Learning to forecast vegetation greenness at fine resolution over Africa with ConvLSTMs

Claire Robin[1], **Christian Requena-Mesa**[1], **Vitus Benson**[1], **Lazaro Alonso**[1], **Jeran Poehls**[1],
Nuno Carvalhais[1,2], and **Markus Reichstein**[1,2]

[1]Biogeochemical Integration, Max-Planck-Institute for Biogeochemistry, Jena, Germany
[2]ELLIS Unit Jena, Michael-Stifel-Center Jena for Data-driven and Simulation Science, Jena, Germany
Corresponding author: *{crobin, crequ}@bgc-jena.mpg.de*

Abstract

Forecasting the state of vegetation in response to climate and weather events is a major challenge. Its implementation will prove crucial in predicting crop yield, forest damage, or more generally the impact on ecosystems services relevant for socio-economic functioning, which if absent can lead to humanitarian disasters. Vegetation status depends on weather and environmental conditions that modulate complex ecological processes taking place at several timescales. Interactions between vegetation and different environmental drivers express responses at instantaneous but also time-lagged effects, often showing an emerging spatial context at landscape and regional scales. We formulate the land surface forecasting task as a strongly guided video prediction task where the objective is to forecast the vegetation developing at very fine resolution using topography and weather variables to guide the prediction. We use a Convolutional LSTM (ConvLSTM) architecture to address this task and predict changes in the vegetation state in Africa using Sentinel-2 satellite NDVI, having ERA5 weather reanalysis, SMAP satellite measurements, and topography (DEM of SRTMv4.1) as variables to guide the prediction. Ours results highlight how ConvLSTM models can not only forecast the seasonal evolution of NDVI at high resolution, but also the differential impacts of weather anomalies over the baselines. The model is able to predict different vegetation types, even those with very high NDVI variability during target length, which is promising to support anticipatory actions in the context of drought-related disasters.

1 Introduction

Climate change is leading to an increase in extreme weather events, affecting both ecosystem and human livelihoods. Africa is one of the most vulnerable continents to climate change with droughts in the region expected to increase in severity according to the IPCC [1]. Given current projections of population growth and impacts from climate change, many more people will be affected by extreme drought events in the future [2]. Any insight into predicting their occurrences can help prepare short-term solutions to alleviate potential impacts on local and regional communities [3].

The local scale response to extreme events is often not homogeneous [4]. Abiotic (*e.g.* soil type, topography, water bodies) and biotic (*e.g.* vegetation type, plant rooting depth) factors and the interactions between them affect how vegetation responds to atmospheric extreme events [5, 6].In addition, weather impacts with temporal lagged responses can have a significant influence on ecosystem responses to weather variability [7–9]. This so-called 'memory effect', arises from the complex
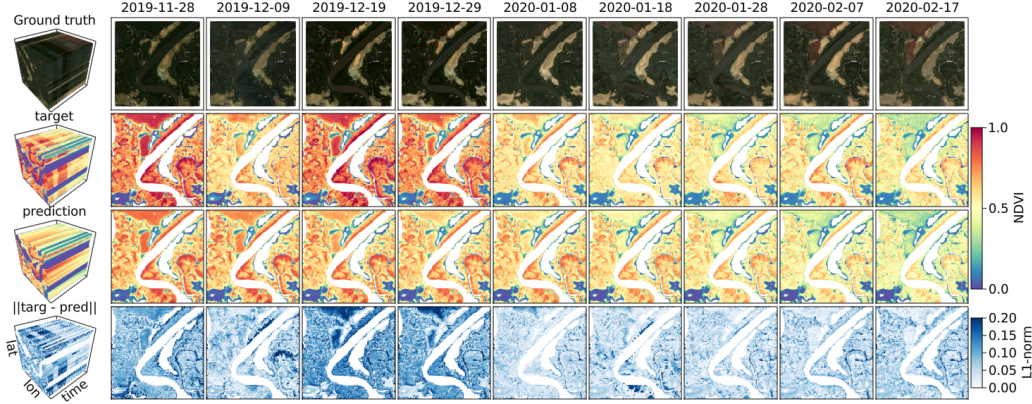
Figure 1: **Model prediction frame by frame**. The frames are temporally interpolated in a spatio-temporal cube (**Left side**). (**Top row**) RGB satellite imagery. (**Second row**) NDVI target. (**Third row**) ConvLSTM predictions. (**Bottom row**) L1-norm of the difference between the target and the prediction. **Location:** 13°16'34.8"N 16°07'10.4"W, The Gambia.

processes involved in vegetation dynamics leading to non-linear interaction on several time scales [10, 11] (*e.g.* the time-lagged effect of precipitation on vegetation due to the water soil recharge [12], legacy effect of earlier drought [6, 13] or early starting season [14]).

Predicting the evolution and impacts of vegetation at the very local level is therefore a challenge to support anticipatory action before a disaster In this paper, we aim to create a model capable of learning the relationship between vegetation states, local factors and weather conditions, at a very fine resolution in order to characterise and forecast the impact of weather extremes from an ecosystem perspective. We use the Normalized Difference Vegetation Index (NDVI) [15] as a proxy for vegetation health monitoring. Our work can be easily extended to predict other vegetation indices, depending on the ecological process of interest.

Contributions Our main contributions are summarized as follows:

- A proof-of-concept of forecasting vegetation greenness in Africa at high spatial resolution.
- A spatio-temporal analysis of the prediction, both of the dataset and of individual samples.

Related work At low spatial resolutions, machine learning models have been proposed for both vegetation modeling [11, 16–22] and crop forecasting [23–25]. Requena-Mesa et al. [26] introduced Earth surface forecasting as modeling the future spectral reflectance of the Earth surface and provided the first dataset in this respect, EarthNet2021. Concurrent to our work, Diaconu et al. [27] and Kladny et al. [28] have built ConvLSTM variants for EarthNet2021. Both study the influence of weather on the predictions by providing artificial meteorological inputs.

2 Method

This paper follows the works of Requena-Mesa et al. [26], who introduced EarthNet2021, focusing instead on Africa and a specific deep learning model: the Convolutional LSTM (ConvLSTM).

Dataset Similar to EarthNet2021 [26], our dataset contains Sentinel 2 satellite imagery (bands red, blue, green and near-infrared), weather variables from ERA5 reanalysis (evaporation, surface pressure, surface net solar radiation, 2m temperature, total precipitation, potential evaporation) [29] and SMAP (latent and sensible heat flux, rootzone and surface soil moisture, surface pressure) [30] and topography from SRTMv4.1 DEM [31]. The data is collected at high spatial resolution for over 40000 locations in Africa, each resulting sample we call a *minicube*. For more information, see supplementary materials 5.1 and 5.2.

Task We define an Earth surface forecasting task [26] as a strongly guided video prediction task. The objective is to forecast a length-k sequence of future NDVI satellite imagery $(\widehat{S_{n+1}}, ..., \widehat{S_{n+k}})$

Model	$RMSE\downarrow$	$NSE\uparrow$	α	β	r
Constant baseline	0.3365	-1.3922	0.0	0.1559	0.0
Previous season baseline	0.2937	-1.0561	1.0169	-0.0084	0.5504
ConvLSTM without weather	0.2331	-0.3356	0.6512	0.1699	0.7348
ConvLSTM (ours)	**0.1882**	**0.0270**	0.7570	0.0628	0.8024

Table 1: **Test set model performance.** All values are computed on minicubes with less than 75% masked values, shown are the median values over all pixels. We additionally exclude the 5% worst pixels as outliers because the metrics are unbounded.

based on previous length-n sequence of satellite imagery $(S_1, ..., S_n)$, topography T and guiding environmental variables during both the context and prediction periods $(E_1, ..., E_n, ..., E_{n+k})$. Formally, the task is to learn a function f such that:

$$(\widehat{S_{n+1}}, ..., \widehat{S_{n+k}}) = f(S_1, .., S_n, E_1, ..., E_n, ..., E_{n+k}, T)$$

In this paper, deep learning models use a context period of one year and then forecast the next three months of NDVI at a 10-daily timestepping.

Model We use a ConvLSTM [32], that is an LSTM [33] using convolution to operate in the spatial domain. We stack two ConvLSTM units (each designed as in Patraucean et al. [34]) into an encoder and another two into a decoder (for forecasting). The encoder is used during the context period, it gets as inputs the concatenated satellite imagery, environmental variables and topography. The decoder uses only the environmental variables and the topography as inputs, but gets the hidden states from the encoder to propagate information from the context period. The model is visualized in supplementary material fig. 3. The training procedure is detailed in supplementary material 5.4.

Baselines and ablation We compare our model against a *constant baseline* (last valid pixel from context period) and a *previous season baseline* (observations from previous year). Furthermore we ablate our model by removing the environmental variables (*ConvLSTM /wo weather*), i.e. doing an ablation without weather to see how much the model learns from just the first order process underlying vegetation dynamics: the seasonal cycle.

Evaluation We evaluate the models using two mean squared error (MSE) derived scores: the root mean squared error ($RMSE$) and the Nash-Sutcliffe model efficiency (NSE) [35]. The latter rescales the MSE with the variance of the observations σ_0^2, it is defined as:

$$NSE = 1 - MSE/\sigma_0^2, \tag{1}$$

Both the MSE and the NSE can be decomposed into parts:

$$MSE = phase\_err + var\_err + bias\_sq \tag{2}$$

$$NSE = 2\alpha r - \alpha^2 - \beta^2, \tag{3}$$

Here we introduced the bias $bias\_sq = (\mu_0 - \hat{\mu})^2$, the variance error $var\_err = (\sigma_0 - \hat{\sigma})^2$, the phase error $phase\_err = (1 - r) * 2 * \sigma_0 * \hat{\sigma}$, the correlation r, the rescaled bias $\beta = (\hat{\mu} - \mu_0)/\sigma_0$ and a measure of relative variability $\alpha = \hat{\sigma}/\sigma_0$. μ represents means, σ^2 represents variances, $\cdot_0$ refers to observations and $\hat{\cdot}$ stands for predicted quantities. For the NSE decomposition, the components' ideal values are $r = 1$, $\alpha = 1$ and $\beta = 0$ [36].

3 Results

Central estimates of our models' performance are reported in tab. 1. Our ConvLSTM model beats the baselines by a large margin in both $RMSE$ and NSE. The previous season baseline has the same NDVI distribution over the whole dataset as the target (since it is the same season one year ago), so we have $\sum \hat{\sigma} \simeq \sum \sigma_0$ and $\sum \hat{\mu} \simeq \sum \mu_0$. Negative errors compensate for positive ones, which explains a median of α and β very close to their ideal value (but not for r). An ablation of it not using weather covariates performs worse than our model using them. This supports the intuition that weather should
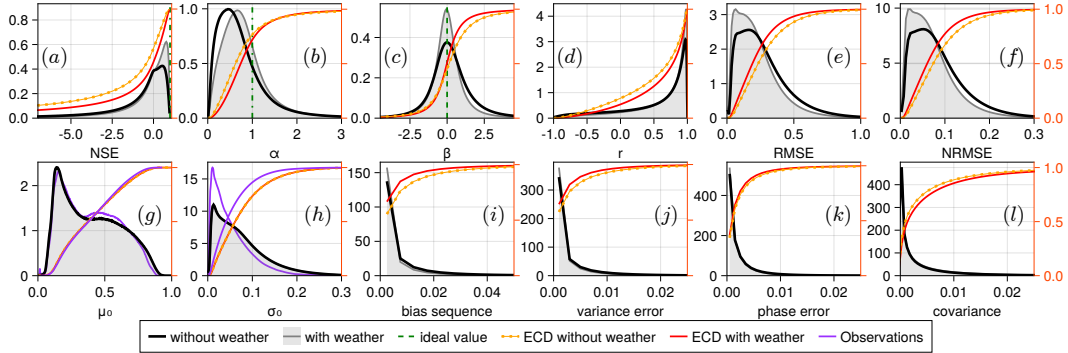
Figure 2: **Probability density plots of pixelwise test set performance.** ECD is the empirical cumulative distribution function. **(First row)** NSE decomposition. Performance distributions are closer to ideal values for NSE α, β and r with weather than without it. The same pattern is observed for $RMSE$, normalized $RMSE$ ($NRMSE$) and the decomposition of MSE into bias sequence, variance error and phase error.

be driving vegetation dynamics. Supplementary figs. 5 and 6 visualize the performance against the baselines clustered by land cover type. Our model is stronger than the baselines across all classes.

Fig. 2 shows the decomposition for both NSE and $RMSE$ on the test set. We find 50% of the pixels in our model reach $NSE \geq 0$, while this is only 35% in the ablation without weather. This difference mostly stems from a reduction in bias (fig. 2c) and a reduction in relative variability (fig. 2b).

While already good, our ConvLSTM does have limitations. Fig. 1 visualizes a prediction on one example minicube. One visible limitation is due the dataset. On 2019-12-09, the gapfilled RGB satellite image has noise due to undetected clouds. Clouds lower the target NDVI and lead to an apparent overestimation of the model on that day. Additionally, since such artifacts are common in the training dataset, the model learns a biased estimate, slightly underestimating NDVI to account for occasional clouds. Such underestimation is visible on 2019-12-19 in fig. 1. We present the pixelwise decomposition of the NSE for this pixel in the supplementary material, fig. 4.

4 Conclusion

This work presents a ConvLSTM deep learning model to predict vegetation greenness in Africa at high spatial resolution from coarse-scale weather. Our model obtains a higher Nash-Sutcliffe model efficiency than two baselines. In an ablation we train a model variant without weather covariates with lower performance. Hence, our final model is able to extract information from meteorology. We trained our model on a diverse dataset, thus it is applicable to a wide variety of landcover classes and climate zones. Decomposing the NSE into parts, we find a vulnerability of our model to data noise: some clouds were not flagged during pre-processing, thereby biasing our model towards lower greenness values. Nevertheless, our model is a proof-of-concept of high resolution vegetation modeling in Africa. Future work should now focus on improving predictions by better understanding of spatial context, on the applicability of our model during extreme events and on building a bridge to the anticipatory action community.

References

[1] Hans O Pörtner, Debra C Roberts, Helen Adams, Carolina Adler, Paulina Aldunce, Elham Ali, Rawshan Ara Begum, Richard Betts, Rachel Bezner Kerr, Robbert Biesbroek, et al. Climate change 2022: impacts, adaptation and vulnerability. 2022.

[2] N Yaghmaei and R Below. Disasters in africa: 20 year review (2000-2019). *CRED Crunch*, (56), 2019.

[3] Zia Mehrabi, Simon Donner, Patricia Rios, Debarati Guha-Sapir, Pedram Rowhani, Milind Kandlikar, and Navin Ramankutty. Can we sustain success in reducing deaths to extreme weather in a hotter world? *World Development Perspectives*, 14:100107, 2019.

[4] Felix N Kogan. Remote sensing of weather impacts on vegetation in non-homogeneous areas. *International Journal of remote sensing*, 11(8):1405–1419, 1990.

[5] Jon D Pelletier, A Brad Murray, Jennifer L Pierce, Paul R Bierman, David D Breshears, Benjamin T Crosby, Michael Ellis, Efi Foufoula-Georgiou, Arjun M Heimsath, Chris Houser, et al. Forecasting the response of earth's surface to future climatic and land use changes: A review of methods and research needs. *Earth's Future*, 3(7):220–251, 2015.

[6] Joan Sturm, Maria J Santos, Bernhard Schmid, and Alexander Damm. Satellite data reveal differential responses of swiss forests to unprecedented 2018 drought. *Global Change Biology*, 28(9):2956–2978, 2022.

[7] Kiona Ogle, Jarrett J Barber, Greg A Barron-Gafford, Lisa Patrick Bentley, Jessica M Young, Travis E Huxman, Michael E Loik, and David T Tissue. Quantifying ecological memory in plant and ecosystem processes. *Ecology letters*, 18(3):221–235, 2015.

[8] Christina Papagiannopoulou, Diego G Miralles, Stijn Decubber, Matthias Demuzere, Niko EC Verhoest, Wouter A Dorigo, and Willem Waegeman. A non-linear granger-causality framework to investigate climate–vegetation dynamics. *Geoscientific Model Development*, 10(5):1945–1960, 2017.

[9] Jill F Johnstone, Craig D Allen, Jerry F Franklin, Lee E Frelich, Brian J Harvey, Philip E Higuera, Michelle C Mack, Ross K Meentemeyer, Margaret R Metz, George LW Perry, et al. Changing disturbance regimes, ecological memory, and forest resilience. *Frontiers in Ecology and the Environment*, 14(7):369–378, 2016.

[10] Wanda De Keersmaecker, Stef Lhermitte, Laurent Tits, Olivier Honnay, Ben Somers, and Pol Coppin. A model quantifying global vegetation resistance and resilience to short-term climate anomalies and their relationship with vegetation cover. *Global Ecology and Biogeography*, 24 (5):539–548, 2015.

[11] Basil Kraft, Martin Jung, Marco Körner, Christian Requena Mesa, José Cortés, and Markus Reichstein. Identifying dynamic memory effects on vegetation state using recurrent neural networks. *Frontiers in big Data*, page 31, 2019.

[12] Alistair WR Seddon, Marc Macias-Fauria, Peter R Long, David Benz, and Kathy J Willis. Sensitivity of global terrestrial ecosystems to climate variability. *Nature*, 531(7593):229–232, 2016.

[13] Ana Bastos, René Orth, Markus Reichstein, Philippe Ciais, Nicolas Viovy, Sönke Zaehle, Peter Anthoni, Almut Arneth, Pierre Gentine, Emilie Joetzjer, et al. Vulnerability of european ecosystems to two compound dry and hot summers in 2018 and 2019. *Earth system dynamics*, 12(4):1015–1035, 2021.

[14] Sebastian Sippel, Jakob Zscheischler, and Markus Reichstein. Ecosystem impacts of climate extremes crucially depend on the timing. *Proceedings of the National Academy of Sciences*, 113(21):5768–5770, 2016.

[15] Compton J Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2):127–150, 1979.

[16] Jonas Koehler and Claudia Kuenzer. Forecasting spatio-temporal dynamics on the land surface using earth observation data—a review. *Remote Sensing*, 12(21):3513, 2020.

[17] Andrea Gobbi, Marco Cristoforetti, Giuseppe Jurman, and Cesare Furlanello. High resolution forecasting of heat waves impacts on leaf area index by multiscale multitemporal deep learning. *arXiv preprint arXiv:1909.07786*, 2019.

[18] Monidipa Das and Soumya K Ghosh. Deep-step: A deep learning approach for spatiotemporal prediction of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 13(12): 1984–1988, 2016.

[19] Joshua Fan, Junwen Bai, Zhiyun Li, Ariel Ortiz-Bobea, and Carla P Gomes. A gnn-rnn approach for harnessing geospatial and temporal information: application to crop yield prediction. *arXiv preprint arXiv:2111.08900*, 2021.

[20] Changlu Cui, Wen Zhang, ZhiMing Hong, and LingKui Meng. Forecasting ndvi in multiple complex areas using neural network techniques combined feature engineering. *International Journal of Digital Earth*, 13(12):1733–1749, 2020.

[21] Lingjun Kang, Liping Di, Meixia Deng, Eugene Yu, and Yang Xu. Forecasting vegetation index based on vegetation-meteorological factor interactions with artificial neural network. In *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pages 1–6. IEEE, 2016.

[22] Thomas Lees, Gabriel Tseng, Clement Atzberger, Steven Reece, and Simon Dadson. Deep learning for vegetation health forecasting: a case study in kenya. *Remote Sensing*, 14(3):698, 2022.

[23] Andreas Kamilaris and Francesc X Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147:70–90, 2018.

[24] Bernhard Schauberger, Jonas Jägermeyr, and Christoph Gornott. A systematic review of local to regional yield forecasting approaches and frequently used data resources. *European Journal of Agronomy*, 120:126153, 2020.

[25] Saeed Khaki, Lizhi Wang, and Sotirios V Archontoulis. A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science*, 10:1750, 2020.

[26] Christian Requena-Mesa, Vitus Benson, Markus Reichstein, Jakob Runge, and Joachim Denzler. Earthnet2021: A large-scale dataset and challenge for earth surface forecasting as a guided video prediction task. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1132–1142, 2021.

[27] Codruț-Andrei Diaconu, Sudipan Saha, Stephan Günnemann, and Xiao Xiang Zhu. Understanding the role of weather data for earth surface forecasting using a convlstm-based model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1362–1371, 2022.

[28] Klaus-Rudolf William Kladny, Marco Milanta, Oto Mraz, Koen Hufkens, and Benjamin David Stocker. Deep learning for satellite image forecasting of vegetation greenness. *bioRxiv*, 2022.

[29] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

[30] Dara Entekhabi, Eni G Njoku, Peggy E O'Neill, Kent H Kellogg, Wade T Crow, Wendy N Edelstein, Jared K Entin, Shawn D Goodman, Thomas J Jackson, Joel Johnson, et al. The soil moisture active passive (smap) mission. *Proceedings of the IEEE*, 98(5):704–716, 2010.

[31] Hannes Isaak Reuter, Andy Nelson, and Andrew Jarvis. An evaluation of void-filling interpolation methods for srtm data. *International Journal of Geographical Information Science*, 21(9): 983–1008, 2007.

[32] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28, 2015.

[33] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

[34] Viorica Patraucean, Ankur Handa, and Roberto Cipolla. Spatio-temporal video autoencoder with differentiable memory. *arXiv preprint arXiv:1511.06309*, 2015.

[35] J Eamonn Nash and Jonh V Sutcliffe. River flow forecasting through conceptual models part i—a discussion of principles. *Journal of hydrology*, 10(3):282–290, 1970.

[36] Hoshin V Gupta, Harald Kling, Koray K Yilmaz, and Guillermo F Martinez. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, 377(1-2):80–91, 2009.

[37] Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, 120:25–36, 2012.

[38] Rolf H Reichle, Qing Liu, Randal D Koster, Wade T Crow, Gabrielle JM De Lannoy, John S Kimball, Joseph V Ardizzone, David Bosch, Andreas Colliander, Michael Cosh, et al. Version 4 of the smap level-4 soil moisture algorithm and data product. *Journal of Advances in Modeling Earth Systems*, 11(10):3106–3130, 2019.

[39] Krishna Karra, Caitlin Kontgis, Zoe Statman-Weil, Joseph C Mazzariello, Mark Mathis, and Steven P Brumby. Global land use/land cover with sentinel 2 and deep learning. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 4704–4707. IEEE, 2021.

[40] Uwe Muller-Wilm, Jerome Louis, Rudolf Richter, Ferran Gascon, and Marc Niezette. Sentinel-2 level 2a prototype processor: Architecture, algorithms and first results. In *Proceedings of the ESA Living Planet Symposium, Edinburgh, UK*, pages 9–13, 2013.

[41] Jin Chen, Per Jönsson, Masayuki Tamura, Zhihui Gu, Bunkei Matsushita, and Lars Eklundh. A simple method for reconstructing a high-quality ndvi time-series data set based on the savitzky–golay filter. *Remote sensing of Environment*, 91(3-4):332–344, 2004.

[42] William Falcon, Jirka Borovec, Adrian Wälchli, N Eggert, J Schock, J Jordan, N Skafte, V Bereznyuk, E Harris, T Murrell, et al. Pytorchlightning/pytorch-lightning: 0.7. 6 release. *Zenodo: Geneva, Switzerland*, 2020.

[43] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

5 Supplementary material

5.1 Dataset description

Following the EarthNet2021 challenge [26], we design a dataset for landsurface forecasting task on the African continent. The dataset contains over 40,000 samples, each sample is called a *minicubes* given that they are three dimensional arrays. The temporal resolution is 10 days, with 45 time-step. There is 36 time-step per year and each minicube contains 5 seasons. The data come from 2016 to 2021, with a start in four different dates: first of March, June, September or December. Minicubes's length is 15 months. The resolution of the spatial data is around 30 meters per pixels, so an image of with a size of 128×128 pixels represent an area of 3.84×3.84 km. Minicubes in the same location never overlap temporally, but the same location can be in the test and the train set (for different years). Half of the samples are dominated by croplands, the other half is dominated by other land cover classes present in Africa (e.g. forest, mangroves, savanna, meadows, scrub, shrubs or bare land).
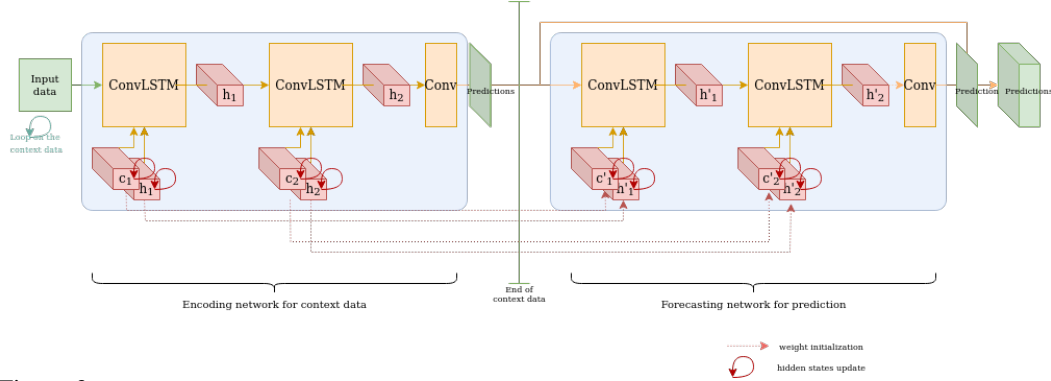
Figure 3: **An Encoding-forecasting architecture**. The encoding network is a stack of two ConvLSTM, the forecasting network is a stack of two ConvLSTM. The hidden states of the first network are initialized at zero, the hidden states of the second network are initialized with the the hidden states' values of the first network. We use skip connections between two predictions, to predict only the difference. The first network is useful to to store as much information as possible during the context period while the second focuses on prediction..
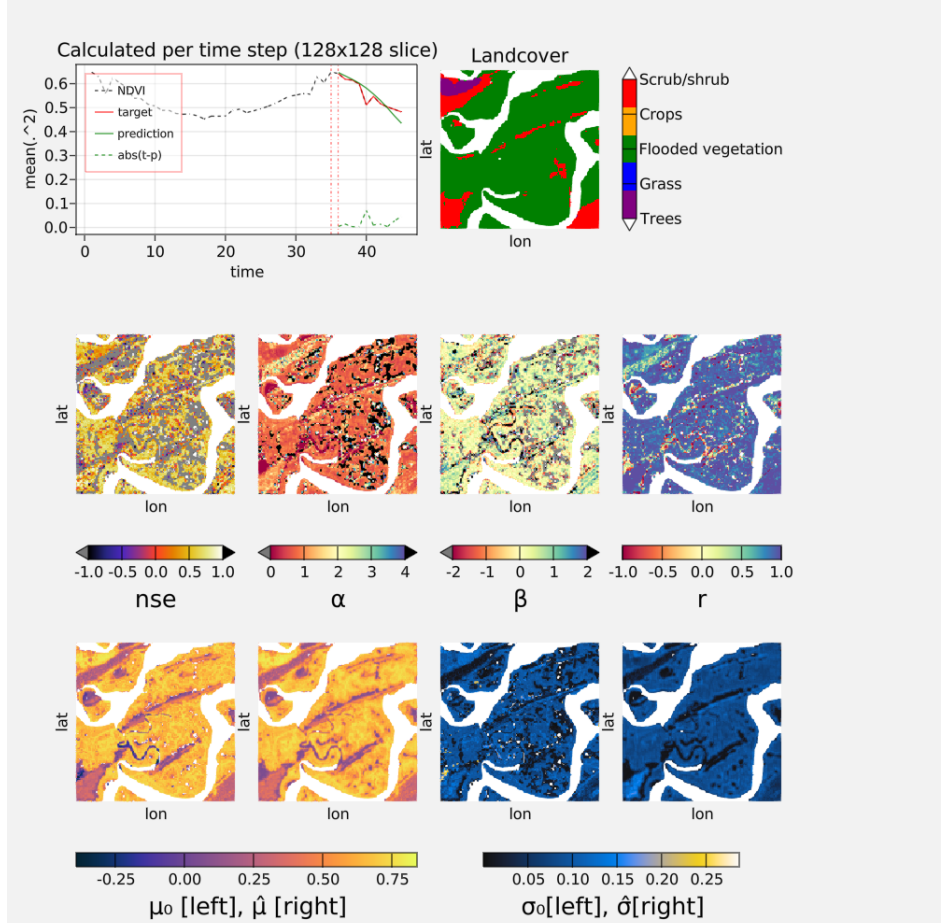

Figure 4: **Nash-Sutcliffe Efficiency decomposition (Top left)** Calculation of the average of each frame, to study the temporal variability. Before the red lines, it is the context, after, the prediction. **(Top right)** Landcover. **Bottom** NSE decomposition, the NSE is per pixel, computed on the time series.

5.2 Data sources

Each minicube contains 45 time-step of high resolution Sentinel-2 satellite imagery in the red, blue, green, and near-infrared, so respectively the B02, B03, B04 and B08 reflectance bands [37]. We
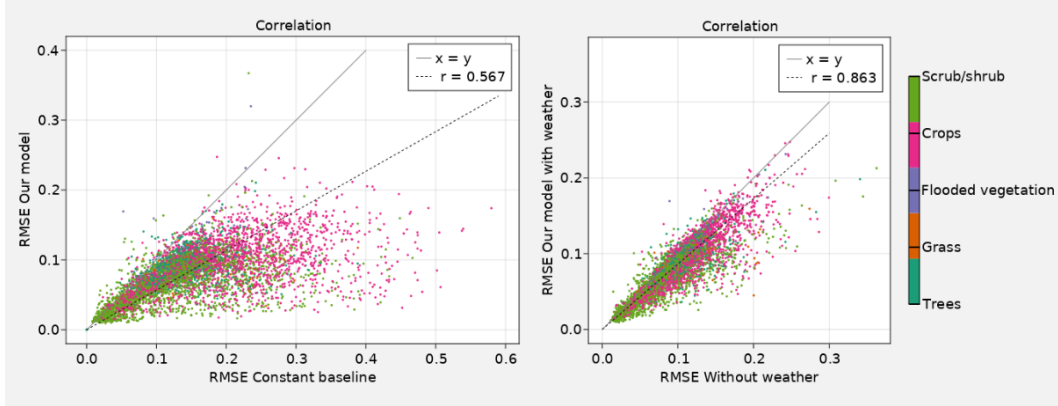
Figure 5: **Correlation between the error of our model against the ablation study and the constant baseline**. The x-axis is the error of the ablation study, the y-axis is the error of our model. Each point below equation x = y is better predicted by our model, each point above is less well predicted. **(Right) Correlation between the error prediction of our model and the constant baseline**. Constant baseline predicts constantly the last NDVI values seen during the context period. The worse the baseline performance, the greater the variation in NDVI during the target period, and thus the more difficult the prediction task can be considered.**(Left) Correlation between the error prediction of our model and the ablation study without weather information**.
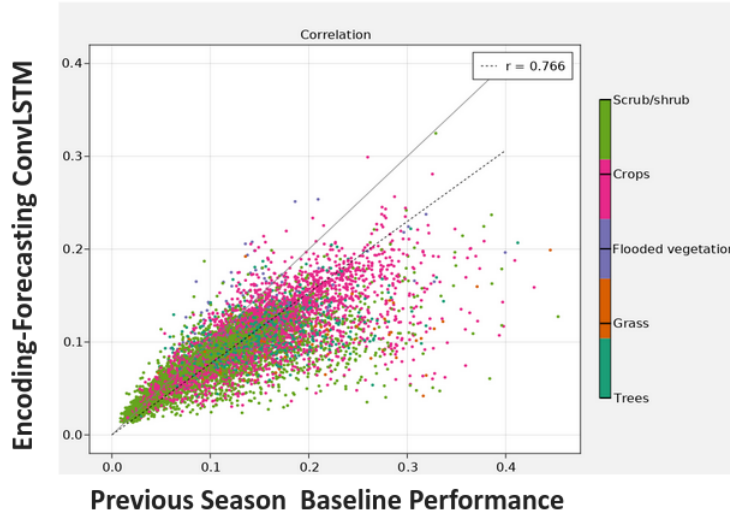

Figure 6: **Correlation between the error prediction of our model and the previous season baseline**. Previous baseline predicts the NDVI values of the previous season. The x-axis is the error of the baseline, the y-axis is the error of our model. The worse the baseline performance, the greater the evolution of the NDVI is different from the previous year, which may be due to different weather conditions.

additionally compute the Normalized Difference Vegetation Index (**NDVI**) [15], a vegetation index (VI) for vegetation health monitoring based on the red and near-infrared bands. This is the target for the landsurface forecasting task. Each minicube contains the **topography** of the location based on a digital elevation model (DEM) of the Shuttle Radar Topography Mission (SRTMv4.1) [31]. The original spatial resolution is 90 meters, we use interpolation to get a 30 meters resolution for each of the 128×128 pixels. For the weather, we use the reanalysis dataset ERA5 [29]. We use **observational weather** to reduce both the uncertainty and the computation cost of a seasonal weather forecasting, the only information available when used for real forecasts. We use the following weather variables: surface pressure, surface net solar radiation, 2m temperature, potential evaporation, and total precipitation. We additionally use the SMAP satellite measurements of the **land surface soil moisture** [38], the soil moisture surface and soil moisture rootzone variables. Finally, each minicube also contains the ESRI2020 **landcover map** [39]. We use the landcover to detect the non-vegetation

9

pixels (water, build area, bare ground, snow or cloud) and mask them during training and evaluation, we also use it for results analysis, but not during the training.

5.3 Pre-processing

The data has been processed using the Level-2A algorithm [40] to the cloud detection following by interpolation on the cloudy data. Then we apply a custom timeseries jump filter based on [41] but adapted for the sentinel-2 data to reduce the noise due to the undetected clouds. However, the pre-processing is imperfect, due to the difficulty of correctly detecting clouds and their shadows. This issue leads to artifacts in the data. The undetected clouds distort the subsequent calculation of NDVI, additionally, when the interpolation is made from undetected cloud images, the result propagates the clouds in the following time steps.

5.4 Training procedure

The loss function is the L2-norm. We apply during training and evaluation a mask to remove all the missing and non-vegetation pixels based on the landcover map. The target is the NDVI and the prediction is done timestep by timestep recurrently. When weather and environmental variables are in input of the ConvLSTM, their dimensions are upscaled to 128×128, moreover, future weather is on input even during the prediction period since it is a strongly guided task. The topography is attached to each timestep. We train our models for 150 epochs with a batch size of 32 and a learning rate of 10^{-6}. Models were implemented using the deep learning framework PyTorch Lightning [42] which is built on top of PyTorch [43].