

PREPRINT

Author-formatted, not peer-reviewed document posted on 10/04/2024

DOI: <https://doi.org/10.3897/arphapreprints.e125077>

Prototype Biodiversity Digital Twin: Grassland Biodiversity Dynamics

 Franziska Taubert,  Tuomas Rossi,  Christoph Wohner, Sarah Venier, 
Tomáš Martinovič,  Taimur Khan,  Julian Gordillo, Thomas Banitz

Prototype Biodiversity Digital Twin: Grassland Biodiversity Dynamics

Franziska Taubert^{‡,§}, Tuomas Rossil, Christoph Wohner[¶], Sarah Venier[¶], Tomáš Martinovič[#], Taimur Haider Khan[□], Julian Lopez Gordillo[«], Thomas Banitz[‡]

[‡] Department of Ecological Modelling, Helmholtz Centre for Environmental Research GmbH – UFZ, Leipzig, Germany

[§] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

| CSC – IT Center for Science Ltd., Espoo, Finland

[¶] Department for Ecosystem Research and Monitoring, Environment Agency Austria, Vienna, Austria

[#] IT4Innovations, VSB – Technical University of Ostrava, Ostrava-Poruba, Czech Republic

[□] Department of Community Ecology, Helmholtz Centre for Environmental Research GmbH – UFZ, Leipzig, Germany

[«] Naturalis Biodiversity Center, Leiden, Netherlands

Corresponding author: Franziska Taubert (franziska.taubert@ufz.de), Thomas Banitz (thomas.banitz@ufz.de)

Reviewable

v 1

Abstract

European grassland management has often favored high production through frequent mowing and heavy fertilization over biodiversity conservation, which is typically supported by less intensive management. Besides management, climate change and extremes are increasingly affecting grassland productivity and biodiversity, requiring timely adaptation of management practices. Here, we describe the development of a prototype Digital Twin (pDT) of grassland biodiversity dynamics intended to support researchers, farmers or regulatory decision-makers in monitoring the current state of selected grassland sites and projecting their future state under various management and climate scenarios.

Keywords

Ecological modeling, Ecosystem service, Ecosystem management, Model-data fusion, High performance computing

Introduction

Approximately 30% of Europe's agricultural land area is covered by grassland (European Commission et al. 2020). Grasslands often occur as agricultural sites, managed by farmers according to their goals and traditions, site conditions, subsidies and regulations. Common practices include:

1. grazing of sheep, cattle and other livestock for food production:

2. mowing for hay and biogas production or for grassland conservation (e.g. to prevent shrub encroachment)
3. fertilization to increase productivity or
4. irrigation, for example to mitigate water scarcity due to climate change.

Grassland farmers mostly favor production (i.e. high yields). Therefore, frequent mowing (up to six times per year) and intense fertilization are common practices on agriculturally used grassland sites. Such intensive management is often at the expense of biodiversity, as it likely favors the dominance of only a few grass species (and the suppression of forbs and legumes). By contrast, low to moderate management intensities rather favor biodiversity. Grassland sites of nature conservation areas, for example, are usually mowed only once or twice per year. They can show a richness of hundreds of different plant species per hectare (Öster et al. 2007). High biodiversity can be critical for grassland persistence under changing conditions, but also for the habitat quality and vitality of other trophic organisms, such as honeybees and wild bees (Decourtye et al. 2010, Evans et al. 2018), earthworms (Piotrowska et al. 2013) and butterflies (Kruse et al. 2016). Besides this tradeoff between management for high grassland productivity or high biodiversity, climate change and extremes increasingly impede both of them. This can have cascading negative effects on other trophic species, but also on fodder and bioenergy supply, or food quality and security (Berauer et al. 2020).

Consequently, well-established management practices may no longer be suitable to achieve their goals and may require adaptation. Farmers increasingly face the question of how best to manage their grassland to achieve high yields, while conserving (or enhancing) biodiversity, and how to adapt management practices to climate change to secure both in the future. However, we still lack a comprehensive understanding of how grassland dynamics and biodiversity respond to changing anthropogenic, environmental and climatic drivers, even more so as these drivers interact. Scientific knowledge and insights gained from observations at specific locations or short-term experiments cannot be directly transferred to other sites with different environmental conditions and can hardly be used to project grassland dynamics under (uncertain) future conditions. Moreover, available observations on plant diversity and productivity in grasslands are still scarce and heterogeneously distributed across Europe. This complicates deriving generalizable knowledge and recommendations for action.

An important complement to observational studies are computational models of grassland dynamics. Especially mechanistic simulation models that capture relevant processes and drivers such as climate, soil conditions and management on grassland yield and biodiversity can help to close our knowledge gaps. Appropriately designed and analyzed, these models can allow to generally assess the role and importance of specific drivers, and also to project dynamics under various (future) scenarios (Gustafson 2013). But we need to ensure that model projections are robust, reliable and realistic, especially when we use them to derive management recommendations. Therefore, model projections should be frequently confronted with available observation data. To this end, the Digital Twin approach provides a highly suitable framework (De Koning et al. 2023).

Objectives

Our mission is a consistent scientific knowledge base on grassland dynamics and biodiversity under different environmental conditions. This will allow reliable recommendations for grassland management under prevailing (e.g. improving biodiversity while maintaining yields) and changing conditions (e.g. securing yield and biodiversity under climate extremes like drought).

Therefore, we develop a prototype Digital Twin (pDT) of grassland biodiversity dynamics that allows end-users (e.g. farmers, regulatory decision-makers) to select a specific grassland site, monitor its current state (including uncertainty measures depending on data availability), and project its future state under pre- or self-defined management and climate scenarios.

Although the ultimate end-users will be farmers and regulators, our primary audience at the current stage of pDT development are grassland biodiversity researchers. Besides advancing the pDT workflow and implementation, we aim to improve its predictive capacity and accuracy for specific sites. Therefore, a close exchange with grassland experts and researchers managing observation sites (as organized in the Integrated European Long-Term Ecosystem, critical zone and socio-ecological Research infrastructure *eLTER*) is crucial.

Workflow

The grassland biodiversity pDT workflow includes retrieving and processing required data, running simulations with the model GRASSMIND (cf. Model), exploring the simulation output by users, and comparing simulation output with observation data (Fig. 1). In particular, weather, soil and management input data for a desired location and time period need to be prepared to run simulations. The simulated dynamics of grassland vegetation allow computation and visualization of multiple metrics of biodiversity and productivity as time series, which can be analyzed by users and spark demand for new simulation scenarios or new data. If available, the simulated vegetation dynamics can also be compared to observations. This may lead to recalibration of model parameters or to changes of the model itself.

Data

Input data on weather, soil characteristics and grassland management events are required to run simulations with GRASSMIND. If available, observation data on grassland vegetation can be used for model validation, recalibration and further improvements. All data refer to the location of a grassland site or a particular plot to be simulated (i.e. spatial point coordinates). The target format of all input and observation data is described in a public guideline (Taubert et al. 2023). Specific scripts have been and will be developed for

streaming these data from different sources (Table 1) and processing the variables as needed for the pDT workflow (e.g. unit conversion, modification of temporal resolution, calculation of additional variables, completion of missing data with default assumptions). Available data sources for weather and soil characteristics cover most of Europe (Table 1). However, data on management is still scarce and often only covers specific regions, specific years or specific aspects (e.g. only mowing, Table 1). For locations where no such data is available, default scenarios of extensive and intensive management will be used.

For observation data, we launched a call to data holders of all eLTER grassland sites. Responses to this call get processed and suitable grassland vegetation data sets published (e.g. [Unknown] 2024a, [Unknown] 2024b).

Model

The pDT employs the individual-based grassland model GRASSMIND (Taubert et al. 2020). For a given grassland plot (typically an area of several m²), GRASSMIND explicitly simulates the processes that let biodiversity dynamics emerge (at a resolution of 1m², Fig. 2). Individual plants can establish, grow and die. These processes are influenced by the plants' interaction and competition for light, space and other limited resources. Thereby, the individuals differ in their traits (as they belong to different plant functional types (PFTs)) and their state (i.e. size). The simulated processes are further affected by external drivers such as weather, soil conditions, mowing events or fertilization (cf. Data).

The simulation results provide multiple vegetation characteristics at different organizational levels (individual plants, populations of PFTs, plant community) as time series (daily resolution). We focus on output characteristics related to biodiversity and productivity (e.g. biomass of different PFTs, Fig. 2), and additional output for which observations are available during model calibration (e.g. vegetation cover or leaf area index). The model is programmed in C++. It runs on Windows and Linux systems.

FAIRness

The pDT aims at a high level of FAIRness (Wilkinson 2016) by releasing its digital objects on relevant open repositories with a persistent identifier and descriptive metadata. To this end, we follow the FAIR Digital Objects framework for interoperability (De Smedt et al. 2020), implemented through the Research Object Crate format (Soiland-Reyes et al. 2022). The GRASSMIND model code will be provided as an open-source repository, supported by documentation and a technical guide. Pipeline scripts for different workflow steps (Fig. 1; e.g. retrieving and processing input data, model calibration) will be available as open source on the BioDT repository on GitHub (<https://github.com/BioDT>) and on the BioDT Space on the WorkflowHub registry (<https://workflowhub.eu/programmes/22>, Goble et al. 2021). Input data comes from various openly accessible sources (cf. Table 1). Observation data (from eLTER grassland sites, cf. Data) is partly published and openly accessible on B2Share (e.g. [Unknown] 2024a, [Unknown] 2024b) and more datasets will follow.

Performance

We expect to run tens of thousands GRASSMIND simulations in order to model many different grassland sites as well as climate and management scenarios, which will highly benefit from the parallel processing capabilities in LUMI-C. The runtime for preparing input files and simulating 160 instances of GRASSMIND (10 year simulation period, 1m² area) is 8 minutes on a local machine without parallelization, 2 minutes with parallelization (10 cores), 25 seconds on a Windows-based HPC system Model Server Grid with parallelization (56 cores), and 5 seconds on a single LUMI-C node (128 cores). Test runs on LUMI were used to assess the number of single stochastic replicate simulations required (160 replicates) such that the mean outcome over all replicates becomes approximately invariant and, thus, can be reasonably considered as representative (e.g. during model calibration).

Interface and outputs

The pDT interface for end-user interaction is designed as an R Shiny App. End-users can submit the site location (spatial coordinates, or DEIMS.iD if the site is listed at the eLTER DEIMS-SDR, Wohner et al. 2022, Wohner et al. 2019). Based on this user input, the pDT workflow will be run and provide an output summary for the desired location (Fig. 3). The selection and visualization of output is under development and will comprise different figures on grassland dynamics and composition (cf. Model) for the simulated environmental and management conditions. If available, observation data are included and compared to the simulation results. Functionalities for users to simulate and explore various future climate scenarios and management regimes (e.g. for one year, five years or decades) are under development. They will include the choice and combination of pre-defined scenarios and novel user-specified scenarios within the interface (e.g. mowing events).

Integration and sustainability

To integrate climate projections developed by the Destination Earth initiative (European Commission 2023), climatic variables that get available can be retrieved and processed similarly to the current way for weather input data (cf. Data).

Some elements of the workflow can be used beyond the context of this grassland pDT. Such 'generic building blocks' include, for example, the scripts to retrieve location-specific weather data from the Copernicus ERA5-Land data set (Muñoz Sabater 2019) or soil data from the SoilGrids 2.0 (Poggio et al. 2021) and HiHydroSoil v2.0 (Simons et al. 2020) data sets (cf. Table 1). These scripts will be publicly available on the BioDT repository on GitHub (<https://github.com/BioDT>).

Application and impact

The fully developed pDT, including technical implementation as well as robust reliable model projections, can serve as an information and decision-support tool for farmers and regulators, for example, to test different management regimes. To this end, expanding the pDT scope from local sites to larger regions (the scales of regulatory measures) and accounting for regional-scale effects of land use change will be essential.

Scaling up the pDT to cover even larger regions, countries or Europe opens another perspective: comprehensive assessment of grassland dynamics in response to environmental factors (weather, soil, management). A Digital Twin map covering grassland across Europe could reveal potentials and limits for yield, plant diversity or other variables of interest. Generalized relationships among these variables and between them and environmental conditions could be derived. The map could also help identify vulnerable sites that need specific attention (e.g. sites at risk of biodiversity and/or productivity loss that require protection, or sites with high projection variability that require more monitoring).

At the core of the pDT, model predictions will be frequently checked with available observations. To avoid computationally expensive calibration for various local sites across Europe, the pDT shall capture grassland dynamics in a generic and regionally transferable manner. Therefore, one set of generic model parameters (especially the PFT traits) will be calibrated using observation data from multiple sites across Europe and model simulations for each site's specific conditions at once (Schmid 2022). However, since the species behind PFTs and their traits can differ across Europe, we also work on transfer functions that represent this flexibility by functional relations of certain trait values to environmental conditions (Rödig et al. 2017).

Acknowledgements

This study has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101057437 (BioDT project, <https://doi.org/10.3030/101057437>). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. We acknowledge the EuroHPC Joint Undertaking for awarding this project access to the EuroHPC supercomputer LUMI, hosted by CSC (Finland) and the LUMI consortium through a EuroHPC Development Access call.

Author contributions

Conceptualization: TB, FT. *Data Curation:* TB, FT, SV, CW. *Methodology:* TB, FT, JLG, TR, TM, THK, CW. *Software:* TR, TM, THK. *Supervision:* TB, FT. *Visualization:* TB, FT. *Writing - original draft:* TB, FT. *Writing - review & editing:* all co-authors.

Conflicts of interest

The authors have declared that no competing interests exist.

Disclaimer: This article is (co-)authored by any of the Editors-in-Chief, Managing Editors or their deputies in this journal.

References

- Berauer BJ, Wilfahrt PA, Reu B, Schuchardt MA, Garcia-Franco N, Zistl-Schlingmann M, Dannenmann M, Kiese R, Kühnel A, Jentsch A (2020) Predicting forage quality of species-rich pasture grasslands using vis-NIRS to reveal effects of management intensity and climate change. *Agriculture, Ecosystems & Environment* 296: 106929. <https://doi.org/10.1016/j.agee.2020.106929>
- Decourtye A, Mader E, Desneux N (2010) Landscape enhancement of floral resources for honey bees in agro-ecosystems. *Apidologie* 41: 264-277. <https://doi.org/10.1051/apido/2010024>
- De Koning K, Broekhuijsen J, Kühn I, Ovaskainen O, Taubert F, Endresen D, Schigel D, Grimm V (2023) Digital twins: dynamic model-data fusion for ecology. *Trends in Ecology & Evolution* S0169534723000903 <https://doi.org/10.1016/j.tree.2023.04.010>
- De Smedt K, Koureas D, Wittenburg P (2020) FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. *Publications* 8: 21. <https://doi.org/10.3390/publications8020021>
- European Commission, Eurostat, Cook E (2020) *Agriculture, forestry and fishery statistics – 2020 edition*. Publications Office of the European Union <https://doi.org/10.2785/143455>
- European Commission (2023) Destination Earth | Shaping Europe's digital future. <https://digital-strategy.ec.europa.eu/en/policies/destination-earth>. Accessed on: 2024-3-11.
- Evans E, Smart M, Cariveau D, Spivak M (2018) Wild, native bees and managed honey bees benefit from similar agricultural land uses. *Agriculture, Ecosystems & Environment* 268: 162-170. <https://doi.org/10.1016/j.agee.2018.09.014>
- Goble C, Soiland-Reyes S, Bacall F, Owen S, Williams A, Eguinoa I, Droesbeke B, Leo S, Pireddu L, Rodríguez-Navas L, Fernández J, Capella-Gutierrez S, Ménager H, Grüning B, Serrano-Solano B, Ewels P, Coppens F (2021) Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory. Zenodo <https://doi.org/10.5281/zenodo.4605654>
- Gustafson EJ (2013) When relationships estimated in the past cannot be used to predict the future: using mechanistic models to predict landscape ecological dynamics in a changing world. *Landscape Ecology* 28: 1429-1437. <https://doi.org/10.1007/s10980-013-9927-4>
- Kruse M, Stein-Bachinger K, Gottwald F, Schmidt E, Heinken T (2016) Influence of grassland management on the biodiversity of plants and butterflies on organic suckler cow farms. *Tuexenia* 36: 97-119. <https://doi.org/10.14471/2016.36.006>

- Lange M, Feilhauer H, Kühn I, Doktor D (2022) Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series. *Remote Sensing of Environment* 277: 112888. <https://doi.org/10.1016/j.rse.2022.112888>
- Muñoz Sabater J (2019) ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) <https://doi.org/10.24381/cds.e2161bac>
- Öster M, Cousins S, Eriksson O (2007) Size and heterogeneity rather than landscape context determine plant species richness in semi-natural grasslands. *Journal of Vegetation Science* 18: 859-868. <https://doi.org/10.1111/j.1654-1103.2007.tb02602.x>
- Piotrowska K, Connolly J, Finn J, Black A, Bolger T (2013) Evenness and plant species identity affect earthworm diversity and community structure in grassland soils. *Soil Biology and Biochemistry* 57: 713-719. <https://doi.org/10.1016/j.soilbio.2012.06.016>
- Poggio L, Sousa LM, Batjes NH, Heuvelink GB, Kempen B, Ribeiro E, Rossiter D (2021) SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL* 7: 217-240. <https://doi.org/10.5194/soil-7-217-2021>
- Rödiger E, Cuntz M, Heinke J, Rammig A, Huth A (2017) Spatial heterogeneity of biomass and forest structure of the Amazon rain forest: Linking remote sensing, forest modelling and field inventory. *Global Ecology and Biogeography* 26: 1292-1302. <https://doi.org/10.1111/geb.12639>
- Schmid JS (2022) Modeling species-rich ecosystems to understand community dynamics and structures emerging from individual plant interactions. PhD thesis. Universität Osnabrück <https://doi.org/10.48693/160>.
- Schwieder M, Wesemeyer M, Frantz D, Pfoch K, Erasmi S, Pickert J, Nendel C, Hostert P (2022) Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. *Remote Sensing of Environment* 269: 112795. <https://doi.org/10.1016/j.rse.2021.112795>
- Simons G, Koster R, Droogers P (2020) HiHydroSoil v2.0 - A high resolution soil map of global hydraulic properties. Wageningen, The Netherlands.
- Soiland-Reyes S, Sefton P, Crosas M, Castro L, Coppens F, Fernández J, Garijo D, Grüning B, La Rosa M, Leo S, Ó Carragáin E, Portier M, Trisovic A, RO-Crate Community, Groth P, Goble C (2022) Packaging research artefacts with RO-Crate. *Data Science* 5: 97-138. <https://doi.org/10.3233/DS-210053>
- Taubert F, Hetzer J, Schmid JS, Huth A (2020) The role of species traits for grassland productivity. *Ecosphere* 11: 03205. <https://doi.org/10.1002/ecs2.3205>
- Taubert F, Banitz T, Wohner C (2023) Guideline on how to provide data from a grassland site for simulations with the GRASSMIND model. <https://zenodo.org/records/10125790>. Accessed on: 2024-3-04.
- [Unknown] (2024a) Grassland Dynamics - Hochschwab (Austria) - 2001, 2008, 2015 (Data set). URL: <https://doi.org/10.23728/B2SHARE.A1C76E16D565459D935B0AE904C5F1DB>
- [Unknown] (2024b) Grassland Dynamics - Schrankogel (Austria) - 1994, 2004, 2014 (Data set). URL: <https://doi.org/10.23728/B2SHARE.57315B8581CF45F6A5686B8EC1E0A788>
- Wilkinson M, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018. <https://doi.org/10.1038/sdata.2016.18>

- Wohner C, Peterseil J, Poursanidis D, Kliment T, Wilson M, Mirtl M, Chrysoulakis N (2019) DEIMS-SDR - A web portal to document research sites and their associated data. *Ecological Informatics* 51: 15-24. <https://doi.org/10.1016/j.ecoinf.2019.01.005>
- Wohner C, Peterseil J, Klug H (2022) Designing and implementing a data model for describing environmental monitoring and research sites. *Ecological Informatics* 70: 101708. <https://doi.org/10.1016/j.ecoinf.2022.101708>

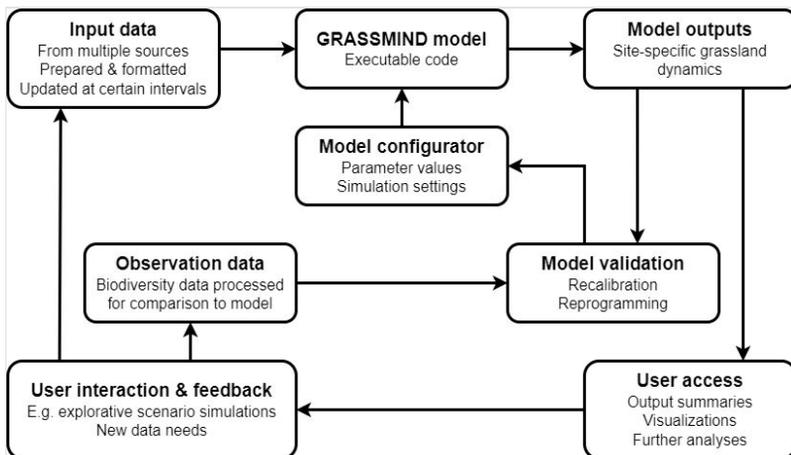


Figure 1.

Major components and steps of the pDT workflow. Arrows from one element to another show direct influence on that element .

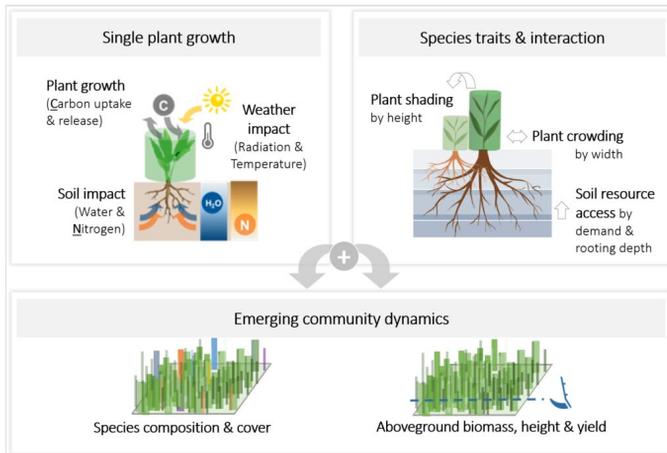


Figure 2.

Mechanistic simulation of grassland community dynamics with GRASSMIND. Growth and competition of single plant individuals, affected by traits and external drivers, lead to trajectories of community composition and properties.

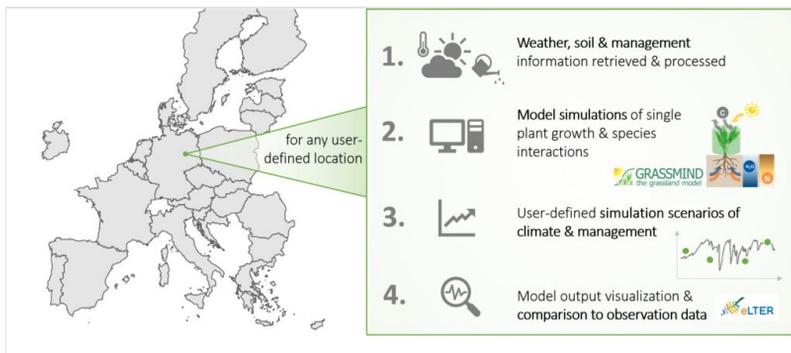


Figure 3.
pDT workflow elements in the end-user interface.

Table 1.

Data streams, variables and sources.

Data category	Streamed variables	Calculated variables	Temporal resolution
Input, weather	Precipitation, Air temperature (at 2m), Dewpoint temperature (at 2m), Surface solar radiation downwards, Surface net solar radiation, Soil heat flux density, Eastward wind component (at 10m), Northward wind component (at 10m) Surface pressure	Photosynthetically active radiation, Potential evapotranspiration	Daily (for full time period to be simulated)
Input, soil	Silt fraction, Clay fraction, Sand fraction	Mean over soil depth 0-200cm	None
Input, soil	Field capacity, Permanent wilting point, Soil porosity, Saturated hydraulic conductivity	Mapping from 6 SoilGrids depth layers to 20 GRASSMIND depth layers (both cover 0-200cm soil depth)	None
Input, management	Mowing events		Dates (2017-2021)
Input, management	Mowing events		Dates (2017-2022)
Observation, vegetation	Cover, Abundance, Biomass, Yield, Leaf area index	Mapping from species to plant functional types (if applicable)	Dates (1 to several time points)