3281 (+)	3405	3372 (−)	3372	3254 (−)
	9 × 9	3270	3246 (−)	3297	3227 (−)	3343	3291 (−)	3308	3206 (−)
	11 × 11	3316	<b>2969</b> * (−)	3282	<b>2997</b> * (−)	3222	<b>3124</b> * (−)	3018	<b>2974</b> * (−)
R4	1 × 1	3822	3794 (−)	3803	3747 (−)	3830	3678 (−)	3806	3590 (−)
	3 × 3	3521	3489 (−)	3525	3504 (−)	3579	3382 (−)	3503	3341 (−)
	5 × 5	3431	3275 (−)	3474	3289 (−)	3210	3235 (+)	3229	3251 (+)
	7 × 7	3257	3136 (−)	3225	3093 (−)	3130	3080 (−)	3157	3291 (+)
	9 × 9	3060	2989 (−)	3109	3063 (−)	3160	<b>2882</b> * (−)	3023	2868 (−)
	11 × 11	3187	<b>2934</b> * (−)	2976	<b>2803</b> * (−)	3061	2905 (−)	<b>2752</b> *	2807 (+)

**Table 9.** Area associated with error due to allocation (EA) obtained for (R1) the Regional District of Bulkley-Nechako, (R2) the Regional District of Central Kootenay, (R3) the Northern Rockies Regional Municipality, and (R4) the Cariboo Regional District. “LC Only” refers to models considering the LC data as the only input (Figure 2), while “LC+SV” refers to models provided the added spatial variables input (Figure 3). ± indicates if LC+SV added or reduced EA.

Effect of Adding Spatial Variables on Error due to Allocation (km <sup>2</sup> )									
	$M \times M$	LSTM		TCN		CNN-LSTM		CNN-TCN	
		LC Only	LC+SV	LC Only	LC+SV	LC Only	LC+SV	LC Only	LC+SV
R1	1 × 1	52	50 (−)	118	137 (+)	46	132 (+)	118	194 (+)
	3 × 3	205	201 (−)	212	197 (−)	119	255 (+)	137	371 (+)
	5 × 5	266	291 (+)	234	184 (−)	199	567 (+)	331	545 (+)
	7 × 7	311	350 (+)	306	268 (−)	289	588 (+)	228	527 (+)
	9 × 9	402	410 (+)	267	428 (+)	347	617 (+)	565	735 (+)
	11 × 11	330	578 (+)	521	690 (+)	517	651 (+)	409	699 (+)
R2	1 × 1	83	100 (+)	97	173 (+)	74	73 (−)	73	122 (+)
	3 × 3	140	176 (+)	100	183 (+)	92	119 (+)	117	137 (+)
	5 × 5	231	203 (−)	173	147 (−)	126	179 (+)	202	181 (−)
	7 × 7	251	292 (+)	236	247 (+)	186	212 (+)	255	220 (−)
	9 × 9	240	313 (+)	240	270 (+)	303	279 (−)	371	364 (−)
	11 × 11	223	416 (+)	265	423 (+)	420	509 (+)	556	592 (+)
R3	1 × 1	5	27 (+)	474	344 (−)	27	140 (+)	55	260 (+)
	3 × 3	176	206 (+)	288	231 (−)	127	186 (+)	197	343 (+)
	5 × 5	242	249 (+)	226	184 (−)	150	226 (+)	137	292 (+)
	7 × 7	389	273 (−)	367	383 (+)	219	245 (+)	255	431 (+)
	9 × 9	369	416 (+)	332	422 (+)	299	355 (+)	340	494 (+)
	11 × 11	319	983 (+)	367	897 (+)	519	651 (+)	815	920 (+)
R4	1 × 1	31	87 (+)	66	154 (+)	26	276 (+)	65	397 (+)
	3 × 3	379	428 (+)	387	434 (+)	324	552 (+)	444	635 (+)
	5 × 5	483	698 (+)	425	700 (+)	812	740 (−)	805	762 (−)
	7 × 7	709	866 (+)	772	957 (+)	890	953 (+)	882	698 (−)
	9 × 9	969	1139 (+)	912	999 (+)	890	1298 (+)	1102	1318 (+)
	11 × 11	765	1251 (+)	1120	1485 (+)	1038	1279 (+)	1567	1451 (−)



**Figure 7.** Simulation maps of forecasts with the highest FOM for each region: (a) the Regional District of Bulkley Nechako forecasted by CNN-TCN<sub>9×9</sub>(LC+SVs), (b) the Regional District of Central Kootenay forecasted by LSTM<sub>9×9</sub>, (c) the Northern Rockies Regional Municipality forecasted by TCN<sub>9×9</sub>(LC+SVs), and (d) the Cariboo Regional District forecasted by CNN-TCN<sub>9×9</sub>(LC+SVs).

#### 4. Discussion

##### 4.1. Influence of Neighborhood Size

Overall, the general trend across Cases 1–3 was that increasing neighborhood dimensions improved the model's capacity to forecast changes. While UA measures associated with smaller neighborhood settings ( $3 \times 3$  and  $5 \times 5$ ) in Cases 1 and 2 were typically higher,

UA values associated with larger neighborhoods were also affected by more numerous persistent cells forecasted incorrectly as changed. Meanwhile, the FOM and PA measures indicated more real-world changed areas were forecasted correctly with larger neighborhood specifications of  $7 \times 7$ ,  $9 \times 9$ , and  $11 \times 11$  across the comparison of full region forecasts explored in Cases 2 and 3. The highest FOM values were obtained with  $9 \times 9$  neighborhood specifications for each full region.

In Case 1, it was observed that influence of neighborhood size increased within areas exhibiting greater amounts of LCCs. For example, FOM values for Subareas A and C mostly increased with neighborhood size until the  $9 \times 9$  neighborhood specification. With fewer LCCs in Subarea A, the effect of neighborhood sizes was less dramatic. It is also suspected that the LC classes characterizing Subarea A introduced challenges, as imbalances of LC classes and changes are open problems for LCC forecasting with DL [67]. It has also been stated that when real-world maps contain a small percentage of changed areas, it becomes increasingly challenging to forecast correct changes [68]. Thus, further work is needed to explore the size and composition of training datasets, the number of per-category samples available, and expanded neighborhood sizes to improve the forecasting of scarcer changes.

In Case 2, the trend of increasing FOM values was also maintained. The outcomes obtained for the study regions compared in Case 2 aligned with the initial hypothesis that expanding neighborhood sizes would improve the capacity of all DL model types to forecast changes. In Case 3, this trend was not affected by the addition of spatial variables.

#### 4.2. Influence of Model Selection

In Case 1, it was observed that LSTM models yielded high FOM values for low to moderately changed subareas (A and B), while TCN and CNN–TCN models yielded higher FOM values for moderate to highly-changed subareas (B and C) of the Regional District of Bulkley-Nechako (R1). The LSTM models forecasted more real-world LCCs with smaller neighborhood specifications, aligning with findings from previous research that showed LSTMs benefited from small neighborhoods [17]. A similar trend of LSTM benefitting from small neighborhood dimensions more than other model types was observed in Cases 2 and 3 for the Regional District of Central Kootenay (R2), where the amount of change was less than the other full study regions (Tables 3, 4 and 7). This may be indicative that the LSTM models may be suitable for smaller study areas exhibiting sparse or lesser amounts of changes.

In Cases 1–3, it was observed that CNN–LSTM and CNN–TCN generally benefitted from larger neighborhood settings, including  $7 \times 7$ ,  $9 \times 9$ , and  $11 \times 11$ . For example, FOM values obtained for Subareas B and C in Case 1 with CNN–TCN $_{9 \times 9}$  and CNN–TCN $_{11 \times 11}$  were higher than those obtained with LSTM. In Cases 2 and 3, the highest FOM measures were obtained with CNN–TCN $_{9 \times 9}$  (R1) and CNN–TCN $_{11 \times 11}$  (R3). TCN models also routinely obtained FOM, PA, and UA measures comparable to LSTM models in Cases 1 and 2. For example, the TCN models also yielded some of the highest FOM values with the TCN $_{5 \times 5}$  and TCN $_{7 \times 7}$  for Subarea C of Case 1 and R3 of Case 2. This suggests TCN and CNN–TCN models are viable alternatives to LSTM and CNN–LSTM for forecasting LCCs. Consistent improvements in FOM were observed for TCN and CNN–TCN for R2, R3, and R4 of Cases 2 and 3. However, the inconsistent behavior of TCN and CNN–TCN for the full Regional District of Bulkley-Nechako (R1) with increasing neighborhood size requires further consideration. For example, while the CNN–TCN $_{9 \times 9}$  model yielded the highest FOM and PA in Case 1 for the full region, CNN–TCN $_{7 \times 7}$  and CNN–TCN $_{11 \times 11}$  models showed distinctly reduced values for these metrics. Given that the Regional District of Bulkley-Nechako (R1) exhibited the highest amount of changed area among the four study regions, future work involving TCN and CNN–TCN models may benefit from differentiating real-world changes from classification errors to improve the training datasets of areas that exhibit more numerous LCCs.

#### 4.3. Influence of Spatial Variables

In Case 3, FOM measures indicated adding spatial variables improved the capacity to forecast LCCs while simultaneously reducing the quantity of error (EQ) for most models and neighborhood settings. This suggests that the addition of spatial variables reduced biases toward persistent areas, aligning with the expectation that adding the static spatial data of LC drivers would benefit the ability of models to forecast LCCs. The effect of adding spatial variables was generally consistent for CNN–LSTM and CNN–TCN models. However, the additional spatial variables had a lesser effect on LSTM’s capacity to forecast changes with respect to all study regions (Table 7). The increases of FOM and decreases of EQ were also observed as neighborhood sizes expanded, indicating the spatial variables did not cause deviation from the trends observed in Cases 1 and 2. However, reducing the disagreement between quantities of projected and observed changed areas simultaneously increased allocation errors (EA) in many model configurations with the additional spatial variables. Likewise, there were still many changed areas forecasted incorrectly as persistent observed in the simulation maps (Figure 7). The increases of EA with neighborhood size correspond with a previous research study which illustrated that models considering neighborhood effects are not always associated with more precise allocations of changes [56]. Despite the inverse behavior of EQ and EA measures, both are equally important to evaluating LCC model forecasts [63]. This necessitates further work to improve the allocation of changed areas. Future work may also consider the effect of adding spatiotemporally varying drivers of LCCs, such as distance to roads as infrastructure changes over time.

#### 5. Conclusions

This research study demonstrated that increasing neighborhood size generally increase model capacity to forecast LCCs in study areas exhibiting moderate amounts of changes. The key models examined were temporal convolutional networks (TCNs) and CNN–TCNs. Overall, increasing neighborhood sizes generally improved change-focused measures. LSTM showed better capacity to forecast LCCs with smaller neighborhoods in low to moderately changed areas, while CNN–TCN exhibited the highest capacity to forecast LCCs with  $9 \times 9$  and  $11 \times 11$  neighborhoods in moderate to largely changed areas. TCN and CNN–TCN obtained similar change-focused measures as LSTM and CNN–LSTM models for various settings, suggesting these models are a viable choice for LCC modeling tasks in future research studies. Overall, adding spatial variables was generally beneficial for the CNN–LSTM and CNN–TCN models, with improvements observed across in model capacity to forecast LCCs. Likewise, adding the static spatial variables as inputs to the CNN–LSTM and CNN–TCN models reduced errors due to quantity across most neighborhood dimensions explored for each study region. However, the benefits of adding spatial variables were less for the LSTM models.

It is acknowledged that outcomes of this research study hinge upon the datasets selected for this work. For instance, the amount of change forecasted by a model depends upon the amount of change observed in the data used to develop it [62]. Future work can consider exploring the impact of spatial resolution and neighborhood specifications on DL model capacity to forecast LCCs, as these settings may impact LCC modeling with DL techniques. In addition, the selection of the initial map used in three-map comparison measures should be carefully considered for evaluating DL models’ capacities to forecast LCCs, especially with rolling-window forecasting techniques. For instance, the duration of LC class persistence should be explicitly considered, as a one-year forecast may not aptly capture more dynamic areas exhibiting changes over longer time periods. Future work may also consider evaluation strategies to determine how well sequential components of the models capture changes that occur at varying temporal scales. From all EQ and EA measures obtained, it is observed that work is still required to improve model capacity to forecast realistic quantities and allocations LCCs simultaneously before attempting to forecast LC for multiple years. Likewise, the gains and losses of each LC class should be explicitly considered in future research studies.



With the smallest neighborhood dimensions, LCC projections were often little better than utilizing the previous timestep as the forecasted map. However, increasing neighborhood size reduced errors due to quantity of LCCs, suggesting future models and training schemes should first consider neighborhood effects prior to using DL models for forecasting LCCs. Integrating further consideration of the unique characteristics of geographic data may also improve change-focused measures and reduce allocation errors. The results collectively indicate that future LCC modeling studies should investigate adequate neighborhood sizes with respect to new datasets, quantities of observed changes, and model specifications.

**Author Contributions:** Conceptualization, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review & editing, A.v.D. and S.D.; Funding acquisition, Supervision, S.D.; Software, A.v.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Natural Sciences and Engineering Research Council (NSERC) of Canada Postgraduate Scholarship-Doctoral Grant (PGS-D) and Discovery Grant RGPIN-2017-03939.

**Data Availability Statement:** Publicly available datasets were used in this study. The datasets can be found at: <https://lpdaac.usgs.gov/products/mcd12c1v006> (MODIS/Terra+Aqua Land Cover Global Land Cover Dataset), <https://asterweb.jpl.nasa.gov/gdem.asp> (ASTER Digital Elevation Model), <https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm> (Population Centres, Coastal Waters, Rivers, and Digital Boundary Files of Canadian Provinces and Territories), <https://open.canada.ca/data/en/dataset/ac26807e-a1e8-49fa-87bf-451175a859b8> (National Railway Network), and <https://open.canada.ca/data/en/dataset/57d5ffae-3048-4a19-9b4c-eab12f6322c5> (Canadian Census Road Network File).

**Acknowledgments:** The authors are grateful for the full support of this study by Natural Sciences and Engineering Research Council (NSERC) of Canada Postgraduate Scholarship-Doctoral Grant (PGS-D) and Discovery Grant awarded to the first and second authors, respectively. The authors also appreciate the constructive feedback from the three anonymous reviewers. The authors are grateful to the SFU Open Access Fund for sponsoring the publication of this paper.

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

## Appendix A

To compute the change-focused measures in this research study, three-map comparison metrics are selected. Given an initial map, a reference real-world map for the next timestep, and a forecasted map of the next timestep, three-map comparison measures consider changed locations explicitly by computing the differences between the initial and reference real-world maps. The measures, formulations, and interpretations are included in Table A1, where each are comprised of the following components as expressed in previous work [62,63]:

- $A$  = amount of area that underwent real-world change but was forecasted incorrectly as persistent.
- $B$  = amount of area that underwent real-world change and was forecasted correctly as changed.
- $C$  = amount of area that underwent real-world change but was forecasted incorrectly to the wrong land cover class.
- $D$  = amount of area that was remained persistent in the real-world but was forecasted incorrectly as changed.

**Table A1.** Equations of measures used in model assessment.

Measure	Equation	Description and Interpretation	Reference
Figure of Merit (FOM)	$FOM = \frac{B}{(A + B + C + D)} \times 100$	Measure of overlap between real-world and forecasted changes. It provides the ratio of correctly forecasted changes (B) versus the union of projected and reference changes. FOM values assume values from 0–100%, where 0% indicates complete disagreement between real-world and forecasted changes, and 100% indicates perfect agreement between real-world and forecasted changes.	[62,64,65]
Producer's Accuracy (PA)	$PA = \frac{B}{(A + B + C)} \times 100$	Measure indicating the proportion of correctly changed area (B) versus all real-world changes observed. PA values closer to 0% indicate few correctly forecasted areas versus the observed real-world changes, while values closer to 100% suggest that high amounts of observed real-world changes were forecasted correctly.	[62,63]
User's Accuracy (UA)	$UA = \frac{B}{(B + C + D)} \times 100$	Measure indicating the proportion of correctly changed area (B) versus all forecasted changes produced by the model. UA values closer to 0% suggest few correctly forecasted areas versus all projected changes, while values closer to 100% suggest that high amounts of the projected changes intersected with their real-world change locations.	[62,63]
Error due to Quantity (EQ)	$EQ =  D - (A + C) $	Measure of error associated with the amount of changed area forecasted. EQ is expressed as the difference between amounts of area that have undergone changes. If low amounts of changed areas are forecasted compared to the real-world reference data, the EQ will be high. If a model forecasted similar amounts of changed areas to that observed in the real-world, the EQ will be low.	[64–66]
Error due to Allocation (EA)	$EA = 2 \times \min(D, (A + C))$	Measure of error associated with the locations of changed area forecasted. EA is expressed as the area that has been wrongly allocated. If many changes are forecasted at incorrect locations, the EA will be higher. If forecasted changes are allocated to the right locations, then EA will be lower.	[64–66]

## References

1. Foley, J.A.; Defries, R.; Asner, G.P.; Barford, C.; Bonan, G.; Carpenter, S.R.; Chapin, F.S.; Coe, M.T.; Daily, G.C.; Gibbs, H.K.; et al. Global consequences of land use. *Science* **2005**, *309*, 570–574. [[CrossRef](#)] [[PubMed](#)]
2. Van Asselen, S.; Verburg, P.H. Land cover change or land-use intensification: Simulating land system change with a global-scale land change model. *Glob. Chang. Biol.* **2013**, *19*, 3648–3667. [[CrossRef](#)] [[PubMed](#)]
3. Meyer, W.B.; Turner, B.L. Land-use/land-cover change: Challenges for geographers. *GeoJournal* **1996**, *39*, 237–240. [[CrossRef](#)]
4. Gibbard, S.; Caldeira, K.; Bala, G.; Phillips, T.J.; Wickett, M. Climate effects of global land cover change. *Geophys. Res. Lett.* **2005**, *32*, 1–4. [[CrossRef](#)]
5. Seifin, O.; Riese, F.M.; Keller, S. Deep learning for land cover change detection. *Remote Sens.* **2021**, *13*, 78. [[CrossRef](#)]
6. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [[CrossRef](#)]
7. Sun, Z.; Di, L.; Fang, H. Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *Int. J. Remote Sens.* **2019**, *40*, 593–614. [[CrossRef](#)]
8. 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](#)]
9. 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](#)]

10. Luo, C.; Meng, S.; Hu, X.; Wang, X.; Zhong, Y. Cropnet: Deep Spatial-Temporal-Spectral Feature Learning Network for Crop Classification from Time-Series Multi-Spectral Images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Waikoloa, HI, USA, 26 September–2 October 2020; pp. 4187–4190. [\[CrossRef\]](#)
11. 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. [\[CrossRef\]](#)
12. 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. *GISci. Remote Sens.* **2022**, *59*, 501–519. [\[CrossRef\]](#)
13. 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. [\[CrossRef\]](#)
14. 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. [\[CrossRef\]](#)
15. Cao, C.; Dragičević, S.; Li, S. Short-term forecasting of land use change using recurrent neural network models. *Sustainability* **2019**, *11*, 5376. [\[CrossRef\]](#)
16. Liu, Q.; Zhou, F.; Hang, R.; Yuan, X. Bidirectional-Convolutional LSTM Based Spectral-Spatial Feature Learning for Hyperspectral Image Classification. *Remote Sens.* **2017**, *9*, 1330. [\[CrossRef\]](#)
17. Sharma, A.; Liu, X.; Yang, X. Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks. *Neural Netw.* **2018**, *105*, 346–355. [\[CrossRef\]](#)
18. Sulla-Menashe, D.; Friedl, M. The Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) Version 6 data product. In NASA EOSDIS L. Process. DAAC. 2018. Available online: [https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/mcd12q1\\_v006](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1_v006) (accessed on 30 January 2022).
19. Ministry of Municipal Affairs. Regional Districts—Legally Defined Administrative Areas of BC. In Br. Columbia Data Cat. 2020. Available online: <https://catalogue.data.gov.bc.ca/dataset/regional-districts-legally-defined-administrative-areas-of-bc> (accessed on 3 October 2021).
20. van Duynhoven, A.; Dragičević, S. Exploring the sensitivity of recurrent neural network models for forecasting land cover change. *Land* **2021**, *10*, 282. [\[CrossRef\]](#)
21. Kleemann, J.; Baysal, G.; Bulley, H.N.N.; Fürst, C. Assessing driving forces of land use and land cover change by a mixed-method approach in north-eastern Ghana, West Africa. *J. Environ. Manag.* **2017**, *196*, 411–442. [\[CrossRef\]](#)
22. Cao, M.; Zhu, Y.; Quan, J.; Zhou, S.; Lü, G.; Chen, M.; Huang, M. Spatial sequential modeling and predication of global land use and land cover changes by integrating a global change assessment model and cellular automata. *Earth's Futur.* **2019**, *7*, 102–1116. [\[CrossRef\]](#)
23. Phiri, D.; Morgenroth, J.; Xu, C. Long-term land cover change in Zambia: An assessment of driving factors. *Sci. Total Environ.* **2019**, *697*, 134206. [\[CrossRef\]](#)
24. 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. [\[CrossRef\]](#)
25. Guo, A.; Zhang, Y.; Hao, Q. Monitoring and Simulation of Dynamic Spatiotemporal Land Use/Cover Changes. *Complexity* **2020**, *2020*, 3547323. [\[CrossRef\]](#)
26. Statistics Canada. 2016 Census—Boundary Files. 2016. Available online: <https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm> (accessed on 10 May 2022).
27. Statistics Canada. 2016 Census Road Network File. 2016. Available online: <https://open.canada.ca/data/en/dataset/57d5ffae-3048-4a19-9b4c-eab12f6322c5> (accessed on 29 July 2022).
28. Hakim, A.M.Y.; Matsuoka, M.; Baja, S.; Rampisela, D.A.; Arif, S. Predicting land cover change in the mamminasata area, indonesia, to evaluate the spatial plan. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 481. [\[CrossRef\]](#)
29. NASA/METI/AIST/Japan Spacesystems and U.S./Japan ASTER Science Team “ASTER Global Digital Elevation Model V003”. In NASA EOSDIS L. Process. DAAC. Available online: <https://lpdaac.usgs.gov/products/astgtmv003/> (accessed on 10 June 2022).
30. Fotheringham, A.S.; Rogerson, P.A. GIS and spatial analytical problems. *Int. J. Geogr. Inf. Syst.* **1993**, *7*, 3–19. [\[CrossRef\]](#)
31. van Rossum, G. *Python Language Reference*; Python Software Foundation: Amsterdam, The Netherlands, 2009; ISBN 9780954161781.
32. Chollet, F. Keras: The Python Deep Learning library. In Keras.io. 2015. Available online: <https://keras.io/> (accessed on 26 May 2022).
33. 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. *arXiv* **2016**, arXiv:1603.04467.
34. Remy, P. Temporal Convolutional Networks for Keras. 2020. Available online: <https://github.com/philipperemy/keras-tcn> (accessed on 1 May 2022).
35. Oprea, S.; Martinez-Gonzalez, P.; Garcia-Garcia, A.; Castro-Vargas, J.A.; Orts-Escolano, S.; Garcia-Rodriguez, J.; Argyros, A. A Review on Deep Learning Techniques for Video Prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2806–2826. [\[CrossRef\]](#)
36. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#)
37. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [\[CrossRef\]](#)

38. Donahue, J.; Hendricks, L.A.; Guadarrama, S.; Rohrbach, M.; Venugopalan, S.; Darrell, T.; Saenko, K. Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 8–12 June 2015; pp. 2625–2634. [\[CrossRef\]](#)
39. Rawat, A.; Kumar, A.; Upadhyay, P.; Kumar, S. Deep learning-based models for temporal satellite data processing: Classification of paddy transplanted fields. *Ecol. Inform.* **2021**, *61*, 101214. [\[CrossRef\]](#)
40. Pham, V.; Bluche, T.; Kermorvant, C.; Louradour, J. Dropout Improves Recurrent Neural Networks for Handwriting Recognition. In Proceedings of the 2014 14th International Conference on Frontiers in Handwriting Recognition, Hersionissos, Greece, 1–4 September 2014; IEEE: Crete, Greece, 2014; pp. 285–290. [\[CrossRef\]](#)
41. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006; ISBN 0387310738.
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. Gan, Z.; Li, C.; Zhou, J.; Tang, G. Temporal convolutional networks interval prediction model for wind speed forecasting. *Electr. Power Syst. Res.* **2021**, *191*, 106865. [\[CrossRef\]](#)
44. Lara-Benítez, P.; Carranza-García, M.; Riquelme, J.C. An Experimental Review on Deep Learning Architectures for Time Series Forecasting. *Int. J. Neural Syst.* **2021**, *31*, 2130001. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Hewage, P.; Behera, A.; Trovati, M.; Pereira, E.; Ghahremani, M.; Palmieri, F.; Liu, Y. Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. *Soft Comput.* **2020**, *24*, 16453–16482. [\[CrossRef\]](#)
46. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [\[CrossRef\]](#)
47. Hsu, H.K.; Tsai, Y.H.; Mei, X.; Lee, K.H.; Nagasaka, N.; Prokhorov, D.; Yang, M.H. Learning to tell brake and turn signals in videos using CNN-LSTM structure. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–6. [\[CrossRef\]](#)
48. Huang, C.-J.; Kuo, P.-H. A Deep CNN-LSTM Model for Particulate Matter (PM<sub>2.5</sub>) Forecasting in Smart Cities. *Sensors* **2018**, *18*, 2220. [\[CrossRef\]](#)
49. 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. [\[CrossRef\]](#)
50. Nikparvar, B.; Thill, J.-C. Machine Learning of Spatial Data. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 600. [\[CrossRef\]](#)
51. Wan, L.; Zhang, H.; Lin, G.; Lin, H. A small-patched convolutional neural network for mangrove mapping at species level using high-resolution remote-sensing image. *Ann. GIS* **2019**, *25*, 45–55. [\[CrossRef\]](#)
52. Memon, N.; Parikh, H.; Patel, S.B.; Patel, D.; Patel, V.D. Automatic land cover classification of multi-resolution dualpol data using convolutional neural network (CNN). *Remote Sens. Appl. Soc. Environ.* **2021**, *22*, 100491. [\[CrossRef\]](#)
53. Lee, H.; Kwon, H. Going Deeper with Contextual CNN for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [\[CrossRef\]](#)
54. Liu, C.; Zeng, D.; Wu, H.; Wang, Y.; Jia, S.; Xin, L. Urban land cover classification of high-resolution aerial imagery using a relation-enhanced multiscale convolutional network. *Remote Sens.* **2020**, *12*, 311. [\[CrossRef\]](#)
55. Kastaniotis, D.; Tsourounis, D.; Fotopoulos, S. Lip Reading modeling with Temporal Convolutional Networks for medical support applications. In Proceedings of the 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Chengdu, China, 17–19 October 2020; pp. 366–371. [\[CrossRef\]](#)
56. 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. [\[CrossRef\]](#)
57. Kocabas, V.; Dragicevic, S. Assessing cellular automata model behaviour using a sensitivity analysis approach. *Comput. Environ. Urban Syst.* **2006**, *30*, 921–953. [\[CrossRef\]](#)
58. 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. [\[CrossRef\]](#)
59. Kong, Y.L.; Huang, Q.; Wang, C.; Chen, J.; Chen, J.; He, D. Long short-term memory neural networks for online disturbance detection in satellite image time series. *Remote Sens.* **2018**, *10*, 452. [\[CrossRef\]](#)
60. Foody, G.M. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. *Remote Sens. Environ.* **2020**, *239*, 111630. [\[CrossRef\]](#)
61. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [\[CrossRef\]](#)
62. Pontius, R.G.; Boersma, W.; Castella, J.C.; Clarke, K.; 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. [\[CrossRef\]](#)
63. Paegelow, M.; Camacho Olmedo, M.T.; Mas, J.; Houet, T. Benchmarking of LUCC modelling tools by various validation techniques and error analysis. *CyberGeo* **2015**, *2014*. [\[CrossRef\]](#)
64. Yubo, Z.; Zhuoran, Y.; Jiuchun, Y.; Yuanyuan, Y.; Dongyan, W.; Yucong, Z.; Fengqin, Y.; Lingxue, Y.; Liping, C.; Shuwen, Z. A Novel Model Integrating Deep Learning for Land Use/Cover Change Reconstruction: A Case Study of Zhenlai County, Northeast China. *Remote Sens.* **2020**, *12*, 3314. [\[CrossRef\]](#)



- 
65. Shoyama, K. Assessment of land-use scenarios at a national scale using intensity analysis and figure of merit components. *Land* **2021**, *10*, 379. [[CrossRef](#)]
  66. Yang, Y.; Zhang, S.; Liu, Y.; Xing, X.; De Sherbinin, A. Analyzing historical land use changes using a Historical Land Use Reconstruction Model: A case study in Zhenlai County, northeastern China. *Sci. Rep.* **2017**, *7*, 41275. [[CrossRef](#)] [[PubMed](#)]
  67. 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](#)]
  68. 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](#)]