International Journal of Applied Earth Observations and Geoinformation 114 (2022) 103060

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Forecasting vegetation dynamics in an open ecosystem by integrating deep learning and environmental variables

Yue Ma [a], Yingjie Hu [a,*], Glenn R. Moncrieff [b,c], Jasper A. Slingsby [d,b], Adam M. Wilson [a], Brian Maitner [a], Ryan Zhenqi Zhou [a]

[a] *Department of Geography, University at Buffalo, The State University of New York, Buffalo, NY 14260, USA*
[b] *Fynbos Node, South African Environmental Observation Network, Private Bag X7, Rhodes Drive, Claremont 7735, South Africa*
[c] *Centre for Statistics in Ecology, Environment and Conservation, Department of Statistical Sciences, University of Cape Town, Private Bag X3, Cape Town 7701, South Africa*
[d] *Department of Biological Sciences and Centre for Statistics in Ecology, Environment and Conservation, University of Cape Town, 7701 Cape Town, South Africa*

ARTICLE INFO

Keywords:
Open ecosystems
NDVI
RNN
ConvLSTM
Cape Floristic Region
GeoAI

ABSTRACT

Open (i.e., non-forest) ecosystems, such as savannas, shrublands, and grasslands, contain over 40 % of the global total ecosystem organic carbon and harbor a substantial portion of the world's biodiversity. Accurately forecasting vegetation dynamics is critical for managing biodiversity, fire, water, and carbon in these open ecosystems. Unlike forests or other relatively stable ecosystems, open ecosystems can have dramatically changing vegetation states since they are prone to natural disturbances, long-term trends, and short-term events. Consequently, it is challenging to accurately predict vegetation state in this type of ecosystems. This paper investigates the use of deep learning based approaches for forecasting vegetation dynamics in an open ecosystem, the fynbos shrublands of the Cape Floristic Region of South Africa, a global biodiversity hotspot. We experiment with different deep learning models and examine the ability of thirteen environmental variables, such as precipitation, fire history, and temperature, to enhance the forecasting. We find that the ConvLSTM model can forecast vegetation state more accurately than four other compared baseline approaches. The environmental variable mean precipitation in July (winter) provides the most prominent enhancement for forecasting among the tested variables. Finally, we discuss the pros and cons of using a deep learning based approach for vegetation forecasting in open ecosystems from a conservation management perspective.

1. Introduction

Open ecosystems, such as savannas, shrublands, and grasslands, cover about 33.9 % of the total vegetated areas on the surface of the Earth (Bond, 2019). They make up over 40 % of the global total ecosystem organic carbon (Prentice et al., 2001) and harbor a substantial proportion of the world's biodiversity. Open ecosystems are often more sensitive to climate change than forest ecosystems but have received relatively less attention so far (McNicol et al., 2018; Sleeter et al., 2018; Duncanson et al., 2019). Accurately forecasting vegetation states in open ecosystems is critical for effectively managing their biodiversity, fire, water, and carbon, which in turn contributes to a number of the United Nations Sustainable Development Goals (SDGs), including Goal 3: *Good Health and Well-being*, Goal 6: *Clean Water and Sanitation*, Goal 13: *Climate Action*, and Goal 15: *Life on Land*.

Forecasting vegetation dynamics in open ecosystems, however, is challenging. Unlike forests or other ecosystems with relatively stable vegetation states, open ecosystems have complex natural dynamics as they are prone to various natural disturbances and seasonality (Slingsby et al., 2020). In open ecosystems, trees may be present but are not dominant, and this type of vegetation composition makes open ecosystems more susceptible to short-term environmental events, such as fire, drought, and temperature extremes. Accordingly, the vegetation states of open ecosystems can have dramatic changes, making them more difficult to forecast.

Multidecadal and continuous Earth observations from satellites and airplanes have provided rich data sets for studying open ecosystems and forecasting their vegetation dynamics. Based on the Moderate Resolution Imaging Spectroradiometer (MODIS) Enhanced Vegetation Index (EVI) time series data, Watts and Laffan (2014) assessed the ability of the

* Corresponding author.
E-mail address: yhu42@buffalo.edu (Y. Hu).

https://doi.org/10.1016/j.jag.2022.103060
Received 15 July 2022; Received in revised form 3 October 2022; Accepted 11 October 2022
Available online 23 October 2022

Breaks for Additive Seasonal and Trend (BFAST) algorithm to detect abrupt changes in an open ecosystem, a semi-arid region of New South Wales, Australia. They found that BFAST was able to detect between 68 % and 79 % abrupt vegetation changes caused by flooding but missed most changes caused by fires (only detected 3 % of the known changes). Browning et al. (2017) also examined the ability of the BFAST algorithm to detect vegetation changes in a grassland open ecosystem in southern New Mexico based on MODIS Normalized Difference Vegetation Index (NDVI) time series data from 2000 to 2013, and found that BFAST was able to detect about 75 % of the known abrupt changes in vegetation states. Leveraging a 10-year archive of biweekly MODIS satellite imagery data, Wilson et al. (2015) developed a Hierarchical Bayesian (HB) model to monitor and forecast the vegetation dynamics in an open ecosystem, which can explain about half (with an $R^2 = 0.47$) of the vegetation dynamics following a fire. Slingsby et al., (2020) further designed a workflow based on this HB model to detect the abnormal changes in the vegetation states (e.g., due to illegal vegetation removal, drought, or invasion by alien trees) by examining the deviation of the observed NDVI signals from the model predictions.

Along with the availability of Earth observation data, the rapid advancements of artificial intelligence (AI) techniques provide new opportunities for enhancing forecasting of vegetation dynamics of open ecosystems. Researchers have developed and trained various deep learning models to address a wide range of social and environmental challenges, such as poverty estimation (Burke et al., 2021), precipitation forecasting (Shi et al., 2015), land cover classification (Rußwurm and Körner, 2018), population mapping (Huang et al., 2021), urban functional region detection (Yang et al., 2022), and many others. These studies have shown that deep learning models have good performance in many tasks, especially when the objective of a task can be formalized as classification or forecasting. Some studies specifically focused on the use of deep learning models for forecasting vegetation states and detecting abnormal vegetation changes. (Reddy and Prasad, 2018) used a long short-term memory (LSTM) model to predict NDVI in the evergreen forest in the interior of the Great Nicobar Islands and mangroves on the coastline of the island. Irvin et al. (2020) developed ForestNet, a convolutional neural network (CNN) model, for classifying the drivers of deforestation based on Landsat 8 images. Ahmad et al. (2020) used a convolutional long short-term memory (ConvLSTM) model to forecast the NDVI of soybean crop fields with the goal of better predicting crop yield and planning agricultural activities. While deep learning based approaches have been used for forecasting vegetation dynamics in an increasing number of studies (Ferchichi et al., 2022), there is a lack of research that investigates their use in open ecosystems where vegetation states can change dramatically due to natural disturbances and that examines the pros and cons of these approaches from a conservation management perspective.

This paper aims to address such a knowledge gap by investigating the use of different deep learning models for forecasting vegetation dynamics in an open ecosystem, the fynbos shrublands of the Cape Peninsula in the Cape Floristic Region (CFR) of South Africa. The CFR is a Global Biodiversity Hotspot (Myers et al., 2000; MacFadyen et al., 2022) and a UNESCO World Heritage Site. The vegetation variable that we aim to predict is NDVI derived from MODIS imagery. Since the NDVI data is in the form of a time series of 2D images, we formalize this vegetation dynamics forecasting task into a time series prediction problem: *given a time series of NDVI images for the study region in the past time steps, predict NDVI images in future time steps.* The main model of interest that we investigate is the ConvLSTM model proposed by Shi et al. (2015) which has shown outstanding performance in prediction tasks based on time series of satellite images (Xiao et al., 2019; Ahmad et al., 2020; Boulila et al., 2021). In addition, we also test two other deep learning models commonly used for time series forecasting: a vanilla RNN model and a fully-connected LSTM (FC-LSTM) model. Both models have been used for NDVI time-series predictions previously (Stepchenko, 2015; Reddy and Prasad, 2018; Rhif et al., 2020), but they did not use convolutional operations to analyze the image at each time step. We further examine whether and to what extent additional environmental variables, such as precipitation, fire history, and vegetation type, can help improve forecasts, and how accurately the best deep learning-based approach can forecast vegetation dynamics in longer time periods ahead (e.g., 1 year ahead).

2. Study area and data

The study area is the Cape Peninsula located at the south-western tip of South Africa (Fig. 1). This area is part of the CFR, and most of its vegetation is fynbos that is subject to various natural disturbances such as fire and seasonality. The indigenous vegetation in this region is threatened by climate change, habitat loss, and invasion of alien species (Rouget et al., 2003; Ntshanga et al., 2021; Skowno et al., 2021). Accurately forecasting the vegetation dynamics in this area can help conservation managers quickly identify abnormal changes and inform conservation management decisions.

We downloaded the MODIS NDVI scenes from the MOD13Q1 (Didan, 2015a) and MYD13Q1 (Didan, 2015b) Version 6 products for the study area from Google Earth Engine for the time period of 2000-02-18 to 2017-07-20. In total, we obtained 748 NDVI images covering the study area organized in a time series (i.e., 748 time steps), and the time interval between two consecutive images is 8 days. All the NDVI images have the same size of 172 (height) by 63 (width) with a spatial resolution of 250 m. The original NDVI data contain noise but come with a quality assessment (QA) file indicating the quality of each pixel. We used the pixels with a QA value of 0 ("Good data; use with confidence"), and applied the savgol filter to smooth the noise pixel values based on the values in nearby time steps of the same pixels (Jonsson and Eklundh, 2002). These 748 NDVI images have different numbers of good pixels and noise pixels due to different cloud covers and other noise factors at each time step. The savgol filter is applied to only the noise pixel values, and its window length is set to seven, i.e., a noise value is smoothed based on the values in three nearby time steps on its two sides. The smoothed results of nine randomly selected pixels are visualized in Supplementary Figure S1. To further quantify the effect of the savgol filter, we measure the means and standard deviations of the original and smoothed NDVI series. The original NDVI has a mean of 0.510 and a standard deviation of 0.167, and the smoothed NDVI has a mean of 0.512 and a standard deviation of 0.142. Thus, the savgol filter largely keeps the mean value of the original NDVI data while reducing the overall data variance by smoothing out noise values. In addition, pixels covered with clouds have a value of 3 ("Cloudy; target not visible, covered with cloud") in the QA file, and their values are also smoothed by the savgol filter. With the smoothed NDVI time series images, we then used the first 75 % of the images as the training data and the remaining 25 % images as the test data.

In addition to NDVI images, we also used environmental data to examine whether and to what extent environmental variables can help a deep learning model make better predictions. Based on previous studies (Wilson et al., 2015; Slingsby et al., 2020), we included 13 environmental variables that are likely to influence vegetation dynamics in the studied open ecosystem, which are: (1) elevation, (2) slope, (3) aspect, (4) topographic position index (TPI), (5) topographic roughness index (TRI), (6) highest temperature in January (summer in South Africa), (7) lowest temperature in July (winter), (8) mean precipitation in January, (9) mean precipitation in July, (10) mean solar radiation in January, (11) mean solar radiation in July, (12) vegetation type, and (13) fire history. The environmental variables related to temperature, precipitation, and solar radiation can change dramatically within short time periods. We choose to use static rather than dynamic measurements for these variables based on a theoretical rationale. The plant species that live in our studied open ecosystem vary across space, and that variability is driven in large part by the environmental spatial heterogeneity. For example, different species live in shaded ravines (low solar radiation)

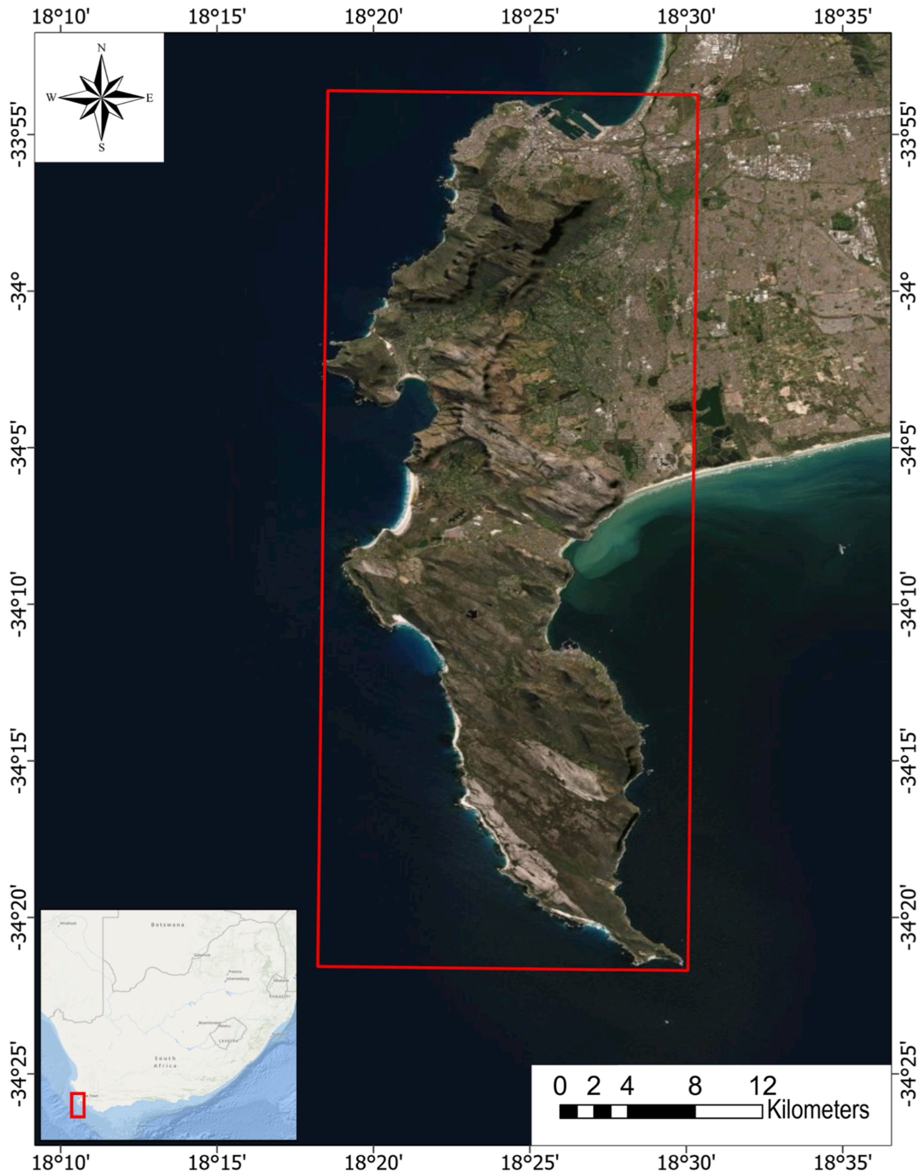

Fig. 1. The study area of the Cape Peninsula located at the south-western tip of South Africa.

compared with mountain tops (high solar radiation). Using static environmental variables allows the model to better focus on spatial heterogeneity, and similar static variables have also been used in previous ecological studies (Guisan and Zimmermann, 2000; Bucklin et al., 2015; Wilson et al., 2015). While adding dynamic solar radiation and other dynamic environmental variables may help the model predict some short-term variability of vegetation activity captured by NDVI, these dynamic variables could also introduce substantial noise due to their large value changes in short time intervals, which may not directly contribute to the vegetation growth in the studied open ecosystem at longer time scales. With these considerations, we choose to focus on static environmental variables in this study.

We obtained data for the 13 environmental variables in the following ways. For elevation, we downloaded the 10 m Digital Elevation Model (DEM) data from the open data portal of the City of Cape Town. The variables of slope, aspect, and topographic indices were then derived from this DEM data. The variables related to temperature, precipitation, and solar radiation were extracted from the CHELSA climatologies (Karger and Zimmermann, 2019). We obtained the vegetation type layer from the City of Cape Town's Environmental Management Department through the city's open data portal. The fire history variable was included in the form of vegetation age calculated using the Table Mountain National Park (TMNP) fire scar database which has been maintained by South African National Parks (Forsyth and Van Wilgen, 2008). All of the 13 environmental variable data layers were resampled to match the resolution of the NDVI pixels using bilinear interpolation. Because each environmental data layer has missing values located at different pixels, we designed a simple mask to keep only those pixels that have valid values in all environmental data layers. Fig. 2 shows 12 of these 13 environmental data layers, excluding fire history. Fire history data are in the form of a time series of images (the same as the NDVI data).

3. Methods

3.1. Overview of experimental design

In this study, we aim to answer three research questions (RQs) through three sets of experiments. The three questions are: RQ1: how accurately can different deep learning models forecast vegetation dynamics in the studied open ecosystem based on NDVI time series data? RQ2: whether and to what extent can different environmental variables

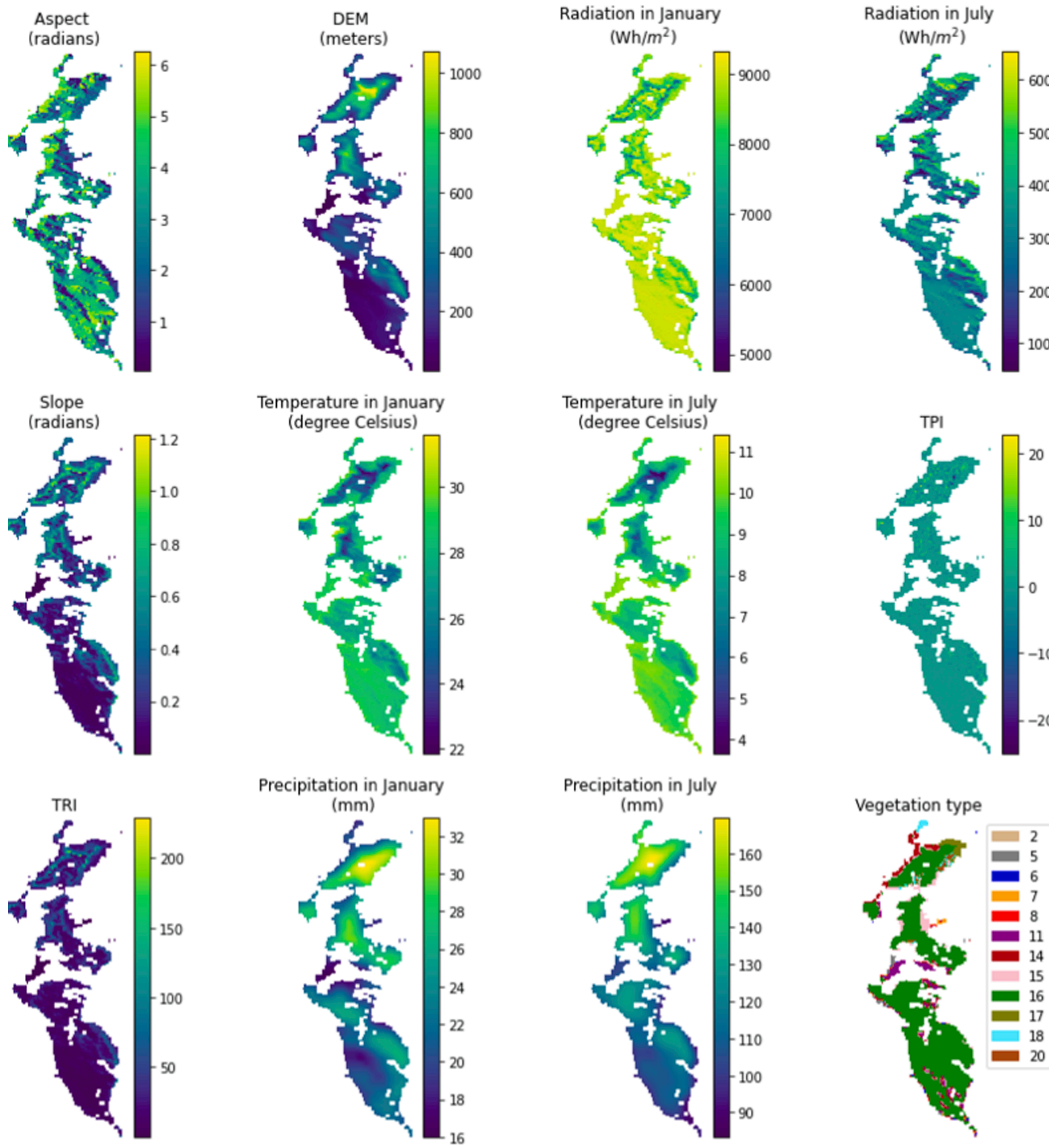

Fig. 2. Maps of twelve environmental data layers used in this study (excluding fire history). The meanings of the vegetation type codes (in the lower-right vegetation type subfigure) are: ***2***: Beach; ***5***: Cape Flats Dune Strandveld - False Bay; ***6***: Cape Flats Dune Strandveld - West Coast; ***7***: Cape Flats Sand Fynbos; ***8***: Cape Lowland Freshwater Wetlands; ***11***: Hangklip Sand Fynbos; ***14***: Peninsula Granite Fynbos - North; ***15***: Peninsula Granite Fynbos - South; ***16***: Peninsula Sandstone Fynbos; ***17***: Peninsula Shale Fynbos; ***18***: Peninsula Shale Renosterveld; ***20***: Southern Afrotemperate Forest. See Fig. 1 for region location and scale.

help enhance vegetation forecasting? RQ3: how accurately can the best deep learning based approach forecast vegetation dynamics in longer time periods ahead? To answer RQ1, we train and test three deep learning models, i.e., RNN, FC-LSTM, and ConvLSTM, using NDVI time series data. To answer RQ2, we add the 13 environmental variables to the input of the best deep learning model identified in RQ1 and conduct an ablation study by removing one environmental variable at a time and measuring how the forecasting accuracy changes. To answer RQ3, we use the best deep learning model with the best set of environmental variables identified in RQ2, and examine its forecasting accuracy at different time steps ahead, from 1 step ahead (8 days) to 46 steps ahead (about 1 year). An overview of the experimental design is provided in Fig. 3.

3.2. *Deep learning models*

Three different RNN models are used in this study. In the following, we provide brief descriptions about their core ideas, and more details about these models are available in related textbooks and papers (Hochreiter and Schmidhuber, 1997; Shi et al., 2015; Bengio et al., 2017; Géron, 2019).

- *Vanilla recurrent neural network (RNN)*: The vanilla RNN model, or simply RNN model, is a neural network model suitable for handling sequential data (Bengio et al., 2017), such as the NDVI time series data in this study. The RNN model processes data through a sequence of time steps, and at each time step, the neurons of the model take into account not only the input data from the current time step but

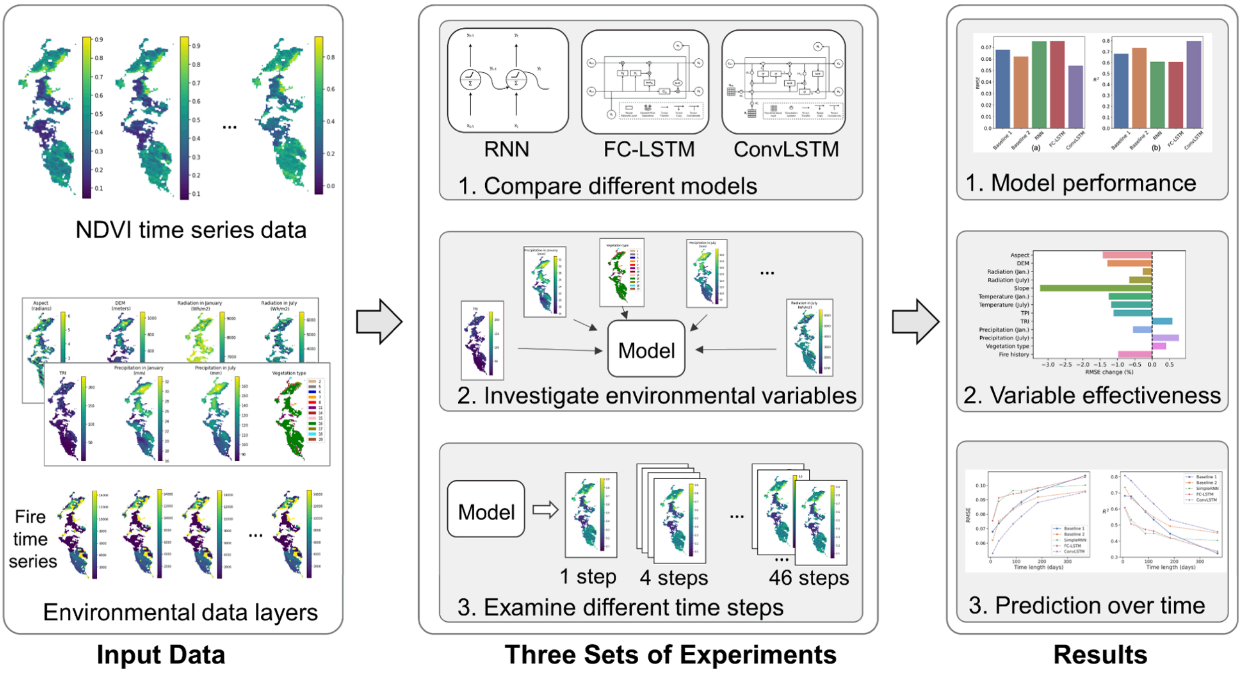

Fig. 3. An overview of the experimental design of this study.

also the model output from the previous time step. In this way, the RNN model can be considered as having a "memory" of the past, and such memory can help the RNN model make effective predictions about the future when there exists a temporal autocorrelation among the data at different time steps, such as a time series of NDVI values from the same pixel. To use RNN in this study, the 2D NDVI image at each time step needs to be flattened into a 1D vector before it is fed to the model.

- *Fully connected long short-term memory (FC-LSTM)*: While vanilla RNN already considers previous time steps, its ability to handle longer time series is rather limited due to gradient vanishing or exploding problems (Bengio et al., 1994). To overcome this limitation, a variant of RNN, long short-term memory model, was proposed (Hochreiter and Schmidhuber, 1997). In this model, neurons and computational operations are organized into three gates: input gate, forget gate, and output gate. The input gate decides which part of the information from the current and previous time steps should be added to the memory of the model, the forget gate decides which part of the information should be forgotten, and the output gate decides the output to be passed to the next step. This design allows the LSTM model to keep both long-term information from many previous steps away and short-term information from more recent time steps. A FC-LSTM model has each neuron in one layer fully connected to the neurons or input values in the previous layer. To use FC-LSTM in this study, the 2D NDVI image at each time step also needs to be flattened into a 1D vector.
- *Convolutional long short-term memory (ConvLSTM):* While the FC-LSTM model can handle longer time series data, it needs to process the data at each time step as a 1D vector. Consequently, it loses important spatial information among the input values when the data at each time step is a 2D image, such as video data or a time series of satellite images. In these and other 2D image data, spatial autocorrelation often exists as pixels nearby tend to have similar values. To capture the potential spatial autocorrelation among the input values, the ConvLSTM model was proposed (Shi et al., 2015), which uses the backbone of a LSTM model but leverages convolutional operations to process the 2D image at each time step without first flattening it into a 1D vector. In this way, the ConvLSTM model preserves the spatial positions of the input pixels, and can learn their potential spatial autocorrelation at the model training stage. Meanwhile, the LSTM backbone allows ConvLSTM to capture the temporal autocorrelation of data in current and previous time steps. To use ConvLSTM in this study, the 2D NDVI image at each time step does not need to be flattened, and the model directly takes a time series of NDVI images as the input and predicts NDVI images in one or multiple future time steps.

For these three models, we conduct hyperparameter tuning to identify the best model architectures using the random search function. The search spaces set for the hyperparameters are: [1, 10] (with a step of 1) for the number of hidden layers, [10, 80] (with a step of 10) for the number of filters for the ConvLSTM layers, {$P/100$, $P/50$, $P/20$, $P/10$} (where P is the total number of pixels in an NDVI image) for the number of neurons for the RNN and FC-LSTM layers, {3, 5, 7} for the kernel size, and {'linear', 'sigmoid'} for the activation function. The Python package KerasTuner is used to implement this tuning process. The best model architectures are then identified based on the hyperparameter tuning results.

3.3. Evaluation metrics

To evaluate the performance of different approaches, we use two metrics: root mean square error (RMSE) and R squared (R^2). RMSE measures the average deviation of the predicted NDVI values from the observed NDVI values (Equation (1)). The lower the RMSE, the better a forecasting approach is.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{y}_i - y_i)^2}, \quad (1)$$

where N is the total number of test pixels in the test NDVI images, $\widehat{y}_i$ is the predicted NDVI value of the ith pixel, and y_i is the observed NDVI value. The second metric R^2 is calculated using Equation (2):

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2}, \quad (2)$$

It measures the overall consistency between the predicted NDVI values and the observed NDVI values. The higher the R^2 value, the better a forecasting approach is. The test pixels are from the NDVI images completely held out from training, and we only use those pixel values with a QA value of 0 (i.e., "Good data; use with confidence") for calculating RMSE and R^2. In addition, since we used the savgol filter to smooth noise values, some test pixel values could be "leaked" to the input data when noise values exist in the previous three steps of a test value (because a window size of seven was used for the savgol filter). We excluded these test values from evaluation to enhance the robustness of results.

4. Results

4.1. Model performance

We first present results regarding the performance of different models in vegetation dynamics forecasting in the studied open ecosystem. The model performance is measured based on a 15-to-1 forecasting task, in which each model is given the NDVI images of the study area in the previous 15 time steps (about 120 days), and is asked to forecast the NDVI in the next time step (i.e., the 16th time step). We use the setting of previous 15 time steps in this forecasting task, because it provides a moderate computational efficiency and also a lot of information is already provided in a 120-day time period for a model to make predictions. All three models were trained using the same NDVI images in the first 561 time steps (75 % of the total data) and were tested on the same remaining 187 time steps (25 % of the data). In addition to the three deep learning models, we also use two naive forecasting baselines, which are: Baseline 1: it simply predicts the NDVI at the next step as the same as the NDVI at the last step, and Baseline 2: it predicts the NDVI at the next step as the average of the previous 15 time steps. While the two baselines can be considered as naive forecasting, they are not that naive since the NDVI in the next time step in most cases will not be largely different from its previous time step or the average of the previous 15 time steps. Fig. 4 shows the RMSE, R^2, training time, and prediction time of the five different approaches. All experiments are conducted on the Microsoft Azure cloud computing platform with the Standard NC6 virtual machine with Intel Xeon E5-2690 v3 processor (6 cores), 56 GB

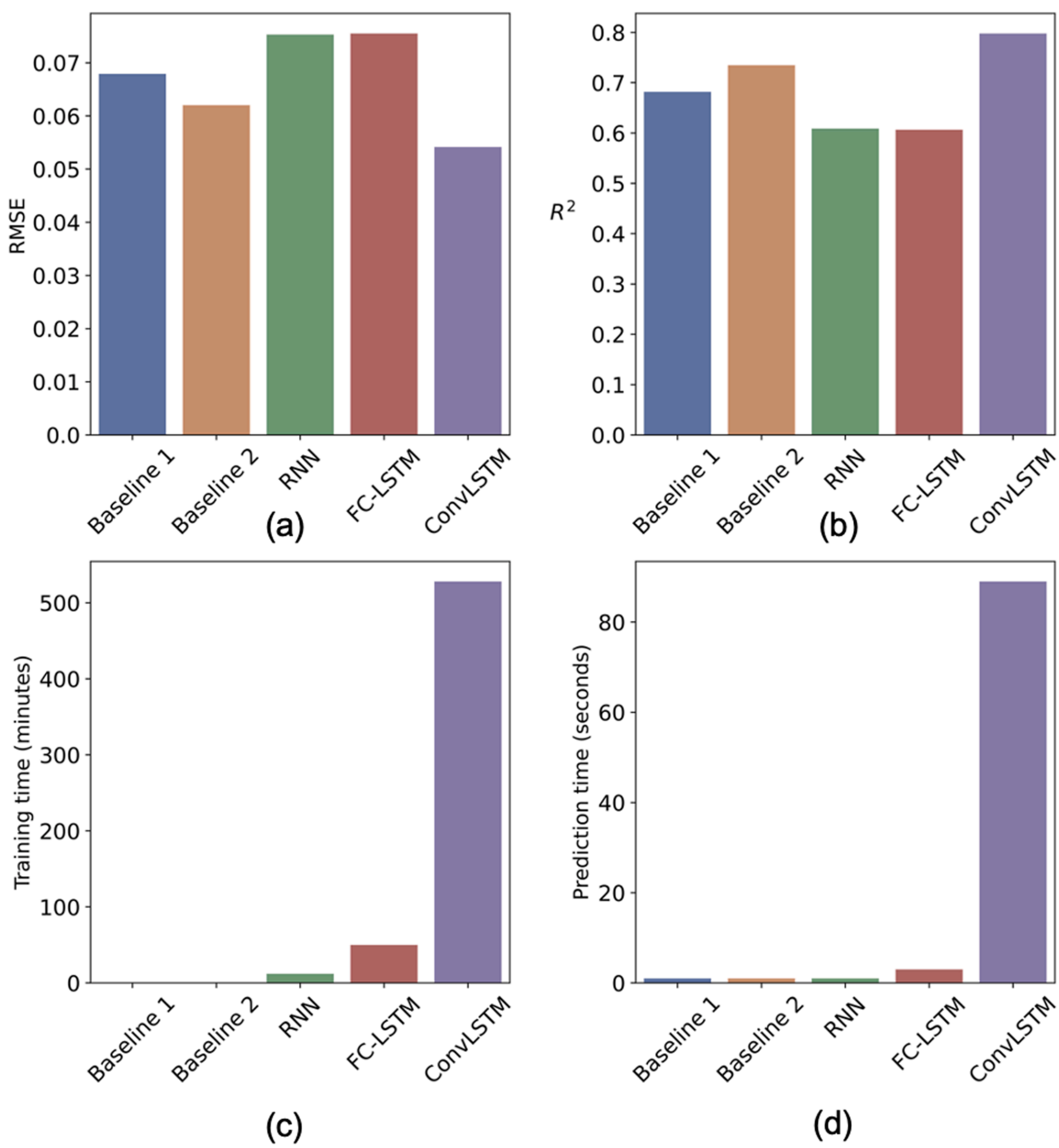

Fig. 4. Performance of the five different approaches for vegetation dynamics forecasting for the study area in a 15-to-1 forecasting task: (a) RMSE; (b) R^2; (c) training time; and (d) prediction time.

RAM, 380 GB disk, and NVIDIA Tesla K80 GPU.

The two naive baselines had RMSEs between 0.062 and 0.068 and R^2 between 0.68 and 0.73 when forecasting the next-step vegetation state. Both baselines performed better than RNN and FC-LSTM which achieved an R^2 of about 0.61. We believe that these two simple baselines achieved better performance than the two more complex models largely because these two baselines are particularly fit for the task of next-step prediction. Despite the strong baselines, the ConvLSTM model makes the most accurate predictions by achieving the highest R^2 of 0.798. In terms of training time, the ConvLSTM model requires the longest training time of over 8 h, while the RNN and FC-LSTM models can be trained within an hour (the two baselines do not need to be trained). In terms of prediction time, the ConvLSTM model takes 89 s to predict all the NDVI values in the future 187 time steps, while the other approaches can finish their predictions within 3 s.

Given the overall higher forecasting accuracy of ConvLSTM, we ask the question: does this higher forecasting accuracy come from one (or a few) subregions of the study area, or is this higher forecasting accuracy more evenly distributed across the entire study area? To answer this question, we compute the spatial RMSE for each of the five approaches, in which RMSE is computed for each pixel over all the test time steps. The results are shown in Fig. 5, in which lighter gray indicates higher RMSE and darker gray indicates lower RMSE. The spatial RMSE figure of ConvLSTM is darker than the other four approaches across the entire study area, suggesting that the higher forecasting accuracy does not come from merely one or a few subregions.

To intuitively see the predictions of ConvLSTM, we randomly select six pixels and visualize their observed and predicted NDVI time series in Fig. 6. We also indicate the locations of the pixels on the map of the study area at the center of the figure, and indicate the low-quality pixel values (based on the QA file) in the time series plots. The upper left and the lower right time series plots contain high percentages (over 75 %) of low-quality values. Their corresponding pixels are located close to the coastline, and these low-quality values are likely due to influences from the ocean. The other four time series plots contain small percentages (within 25 %) of low-quality values and their corresponding pixels are located more toward the inland area. In the middle subfigure on the left side of Fig. 6, there is a sudden drop of NDVI followed by a gradual vegetation recovery. The fire history data of the corresponding pixel show that a fire occurred at that time step, likely causing this sudden drop of NDVI. This example shows that the trained ConvLSTM model can still provide fairly accurate predictions when there is a major disturbance (i.e., fire). Overall, the predicted NDVI time series of ConvLSTM are consistent with the observed NDVI time series, when the percentages of low-quality pixels are small. Note that low-quality pixel values are not used for calculating R^2 and RMSE since their true NDVI values are unknown.

It is worth noting that the three deep learning models were trained with time series of NDVI images only, and no environmental variables were used in this set of experiments. The reason that we did so is that other environmental data layers may not always be available. For example, most of the environmental data layers used in this study were created and maintained by local authorities or organizations in CFR of South Africa, such as the City of Cape Town's Environmental Management Department. When there is a lack of relevant local authorities or organizations, or when the relevant local authorities or organizations lack sufficient resources to create and maintain such data sets, these environmental data layers may not be available. By contrast, global-scale NDVI data are readily available from public satellite data sources. Thus, testing the performance of models trained with NDVI data only can help us understand how accurately these models can forecast vegetation states in more general situations.

4.2. The ability of environmental variables for enhancing forecasting

In our study area of the Cape Peninsula, we are fortunate to have a number of environmental data layers created and maintained by relevant local authorities and organizations. Here, we present the results of including the 13 environmental variables into the forecasting model (in addition to NDVI time series) to understand whether and to what extent these environmental variables can help further improve the forecasting accuracy. We focus on the ConvLSTM model which has shown the best performance in our previous set of experiments, and we use the same 15-to-1 forecasting task as used previously in order to compare the new performance of the model with not using environmental variables. To include environmental variables, we add them as additional channels into the input of the ConvLSTM model. Thus, the ConvLSTM model is aware of both NDVI and environmental conditions at each time step.

We first present the result of including all 13 environmental variables all together into the ConvLSTM model. Surprisingly, the model does not show a clear improvement in forecasting accuracy. In fact, the R^2 of the model slightly decreases (a decrease of 0.003 in R^2) after all 13 environmental variables are included, compared with using NDVI time series only. With curiosity, we further conduct an ablation study by removing one environmental variable at a time and measuring the forecasting accuracy change after the environmental variable is removed. Fig. 7 shows the changes of the model performance in RMSE and R^2 when each of the environmental variables is removed. Note that we control the experiments by ensuring that all the other environmental variables are kept the same.

In Fig. 7, a positive RMSE change indicates that the forecasting accuracy of the model decreases when the corresponding environmental

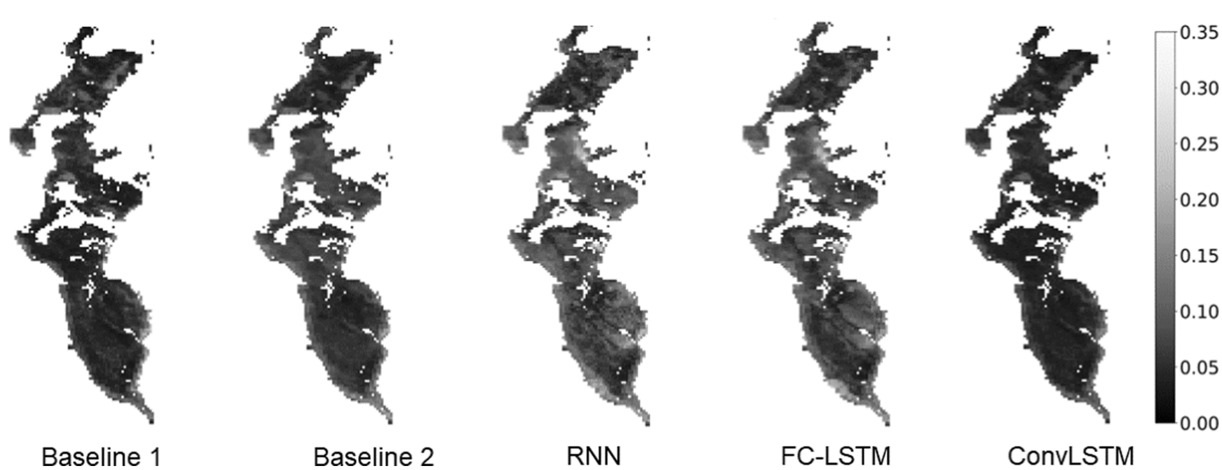

Fig. 5. Spatial RMSE of the five approaches for vegetation dynamics forecasting.

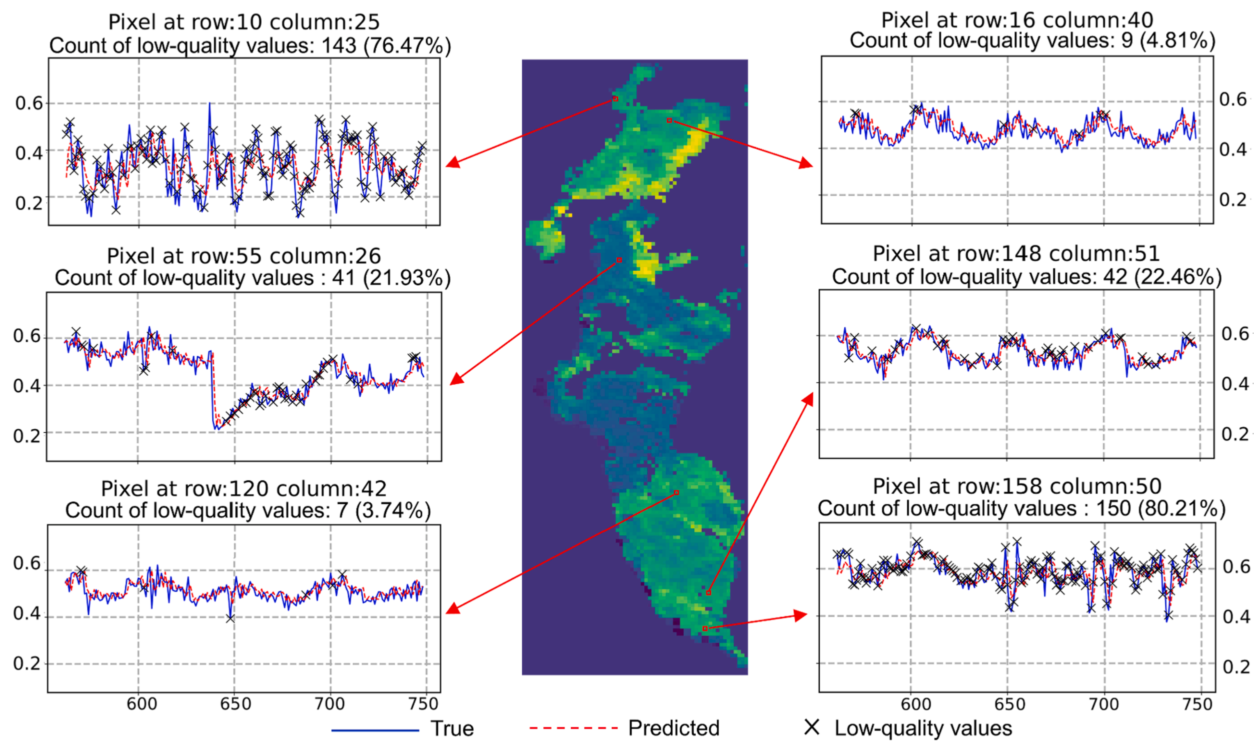

Fig. 6. Predicted NDVI of ConvLSTM and observed NDVI in six randomly selected pixels and their locations in the study area.

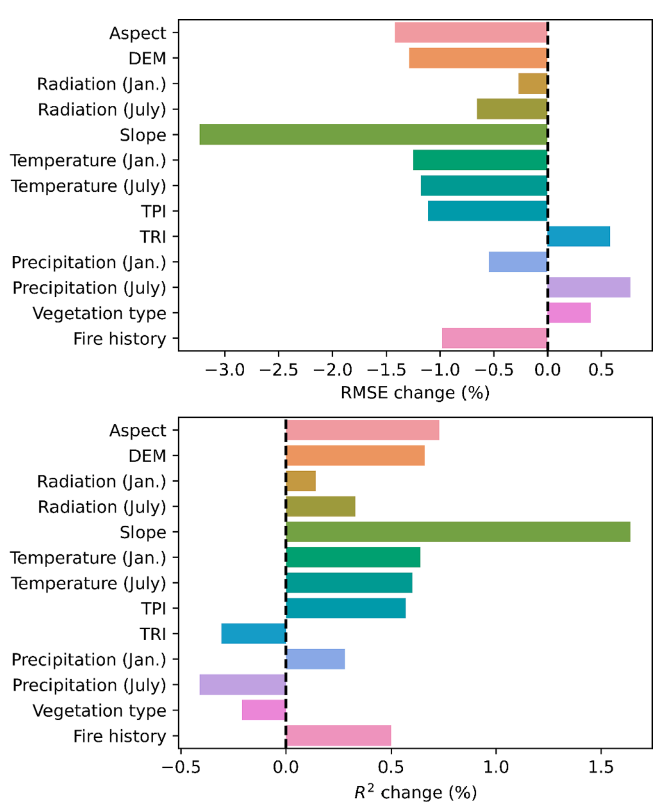

Fig. 7. RMSE and R^2 changes of the ConvLSTM model when different environmental variables are removed, compared with using all environmental variables.

variable is removed, suggesting that this environmental variable is important for the model to make correct predictions. By contrast, a negative RMSE change indicates that the forecasting accuracy of the model, in fact, increases when the corresponding environmental variable is removed, suggesting that the environmental variable is probably less useful for the model to predict vegetation state. The R^2 changes should be interpreted in a reverse manner, i.e., negative changes indicate that the corresponding environmental variables are useful, while positive changes indicate that the corresponding environmental variables are probably less useful. As can be seen, three environmental variables, namely topographic roughness index, mean precipitation in July (winter), and vegetation type, positively contribute to the forecasting of vegetation dynamics of the fynbos open ecosystem. The other ten environmental variables have small negative effects on the forecasting accuracy.

Since only three environmental variables have shown the effect of improving forecasting accuracy, could we obtain a better model by including these three environmental variables only? To answer this question, we train another model using only these three environmental variables and NDVI images. However, keeping only these three environmental variables does not achieve the best performance either, with a slight decrease in R^2 compared with using all 13 environmental variables. This result suggests that there may exist certain interactions among the environmental variables inside the model, and we cannot simply remove all the other nine variables. Realizing this complexity, we use a greedy approach to identify a better forecasting setting, in which we gradually remove the environmental variables starting from those that have shown the largest negative effects on forecasting accuracy. If removing one environmental variable leads to improved forecasting accuracy, we remove that variable and continue to test the next variable; otherwise, we put that variable back and move to testing the next variable. Through this greedy approach, we find that removing the variable slope achieves the highest forecasting accuracy, with an RMSE of 0.053 and an R squared of 0.807.

4.3. Forecasting over long time periods

In the previous two sets of experiments, we have focused on understanding the forecasting accuracy of different approaches based on a next-step prediction task. Here, we present the results of multi-step prediction, and we focus on the best approach identified so far, i.e., the ConvLSTM model integrated with the environmental variables excluding slope. The forecasting steps tested include: 1 step ahead (8 days), 4 steps ahead (about 1 month), 11 steps ahead (about 3 months), 15 steps ahead (120 days), 23 steps ahead (about half a year), and 46 steps ahead (about 1 year). We also compare the ConvLSTM with environmental variables with the same four baselines as used in the first set of experiments, which are: Baseline 1 (same as the last step), Baseline 2 (average of the previous 15 steps), RNN, and FC-LSTM. Note that the two deep learning models are trained using NDVI images only and serve as two additional baselines for the ConvLSTM with environmental variables. The results are shown in Fig. 8.

As can be seen, all approaches have decreased forecasting accuracies when they are asked to predict longer time steps ahead. While Baseline 1 is a strong baseline for predicting 1 step ahead, its performance decreases substantially over longer time steps and becomes the worst approach for predicting 46 steps ahead. The performances of RNN and LSTM change over different time steps, and are generally worse than the two naive forecasting approaches for predicting shorter time steps and better for predicting longer times steps than Baseline 1. The ConvLSTM model with environmental variables has performed the best among all the five approaches in all the tested time steps, although its performance is only slightly better than Baseline 2 for predicting 46 steps ahead.

5. Discussion

5.1. Forecasting vegetation dynamics in an open ecosystem using different approaches

The results of our experiments suggest that the ConvLSTM model can forecast vegetation dynamics in the studied open ecosystem with fairly high accuracy. While ConvLSTM has been used previously for NDVI forecasting such as in agricultural land (Ahmad et al., 2020), it is unknown how accurately it can forecast vegetation dynamics in open ecosystems where the vegetation states can change dramatically. Our study therefore provides one piece of evidence by demonstrating that the ConvLSTM model is overall effective in the studied open ecosystem in South Africa which is a global biodiversity hotspot. Our study also reveals the roles of 13 different environmental variables in enhancing vegetation dynamics forecasting, and the performance of different models in predicting vegetation states in longer time periods ahead. In addition, the ConvLSTM model has shown a higher forecasting accuracy than that of the RNN and FC-LSTM models. While the other two models take into account the past vegetation states for making predictions, they process 2D NDVI images in the form of 1D vectors and do not make use of the important spatial information related to pixel locations. By contrast, the ConvLSTM model preserves such spatial information by directly processing 2D NDVI images and leverages convolutional operations to capture the likely similarity of vegetation states of nearby pixels. Compared with Baseline 2 which has achieved the second best performance in most experiments, ConvLSTM shows an improvement of about 0.01 in RMSE. While this is only a small improvement in terms of its absolute value, NDVI values on land typically range between [0, 1], and an RMSE improvement of 0.01 can still be useful. From a percentage perspective, it is about 15 % improvement compared with the RMSE of Baseline 2. In addition, the 0.01 RMSE improvement is an average value for an individual pixel. When applying this improvement to the many pixels in the entire study area, we may still gain a large improvement collectively in predicting the vegetation states across many different locations.

5.2. Pros and cons of using a deep learning based approach for conservation management

The results of our experiments suggest that the approach of integrating ConvLSTM with environmental variables has potential to be used as a forecasting tool to support conservation management. Then, what would be some of its pros and cons? We identify three pros for such an approach. First, as shown in the experiments, the ConvLSTM model integrated with environmental variables can achieve a relatively high prediction accuracy, such as an R^2 of 0.680 for predicting three months ahead. A higher forecasting accuracy can help conservation managers better prepare for the future and identify the abnormal changes which may otherwise be missed. Second, such a vegetation dynamics forecasting model can be run frequently (e.g., once every 8 days) and therefore produces predictions with a high temporal resolution which is also critical for supporting conservation management decisions (MacFadyen et al., 2022). Third, the vegetation states forecasted by such an approach are spatially explicit, i.e., in the form of 2D images. This spatially explicit forecasting can help conservation managers identify the locations of potential abnormal vegetation changes, so that field trips could be arranged to investigate the underlying issues.

We expect two likely cons of using this deep learning based approach. First, the explainability of such an AI based model is limited. While we have conducted an ablation study to understand the ability of different environmental variables to enhance forecasting, a full understanding of the ways that these variables interact inside the model is yet to be achieved. Explainable AI is currently an active research topic (Samek et al., 2019), and with research effort on this topic, we hope that we can improve the explainability of this and other deep learning models for conservation management in the near future. Second, the prediction accuracy of the current approach for very long time periods is still limited. The ConvLSTM model integrated with environmental

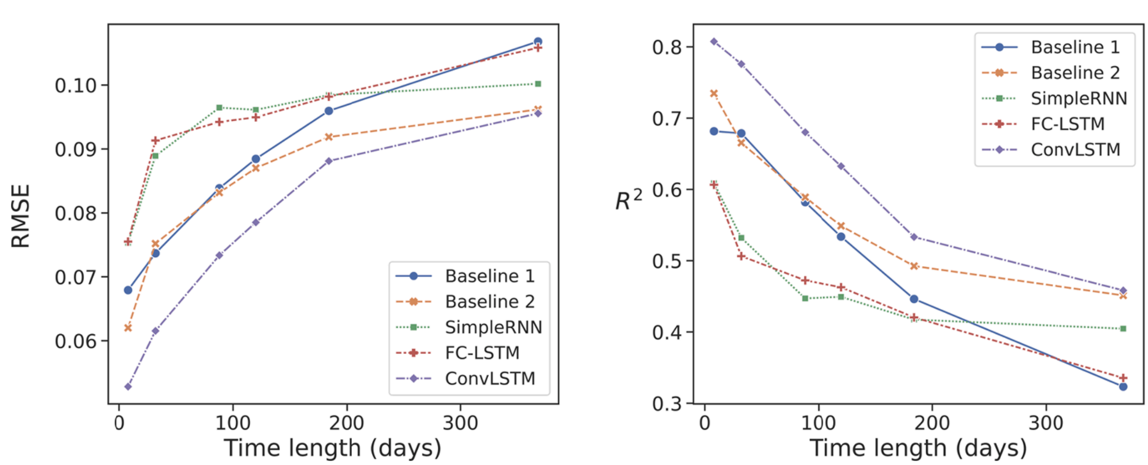

Fig. 8. Forecasting accuracy of different approaches in predicting longer time periods ahead.

variables achieves an R^2 of 0.458 when predicting one year ahead, which has a similar performance as using the average of the previous 15 steps. The current approach may still need to be improved for supporting conservation management decisions that require forecasting into a long time ahead.

5.3. Limitations and future work

This study is not without limitations. First, while we have examined the forecasting accuracy of ConvLSTM and other two deep learning approaches in one open ecosystem, more research is needed to understand the performance of these models in other open ecosystems with similar complex vegetation dynamics, such as the Californian Chaparral, Australian Kwongan, and parts of the Mediterranean Basin. Studies in these other open ecosystems, together with this current study on the fynbos shrubland, can help form evidence on the ability of ConvLSTM and other deep learning models to forecast vegetation dynamics in open ecosystems. Second, due to data availability constraints, we have examined only 13 environmental variables in this study. Other variables, such as soil conditions, may also help the model make better predictions. In addition, we have used only static environmental variables, and dynamic variables in time series measurements could help inform the model about short-term changes in the environment. While those short-term changes could also bring in noise less relevant to vegetation growth, it would still be interesting to investigate and understand whether the additional information brought by the dynamic variables could outweigh their noise and ultimately improve forecasting accuracy.

6. Conclusions

Accurately forecasting vegetation dynamics in open ecosystems is critical for managing their biodiversity, fire, water, and carbon. In this work, we investigated the use of deep learning based approaches for forecasting vegetation dynamics in an open ecosystem, the fynbos shrubland of the Cape Peninsula located in the Cape Floristic Region of South Africa. We also examined the ability of a number of environmental variables in enhancing forecasting, including precipitation, fire history, and vegetation types. We found that the ConvLSTM model can forecast vegetation state more accurately than RNN and FC-LSTM as well as two naive forecasting baselines based on NDVI time-series data. Environmental variables showed different ability to further improve the accuracy of vegetation forecasting. By integrating the ConvLSTM model and selected environmental variables, we obtained the best forecasting approach that can achieve an R^2 of 0.807 for predicting one step ahead (about 8 days) and an R^2 of 0.458 for predicting 46 steps ahead (about one year). Finally, we discussed the pros and cons of using such a deep learning based approach for supporting conservation management.

CRediT authorship contribution statement

Yue Ma: Methodology, Formal analysis, Data curation, Visualization, Writing – original draft. **Yingjie Hu:** Conceptualization, Methodology, Formal analysis, Validation, Supervision, Project administration, Funding acquisition, Writing – original draft, Writing – review & editing. **Glenn R. Moncrieff:** Conceptualization, Resources, Validation, Writing – review & editing. **Jasper A. Slingsby:** Conceptualization, Resources, Validation, Writing – review & editing. **Adam M. Wilson:** Conceptualization, Resources, Validation, Supervision, Project administration, Funding acquisition, Writing – review & editing. **Brian Maitner:** Resources, Validation, Writing – review & editing. **Ryan Zhenqi Zhou:** Data curation, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the US National Aeronautics and Space Administration under grant 80NSSC21K1183. It was also supported by the Microsoft AI for Earth program.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2022.103060.

References

Ahmad, R., Yang, B., Ettlin, G., Berger, A., Rodríguez-Bocca, P., 2020. A machine-learning based ConvLSTM architecture for NDVI forecasting. Int. Trans. Oper. Res. https://doi.org/10.1111/itor.12887.
Bengio, Y., Goodfellow, I., Courville, A., 2017. Deep learning. MIT press Cambridge, MA, USA.
Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw. 5, 157–166.
Bond, W.J., 2019. Open Ecosystems: ecology and evolution beyond the forest edge. Oxford University Press.
Boulila, W., Ghandorh, H., Khan, M.A., Ahmed, F., Ahmad, J., 2021. A novel CNN-LSTM-based approach to predict urban expansion. Ecol. Inform. 64, 101325 https://doi.org/10.1016/j.ecoinf.2021.101325.
Browning, D.M., Maynard, J.J., Karl, J.W., Peters, D.C., 2017. Breaks in MODIS time series portend vegetation change: verification using long-term data in an arid grassland ecosystem. Ecol. Appl. 27, 1677–1693. https://doi.org/10.1002/eap.1561.
Bucklin, D.N., Basille, M., Benscoter, A.M., Brandt, L.A., Mazzotti, F.J., Romanach, S.S., Speroterra, C., Watling, J.I., 2015. Comparing species distribution models constructed with different subsets of environmental predictors. Divers. Distrib. 21, 23–35. https://doi.org/10.1111/ddi.12247.
Burke, M., Driscoll, A., Lobell, D.B., Ermon, S., 2021. Using satellite imagery to understand and promote sustainable development. Science 371. https://doi.org/10.1126/science.abe8628.
Didan, Kamel, 2015a. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. https://doi.org/10.5067/MODIS/MOD13Q1.006.
Didan, Kamel, 2015b. MYD13Q1 MODIS/Aqua Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. https://doi.org/10.5067/MODIS/MYD13Q1.006.
Duncanson, L., Armston, J., Disney, M., Avitabile, V., Barbier, N., Calders, K., Carter, S., Chave, J., Herold, M., Crowther, T.W., Falkowski, M., Kellner, J.R., Labrière, N., Lucas, R., MacBean, N., McRoberts, R.E., Meyer, V., Næsset, E., Nickeson, J.E., Paul, K.I., Phillips, O.L., Réjou-Méchain, M., Román, M., Roxburgh, S., Saatchi, S., Schepaschenko, D., Scipal, K., Siqueira, P.R., Whitehurst, A., Williams, M., 2019. The Importance of Consistent Global Forest Aboveground Biomass Product Validation. Surv. Geophys. 40, 979–999. https://doi.org/10.1007/s10712-019-09538-8.
Ferchichi, A., Abbes, A.B., Barra, V., Farah, I.R., 2022. Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review. Ecol. Inform. 101552 https://doi.org/10.1016/j.ecoinf.2022.101552.
Forsyth, G.G., Van Wilgen, B.W., 2008. The recent fire history of the Table Mountain National Park and implications for fire management. Koedoe Afr. Prot. Area Conserv. Sci. 50, 3–9.
Géron, A., 2019. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media.
Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.
Huang, X., Zhu, D., Zhang, F., Liu, T., Li, X., Zou, L., 2021. Sensing Population Distribution from Satellite Imagery Via Deep Learning: Model Selection, Neighboring Effects, and Systematic Biases. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 5137–5151. https://doi.org/10.1109/JSTARS.2021.3076630.
Irvin, J., Sheng, H., Ramachandran, N., Johnson-Yu, S., Zhou, S., Story, K., Rustowicz, R., Elsworth, C., Austin, K., Ng, A.Y., 2020. ForestNet: Classifying Drivers of Deforestation in Indonesia using Deep Learning on Satellite Imagery, in: NeurIPS 2020.
Jonsson, P., Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans. Geosci. Remote Sens. 40, 1824–1832. https://doi.org/10.1109/TGRS.2002.802519.
Karger, D.N., Zimmermann, N.E., 2019. Climatologies at high resolution for the earth land surface areas CHELSA V1. 2: Technical specification. Swiss Fed. Res. Inst. WSL Switz.

MacFadyen, S., Allsopp, N., Altwegg, R., Archibald, S., Botha, J., Bradshaw, K., Carruthers, J., De Klerk, H., de Vos, A., Distiller, G., 2022. Drowning in data, thirsty for information and starved for understanding: A biodiversity information hub for cooperative environmental monitoring in South Africa. Biol. Conserv. 274, 109736 https://doi.org/10.1016/j.biocon.2022.109736.

McNicol, I.M., Ryan, C.M., Mitchard, E.T.A., 2018. Carbon losses from deforestation and widespread degradation offset by extensive growth in African woodlands. Nat. Commun. 9, 3045. https://doi.org/10.1038/s41467-018-05386-z.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858. https://doi.org/10.1038/35002501.

Ntshanga, N.K., Procheş, S., Slingsby, J.A., 2021. Assessing the threat of landscape transformation and habitat fragmentation in a global biodiversity hotspot. Austral Ecol. https://doi.org/10.1111/aec.13037.

Prentice, I.C., Farquhar, G.D., Fasham, M.J.R., Goulden, M.L., Heimann, M., Jaramillo, V. J., Kheshgi, H.S., Quéré, C.L., Scholes, R.J., Wallace, D.W.R., Archer, D., Ashmore, M.R., Aumont, O., Baker, D., Battle, M., Bender, M., Bopp, L.P., Bousquet, P., Caldeira, K., Ciais, P., Cox, P.M., Cramer, W., Dentener, F., Entling, I.G., Field, C.B., Friedlingstein, P., Holland, E.A., Houghton, R.A., House, J.I., Ishida, A., Jain, A.K., Janssens, I.A., Joos, F., Kaminski, T., Keeling, C.D., Keeling, R.F., Kicklighter, D.W., Kohfeld, K.E., Knorr, W., Law, R., Lenton, T., Lindsay, K., Maier-Reimer, E., Manning, A.C., Matear, R.J., McGuire, A.D., Melillo, J.M., Meyer, R., Mund, M., Orr, J.C., Piper, S., Platner, K., Rayner, P.J., Sitch, S., Slater, R., Taguchi, S., Tans, P.P., Tian, H.Q., Weirig, M.F., Whorf, T., Yool, A., 2001. The carbon cycle and atmospheric CO2, Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Reddy, D.S., Prasad, P., 2018. Prediction of vegetation dynamics using NDVI time series data and LSTM. Model. Earth Syst. Environ. 4, 409–419. https://doi.org/10.1007/s40808-018-0431-3.

Rhif, M., Abbes, A.B., Martínez, B., Farah, I.R., 2020. Deep learning models performance for NDVI time series prediction: a case study on north west Tunisia. Presented at the 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS), IEEE, pp. 9–12. https://doi.org/10.1109/M2GARSS47143.2020.9105149.

Rouget, M., Richardson, D.M., Cowling, R.M., Lloyd, J.W., Lombard, A.T., 2003. Current patterns of habitat transformation and future threats to biodiversity in terrestrial ecosystems of the Cape Floristic Region, South Africa. Biol. Conserv. Conservation Planning in the Cape Floristic Region 112, 63–85. https://doi.org/10.1016/S0006-3207(02)00395-6.

Rußwurm, M., Körner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. ISPRS Int. J. Geo-Inf. 7, 129. https://doi.org/10.3390/ijgi7040129.

Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R., 2019. Explainable AI: interpreting, explaining and visualizing deep learning. Springer Nature.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Adv. Neural Inform. Process. Syst. 802–810.

Skowno, A.L., Jewitt, D., Slingsby, J.A., 2021. Rates and patterns of habitat loss across South Africa's vegetation biomes. South Afr. J. Sci. 117, 1–5. https://doi.org/10.17159/sajs.2021/8182.

Sleeter, B.M., Liu, J., Daniel, C., Rayfield, B., Sherba, J., Hawbaker, T.J., Zhu, Z., Selmants, P.C., Loveland, T.R., 2018. Effects of contemporary land-use and land-cover change on the carbon balance of terrestrial ecosystems in the United States. Environ. Res. Lett. 13, 045006 https://doi.org/10.1088/1748-9326/aab540.

Slingsby, J.A., Moncrieff, G.R., Wilson, A.M., 2020. Near-real time forecasting and change detection for an open ecosystem with complex natural dynamics. ISPRS J. Photogramm. Remote Sens. 166, 15–25. https://doi.org/10.1016/j.isprsjprs.2020.05.017.

Stepchenko, A., 2015. Normalized difference vegetation index forecasting using a regularized layer recurrent neural network. Presented at the Proc. of the 3rd Virtual Multidisciplinary Conf. QUAESTI, December 7-11, 2015, Zilina, Slovakia, pp. 261–266. https://doi.org/10.18638/quaesti.2015.3.1.192.

Watts, L.M., Laffan, S.W., 2014. Effectiveness of the BFAST algorithm for detecting vegetation response patterns in a semi-arid region. Remote Sens. Environ. 154, 234–245. https://doi.org/10.1016/j.rse.2014.08.023.

Wilson, A.M., Latimer, A.M., Silander, J.A., 2015. Climatic controls on ecosystem resilience: Postfire regeneration in the Cape Floristic Region of South Africa. Proc. Natl. Acad. Sci. 112, 9058–9063. https://doi.org/10.1073/pnas.1416710112.

Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., Xu, L., Chen, Z., Gong, J., 2019. A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. Environ. Model. Softw. 120, 104502 https://doi.org/10.1016/j.envsoft.2019.104502.

Yang, M., Kong, B., Dang, R., Yan, X., 2022. Classifying urban functional regions by integrating buildings and points-of-interest using a stacking ensemble method. Int. J. Appl. Earth Obs. Geoinformation 108, 102753. https://doi.org/10.1016/j.jag.2022.102753.