

APPLICATION PAPER  

# Identifying compound weather drivers of forest biomass loss with generative deep learning

Mohit Anand<sup>1,2</sup> , Friedrich J. Bohn<sup>3</sup>, Gustau Camps-Valls<sup>4</sup>, Rico Fischer<sup>5</sup>, Andreas Huth<sup>5</sup>, Lily-belle Sweet<sup>1</sup> and Jakob Zscheischler<sup>1,2</sup>

<sup>1</sup>Helmholtz Centre for Environmental Research—UFZ, Department of Compound Environmental Risks, Leipzig, Germany

<sup>2</sup>Technische Universität Dresden, Dresden, Germany

<sup>3</sup>Helmholtz Centre for Environmental Research—UFZ, Department of Computational Hydrosystems, Leipzig, Germany

<sup>4</sup>Image Processing Laboratory (IPL), Universitat de València, València, Spain

<sup>5</sup>Helmholtz Centre for Environmental Research—UFZ, Department of Ecological Modelling, Leipzig, Germany

**Corresponding author:** Mohit Anand; Email: [mohit.anand@ufz.de](mailto:mohit.anand@ufz.de)

**Received:** 10 March 2023; **Revised:** 22 November 2023; **Accepted:** 11 January 2024



**Keywords:** compound events; extreme events; forest mortality; generative deep learning; variational autoencoder

## Abstract

Globally, forests are net carbon sinks that partly mitigate anthropogenic climate change. However, there is evidence of increasing weather-induced tree mortality, which needs to be better understood to improve forest management under future climate conditions. Disentangling drivers of tree mortality is challenging because of their interacting behavior over multiple temporal scales. In this study, we take a data-driven approach to the problem. We generate hourly temperate weather data using a stochastic weather generator to simulate 160,000 years of beech, pine, and spruce forest dynamics with a forest gap model. These data are used to train a generative deep learning model (a modified variational autoencoder) to learn representations of three-year-long monthly weather conditions (precipitation, temperature, and solar radiation) in an unsupervised way. We then associate these weather representations with years of high biomass loss in the forests and derive weather prototypes associated with such years. The identified prototype weather conditions are associated with 5–22% higher median biomass loss compared to the median of all samples, depending on the forest type and the prototype. When prototype weather conditions co-occur, these numbers increase to 10–25%. Our research illustrates how generative deep learning can discover compounding weather patterns associated with extreme impacts.

## Impact Statement

Tree mortality is a complex phenomenon involving multiple processes at different temporal scales. Here, we rely on very long simulations of a forest model and develop a method based on generative deep learning to find the relationship between complex weather patterns and tree mortality. The generative nature of the method allows for the generation of new realistic weather conditions outside of the provided samples, which are associated with high biomass loss in a forest. Furthermore, the method can be applied to different weather-driven impacts, adding to the growing pool of explainable/interpretable data-driven methods for Earth system sciences.

  This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

## 1. Introduction

Forests are the predominant atmospheric carbon sink on land that play a crucial role in mitigating anthropogenic climate change. Forests sequestered 7.6 GtCO<sub>2</sub>e/yr on average between 2000 and 2019 (Harris et al., 2021), which amounted to 22% of the total fossil CO<sub>2</sub> emissions during 2020 (Friedlingstein et al., 2022). Additionally, forests regulate microclimates, sustain biodiversity, and provide ecosystem services such as the production of timber, edible wild plants and fungi, pest control, and increasing nitrogen availability (Felipe-Lucia et al., 2018). However, sudden and unexpected weather-related forest mortality events have been observed in recent years across the globe, including in ecosystems that were not previously considered at risk (Hartmann et al., 2022). In the European Union, the net carbon sink from forest land decreased by 12% between 2010 and 2018 (Pilli et al., 2022). Weather conditions often interact with disturbances such as insect outbreaks, diseases and forest fires to cause forest mortality (Sturrock et al., 2011; Millar and Stephenson, 2015; Jain et al., 2022). With weather extremes expected to increase in intensity and frequency in the future (Seneviratne et al., 2021), an improved understanding of the weather drivers of forest mortality is necessary for the development of effective forest conservation and management strategies.

Understanding the influence of weather drivers on large-scale tree mortality is challenging because forest mortality depends on complex interactions between weather events, the presence of pests and diseases, and forest characteristics such as canopy height, tree cover, biomass, and diversity (Bastos et al., 2023). The forest structure and diversity determine forest resistance to climatic or biotic stressors (for example, a forest with less biomass requires less water for its growth), and this structure, in turn, depends on previous weather conditions (López-Serrano et al., 2005; Bequet et al., 2011; Ježik et al., 2011; Rais et al., 2021; Mahecha et al., 2022; Li et al., 2023). This means that present-day tree mortality can be influenced by weather conditions from previous years. Extreme climate events such as droughts and heatwaves are common drivers of tree mortality (Allen et al., 2010; Anderegg et al., 2013; McDowell et al., 2013; Park Williams et al., 2013; Choat et al., 2018; Senf et al., 2018; Brodrigg et al., 2020). However, due to the complex interactions between climatic, biotic, and forest state factors, forest mortality events can also be driven by the temporally compounding effects of multiple weather events, such as observed in sequential droughts (Zscheischler et al., 2018, 2020; Sánchez-Pinillos et al., 2022; Bastos et al., 2023).

Modeling forests with process-based models can aid in improving the scientific understanding of forest dynamics. There is a long tradition in ecology of using individual-based forest models to answer a wide range of scientific questions (Shugart et al., 2018), which simulate each tree's growth within a forest stand. These forest gap models were first developed for North America and have become one of the most widely used model types in ecology (Bugmann, 2001). Various process-based models are currently employed, each with different strengths and weaknesses (Mahnken et al., 2022).

Due to the complexity of the simulated processes and their convolution through time, it is challenging to disentangle weather conditions associated with high forest mortality, even in a model environment. Hence, a better understanding of such processes is needed (Bugmann et al., 2019). Currently, two widely used approaches exist to identify the drivers of tree mortality in forests. First, one can gain insight through forward simulations from models (Taccoen et al., 2019; Scheel et al., 2022). Forcing the model with perturbed weather data and analyzing the output in a conventional sensitivity analysis setting has two main challenges. Due to the long-range temporal dependency of weather variables, the forcing data have many input features. The large number of features translates into an even larger number of combinations of weather, making a comprehensive analysis computationally infeasible. Furthermore, defining ways to modify multivariate weather data so that the modified data are physically realistic is not trivial due to structured correlation within the variables. Nevertheless, forward simulations are helpful in analyzing individual events. Second, statistical methods such as linear regression, generalized additive models, principal component analysis, and random forests can be used to understand the impact of weather drivers (Park Williams et al., 2013; Matusick et al., 2018; Xu et al., 2018; Crouchet et al., 2019; Senf et al., 2020; Forzieri et al., 2022; Jaime et al., 2022; Le Grix et al., 2023). Linear methods typically fail to capture

complex nonlinear forest dynamics, and conventional machine learning methods like random forests or support vector machines require predefining predictors from an often extensive set of potential candidates.

Deep learning (DL) methods can capture the complex temporal dynamics in high-dimensional data and have been extensively used for time series forecasting and classification problems (Wang et al., 2017; Zhao et al., 2017; Wu et al., 2020; Lim and Zohren, 2021; Marcolongo et al., 2022). Despite high predictive accuracy, DL methods are often criticized for their lack of explainability. Most current weather-induced impact research utilizes post hoc methods, which interpret an existing, trained model (for example SHapley Additive exPlanations (SHAP) values, Layer-wise Relevance Propagation (LRP), feature importance analysis, and regression activation maps (Wolanin et al., 2020; Labe and Barnes, 2021; Yan et al., 2022; Liu et al., 2023; Mamalakis et al., 2023; Sweet et al., 2023; Wadoux et al., 2023). The advantage of post hoc methods is that they can be used with any model architecture and, therefore, do not require a trade-off of predictive model performance. However, such approaches explain model decisions on specified data points and do not provide a holistic view of the entire decision function (Bordt et al., 2022). An alternative approach is to use intrinsically explainable machine learning models (Rudin, 2019). A method in this approach is to learn a lower-dimensional representation of the input and relate it to the output with a more explainable function (Alvarez-Melis and Jaakkola, 2018; Kim et al., 2018; Chen et al., 2019; Varando et al., 2021; Yang et al., 2021; Gautam et al., 2022). A generative DL architecture like the variational autoencoder (VAE) is one way to learn a lower-dimensional explainable data representation. VAEs learn a lower-dimensional representation (latent variables) of a very high-dimensional dataset by minimizing a reconstruction error and constraining the latent distribution.

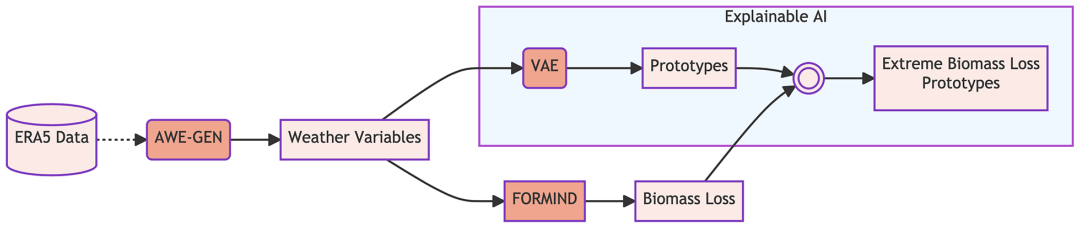
In this study, we develop a VAE-based methodology to generate lower-dimensional representations of complex and correlated weather conditions. We apply and evaluate its performance on simulated data as a proof-of-concept, making use of long simulations of multiple forest types driven by synthetic weather data from a weather generator. By using modeled data, we are able to compare our results to the known processes of the model and thereby validate our novel, data-driven approach. With the help of the developed method, we can search for representations (weather prototypes) that are associated with high biomass loss events and thereby identify compound weather drivers of tree mortality. We also explore the relationship between compounding weather prototypes and biomass loss. Our developed methodology could, in future, be applied to real-world data (such as inventory datasets from national forest inventories), for which no ground truth is typically available for method validation.

## 2. Data and methods

An overview of the methodology is shown in [Figure 1](#). We use simulated weather from a weather generator to run a parameterized forest model for the Hainich beech forest in Germany at 51.08° N and 10.43° E. We use composites and logistic regression analysis to obtain a general overview of the weather and biomass loss relationships. Next, a VAE is used to learn a low-dimensional representation of weather conditions. These representations are then used to identify weather prototypes associated with high levels of biomass loss based on the forest simulation for three different types of forests, namely beech, pine, and spruce.

### 2.1. Weather simulation

AWE-GEN is an advanced stochastic weather generator that can generate realistic high-resolution (hourly) weather variables (Fatichi et al., 2011). It can reproduce low- and high-frequency characteristics of weather variables, such as the interannual variability of precipitation and temperature, and has been widely used in hydrology and ecological modeling (Mastrotheodoros et al., 2020; Meili et al., 2020; Hirschberg et al., 2021; Marcolongo et al., 2022). For this study, AWE-GEN is calibrated on 40 years (1979–2019) of bias-adjusted ERA5 reanalysis data designed explicitly for impact studies (Cucchi et al., 2020) for the location of the Hainich forest. AWE-GEN requires precipitation, cloud cover, air



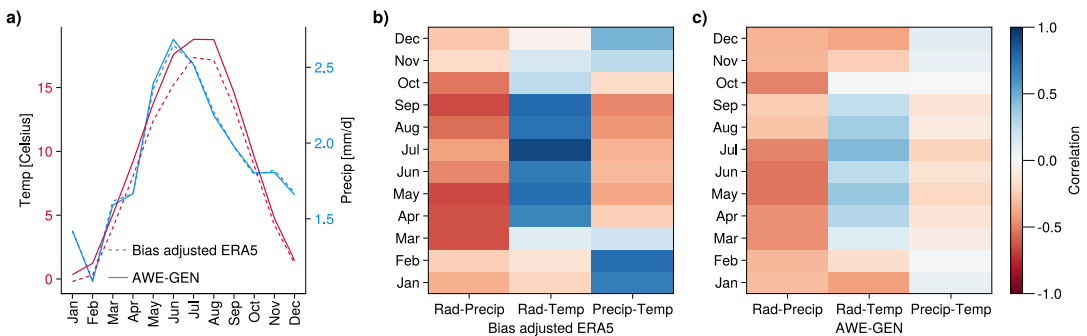
**Figure 1.** Illustration of the methodology. ERA5 data are used to calibrate AWE-GEN. AWE-GEN is used to generate realistic weather conditions. A VAE is used to learn low-dimensional weather representations and generate prototypes. Based on simulated weather conditions, FORMIND simulates forest dynamics, including mortality and biomass loss. FORMIND-derived biomass loss is used to select weather prototypes associated with high biomass loss. The explainable AI method used here consists of the training of VAE together with the generation of weather prototypes relevant for high biomass loss induced by forest mortality.

temperature, shortwave radiation, relative humidity, wind speed, and atmospheric pressure at hourly time step for calibration. For the cloud cover, a single level of hourly ERA5 data is used (Hersbach et al., 2020). We compute relative humidity and specific humidity from pressure and temperature. Ultimately, 200,000 years of hourly stationary weather is generated after calibration.

The simulation’s mean annual temperature and precipitation are 8.5 C and 693 mm, respectively. The monthly seasonal cycle of temperature and precipitation is well captured by AWE-GEN, with slightly too warm temperatures in the weather generator (Figure 2a). Reanalysis data show strong negative correlations between monthly radiation and precipitation throughout the year and positive correlations between radiation and temperature during the summer (Figure 2b). The weather generator generally captures these correlations, albeit with weaker magnitude (Figure 2c).

**2.2. Simulation of forest dynamics**

We simulate forest dynamics with the forest gap model FORMIND, a process-based, individual-oriented, grid-based forest growth model (Fischer et al., 2015, 2016). It describes forest succession in small-scale forest patches (20 m × 20 m). The main processes considered are tree growth, mortality, recruitment, and competition; the trees within a forest patch compete for space, light, and water.



**Figure 2.** (a) Monthly mean temperature (red) and precipitation (blue) of the bias-adjusted ERA5 (dotted line) and simulated by AWE-GEN (continuous line). Correlation between monthly means across all three variable combinations for all months for the bias-adjusted ERA5 data (b) and data simulated by AWE-GEN (c).

Carbon balance-based ecophysiological processes, such as photosynthesis, respiration, and carbon allocation, are used to calculate individual tree growth. Photosynthesis is related to temperature with a bell-shaped curve, whereas water shortage linearly reduces photosynthesis. In addition, deciduous species only sequester CO<sub>2</sub> during the temperature-dependent growing season. Maintenance respiration is also modified by temperature following the well-established  $Q_{10}$  approach (full model description in Bohn et al., 2014).

In the FORMIND model, mortality processes occur due to competition for space and reduced growth rate. Space-competition mortality occurs under optimal growing conditions when crowns occupy more space than is available. The model then randomly removes trees until the remaining crowns have sufficient space. In addition, temperature also affects Net Primary Productivity (NPP) through maintenance respiration. NPP is used to calculate the diameter increment of each tree. The increment in diameter affects the biomass loss as there is stochastic diameter increment-dependent mortality included in FORMIND (Bohn et al., 2014, [www.formind.org](http://www.formind.org)). FORMIND has been frequently used to simulate temperate forest dynamics (Bohn et al., 2014; Rödig et al., 2017; Bohn et al., 2018; Bugmann et al., 2019; Holtmann et al., 2021).

Here, we conduct three simulations, each time simulating only beech, pine, or spruce trees. Beech trees are late-successional, shade-tolerant, and deciduous, whereas pine is a fast-growing, light-demanding, evergreen needle-leaf tree. Spruce is a moderately fast-growing, shade-tolerant, evergreen, needle-leaf tree. The 200,000 years of AWE-GEN-generated precipitation and temperature data, together with solar irradiance (computed as a fraction of shortwave radiation), are used as an input to FORMIND. The specific parameters of the model are derived from various sources. For instance, we use yield tables to fit allometric and growth functions. Other parameters are obtained by taking the average over several field measurements, like wood density or leaf area index. For more details, see Bohn et al. (2014). The simulations are conducted for 200 hectares of land to reduce the internal stochasticity of the model. The forest model starts simulating the growth of trees from barren land and thus requires a spin-up period of roughly 2,000 years to reach a quasi-equilibrium. The simulations are conducted in chunks of 10,000 years for computational efficiency. Therefore, we do not analyze the first 2,000 years for each of these parts and are finally left with 160,000 (8,000 × 20) years of forest data. Only the 160,000 years of weather is then used for further analysis. The simulations, even though potentially strongly differing from real-world behavior, have enough complexity to validate our proof-of-concept.

FORMIND simulates multiple variables at a daily and yearly time scale. For this analysis, we use biomass loss (BL) on a yearly scale. Instead of the fraction of trees dying yearly, BL refers to the fraction of biomass lost yearly. For example, the BL may be low when some young (therefore small) trees die but high when the same number of old (therefore giant) trees die. BL combines different types of mortality, namely background mortality, which depends on the genus; size-dependent mortality, which accounts for the mortality of young trees; and diameter increment-dependent mortality, which accounts for the fact that older trees have a higher mortality rate. We also consider five forest structure variables at a yearly time scale: age, stem volume, cumulative leaf area index, height, and diameter of each tree. The information on the state variables is condensed as histograms with 51 bins.

As this study focuses on extreme biomass loss, we convert BL to a binary value by thresholding at the 90th percentile. To understand extreme BL ( $y_t$ ), the weather conditions in the current year and the two previous years ( $\mathbf{x}_{t-2}^d, \mathbf{x}_{t-1}^d, \mathbf{x}_t^d$ ) are considered in the analysis. By definition,  $\mathbf{x}_t^d$  has a dimension of  $12 \times 3$  for 12 months and three variables (temperature, precipitation, and radiation). The forest structure variables at the end of the third year before the considered BL year ( $\mathbf{x}_{t-3}^s$ ) are also used in some of our analyses.  $\mathbf{x}_t^s$  thus refers to the five forest structure variables described by 51 binned distributions each. When the subscript is not specified,  $\mathbf{x}^d$  ( $36 \times 3$ ) refers to  $[\mathbf{x}_{t-2}^d, \mathbf{x}_{t-1}^d, \mathbf{x}_t^d]$ ,  $\mathbf{x}^s$  ( $51 \times 5$ ) refers to  $\mathbf{x}_{t-3}^s$  and  $y$  ( $1 \times 1$ ) refers to  $y_t$ .

### 2.3. Composites and logistic regression

Composites are the difference of the mean weather for years with extremely high BL with respect to the mean weather conditions for all the years. We first construct three years of extreme BL weather

composites ( $BL \geq 90$ th percentile) to identify general weather conditions associated with high BL. We also train a logistic regression model, which can be used to identify meteorological drivers of large impacts (Bevacqua et al., 2021; Vogel et al., 2021):

$$\mathbb{P}(BL \geq 90\text{th percentile}) = \frac{1}{1 + e^{-(\mathbf{w}\mathbf{x}^T + b)}}$$

where  $\mathbb{P}(BL \geq 90\text{th percentile})$  is the probability of extremely high BL and  $\mathbf{x} = [\text{vectorize}(\mathbf{x}^d)]$  or  $\mathbf{x} = [\text{vectorize}(\mathbf{x}^d), \text{vectorize}(\mathbf{x}^s)]$ . The logistic regression model is used to understand the relationship between three years of monthly weather ( $\mathbf{x}^d$ ) and BL  $y$  at the end of the third year. In addition to dynamic weather conditions ( $\mathbf{x}^d$ ), the BL also depends on the forest structure. Therefore, we also train a logistic regression model with the forest structure variables ( $\mathbf{x}^s$ ) to quantify their importance. We assess the goodness of fit of the regression models using the Critical Success Index (CSI, Donaldson et al., 1975),  $F_1$ -Score and Average Precision (with definitions provided in Appendix B).

#### 2.4. Variational autoencoder as a generative model

A variational autoencoder (VAE) is a form of DL model that can extract nonlinear feature representations from high-dimensional data, which has been widely used in Earth sciences, weather applications and remote sensing (Camps-Valls et al., 2021; Tibau et al., 2021). A VAE is a generative model that can learn low-dimensional representations of complex data in an unsupervised way. It comprises an encoder, with the parameters  $\phi$ , that estimates the latent representation and a decoder part, with parameters  $\theta$ , that reconstructs the input samples from the latent dimensions. Overall, the aim is to estimate the log-likelihood of some observed data  $\mathbf{x}$ .

This can be done by using variational Bayesian inference. We are interested in the posterior distribution  $p(\mathbf{z}|\mathbf{x})$ , where  $\mathbf{x}$  are our observations and  $\mathbf{z}$  are latent variables. However, this posterior is often intractable because we cannot compute the evidence or denominator of Bayes' theorem,  $p(\mathbf{x})$ , because the latent variables,  $\mathbf{z}$ , must be marginalized out. The central concept of variational inference (VI) is to use optimization to find a more tractable distribution  $q(\mathbf{z})$  from a family of distributions  $\mathcal{Q}$  such that it is close to the desired posterior distribution  $p(\mathbf{z}|\mathbf{x})$ . In VI, proximity is defined using the Kullback–Leibler Divergence ( $D_{KL}[\cdot|\cdot]$ ), so the goal is to retrieve an optimal  $q(\mathbf{z})$  that minimizes  $D_{KL}[q(\mathbf{z})||p(\mathbf{z}|\mathbf{x})]$ , which can be interpreted as minimizing the relative entropy between the two distributions. Mathematically, this is equivalent to maximizing the evidence lower bound (ELBO) (also sometimes called the variational lower bound or the negative variational free energy) (Kingma and Welling, 2013):

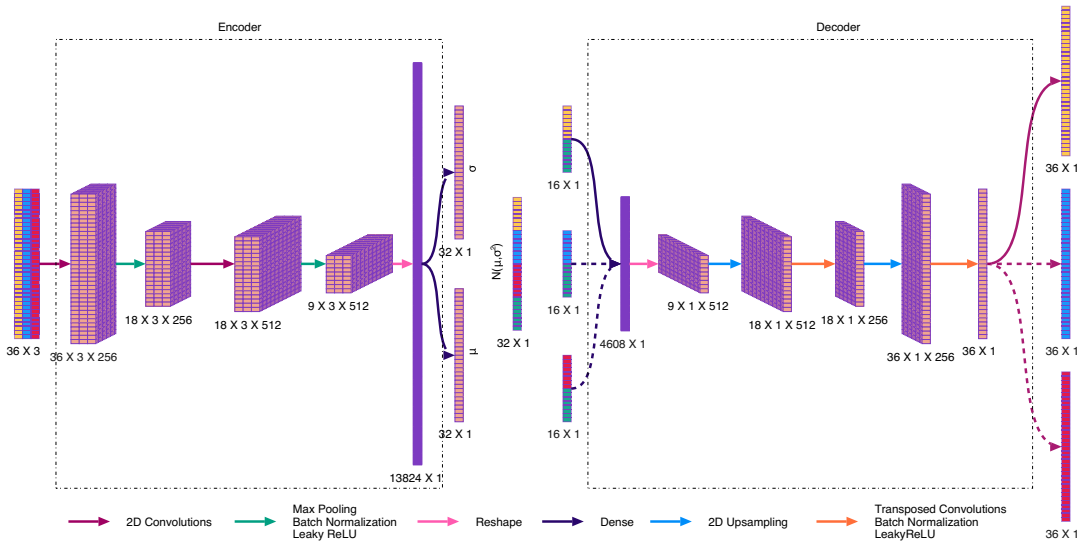
$$\log p_{\theta}(\mathbf{x}|\mathbf{z}) \geq -D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x})||p(\mathbf{z})) + \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}[\log p_{\theta}(\mathbf{x}|\mathbf{z})] = \text{ELBO},$$

where  $\mathbf{x}$  and  $\mathbf{z}$  represent the input data and latent variables respectively. The ELBO has two components. The KL-Divergence term enforces how close the latent dimensions  $\mathbf{z}$  distribution should be to a specified prior distribution  $p_{\phi}(\mathbf{z})$ , while the other term aims to minimize reconstruction error,  $\log p_{\theta}(\mathbf{x}|\mathbf{z})$ . In the case of  $\beta$ -VAE, the hyperparameter  $\beta$  weighs the KL-Divergence term of the loss function:

$$\mathbb{L}_{\text{VAE}}(\theta, \phi) = \beta D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x})||p(\mathbf{z})) - \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})}[\log p_{\theta}(\mathbf{x}|\mathbf{z})]. \quad (1)$$

Therefore, a value of  $\beta > 1$  pushes the model to learn a latent distribution closer to the specified distribution at the cost of a higher reconstruction error (Higgins et al., 2017) and vice versa for  $\beta < 1$ . Different latent distributions can be learned through VAE (Bond-Taylor et al., 2022), and here we choose a normally distributed prior with a diagonal covariance matrix with zero mean and unit variance. This particular choice of prior helps compute the KL-Divergence analytically. The input data ( $\mathbf{x} = \mathbf{x}^d$ ) consists of three weather variables, each a monthly time series of three years (Section 2.2).

We have modified the standard  $\beta$ -VAE by making a shared decoder for each weather variable (Figure 3). This is because each weather variable is a time series, and we expect the decoder to learn the general features of time series data. The encoder outputs two 32-dimensional vectors, one for the mean



**Figure 3.** The VAE architecture with encoder and decoder blocks. The encoder takes in  $36 \times 3$  dimensional weather variables and outputs two 32-dimensional vectors of mean and log variance of a normal distribution, which define the latent space. The decoder takes samples in the latent variable space associated with specific weather variables in the input space (distinguished by different colors) to generate reconstructions of each of the individual weather variables. Three passes of the decoder are required to reconstruct all three weather variables.

and the other for the log variance. Then, 32-dimensional latent values (following a multivariate normal distribution) can be sampled from these parameters. The decoder always takes in 16-dimensional latent variables to reconstruct one of the weather variables. The decoder takes eight weather-specific and eight shared dimensions and returns one reconstructed weather variable. This is done for each weather variable, and to obtain the full reconstructed output, three passes of the decoder are required. The eight shared dimensions capture the correlation between the weather variables (Figure 2b). As a design choice, the first eight dimensions are selected to represent radiation, the next eight for temperature, and the last eight are shared between all variables.

The encoders and decoders can have different architectures. In the encoder, a convolution layer is used to learn the different patterns in the data, increasing the number of dimensions, followed by max-pooling (Scherer et al., 2010), which reduces the number of dimensions. Batch normalization (Ioffe and Szegedy, 2015) is used, followed by LeakyReLU (Maas et al., 2013) as the nonlinear function. This step is repeated twice for 256 and 512 feature kernels. After reshaping the output, two dense layers generate the mean and the log variance vector. Samples from the latent dimensions are then fed to the decoder. First, a dense layer increases the dimensions to reshape it into  $9 \times 1 \times 32$ . This is followed by upsampling, transposed convolution (Zeiler et al., 2010), and batch normalization with Leaky-ReLU as an activation function. Like in the encoder, this is repeated twice, leading to an output of dimensions  $36 \times 1 \times 1$ . Finally, a convolutional layer is used without any activation function to get the output (see Figure 3 for more details).

The entire network is trained through backpropagation (Rumelhart et al., 1986). The inputs are normalized based on the statistics of the training set. Adam optimizer is used with a learning rate of 0.0001, the upsampling and downsampling size is (2, 1), and the kernel size is (3, 1) with stride as 1. LeakyReLU is used with the value  $\alpha = 0.3$ . The train, validation, and test split is 80%, 10%, and 10%, respectively. The total sample size available for training is 126,976. The model is trained for 150 epochs with a  $\beta = 1.5$  (cf. Equation 1), and the batch size is 256. The computation was done on NVIDIA A100 GPU using TensorFlow v2.9 (TensorFlow Developer (2022)). The model training time is less than 20 minutes.

### 2.5. Identifying climate prototypes associated with extremely high BL

To find the latent dimensions that are associated with extremely high BL, for each latent dimension, we compute the mean over samples that are associated with high BL and the mean over the remaining samples. We consider latent dimensions relevant for increasing BL when the difference between the two sample means exceeds half of the standard deviation (0.5).

We create weather prototypes associated with high BL based on the identification step above. For each identified latent dimension of interest, we change its value by a fixed amount  $s$  in the direction that increases the BL while setting all other dimensions to 0. If the absolute magnitude of the correlation between two dimensions is greater than 0.3. In that case, we also change the value of the correlated dimension, assuming a bivariate Gaussian distribution with a covariance that results in this correlation. Finally, we use the decoder to generate the climate prototype  $P_{i|s}$  for a given change  $s$  in dimensions  $i$  of the latent space. Compound climate prototypes are generated similarly by selecting the two latent dimensions of interest and changing both in the direction that increases BL, thus generating prototypes  $P_{i\wedge j}$ . We generate univariate prototypes with  $s = 2\sigma$  and compound prototypes with  $s = 2\sigma/\text{number of dimensions}$ , denoted by  $P_{i|2\sigma}$  and  $P_{i\wedge j|\sigma}$ , respectively.

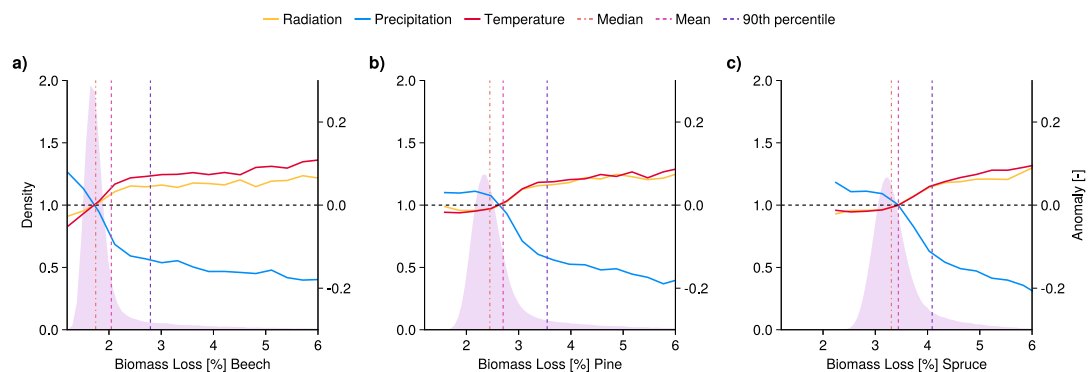
We further test whether samples corresponding to the prototypes are associated with generally higher levels of BL. To this end, we assign weather conditions in the original input space to a univariate prototype if, in the latent space, the value of the corresponding latent dimension exceeds  $s$ , i.e., two standard deviations. Similarly, a sample corresponds to a compound prototype if, in latent space, values in the two corresponding latent dimensions concurrently exceed  $s$ , i.e., one standard deviation.

## 3. Results

### 3.1. Composites and logistic regression

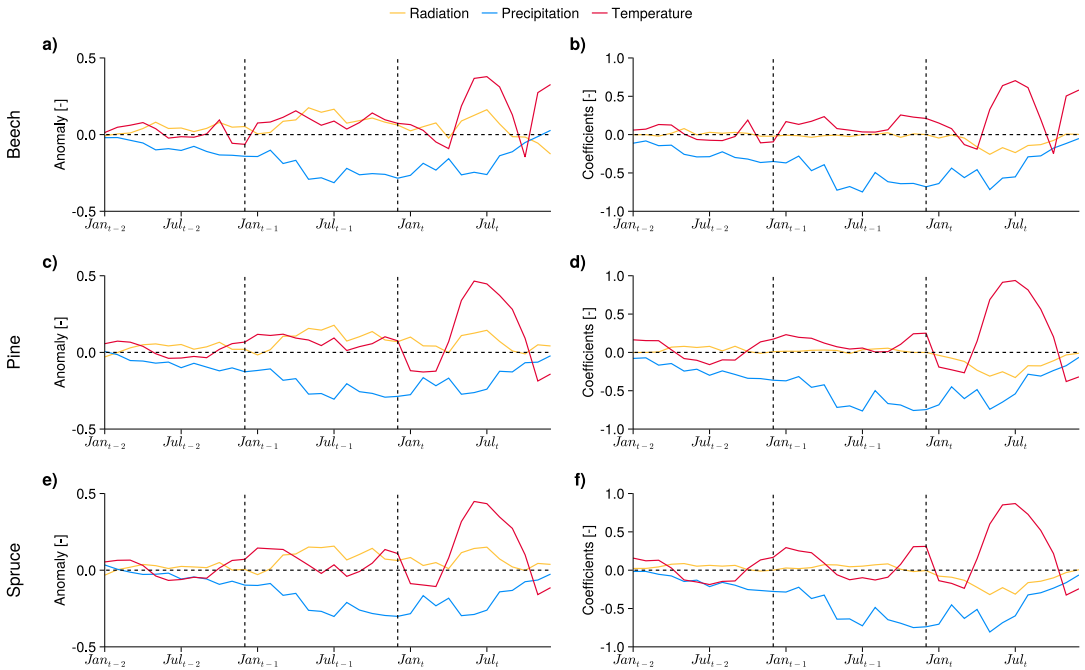
The median BL is 1.8%, 2.4%, and 3.3% per year for beech, pine, and spruce, respectively (Figure 4). The BL distribution is highly right-skewed, with a few years of experiencing extremely high BL values. Excess kurtosis for BL values is 24.8, 17.4, and 8.4 for beech, pine, and spruce forests. Generally, a lower BL is associated with lower radiation and temperature averaged over the preceding three years but higher precipitation, and vice versa for higher BL.

When we increase the temporal resolution of the composites to monthly values, we find that extremely high BL is associated with dry conditions in all forests, especially over the previous 18 months (Figure 5a,c,e). Another prominent feature is the strong positive temperature anomaly from May to September of the BL year. For beech, composites show no temperature anomalies in October<sub>t</sub> but positive anomalies in November<sub>t</sub> and December<sub>t</sub>. For pine and spruce, the signal is reversed for these



**Figure 4.** Probability density estimate of biomass loss (left axis), along with three-year average standardized temperature (red), radiation (yellow) and precipitation (blue) (right axis). Vertical lines indicate the median, mean, and 90th percentile of the biomass loss distribution.





**Figure 5.** Composite of monthly standardized weather anomalies associated with the 90th percentile of biomass loss (left) for beech (a), pine (c), and spruce (e). Coefficients of logistic regression without structure variables (right) for beech (b), pine (d), and spruce (f).

three months; anomalies are strongly positive in October<sub>t</sub> but negative in November<sub>t</sub> and December<sub>t</sub>. Radiation has an overall weaker signal that is negatively correlated with precipitation.

Predicting BL based on monthly weather predictors reaches average precision values between 0.64 (beech) and 0.70 (pine) when no structure variables are included (Table 1). Including structure variables as predictors increases accuracy substantially, reaching values between 0.76 and 0.85. Generally, the performance of the linear model is similar for pine and spruce forests and higher than for the beech forest. The coefficients of the logistic regression (without structure variables) generally mirror the behavior of the composites (Figure 5b,d,e). Coefficients associated with precipitation for high BL years are negative, increasing in magnitude until May<sub>t-1</sub> and staying constant until July<sub>t</sub>, finally decreasing until the end of the time series. The coefficients for temperature are relatively small, except between April<sub>t</sub> and September<sub>t</sub>. The coefficients associated with radiation are close to zero most of the time. The logistic regression coefficients and composite plots show similar trends in all the variables except for radiation in year<sub>t</sub>.

**Table 1.** Performance metrics for the logistic regression with weather variables only ( $\mathbf{x}_d$ ) and with structure variables ( $\mathbf{x}_d$  and  $\mathbf{x}_s$ )

| Metrics   | CSI  | F <sub>1</sub> -score | Average precision |
|---|------|-----------------------|-------------------|
| Beech logistic regression with $\mathbf{x}_d$                     | 0.43 | 0.60                  | 0.64              |
| Beech logistic regression with $\mathbf{x}_d$ and $\mathbf{x}_s$  | 0.52 | 0.68                  | 0.76              |
| Pine logistic regression with $\mathbf{x}_d$                      | 0.46 | 0.64                  | 0.70              |
| Pine logistic regression with $\mathbf{x}_d$ and $\mathbf{x}_s$   | 0.60 | 0.75                  | 0.85              |
| Spruce logistic regression with $\mathbf{x}_d$                    | 0.46 | 0.63                  | 0.68              |
| Spruce logistic regression with $\mathbf{x}_d$ and $\mathbf{x}_s$ | 0.59 | 0.74                  | 0.84              |

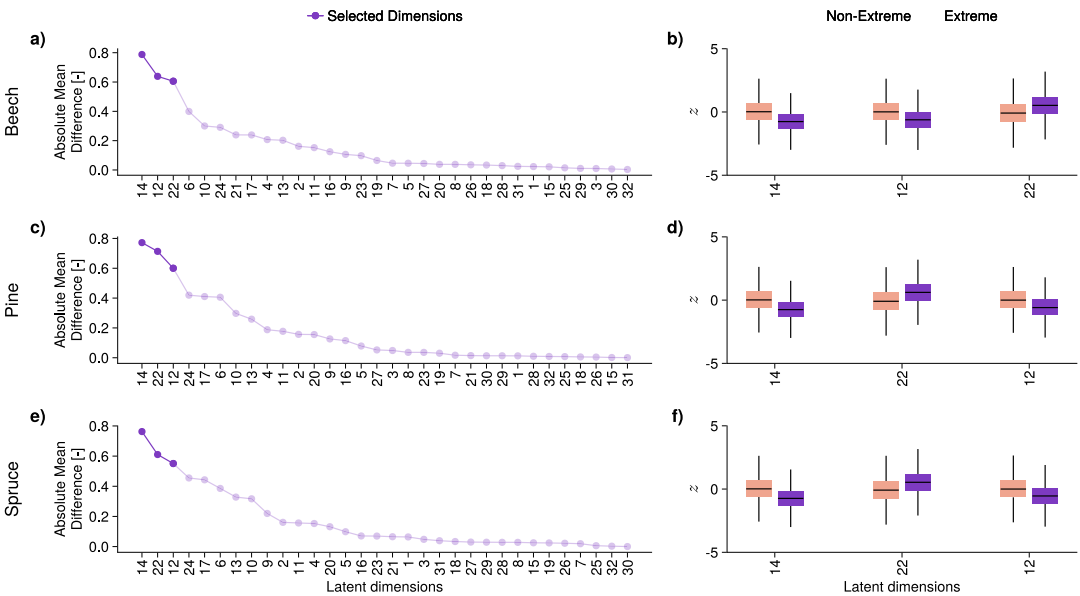
**Table 2.** Training, validation and test losses and mean standard errors (MSE) for  $\beta$ -VAE

| Name       | Total loss | KLD loss | Reconstruction | Radiation | Precipitation | Temperature |
|------------|------------|----------|----------------|-----------|---------------|-------------|
|            |            |          | MSE            | MSE       | MSE           | MSE         |
| Training   | 37.76      | 24.65    | 0.79           | 0.79      | 0.78          | 0.79        |
| Validation | 37.73      | 24.63    | 0.79           | 0.79      | 0.78          | 0.79        |
| Test       | 37.83      | 24.70    | 0.79           | 0.79      | 0.78          | 0.79        |

**3.2. Weather representations via  $\beta$ -VAE**

Initial tests with vanilla  $\beta$ -VAE with similar layers showed better reconstruction for the correlated variables (precipitation and radiation) under-representing temperature. The modified  $\beta$ -VAE achieves similar reconstruction errors for all three weather variables (Table 2) by introducing a shared decoder (Section 2.4). Losses and mean standard errors are similar between training, validation, and test datasets, suggesting that the model does not overfit.

We explore the latent dimensions ( $z_i$ ) of the trained  $\beta$ -VAE. Each three-year input sample is associated with one point in the 32-dimensional latent space. We test whether certain dimensions of this space are associated with weather conditions that lead to higher BL. Generally, we find high variability in the shift in distribution between the samples associated with extremely high BL and the remaining samples across the 32 dimensions (Figure 6, left). For all forests,  $z_{14}$  exhibits the largest absolute difference in the means (Figure 6, right) followed by  $z_{22}$  and  $z_{12}$ . Here,  $z_{12}$  shows a higher absolute mean difference than  $z_{14}$  for beech and vice versa for pine and spruce. All three dimensions show significant ( $p < 0.01$ ) mean absolute differences between years with extreme and nonextreme BL  $> 0.5$  for all forests and are selected to construct the prototypes of weather patterns associated with the high BL in the following. We also test



**Figure 6.** Absolute mean difference between samples in the latent dimensions representative of three-year periods leading to extreme and nonextreme BL for beech (a), pine (c), and spruce (e). The latent dimensions are sorted in descending order with respect to the absolute mean difference. Box plot of three latent dimensions with the highest absolute mean difference associated with nonextreme (orange) and extreme BL (purple) for beech (b), pine (d), and spruce (f).

**Table 3.** Logistic regression with  $z_{14}$ ,  $z_{22}$  and  $z_{12}$

| Type of forest | CSI  | $F_1$ -score | Average precision |
|----------------|------|--------------|-------------------|
| Beech          | 0.24 | 0.39         | 0.33              |
| Pine           | 0.24 | 0.39         | 0.34              |
| Spruce         | 0.22 | 0.36         | 0.30              |

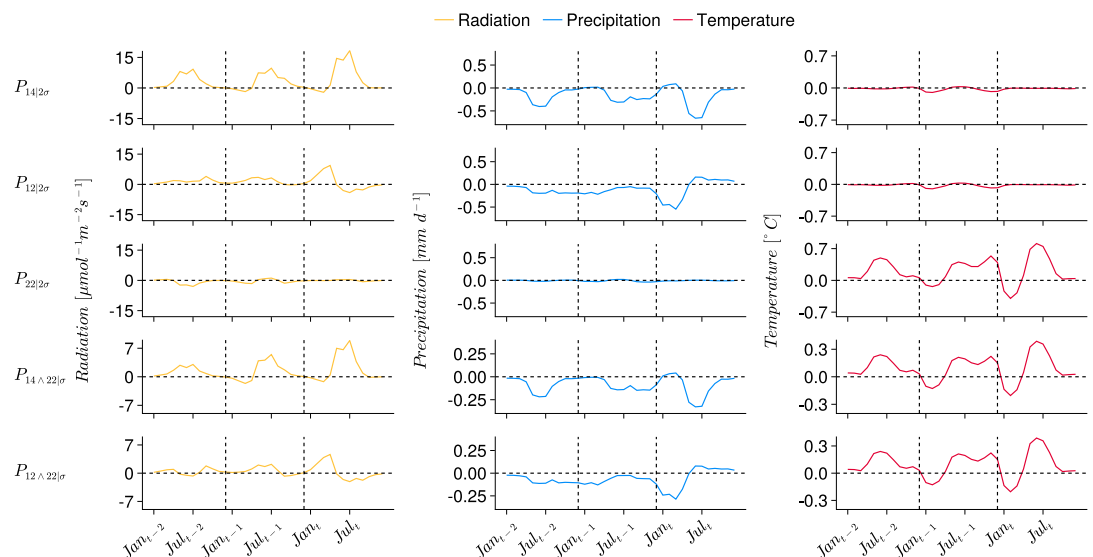
whether extreme BL is predictable when only relying on these three dimensions using a logistic regression model. Prediction accuracy is much lower than when monthly weather predictors are used, reaching average precision between 0.30 and 0.34, but is substantially higher than random (compare Table 3 with Table 1). When relying only on these three dimensions, beech and pine BL are better explained by the latent dimensions than spruce BL.

Our VAE architecture allows us to generate precipitation, temperature, and radiation independently from a sample in the latent space. Still, correlations exist between weather variables in the input space (Section 2.1). These correlations propagate to correlations between latent dimensions. We observe that the latent dimensions responsible solely for radiation ( $z_{1,8}$ ) are negatively correlated (correlation  $< -0.3$ ) with the latent dimensions responsible solely for precipitation ( $z_{9,16}$ ), respectively (Figure A1 in Appendix A). All other dimensions are not correlated.

### 3.3. Weather prototypes and link to BL

The three selected dimensions  $z_{14}$ ,  $z_{12}$  and  $z_{22}$  are used to generate weather prototypes  $P_{14|2\sigma}$ ,  $P_{12|2\sigma}$ , and  $P_{22|2\sigma}$ . By construction,  $z_{14}$  and  $z_{12}$  only affect the reconstruction of precipitation (Section 2.4). However, because of the correlation between  $z_{14}$  and  $z_6$ , as well as  $z_{12}$  and  $z_4$  (Figure A1), changing  $z_{14}$  and  $z_{12}$  also requires changing the values of  $z_6$  and  $z_4$  to obtain realistic prototypes. The prototype based on  $z_{22}$  is generated by only changing  $z_{22}$ .

Prototypes generated from the selected latent dimensions represent basic weather conditions associated with increased BL (Figure 7). Prototype  $P_{14|2\sigma}$  and  $P_{12|2\sigma}$ , by design, only show changes in precipitation and radiation. Both prototypes feature negative precipitation anomalies and positive radiation anomalies that are



**Figure 7.** Weather prototypes. Note that the y-axis ranges differ between the univariate prototypes (first three rows) and the compound prototypes (last two rows).

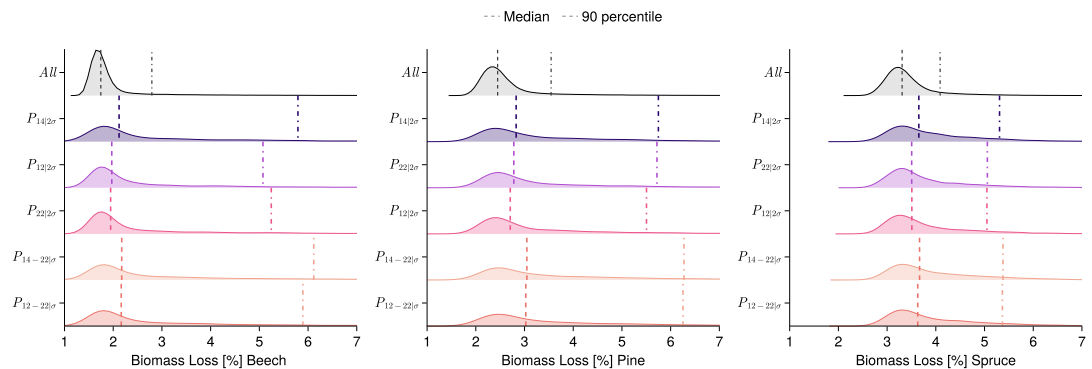
correlated.  $P_{14|2\sigma}$  is characterized by negative precipitation anomalies in all three summers, coincident with higher radiation. Precipitation is also lower in the autumn of the second year. The lowest precipitation values occur in the year  $t_i$  followed by the year  $t_{i-2}$  and the year  $t_{i-1}$ .  $P_{12|2\sigma}$  has a different pattern and is characterized by two periods of negative precipitation anomalies. The first one occurs between March  $t_{i-2}$  and July  $t_{i-1}$  and is less pronounced than the negative anomaly in the second period, which occurs in the first five months of the last year. We only see slightly positive precipitation anomalies in the last half of year  $t_i$ . In both prototypes, maximum precipitation anomalies are around 0.5 mm/d. Prototype  $P_{22|2\sigma}$  only features temperature variations, with positive temperature anomalies in three periods. The first period starts from March  $t_{i-2}$  to August  $t_{i-2}$ , the second from the March  $t_{i-1}$  and continues for the entire year. The last high-temperature period starts from March  $t_i$  and ends in September  $t_i$ . There is a short period at the beginning of the last year where we see negative temperature anomalies. Based on their anomaly patterns, we call prototypes  $P_{14|2\sigma}$  Dry Summers,  $P_{12|2\sigma}$  Dry Winters and  $P_{22|2\sigma}$  Hot Summers.

We generate two compound prototypes based on the selected three mortality-influencing latent dimensions, namely  $P_{12\wedge 22|2\sigma}$  and  $P_{14\wedge 22|2\sigma}$ . While the univariate prototypes are created by moving two standard deviations along a single latent dimension in the direction of increasing BL, compound prototypes are created by moving one standard deviation in two latent dimensions (Section 2.5).  $P_{14\wedge 22|2\sigma}$  corresponds to dry and hot summers, while  $P_{12\wedge 22|2\sigma}$  corresponds to dry winters and hot summers. Weather conditions associated with compound prototypes occur more frequently in the training dataset (2.5–3.4% of all samples) than those associated with univariate prototypes (1.7–2.2%).

Weather conditions that are associated with any of the prototypes (Section 2.5) are generally associated with a higher BL in all forest types when compared to all weather conditions (Figure 8). The effect is particularly strong in the tail of the distribution; the 90th percentile in the BL distribution is about twice as large for the prototypes compared to all samples (Table 4). From the univariate prototypes,  $P_{14|2\sigma}$  (Dry Summers) is associated with the highest increase in both median and 90th-percentile BL across all forest types (10.4–21.5% and 29.8–107.3%, respectively). Weather conditions associated with the compound prototypes  $P_{14\wedge 22|2\sigma}$  and  $P_{12\wedge 22|2\sigma}$  are generally associated with higher BL compared to weather conditions associated with any of the univariate prototypes, with  $P_{14\wedge 22|2\sigma}$  (Dry and Hot Summers) leading to slightly higher median BL (9.7–24.0%) compared to  $P_{12\wedge 22|2\sigma}$  (dry winters and hot summers, 10.8–24.6%).

#### 4. Discussion

This study illustrates how novel VAE-based architectures can learn weather prototypes associated with high BL in a simulated forest setting based on the weather for a given time period. Working with simulated



**Figure 8.** Biomass loss density plot for all samples (top), samples contributing to univariate prototypes (2nd to 4th row) and samples contributing to compound prototypes (last two rows) for beech (left), pine (middle), and spruce (right). Vertical dashed and dash-dotted lines represent the median and 90th percentile in each distribution.

**Table 4.** Percentage increase in the median and the 90th percentile biomass loss for different weather prototypes and different types of forests

| Genus  | Median           |                  |                  |                           |                           | 90th percentile  |                  |                  |                           |                           |
|--------|------------------|------------------|------------------|---------------------------|---------------------------|------------------|------------------|------------------|---------------------------|---------------------------|
|        | $P_{14 2\sigma}$ | $P_{12 2\sigma}$ | $P_{22 2\sigma}$ | $P_{14\wedge 22 2\sigma}$ | $P_{12\wedge 22 2\sigma}$ | $P_{14 2\sigma}$ | $P_{12 2\sigma}$ | $P_{22 2\sigma}$ | $P_{14\wedge 22 2\sigma}$ | $P_{12\wedge 22 2\sigma}$ |
| Beech  | 21.5             | 12.9             | 11.5             | 24.6                      | 24.0                      | 107.3            | 81.7             | 87.8             | 119.0                     | 111.0                     |
| Pine   | 15.5             | 13.6             | 10.6             | 24.5                      | 23.6                      | 62.1             | 61.3             | 55.2             | 76.9                      | 76.3                      |
| Spruce | 10.4             | 5.9              | 6.0              | 10.8                      | 9.7                       | 29.8             | 23.7             | 23.5             | 31.6                      | 31.3                      |

data allows us to avoid the complexities of real-world data arising from nonstationarity and confounders' influence. Applicability of these methods beyond stationary weather requires careful validation and potential modifications to the developed methodology, which is beyond the scope of this paper. Lastly, the reconstruction errors, though small and similar for all the weather variables, impact generating realistic weather prototypes. Very high reconstruction errors would lead to reconstructions different from the input. Within these limits, we test the applicability of our approach. The method could further be adapted for use with observational data in future research.

Here, we learn the weather representations in a completely unsupervised setting and then try to find the latent dimensions associated with high BL. Our generative model can produce realistic weather prototypes. Of the three univariate prototypes generated, dry summers and dry winters feature negative precipitation anomalies and positive radiation anomalies, while hot summers feature positive temperature anomalies in summers. All identified weather prototypes are associated with higher BL overall, particularly in the tail of the BL distribution. Dry Summers lead to the highest increase in BL within the univariate prototypes. However, when combined, dry and hot summers lead to the highest increase in BL overall. Gazol and Camarero (2022) also found that European forest mortality frequently cooccurs with dry and hot summers. Vegetation models and Earth system models also generally support this finding (Zscheischler et al., 2014a, 2014b; Bastos et al., 2021; Tschumi et al., 2022, 2023). Our analysis shows that compound weather prototypes lead to higher BL than univariate weather prototypes despite occurring more frequently. This is consistent with real-world observations, where concurrent hot and dry conditions have often been identified as primary drivers of forest mortality (Allen et al., 2010; Anderegg et al., 2013; Crouchet et al., 2019; Hammond et al., 2022; Yi et al., 2022).

Some weather variables used are correlated (Figure 2), so we modify the traditional  $\beta$ -VAE architecture to work well with correlated time series. In particular, we design our encoder to take multivariate, high-dimensional weather data and learn three sets of eight weather-specific latent dimensions, each corresponding to one of precipitation, radiation, and temperature. The final eight latent dimensions contain information about all three weather variables. The decoder takes 16 (eight weather-specific and eight shared) latent dimensions and outputs one of the weather variables. The rationale is that our decoder only learns the structure of the weather time series, and the latent representation has all the information about the time series. While working with correlated weather variables, our architecture helps us learn unbiased weather representations (Table 2). In earlier tests with the standard  $\beta$ -VAE architecture, we found that the two more correlated weather variables (precipitation and radiation) were better represented than the less correlated ones (temperature). Though we expected all the correlations between weather variables to be captured in the eight shared latent dimensions, this was not the case. We find a one-to-one correlation in all latent dimensions related to solar radiation and precipitation ( $z_{1:8}$  and  $z_{9:16}$ , Figure A1). The correlations in latent dimensions highlight that radiation and precipitation are correlated for most months of the year. Our method respects this correlation when generating prototypes (Section 2.5). Autoencoders have been used to learn representations of weather variables (Guevara et al., 2021; Heinze-Deml et al., 2021; Oliveira et al., 2022; Ahn et al., 2023). However, so far, most work focuses on learning spatial representations of weather variables or univariate spatiotemporal representations instead of a multivariate temporal latent representation of variables. Huamin et al. (2020) try

to learn time series representation, but only after transforming the time series to a recurrence plot, learning the latent representation of 2-D images for a univariate time series. The method we develop in this study can learn representations of multivariate time series and separate the representations specific to weather variables while considering the correlation between the variables.

The latent variables can capture the patterns in the weather variables, but they have no inherent information about forest structure or BL. Thus, they do not learn representations specific to high BL or given forest structures. In unsupervised settings, VAEs cannot learn the disentangled representation without inductive biases (Locatello et al., 2019). Including BL and forest structure information would introduce an inductive bias and possibly improve representation learning. In contrast, by not including forest structure, we identify weather prototypes leading to higher overall BL rather than conditioned to specific forest structures. The use of  $\beta$ -VAE may also be a limitation, as the reconstruction error is higher than that of standard VAE for  $\beta > 1$  (in Equation 1), possibly due to a trade-off of reconstruction error against disentanglement aligned with human intuition (Burgess et al., 2018). Fil et al. (2021) show how  $\beta > 1$  does not always yield the best qualitative disentanglement for complex datasets. There is no clear consensus on the correct choice of  $\beta$ , and one must subjectively balance the trade-off between disentanglement and reconstruction error.

Compared to simple composite analysis and logistic regression, the VAE allows us to understand better the diverse weather patterns associated with high BL. In particular, we identify three weather prototypes and their possible combinations, which would not be possible using simple analysis. Nevertheless, the composite plots and logistic regression coefficients also give interesting insights. The composites of anomalies (Figure 5) show drier weather on average before BL years alongside higher temperatures in the last year. Unlike pine and spruce, we see negative temperature anomalies for beech forests in October, which rise in November and December of the previous year. This may be explained by the difference in vegetation period between the forest types. In the FORMIND beech forest model, the dynamic vegetation period ends when the ten-day mean air temperature falls below 9 °C. Therefore, a negative temperature in October would lead to an early end of the vegetation period. Higher temperatures after this may lead to high BL by increasing maintenance respiration and inducing plant water stress. For needle-leaf forests like pine and spruce, the simulated vegetation period is year-round, so the effect of lower temperature leading to higher BL dominates. In general, temperature has both positive and negative effects on plant productivity. Increasing temperature has been reported to increase phenology-driven carbon uptake for temperate forests (Keenan et al., 2014), and the higher the carbon uptake, the healthier the forest. At the same time, high temperatures increase maintenance respiration for plants and can increase the atmosphere's evaporative demand (Millar and Stephenson, 2015), leading to water stress.

Using a DL model that can capture high-dimensional relationships enables us to study the relationship between BL and the preceding years of weather in more detail. There is strong evidence that it is essential to consider multiple years of weather when analyzing tree mortality in the real world. For instance, an assessment of many different forest mortality events across the globe showed that those are typically associated with negative precipitation anomalies two years before the mortality year (Hammond et al., 2022). In response to the Europe-wide 2003 drought, mortality peaked two years after the driest year (Senf et al., 2020; George et al., 2021). On the other hand, the drought in 2018 led to peak forest mortality in the same year (George et al., 2021). We chose the weather of the previous two years along with the current year to analyze the BL in the current year. Accounting for three years of weather emerged as a good trade-off between model complexity and accuracy in our model simulations.

Adding information about the forest structure ( $\mathbf{x}^s$ ) alongside the weather variables ( $\mathbf{x}^d$ ) increases the predictive accuracy for all forest types (Table 1). It is well known that forest structures play an essential role in forest mortality (van Mantgem et al., 2009; Lines et al., 2010; Young et al., 2017; Restaino et al., 2019). Our VAE does not use any information about the forest structure or forest mortality in the current setting, which may limit its usability. Adding this information to the input or the latent space might lead to a better representation learning for characterizing the extreme BL years.

In our study, we run FORMIND for 200 ha, for which we then see a somewhat deterministic response of forest dynamics to weather conditions. For smaller areas, stochasticity dominates forest mortality (Eid

and Øyen, 2003; Vanderwel et al., 2013). It is important to highlight that in this study, each simulated forest contains only one variety of trees. Most forests are more diverse and include multiple plant functional types. Thus, our interpretations of weather drivers apply to extensive homogeneous forests with a particular kind of tree.

## 5. Conclusion

In this study, we develop an interpretable machine learning-based method to identify compound weather drivers of BL in forests using generative DL on simulated forest dynamics. The proposed method is designed to work well with correlated time series variables, which are often present in climate-related data. In our application, where we use simulations of beech, spruce, and pine forests in a temperate climate, we find that sequential hot summers, sequential dry summers, and dry winters are the weather patterns chiefly associated with high BL. Furthermore, we demonstrate that combinations of these weather patterns are associated with even higher BL. The approach thus enables us to generate hypotheses for compound drivers of extreme impacts, which can then be verified by using, for instance, real-world observations. In principle, the method is able to generate new realistic weather conditions associated with high BL. Overall, the method is a step toward an improved understanding of the weather drivers of significant impacts by using deep generative models.

**Acknowledgments.** We thank Nadav Peleg for the help with setting up the weather generator (AWE-GEN). We also thank Miguel Mahecha for the fruitful discussions on the manuscript.

**Author contribution.** Conceptualization: M.A., J.Z.; Data curation: M.A.; Data visualization: M.A.; Methodology: M.A., J.Z.; Writing original draft: M.A.; All authors contributed to Writing – Review & Editing and approve the final submitted draft.

**Competing interest.** The authors declare none.

**Data availability statement.** The AWE-GEN and FORMIND simulations are available at zenodo <https://doi.org/10.5281/zenodo.7708544>. Deep Learning architecture and weights are available at <https://github.com/AIM4ES/weather-vae-fomo>.

**Funding statement.** M.A., L-b.S. and J.Z. acknowledge funding from the Helmholtz Initiative and Networking Fund (Young Investigator Group COMPOUNDX, Grant Agreement VH-NG-1537). G.C-V. was supported by the European Research Council (ERC) under the ERC Synergy Grant USMILE (grant agreement 855187). G.C-V. and J.Z. acknowledge the European Union's Horizon 2020 research and innovation program within the project 'XAIDA: Extreme Events – Artificial Intelligence for Detection and Attribution' under grant agreement No 101003469. FB was supported by the Collaborative Research Centre AquaDiva 559 of the Friedrich Schiller University Jena, funded by the Deutsche Forschungsgemeinschaft 560 (DFG, German Research Foundation) – SFB 1076 – “Project Number 218627073”.

## References

- Ahn H, Lee S, Ko H, Kim M, Han SW and Seok J (2023) Searching similar weather maps using convolutional autoencoder and satellite images. *ICT Express* 9(1), 69–75. <https://doi.org/10.1016/j.ict.2022.03.013>.
- Allen CD, Macalady AK, Chenchouni H, Bachelet D, McDowell N, Venetier M, Kitzberger T, Rigling A, Breshears DD, Hogg EHT, Gonzalez P, Fensham R, Zhang Z, Castro J, Demidova N, Lim J-H, Allard G, Running SW, Semerci A and Cobb N (2010) A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management* 259(4), 660–684.
- Alvarez-Melis D and Jaakkola TS (2018) Towards robust interpretability with self-explaining neural networks. In Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N and Garnett R (eds.), *Advances in Neural Information Processing Systems*, Vol. 31. Curran Associates, Inc. Available at [https://proceedings.neurips.cc/paper\\_files/paper/2018/file/3e9f0fc9b2f89e043bc6233994dfcf76-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2018/file/3e9f0fc9b2f89e043bc6233994dfcf76-Paper.pdf).
- Anderegg WRL, Kane JM and Anderegg LDL (2013) Consequences of widespread tree mortality triggered by drought and temperature stress. *Nature Climate Change* 3(1), 30–36.
- Bastos A, Orth R, Reichstein M, Ciais P, Viovy N, Zaehle S, Anthoni P, Arneth A, Gentine P, Joetzer E, Lienert S, Loughran T, McGuire PC, O S, Pongratz J and Sitch S (2021) Vulnerability of European ecosystems to two compound dry and hot summers in 2018 and 2019. *Earth System Dynamics* 12(4), 1015–1035.
- Bastos A, Sippel S, Frank D, Mahecha MD, Zaehle S, Zscheischler J and Reichstein M (2023) A joint framework for studying compound ecoclimatic events. *Nature Reviews Earth & Environment* 4, 333–350.

- Bequet R, Campioli M, Kint V, Vansteenkiste D, Muys B and Ceulemans R (2011) Leaf area index development in temperate oak and beech forests is driven by stand characteristics and weather conditions. *Trees* 25(5), 935–946.
- Bevacqua E, De Michele C, Manning C, Couason A, Ribeiro AFS, Ramos AM, Vignotto E, Bastos A, Blesić S, Durante F, Hillier J, Oliveira SC, Pinto JG, Ragno E, Rivoire P, Saunders K, van der Wiel K, Wu W, Zhang T and Zscheischler J (2021) Guidelines for studying diverse types of compound weather and climate events. *Earth's Future* 9(11), e2021EF002340.
- Bohn FJ, Frank K and Huth A (2014) Of climate and its resulting tree growth: Simulating the productivity of temperate forests. *Ecological Modelling* 278, 9–17.
- Bohn FJ, May F and Huth A (2018) Species composition and forest structure explain the temperature sensitivity patterns of productivity in temperate forests. *Biogeosciences* 15(6), 1795–1813.
- Bond-Taylor S, Leach A, Long Y and Willcocks CG (2022) Deep generative modelling: A comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44(11), 7327–7347.
- Bordt S, Finck M, Raidl E and von Luxburg U (2022) Post-hoc explanations fail to achieve their purpose in adversarial contexts. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 891–905. <https://doi.org/10.1145/3531146.3533153>.
- Brodribb TJ, Powers J, Cochard H and Choat B (2020) Hanging by a thread? Forests and drought *Science* 368(6488), 261–266.
- Bugmann H (2001) A review of forest gap models. *Climatic Change* 51(3), 259–305.
- Bugmann H, Seidl R, Hartig F, Bohn F, Brúna J, Cailleret M, François L, Heinke J, Henrot A-J, Hickler T, Hülsmann L, Huth A, Jacquemin I, Kollas C, Lasch-Born P, Lexer MJ, Merganič J, Merganičová K, Mette T, Miranda BR, Nadal-Sala D, Rammer W, Rammig A, Reineking B, Roedig E, Sabaté S, Steinkamp J, Suckow F, Vacchiano G, Wild J, Xu C and Reyher CPO (2019) Tree mortality submodels drive simulated long-term forest dynamics: Assessing 15 models from the stand to global scale. *Ecosphere* 10(2), e02616.
- Burgess CP, Higgins I, Pal A, Matthey L, Watters N, Desjardins G and Lerchner A (2018) Understanding disentangling in  $\beta$ -VAE. <https://doi.org/10.48550/arXiv.1804.03599>.
- Camps-Valls G, Tuia D, Zhu X and Reichstein M (2021) Deep learning for the earth sciences: A comprehensive approach to remote sensing. In *Climate Science and Geosciences*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119646181>.
- Chen C, Li O, Tao C, Barnett AJ, Su J and Rudin C (2019) This looks like that: Deep learning for interpretable image recognition. In Wallach H, Larochelle H, Beygelzimer A, dAlché-Buc F, Fox E and Garnett R (eds.), *Advances in Neural Information Processing Systems*, Vol. 32. Curran Associates, Inc. Available at [https://proceedings.neurips.cc/paper\\_files/paper/2019/file/adf7ee2dcf142b0e11888e72b43fcb75-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/adf7ee2dcf142b0e11888e72b43fcb75-Paper.pdf).
- Choat B, Brodribb TJ, Brodersen CR, Duursma RA, López R and Medlyn BE (2018) Triggers of tree mortality under drought. *Nature* 558(7711), 531–539.
- Crouchet SE, Jensen J, Schwartz BF and Schwinning S (2019) Tree mortality after a hot drought: Distinguishing density-dependent and -independent drivers and why it matters. *Frontiers in Forests and Global Change* 2, 21. <https://doi.org/10.3389/ffgc.2019.00021>.
- Cucchi M, Weedon GP, Amici A, Bellouin N, Lange S, Müller Schmied H, Hersbach H and Buontempo C (2020) WFDE5: Bias-adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data* 12(3), 2097–2120.
- Donaldson RJ, Dyer RM and Kraus MJ (1975) An objective evaluator of techniques for predicting severe weather events. Technical Report ADA015773, Air Force Cambridge Research Labs, Hanscom AFB, MA.
- Eid T and Øyen B-H (2003) Models for prediction of mortality in even-aged forest. *Scandinavian Journal of Forest Research* 18(1), 64–77.
- Faticchi S, Ivanov VY and Caporali E (2011) Simulation of future climate scenarios with a weather generator. *Advances in Water Resources* 34(4), 448–467.
- Felipe-Lucia MR, Soliveres S, Penone C, Manning P, van der Plas F, Boch S, Prati D, Ammer C, Schall P, Gossner MM, Bausch J, Buscot F, Blaser S, Blüthgen N, de Frutos A, Ehbrecht M, Frank K, Goldmann K, Hänsel F, Jung K, Kahl T, Nauss T, Oelmann Y, Pena R, Polle A, Renner S, Schloter M, Schöning I, Schrumpp M, Schulze E-D, Solly E, Sorkau E, Stempfhuber B, Tschapka M, Weisser WW, Wubet T, Fischer M and Allan E (2018) Multiple forest attributes underpin the supply of multiple ecosystem services. *Nature Communications* 9(1), 4839.
- Fil M, Mesinovic M, Morris M and Wildberger J (2021) Beta-VAE reproducibility: Challenges and extensions. <https://doi.org/10.48550/arXiv.2112.14278>.
- Fischer R, Bohn F, Dantas de Paula M, Dislich C, Groeneveld J, Gutiérrez AG, Kazmierczak M, Knapp N, Lehmann S, Paulick S, Pütz S, Rödig E, Taubert F, Köhler P and Huth A (2016) Lessons learned from applying a forest gap model to understand ecosystem and carbon dynamics of complex tropical forests. *Ecological Modelling* 326, 124–133.
- Fischer R, Ensslin A, Rutten G, Fischer M, Costa DS, Kleyer M, Hemp A, Paulick S and Huth A (2015) Simulating carbon stocks and fluxes of an African tropical montane forest with an individual-based forest model. *PLoS One* 10(4), e0123300.
- Forzieri G, Dakos V, McDowell NG, Ramdane A and Cescatti A (2022) Emerging signals of declining forest resilience under climate change. *Nature* 608(7923), 534–539.
- Friedlingstein P, Jones MW, O'Sullivan M, Andrew RM, Bakker DCE, Hauck J, Le Quéré C, Peters GP, Peters W, Pongratz J, Sitch S, Canadell JG, Ciais P, Jackson RB, Alin SR, Anthoni P, Bates NR, Becker M, Bellouin N, Bopp L, Chau TTT, Chevallier F, Chini LP, Cronin M, Currie KI, Decharme B, Djéutchouang LM, Dou X, Evans W, Feely RA, Feng L, Gasser T, Gilfillan D, Gkritzalis T, Grassi G, Gregor L, Gruber N, Gürses Ö, Harris I, Houghton RA, Hurtt GC, Iida Y, Ilyina T,



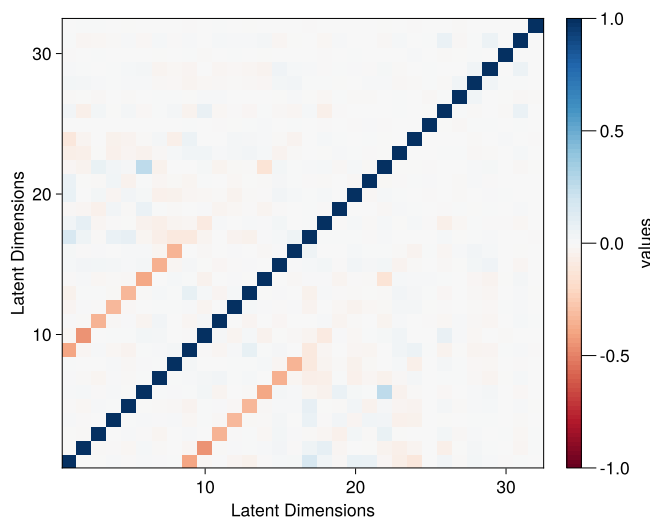
- Luijckx IT, Jain A, Jones SD, Kato E, Kennedy D, Klein Goldewijk K, Knauer J, Korsbakken JI, Körtzinger A, Landschützer P, Lauvset SK, Lefèvre N, Lienert S, Liu J, Marland G, McGuire PC, Melton JR, Munro DR, Nabel JEMS, Nakaoka S-I, Niwa Y, Ono T, Pierrot D, Poulter B, Rehder G, Resplandy L, Robertson E, Rödenbeck C, Rosan TM, Schwinger J, Schwingshackl C, Séférian R, Sutton AJ, Sweeney C, Tanhua T, Tans PP, Tian H, Tilbrook B, Tubiello F, van der Werf GR, Vuichard N, Wada C, Wanninkhof R, Watson AJ, Willis D, Wiltshire AJ, Yuan W, Yue C, Yue X, Zaehle S and Zeng J (2022) Global carbon budget 2021. *Earth System Science Data* 14(4), 1917–2005.
- Gautam S, Boubekki A, Hansen S, Salahuddin SA, Jossen R, Höhne MM and Kampffmeyer M (2022) ProtoVAE: A trustworthy self-explainable prototypical variational model. In Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K and Oh A (eds.), *Advances in Neural Information Processing Systems*, Vol. 35. Curran Associates, Inc., pp. 17940–17952. Available at [https://proceedings.neurips.cc/paper\\_files/paper/2022/file/722f3f9298a961d2639eadd3f14a2816-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/722f3f9298a961d2639eadd3f14a2816-Paper-Conference.pdf).
- Gazol A and Camarero JJ (2022) Compound climate events increase tree drought mortality across European forests. *Science of the Total Environment* 816, 151604.
- George J-P, Neumann M, Vogt J, Cammalleri C and Lang M (2021) Assessing effects of drought on tree mortality and productivity in European forests across two decades: A conceptual framework and preliminary results. In *IOP Conference Series: Earth and Environmental Science*, Vol. 932, p. 012009. <https://doi.org/10.1088/1755-1315/932/1/012009>.
- Guevara J, Borges D, Watson C and Zadrozny B (2021) A comparative study of stochastic and deep generative models for multisite precipitation synthesis. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*. Available at <https://www.climatechange.ai/papers/icml2021/47>.
- Hammond WM, Williams AP, Abatzoglou JT, Adams HD, Klein T, López R, Sáenz-Romero C, Hartmann H, Breshears DD and Allen CD (2022) Global field observations of tree die-off reveal hotter-drought fingerprint for Earth's forests. *Nature Communications* 13(1), 1761.
- Harris NL, Gibbs DA, Baccini A, Birdsey RA, de Bruin S, Farina M, Fatoyinbo L, Hansen MC, Herold M, Houghton RA, Potapov PV, Suarez DR, Roman-Cuesta RM, Saatchi SS, Slay CM, Turubanova SA and Tyukavina A (2021) Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change* 11(3), 234–240.
- Hartmann H, Bastos A, Das AJ, Esquivel-Muelbert A, Hammond WM, Martínez-Vilalta J, McDowell NG, Powers JS, Pugh TA, Ruthrof KX and Allen CD (2022) Climate change risks to global forest health: Emergence of unexpected events of elevated tree mortality worldwide. *Annual Review of Plant Biology* 73(1), 673–702.
- Heinz-Deml C, Sippel S, Pendergrass AG, Lehner F and Meinshausen N (2021) Latent linear adjustment autoencoder v1.0: A novel method for estimating and emulating dynamic precipitation at high resolution. *Geoscientific Model Development* 14(8), 4977–4999.
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S and Thépaut J-N (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146(730), 1999–2049.
- Higgins I, Matthay L, Pal A, Burgess C, Glorot X, Botvinick M, Mohamed S, and Lerchner A (2017) Beta-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*. Available at <https://openreview.net/forum?id=Sy2fzU9gl>.
- Hirschberg J, Faticchi S, Bennett GL, McArdeell BW, Peleg N, Lane SN, Schlunegger F and Molnar P (2021) Climate change impacts on sediment yield and debris-flow activity in an alpine catchment. *Journal of Geophysical Research: Earth Surface* 126(1), e2020JF005739.
- Holtmann A, Huth A, Pohl F, Rebmann C and Fischer R (2021) Carbon sequestration in mixed deciduous forests: The influence of tree size and species composition derived from model experiments. *Forests* 12(6), 726.
- Huamin T, Qiuqun D and Shanzhu X (2020) Reconstruction of time series with missing value using 2D representation-based denoising autoencoder. *Journal of Systems Engineering and Electronics* 31(6), 1087–1096.
- Ioffe S and Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37. PMLR, pp. 448–456. Available at <https://doi.org/10.5555/3045118.3045167>.
- Jaime L, Batllori E, Ferretti M and Lloret F (2022) Climatic and stand drivers of forest resistance to recent bark beetle disturbance in European coniferous forests. *Global Change Biology* 28(8), 2830–2841.
- Jain P, Castellanos-Acuna D, Coogan SCP, Abatzoglou JT and Flannigan MD (2022) Observed increases in extreme fire weather driven by atmospheric humidity and temperature. *Nature Climate Change* 12(1), 63–70.
- Ježík M, Blaženeč M, Štřelcová K and Dítmarová L' (2011) The impact of the 2003–2008 weather variability on intra-annual stem diameter changes of beech trees at a submontane site in Central Slovakia. *Dendrochronologia* 29(4), 227–235.
- Keenan TF, Gray J, Friedl MA, Toomey M, Bohrer G, Hollinger DY, Munger JW, O'Keefe J, Schmid HP, Wing IS, Yang B and Richardson AD (2014) Net carbon uptake has increased through warming-induced changes in temperate forest phenology. *Nature Climate Change* 4(7), 598–604.
- Kim B, Wattenberg M, Gilmer J, Cai C, Wexler J, Viegas F and Sayres R (2018) Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV). In Dy J and Krause A (eds.), *Proceedings of the 35th International Conference on Machine Learning*, Vol. 80. PMLR, pp. 2668–2677. Available at <https://proceedings.mlr.press/v80/kim18d.html>.

- Kingma DP and Welling M** (2013) Auto-encoding variational bayes, 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, Conference Track Proceedings.
- Labe ZM and Barnes EA** (2021) Detecting climate signals using explainable AI with single-forcing large ensembles. *Journal of Advances in Modeling Earth Systems* 13(6), e2021MS002464.
- Le Grix N, Cheung WL, Reygondeau G, Zscheischler J and Frölicher TL** (2023) Extreme and compound ocean events are key drivers of projected low pelagic fish biomass. *Global Change Biology* 29(23), 6478–6492.
- Li W, Guo W-Y, Pasgaard M, Niu Z, Wang L, Chen F, Qin Y and Svenning J-C** (2023) Human fingerprint on structural density of forests globally. *Nature Sustainability* 6, 368–379.
- Lim B and Zohren S** (2021) Time-series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379(2194), 20200209.
- Lines ER, Coomes DA and Purves DW** (2010) Influences of forest structure, climate and species composition on tree mortality across the Eastern US. *PLoS One* 5(10), e13212.
- Liu Y, Duffy K, Dy JG and Ganguly AR** (2023) Explainable deep learning for insights in El Niño and river flows. *Nature Communications* 14(1), 339.
- Locatello F, Bauer S, Lucic M, Rätsch G, Gelly S, Schölkopf B and Bachem O** (2019) Challenging common assumptions in the unsupervised learning of disentangled representations. In Chaudhuri K and Salakhutdinov R (eds.), *Proceedings of the 36th International Conference on Machine Learning*, Vol. 97. PMLR, pp. 4114–4124. Available at <https://proceedings.mlr.press/v97/locatello19a.html>.
- López-Serrano FR, García-Morote A, Andrés-Abellán M, Tendero A and del Cerro A** (2005) Site and weather effects in allometries: A simple approach to climate change effect on pines. *Forest Ecology and Management* 215(1), 251–270.
- Maas AL, Hannun AY and Ng AY** (2013) Rectifier nonlinearities improve neural network acoustic models. In *Proceedings of the 30th International Conference on Machine Learning*, Vol. 30. Available at [https://ai.stanford.edu/~amaas/papers/relu\\_hybrid\\_icml2013\\_final.pdf](https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf).
- Mahecha MD, Bastos A, Bohn FJ, Eisenhauer N, Feilhauer H, Hartmann H, Hickler T, Kalesse-Los H, Migliavacca M, Otto FEL, Peng J, Quaas J, Tegen I, Weigelt A, Wendisch M and Wirth C** (2022) Biodiversity loss and climate extremes — Study the feedbacks. *Nature* 612(7938), 30–32.
- Mahnken M, Cailleret M, Collalti A, Trotta C, Biondo C, D’Andrea E, Dalmonech D, Marano G, Mäkelä A, Minunno F, Peltoniemi M, Trotsiuk V, Nadal-Sala D, Sabaté S, Vallet P, Aussenac R, Cameron DR, Bohn FJ, Grote R, Augustynczyk ALD, Yousefpour R, Huber N, Bugmann H, Merganičová K, Merganic J, Valent P, Lasch-Born P, Hartig F, Vega del Valle ID, Volkholz J, Gutsch M, Matteucci G, Krejza J, Ibrom A, Meesenburg H, Rötzer T, van der Maaten-Theunissen M, van der Maaten E and Reyer CPO** (2022) Accuracy, realism and general applicability of European forest models. *Global Change Biology* 28(23), 6921–6943.
- Mamalakis A, Barnes EA and Ebert-Uphoff I** (2023) Carefully choose the baseline: Lessons learned from applying XAI attribution methods for regression tasks in geoscience. *Artificial Intelligence for the Earth Systems* 2(1), e220058, <https://doi.org/10.1175/AIES-D-22-0058>.
- Marcolongo A, Vladymyrov M, Lienert S, Peleg N, Haug S and Zscheischler J** (2022) Predicting years with extremely low gross primary production from daily weather data using convolutional neural networks. *Environmental Data Science* 1, e2.
- Mastrotheodoros T, Pappas C, Molnar P, Burlando P, Manoli G, Parajka J, Rigon R, Szeles B, Bottazzi M, Hadjidoukas P and Faticchi S** (2020) More green and less blue water in the Alps during warmer summers. *Nature Climate Change* 10(2), 155–161.
- Matusick G, Ruthrof KX, Kala J, Brouwers NC, Breshears DD and Hardy GESJ** (2018) Chronic historical drought legacy exacerbates tree mortality and crown dieback during acute heatwave-compounded drought. *Environmental Research Letters* 13(9), 095002.
- McDowell NG, Ryan MG, Zeppel MJB and Tissue DT** (2013) Improving our knowledge of drought-induced forest mortality through experiments, observations, and modeling. *The New Phytologist* 200(2), 289–293.
- Meili N, Manoli G, Burlando P, Bou-Zeid E, Chow WTL, Coutts AM, Daly E, Nice KA, Roth M, Tapper NJ, Velasco E, Vivoni ER and Faticchi S** (2020) An urban ecohydrological model to quantify the effect of vegetation on urban climate and hydrology (UT&C v1.0). *Geoscientific Model Development* 13(1), 335–362.
- Millar CI and Stephenson NL** (2015) Temperate forest health in an era of emerging megadisturbance. *Science* 349(6250), 823–826.
- Oliveira DAB, Diaz JG, Zadrozny B, Watson CD and Zhu XX** (2022) Controlling weather field synthesis using variational autoencoders. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5027–5030. <https://doi.org/10.1109/IGARSS46834.2022.9884668>.
- Park Williams A, Allen CD, Macalady AK, Griffin D, Woodhouse CA, Meko DM, Swetnam TW, Rauscher SA, Seager R, Grissino-Mayer HD, Dean JS, Cook ER, Gangodagamage C, Cai M and McDowell NG** (2013) Temperature as a potent driver of regional forest drought stress and tree mortality. *Nature Climate Change* 3(3), 292–297.
- Pilli R, Alkama R, Cescatti A, Kurz WA and Grassi G** (2022) The European forest carbon budget under future climate conditions and current management practices. *Biogeosciences* 19(13), 3263–3284.
- Rais A, Uhl E, van de Kuilen J-WG and Pretzsch H** (2021) Short-term reaction of European beech stem taper due to weather extremes. *Forest Ecology and Management* 480, 118653.

- Restaino C, Young DJN, Estes B, Gross S, Wuenschel A, Meyer M and Safford H** (2019) Forest structure and climate mediate drought-induced tree mortality in forests of the Sierra Nevada, USA. *Ecological Applications* 29(4), e01902.
- Rödiger E, Huth A, Bohn F, Rebmann C and Cuntz M** (2017) Estimating the carbon fluxes of forests with an individual-based forest model. *Forest Ecosystems* 4(1), 4.
- Rudin C** (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1(5), 206–215.
- Rumelhart DE, Hinton GE and Williams RJ** (1986) Learning representations by back-propagating errors. *Nature* 323(6088), 533–536.
- Sánchez-Pinillos M, D'Orangeville L, Boulanger Y, Comeau P, Wang J, Taylor AR and Kneeshaw D** (2022) Sequential droughts: A silent trigger of boreal forest mortality. *Global Change Biology* 28(2), 542–556.
- Scheel M, Lindeskog M, Smith B, Suvanto S and Pugh TAM** (2022) Increased central European forest mortality explained by higher harvest rates driven by enhanced productivity. *Environmental Research Letters* 17(11), 114007.
- Scherer D, Müller A and Behnke S** (2010) Evaluation of pooling operations in convolutional architectures for object recognition. In Diamantaras K, Duch W and Iliadis LS (eds.), *Artificial Neural Networks – ICANN 2010, Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, pp. 92–101.
- Seneviratne S, Zhang X, Adnan M, Badi W, Dereczynski C, Di Luca A., Ghosh S, Iskandar I, Kossin J, Lewis S, Otto F, Pinto I, Satoh M, Vicente-Serrano S, Wehner M and Zhou B** (2021) Weather and climate extreme events in a changing climate supplementary material. In Masson-Delmotte V, Zhai P, Pirani A, Connors S, Pean C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis M, Huang M, Leitzell K, Lonnoy E, Matthews J, Maycock T, Waterfield T, Yelekci O, Yu R and Zhou B (eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK; New York, NY: Cambridge University Press, pp. 1513–1766. <https://doi.org/10.1017/9781009157896.013>.
- Senf C, Buras A, Zang CS, Rammig A and Seidl R** (2020) Excess forest mortality is consistently linked to drought across Europe. *Nature Communications* 11(1), 6200.
- Senf C, Pflugmacher D, Zhiqiang Y, Sebald J, Knorn J, Neumann M, Hostert P and Seidl R** (2018) Canopy mortality has doubled in Europe's temperate forests over the last three decades. *Nature Communications* 9(1), 4978.
- Shugart HH, Wang B, Fischer R, Ma J, Fang J, Yan X, Huth A and Armstrong AH** (2018) Gap models and their individual-based relatives in the assessment of the consequences of global change. *Environmental Research Letters* 13(3), 033001.
- Sturrock RN, Frankel SJ, Brown AV, Hennon PE, Kliejunas JT, Lewis KJ, Worrall JJ and Woods AJ** (2011) Climate change and forest diseases. *Plant Pathology* 60(1), 133–149.
- Sweet L-b, Müller C, Anand M and Zscheischler J** (2023) Cross-validation strategy impacts the performance and interpretation of machine learning models. *Artificial Intelligence for the Earth Systems* 2(4), e230026.
- Tacoen A, Piedallu C, Seynave I, Perez V, Gégout-Petit A, Nageleisen L-M, Bontemps J-D and Gégout J-C** (2019) Background mortality drivers of European tree species: Climate change matters. *Proceedings of the Royal Society B: Biological Sciences* 286(1900), 20190386.
- TensorFlow Developer** (2022) TensorFlow (v2.9.1). Zenodo. <https://doi.org/10.5281/zenodo.6574233>.
- Tibau X-A, Reimers C, Requena-Mesa C and Runge J** (2021) Spatio-temporal autoencoders in weather and climate research. In *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*. John Wiley & Sons, Ltd., pp. 186–203. <https://doi.org/10.1002/9781119646181.ch1>.
- Tschumi E, Lienert S, Bastos A, Ciais P, Gregor K, Joos F, Knauer J, Papastefanou P, Rammig A, van der Wiel K, Williams K, Xu Y, Zaehle S and Zscheischler J** (2023) Large variability in simulated response of vegetation composition and carbon dynamics to variations in drought-heat occurrence. *Journal of Geophysical Research: Biogeosciences* 128(4), e2022JG007332.
- Tschumi E, Lienert S, van der Wiel K, Joos F and Zscheischler J** (2022) The effects of varying drought-heat signatures on terrestrial carbon dynamics and vegetation composition. *Biogeosciences* 19(7), 1979–1993.
- van Mantgem PJ, Stephenson NL, Byrne JC, Daniels LD, Franklin JF, Fulé PZ, Harmon ME, Larson AJ, Smith JM, Taylor AH and Veblen TT** (2009) Widespread increase of tree mortality rates in the Western United States. *Science* 323(5913), 521–524.
- Vanderwel MC, Coomes DA and Purves DW** (2013) Quantifying variation in forest disturbance, and its effects on aboveground biomass dynamics, across the eastern United States. *Global Change Biology* 19(5), 1504–1517.
- Varando G, Fernández-Torres M-A and Camps-Valls G** (2021) Learning granger causal feature representations. In *International Conference on Machine Learning - Workshop on Tackling Climate Change with Machine Learning*. ICML. Available at <https://www.climatechange.ai/papers/icml2021/34>.
- Vogel J, Rivoire P, Deidda C, Rahimi L, Sauter CA, Tschumi E, van der Wiel K, Zhang T and Zscheischler J** (2021) Identifying meteorological drivers of extreme impacts: An application to simulated crop yields. *Earth System Dynamics* 12(1), 151–172.
- Wadoux AMJ-C, Saby NPA and Martin MP** (2023) Shapley values reveal the drivers of soil organic carbon stock prediction. *The Soil* 9(1), 21–38.
- Wang Z, Yan W and Oates T** (2017) Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1578–1585. <https://doi.org/10.1109/IJCNN.2017.7966039>.

- Wolanin A, Mateo-García G, Camps-Valls G, Gómez-Chova L, Meroni M, Duveiller G, Liangzhi Y and Guanter L (2020)** Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. *Environmental Research Letters* 15(2), 024019.
- Wu Z, Pan S, Long G, Jiang J, Chang X and Zhang C (2020)** Connecting the Dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20. New York, NY: Association for Computing Machinery, pp. 753–763.
- Xu P, Zhou T, Yi C, Fang W, Hendrey G and Zhao X (2018)** Forest drought resistance distinguished by canopy height. *Environmental Research Letters* 13(7), 075003.
- Yan L, Lei Q, Jiang C, Yan P, Ren Z, Liu B and Liu Z (2022)** Climate-informed monthly runoff prediction model using machine learning and feature importance analysis. *Frontiers in Environmental Science* 10, 1049840, <https://doi.org/10.3389/fenvs.2022.1049840>.
- Yang M, Liu F, Chen Z, Shen X, Hao J and Wang J (2021)** CausalVAE: Disentangled representation learning via neural structural causal models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9588–9597. <https://doi.org/10.1109/CVPR46437.2021.00947>.
- Yi C, Hendrey G, Niu S, McDowell N and Allen CD (2022)** Tree mortality in a warming world: Causes, patterns, and implications. *Environmental Research Letters* 17(3), 030201.
- Young DJN, Stevens JT, Earles JM, Moore J, Ellis A, Jirka AL and Latimer AM (2017)** Long-term climate and competition explain forest mortality patterns under extreme drought. *Ecology Letters* 20(1), 78–86.
- Zeiler MD, Krishnan D, Taylor GW and Fergus R (2010)** Deconvolutional networks. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2528–2535. <https://doi.org/10.1109/CVPR.2010.5539957chat>.
- Zhao B, Lu H, Chen S, Liu J and Wu D (2017)** Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics* 28(1), 162–169.
- Zscheischler J, Martius O, Westra S, Bevacqua E, Raymond C, Horton RM, van den Hurk B, AghaKouchak A, Jézéquel A, Mahecha MD, Maraun D, Ramos AM, Ridder NN, Thiery W and Vignotto E (2020)** A typology of compound weather and climate events. *Nature Reviews Earth & Environment* 1(7), 333–347.
- Zscheischler J, Michalak AM, Schwalm C, Mahecha MD, Huntzinger DN, Reichstein M, Berthier G, Ciais P, Cook RB, El-Masri B, Huang M, Ito A, Jain A, King A, Lei H, Lu C, Mao J, Peng S, Poulter B, Ricciuto D, Shi X, Tao B, Tian H, Viovy N, Wang W, Wei Y, Yang J and Zeng N (2014a)** Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MSTMIP data. *Global Biogeochemical Cycles* 28(6), 585–600.
- Zscheischler J, Reichstein M, von Buttlar J, Mu M, Randerson JT and Mahecha MD (2014b)** Carbon cycle extremes during the 21st century in CMIP5 models: Future evolution and attribution to climatic drivers. *Geophysical Research Letters* 41(24), 8853–8861.
- Zscheischler J, Westra S, van den Hurk BJJM, Seneviratne SI, Ward PJ, Pitman A, AghaKouchak A, Bresch DN, Leonard M, Wahl T and Zhang X (2018)** Future climate risk from compound events. *Nature Climate Change* 8(6), 469–477.

## A. Figures



**Figure A1.** Pearson's correlation coefficients among all latent dimensions.

## B. Metrics

Three metrics are used for measuring the performance of the logistic regression models. CSI is equal to the total number of correct events predicted (true-positives) divided by the total number of events predicted (true-positives + false-positives) plus the number of misses (false-negatives).

$$CSI = \frac{TP}{TP + FP + FN}.$$

$F_1$ -score is the harmonic mean of precision and recall, which can also be written as a function of true-positives, false-positives, and false-negatives.

$$F_1\text{-Score} = \frac{2TP}{2TP + FP + FN}.$$

Average Precision is the area under the precision and recall curve. It is typically used for imbalanced datasets. CSI,  $F_1$ -Score and Average Precision range between 0 and 1, with higher values indicating a better model.