Multi-scale assessment of a grassland productivity model

Shawn D. Taylor[1] and Dawn M. Browning[1]

[1]U.S. Department of Agriculture, Agricultural Research Service, Jornada Experimental Range, New Mexico State University, Las Cruces, New Mexico, United States

Correspondence: Shawn D. Taylor (shawn.taylor@usda.gov)

Abstract. Grasslands provide many important ecosystem services globally and projecting grassland productivity in the coming decades will provide valuable information to land managers. Productivity models can be well-calibrated at local scales, but generally have some maximum spatial scale in which they perform well. Here we evaluate a grassland productivity model to find the optimal spatial scale for parameterization, and thus for subsequently applying it in future productivity projections for North America. We also evaluated the model on new vegetation types to ascertain its potential generality. We find the model most suitable when incorporating only grasslands, as opposed to also including agriculture and shrublands, and only in the Great Plains and Eastern Temperate Forest ecoregions of North America. The model was not well suited to grasslands in North American Deserts or Northwest Forest ecoregions. It also performed poorly in agriculture vegetation, likely due to management activities, and shrubland vegetation, likely because the model lacks representation of deep water pools. This work allows us to perform long-term projections in areas where model performance has been verified, with gaps filled in by future modelling efforts.

1 Introduction

Grassland systems span nearly 30% of the global land surface (Adams et al., 1990) and play a prominent role in terrestrial carbon cycles (Parton et al., 2012). Grasslands in North America provide a large proportion of food and fiber agricultural products for the region. Annual productivity of grasslands in central and western North America is driven in large part by precipitation (Sala et al., 2012). Future changes in the amount, intensity, and timing of precipitation will be heterogeneous across North America (Easterling et al., 2017), resulting in heterogeneous changes to grassland productivity. For example, even with consistent shifts in climate, different locations can experience different changes in productivity due to local-scale responses (Zhang et al., 2011; Sala et al., 2012; Knapp et al., 2017). This highlights the need for models which can be resolved at small spatial and temporal scales, thus making long-term grassland productivity projections as informative as possible.

A promising method for this is low-dimensional models, which are process based models with some simplified components (Choler et al., 2010, 2011). For example, a low-dimensional model might approximate transpiration as a function of potential evapotranspiration, soil available water, and live vegetation cover along with a single parameter. As opposed to a

high-dimensional model with multiple functions accounting for leaf area index, stomatal conductance, rooting depth and surface area, etc (Caylor et al., 2009; Asbjornsen et al., 2011). The low-dimensional model is advantageous since it can generalize across broad regions with relatively few inputs. Yet they are still susceptible to over-fitting to local conditions since parameters or model structure can be tied to specific locations or plant functional groups (Fisher and Koven, 2020). Thus parameterizing low-dimensional models must be done with care such that they are applicable to a broad area while maintaining an acceptable level of accuracy.

Here we evaluate a low-dimensional model with the intention of it driving productivity projections. The PhenoGrass model developed by Hufkens et al. (2016) is a pulse-response productivity model with temperature and precipitation as the primary drivers. The model is parameterized using observations from the PhenoCam network, which have a small spatial resolution (footprints of < 1ha), sub-daily sampling and sites across all major biomes. These attributes make the PhenoGrass model potentially widely applicable. We expand on the evaluation of the original study by using 84 PhenoCam sites, totalling 89 distinct time series, with 463 site-years of data. ~~We~~ Using the original methods and input datasets we test the model's performance ~~across varying~~ with this expanded dataset. We also evaluate the model across different combinations of North American ecoregions and vegetation types to find an optimal spatial scale in which to parameterize and apply the model. Finally we address where the model performs poorly and how productivity projections for these areas could be implemented or improved.

2 Methods

2.1 PhenoGrass Model

The PhenoGrass model is an ecohydrology model which has interacting state variables for soil water, plant available water, and plant fractional cover (Hufkens et al., 2016). Model inputs are daily precipitation, temperature, potential evapotranspiration (derived from the Hargreaves equation, Hargreaves and Samani (1985)), and solar radiation. The primary output is fractional vegetation cover (fCover). The original model form, derived in Choler et 2010 and Choler et al. 2011, used only temperature and potential evapotranspiration and was parameterized using satellite-derived NDVI data. Hufkens et al. (2016) expanded on the original Choler model by incorporating growth and senescence restraints from temperature and solar radiation, and also included a scaling factor to convert PhenoCam G_{cc} data to a fractional cover estimate (see supplemental equations). Hufkens et al. (2016) evaluated the PhenoGrass model using 14 grassland PhenoCam sites across Western North America with a total of 34 site years. They found the modelled fractional cover correlated well with annual productivity at both a daily and annual timescale. Despite it's name the PhenoGrass model can theoretically apply to any vegetation type with a distinct growth signal in response to precipitation, as hypothesized in the original threshold-delay model (Ogle and Reynolds, 2004). Here we evaluate two additional vegetation types, shrubs and agricultural plots, to test how applicable it is beyond grasslands.
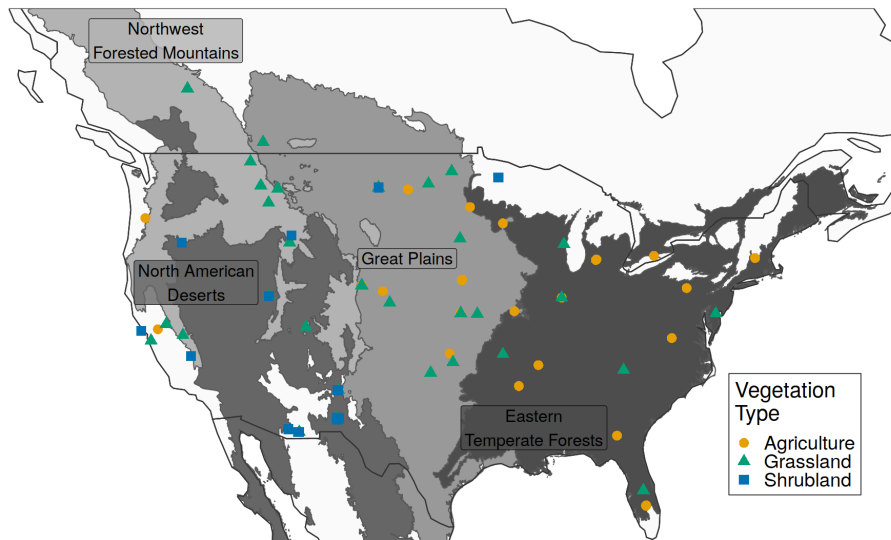
Figure 1. Locations of Phenocam sites. Color indicates the vegetation type represented at each site. Vegetation type is defined by the PhenoCam Network. Shading indicates E.P.A. North American Level 1 Ecoregions.

2.2 Phenocam Data

The PhenoCam network is a global network of fixed, near-surface cameras capturing true-color images of vegetation throughout the day (Richardson et al., 2018a). Using a ratio of the three RGB bands a greenness metric (green chromatic coordinate, G_{cc}) is calculated from each image, resulting in a daily scale time series of canopy greeness. G_{cc} is a unitless metric which is highly correlated with satellite derived NDVI (Richardson et al., 2018b) and flux tower derived primary productivity (Yan et al., 2019; Toomey et al., 2015). Each Phenocam image is subset to one to several different plant vegetation types based on the field of view. These regions of interest (ROI) serve as the basis for the G_{cc} calculation and subsequent post-processing (Seyednasrollah et al., 2019).

We downloaded all Phenocam data with ROIs of the grasslands (GR), shrublands (SH), and agricultural (AG) vegetation types for the years 2012 to 2018, totalling 89 distinct time series and 463 site-years (Fig. 1, Table S2). As input to the PhenoGrass model we used the 3-day smoothed G_{cc} scaled, for each ROI, from 0-1. In the model parameterization each ROI time series is further transformed to fractional vegetation cover (*fCover*, see supplmental equations) using the local mean annual precipitation (MAP) combined with a scaling factor (Hufkens et al., 2016; Donohue et al., 2013).

2.3 Environmental Data

For historic precipitation and temperature we used the 1-km resolution Daymet dataset (Thornton et al., 2018), extracting daily time series for the pixel at each PhenoCam tower location. Daily mean temperature was calculated as the average between the

Daymet daily minimum and maximum temperature, and smoothed with a 15 day moving average. Potential evapotranspiration was calculated using the Hargreaves equation (Hargreaves and Samani, 1985). Soil wilting point and field capacity were extracted at each Phenocam location from a global dataset (Global Soil Data Task Group, 2000). Despite other options being available we chose to use the same climate and soil datasets as those used in Hufkens et al. (2016) so that any differences in results can be attributed to the expanded PhenoCam dataset used here.

2.4 Model Evaluation

To find the most appropriate spatial scale we evaluated the model using three different combinations of ecoregion and vegetation type, with 11 total model parameterizations (Fig. 2). Here we use the term "spatial scale" to refer to the combination of ecoregion/s and vegetation type/s used within each model. This includes using all sites of one vegetation types within an ecoregion, or all sites of a specific vegetation type from several ecoregions (Fig. 2). The largest spatial scale used all Phenocam locations described above (89 sites). Next were all sites, respectively, within the three vegetation types indicated by the ROI (grasslands, shrublands, and agricultural). Finally, we parameterized models for each vegetation type within each Level 1 North American Ecoregion (eg. All grassland sites within the Great Plains ecoregion). All sets of parameterized models were limited to have at least five sites.

We evaluated each of the 11 models using the Nash Sutcliffe coefficient of efficiency (NSE, Eq. 2) as well as the mean coefficient of variation of the mean absolute error (F, Eq. 3) of the daily fractional cover estimates. NSE was calculated for each site and then averaged across all sites within the respective spatial scale ($\overline{NSE}$, Eq. 1). There was no cross-validation using out of sample data in the initial fitting as it would have been computationally expensive. Rather, error metrics from these in-sample tests were treated as a best case scenario in what each model parameterization can achieve. From these results we used a threshold to select which models to evaluate further using cross-validation. The threshold value was an $\overline{NSE}$ of 0.65, which is viewed as "acceptable" for time-series models (Ritter and Muñoz-Carpena, 2013). Parameterization was done using differential evolution, a global optimization algorithm, to minimize F.

$$\overline{NSE} = \frac{1}{N} \sum_{j=1}^{N} NSE_j \tag{1}$$

$$NSE_j = 1 - \frac{\sum_{i=1}^{n} (fCover_{i,obs} - fCover_{i,pred})^2}{\sum_{i=1}^{n} (fCover_{i,obs} - \overline{fCover_{obs}})^2} \tag{2}$$

$$F = \frac{1}{N} \sum_{j=1}^{N} CVMAE_j \tag{3}$$

$$CVMAE_j = \frac{\frac{1}{n} \sum_{i=1}^{n} |fCover_{i,obs} - fCover_{i,pred}|}{\overline{fCover_{obs}}} \tag{4}$$
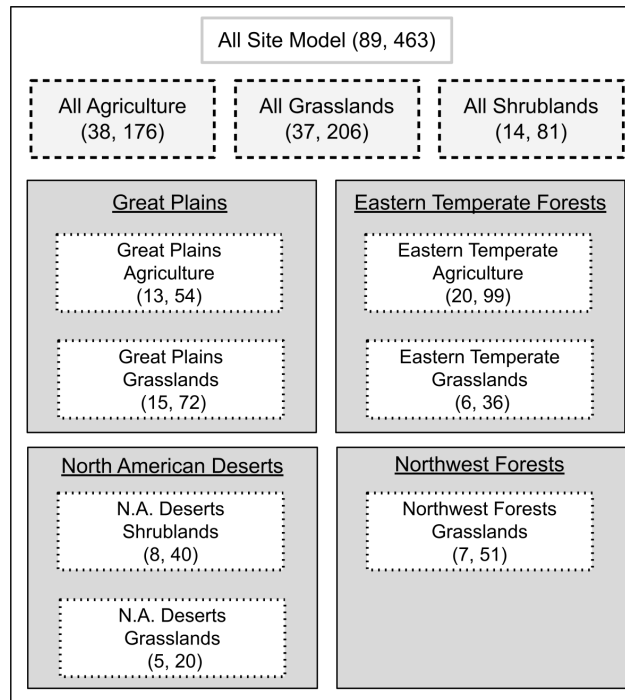
Figure 2. Scaling representation of the 11 model parameterizations. Numbers in parentheses represent the number of sites and site years, respectively. Each model uses a different subset of sites ranging from the entire dataset (All Site Model) to one vegetation type within an ecoregion (e.g., Eastern Temperate Forest Grasslands).

where N is the number of sites within the spatial scale evaluated and n is the number of daily values at each site, $fCover_{i,obs}$ and $fCover_{i,pred}$ are observed and predicted values, respectively. $\overline{fCover_{obs}}$ is the average observed $fCover$ at each site.

Models which exceeded the threshold were subject to further evaluation. For each model we performed a leave one out cross-validation, where the model was re-fit with one Phenocam site not included in the training data, and then evaluated against this left out site. In this step a scaling coefficient to link mean annual precipitation with PhenoCam G_{cc} was held constant at the value obtained in the first fitting. The resulting $\overline{NSE}$ and F are from all modelled sites using their respective out of sample test.

All phenocam data were downloaded using the phenocamr R package (Hufkens et al., 2018). Other packages used in the R 3.6 language were dplyr (Wickham et al., 2017), tidyr (Wickham and Henry, 2018), ggplot2 (Wickham, 2016), daymetr (Hufkens et al., 2018), rgdal (Bivand et al., 2019), and sf (Pebesma, 2018). Python 3.7 packages included scipy (Virtanen et al., 2020), numpy (van der Walt et al., 2011), pandas (McKinney, 2010), and dask (Team, 2016). All code and data used in the analysis is available in the repository at https://github.com/sdtaylor/PhenograssReplication, the PhenoGrass model is implemented in a python package https://github.com/sdtaylor/GrasslandModels. Both are archived permanently on Zenodo (https://doi.org/10.5281/zenodo.3897319).
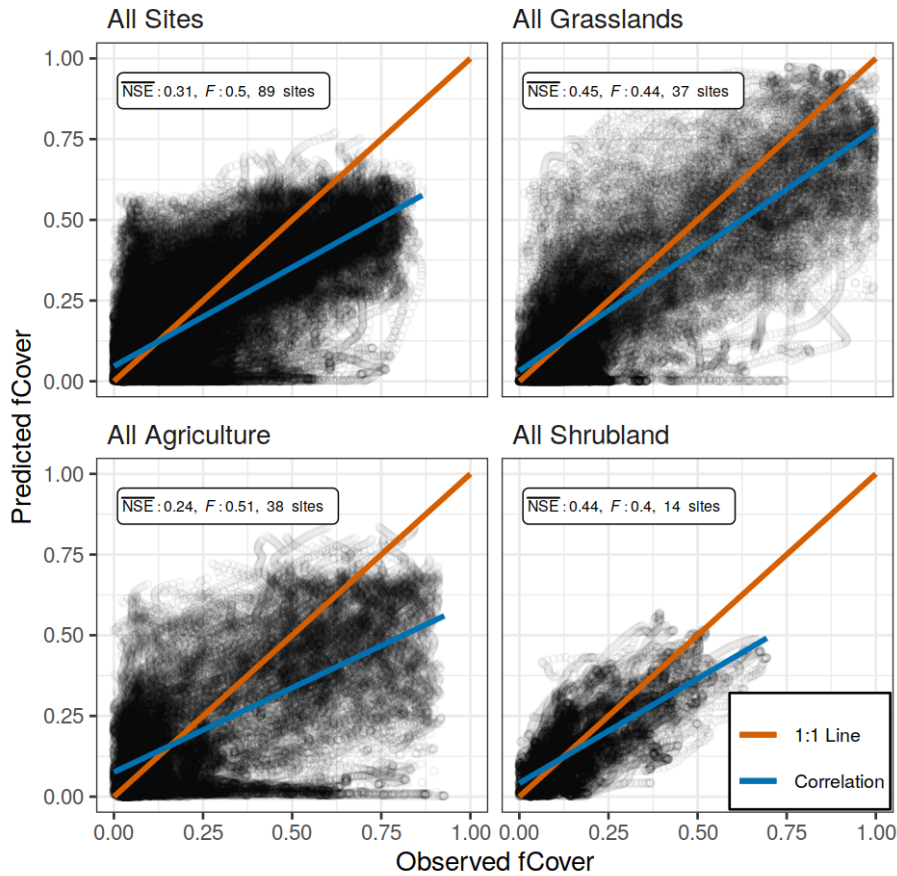
Figure 3. Observed and predicted daily fCover values of the All Site model and the three vegetation type models, each using all available sites and years with the respective spatial scale. NSE is the average Nash Sutcliffe coefficient of efficiency among sites, and F is the mean coefficient of variation of the mean absolute error. The correlation line (blue) represents the overall trend in predicted versus observed values, while the 1:1 line (red) represents a perfect fit.

3 Results

At the largest spatial scale, where the PhenoGrass model was parameterized using all 89 sites, the model performed poorly with an $\overline{NSE}$ value of 0.31 (Table 1; Fig. 3). Models built using all sites of a respective vegetation type performed poorly as well, though were slightly better than the all site model (Fig. 3). The best model performance was achieved when models were built using a single vegetation type subset to a single ecoregion. Grasslands within the Great Plains and Eastern Temperate Forests ecoregions were the only instances where $\overline{NSE}$ exceeded the 0.65 threshold, though Grasslands within N.W. Forests came close (NSE = 0.64).
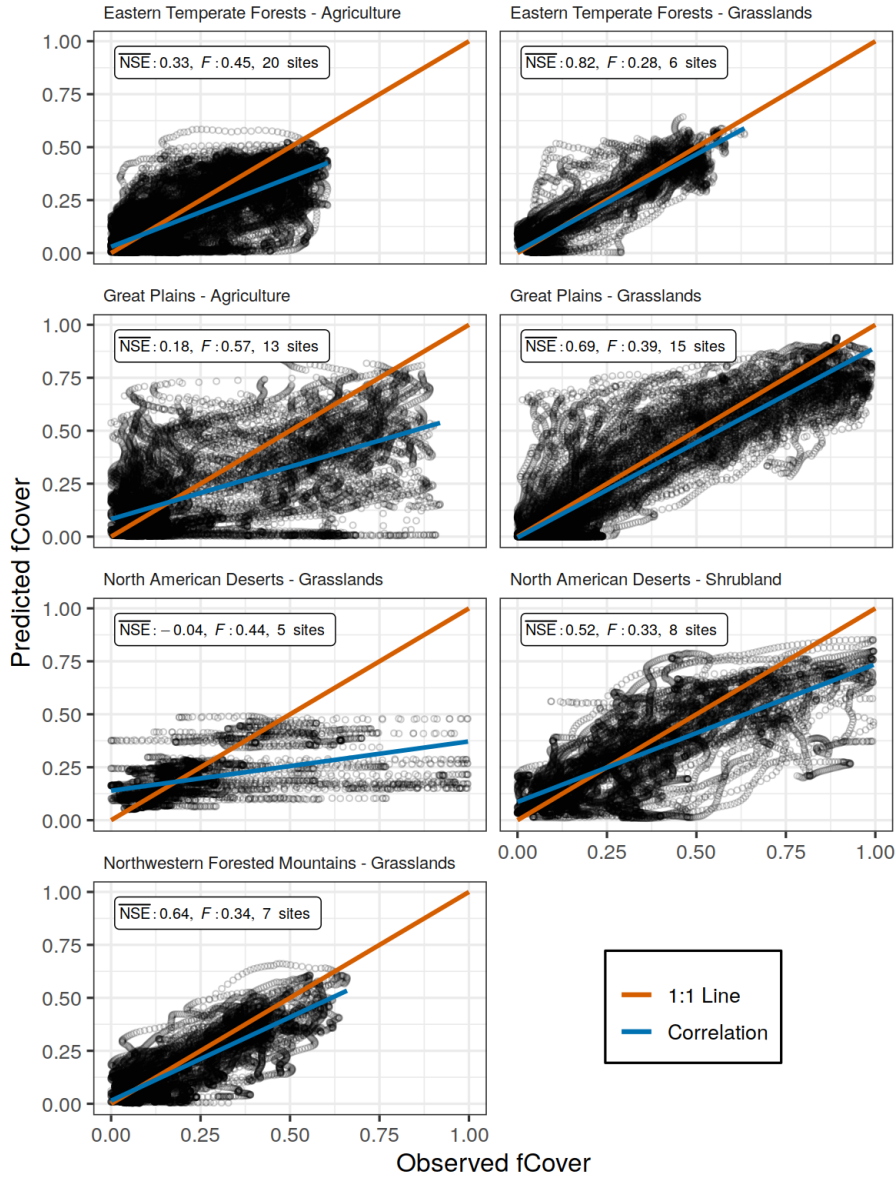
Figure 4. Observed and predicted daily fCover values for models from seven spatial scales, where only specific vegetation types within a single ecoregion were used in model fitting. Each uses all available sites and years with the respective spatial scale. NSE is the average Nash Sutcliffe coefficient of efficiency among sites, and F is the mean coefficient of variation of the mean absolute error. The correlation line (blue) represents the overall trend in predicted versus observed values, while the 1:1 line (red) represents a perfect fit.

In all 11 models the PhenoGrass model tended to underestimate the highest fCover values, and to a lesser degree overpredict the lowest values (Figs. 3,4). The best performing models (Grasslands in the Great Plains and Eastern Temperate forests) minimized this effect (Fig. 4). The worst performing models, Grasslands in N.A. Deserts, had little variation in predicted fCover values, resulting in the lowest $\overline{NSE}$ overall.

5 The grassland vegetation type, subset to specific ecoregions, predominantly outperformed models built with other spatial scales (Table 1). Models built using grasslands within the Eastern Temperate Forest and Great Plains ecoregions had the highest $\overline{NSE}$ values of 0.82 and 0.69, respectively. Using leave one out cross-validation on these two grassland models resulted in similar errors of 0.79 and 0.67 for the Eastern Temperate Forest and Great Plains, respectively. Though N.W. Forests grasslands had an in-sample $\overline{NSE}$ just below the 0.65 threshold, the cross-validation was well below it (0.52). Grasslands in the North

10 American deserts were not modelled well at any spatial scale and had the lowest $\overline{NSE}$ values in the entire analysis. The observed greenness patterns of these desert grasslands had extremely high variability in their magnitude and timing, with short distinct peaks in greenness and numerous off-peak fluctuations. The fitted model, which minimized F among the 5 sites, was not able to reproduce this high variability and instead produced fCover values that were severely constrained to a narrow range (Fig. S1).

15 Agriculture and shrubland sites were poorly modelled at all spatial scales. Performance of agriculture within the E. Temperate Forest ecoregion ($\overline{NSE} = 0.33$) improved over the All Agriculture model ($\overline{NSE} = 0.18$), but decreased in the Great Plains ($\overline{NSE}$ from 0.24 to 0.18). There was only a single ecoregion with a minimum of five shrubland sites, N.A. Deserts, and it performed only slightly better than the All Shrubland model. Shrublands in N.A. Deserts did not have the high variability seen in desert grasslands.

20 **4 Discussion**

We performed an extensive evaluation of the PhenoGrass model across ecoregions and vegetation types to determine the best spatial scale at which to parameterize and apply the model. We found the model most suitable to grassland vegetation when constrained to the ecoregion level, though it did not perform well in grasslands in the North American Deserts ecoregion. Shrublands and agriculture were not well represented by the model regardless of the spatial scale. Results from this study will

25 facilitate long-term projections of grassland productivity constrained to an appropriate vegetation type and scale.

The PhenoGrass model performed best in grassland sites embedded within ecoregions. Studies using earlier forms of the model applied it exclusively to grasslands (Choler et al., 2010, 2011; Hufkens et al., 2016), and results here confirm that it performs well in grassland vegetation with two exceptions. The model did not work in the desert grasslands, nor did it generalize well when built using all North American grasslands simultaneously. Grasslands in the N.A. Desert biome coexist with shrubs,

30 resulting in complex water use dynamics described in more detail below. The pulse-response design of PhenoGrass, which makes it well suited in areas with high cover of perennial grass, is likely not applicable when grasses are interspersed with woody plants.

Table 1. Average site-level Nash Sutcliffe coefficient of efficiency ($\overline{NSE}$) and mean coefficient of variation of the mean absolute error (F) for each model parameterization. Bold indicates when the $\overline{NSE}$ was greater than the acceptable threshold of 0.65. Values in parentheses represent the $\overline{NSE}$ or F in leave-1-out cross validation.

	$\overline{NSE}$	F	Num. Sites	Site Years
All Sites	0.31	0.50	89	462.5
All Agriculture	0.24	0.51	38	175.7
All Grasslands	0.45	0.44	37	205.8
All Shrublands	0.44	0.40	14	81.0
E. Temperate Forests				
Agriculture	0.33	0.45	20	99.2
Grasslands	**0.82 (0.79)**	0.28 (0.30)	6	35.7
Great Plains				
Agriculture	0.18	0.57	13	54.3
Grasslands	**0.69 (0.67)**	0.39 (0.40)	15	71.5
N. American Deserts				
Grasslands	-0.04	0.44	5	20.3
Shrublands	0.52	0.33	8	39.7
N.W. Forests				
Grasslands	0.64	0.34	7	50.7

Shrublands were not well modelled at any spatial scale. Dryland shrubs, representing 8 of the 15 shrubland PhenoCams analysed here, coexist with grasses by accessing different pools of soil water (Weltzin and McPherson, 2000; Muldavin et al., 2008), thus have different responses to precipitation and resulting greenness patterns (Browning et al., 2017; Yan et al., 2019). A prior form of the PhenoGrass model was designed to work with dryland shrubs by using two soil water pools (Ogle and Reynolds, 2004), yet here PhenoGrass, with a single soil water pool, was less effective for shrubland vegetation. The single pool of the PhenoGrass model is coupled with fluxes from precipitation and evapotranspiration, thus is not well-suited for representing the deeper water pools that shrubs can routinely access (Schenk and Jackson, 2002; Ward et al., 2013). Potential improvements would likely need to incorporate a deep soil water pool, in addition to the shallow, which are each utilized by the respective plant functional groups. This has already been implemented in highly parameterized ecohydrology models (Scanlon et al., 2005; Lauenroth et al., 2014) and could potentially be used here to make a more generalized PhenoShrub model to apply across large scales. This approach could also help in modelling N.A. Desert grasslands which coexist among shrubs.

Agriculture areas performed poorly with the PhenoGrass model. Management practices of crops artificially increase productivity beyond what would naturally occur, and planting and harvest result in abrupt changes in greenness metrics (Bégué et al., 2018). While the results were not necessarily surprising, to our knowledge this is the first attempt to use near-surface images to drive an productivity model for agricultural vegetation. We have shown that the PhenoGrass model, designed for natural sys-
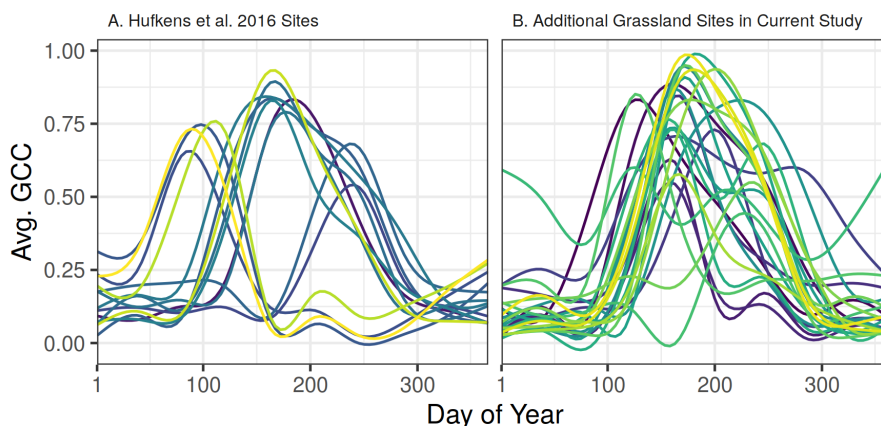
Figure 5. Smoothed time series for all 14 grassland sites used in Hufkens et al 2016 (A) and 24 additional grassland sites added in the current study (B). Each line represents the long-term average Green Chromatic Coordinate of a single site across all available years, smoothed using a GAM model.

tems, does not generalize to actively managed agricultural systems. Future work in using PhenoCam data to model agricultural productivity would likely need to incorporate crop specific parameters and management activity, which other cropland modelling systems use (Fritz et al., 2019). The integration of the PhenoCam network within the Long-Term Agricultural Research (LTAR) will likely be beneficial for this, as the timing and intensity of management activities or experimental treatments can
5 be incorporated into modelling efforts.

   Hufkens et al. (2016) originally evaluated the PhenoGrass model using 14 grassland sites distributed among seven North American ecoregions. In their evaluation they had an average ~~R²~~ NSE of 0.71, while here the model performed poorly when using more than one ecoregion. It's likely that the original 14 grassland sites were ideal locations for the PhenoGrass model, since on average they have a single greenup season every year in the spring or summer (Fig. 5A). The additional 24 grassland
10 sites used in the current study have high seasonal variability and elongated growing seasons (Fig. 5B, Fig. S1), and were thus more difficult to represent in a single continental scale grassland model. This highlights the need for longer time series in evaluating low-dimensional models as it may take many years for a single location to experience the full range of variability. As the PhenoCam data archive grows, temporally out of sample validation can be done to better evaluate performance into novel conditions. Further work could use the expanded PhenoCam timeseries to compare with flux tower derived productivity,
15 which Hufkens et al. (2016) found was well correlated at the five original locations with available flux data.

5   Conclusions

Replication is an important step in the scientific process, especially given newly available data. Here we have validated prior modelling work and highlighted its limitations. Newer vegetation models can be validated in the same framework and applied

to areas where PhenoGrass performs poorly. This can result in a spatial ensemble where the output for any one location and vegetation type is represented by the most appropriate model. Our current work will allow for long-term projections of grassland productivity for a large fraction of North America.

Code and data availability. All code and data used in the analysis is available in the repository at https://github.com/sdtaylor/PhenograssReplication,
5   the PhenoGrass model is implemented in a python package https://github.com/sdtaylor/GrasslandModels. Both are archived permanently on Zenodo (https://doi.org/10.5281/zenodo.3897319).

Author contributions. Shawn concieved of the project. Shawn and Dawn wrote the paper.

Competing interests. The authors declare no competing interests.

Acknowledgements. This research was a contribution from the Long-Term Agroecosystem Research (LTAR) network. LTAR is supported by
10 the United States Department of Agriculture. The authors acknowledge the USDA Agricultural Research Service (ARS) Big Data Initiative and SCINet high performance computing resources (https://scinet.usda.gov) made available for conducting the research reported in this paper.
    We thank our many collaborators, including site PIs and technicians, for their efforts in support of PhenoCam. The development of PhenoCam has been funded by the Northeastern States Research Cooperative, NSF's Macrosystems Biology program (awards EF-1065029
15 and EF-1702697), and DOE's Regional and Global Climate Modeling program (award DE-SC0016011).

References

Adams, J. M., Faure, H., Faure-Denard, L., McGlade, J. M., and Woodward, F. I.: Increases in terrestrial carbon storage from the Last Glacial Maximum to the present, Nature, 348, 711–714, https://doi.org/10.1038/348711a0, 1990.

Asbjornsen, H., Goldsmith, G. R., Alvarado-Barrientos, M. S., Rebel, K., Van Osch, F. P., Rietkerk, M., Chen, J., Gotsch, S., Tobon, C., Geissert, D. R., Gomez-Tagle, A., Vache, K., and Dawson, T. E.: Ecohydrological advances and applications in plant-water relations research: a review, Journal of Plant Ecology, 4, 3–22, https://doi.org/10.1093/jpe/rtr005, https://academic.oup.com/jpe/article-lookup/doi/10.1093/jpe/rtr005, 2011.

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., de Abelleyra, D., P. D. Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., and R. Verón, S.: Remote Sensing and Cropping Practices: A Review, Remote Sensing, 10, 99, https://doi.org/10.3390/rs10010099, 2018.

Bivand, R., Keitt, T., and Rowlingson, B.: rgdal: Bindings for the 'Geospatial' Data Abstraction Library, https://cran.r-project.org/package=rgdal, 2019.

Browning, D. M., Maynard, J. J., Karl, J. W., and Peters, D. C.: Breaks in MODIS time series portend vegetation change: verification using long-term data in an arid grassland ecosystem, Ecological Applications, 27, 1677–1693, https://doi.org/10.1002/eap.1561, 2017.

Caylor, K. K., Scanlon, T. M., and Rodriguez-Iturbe, I.: Ecohydrological optimization of pattern and processes in water-limited ecosystems: A trade-off-based hypothesis, Water Resources Research, 45, 1–15, https://doi.org/10.1029/2008WR007230, http://doi.wiley.com/10.1029/2008WR007230, 2009.

Choler, P., Sea, W., Briggs, P., Raupach, M., and Leuning, R.: A simple ecohydrological model captures essentials of seasonal leaf dynamics in semi-arid tropical grasslands, Biogeosciences, 7, 907–920, https://doi.org/10.5194/bg-7-907-2010, 2010.

Choler, P., Sea, W., and Leuning, R.: A Benchmark Test for Ecohydrological Models of Interannual Variability of NDVI in Semi-arid Tropical Grasslands, Ecosystems, 14, 183–197, https://doi.org/10.1007/s10021-010-9403-9, 2011.

Donohue, R. J., Roderick, M. L., McVicar, T. R., and Farquhar, G. D.: Impact of CO_2 fertilization on maximum foliage cover across the globe's warm, arid environments, Geophysical Research Letters, 40, 3031–3035, https://doi.org/10.1002/grl.50563, 2013.

Easterling, D. R., Kunkel, K., Arnold, J., Knutson, T. R., LeGrande, A., Leung, L., Vose, R. S., Waliser, D., and Wehner, M.: Precipitation change in the United States, in: Climate Science Special Report: Fourth National Climate Assessment, Volume I, vol. I, pp. 207–230, U.S. Global Change Research Program, Washington, D.C., USA, https://doi.org/10.7930/J0H993CC.U.S., 2017.

Fisher, R. A. and Koven, C. D.: Perspectives on the future of Land Surface Models and the challenges of representing complex terrestrial systems, Journal of Advances in Modeling Earth Systems, pp. 0–3, https://doi.org/10.1029/2018ms001453, 2020.

Fritz, S., See, L., Bayas, J. C. L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J., Rembold, F., Rojas, O., Schucknecht, A., Van der Velde, M., Verdin, J., Wu, B., Yan, N., You, L., Gilliams, S., Mücher, S., Tetrault, R., Moorthy, I., and McCallum, I.: A comparison of global agricultural monitoring systems and current gaps, Agricultural Systems, 168, 258–272, https://doi.org/10.1016/j.agsy.2018.05.010, 2019.

Global Soil Data Task Group: Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS), https://doi.org/10.3334/ORNLDAAC/569, 2000.

Hargreaves, G. H. and Samani, Z. A.: Reference Crop Evapotranspiration from Temperature, Applied Engineering in Agriculture, 1, 96–99, https://doi.org/10.13031/2013.26773, 1985.

Hufkens, K., Keenan, T. F., Flanagan, L. B., Scott, R. L., Bernacchi, C. J., Joo, E., Brunsell, N. A., Verfaillie, J., and Richardson, A. D.: Productivity of North American grasslands is increased under future climate scenarios despite rising aridity, Nature Climate Change, 6, 710–714, https://doi.org/10.1038/nclimate2942, 2016.

Hufkens, K., Basler, D., Milliman, T., Melaas, E. K., and Richardson, A. D.: An integrated phenology modelling framework in <scp>r</scp>, Methods in Ecology and Evolution, 9, 1276–1285, https://doi.org/10.1111/2041-210X.12970, 2018.

Knapp, A. K., Ciais, P., and Smith, M. D.: Reconciling inconsistencies in precipitation-productivity relationships: implications for climate change, New Phytologist, 214, 41–47, https://doi.org/10.1111/nph.14381, 2017.

Lauenroth, W. K., Schlaepfer, D. R., and Bradford, J. B.: Ecohydrology of Dry Regions: Storage versus Pulse Soil Water Dynamics, Ecosystems, 17, 1469–1479, https://doi.org/10.1007/s10021-014-9808-y, 2014.

McKinney, W.: Data Structures for Statistical Computing in Python, in: Proceedings of the 9th Python in Science Conference, pp. 51–56, SciPy, Austin, Texas, USA, http://conference.scipy.org/proceedings/scipy2010/mckinney.html, 2010.

Muldavin, E. H., Moore, D. I., Collins, S. L., Wetherill, K. R., and Lightfoot, D. C.: Aboveground net primary production dynamics in a northern Chihuahuan Desert ecosystem, Oecologia, 155, 123–132, https://doi.org/10.1007/s00442-007-0880-2, 2008.

Ogle, K. and Reynolds, J. F.: Plant responses to precipitation in desert ecosystems: integrating functional types, pulses, thresholds, and delays, Oecologia, 141, 282–294, https://doi.org/10.1007/s00442-004-1507-5, 2004.

Parton, W., Morgan, J., Smith, D., Del Grosso, S., Prihodko, L., LeCain, D., Kelly, R., and Lutz, S.: Impact of precipitation dynamics on net ecosystem productivity, Global Change Biology, 18, 915–927, https://doi.org/10.1111/j.1365-2486.2011.02611.x, 2012.

Pebesma, E.: Simple Features for R: Standardized Support for Spatial Vector Data, The R Journal, 10, 439–446, https://doi.org/10.32614/RJ-2018-009, 2018.

Richardson, A. D., Hufkens, K., Milliman, T., Aubrecht, D. M., Chen, M., Gray, J. M., Johnston, M. R., Keenan, T. F., Klosterman, S. T., Kosmala, M., Melaas, E. K., Friedl, M. A., and Frolking, S.: Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery, Scientific Data, 5, 180 028, https://doi.org/10.1038/sdata.2018.28, 2018a.

Richardson, A. D., Hufkens, K., Milliman, T., and Frolking, S.: Intercomparison of phenological transition dates derived from the PhenoCam Dataset V1.0 and MODIS satellite remote sensing, Scientific Reports, 8, 5679, https://doi.org/10.1038/s41598-018-23804-6, 2018b.

Ritter, A. and Muñoz-Carpena, R.: Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments, Journal of Hydrology, 480, 33–45, https://doi.org/10.1016/j.jhydrol.2012.12.004, 2013.

Sala, O. E., Gherardi, L. A., Reichmann, L., Jobbagy, E., and Peters, D.: Legacies of precipitation fluctuations on primary production: theory and data synthesis, Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 3135–3144, https://doi.org/10.1098/rstb.2011.0347, 2012.

Scanlon, T. M., Caylor, K. K., Manfreda, S., Levin, S. A., and Rodriguez-Iturbe, I.: Dynamic response of grass cover to rainfall variability: implications for the function and persistence of savanna ecosystems, Advances in Water Resources, 28, 291–302, https://doi.org/10.1016/j.advwatres.2004.10.014, 2005.

Schenk, H. J. and Jackson, R. B.: Rooting depths, lateral root spreads and below-ground/above-ground allometries of plants in water-limited ecosystems, Journal of Ecology, 90, 480–494, https://doi.org/10.1046/j.1365-2745.2002.00682.x, https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2745.2002.00682.x, 2002.

Seyednasrollah, B., Young, A. M., Hufkens, K., Milliman, T., Friedl, M. A., Frolking, S., and Richardson, A. D.: Tracking vegetation phenology across diverse biomes using Version 2.0 of the PhenoCam Dataset, Scientific data, 6, 222, https://doi.org/10.1038/s41597-019-0229-9, 2019.

Team, D. D.: Dask: Library for dynamic task scheduling, https://dask.org, 2016.

Thornton, P. E., Thornton, M. M., Mayer, B. W., Wei, Y., Devarakonda, R., Vose, R., and Cook, R. B.: Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3., Tech. rep., ORNL DAAC, Oak Ridge, Tennessee, USA, Oak Ridge, Tennessee, USA, https://doi.org/10.3334/ORNLDAAC/1328, 2018.

5   Toomey, M., Friedl, M. A., Frolking, S., Hufkens, K., Klosterman, S., Sonnentag, O., Baldocchi, D. D., Bernacchi, C. J., Biraud, S. C., Bohrer, G., Brzostek, E., Burns, S. P., Coursolle, C., Hollinger, D. Y., Margolis, H. A., McCaughey, H., Monson, R. K., Munger, J. W., Pallardy, S., Phillips, R. P., Torn, M. S., Wharton, S., Zeri, M., and Richardson, A. D.: Greenness indices from digital cameras predict the timing and seasonal dynamics of canopy-scale photosynthesis, Ecological Applications, 25, 99–115, https://doi.org/10.1890/14-0005.1, 2015.

10 van der Walt, S., Colbert, S. C., and Varoquaux, G.: The NumPy Array: A Structure for Efficient Numerical Computation, Computing in Science and Engineering, 13, 22–30, https://doi.org/10.1109/MCSE.2011.37, 2011.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald,

15 A. M., Ribeiro, A. H., Pedregosa, F., and van Mulbregt, P.: SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Ward, D., Wiegand, K., and Getzin, S.: Walter's two-layer hypothesis revisited: back to the roots!, Oecologia, 172, 617–630, https://doi.org/10.1007/s00442-012-2538-y, http://link.springer.com/10.1007/s00442-012-2538-y, 2013.

Weltzin, J. F. and McPherson, G. R.: Implications of precipitation redistribution for shifts in temperate savanna ecotones, Ecology, 81,

20   1902–1913, 2000.

Wickham, H.: ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, http://ggplot2.org, 2016.

Wickham, H. and Henry, L.: tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions, https://cran.r-project.org/package=tidyr, 2018.

Wickham, H., Francois, R., Henry, L., and Müller, K.: dplyr: A Grammar of Data Manipulation, https://cran.r-project.org/package=dplyr, 2017.

25 Yan, D., Scott, R., Moore, D., Biederman, J., and Smith, W.: Understanding the relationship between vegetation greenness and productivity across dryland ecosystems through the integration of PhenoCam, satellite, and eddy covariance data, Remote Sensing of Environment, 223, 50–62, https://doi.org/10.1016/j.rse.2018.12.029, 2019.

Zhang, L., Wylie, B. K., Ji, L., Gilmanov, T. G., Tieszen, L. L., and Howard, D. M.: Upscaling carbon fluxes over the Great Plains grasslands: Sinks and sources, Journal of Geophysical Research: Biogeosciences, 116, 1–13, https://doi.org/10.1029/2010JG001504, 2011.