Metadata form of Silva Fennica / DATA

This form is designed for writing the elements of metadata, which are used in the description of research materials such as data and codes. The form is based on the work done in the Work Group "Description of research materials" under the Finnish Open Science Coordination.

Item	Description	Responsible
Name of the code	The training data set to train the time series prediction models: **prebasdata_v4_0_subSet_1e_imp_setLabel.csv**: Climate data yearly aggregates: **WEATHER_3_1_hadgem3_gc31_ll_2020_2100_20240205_YMD_grSum_Y_norm.csv** Climate data monthly aggregates: **WEATHER_3_1_hadgem3_gc31_ll_2020_2100_20240205_YMD_grSum_M_norm.csv** Program parameter text file for defining the neural network model and the training process: **FC_RNN_GRU_GPU_0_HDBA_yyyymmdd_example.txt**	Author
Author & ORCID	Astola, Heikki ORCID ID: 0000-0003-4989-9910	Author
Authors' affiliation(s)	VTT technical research Centre of Finland, ROR: **https://ror.org/04b181w54**	Author
Owner of the material	VTT technical research Centre of Finland, ROR: **https://ror.org/04b181w54**	Author
Publisher	VTT technical research Centre of Finland, ROR: **https://ror.org/04b181w54**	Author
Funder	Research Council of Finland, ROR: **https://ror.org/05k73zm37** European Union (EU), ROR: **https://ror.org/019w4f821**	Author
Description	The training data set to train the time series prediction models: **prebasdata_v4_0_subSet_1e_imp_setLabel.csv**: This data set includes the input variables (except climate data) and targets for training the prediction models. The data has been arranged into a tabular form (CSV text file) with variable headers to enable the selection of different input and target variables, and for selecting desired subsets of the total data set (i.e. filtering). Each row of the data set corresponds to a single forest site (field plot). The original forest variable data for the field inventory plots was provided by Finnish Forest Centre (FFC).	Author

| | Climate data yearly or monthly aggregates: **WEATHER_3_1_hadgem3_gc31_ll_2020_2100_20240205_YMD_grSum_[Y,M]_ norm.csv** The climate data set includes the CMIP6 climate projections data downloaded from Copernicus Data Store (CDS) for period 2020 – 2100. The downloaded data included climate model hadgem3_gc31_ll daily data sets for near surface air temperature, near surface specific humidity and precipitation, and the monthly photosynthetically active radiation (PAR) data set for three climate scenarios4: ssp1_2_6, ssp2_4_5, and ssp5_8_5. The yearly CO_2 concentration (CMIP6 Data) was downloaded using the link given at https://gmd.copernicus.org/articles/10/2057/2017/, but the data server link to these data seems to be broken (4/2025).

The data has been arranged into a tabular form (CSV text file). The file with the identifier 'Y' in the name contains the yearly aggregates (sum, mean & standard deviation) of the climate & CO_2 data, and the file with identifier 'M' contains the monthly aggregates. In addition, three metadata columns indicating the data year (YEAR), the climate scenario (scenario), and the climate data zone (climID) are included.

Links:
Copernicus Data Store : https://cds.climate.copernicus.eu/#!/home
FFC field inventory data: https://www.metsakeskus.fi/fi/avoin-metsa-ja-luontotieto/aineistot-paikkatieto-ohjelmille/paikkatietoaineistot | |
| *Methods* | We used the Rprebasso (Mäkelä, 1997; Minunno et al., 2016, 2019) software available on GitHub to produce yearly forest structure and carbon balance variable predictions. These predictions were used as reference data (or targets) to train the deep neural network models. The time range for the variable prediction was set to 25 years, considered adequate for demonstrating the time series prediction.

The Rprebasso program was used to produce 25-year predictions for the field inventory plots for a set of selected forest variables: tree height, stem diameter, basal area, stem volume, and a subset of carbon balance variables: net primary production (NPP), gross primary production per tree layer (GPP), net ecosystem exchange (NEP) and gross growth (GGR). The variables for different tree species were considered separately, such as the tree height for pine (H_{pine}), spruce (H_{spr}), and broadleaved trees (H_{bl}).

Mäkelä, A. (1997). A Carbon Balance Model of Growth and Self-Pruning in Trees Based on Structural Relationships. *Forest Science*, *43*(1), 7–24. https://doi.org/10.1093/forestscience/43.1.7

Minunno, F., Peltoniemi, M., Launiainen, S., Aurela, M., Lindroth, A., Lohila, A., Mammarella, I., Minkkinen, K., & Mäkelä, A. (2016). Calibration and validation of a semi-empirical flux ecosystem model for coniferous forests in the Boreal region. *Ecological Modelling*, *341*, 37–52. https://doi.org/10.1016/j.ecolmodel.2016.09.020 | Author |

Data set: prebasdata_v4_0_subSet_1e_imp_setLabel.csv

	Column header	Unit	Description	
Variables	siteID		Site/field plot identification number	Author
	age_pine	y	tree mean age - pine (initial condition from field plot data)	
	age_spr	y	tree mean age - spruce (initial condition from field plot data)	

	age_bl	y	tree mean age - broadleaved (initial condition from field plot data)	
	H_pine	m	tree mean height - pine (initial condition from field plot data)	
	H_spr	m	tree mean height - spruce (initial condition from field plot data)	
	H_bl	m	tree mean height - broadleaved (initial condition from field plot data)	
	D_pine	cm	tree mean diameter - pine (initial condition from field plot data)	
	D_spr	cm	tree mean diameter - spruce (initial condition from field plot data)	
	D_bl	cm	tree mean diameter - broadleaved (initial condition from field plot data)	
	BA_pine	m2/ha	basal area - pine (initial condition from field plot data)	
	BA_spr	m2/ha	basal area - spruce (initial condition from field plot data)	
	BA_bl	m2/ha	basal area - broadleaved (initial condition from field plot data)	
	climID		climate data identification number (shuffled data)	
	siteType		Site fertility class	
	SWinit		Initial soil water	
	CWinit		Initial crown water	
	SOGinit		Initial snow on ground	
	Sinit		Initial temperature acclimation state	
	nLayers		Number of layers	
	nSpecies		Number of species	
	soildepth		Soil depth	
	effective field capacity		Effective field capacity	
	permanent wilting point		Permanent wilting point	
	climID_orig		climate data identification number (original data)	
	year_start		climate data/prediction starting year	
	year_end		climate data/prediction end year	
	model		Climate data model (hadgem3_gc31_ll)	
	scenario		Climate data scenario identifier (126 = ssp1_2_6; 245 = ssp2_4_5; 585 = ssp5_8_5)	
	runID		Identifier for forestry district. May be used to select different sized data sets.	
	runID_i		Identifier for data shuffling/forestry district (not used)	
	exst_pine		Boolean indicating if this species is present in the site (0 = not present, 1 = present)	
	exst_spr		Boolean indicating if this species is present in the site (0 = not present, 1 = present)	
	exst_bl		Boolean indicating if this species is present in the site (0 = not present, 1 = present)	
	H_pine_err		Indicator of constant growth of height/pine.	
	H_spr_err		Indicator of constant growth of height/spruce.	
	H_bl_err		Indicator of constant growth of height/BL.	
	setLabel		A label identifying the training, validation and test sets (1, 2, 3 respectively)	

The rest of the data set columns include the yearly predictions (Rprebasso) for the following forest and carbon balance variables. All variables include the yearly prediction for 25 years (indicated with suffix '_y', with y = year (decimal number) in the column header), and for tree species categories: pine, spruce (spr), and broadleaved (BL). E.g. the variable header H_pine_25.0 = tree height/pine prediction for the 25. year from 'year_start'.

The variables with 25-year predictions:

Variable	Unit	Description
H	m	Tree mean height
D	cm	Tree mean diameter
BA	m2/ha	Basal area
V	m3/ha	Stem volume
npp	gC m-2 y-1	Net primary production
GGR	m3 ha-1 y-1	Gross growth
GPPtrees	gC m-2 y-1	Gross primary production (tree layer) - abreviated as GPP in the manuscr
NEP	gC m-2 y-1	Net ecosystem exchange
ET_preles	mm y-1	Annual evapotranspiration (not used in the study)

Data set:
**WEATHER_3_1_hadgem3_gc31_ll_2020_2100_20240205_YMD_grSum_Y_norm
.csv**

Column header	Unit	Description
climID		Climate zone identifier number
YEAR		The year of the aggregated data (2020 - 2100)
PAR_sum	mmol/m2	Yearly sum of photosynthetically active radiation
TAir_sum	degC	Yearly sum of air temperature
Precip_sum	mm	Yearly sum of precipitation
VPD_sum	kPa	Yearly sum of vapor pressure deficit
CO2_sum	ppm	Yearly sum of CO2 concentration
PAR_mean	mmol/m2	Yearly mean of photosynthetically active radiation
TAir_mean	degC	Yearly mean of air temperature
Precip_mean	mm	Yearly mean of precipitation
VPD_mean	kPa	Yearly mean of vapor pressure deficit
CO2_mean	ppm	Yearly mean of CO2 concentration
PAR_std	mmol/m2	Yearly standard deviation of photosynthetically active
TAir_std	degC	Yearly standard deviation of air temperature
Precip_std	mm	Yearly standard deviation of precipitation
VPD_std	kPa	Yearly standard deviation of vapor pressure deficit
CO2_std	ppm	Yearly standard deviation of CO2 concentration
scenario		Climate data scenario ID (126 = ssp1_2_6; 245 = ssp2 585 = ssp5_8_5)

Data set:
**WEATHER_3_1_hadgem3_gc31_ll_2020_2100_20240205_YMD_grSum_M_nor
m.csv**

Column header	Unit	Description
climID		Climate zone identifier number
YEAR		The year of the aggregated data (2020 - 2100)
MONTH		Month (1–12)
PAR_sum	mmol/m2	Monthly sum of photosynthetically active radiation
TAir_sum	degC	Monthly sum of air temperature
Precip_sum	mm	Monthly sum of precipitation
VPD_sum	kPa	Monthly sum of vapor pressure deficit
CO2_sum	ppm	Monthly sum of CO2 concentration
PAR_mean	mmol/m2	Monthly mean of photosynthetically active radiation
TAir_mean	degC	Monthly mean of air temperature
Precip_mean	mm	Monthly mean of precipitation
VPD_mean	kPa	Monthly mean of vapor pressure deficit
CO2_mean	ppm	Monthly mean of CO2 concentration
PAR_std	mmol/m2	Monthly standard deviation of photosynthetically active
TAir_std	degC	Monthly standard deviation of air temperature
Precip_std	mm	Monthly standard deviation of precipitation
VPD_std	kPa	Monthly standard deviation of vapor pressure deficit
CO2_std	ppm	Monthly standard deviation of CO2 concentration
scenario		Climate data scenario ID (126 = ssp1_2_6; 245 = ssp2 585 = ssp5_8_5)

The program parameter file example will be available in the code repository along the
publication of the manuscript.
(FC_RNN_GRU_GPU_0_HDBA_yyyymmdd_example.txt)

Author keywords	Copernicus Climate data Store (CDS); The Finnish Forest Centre	Author
Vocabulary keywords (community standard)	NA	Author
Discipline	Field(s) of study to which the material is related. This is generally given by the repository as they use specific classifications.	Archive/Rep ository/Publi sher
Type of material	Research data	Author
Language	GBR	Author
Time range covered	2020 – 2100	Author
Geographic region	The training data set **prebasdata_v4_0_subSet_1e_imp_setLabel.csv** covers the area of continental Finland (ISO ISO 3166 Alpha-3 code/Numeric: **FIN/246**) The climate/weather data covers the region indicated by the ESRI Shapefile: **WEATHER_3_1_ssp1_2_6_INT16.shp.** The shapefile field 'gridcode' corresponds to the variable 'climID' in the other data sets.	Author
Version	NA	Author
File format(s)	Data: text formats (*.csv, *.txt)	Author
Availability of the materials (open, embargo, registration, limited, registration required)	Open	Author
Justification for access restrictions	NA	Author
Licence	**CC BY 4.0**	Author
Connections with other research materials	1) The material is derived from another material e.g., research data is derived from raw data (IsBasedOn)	Author

Access to the connected research materials	**Field inventory data: The Finnish Forest Centre** https://www.metsakeskus.fi/fi/avoin-metsa-ja-luontotieto/aineistot-paikkatieto-ohjelmille/paikkatietoaineistot **Climate data: Copernicus Data Store** https://cds.climate.copernicus.eu/#!/home	Author
Codes only: hardware/software requirements for running the code	NA	Author
Connections to other products of research	NA	Author
Personal data	NA	Author
Confidential or secret data	No	Author
Publication date	08.01.2026	Archive/Repository/Publisher
Preservation policy	Decision and plan for preserving the material permanently or for a specific time. Justify why this decision was made. Please note that the author is responsible on removing the material after the predefined preservation time.	Author
Permanent identifier (PID)	10.5281/zenodo.18186624	Archive/Repository/Publisher