MIRANDA: MId-feature RANk-adversarial Domain Adaptation toward climate change-robust ecological forecasting with deep learning

Yuchang Jiang
yuchang.jiang@uzh.ch

Jan Dirk Wegner
jandirk.wegner@uzh.ch

Vivien Sainte Fare Garnot
vsaint@ics.uzh.ch

DM3L, University of Zurich, Zurich, Switzerland

Abstract

Plant phenology modelling aims to predict the timing of seasonal phases, such as leaf-out or flowering, from meteorological time series. Reliable predictions are crucial for anticipating ecosystem responses to climate change. While phenology modelling has traditionally relied on mechanistic approaches, deep learning methods have recently been proposed as flexible, data-driven alternatives with often superior performance. However, mechanistic models tend to outperform deep networks when data distribution shifts are induced by climate change. Domain Adaptation (DA) techniques could help address this limitation. Yet, unlike standard DA settings, climate change induces a temporal continuum of domains and involves both a covariate and label shift, with warmer records and earlier start of spring. To tackle this challenge, we introduce Mid-feature Rank-adversarial Domain Adaptation (MIRANDA). Whereas conventional adversarial methods enforce domain invariance on final latent representations, an approach that does not explicitly address label shift, we apply adversarial regularization to intermediate features. Moreover, instead of a binary domain-classification objective, we employ a rank-based objective that enforces year-invariance in the learned meteorological representations. On a country-scale dataset spanning 70 years and comprising 67,800 phenological observations of 5 tree species, we demonstrate that, unlike conventional DA approaches, MIRANDA improves robustness to climatic distribution shifts and narrows the performance gap with mechanistic models.

1. Introduction

Plant phenology, the study of seasonal vegetation cycles, focuses on the timing of events such as leaf emergence, flowering, and fruit maturation, which are driven by environmental and climatic factors [32]. Climate change
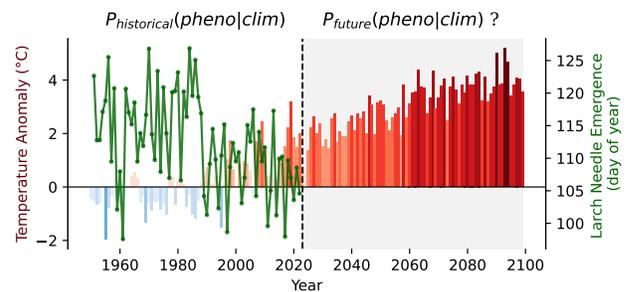

Figure 1. **Climate change-robust machine learning.** Deep learning models trained on historical climatic conditions face a distribution shift when applied to future climate projections. In this paper we consider models that predict plant phenological dates from meteorological time series and devise a domain adaptation method to enhance their robustness to such shifts. (Temperature data: CMIP6 CNRM-CM6-1 SSP2-4.5 [12, 38], phenological data: Swiss Phenology Network [13])

has already induced widespread shifts in phenological timing, providing some of the most visible biological evidence of ongoing global change [8]. These shifts have far-reaching ecological consequences, affecting trophic interactions [3, 40] and ecosystem functioning, and they feed back to the climate system through their influence on biomass production, surface albedo, and carbon and water fluxes [26, 28]. Beyond ecology and climate science, phenology also underpins many Earth observation applications, including crop type mapping, biomass estimation, and species distribution modelling [7, 9, 18, 27]. Reliable phenology models are therefore essential for anticipating the impacts of future climate change on ecosystems, land-climate interactions, and remote sensing applications.

In this work, we consider the task of climate-phenology modelling, which consists of predicting species-specific phenological dates from one year of daily meteorological time series (Fig. 2). Such regression models are typ-

ically trained on phenological observations derived from near-surface and satellite remote sensing [6, 41], or citizen-science networks [13]. Traditionally, phenology modelling has relied on mechanistic, hypothesis-driven models that encode predefined physiological responses to temperature and day length [19]. More recently, deep learning approaches have been proposed, achieving superior performance under stationary climatic conditions [15]. Unlike process-based models, data-driven approaches can flexibly incorporate a broader range of predictors and scale more readily to large spatial extents or diverse species sets [15]. However, deep models are prone to learning correlations that are specific to the historical climate distribution on which they are trained. As a result, they often underperform mechanistic models when evaluated under climatic conditions that deviate substantially from training data [2, 15]. This limitation is critical for ecological forecasting, since future climate projections include environmental conditions that may lie outside the historical training distribution [11], as illustrated in Fig. 1.

Our goal in this paper is to improve the robustness of deep phenology models under such distribution shifts. We frame this challenge as an Unsupervised Domain Adaptation (UDA) problem. Historical observations define a *source* domain, where both meteorological inputs and phenological labels are available. Future climate simulations define a *target* domain, for which meteorological inputs are accessible but phenological labels are unavailable [12]. This setting naturally enables unsupervised adaptation using unlabeled future climate data. However, phenology forecasting presents challenges that are not well addressed by existing UDA methods. Phenology modelling under climate change is subject to both a covariate shift, with a trend of warming climatic records, and a label shift with a trend of advancing date of spring phenological events. While some UDA methods address this combined shift with distribution re-weighting strategies, these frameworks are designed for classification settings and are ill-suited here. Furthermore, climate change induces a temporal continuum of domains rather than two clearly separated domains. Applying existing methods could thus result in data invariances that are harmful to the model's robustness. To address these challenges, we introduce MId-feature RANk-adversarial Domain Adaptation (MIRANDA), to the best of our knowledge, this is the first domain adaptation method designed explicitly for phenology modelling under climate change. To address the combined covariate and label shift, we apply an adversarial regularizer on the intermediate features of a transformer model that enforce domain-invariance. Applying this regularizer here instead of on the final latent features allows for domain-specific variations in the last part of the model, which we further expand with a hybrid normalization layer. Secondly, instead of a binary domain classifica-
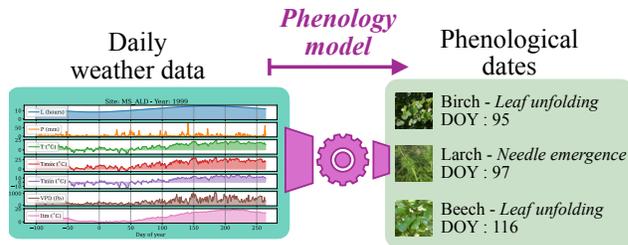

Figure 2. **Phenology modeling (conceptual illustration).** A phenology model takes a set of climatic and environmental covariates and predicts phenological dates. In this paper we consider models operating on daily meteorological time series to predict the date of 50% leaf development of 5 different tree species.

tion objective for the adversarial network, we implement a rank-based loss that pushes the main network to learn year-invariant intermediate features. We emulate the UDA setting using a 70-year historical dataset from a country-scale citizen-science phenology network with $67,800$ different phenological observation. We construct three different dataset splits mimicking three different intensities of distribution shift. Through extensive experiments, we show that, unlike conventional UDA methods, MIRANDA improves performance under strong distribution shifts and narrows the performance gap between deep learning and process-based mechanistic models in extrapolative settings.

2. Related Work

Phenology modelling Process-based models implement clear causal mechanisms linking climatic drivers to plant phenological response [5, 19]. Such models, like the M1 model [4], typically have high interpretability, small number of parameters, and robust performance. However, they hinge on theoretical knowledge of the pathway that links a driver to plant phenology. This requires extensive controlled experiments in growth chambers [16]. In contrast, data-driven approaches can learn predictive relationships directly from observational data, enabling the incorporation of diverse drivers without requiring explicit mechanistic assumptions [10, 21, 29]. Recent deep learning architectures such as PhenoFormer [15] have been developed for plant phenology and showed significantly superior performance than process-based models when tested in stationary climatic conditions. However, when evaluated under shifted climate conditions, mimicking global warming, the M1 process-based models still outperformed PhenoFormer. This challenge is critical in realistic forecasting scenarios where models are trained on historical observations, and deployed on simulated future meteorological data to anticipate phenological trends for the next century [12, 25]. The versatility and scalability of data-driven methods could benefit phenological forecasting, enabling, for example, the inclu-

sion of drivers like water availability which are currently not embedded in process models [26]. However, this climatic robustness challenge needs to be addressed first.

Domain adaptation Unsupervised domain adaptation (UDA) aims to transfer predictive models from a labeled source domain to an unlabeled target domain under distribution shift, and has become a central topic in machine learning when annotation is scarce or expensive [23, 39]. A common principle of many UDA methods is to use the unlabelled target samples to encourage the extraction of *domain-invariant* representations. Indeed, if the representations are domain-invariant, the mapping from these representations to the label space will likely remain valid in the target domain. There are broadly three types of methods to achieve this. *Divergence-based* approaches regularize model training with a loss term which aims at minimizing the statistical divergence between the distributions of source and target representations, such as Maximum Mean Discrepancy [24] or second order statistics [35]. *Adversarial approaches* implement an adversarial training objective where a discriminator network tries to predict the domain of latent representations while the main network tries to make this task impossible, effectively erasing domain-specific patterns in the representations [1, 37]. *Normalization-based* approaches rely on the statistics of normalization layers to make the representations domain-agnostic [30]. Beyond these methods focusing on domain-invariance, some works incorporate self-supervised objectives aiming at learning features that are more robust to domain shifts [31, 36]. Although much of the UDA literature originates from computer vision, domain adaptation techniques have also been explored for time series data, which introduce additional challenges [14, 17, 33]. However, most existing approaches focus on classification tasks or multi-source adaptation, and are thus not directly suited for phenology modeling. Moreover, many UDA methods assume covariate shift, where only the input distribution changes. In phenology forecasting, climatic changes induce not only covariate shift but also observable label shift in phenological dates (see Fig. 4). While some recent works address label-related shifts [14, 17], they do so in classification settings. In contrast, we propose a domain adaptation framework tailored to time-series regression under combined covariate and label shift, specifically designed for phenology modelling, a regression problem.

3. Methods

3.1. Problem statement

Phenology modeling We consider the task of phenology modelling from meteorological time series. A model $\mathbf{F}_\theta$, parametrized by θ, operates on a meteorological time series $X_i \in \mathbb{R}^{C \times T}$, with C the number of climate variables

and T the length of the temporal sequence to predict a set of phenological dates $\hat{Y}_i \in \mathbb{R}^S$ with S the number of different tree species under consideration (across our experiments $T = 365$, $C = 7$, and $S = 5$). Formally, this is a multivariate regression problem from multi-temporal inputs. The dataset is organized in *site-years*: the same sites are monitored over multiple years and sample i correspond to the meteorological conditions and phenological observations of site s_i during year g_i.

Unsupervised domain adaptation We cast the task of future phenology projection as an unsupervised domain adaptation problem. In this setting, we make use of a *source* dataset $\mathcal{D}_\sigma$ and a *target* dataset $\mathcal{D}_\tau$. The source dataset is fully labeled: $\mathcal{D}_\sigma = \{(X_i, Y_i) \in \mathbb{R}^{C \times T} \times \mathbb{R}^S\}_{i=1}^{N_\sigma}$, with N_σ the number of training samples and Y_i the observed phenophase dates for sample i. In contrast, only unlabeled inputs are available in the target domain $\mathcal{D}_\tau = \{(X_i) \in \mathbb{R}^{C \times T}\}_{i=1}^{N_\tau}$, with N_τ the number of target samples.

3.2. Model architecture

We use PhenoFormer [15], an established architecture for the task at hand, as backbone for our method. PhenoFormer is a lightweight architecture composed of a shared linear encoder, a transformer encoder layer with learnt tokens, and a linear decoder. Our only modification is to use two consecutive transformer layers instead of one, leaving all other hyperparameters unchanged.

$$X_i \xrightarrow{\mathbf{e}} E_i \in \mathbb{R}^{D \times T} \tag{1}$$

$$[E_i + p_i, \mathbf{L}] \xrightarrow{\mathbf{t_1}} [H_i, Z_i] \in \mathbb{R}^{D \times (T+S)} \tag{2}$$

$$[H_i, Z_i] \xrightarrow{\mathbf{t_2}} [\tilde{H}_i, G_i] \in \mathbb{R}^{D \times (T+S)} \tag{3}$$

$$G_i \xrightarrow{\mathbf{d}} \hat{Y}_i \in \mathbb{R}^S \tag{4}$$

A given input multi-variate time series X_i is firstly embedded by a linear layer $\mathbf{e}$, applied in parallel to all time points, resulting in a sequence of embedding vectors E_i of dimension D, with D a hyperparameter. Temporal positional encodings p_i are added to these embeddings, and learnt tokens $\mathbf{L}$ are appended to the sequence. We use one learnt token per predicted species so $\mathbf{L} \in \mathbb{R}^{D \times S}$. The resulting sequence is processed by two successive transformer encoder layers $\mathbf{t}_1$ and $\mathbf{t}_2$. Lastly, the output embeddings $G_i \in \mathbb{R}^{D \times S}$ corresponding to the learnt tokens' positions are used as global embedding for each species and decoded into a date prediction with a linear decoder layer $\mathbf{d}$. See Garnot et al. [15] for more details on this architecture.

3.3. MIRANDA

We introduce MIRANDA a domain adaptation method specifically designed for climate change robustness in phenology modelling. As shown in Fig. 3, our method relies
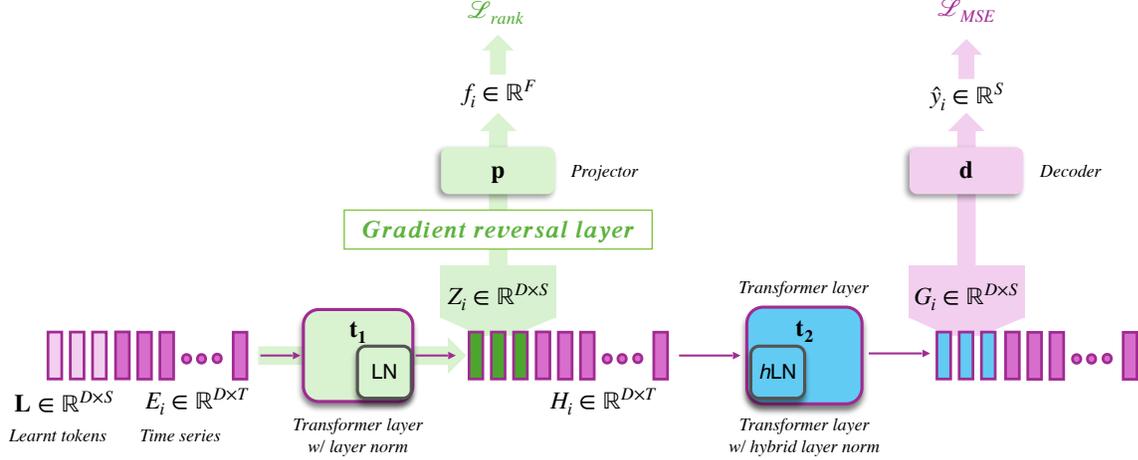
Figure 3. **Method overview**. Our framework addresses domain shifts in phenology modelling via two key components: rank-based adversarial training on intermediate features and hybrid layer normalization. Purple elements correspond to the **main phenology prediction pathway**: the learnable tokens $\mathbf{L}$ are concatenated with the input time series E_i and processed by two transformer encoder layers ($\mathbf{t_1}$ and $\mathbf{t_2}$). The resulting global embeddings of the learnable tokens, denoted as G_i, are then fed into the decoder $\mathbf{d}$ to predict the target date using the regression loss $\mathcal{L}_{\mathrm{MSE}}$. We apply **rank-based adversarial learning** on the mid-level features Z_i (green), where a discriminator $\mathbf{p}$ is trained with a ranking loss $\mathcal{L}_{\mathrm{rank}}$ through a gradient reversal layer to encourage domain-invariant mid-level features. Meanwhile, we replace the standard layer normalization in t_2 with our **hybrid layer normalization** (blue), which preserves domain-dependent variations in high-level representations.

on 1. a rank-based adversarial objective that is better suited to the continuously changing climatic shift, 2. the adversarial regularization of the intermediate (instead of late) features, which only enforces domain-invariance in the earlier part of the network, 3. a hybrid normalisation layer in the last transformer layer of the architecture, which further promotes better extrapolation.

Rank-based adversarial objective Classical adversarial domain adaptation formulates the discriminator objective as binary domain classification: the discriminator predicts whether a feature originates from the source or the target domain, while the feature extractor is trained to confuse it. This formulation implicitly assumes two clearly separated and discrete domains. However, as visible in Fig 1, in climate-phenology forecasting, domain shift occurs progressively over time, resulting in a continuum of intermediate regimes rather than a sharp source-target dichotomy. Moreover, what constitutes a domain-shift might be different for different tree species. Collapsing this structure into a binary objective discards valuable information about the relative proximity between climatic conditions over time. To better capture this gradual shift, we replace the binary adversarial objective with a Rank-N-Contrast loss [42]. Instead of enforcing indiscriminate domain invariance, this objective encourages features to be more similar when they are coming from closer years, and less similar otherwise. This ranking-based formulation naturally accommodates continuous domain shifts and provides a smoother regular-

ization signal than binary discrimination.

Within a mini-batch of size N, each pair of feature vectors (f_i, f_j), is encouraged by a contrastive term, to be closer than a set of negative pairs. The negative set $S_{i,j}$ is defined as the set of samples k that are more distant in time than samples i and j:

$$S_{i,j} = \{f_k \mid k \neq i, \; |g_i - g_k| \geq |g_i - g_j|\}, \quad (5)$$

where g_i is the year of observation of sample i. We use the cosine similarity $\mathrm{sim}(\cdot, \cdot)$ to measure *closeness* in feature space such that the contrastive term for sample i is :

$$l_{\mathrm{rank}}^{(i)} = \frac{1}{N} \sum_{j=1, \, j \neq i}^{N} - \log \frac{\exp\left(\mathrm{sim}(f_i, f_j)/\tau\right)}{\sum_{f_k \in S_{i,j}} \exp\left(\mathrm{sim}(f_i, f_k)/\tau\right)}, \quad (6)$$

with τ is a temperature hyperparameter. The final adversarial loss $\mathcal{L}_{\mathrm{rank}}$ is then:

$$\mathcal{L}_{\mathrm{rank}} = \frac{1}{N} \sum_{i=1}^{N} l_{\mathrm{rank}}^{(i)} . \quad (7)$$

Mid-feature regularization Unlike prior adversarial adaptation methods such as DANN [1], which apply domain discrimination on late features (G_i), we perform adversarial alignment on the mid-level representations (Z_i) of the learnt tokens. Indeed, if the late features were perfectly aligned across domains, the distributions of the predictions would also be aligned across domains, and thus not be able

to track the label shift. Instead, we propose to only enforce domain invariance in the intermediate representations of our model. Specifically, we flatten $Z_i \in \mathbb{R}^{D \times S}$ into a vector and process it with a multi-layer perceptron $\mathbf{p}$ to obtain an embedding $f_i \in \mathbb{R}^F$, with F a hyperparameter. The discriminator $\mathbf{p}$ is trained to minimize $\mathcal{L}_{rank}$, while a gradient reversal layer (GRL) from DANN [1] is inserted between $\mathbf{p}$ and the feature extractor $\mathbf{t_1}$. As a result, $\mathbf{t_1}$ is optimized adversarially to produce embeddings that cannot be separated according to their year order, leading to domain-invariant mid-level representations. The overall training objective is defined as

$$\mathcal{L} = \mathcal{L}_{\mathrm{MSE}} + \lambda \mathcal{L}_{\mathrm{rank}}, \tag{8}$$

where $\mathcal{L}_{\mathrm{MSE}}$ is the supervised regression loss computed on labeled source samples,

$$\mathcal{L}_{\mathrm{MSE}} = \|y - \hat{y}\|_2^2, \tag{9}$$

and $\mathcal{L}_{\mathrm{rank}}$ is applied to both source and target samples since it does not rely on phenology labels. Each mini-batch is composed of an equal number of source and target samples.

Hybrid layer normalization While rank-based adversarial training enforces domain invariance at the mid-level representations, the final task-specific features should retain sufficient flexibility to accommodate label shifts in phenological dates. To enhance flexibility, we propose *hybrid layer normalization* (hLN), which replaces the standard layer normalization in the second transformer layer t_2. Our approach is inspired by [30]. Given a generic input feature vector $x \in \mathbb{R}^D$, standard layer normalization computes the mean and standard deviation from the current sample:

$$\mu(x) = \frac{1}{D} \sum_{d=1}^{D} x_d, \qquad \sigma(x) = \sqrt{\frac{1}{D} \sum_{d=1}^{D} (x_d - \mu(x))^2} \tag{10}$$

and applies normalization as

$$\mathrm{LN}(x) = \gamma \frac{x - \mu(x)}{\sigma(x)} + \beta, \tag{11}$$

where γ and β are learnable affine parameters. Our hLN follows the same standard functioning in the source domain and differs in the target domain. hLN maintains *source-domain population statistics* during training in a running-average manner similar to batch normalization. At inference time on the target domain, hLN uses these running statistics to normalize target feature vectors.

For each training batch, we compute the layer normalization statistics $\mu(x)$ and $\sigma(x)$ of the source domain samples and update the running estimates as:

$$\mu_s \leftarrow (1-m)\mu_s + m\,\mu(x), \qquad \sigma_s \leftarrow (1-m)\sigma_s + m\,\sigma(x), \tag{12}$$

where m is the momentum coefficient. During inference, hLN normalizes target features using the source-domain running statistics (μ_s, σ_s):

$$h\mathrm{LN}(x) = \gamma \frac{x - \mu_s}{\sigma_s} + \beta, \tag{13}$$

This domain-specific normalization, enables the late features G_i of the target domain to have a different range than those of the source domain.

4. Experiments

4.1. Data and domain definition

PhenoFormer dataset. We use the dataset introduced in [15], based on phenological records from the Swiss Phenology Network (SPN) [13]. The dataset comprises 67,800 observations collected from 175 sites across Switzerland over up to 70 years. We focus on a subset of five tree species that offers the longest record: horse chestnut, european beech, european larch, common spruce, and hazel. We predict the date of needle emergence for european larch and common spruce, and the date of leaf unfolding for the other species. Each sample includes daily meteorological time series with 365 data points for seven variables between day of year -100 and 265 of the year at hand. The meteorological variables cover temperature, precipitation, pressure, and day length.

Dataset splits We construct three dataset splits. For all splits, the training set is treated as the source domain, while the validation and test sets are treated as target domains.

- **Chronological** First, we partition the dataset chronologically: observations from 1951–2002 are used for training, 2003–2012 for validation, and 2013–2022 for testing. As shown in Fig. 4, this split induces distribution shift, where the mean annual temperature of the test set is $0.90\,^\circ\mathrm{C}$ higher than the training set, corresponding to a phenological date shift (e.g., larch needle emergence occurs 6.22 days earlier on average).
- **Annual temperature** Second, we rank years by their mean annual temperature. We use the 15% warmest years as test set, the next warmest 15% as validation, and the remaining as training. This creates a stronger shift with a $+1.23\,^\circ\mathrm{C}$ difference between training and testing, and an almost 8 day advancement of larch needle emergence.
- **Elevation** For the third configuration, we split the data based on elevation to simulate stronger climate changes. We rank all site-years by elevation – available in the site metadata – and assign the highest 25% to the test set, the next 15% to validation, and the remaining to training. This setting yields the most challenging shift among all splits: as shown in Fig. 4, both the temperature and phenology date distribution differ substantially, and reach values outside of the training range.
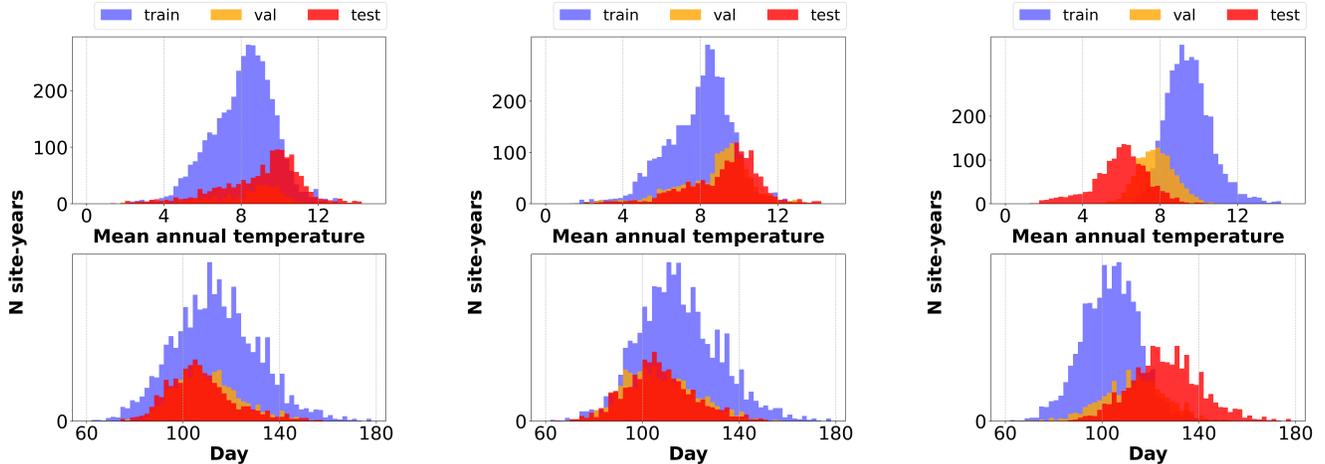
(a) Chronological split.
$\Delta T = 0.90°C, \Delta Date = -6.22\ day$

(b) Annual temperature split
$\Delta T = 1.23°C, \Delta Date = -7.92\ day$

(c) Elevation split
$\Delta T = -3.51°C, \Delta Date = 22.96\ day$

Figure 4. **Domain distribution shifts.** We show the histograms of input (top) and predicted (bottom) variables on the train, validation, and test sets for each of the three dataset splits. We plot the histogram of annual temperature in the top row to represent the climatic distribution, and the date of Larch needle emergence in the bottom row to illustrate the shift in predicted variables.

4.2. Implementation details

All models are trained with MSE loss and optimized using Adam [20] with learning rate 10^{-4} and batch size 16. We use vector dimensions of $D = 64$ and $F = 128$. The discriminator **p** consists of a three-layer MLP with ReLU activations. For our method, we set $\tau = 0.1$ in $\mathcal{L}_{rank}$ and adopt the dynamic λ scheduler from [1]. Models are trained for up to 300 epochs with model selection based on validation performance. Each experiment is repeated 10 times with different random initializations, and we report the mean results across runs. We evaluate performance using the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE). All experiments are conducted on one NVIDIA GeForce RTX 3080 GPU. Additional implementation details are provided in the Supplementary Material (Sec. S1).

4.3. Results

Competing methods We evaluate MIRANDA against five other domain adaptation approaches and a process model. We compare against divergence-based (CORAL [35]), adversarial (DANN [1], ADDA [37]), and normalization-based approaches (DANL [30], AdaBN [22]). We also compare against a vanilla PhenoFormer without adaptation. For fairness the vanilla model is set with two transformer layers as in MIRANDA , instead of one. Lastly, we evaluate the M1 model [4] which was the strongest process-based model in a previous study using the same dataset [15].

Main results Overall, MIRANDA achieves the strongest overall performance among deep learning methods. On the chronological and annual temperature splits (moderate shifts), MIRANDA achieves competitive or superior results compared to all baselines. In the elevation split, the most challenging setting with substantial climatic and phenological distribution shift, MIRANDA significantly outperforms prior adaptation methods. Compared to the vanilla PhenoFormer, MIRANDA improves R^2 by approximately 1% (chronological), 2% (annual temperature), and 5% (elevation). In contrast, existing adaptation methods degrade severely under elevation shift. For example, DANN drops from $R^2 = 0.49$ in chronological to $R^2 = -0.20$ in elevation split. Similar instability is observed for ADDA and CORAL. These results suggest that conventional domain adaptation techniques, which primarily align final-layer representations or rely solely on normalization adjustments, are insufficient for phenology forecasting under climate change. The process-based M1 model outperforms all deep learning approaches including ours, showcasing the robustness of mechanistic models. In terms of R^2, MIRANDA closes the performance gap on the chronological split, and reduces it to $2 - 3\%$ on the other two splits. Notably, our method also exhibits lower standard deviation across runs and species than M1 (e.g., 0.04 vs. 0.15 on the elevation split). This suggests that, although M1 can perform well in settings aligned with its mechanistic assumptions, its behavior is less consistent under unseen conditions. In contrast, MIRANDA achieves comparable robustness with greater stability. This brings deep phenol-

Table 1. **Main results of three data splits.** We compare MIRANDA with PhenoFormer and five domain adaptation baselines across three data splits representing progressively stronger distribution shifts: Chronological, Annual Temperature, and Elevation. For reference, we also report the performance of the process-based model (M1), highlighted in blue. Evaluation metrics include R^2 ($\uparrow$), RMSE ($\downarrow$), and MAE ($\downarrow$). Results are averaged over 10 repetitions and 5 tree species, and we also report the standard deviation of R^2 in brackets.

Dataset split → Model	1. Chronological			2. Annual Temperature			3. Elevation			Training time (min)
	$R^2 \uparrow$	RMSE $\downarrow$	MAE $\downarrow$	$R^2 \uparrow$	RMSE $\downarrow$	MAE $\downarrow$	$R^2 \uparrow$	RMSE $\downarrow$	MAE $\downarrow$	
PhenoFormer Vanilla [15]	0.50 (0.01)	8.96	6.82	0.41 (0.02)	9.56	7.26	0.17 (0.02)	11.67	9.00	15
+DANN [1]	0.49 (0.01)	9.08	6.98	0.38 (0.06)	9.78	7.44	-0.20 (0.21)	14.03	11.32	17
+ADDA [37]	0.23 (0.15)	11.14	8.80	0.12 (0.10)	11.64	9.18	-0.61 (0.16)	16.22	13.33	26
+CORAL [34, 35]	**0.51** (0.01)	**8.93**	**6.79**	0.42 (0.02)	9.50	7.22	-0.16 (0.15)	13.77	10.98	18
+DANL [30]	0.48 (0.02)	9.17	7.03	0.30 (0.03)	10.44	8.06	-0.01 (0.10)	12.90	10.24	18
+AdaBN [22]	0.49 (0.01)	9.05	6.89	0.40 (0.04)	9.64	7.34	-0.14 (0.09)	13.71	10.92	15
MIRANDA (ours)	**0.51** (0.01)	8.97	6.80	**0.43** (0.02)	**9.43**	**7.16**	**0.22** (0.04)	**11.32**	**8.63**	26
M1 [4]	0.51 (0.05)	8.64	6.77	0.45 (0.05)	9.40	7.04	0.25 (0.15)	11.56	8.53	150

ogy models closer to the extrapolation properties traditionally associated with process-based models while maintaining the flexibility of data-driven modelling. For example process models require one model to be trained per species, while MIRANDA trains a single model to predict the 5 species' dates. As shown on the last column of Tab. 1 this results in a $6\times$ shorter training time. Additionally, while the M1 model is only based on temperature and day length features, MIRANDA also takes into account the pressure and precipitation variables. This did not improve performance on our dataset, but other potential drivers such as soil water content could be easily tested using our framework.

Per-species performance We further examine the per-species performance of PhenoFormer and MIRANDA in Fig. 5. Under the chronological split, the two models perform comparably across species. In the annual temperature split, the gains become more noticeable for european beech (BEE), and horse chestnut to a lesser extent. The most substantial differences appear in the elevation split, which represents the strongest climatic shift. Here, MIRANDA consistently outperforms PhenoFormer across all species, though the magnitude of improvement varies. For example, for horse chestnut (HCH), R^2 increases modestly from approximately 0.25 to 0.26, whereas for european larch (LAR), the improvement is much larger, rising from approximately 0.12 to 0.23. These variations highlight the fact that species respond differently to climatic distribution shifts. Across species and splits, our method is either on par or improves over the vanilla model.

Qualitative results Fig. 6 provides a qualitative comparison of prediction distributions before and after adaptation in the elevation split. Wihout MIRANDA , the vanilla model's prediction saturate and struggle to predict phenological dates later than day of year 150 leading to systematic underestimation. In comparison, the predctions of MIRANDA do not seem to saturate and display an improved
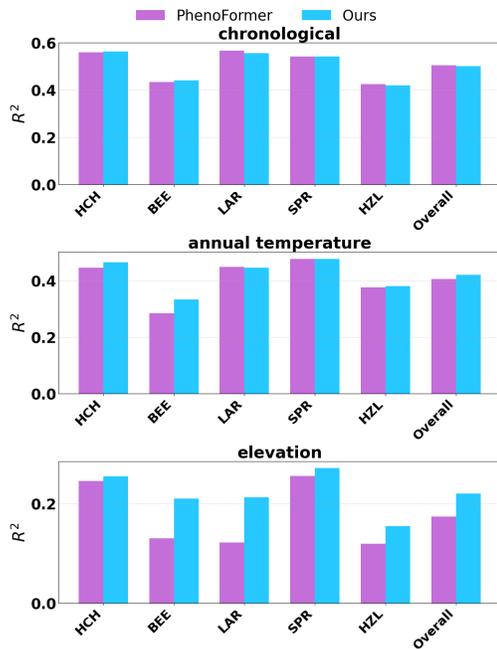

Figure 5. **Per-species performance** Histogram of R^2 per species for PhenoFormer and our model. Five species are: horse chestnut (HCH), european beech (BEE), european larch (LAR), common spruce (SPR), and hazel (HZL).

correlation with reference dates. This illustrates how our method successfully addresses the label shift induced by climate change on phenology modelling.

4.4. Ablation study

We conduct an ablation study to analyze the contribution of each component in our method. Starting from the PhenoFormer backbone with DANN [1] adversarial training as the baseline, we progressively introduce three components: 1) mid-level feature alignment (mid-feat) with binary domain classification as adversarial task, 2) rank-based adver-

Table 2. Ablations study of MIRANDA . Starting from the PhenoFormer backbone and DANN adversarial training, we progressively introduce (1) mid-level feature alignment (mid-feat), (2) rank-based adversarial objective $\mathcal{L}_{\text{rank}}$, and (3) hybrid layer normalization (hLN). Results demonstrate that each component contributes to improved robustness under domain shift, with the largest gains observed under the challenging Elevation split.

| Dataset split $\rightarrow$	1. Chronological			2. Annual Temperature			3. Elevation		
	$R^2 \uparrow$	RMSE $\downarrow$	MAE $\downarrow$	$R^2 \uparrow$	RMSE $\downarrow$	MAE $\downarrow$	$R^2 \uparrow$	RMSE $\downarrow$	MAE $\downarrow$
PhenoFormer Vanilla	0.50	8.96	6.82	0.41	9.56	7.26	0.17	11.67	9.00
+DANN [1]	0.49	9.08	6.98	0.38	9.78	7.44	-0.20	14.03	11.32
✓mid-feat	0.51	8.96	6.82	0.40	9.59	7.23	0.12	12.04	9.38
✓mid-feat ✓$\mathcal{L}_{\text{rank}}$	0.50	8.97	6.82	0.42	9.45	7.18	0.19	11.54	8.87
✓mid-feat ✓hLN	0.50	8.97	6.81	0.42	9.51	7.22	0.18	11.66	8.97
MIRANDA	0.51	8.97	6.80	0.43	9.43	7.16	0.22	11.32	8.63
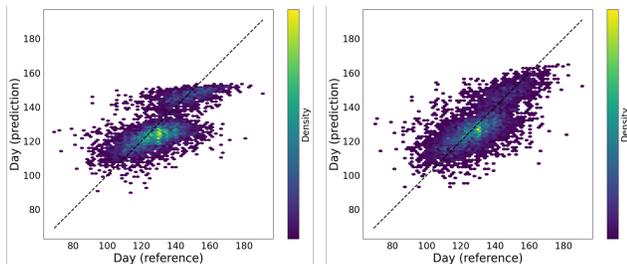
Figure 6. Predicted (y-axis) vs true dates (x-axis) **before (left, with R^2 = 17%) and after (right, with R^2 = 22%) adaptation on the elevation split.** Note that we plot the dates of the 5 species together, which explains the different modes of the distribution.

sarial training $\mathcal{L}_{\text{rank}}$, 3) hybrid layer normalization (hLN). Our method combines all three components. Results are reported in Tab. 2.

Mid-level feature alignment is critical Replacing high-level adversarial alignment (DANN) with mid-level feature alignment (✓mid-feat) is the single most important improvement. This suggests that aligning intermediate representations is more effective than aligning final high-level features when facing severe climatic distribution shift. Conventional adversarial adaptation methods largely assume covariate shift, where the input distribution changes while the label distribution remains stable. Under this assumption, aligning final high-level features is sufficient. However, in phenology forecasting, domain shift also manifests as clear label shift (as shown in Fig. 4). In such cases, forcing alignment at the final representation level may be counterproductive and limit prediction range. Instead, enforcing domain-invariance only on intermediate features seems to strike a good balance to both address the climatic covariate shift, while also leaving space for domain-specific variations. This helps explain why domain adaptation methods that focus solely on high-level feature alignment perform

poorly in Tab. 1.

Rank-based adversarial objective and hybrid layer normalization Introducing the rank-based adversarial loss $\mathcal{L}_{\text{rank}}$ improves performance beyond standard binary adversarial training. By providing structured supervision for domain-invariant features, it yields more informative mid-level alignment, leading to consistent gains across splits. The effect is most pronounced under the elevation split, where R^2 increases from 0.12 to 0.19, surpassing the baseline PhenoFormer. Adding the hybrid normalization layer (hLN) also enhances robustness. By applying source-domain normalization statistics to both domains, hLN enforces consistent feature scaling while preserving domain-specific variations.

5. Conclusion

In this work, we addressed climatic distribution shift in deep phenology modelling through a domain adaptation framework tailored to this setting. Across multiple distribution shift settings, our method outperformed existing adversarial, divergence-based, and normalization-based adaptation approaches. We showed that applying adversarial regularization to intermediate features is key to handling both climate covariate and phenological label shift. Our hybrid normalization layer further improved the model's ability to predict phenological dates outside the range observed in the source domain. Finally, we demonstrated that a rank-based adversarial objective encouraging year-invariant representations is better suited to phenology forecasting than conventional binary domain classification. Altogether, these contributions narrow the robustness gap between deep learning and process-based models, which have traditionally demonstrated stronger extrapolation under shifted climatic conditions. These findings encourage further exploration of this problem in order to achieve even stronger data-driven performance with for example the inclusion of additional drivers, or exploring fourrier-based temporal features for improved robustness.

By enhancing stability and transferability of data-driven phenology models, our work opens the way for a new set of tools for climate change-robust ecological forecasting.

References

[1] Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, and Mario Marchand. Domain-adversarial neural networks. *arXiv preprint arXiv:1412.4446*, 2014. 3, 4, 5, 6, 7, 8, 1

[2] Daphné Asse, Christophe F Randin, Marc Bonhomme, Anne Delestrade, and Isabelle Chuine. Process-based models outcompete correlative models in projecting spring phenology of trees in a future warmer climate. *Agricultural and Forest Meteorology*, 285:107931, 2020. 2

[3] Michael W Belitz, Elise A Larsen, Allen H Hurlbert, Grace J Di Cecco, Naresh Neupane, Leslie Ries, Morgan W Tingley, Robert P Guralnick, and Casey Youngflesh. Potential for bird–insect phenological mismatch in a tri-trophic system. *Journal of Animal Ecology*, 94(4):717–728, 2025. 1

[4] Klaus Blümel and Frank-M Chmielewski. Shortcomings of classical phenological forcing models and a way to overcome them. *Agricultural and Forest Meteorology*, 164:10–19, 2012. 2, 6, 7, 1

[5] Amelia Caffarra, Alison Donnelly, and Isabelle Chuine. Modelling the timing of betula pubescens budburst. ii. integrating complex effects of photoperiod into process-based models. *Climate research*, 46(2):159–170, 2011. 2

[6] Jose A Caparros-Santiago, Victor Rodriguez-Galiano, and Jadunandan Dash. Land surface phenology as indicator of global terrestrial ecosystem dynamics: A systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 171:330–347, 2021. 2

[7] Li Chen, Hui Lin, Jiangping Long, Zhaohua Liu, Peisong Yang, and Tingchen Zhang. Evaluating the transferability of spectral variables and prediction models for mapping forest aboveground biomass using transfer learning methods. *Remote Sensing*, 15(22):5358, 2023. 1

[8] Elsa E Cleland, Isabelle Chuine, Annette Menzel, Harold A Mooney, and Mark D Schwartz. Shifting plant phenology in response to global change. *Trends in ecology & evolution*, 22(7):357–365, 2007. 1

[9] Michele Croci, Manuele Ragazzi, Alessandro Grassi, Giorgio Impollonia, Stefano Amaducci, et al. Assessing the temporal transferability of machine learning models for predicting processing pea yield and quality using sentinel-2 and era5-land data. *Smart Agricultural Technology*, 12(1):1–20, 2025. 1

[10] Wujun Dai, Huiying Jin, Yuhong Zhang, Tong Liu, and Zhiqiang Zhou. Detecting temporal changes in the temperature sensitivity of spring phenology with global warming: Application of machine learning in phenological model. *Agricultural and Forest Meteorology*, 279:107702, 2019. 2

[11] R Deepa, Vinay Kumar, and Suchithra Sundaram. A systematic review of regional and global climate extremes in cmip6 models under shared socio-economic pathways. *Theoretical and Applied Climatology*, 155(4):2523–2543, 2024. 2

[12] Veronika Eyring, Sandrine Bony, Gerald A Meehl, Catherine A Senior, Bjorn Stevens, Ronald J Stouffer, and Karl E Taylor. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016. 1, 2

[13] Federal Office of Meteorology and Climatology MeteoSwiss. Swiss Phenology Network. https://www.meteoswiss.admin.ch/services-and-publications/applications/measurement-values-and-measuring-networks.html#param=messnetz-phaenologie&table=false, 2022. 1, 2, 5

[14] David Gackstetter, Kang Yu, and Marco Körner. Self-attention and frequency-augmentation for unsupervised domain adaptation in satellite image-based time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 224:113–132, 2025. 3

[15] Vivien Sainte Fare Garnot, Lynsay Spafford, Jelle Lever, Christian Sigg, Barbara Pietragalla, Yann Vitasse, Arthur Gessler, and Jan Dirk Wegner. Deep learning meets tree phenology modelling: Phenoformer versus process-based models. *Methods in Ecology and Evolution*, 16(7):1489–1506, 2025. 2, 3, 5, 6, 7, 1

[16] Heikki Hänninen, Koen Kramer, Karen Tanino, Rui Zhang, Jiasheng Wu, and Yongshuo H Fu. Experiments are necessary in process-based tree phenology modelling. *Trends in Plant Science*, 24(3):199–209, 2019. 2

[17] Huan He, Owen Queen, Teddy Koker, Consuelo Cuevas, Theodoros Tsiligkaridis, and Marinka Zitnik. Domain adaptation for time series under feature and label shifts. In *International conference on machine learning*, pages 12746–12774. PMLR, 2023. 3

[18] Hauke Hoppe, Peter Dietrich, Philip Marzahn, Thomas Weiß, Christian Nitzsche, Uwe Freiherr von Lukas, Thomas Wengerek, and Erik Borg. Transferability of machine learning models for crop classification in remote sensing imagery using a new test methodology: A study on phenological, temporal, and spatial influences. *Remote Sensing*, 16(9):1493, 2024. 1

[19] Koen Hufkens, David Basler, Tom Milliman, Eli K Melaas, and Andrew D Richardson. An integrated phenology modelling framework in r. *Methods in Ecology and Evolution*, 9(5):1276–1285, 2018. 2, 1

[20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6, 1

[21] Sehyun Lee, Sujong Jeong, Chang-Eui Park, and Jongho Kim. A simple method of predicting autumn leaf coloring date using machine learning with spring leaf unfolding date. *Asia-Pacific Journal of Atmospheric Sciences*, pages 1–8, 2022. 2

[22] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. *arXiv preprint arXiv:1603.04779*, 2016. 6, 7, 1

[23] Xiaofeng Liu, Chaehwa Yoo, Fangxu Xing, Hyejin Oh, Georges El Fakhri, Je-Won Kang, and Jonghye Woo. Deep

unsupervised domain adaptation: A review of recent advances and perspectives. *APSIPA Transactions on Signal and Information Processing*, 11(1):1, 2022. 3

[24] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015. 3

[25] Stephanie Pau, Elizabeth M Wolkovich, Benjamin I Cook, T Jonathan Davies, Nathan JB Kraft, Kjell Bolmgren, Julio L Betancourt, and Elsa E Cleland. Predicting phenology by integrating ecology, evolution and climate science. *Global change biology*, 17(12):3633–3643, 2011. 2

[26] Shilong Piao, Qiang Liu, Anping Chen, Ivan A Janssens, Yongshuo Fu, Junhu Dai, Lingli Liu, XU Lian, Miaogen Shen, and Xiaolin Zhu. Plant phenology and global climate change: Current progresses and challenges. *Global change biology*, 25(6):1922–1940, 2019. 1, 3

[27] Raquel Ponti and Marco Sannolo. The importance of including phenology when modelling species ecological niche. *Ecography*, 2023(4):e06143, 2023. 1

[28] Andrew D Richardson, T Andy Black, Philippe Ciais, Nicolas Delbart, Mark A Friedl, Nadine Gobron, David Y Hollinger, Werner L Kutsch, Bernard Longdoz, Sebastiaan Luyssaert, et al. Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1555):3227–3246, 2010. 1

[29] VF Rodriguez-Galiano, M Sanchez-Castillo, J Dash, and PM Atkinson. Modelling anomalies in the spring and autumn land surface phenology of the european forest. *Biogeosciences Discussions*, 12(14), 2015. 2

[30] Rob Romijnders, Panagiotis Meletis, and Gijs Dubbelman. A domain agnostic normalization layer for unsupervised adversarial domain adaptation. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1866–1875. IEEE, 2019. 3, 5, 6, 7, 1

[31] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self supervision. *Advances in neural information processing systems*, 33:16282–16292, 2020. 3

[32] Mark Donald Schwartz et al. *Phenology: an integrative environmental science*. Springer, 2003. 1

[33] Yongjie Shi, Xianghua Ying, and Jinfa Yang. Deep unsupervised domain adaptation with time series sensor data: A survey. *Sensors*, 22(15):5507, 2022. 3

[34] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pages 443–450. Springer, 2016. 7

[35] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *AAAI*, 2016. 3, 6, 7, 1

[36] Yu Sun, Eric Tzeng, Trevor Darrell, and Alexei A Efros. Unsupervised domain adaptation through self-supervision. *arXiv preprint arXiv:1909.11825*, 2019. 3

[37] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017. 3, 6, 7, 1

[38] Aurore Voldoire, David Saint-Martin, Stéphane Sénési, Bertrand Decharme, A Alias, Matthieu Chevallier, Jeanne Colin, J-F Guérémy, Martine Michou, M-P Moine, et al. Evaluation of cmip6 deck experiments with cnrm-cm6-1. *Journal of Advances in Modeling Earth Systems*, 11(7): 2177–2213, 2019. 1

[39] Garrett Wilson and Diane J Cook. A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(5):1–46, 2020. 3

[40] Yingying Xie, Hanna T Thammavong, and Daniel S Park. The ecological implications of intra-and inter-species variation in phenological sensitivity. *New Phytologist*, 236(2): 760–773, 2022. 1

[41] Adam M Young, Thomas Milliman, Koen Hufkens, Keith L Ballou, Christopher Coffey, Kai Begay, Michael Fell, Mostafa Javadian, Alison K Post, Christina Schädel, et al. Tracking vegetation phenology across diverse biomes using version 3.0 of the phenocam dataset. *Earth System Science Data*, 17(11):6531–6556, 2025. 2

[42] Kaiwen Zha, Peng Cao, Jeany Son, Yuzhe Yang, and Dina Katabi. Rank-n-contrast: Learning continuous representations for regression. *Advances in Neural Information Processing Systems*, 36:17882–17903, 2023. 4

MIRANDA: <u>MI</u>d-feature <u>RAN</u>k-adversarial <u>D</u>omain <u>A</u>daptation toward climate change-robust ecological forecasting with deep learning

Supplementary Material

1. Implementation details

We follow the implementation of PhenoFormer [15] as our backbone and use two transformer encoder layers (t_1 and t_2). All models are trained using the MSE loss and optimized with Adam [20] with a learning rate of 10^{-4}, batch size of 16, and up to 300 epochs. All domain adaptation methods adopt PhenoFormer as the backbone. For approaches involving adversarial training, the discriminator p is implemented as a three-layer MLP with ReLU activations. Model-specific implementation details are provided below.

- **DANN** [1]: We adopt Phenoformer as the backbone and attach a domain discriminator to the high-level features extracted from the last transformer encoder layer. The discriminator is trained with the standard binary domain classification objective, while the feature extractor is optimized adversarially via a gradient reversal layer, following the original DANN formulation.
- **ADDA** [37]: We initialize ADDA with the pretrained Phenoformer and follow the same experimental protocol as the other adaptation methods. Specifically, we fine-tune the encoder and train a domain discriminator to align feature representations in an adversarial manner.
- **CORAL** [35]: We apply the CORAL loss to the high-level features extracted from the last transformer encoder layer, prior to the regression head, using the original formulation of the CORAL objective. For a fair comparison, we set the CORAL weighting coefficient λ to be the same as that used for our $\mathcal{L}_{\text{rank}}$.
- **DANL** [30]: The original DANL method introduces domain-agnostic normalization for batch normalization layers in convolutional architectures under the DANN framework. Since Phenoformer is transformer-based and relies on layer normalization, we adapt DANL by replacing the standard layer normalization layers in the transformer encoder with our domain-agnostic layer normalization, while keeping the rest of the DANN training procedure unchanged.
- **AdaBN** [22]: AdaBN was originally proposed for models with batch normalization, where target-domain statistics are used during inference. To incorporate AdaBN into Phenoformer, we insert a BatchNorm1d layer before the linear decoder and update its running statistics using unlabeled target samples during adaptation.
- **M1** [4]: We use the implementation of M1 in the R package, Phenor [19] and follow the same implementation setup specified in [15].

- **MIRANDA** : We set $\tau = 0.1$ in $\mathcal{L}_{\text{rank}}$ and adopt the dynamic λ scheduler from [1]. For the embeddings used for rank loss, we set the dimension to 128.

2. Additional experiment results

2.1. Guidance for rank-based adversarial training

We investigate difference guidance for rank-based adversarial training: besides using year information as guidance, we explore the effect of using annual temperature and elevation as the guidance.

Table 3. Different guidance for rank-based adversarial training under the Structured Temporal split.

Guidance	$R^2 \uparrow$	RMSE $\downarrow$	MAE $\downarrow$
Annual temperature	0.50	9.01	6.85
Elevation	0.50	9.01	6.89
Year (ours)	**0.51**	**8.97**	**6.80**

Table 4. Different guidance for rank-based adversarial training under the Annual Temperature split.

Guidance	$R^2 \uparrow$	RMSE $\downarrow$	MAE $\downarrow$
Annual temperature	0.43	9.44	7.18
Elevation	0.42	9.44	7.19
Year (ours)	**0.43**	**9.43**	**7.16**

Table 5. Different guidance for rank-based adversarial training under the Elevation split.

Guidance	$R^2 \uparrow$	RMSE $\downarrow$	MAE $\downarrow$
Annual temperature	0.22	11.36	8.67
Elevation	0.19	11.55	8.86
Year (ours)	**0.22**	**11.32**	**8.63**

2.2. Additional qualitative results

Detailed per-species scatter plots of PhenoFormer and our method of elevation data split are shown in Figure 1 to 5 .
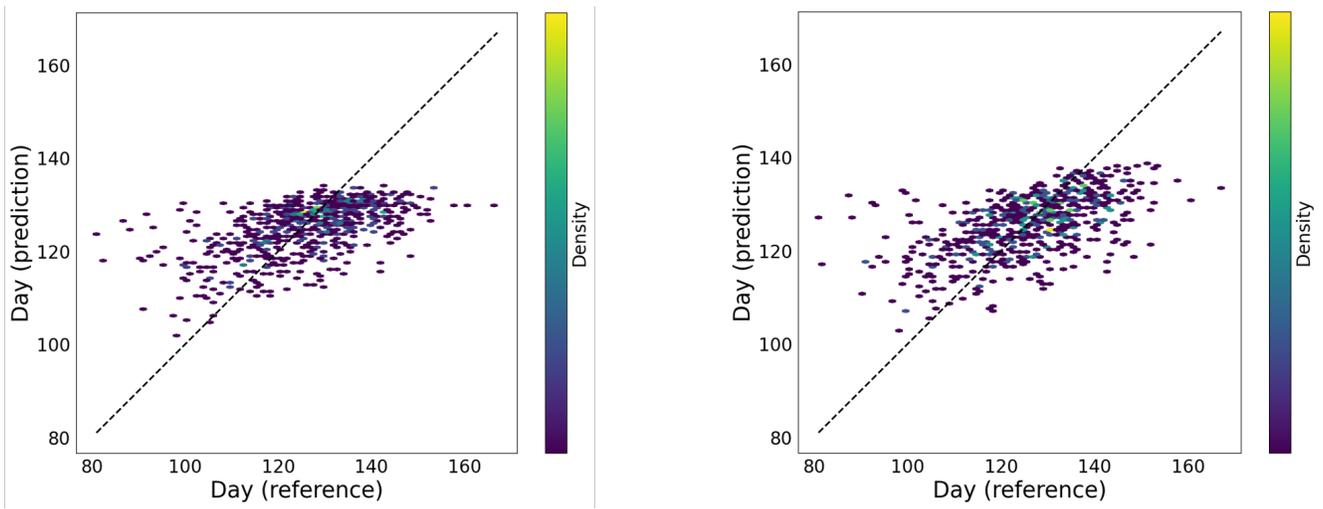
Figure 1. Scatter plot of predicted (y-axis) vs true dates (x-axis) **before (left, with R^2 = 24%) and after (right, with R^2 = 26%) adaptation** for phenology task leaf unholding of horse chestnut in elevation data split.
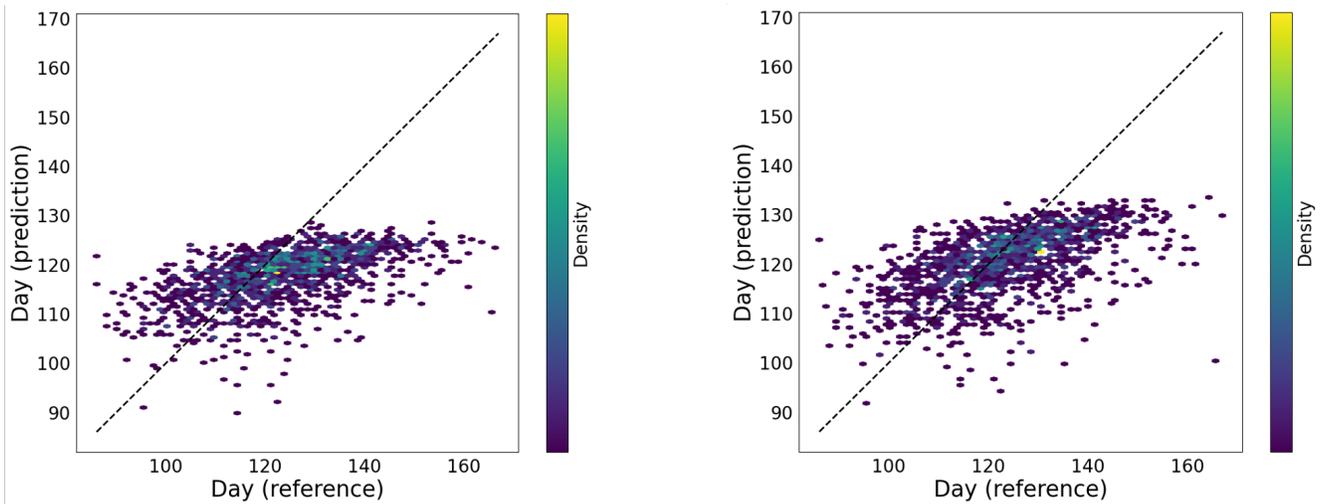

Figure 2. Scatter plot of predicted (y-axis) vs true dates (x-axis) **before (left, with R^2 = 12%) and after (right, with R^2 = 23%) adaptation** for phenology task needle emergence of European larch in elevation data split.
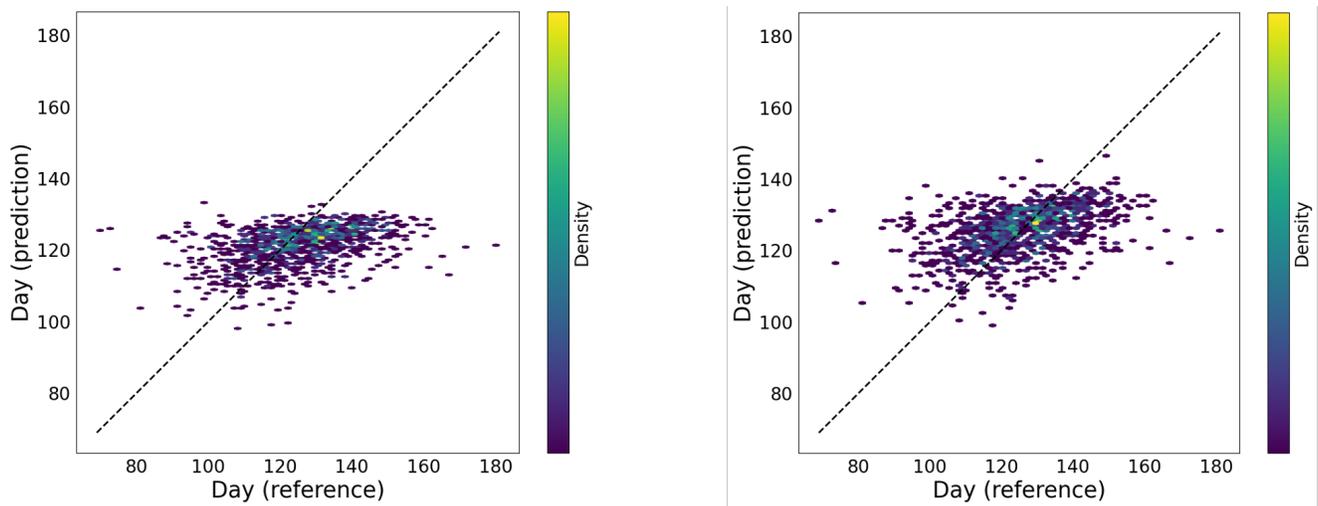
Figure 3. Scatter plot of predicted (y-axis) vs true dates (x-axis) **before (left, with $R^2 = 12\%$) and after (right, with $R^2 = 15\%$) adaptation** for phenology task leaf unholding of hazel in elevation data split.
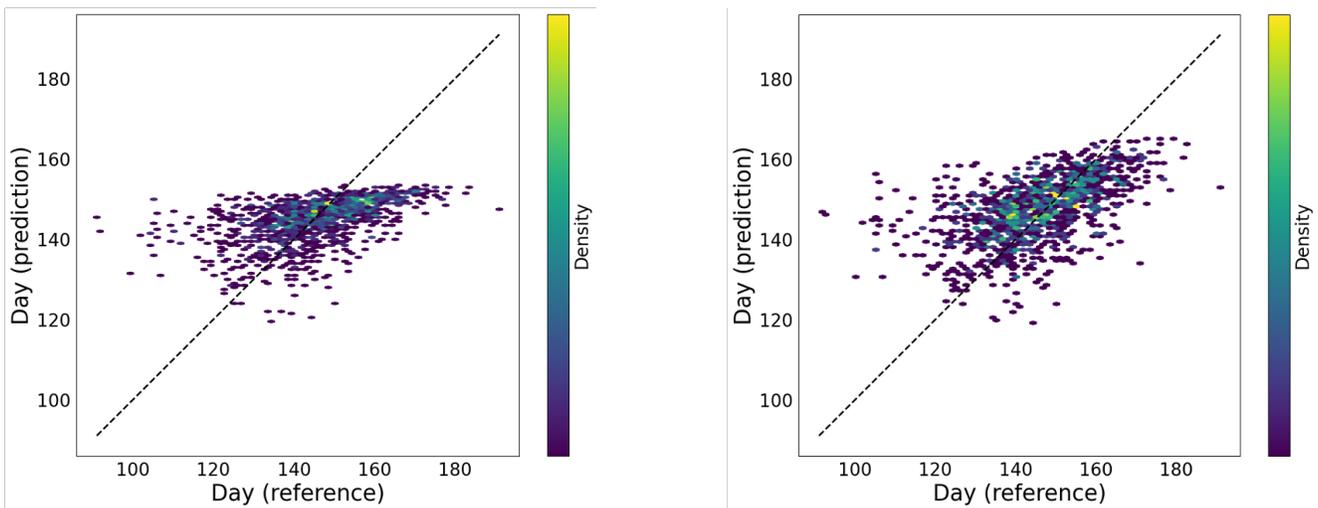

Figure 4. Scatter plot of predicted (y-axis) vs true dates (x-axis) **before (left, with $R^2 = 25\%$) and after (right, with $R^2 = 28\%$) adaptation** for phenology task needle emergence of common spruce in elevation data split.
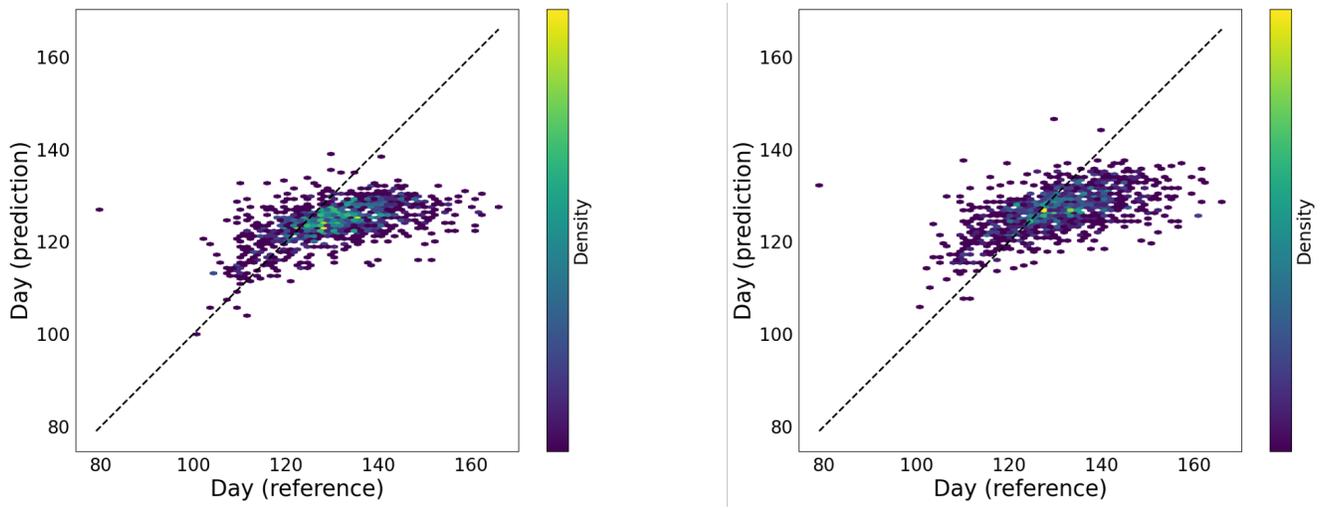
Figure 5. Scatter plot of predicted (y-axis) vs true dates (x-axis) **before (left, with R^2 = 13%) and after (right, with R^2 = 20%) adaptation** for phenology task leaf unfolding of European beech in elevation data split.