EcoDiffusion: Uncertainty-Aware Emulation of Ecosystem Processes with Conditional Diffusion for Long Sequences with Single-Step Initialization

Ruohan Li[1], Zhihao Wang[1], Xiaowei Jia[2], Gengchen Mai[3], Lei Ma[1], George C. Hurtt[1], Quan Shen[1], Zhili Li[1], Yiqun Xie[1]*

[1]University of Maryland
[2]University of Pittsburgh
[3]University of Texas at Austin
{r526li, zhwang1}@umd.edu, xiaowei@pitt.edu, gengchen.mai@austin.utexas.edu, {lma6, gchurtt, qshen, lizhili, xie}@umd.edu

Abstract

Terrestrial ecosystems constitute a major component of the global carbon sink and play a critical role in regulating the global carbon cycle. Although process-based models such as the Ecosystem Demography (ED) model are widely used to simulate these dynamics and widely adopted in research and applications, they remain computationally intensive and are not well suited for large-scale (e.g., global) projections at high spatial and temporal resolution, or under wide-range of future scenarios. AI-based emulators of process-based physical models have emerged as promising ways to accelerate the computation. However, there are several challenges in developing emulators for ecosystem processes, including error accumulation over long sequences, single-step initial conditions, and high-dimensional environmental conditions. Existing works often rely on time-series patterns in look-back windows, which are not well-suited for the problem with single-step initial conditions. Moreover, they often do not consider uncertainty, making it hard to know when the approximations are highly confident and when the results may need to be updated, e.g., by the process-based models. To address these limitations, we introduce EcoDiffusion, a conditional diffusion framework tailored for ecosystem dynamics emulation. We evaluated EcoDiffusion at locations distributed worldwide under different scenarios and showed that it demonstrated significant improvements over existing models.

Introduction

Terrestrial ecosystems are a critical component of the global carbon cycle, regulating key climate-related fluxes of carbon, energy, and water (Bonan 2008). Accurate modeling of ecosystem processes is widely recognized as essential for improving projections of future carbon dynamics and capturing biologically mediated feedbacks in climate modeling. To this end, many process-based models have been developed to simulate ecosystem processes (Hurtt et al. 1998; Fisher et al. 2018; Lawrence et al. 2019; Sato, Itoh, and Kohyama 2007; Prentice et al. 2007). However, the advances also lead to significantly higher costs of computation due to increased model complexity. In particular, simulating ecological demographic processes–such as tracking individual-level trees, cohort competitions, and their connections to global-scale carbon cycles is a highly expensive task (e.g., months or years of computation on supercomputers) (Fisher et al. 2018). As a result, current models are not well-suited for large-scale projections at both high spatial and temporal resolution, or under a wide array of future climate scenarios.

Emulating ecosystem models with machine learning presents several challenges compared to standard time series forecasting tasks. First, these models must handle long sequences that span decades at monthly or higher temporal resolution, where trends and seasonal variability complicate pattern separation and increase error accumulation. Second, ecosystem simulations often rely on single-step initial conditions, instead of a longer look-back window, to forecast all the time steps for the rest of the sequence. Third, ecosystem emulation must effectively account for high-dimensional environment variables (e.g., climate forcings) and capture complex, multivariate dependencies. The emulator also needs to recognize differences in initial conditions and correctly generate different corresponding forecasts under the similar environmental variables over long periods. Moreover, it is highly desired for the emulator to have uncertainty-awareness, so users are informed about the confidence of the approximations under different conditions.

Time-series forecasting models (Zhou et al. 2021, 2022; Wu et al. 2022; Wang et al. 2024a) are naturally good candidates for the emulation which is essentially a forecasting task. However, existing formulations mostly assume rich historical context and rely on patterns in past look-back windows (e.g., tens or over a hundred past steps) to make forecasts, making them less effective for ecosystem forecasting with single-step initial conditions. Moreover, they often do not offer explicit uncertainty quantification. Diffusion models offer a promising avenue for probabilistic forecasting (Ho, Jain, and Abbeel 2020; Rasul et al. 2021; Tashiro et al. 2021; Lin et al. 2024; Meijer and Chen 2024). However, their adaptation to process-based ecosystem emulation is currently underexplored. Autoregressive diffusion models (Rasul et al. 2021) allow for integration of exogenous inputs as additional conditions but suffer from error accumulation across long horizons, which is a major concern for long-term ecosystem projections. Non-autoregressive diffusion models that are forecasting-oriented (Shen and Kwok 2023; Li et al. 2024) often rely on look-back windows like in typical time-series forecasting models. The imputation-based versions (Tashiro et al. 2021; Alcaraz

*Corresponding author.

and Strodthoff 2022) further use bidirectional windows on all inputs that may lead to logically inconsistent results for forecasting. In addition, they are not explicitly designed to incorporate rich contextual information that is essential in ecosystem modeling with external environmental variables.

To address the limitations, we present EcoDiffusion to emulate the Ecosystem Demography (ED) model for carbon monitoring, with the following contributions:

- We propose EcoDiffusion, an uncertainty-aware time-series forecasting model to emulate ED based on conditional diffusion for complex ecosystem processes. These processes have long sequences with high-dimensional environmental variables but an initial condition for a single time step. Unlike models that condition on past time-series patterns, EcoDiffusion conditions on "future" contexts (e.g., different future climate scenarios from CMIP-6) to make corresponding future projections and assist policy making. EcoDiffusion enforces strict temporal causal masks on these conditional variables.
- EcoDiffusion reduces error accumulation using a non-autoregressive process with a multi-scale subset-forecasting and full-set-imputation to enhance the use of the initial condition and environment contexts. We also separate the full-set imputation process of target variables based on their different patterns to avoid confusion in training and incorporate explicit encoding of initial conditions and climate zones to increase the model's sensitivity to these factors.
- We evaluate the performance of EcoDiffusion at 852 locations over the globe under different climate conditions. The results show significant improvements over existing methods and enhanced abilities to capture uncertainty under different scenarios.

Problem Definition

Background: Ecosystem Demography (ED) Model

The ED model has been developed for two decades (Hurtt et al. 1998; Fisher et al. 2018; Ma et al. 2022) and provides comprehensive modeling of the forest ecological processes, including individual-tree level growth, mortality, competition, and many more, which are further integrated with carbon cycle, water cycle, disturbances, etc. It considers environmental variables (e.g., climate forcings, soil properties) to forecast carbon-related variables. The driving data includes meteorology from NASA Daymet and MERRA2, soil properties from the POLARIS dataset, land use change forcing from LUH2 (Hurtt et al. 2020) and CO_2 concentration from NOAA CarbonTracker. The output variables include canopy height, aboveground biomass (AGB), soil carbon, leaf area index (LAI), gross primary production (GPP), net primary production (NPP), and heterotrophic respiration (Rh). ED has been extensively validated at regional and global scales by comparing against in-situ and airborne observations (e.g., GEDI, ICESat-2) (Ma et al. 2022) and widely adopted in both research and applications (e.g., NASA Carbon Monitoring System, NASA GEDI mission, Global Carbon Budget, Maryland Dept. of Env.).

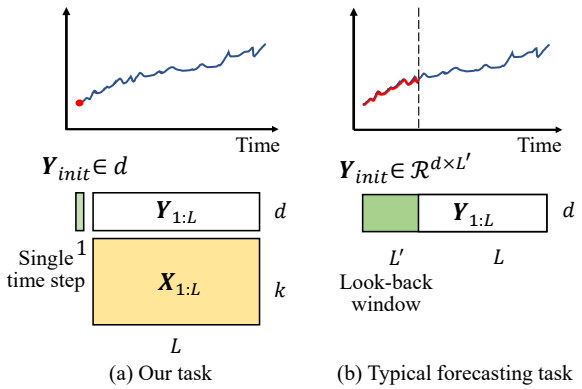

Figure 1: Problem formulation: A comparison.

Task Formulation

Let $\boldsymbol{Y}_{1:L} \in \mathbb{R}^{d \times L}$ denote the simulated output variables, where $d = 7$ is the number of output variables and $L = n_{\text{year}} \times n_{\text{month}}$ is the total length of the time series (e.g., $L = 39 \times 12 = 468$ for 39 years). Let $\boldsymbol{X}_{1:L} \in \mathbb{R}^{k \times L}$ denote the climate forcing variables, where $k = 136$ is the number of input features. Given the initial conditions at a single time step $\boldsymbol{Y}_{init} \in \mathbb{R}^d$, we aim to forecast $\boldsymbol{Y}_{1:L}$ for future time steps based on environment conditions $\boldsymbol{X}_{1:L}$.

Fig. 1 (a) illustrates the task, which has some interesting characteristics that are different from many typical time-series forecasting formulations:

- **Single-step initial condition.** It has only a single-step initial condition $\boldsymbol{Y}_{init} \in \mathbb{R}^d$ instead of a time-window that is employed in most time-series forecasting models shown in Fig. 1 (b). The reason that there is only a single step is due to the objective of the emulator, which aims to approximate and accelerate the process-based physical model. As a result, if the physical model is used to generate a time-sequence (e.g., 10 years out of 40 years), then the acceleration of the emulation is bounded by 4 times. Thus, the emulator follows the same procedure as the physical model, where a single initial condition is given, and then the model needs to forecast the rest of the long sequence.
- **Auxiliary environment context variables $\boldsymbol{X}_{1:L}$.** Future environment variables (e.g., climate forcings) are available for the future time steps and broadly used, because in ecological and carbon process modeling researchers and practitioners often need to estimate the future ecosystem and carbon status (e.g., future carbon sequestration potential) under different future climate conditions (e.g., forecasted by CMIP6) in order to inform policy making. As a result, these conditions $\boldsymbol{X}_{1:L}$ are used in the actual physics-based simulation. Thus, the task formulation adopts the same set of inputs as the physical model so it can learn the same process.
- **Impact of changes in initial condition.** Many data samples share the same high-dimensional environmental conditions $\boldsymbol{X}_{1:L} \in \mathbb{R}^{k \times L}$ but have different targets $\{\boldsymbol{Y}_{1:L} \in \mathbb{R}^{d \times L}\}$ (e.g., $L = 468$) due to changes in single-step initial conditions $\{\boldsymbol{Y}_{init} \in \mathbb{R}^d\}$, where $d << k$ (e.g., $d = 7$

and $k = 136$). This requires the emulator to respond well to the initial condition while still being able to utilize the high-dimensional environment conditions $\boldsymbol{X}_{1:L}$ for the long-sequence forecasting and maintain low errors at the end of the sequence.

Related Work

Deep learning has been widely applied to time series forecasting tasks, initially through variations of recurrent and convolutional neural networks (Lim and Zohren 2021). In recent years, transformer-based models have achieved greater success due to their ability to capture long-term dependencies in multivariate forecasting tasks, such as Informer (Zhou et al. 2021), PatchTST (Nie et al. 2022), FedFormer (Zhou et al. 2022), CrossFormer (Zhang and Yan 2023), and TimeXer (Wang et al. 2024a). However, these methods often rely on time-series patterns from a historical window, whereas in ecological process forecasting the initial condition is often only given at the fist time step. Moreover, they do not inherently consider uncertainty in forecasting. Recently, a DeepED model was also developed specifically for ecological process emulation (Wang et al. 2023). However, it is an autoregressive model that is more prone to error accumulation over time and also does not consider uncertainty.

To explicitly model uncertainty, diffusion-based models have gained increasing attention in time series modeling for their strengths in probabilistic forecasting and imputation. TimeGrad (Rasul et al. 2021) first introduced an autoregressive diffusion framework, where future values are generated step-by-step and conditioned on both past observations and auxiliary covariates. However, its autoregressive nature leads to error accumulation over long horizons. CSDI (Tashiro et al. 2021) addressed this by proposing a non-autoregressive alternative that uses flexible input masking strategies, making it applicable to both imputation and forecasting tasks. SSSD (Alcaraz and Strodthoff 2022) improves upon CSDI by replacing its transformer backbone with structured state space models. To further enhance temporal coherence, TimeDiff (Shen and Kwok 2023) incorporates a look-back window to provide initial conditions, specifically addressing boundary inconsistencies in CSDI. This idea is extended by TMDM (Li et al. 2024), which conditions both forward and reverse diffusion steps on historical context for improved stability. In recent years, to disentangle temporal representation, Diffusion-TS (Yuan and Qiao 2024) adopts a Transformer-based encoder-decoder design and introduces Fourier components to improve both interpretability and realism in generated sequences. mr_Diff uses the seasonal-trend decomposition to sequentially extract fine-to-coarse trends (Shen, Chen, and Kwok 2024). Finally, TSDiff (Kollovieh et al. 2023) introduces a self-guidance mechanism for conditioning during inference, enabling better adaptation to downstream tasks across diverse time series applications. However, similar to typical time-series forecasting models, they condition patterns from past time windows (or bidirectional windows in the case of imputation) to infer future values, which is less effective for this problem with only a single-step initial condition. Moreover, they are not explicitly designed to incorporate rich contextual information that are essential in ecosystem modeling with external environmental variables.

Methods: EcoDiffusion

Background: Denoising Diffusion Probabilistic Model (DDPM)

DDPM (Ho, Jain, and Abbeel 2020) is a latent variable model that converts input data $\boldsymbol{Y}^0$ into noise $\boldsymbol{Y}^T$. Let $\boldsymbol{Y}^t$ for $t = 1, \ldots, T$ be a sequence of latent variables in the sample space as $\boldsymbol{Y}^0$. DDPM achieves this through a forward diffusion process q and then reconstructs it via reverse denoising p_θ. In the forward process, at each step t, the noise-added data $\boldsymbol{Y}^t$ is obtained by scaling the previous state $\boldsymbol{Y}^{t-1}$ by $\sqrt{1-\beta_t}$ and adding Gaussian noise with variance $\beta_t \in [0, 1]$:

$$q(\boldsymbol{Y}^t|\boldsymbol{Y}^{t-1}) = \mathcal{N}(\boldsymbol{Y}^t; \sqrt{1-\beta_t}\boldsymbol{Y}^{t-1}, \beta_t\mathbf{I}) \quad (1)$$

This can be rewritten as $q(\boldsymbol{Y}^t|\boldsymbol{Y}^0) = \mathcal{N}(\boldsymbol{Y}^t; \sqrt{\bar{\alpha}_t}\boldsymbol{Y}^0, (1-\bar{\alpha}_t)\mathbf{I})$, where $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, and $\alpha_t = 1 - \beta_t$. Thus, $\boldsymbol{Y}^t$ can be retrieved following $\boldsymbol{Y}^t = \sqrt{\bar{\alpha}_t}\boldsymbol{Y}^0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is noise from $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

The reverse denoising process follows a Markov chain. At the t th step, $\boldsymbol{Y}^{t-1}$ is sampled from a conditional Gaussian distribution given $\boldsymbol{Y}^t$:

$$p_\theta(\boldsymbol{Y}^{t-1}|\boldsymbol{Y}^t) = \mathcal{N}(\boldsymbol{Y}^{t-1}; \boldsymbol{\mu}_\theta(\boldsymbol{Y}^t, t), \Sigma_\theta(\boldsymbol{Y}^t, t)). \quad (2)$$

In this formulation, the variance $\boldsymbol{\Sigma}_\theta(\boldsymbol{Y}^t, t)$ is typically fixed as $\sigma_t^2\mathbf{I}$, while the mean $\boldsymbol{\mu}_\theta(\boldsymbol{Y}^t, t)$ is modeled by a neural network parameterized by θ. In the noise estimation framework, a neural network $\boldsymbol{\epsilon}_\theta$ predicts the noise added to the diffused input $\boldsymbol{Y}^t$, which is then used to compute $\boldsymbol{\mu}_\theta(\boldsymbol{Y}^t, t)$ as:

$$\boldsymbol{\mu}_\theta(\boldsymbol{Y}^t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{Y}^t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{Y}^t, t)\right) \quad (3)$$

Parameter θ is learned by minimizing:

$$\mathcal{L}_\epsilon = \mathbb{E}_{\boldsymbol{Y}^0,\epsilon,t}\left[\|\epsilon - \epsilon_\theta(\boldsymbol{Y}^t, t)\|^2\right].$$

General Multi-Scale Structure: Forecasting and Imputation

Inspired by the success in high-resolution imputation by conditional diffusion based models such as CSDI (Tashiro et al. 2021), we reformulate the long-sequence forecasting problem as a two-step multi-scale process, where the first step performs forecasts on a subset of uniformly sampled time steps (e.g., once per year) whereas the second step performs imputation on other time steps using that as a condition.

While this is conceptually analogous to multi-scale detrending methods (Zhou et al. 2022; Wang et al. 2023), the trend decomposition methods model forecasted values at each time step as a sum of scale-dependent signals from different scales. In comparison, this forecast-and-imputation approach first generates a subset of predictions along the full forecast horizon to capture the general pattern, but these values are modeled as samples along the sequence and are not added as bases for the remaining time steps. Formally, we denote this subset as $\boldsymbol{A}_{1:n_{\text{year}}} \subset \boldsymbol{Y}_{1:L}$ and

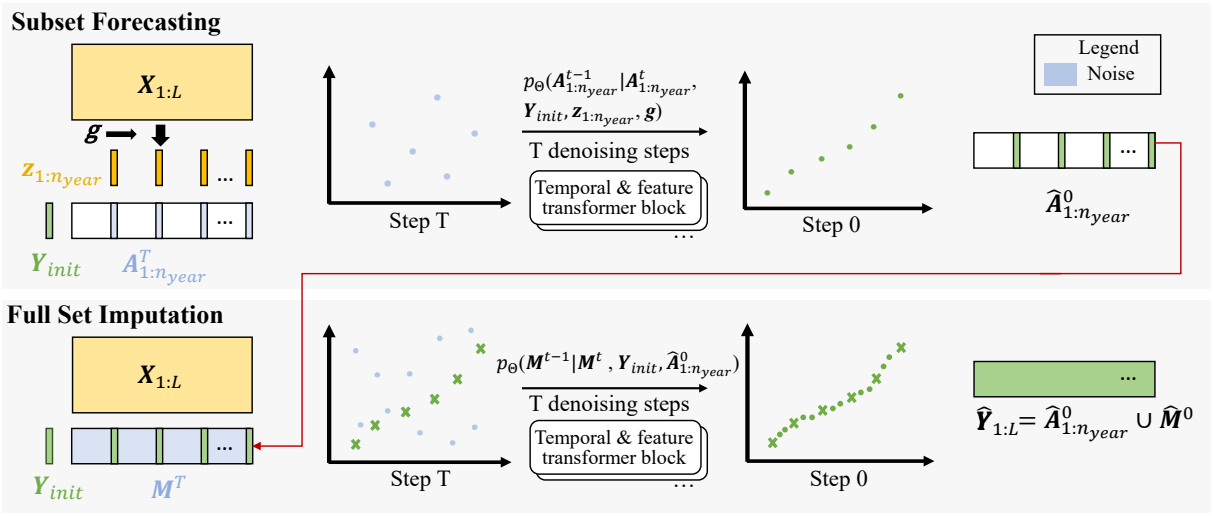

Figure 2: A high-level illustration of the multi-scale structure of EcoDiffusion. In the scatter plot, the number of points in the bottom row (imputation) should be 12 times the above row (from a subset of samples with 12-month distance to full monthly samples), but that will make the points harder to see. Thus, we used a reduced ratio just for illustration purposes.

$\boldsymbol{A}_{1:n_{\text{year}}} \in \mathbb{R}^{d \times n_{\text{year}}} = \{\boldsymbol{Y}_{12 \cdot i}\}, \forall i = 1, ..., n_{year}$, where one time step is sampled per year uniformly over years. In the second step, the other values, denoted as $\boldsymbol{M} = \boldsymbol{Y}_{1:L} \setminus \boldsymbol{A}_{1:n_{year}}$, will be imputed from scratch based on the forecasted samples $\hat{\boldsymbol{A}}_{1:n_{year}}$ and they do not affect the already forecasted values. This imputation modeling instead of aggregation reduces the need for the general trend to be applied at each step and delegates the prediction tasks more directly to the imputation model, making it relatively less sensitive to errors introduced by the high-level trends. In this way, the initial condition can also play a bigger role being one of the subset of samples used as conditions for imputation (i.e., $\boldsymbol{Y}_{init} \cup \hat{\boldsymbol{A}}_{1:n_{year}}$) instead of one of the all samples.

More importantly, this process allows the emulator to rely less on historical patterns in look-back windows but focus more on the environment conditions $\boldsymbol{X}_{1:L}$ at each time step, while still performing predictions at the two scales. In other words, both the forecast and imputation model are less affected by each other, and they need to rely on the corresponding environmental conditions, as shown in Fig. 2.

Overall, for forecasting on the subset, $\boldsymbol{X}_{1:L}$ and $\boldsymbol{Y}_{init}$ will be used as conditions. For imputation, both (1) the environment conditions $\boldsymbol{X}_{1:L}$ and (2) forecasted values with the initial condition $\boldsymbol{Y}_{init} \cup \hat{\boldsymbol{A}}_{1:n_{year}}$ will be used as conditions.

Forecasting: Diffusion with Environment Conditions and Causal Masks

In time series forecasting, we aim to predict the future values $\boldsymbol{A}^0_{1:L}$ given the initial single observations $\boldsymbol{Y}_{init}$.

$$\begin{aligned} &p_\theta(\boldsymbol{A}^{0:T}_{1:n_{\text{year}}} | \boldsymbol{Y}_{init}, \mathbf{z}, \mathbf{g}) \\ =&p_\theta(\boldsymbol{A}^T_{1:n_{\text{year}}}) \prod_{t=1}^{T} p_\theta(\boldsymbol{A}^{t-1}_{1:n_{\text{year}}} | \boldsymbol{A}^t_{1:n_{\text{year}}}, \boldsymbol{Y}_{init}, \mathbf{z}, \mathbf{g}) \end{aligned} \tag{4}$$

$$\mathbf{z}_{1:n_{year}} = \mathcal{F}(\boldsymbol{X}_{1:L}) \tag{5}$$

where $\boldsymbol{A}^T_{1:n_{\text{year}}} \sim \mathcal{N}(0, \mathbf{I})$ is the noise at the beginning of the denoising process; $\mathbf{z} \in \mathbb{R}^{k \times n_{\text{year}}}$ is the latent/learned conditions summarized from the original $\boldsymbol{X}_{1:L} \in \mathbb{R}^{k \times L}$ ($L = n_{year} \times n_{month}$) using a LSTM network F; and $\mathbf{g}$ is an additional global embedding for time, etc. As the performance may vary across different initial conditions (e.g., forest age at the initial step) and climate types, we additionally feed in this information for each sample as part of the global condition $\mathbf{g}$. The climate information is retrieved from the Köppen Climate Classification Map (Beck et al. 2023). Correspondingly, the denoising process at step t is given by:

$$\begin{aligned} &p_\theta(\boldsymbol{A}^{t-1}_{1:n_{\text{year}}} \mid \boldsymbol{A}^t_{1:n_{\text{year}}}, \boldsymbol{Y}_{init}, \mathbf{z}, \mathbf{g}) \\ =&\mathcal{N}\left(\boldsymbol{A}^{t-1}_{1:n_{\text{year}}}; \boldsymbol{\mu}_\theta(\boldsymbol{A}^t_{1:n_{\text{year}}}, t | \boldsymbol{Y}_{init}, \mathbf{z}, \mathbf{g}), \sigma_t^2 \mathbf{I}\right) \end{aligned} \tag{6}$$

During inference, denote $\hat{\boldsymbol{A}}^t_{1:n_{\text{year}}}$ as the generated sample at step t. The process begins by initializing $\hat{\boldsymbol{A}}^T_{1:n_{\text{year}}} \sim \mathcal{N}(0, \mathbf{I})$. Then, by iteratively applying the denoising step (i.e., by predicting the noise) in Equation (6) from $t = T$ down to $t = 1$, we obtain the final generated predictions $\hat{\boldsymbol{A}}^0_{1:n_{\text{year}}}$.

A transformer model is used for the noise prediction. Specifically, for each denoising step, we use a combination of feature and temporal transformers that is commonly employed in diffusion models, as shown in Fig. 2. The temporal transformer learns the temporal attention and the feature transformer learns the relationship between different conditions and between the target variables $\boldsymbol{A}_{1:n_{year}}$ and the conditions $\mathbf{z}_{1:n_{year}}$ generated from $\boldsymbol{X}_{1:L}$, $\boldsymbol{Y}_{init}$ and $\mathbf{g}$. As our model needs to make use of the environmental conditions over time, we apply strict causal masks in the temporal transformer, which is different from several existing diffusion models such as CSDI (Tashiro et al. 2021). For example, when CSDI is used for forecasting, it allows bidirectional temporal attention on available inputs (according

	Height		AGB		Soil		LAI		GPP		NPP		Rh	
Models	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS
Informer	2.83	–	1.13	–	1.15	–	0.57	–	0.34	–	0.17	–	0.18	–
TimeXer	2.85	–	1.17	–	1.20	–	0.60	–	0.40	–	0.20	–	0.20	–
FEDformer	5.29	–	2.25	–	3.31	–	0.89	–	0.61	–	0.30	–	0.27	–
FreTS	3.54	–	1.45	–	1.51	–	0.75	–	0.51	–	0.25	–	0.26	–
DeepED	1.87	–	0.55	–	0.64	–	0.41	–	0.25	–	0.12	–	0.14	–
TimeGrad	4.76	1.70	2.05	0.55	2.23	0.71	0.90	0.27	0.57	0.17	0.28	0.08	0.37	0.10
CSDI	2.43	0.86	1.07	0.34	0.86	0.35	0.57	0.18	0.37	0.11	0.18	0.06	0.19	0.06
TMDM	3.25	1.53	1.36	0.55	1.28	0.61	0.74	0.29	0.48	0.17	0.23	0.08	0.23	0.09
EcoDiffusion	**1.26**	**0.38**	**0.42**	**0.13**	**0.40**	**0.15**	**0.27**	**0.08**	**0.14**	**0.04**	**0.07**	**0.02**	**0.09**	**0.03**

Table 1: Comparison of emulation results at the annual scale (forecasting step) over 39 years. Models with "-" for CRPS do not have uncertainty quantification abilities. Bold numbers indicate the best performances.

	Height		AGB		Soil		LAI		GPP		NPP		Rh	
Models	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS
Informer	3.82	–	1.42	–	1.92	–	1.18	–	0.98	–	0.49	–	0.37	–
TimeXer	2.87	–	1.28	–	1.23	–	0.93	–	0.70	–	0.35	–	0.27	–
FEDformer	6.88	–	3.01	–	3.99	–	1.49	–	1.12	–	0.56	–	0.43	–
FreTS	3.86	–	1.97	–	1.68	–	1.33	–	1.06	–	0.53	–	0.42	–
DeepED	2.20	–	0.72	–	0.95	–	1.39	–	1.16	–	0.58	–	0.46	–
TimeGrad	4.45	2.18	8.37	2.15	4.35	1.75	1.98	0.83	1.62	0.66	0.80	0.32	0.68	0.25
CSDI	2.23	0.81	0.97	0.32	0.82	0.35	0.80	0.30	0.64	0.25	0.32	0.13	0.24	0.10
TMDM	2.99	1.42	1.16	0.48	1.11	0.53	0.95	0.42	0.79	0.34	0.39	0.17	0.30	0.14
EcoDiffusion	**1.26**	**0.50**	**0.42**	**0.16**	**0.38**	**0.18**	**0.36**	**0.15**	**0.23**	**0.09**	**0.11**	**0.04**	**0.10**	**0.05**

Table 2: Comparison of emulation results at the monthly scale (imputation step) over 39 years, i.e., 468 months. Models with "-" for CRPS do not have uncertainty quantification abilities. Bold numbers indicate the best performances.

to the implementation). While this is technically feasible as any already-given inputs (e.g., projected future climate forcings by CMIP-6 for the next decades) or model-generated results are always available at test time, it is not suitable for our problem setting. In particular, environmental conditions $\boldsymbol{X}_{1:L}$ are important input conditions to EcoDiffusion (Fig. 1) and changes in $\boldsymbol{X}_{1:L}$ can have direct and major impact on the forecast values. As a result, without the causal mask, forecastings for earlier years (e.g., 1 to 20) can change if environmental conditions in future years (e.g., 20 to 40) are changed. Thus, we apply strict temporal causal masks (i.e., no attention from a time step to a future input) to avoid logical inconsistency. Moreover, we find that incorporating this constraint improves both the generalizability of EcoDiffusion, as shown later in the experiment results.

This temporal and feature layers form a single transformer block and we stack multiple such blocks, where each block processes the output from the previous one. The final outcome is the predicted noise at the current step t.

Imputation: Heterogeneous Variable Diffusion

For fine-scale imputation, we aim to fill the values at the remaining time steps $\boldsymbol{M} = \boldsymbol{Y}_{1:L} \setminus \boldsymbol{A}_{1:n_{year}}$ given the initial single observations $\boldsymbol{Y}_{init}$ and the predicted values on the subset $\hat{\boldsymbol{A}}_{1:n_{\text{year}}}$:

$$\begin{aligned} &p_\theta(\boldsymbol{M}^{0:T} \mid \boldsymbol{Y}_{init}, \hat{\boldsymbol{A}}_{1:n_{\text{year}}}) \\ =&p_\theta(\boldsymbol{M}^T) \prod_{t=1}^{T} p_\theta(\boldsymbol{M}^{t-1} \mid \boldsymbol{M}^t, \boldsymbol{Y}_{init}, \hat{\boldsymbol{A}}_{1:n_{\text{year}}}) \end{aligned} \tag{7}$$

where $\boldsymbol{M}^T \sim \mathcal{N}(0, \mathbf{I})$. Correspondingly, the denoising process at step t is given by:

$$\begin{aligned} &p_\theta(\boldsymbol{M}^{t-1} \mid \boldsymbol{M}^t, \boldsymbol{Y}_{init}, \hat{\boldsymbol{A}}_{1:n_{\text{year}}}) \\ =&\mathcal{N}\left(\boldsymbol{M}^{t-1};\, \boldsymbol{\mu}_\theta(\boldsymbol{M}^t, t \mid (\boldsymbol{Y}_{init}\hat{\boldsymbol{A}}_{1:n_{\text{year}}}),\, \sigma_t^2 \mathbf{I}\right) \end{aligned} \tag{8}$$

Similarly, during inference, we denote the generated sample at step t as $\hat{\boldsymbol{M}}^t$. The process begins by initializing $\hat{\boldsymbol{M}}^T \sim \mathcal{N}(0, \mathbf{I})$. Then, we gradually generated sample $\hat{\boldsymbol{M}}^0$. By combining $\hat{\boldsymbol{M}}^0$ and $\hat{\boldsymbol{A}}^0_{1:n_{\text{year}}}$, we obtain our final output $\hat{\boldsymbol{Y}}_{1:L}$. At the monthly scale, the output variables exhibit a more immediate and direct response to changes in climate forcings. To better capture this rapid coupling, we hence move the feature transformer layer ahead of the temporal transformer layer to focus more on the local climate forcings, compared to the forecasting diffusion model.

Furthermore, different ecosystem variables exhibit very distinct patterns. For example, vegetation height tends to increase over time whereas other variables such as GPP fluctuates. Thus, we divide the temporal transformer layer into two branches to separately process variables with high monthly variability and those that remain relatively stable over time. Since the overall trend is already established by the yearly model and the input features primarily drive monthly variability, we exclude the climate and age embedding conditions at this stage. For the imputation training, we adopt the masking strategy as CSDI.

	Height		AGB		Soil		LAI		GPP		NPP		Rh	
Models	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS	RMSE	CRPS
EcoDiffusion-X	1.67	0.48	0.47	0.14	0.36	0.13	**0.26**	**0.08**	0.15	**0.04**	**0.07**	**0.02**	**0.08**	**0.02**
EcoDiffusion-causal	1.55	0.44	**0.42**	**0.13**	**0.35**	**0.14**	0.27	**0.08**	**0.14**	**0.04**	**0.07**	**0.02**	0.09	0.03
EcoDiffusion	**1.26**	**0.38**	**0.42**	**0.13**	0.40	0.15	0.27	**0.08**	**0.14**	**0.04**	**0.07**	**0.02**	0.09	0.03

Table 3: Ablation study on annual forecasting component in Table 1.

Model Backbone

Depending on the task, either $\mathbf{z}$ (yearly) or $\boldsymbol{X}$ (monthly) is concatenated with $\boldsymbol{Y}_{init}$ and $\boldsymbol{A}^t_{1:n_{year}}$ (yearly) or $\boldsymbol{Y}_{init} \cup \hat{\boldsymbol{A}}_{1:n_{year}}$ and $\boldsymbol{M}^t$ (monthly). The resulting sequence is then added with the diffusion timestep t via a sinusoidal embedding, which is subsequently passed through a series of residual layers. Each residual layer consists of a two-branch causal masked temporal transformer operating on the predicted target variable, along with a feature-wise attention layer across all predicted variables and the encoded $\mathbf{z}$ (yearly) or $\boldsymbol{X}_{1:L}$ (monthly). Global conditioning variables—including time, feature index, conditional mask, climate type, and initial age—are embedded and added to the output of the transformer layers. More details are provided in the appendix.

Experiments

Candidate methods

We consider a variety of candidate methods as follows, such as widely used time-series forecasting methods and conditional diffusion models, covering designs that are autoregressive/non-autoregressive, multi-scale, ED-tailored, uncertainty-aware, etc. More details of the models and hyperparameter settings (e.g., learning rate) are in the appendix.

- **Informer:** A transformer-based time-series forecasting model with ProbSparse attention to improve parameter efficiency (Zhou et al. 2021).
- **TimeXer:** A transformer-based forecasting model that considers both the relationships between endogenous targets and exogenous dependencies (Wang et al. 2024b).
- **FEDFormer:** A frequency-based transformer with the seasonal-trend decompositions to better capture the global profile of time series (Zhou et al. 2022).
- **FreTS:** A MLP-based model using frequency information to better capture the global specturm of time series, showing comparable performance among transformer-based models while maintaining efficiency (Yi et al. 2023).
- **DeepED:** An autoregressive emulator based on LSTM, which is designed for ED with multi-scale trend decomposition (Wang et al. 2023).
- **TimeGrad:** An autoregressive forecasting model based on conditional diffusion where in each step it make predictions based on the gradients (Rasul et al. 2021).
- **CSDI:** An imputation model based on conditional diffusion where masking strategies are used for training. It has both imputation and forecasting modes and we used the forecasting version here (Tashiro et al. 2021).
- **TMDM:** A transformer-modulated diffusion model for forecasting using relationships between a past window and future values (Li et al. 2024). For this problem the window length is a single step, which is challenging for TMDM-type of models.
- **EcoDiffusion:** Our proposed approach.

Metrics. We consider both metrics for prediction quality and uncertainty. For prediction quality, we use the root-mean-squared-errors (RMSE) and for uncertainty, we use the Continuous Ranked Probability Score (CRPS) (Matheson and Winkler 1976). CRPS measures the difference between the predicted cumulative distribution function (CDF) and the empirical CDF of the observation, and is a commonly employed metric for uncertainty-aware model evaluation.

Data

We generated ED data over the globe, covering different climate conditions, terrains, and forest types, with 15 distinct initial-condition scenarios for each. Test data are sampled at least 1 degree away from the training data, resulting in 852 sites globally. The dataset contains 136 driving features and 7 target variables (Height, AGB, Soil, LAI, GPP, NPP, and Rh). The time span is from 1981 to 2020.

Results

We present the results using three parts: (1) Results at annual level to show if the methods can capture the general trends; (2) Results at monthly level to show if the methods can capture finer-scale changes; and (3) Ablation studies to show the effects of different components of EcoDiffusion.

Yearly Results. Table 1 shows the performance of EcoDiffusion compared with the baseline models. Overall, Ecodiffusion shows the best results with lower RMSE values. Among the other diffusion models, the autoregressive-based TimeGrad has higher errors potentially due to the error accumulation effects. For the other time-series forecasting models, DeepED shows better results, potentially due to its targeted designs of the recurrent network structures based on ED. Interestingly, the frequency-based trend decomposition models do not show improvements and show higher errors. This is likely caused by the trend modeling, which relies on the input conditions. In this task, the conditions can only be on the environment contexts and there is no past windows as inputs, so the trends inferred from environment contexts are likely less relevant for the forecasting. Looking at the uncertainty results, EcoDiffusion also shows distributions most aligned with the target distributions with the lowest CRPS values. The non-diffusion models in the table do not consider uncertainty so their CRPS values are skipped.

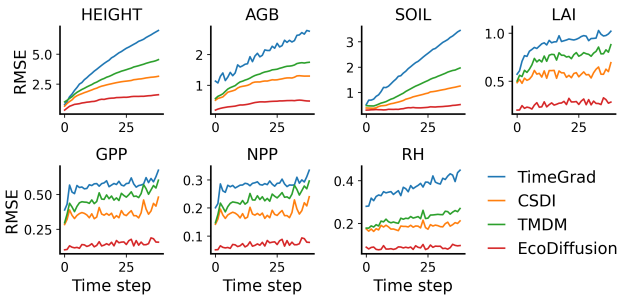

Figure 3: RMSEs over years for candidate models (diffusion-based as examples for visual clarity; more in appendix).

Fig. 3 further shows the RMSE trends over years for different target variables for the diffusion-based models (displaying all together makes the graph hard to read so we use a subset of candidates; more results are available in the appendix). The RMSEs are aggregated over all testing locations. We can see that the patterns are different for different types of variables. As variables such as height and AGB tend to keep increasing over time, it is easier for errors to accumulate as well. Overall, EcoDiffusion shows the least accumulation over years across all the variables.

Monthly Results. Table 2 and Fig. 5 show the results at the monthly level, where the overall rankings of the methods remain the same, with EcoDiffusion consistently showing the best results. According to Fig. 5, at the monthly level, we can see that the errors for several variables get more unstable (e.g., GPP, NPP and Rh), and the pattern is the same for all models. This is likely due to the naturally higher monthly fluctuations of these variables that are harder to capture. This is also reflected by the numbers in the tables. For example, comparing the results of EcoDiffusion in Tables 1 and 2, the RMSE values between annual and monthly predictions are similar for height, AGB, and soil temperature, but increase for LAI, GPP, NPP and Rh.

Furthermore, Fig. 4 shows the forecasting results of several diffusion models for a sample location over time, where the buffer along each predicted line shows the estimated uncertainty range. We can see that buffers along EcoDiffusion can more effectively contain the target values and are tighter around them. More examples are available in the appendix.

Ablation Studies. Table 3 presents the ablation results of the annual forecasting component (i.e., the Table 1 part). Additional ablation results for the monthly imputation component are included in the appendix. We first evaluate a basic EcoDiffusion architecture without a causal mask or global condition as a base with the environment conditions (EcoDiffusion-X). Then, we add the causal mask within the temporal layer (EcoDiffusion-causal), which leads to improvement of quality, particularly for the gradually increasing variables such as height, AGB, and soil carbon. Finally, we include global conditioning by embedding prior knowledge of climate types and initial forest age (used to define initial conditions) into a global context vector $\mathbf{g}$. This particularly enhances the predictions for height, which may be

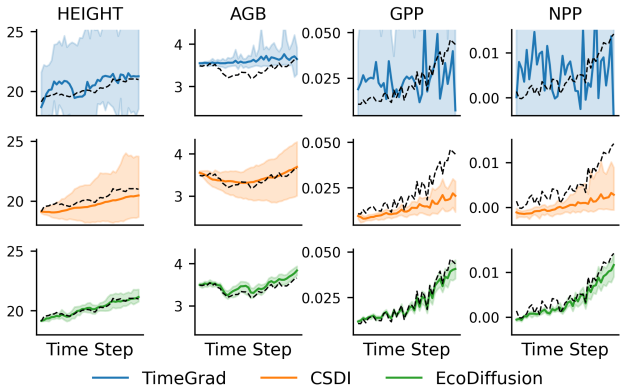

Figure 4: Uncertainty with a sample location over 39 years for a few candidate methods. The dash black lines are target values and buffers show the uncertainty around forecasts. More examples are in the appendix.

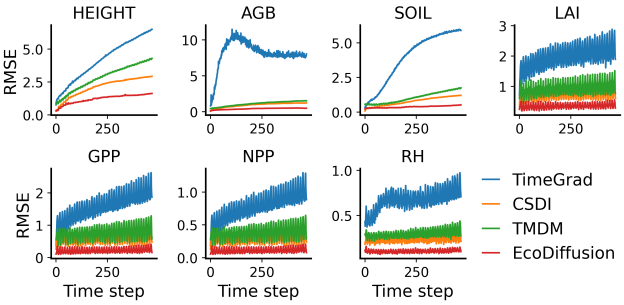

Figure 5: RMSEs over years for candidate models (diffusion-based as examples for visual clarity; more in appendix).

more affected by initial conditions, and the results on the other variables are relatively stable. This suggests future designs can better optimize the usage of different designs to further reduce the errors across the variables. In general, the forecasting components at the annual level are most effective for target variables that have larger values over time and are more prone to error accumulation.

Conclusion

We proposed EcoDiffusion, an uncertainty-aware emulator of the process-based ED model for ecosystem forecasting. EcoDiffusion handles long-term forecasting via designs with subset-forecasting and full set-imputation, which is effective in handling single-step initial conditions with a large number of environmental variables. EcoDiffusion also incorporates strict causal masks and heterogeneous variable separation to enhance the approximation quality and generalizability. The results showed that EcoDiffusion is effective and can significantly reduce errors compared with various types of baselines, with enhanced uncertainty estimation.

In future work, we will examine EcoDiffusion for more challenging scenarios such as cross-domain adaptation and explore different backbone networks for the denoising steps.

Acknowledgments

Ruohan Li, Yiqun Xie, Zhihao Wang, and Zhili Li are supported in part by the NSF under Grant No. 2126474, 2147195, 2425844, and 2530610; NASA under grant 80NSSC25K0013 and 80NSSC25K7221; Google's AI for Social Good Impact Scholars program; and the Zaratan cluster at the University of Maryland. George Hurtt is supported by the NASA CMS under Grant No. 80NSSC25K7221 and NASA EIS under Grant No. 80NSSC22K1733. Lei Ma is supported by the NASA ECIP-ES under Grant No. 80NSSC24K1632. Xiaowei Jia was partially supported by the NSF under Grant No. 2239175, 2147195, 2316305, 2425845, 2530609, 2203581; NASA under Grant No. 80NSSC24K1061 and 80NSSC25K0013; the USGS awards G21AC10564 and G22AC00266; and Pitt Momentum Funds and CRC at the University of Pittsburgh. Gengchen Mai is supported by the NSF under Grant No. 2521631. We would like to acknowledge high-performance computing support from the Derecho system (doi:10.5065/qx9a-pg09) provided by the NSF National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation.

References

Alcaraz, J. M. L.; and Strodthoff, N. 2022. Diffusion-based time series imputation and forecasting with structured state space models. *arXiv preprint arXiv:2208.09399*.

Beck, H. E.; McVicar, T. R.; Vergopolan, N.; Berg, A.; Lutsko, N. J.; Dufour, A.; Zeng, Z.; Jiang, X.; van Dijk, A. I.; and Miralles, D. G. 2023. High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Scientific data*, 10(1): 724.

Bonan, G. B. 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *science*, 320(5882): 1444–1449.

Fisher, R. A.; Koven, C. D.; Anderegg, W. R.; Christoffersen, B. O.; Dietze, M. C.; Farrior, C. E.; Holm, J. A.; Hurtt, G. C.; Knox, R. G.; Lawrence, P. J.; et al. 2018. Vegetation demographics in Earth System Models: A review of progress and priorities. *Global change biology*, 24(1): 35–54.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Hurtt, G. C.; Chini, L.; Sahajpal, R.; Frolking, S.; Bodirsky, B. L.; Calvin, K.; Doelman, J. C.; Fisk, J.; Fujimori, S.; Goldewijk, K. K.; et al. 2020. Harmonization of global land-use change and management for the period 850–2100 (LUH2) for CMIP6. *Geoscientific Model Development Discussions*, 2020: 1–65.

Hurtt, G. C.; Moorcroft, P. R.; And, S. W. P.; and Levin, S. A. 1998. Terrestrial models and global change: challenges for the future. *Global Change Biology*, 4(5): 581–590.

Kollovieh, M.; Ansari, A. F.; Bohlke-Schneider, M.; Zschiegner, J.; Wang, H.; and Wang, Y. B. 2023. Predict, refine, synthesize: Self-guiding diffusion models for probabilistic time series forecasting. *Advances in Neural Information Processing Systems*, 36: 28341–28364.

Lawrence, D. M.; Fisher, R. A.; Koven, C. D.; Oleson, K. W.; Swenson, S. C.; Bonan, G.; Collier, N.; Ghimire, B.; Van Kampenhout, L.; Kennedy, D.; et al. 2019. The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, 11(12): 4245–4287.

Li, Y.; Chen, W.; Hu, X.; Chen, B.; Zhou, M.; et al. 2024. Transformer-modulated diffusion models for probabilistic multivariate time series forecasting. In *The Twelfth International Conference on Learning Representations*.

Lim, B.; and Zohren, S. 2021. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194): 20200209.

Lin, L.; Li, Z.; Li, R.; Li, X.; and Gao, J. 2024. Diffusion models for time-series applications: a survey. *Frontiers of Information Technology & Electronic Engineering*, 25(1): 19–41.

Ma, L.; Hurtt, G.; Ott, L.; Sahajpal, R.; Fisk, J.; Lamb, R.; Tang, H.; Flanagan, S.; Chini, L.; Chatterjee, A.; et al. 2022. Global evaluation of the Ecosystem Demography model (ED v3. 0). *Geoscientific Model Development*, 15(5): 1971–1994.

Matheson, J. E.; and Winkler, R. L. 1976. Scoring rules for continuous probability distributions. *Management science*, 22(10): 1087–1096.

Meijer, C.; and Chen, L. Y. 2024. The rise of diffusion models in time-series forecasting. *arXiv preprint arXiv:2401.03006*.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2022. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*.

Prentice, I. C.; Bondeau, A.; Cramer, W.; Harrison, S. P.; Hickler, T.; Lucht, W.; Sitch, S.; Smith, B.; and Sykes, M. T. 2007. Dynamic global vegetation modeling: quantifying terrestrial ecosystem responses to large-scale environmental change. In *Terrestrial ecosystems in a changing world*, 175–192. Springer.

Rasul, K.; Seward, C.; Schuster, I.; and Vollgraf, R. 2021. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International conference on machine learning*, 8857–8868. PMLR.

Sato, H.; Itoh, A.; and Kohyama, T. 2007. SEIB–DGVM: A new Dynamic Global Vegetation Model using a spatially explicit individual-based approach. *Ecological Modelling*, 200(3-4): 279–307.

Shen, L.; Chen, W.; and Kwok, J. 2024. Multi-resolution diffusion models for time series forecasting. In *The Twelfth International Conference on Learning Representations*.

Shen, L.; and Kwok, J. 2023. Non-autoregressive conditional diffusion models for time series prediction. In *International Conference on Machine Learning*, 31016–31029. PMLR.

Tashiro, Y.; Song, J.; Song, Y.; and Ermon, S. 2021. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in neural information processing systems*, 34: 24804–24816.

Wang, Y.; Wu, H.; Dong, J.; Qin, G.; Zhang, H.; Liu, Y.; Qiu, Y.; Wang, J.; and Long, M. 2024a. Timexer: Empowering

transformers for time series forecasting with exogenous variables. *Advances in Neural Information Processing Systems*, 37: 469–498.

Wang, Y.; Wu, H.; Dong, J.; Qin, G.; Zhang, H.; Liu, Y.; Qiu, Y.; Wang, J.; and Long, M. 2024b. TimeXer: Empowering Transformers for Time Series Forecasting with Exogenous Variables. In Globerson, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J.; and Zhang, C., eds., *Advances in Neural Information Processing Systems*, volume 37, 469–498. Curran Associates, Inc.

Wang, Z.; Xie, Y.; Jia, X.; Ma, L.; and Hurtt, G. 2023. High-fidelity deep approximation of ecosystem simulation over long-term at large scale. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, 1–10.

Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2022. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*.

Yi, K.; Zhang, Q.; Fan, W.; Wang, S.; Wang, P.; He, H.; An, N.; Lian, D.; Cao, L.; and Niu, Z. 2023. Frequency-domain MLPs are More Effective Learners in Time Series Forecasting. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 76656–76679. Curran Associates, Inc.

Yuan, X.; and Qiao, Y. 2024. Diffusion-ts: Interpretable diffusion for general time series generation. *arXiv preprint arXiv:2403.01742*.

Zhang, Y.; and Yan, J. 2023. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The eleventh international conference on learning representations*.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 11106–11115.

Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, 27268–27286. PMLR.