PHYRESTORE: PHYSICS-STRUCTURED LATENT-FACTOR RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Estimating temporal soil-loss change is challenging when physically meaningful input factors are noisy or corrupted, particularly because substantial changes are rare relative to the large number of locations exhibiting little change. We study this problem through the Revised Universal Soil Loss Equation (RUSLE) and introduce PhyRestore, a physics-structured latent-factor restoration framework. Rather than directly predicting soil-loss change or correcting a degraded physical estimate, PhyRestore restores corrupted physical factors and reconstructs temporal change through the known physical relationship. We evaluate PhyRestore in a watershed-scale bitemporal raster setting under isolated and simultaneous corruption of rainfall erosivity and cover management, comparing it with the degraded RUSLE estimate and Direct RF, XGBoost, MLP, and CNN models. Factor restoration improves high-magnitude recovery when the corrupted factors remain identifiable, but its advantage weakens under joint corruption, sparse positive extremes, and factor values outside the training support.

1 INTRODUCTION

Because direct soil-erosion measurements are typically limited to localized plots or field sites and are difficult to extrapolate reliably, watershed-scale assessments commonly combine empirical models with geospatial observations to produce spatially distributed estimates (Stroosnijder, 2005; Boix-Fayos et al., 2006; Phinzi & Ngetar, 2019; Benavidez et al., 2018). The Revised Universal Soil Loss Equation (RUSLE) is widely used for this purpose, integrating factors representing rainfall erosivity, soil erodibility, topography, cover management, and support practices (Renard et al., 1997; Benavidez et al., 2018). Although its factorized formulation is interpretable and practical when dense field observations are unavailable, its estimates remain sensitive to errors in the underlying factors (Benavidez et al., 2018; Phinzi & Ngetar, 2019).

We study this problem in a temporal setting using co-registered factor rasters from 2017 and 2022 over a Colorado watershed. The prediction target is the signed difference between the clean RUSLE estimates for the two years and therefore represents a RUSLE-derived temporal soil-loss proxy rather than independently measured soil-loss change. When a time-varying factor is corrupted at either date, the resulting error propagates through the corresponding annual estimate and can distort both the magnitude and direction of the estimated temporal change.

Three learning pathways are possible: directly predict the clean temporal target, learn a residual or gated correction to the degraded RUSLE estimate, or restore the corrupted factor before applying the known RUSLE relationship. These pathways differ in whether learning replaces the analytical estimate, corrects it after error propagation, or operates on the corrupted factor itself.

The temporal soil-loss changes are strongly imbalanced: most locations exhibit changes near zero, while large positive or negative changes are comparatively rare. Learning the final temporal target is therefore dominated by small-change samples, making rare high-magnitude changes particularly difficult to recover (Ribeiro & Moniz, 2020; Yang et al., 2021) and allowing strong aggregate performance over the full spatial support to obscure substantial errors in the changes of greatest magnitude.

These considerations motivate PhyRestore, which estimates the clean value of a corrupted RUSLE factor from bitemporal spatial context and reconstructs signed temporal change through RUSLE. We evaluate it under isolated and simultaneous factor corruption against the degraded analytical

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

estimate and Direct, formula-correction, and trust-correction strategies implemented with RF, XGBoost, MLP, and CNN models.

Our contributions are:

- We formulate signed temporal soil-loss prediction under controlled corruption of year-specific physical factors and compare analytical, Direct, formula-correction, trust-correction, and factor-restoration pathways under a common spatial evaluation protocol.
- We introduce **PhyRestore**, a physics-structured framework that restores corrupted physical factors from bitemporal spatial observations and reconstructs temporal change through the known RUSLE relationship.
- We characterize when factor-level restoration improves rare high-magnitude recovery and when its advantage weakens because corrupted factors are insufficiently recoverable or lie outside the training support.

2 Related Work

RUSLE-Based Soil-Loss Estimation and Temporal Analysis. RUSLE uses factors representing rainfall erosivity, soil erodibility, topography, cover management, and support practices to estimate spatial soil-loss patterns (Renard et al., 1997; Benavidez et al., 2018), supporting regional and continental assessments and analyses of environmental change (Panagos et al., 2015; Borrelli et al., 2017). Remote sensing and geographic information systems increase factor availability (Li et al., 2023), while machine learning has been used to estimate erosion susceptibility, analyze driving factors, and improve RUSLE input layers (Ge et al., 2023; Samarinas et al., 2024). Because these applications remain sensitive to factor quality (Benavidez et al., 2018), we examine signed temporal-change prediction when specific year-dependent factors are corrupted.

Physics-Guided and Hybrid Scientific Machine Learning. Scientific machine learning incorporates domain knowledge through theory-guided modeling (Karpatne et al., 2017), differential-equation constraints (Raissi et al., 2019), and hybrid analytical–learned systems (Karniadakis et al., 2021; Willard et al., 2022). PhyRestore follows the hybrid paradigm but uses RUSLE as a fixed reconstruction rather than an optimization constraint, restoring corrupted physical factors before analytical reconstruction rather than learning a post-hoc correction after factor error has propagated through RUSLE.

Learning with Corrupted or Missing Covariates. Prior work addresses imperfect inputs through robust prediction, denoising, and imputation, including marginalized corrupted-feature methods that account for stochastic covariate perturbations during training (van der Maaten et al., 2013) and GAIN, which reconstructs missing values from partially observed data (Yoon et al., 2018). Unlike generic imputation, PhyRestore supervises factor-specific restoration and restores each affected factor separately under simultaneous corruption.

Imbalanced Regression and Rare Extreme Recovery. Imbalanced-regression objectives and aggregate metrics can be dominated by densely sampled target regions, motivating weighting, resampling, specialized objectives, and representation learning for rare values (Ribeiro & Moniz, 2020; Yang et al., 2021). Here, temporal changes concentrate near zero; PhyRestore addresses this imbalance indirectly by restoring the corrupted factor rather than reweighting the final target.

3 Temporal Prediction under Factor Corruption

3.1 Temporal RUSLE-Derived Target

Let $\mathbf{s} \in \Omega$ denote a valid spatial location and $t \in \{2017, 2022\}$ the observation year. RUSLE represents soil loss using rainfall erosivity R, soil erodibility K, the topographic factor LS, cover management C, and the support-practice factor P. In our study, R and C are year specific, K and LS are shared across both years, and $P = 1$ (İpek & Kahya, 2026). The factor rasters are derived from PRISM precipitation for R (Daly et al., 2002), USDA SSURGO soils for K (U.S. Department of Agriculture, Natural Resources Conservation Service, 2023), USGS 3DEP elevation data for

LS (U.S. Geological Survey, 2021), and Sentinel-2 land-cover data for C (Karra et al., 2021). The annual RUSLE-derived soil-loss proxy is therefore

$$A_t(\mathbf{s}) = R_t(\mathbf{s})K(\mathbf{s})LS(\mathbf{s})C_t(\mathbf{s}). \tag{1}$$

The prediction target is the signed change between the clean 2022 and 2017 estimates:

$$T(\mathbf{s}) \equiv \Delta A_{\text{clean}}(\mathbf{s}) = A_{2022}(\mathbf{s}) - A_{2017}(\mathbf{s}) = K(\mathbf{s})LS(\mathbf{s})\big[R_{2022}(\mathbf{s})C_{2022}(\mathbf{s}) - R_{2017}(\mathbf{s})C_{2017}(\mathbf{s})\big]. \tag{2}$$

Positive values indicate an increase in the RUSLE-derived soil-loss proxy from 2017 to 2022, whereas negative values indicate a decrease. Thus, $T(\mathbf{s})$ is a factor-derived reference quantity rather than an independently observed field measurement of soil loss.

3.2 Controlled Factor Corruption

We distinguish the clean factors defining the fixed target from the observations available to the predictor. We corrupt R_{2022} and C_{2017} because they are distinct year-specific factors appearing in opposite terms of Eq. 2; their combination therefore perturbs both terms. Uncertainty in the static K and LS factors is outside this study's scope.

Rainfall erosivity is degraded using multiplicative Gaussian noise:

$$\widetilde{R}_{2022}(\mathbf{s}) = R_{2022}(\mathbf{s})\big(1 + \epsilon_R(\mathbf{s})\big), \qquad \epsilon_R(\mathbf{s}) \sim \mathcal{N}(0, 0.15^2). \tag{3}$$

The 2017 cover-management factor is degraded using additive Gaussian noise followed by clipping to its valid range:

$$\widetilde{C}_{2017}(\mathbf{s}) = \text{clip}\big(C_{2017}(\mathbf{s}) + \epsilon_C(\mathbf{s}), 0, 1\big), \qquad \epsilon_C(\mathbf{s}) \sim \mathcal{N}(0, 0.08^2). \tag{4}$$

Perturbations are generated independently across raster cells before spatial samples are extracted. A fixed realization is reused across spatial folds, model seeds, model families, and prediction strategies, ensuring that all methods receive identical degraded observations. We consider four conditions: *Clean*, with no corruption; R_{2022}, with only 2022 rainfall erosivity corrupted; C_{2017}, with only 2017 cover management corrupted; and *Mixed*, with both factors corrupted. All untargeted factors retain their clean values. Corruption affects only the predictor inputs; the reference target in Eq. 2 remains fixed across conditions.

3.3 Prediction under Corrupted Factors

Let $\widetilde{\mathbf{X}}(\mathbf{s})$ denote the bitemporal factor observations available around location $\mathbf{s}$ under a given condition. The task is to estimate the fixed clean target $T(\mathbf{s})$ without access to the clean value of any corrupted factor. The following pathways differ in how learning is combined with the known RUSLE relationship.

Degraded analytical prediction. The analytical reference evaluates RUSLE directly using the available factors:

$$F(\mathbf{s}) \equiv \Delta\widetilde{A}_{\text{formula}}(\mathbf{s}) = \widetilde{A}_{2022}(\mathbf{s}) - \widetilde{A}_{2017}(\mathbf{s}), \tag{5}$$

where each $\widetilde{A}_t(\mathbf{s})$ is computed from the factor values available at year t. Under corruption, factor error therefore propagates directly through RUSLE.

Learned temporal prediction. Direct prediction bypasses the analytical estimate, Formula Correction adds a learned residual to it, and Trust Correction modulates that residual using a learned bounded gate:

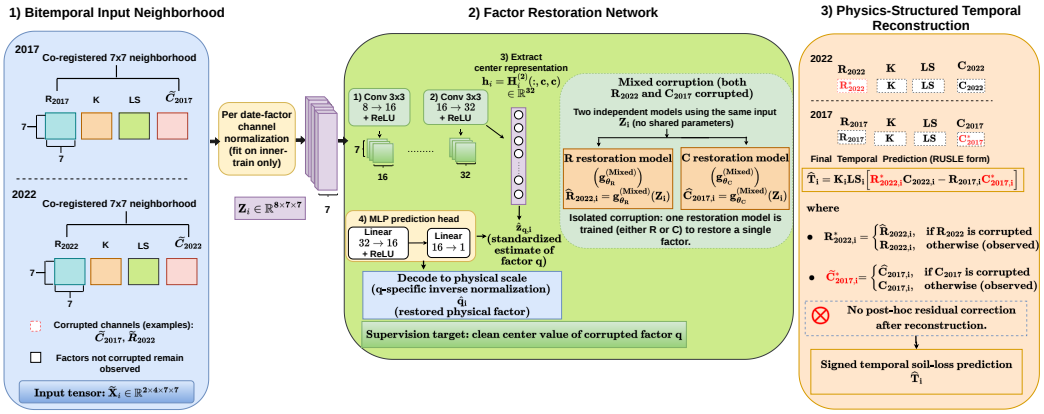


Figure 1: PhyRestore factor restoration and RUSLE-based temporal reconstruction. Under Mixed corruption, separate models restore R_{2022} and C_{2017}.

$$\widehat{T}_{\text{Direct}}(\mathbf{s}) = f_\theta\big(\widetilde{\mathbf{X}}(\mathbf{s})\big), \tag{6}$$

$$\widehat{T}_{\text{Correction}}(\mathbf{s}) = F(\mathbf{s}) + \delta_\phi\big(\widetilde{\mathbf{X}}(\mathbf{s})\big), \tag{7}$$

$$\widehat{T}_{\text{Trust}}(\mathbf{s}) = F(\mathbf{s}) + \alpha_\phi\big(\widetilde{\mathbf{X}}(\mathbf{s})\big)\delta_\phi\big(\widetilde{\mathbf{X}}(\mathbf{s})\big), \qquad 0 \le \alpha_\phi\big(\widetilde{\mathbf{X}}(\mathbf{s})\big) \le 1. \tag{8}$$

Thus, Direct prediction learns the final clean target without using $F(\mathbf{s})$, whereas Formula Correction and Trust Correction operate after factor error has propagated through RUSLE. In Trust Correction, the gate controls how strongly the learned residual modifies the degraded analytical prediction.

Factor-level restoration. Because the controlled degradation affects identifiable physical factors, an alternative is to restore the corrupted factor and apply the known RUSLE relationship afterward. For a corrupted factor q, this pathway is

$$\widetilde{\mathbf{X}}(\mathbf{s}) \xrightarrow{\text{factor restoration}} \widehat{q}(\mathbf{s}) \xrightarrow{\text{RUSLE reconstruction}} \widehat{T}(\mathbf{s}). \tag{9}$$

4 PhyRestore: Physics-Structured Latent-Factor Restoration

PhyRestore learns in physical-factor space, restoring corrupted factors before deterministic reconstruction of the temporal target through RUSLE. Figure 1 summarizes the pipeline.

4.1 Bitemporal Representation and Factor Restoration

For each valid center location i, we extract co-registered 7×7 neighborhoods from 2017 and 2022. Each date contains the four RUSLE channels $[R, K, LS, C]$, producing the condition-specific bitemporal observation

$$\widetilde{\mathbf{X}}_i \in \mathbb{R}^{2\times 4\times 7\times 7}. \tag{10}$$

A corrupted factor appears only in its degraded form; its clean value is not supplied through another input pathway. Separate normalization statistics are fitted for each date–factor channel using only the inner-training partition and held fixed during inner validation and outer-test inference. The dates are then concatenated along the channel dimension:

$$\widetilde{\mathbf{X}}_i \longrightarrow \mathbf{Z}_i \in \mathbb{R}^{8\times 7\times 7}. \tag{11}$$

Let $\mathcal{Q}_d$ denote the set of corrupted factors under condition d:

$$\mathcal{Q}_d = \begin{cases} \varnothing, & d = \text{Clean}, \\ \{R_{2022}\}, & d = R_{2022}, \\ \{C_{2017}\}, & d = C_{2017}, \\ \{R_{2022}, C_{2017}\}, & d = \text{Mixed}. \end{cases} \tag{12}$$

For each $q \in \mathcal{Q}_d$, PhyRestore learns a factor-specific function that predicts the clean center value of q:

$$\widehat{q}_i = g_{\theta_q}^{(d)}(\mathbf{Z}_i). \tag{13}$$

Under Mixed corruption, separate restorers for R_{2022} and C_{2017} are trained using the same jointly corrupted observations:

$$\widehat{R}_{2022,i} = g_{\theta_R}^{(\text{Mixed})}(\mathbf{Z}_i), \qquad \widehat{C}_{2017,i} = g_{\theta_C}^{(\text{Mixed})}(\mathbf{Z}_i). \tag{14}$$

Restorers trained under isolated corruption are not reused under Mixed corruption because the available input information differs between these conditions.

4.2 Restoration Network and Factor-Level Supervision

All factor restorers use the same 6,353-parameter convolutional architecture. Two unit-padded 3×3 convolutions preserve the 7×7 spatial dimensions and map the eight input channels to 16 and then 32 feature maps. The center representation is passed through a $32 \to 16 \to 1$ prediction head:

$$\mathbf{H}_i^{(1)} = \text{ReLU}\left(\text{Conv}_{3\times3}^{8\to16}(\mathbf{Z}_i)\right), \qquad \mathbf{H}_i^{(2)} = \text{ReLU}\left(\text{Conv}_{3\times3}^{16\to32}(\mathbf{H}_i^{(1)})\right), \tag{15}$$

$$\mathbf{h}_i = \mathbf{H}_i^{(2)}(:, c, c) \in \mathbb{R}^{32}, \tag{16}$$

$$\widehat{z}_{q,i} = \mathbf{W}_2\,\text{ReLU}\left(\mathbf{W}_1\mathbf{h}_i + \mathbf{b}_1\right) + \mathbf{b}_2, \tag{17}$$

where c is the center index of the spatial window. Using the same architecture for every restored factor prevents differences in restoration behavior from arising from factor-specific model capacity.

Factor targets are standardized using the mean μ_q and standard deviation σ_q of the clean center values in the corresponding inner-training partition. The network is trained using Huber loss with $\delta = 1$, and its predictions are decoded to physical units before reconstruction:

$$z_{q,i} = \frac{q_i - \mu_q}{\sigma_q}, \tag{18}$$

$$\mathcal{L}_q = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} \ell_{\text{Huber}}\left(\widehat{z}_{q,i}, z_{q,i}\right), \qquad \delta = 1, \tag{19}$$

$$\widehat{q}_i = \mu_q + \sigma_q \widehat{z}_{q,i}. \tag{20}$$

Checkpoint selection uses factor-level mean absolute error in the original physical units on the inner-validation partition. Neither the temporal target nor the held-out outer-test region is used for checkpoint selection.

4.3 Physics-Structured Temporal Reconstruction

No additional learned transformation is applied after restoration. For each condition, reconstruction uses the restored value when a factor is corrupted and the observed value otherwise:

$$R^*_{2022,i} = \begin{cases} \widehat{R}_{2022,i}, & R_{2022} \in \mathcal{Q}_d, \\ R_{2022,i}, & R_{2022} \notin \mathcal{Q}_d, \end{cases} \qquad C^*_{2017,i} = \begin{cases} \widehat{C}_{2017,i}, & C_{2017} \in \mathcal{Q}_d, \\ C_{2017,i}, & C_{2017} \notin \mathcal{Q}_d. \end{cases} \tag{21}$$

The final prediction follows directly from the temporal RUSLE relationship:

$$\widehat{T}_i^{\mathrm{PhyRestore}} = K_i L S_i \big[R^*_{2022,i} C_{2022,i} - R_{2017,i} C^*_{2017,i} \big]. \tag{22}$$

Equation 22 covers isolated R_{2022} corruption, isolated C_{2017} corruption, and Mixed corruption. Under Clean observations, no restoration model is required, and the prediction reduces to the clean temporal RUSLE difference. No learned residual or post-hoc correction follows reconstruction; the final estimate is determined entirely by the restored factors and the known RUSLE relationship.

5 Experiments

5.1 Experimental Setup

Data, baselines, and conditions. We evaluate temporal soil-loss prediction using co-registered 2017 and 2022 RUSLE-factor rasters over a Colorado watershed, with a common valid support of 14,225 locations and 7×7 spatial neighborhoods. The clean signed target T and the Clean, isolated R_{2022}, isolated C_{2017}, and Mixed corruption conditions are defined in Sections 3.1 and 3.2. We compare the degraded analytical estimate with the Direct, Formula Correction, and Trust Correction strategies from Section 3.3, using MLP, random-forest (RF), XGBoost, and CNN models. The complete comparison covers three spatial outer folds and five seeds (101–105).

Targeted high-magnitude controls. We additionally test seven controls that retain T as the prediction target while modifying the spatial representation, loss, or output structure. RF7 replaces the baseline CNN's 3×3–3×3 convolutions, whose theoretical receptive field is 5×5, with 5×5–3×3 convolutions spanning the full 7×7 neighborhood. MLP37 augments 19 center-pixel inputs of the baseline MLP (MLP19)—eight bitemporal factor values, eight validity indicators, and three availability/corruption indicators—with 18 neighborhood summaries: the mean, standard deviation, and valid fraction of six factor surfaces over the 48 surrounding cells.

MAG5-LINEAR-Q95 retains the MLP19 architecture but applies sign-symmetric magnitude weighting. On each inner-training split,

$$s = q_{95,\mathrm{train}}(|T|), \qquad w_i^{\mathrm{raw}} = 1 + 4\,\mathrm{clip}(|T_i|/s, 0, 1), \qquad w_i = \frac{w_i^{\mathrm{raw}}}{\mathrm{mean}_{\mathrm{train}}(w^{\mathrm{raw}})}.$$

Thus, raw weights increase linearly from 1 to 5 and saturate at the inner-training 95th percentile of $|T|$. RF7, MLP37, and MAG5 are confirmatory controls evaluated over all three outer folds, five seeds, and four conditions, yielding complete out-of-fold predictions for each control.

The remaining controls are development-only experiments evaluated on folds 0 and 2 with seed 101 under all four conditions, using inner-validation data only. RankSim regularizes the Direct MLP representation using rank similarity (Gong et al., 2022), with $\gamma = 3 \times 10^{-4}$ and $\lambda = 2$. Auxiliary classification adapts joint classification–regression learning (Pintea et al., 2023): signed-T quartiles define four inner-training classes, and a shared MLP trunk feeds regression and classification heads trained with $\mathcal{L}_{\mathrm{Huber}} + \beta \mathcal{L}_{\mathrm{CE}}$, where $\beta = 0.0657$; only the regression output is used at inference.

SER-MAD adapts segmentation-based expert regression (Rashiwa & Branco, 2026). Three MLP experts model lower, central, and upper target regimes separated at $m \pm 1.5\sigma_{\mathrm{MAD}}$, where $m = \mathrm{median}(T)$ and $\sigma_{\mathrm{MAD}} = 1.4826\,\mathrm{median}(|T - m|)$ on the inner-training split. Deployable routing uses an ordinary Direct MLP prediction, whereas oracle routing uses the true T to test whether perfect regime information would make specialization effective. Empty regimes fall back to the central expert.

The bitemporal Transformer applies self-attention (Vaswani et al., 2017) to 49 spatial tokens, each containing the four factor values from both dates. It uses 32-dimensional embeddings, two encoder layers, four attention heads, and the contextualized center token for regression.

Protocol and metrics. For each seed and condition in the primary and confirmatory experiments, predictions from the three held-out outer regions are concatenated into an out-of-fold prediction over all 14,225 locations. Fitting, normalization, early stopping, and checkpoint selection use only the corresponding training data, and matched methods receive the same frozen corruption realization.

Table 1: Full-support MAE ($\times 10^{-3}$), reported as mean $\pm$ standard deviation across five seeds or once for seed-invariant methods. Bold marks the best learned strategy within each family.

Method	R_{2022}	C_{2017}	Mixed
Degraded Formula	10.271	318.866	318.865
MLP–Direct	$\mathbf{2.007 \pm 0.623}$	$\mathbf{1.717 \pm 0.856}$	$\mathbf{1.932 \pm 0.656}$
MLP–Correction	6.076 ± 0.884	176.305 ± 21.667	173.110 ± 26.649
MLP–Trust	6.045 ± 0.809	145.850 ± 11.079	150.191 ± 12.054
RF–Direct	$\mathbf{0.981 \pm 0.005}$	$\mathbf{1.048 \pm 0.006}$	$\mathbf{0.979 \pm 0.011}$
RF–Correction	7.127 ± 0.085	212.359 ± 9.410	210.358 ± 9.581
RF–Trust	7.087 ± 0.067	186.493 ± 6.035	188.427 ± 6.068
XGB–Direct	$\mathbf{0.913}$	$\mathbf{0.905}$	$\mathbf{0.927}$
XGB–Correction	4.764	117.242	117.663
XGB–Trust	4.850	112.904	114.652
CNN–Direct	$\mathbf{1.094 \pm 0.265}$	$\mathbf{1.163 \pm 0.165}$	$\mathbf{1.278 \pm 0.335}$
CNN–Correction	10.472 ± 0.132	245.604 ± 25.101	361.743 ± 27.592
CNN–Trust	10.342 ± 0.080	283.415 ± 21.067	338.152 ± 23.496

Neural models use Huber loss with $\delta = 1$, validation-based early stopping, and Adam; MAG5 instead selects checkpoints using its weighted validation objective.

We report MAE over the full support and the fixed TAIL90 and TAIL95 subsets. TAIL90 contains the 1,423 locations with $|T| \geq 0.0012514$, while TAIL95 contains the 712 locations with $|T| \geq 0.0019710$. Because TAIL95 contains 61 positive and 651 negative changes, we also evaluate TAIL95+ and TAIL95$-$. Comparisons with the zero predictor $\widehat{T} = 0$, direction accuracy, and the prediction-to-target standard-deviation ratio distinguish low aggregate error from recovery of temporal-change sign and magnitude.

5.2 Direct Prediction versus Formula-Anchored Correction

Table 1 evaluates whether retaining the degraded analytical estimate as a prediction anchor improves robustness.

Direct prediction has lower full-support MAE than Formula Correction and Trust Correction in all 12 family–corruption combinations. For stochastic learners, the ordering holds in all five seeds of every comparison and also holds seed-wise on TAIL95. Formula-anchored strategies nevertheless repair part of the analytical error in some cases: under C_{2017} corruption, XGBoost reduces MAE from 0.318866 to 0.117242 with Formula Correction and 0.112904 with Trust Correction. However, Trust is not consistently better than Formula Correction, and CNN correction can approach or exceed the degraded analytical error. Retaining the corrupted formula as an anchor therefore remains substantially less robust than predicting T directly.

5.3 Robustness Does Not Imply High-Magnitude Recovery

Because T is concentrated near zero, all four primary Direct families have higher full-support MAE than the zero predictor under corrupted observations. Their tail behavior differs: MLP retains the strongest signed-tail signal, whereas RF, XGBoost, and CNN compress their predictions more strongly toward zero. Under C_{2017} corruption on TAIL95, their prediction-to-target standard-deviation ratios are 0.301, 0.054, 0.014, and 0.020, respectively, while direction accuracies are 0.796, 0.417, 0.451, and 0.022. Thus, low aggregate error does not establish recovery of high-magnitude sign or variation. Table 2 summarizes the targeted controls and their evaluation scope.

Among the confirmatory controls, RF7 yields mean TAIL95 MAEs of 0.005624, 0.005591, 0.005591, and 0.005588 under Clean, R_{2022}, C_{2017}, and Mixed observations, respectively, without materially improving signed-tail recovery over Direct CNN. MLP37 likewise provides no reproducible advantage over MLP19, with corresponding TAIL95 MAEs of 0.010718, 0.012408, 0.004714, and 0.006525. MAG5 worsens mean full-support MAE in every condition, and its iso-

Table 2: Controls targeting high-magnitude recovery while retaining T as the prediction target. Confirmatory controls use three folds and five seeds; development controls use folds 0 and 2 and seed 101 with inner-validation data only.

Control	Intervention	Main outcome
RF7	Full theoretical 7×7 CNN receptive field	No material signed-tail recovery relative to Direct CNN.
MLP37	Eighteen summaries of the surrounding 48 cells	No reproducible improvement over center-only MLP19.
MAG5	Sign-symmetric magnitude-weighted regression	No reproducible tail gain; full-support MAE worsens in all conditions.
RankSim	Rank-aware representation regularization	Representation ordering improves, but TAIL90 and TAIL95 errors consistently worsen.
Auxiliary classification	Four-class signed-target auxiliary task	Does not establish a tail-regression gain.
SER-MAD	Lower/central/upper regression experts	Neither deployable nor oracle routing improves the evaluated TAIL95 subsets.
Bitemporal Transformer	Attention over 49 paired-date spatial tokens	No consistent positive-tail advantage after convergence.

lated tail improvements do not meet the prespecified reproducibility criterion of improvement in at least four of five seeds.

In development, RankSim improves mean Spearman agreement by 0.188 but worsens every evaluated TAIL90 and TAIL95 result. Auxiliary classification achieves approximately 0.20 macro-F1 without improving tail regression, while neither deployable nor oracle SER-MAD routing improves TAIL95.

The Transformer's initial 40-epoch study was insufficient for convergence: seven of eight runs reached the maximum epoch. With a 200-epoch maximum, all eight runs converged and selected checkpoints between epochs 8 and 74. The converged Transformer improves over Direct MLP on Fold 2 TAIL95$-$ under C_{2017} corruption (0.004616 versus 0.006761), but is worse under R_{2022} (0.004826 versus 0.004012) and Mixed corruption (0.004756 versus 0.004172), and improves none of the four Fold 0 TAIL95+ comparisons. Additional bitemporal attention therefore does not consistently resolve the tail failure.

These results motivate changing the learned target rather than only the loss or representation. PhyRestore restores the corrupted physical factors and then reconstructs temporal change using the known RUSLE relationship.

5.4 Factor Restoration Improves High-Magnitude Recovery

A preliminary evaluation found recoverable factor-level signal, motivating full PhyRestore evaluation over three folds and five seeds. Because Formula Correction and Trust Correction are dominated on TAIL95 by their matched Direct models in every family and corrupted condition, Figure 2 compares PhyRestore with the stronger Direct baselines and zero predictor.

Under isolated R_{2022} corruption, PhyRestore obtains 0.003954 TAIL95 MAE versus 0.005561 for the strongest Direct baseline, a 28.9% reduction. Under isolated C_{2017} corruption, it obtains 0.001193 versus 0.004989, a 76.1% reduction. Relative to the degraded formula, the reductions are 81.8% and 99.8%, respectively. PhyRestore outperforms Direct MLP in four of five runs for R_{2022} and all five for C_{2017}, and outperforms RF, XGBoost, and CNN in all five runs under both isolated conditions.

Under Mixed corruption, Direct MLP is strongest, obtaining 0.004707 versus 0.005217 for PhyRestore. PhyRestore remains better on average than the degraded formula, zero predictor, RF, XGBoost, and CNN, but outperforms Direct MLP in only one of five seeds. TAIL90 exhibits the same

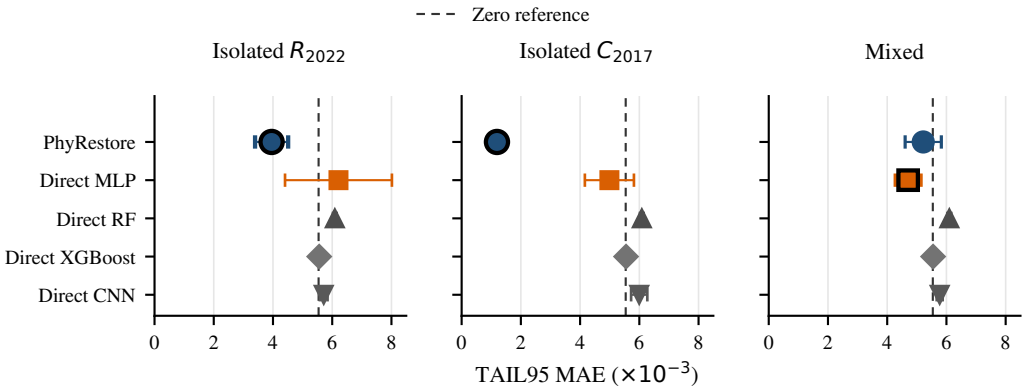


Figure 2: TAIL95 MAE ($\times 10^{-3}$). Error bars show standard deviation across five seeds; dashed lines mark the zero predictor. The degraded formula is omitted because of its substantially larger scale.

pattern: PhyRestore is strongest under isolated corruption, while Direct MLP is strongest under Mixed corruption.

Full-support accuracy. Tail improvements do not always extend to the full distribution. Under R_{2022} corruption, PhyRestore obtains 0.001159 ± 0.000064 full-support MAE, compared with 0.000913 for XGBoost, 0.000981 for RF, and 0.000706 for zero. Under Mixed corruption, the corresponding values are 0.001207 ± 0.000063, 0.000927, 0.000979, and 0.000706. Under C_{2017} corruption, however, PhyRestore is also strongest on full support, reaching 0.000065 ± 0.000006. The near-zero majority can therefore favor compressed predictions globally even when they recover high-magnitude changes poorly.

5.5 Signed Extremes and Limits of Factor Recoverability

Under isolated corruption, PhyRestore improves both signed TAIL95 subsets. For R_{2022}, its TAIL95+ and TAIL95− MAEs are 0.005473 and 0.003811, versus the strongest Direct errors of 0.009031 and 0.005191. For C_{2017}, it obtains 0.006710 and 0.000676, versus 0.009238 and 0.004542. Under Mixed corruption, however, PhyRestore obtains 0.011995 on TAIL95+, compared with 0.009031 for Direct RF and 0.009487 for zero, and 0.004582 on TAIL95−, slightly above Direct MLP's 0.004189. Moreover, under isolated C_{2017} corruption, its positive-tail prediction-to-target standard-deviation ratio remains only 0.061 despite the MAE improvement.

Factor restoration depends on inner-training support. Pooling outer-test predictions across seeds and relevant conditions, R_{2022} restoration MAE is 0.286836 for in-support values ($N = 56{,}350$) and 0.348697 out of support ($N = 85{,}900$). For C_{2017}, seen values yield 1.01×10^{-6} MAE ($N = 142{,}170$), whereas unseen values yield 0.013750 ($N = 80$).

6 Conclusion

We studied temporal prediction under corrupted physical factors and severe target imbalance. Across four model families, Direct prediction was more robust than formula-anchored correction but remained limited in recovering rare high-magnitude changes despite modifications to the objective, representation, spatial context, and architecture.

PhyRestore instead restores corrupted physical factors before reconstructing temporal change through RUSLE. Relative to the strongest Direct baselines, it reduced TAIL95 error by 28.9% and 76.1% under isolated R_{2022} and C_{2017} corruption, respectively. Its advantage weakened under simultaneous corruption and outside the training support, demonstrating that its effectiveness depends critically on factor recoverability.

AI USE STATEMENT

Generative AI tools were used to assist with editing and polishing the manuscript for clarity, grammar, and readability. All AI-assisted text was reviewed and revised by the authors. The authors take full responsibility for the final content of the paper.

REFERENCES

Rubianca Benavidez, Bethanna Jackson, Deborah Maxwell, and Kevin Norton. A review of the (revised) universal soil loss equation ((R)USLE): With a view to increasing its global applicability and improving soil loss estimates. *Hydrology and Earth System Sciences*, 22: 6059–6086, 2018. doi: 10.5194/hess-22-6059-2018. URL https://doi.org/10.5194/hess-22-6059-2018.

Carolina Boix-Fayos, María Martínez-Mena, Eva Arnau-Rosalén, Adolfo Calvo-Cases, Victor Castillo, and Juan Albaladejo. Measuring soil erosion by field plots: Understanding the sources of variation. *Earth-Science Reviews*, 78(3–4):267–285, 2006. doi: 10.1016/j.earscirev.2006.05.005. URL https://doi.org/10.1016/j.earscirev.2006.05.005.

Pasquale Borrelli, David A. Robinson, Larissa R. Fleischer, Emanuele Lugato, Cristiano Ballabio, Christine Alewell, Katrin Meusburger, Sirio Modugno, Brigitta Schütt, Vito Ferro, Vincenzo Bagarello, Kristof Van Oost, Luca Montanarella, and Panos Panagos. An assessment of the global impact of 21st century land use change on soil erosion. *Nature Communications*, 8 (1):2013, 2017. doi: 10.1038/s41467-017-02142-7. URL https://doi.org/10.1038/s41467-017-02142-7.

Christopher Daly, Wayne Gibson, George Taylor, Gary Johnson, and Paul Pasteris. A knowledge-based approach to the statistical mapping of climate. *Climate Research*, 22:99–113, 2002.

Yuankai Ge, Longlong Zhao, Jinsong Chen, Xiaoli Li, Hongzhong Li, Zhengxin Wang, and Yanni Ren. Study on soil erosion driving forces by using (R)USLE framework and machine learning: A case study in southwest china. *Land*, 12(3):639, 2023. doi: 10.3390/land12030639. URL https://doi.org/10.3390/land12030639.

Yu Gong, Greg Mori, and Fred Tung. RankSim: Ranking similarity regularization for deep imbalanced regression. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 7634–7649, 2022. URL https://proceedings.mlr.press/v162/gong22a.html.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021. doi: 10.1038/s42254-021-00314-5. URL https://doi.org/10.1038/s42254-021-00314-5.

Anuj Karpatne, Gowtham Atluri, James H. Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2318–2331, 2017. doi: 10.1109/TKDE.2017.2720168. URL https://doi.org/10.1109/TKDE.2017.2720168.

Krishna Karra, Caitlin Kontgis, Zoe Statman-Weil, Joseph C. Mazzariello, Mark Mathis, and Steven P. Brumby. Global land use / land cover with sentinel 2 and deep learning. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp. 4704–4707, 2021.

Pingheng Li, Aqil Tariq, Qingting Li, Bushra Ghaffar, Muhammad Farhan, Ahsan Jamil, Walid Soufan, Ayman El Sabagh, and Mohamed Freeshah. Soil erosion assessment by RUSLE model using remote sensing and GIS in an arid zone. *International Journal of Digital Earth*, 16(1): 3105–3124, 2023. doi: 10.1080/17538947.2023.2243916. URL https://doi.org/10.1080/17538947.2023.2243916.

Panos Panagos, Pasquale Borrelli, Jean Poesen, Cristiano Ballabio, Emanuele Lugato, Katrin Meusburger, Luca Montanarella, and Christine Alewell. The new assessment of soil loss by water erosion in europe. *Environmental Science & Policy*, 54:438–447, 2015. doi: 10.1016/j.envsci.2015.08.012. URL https://doi.org/10.1016/j.envsci.2015.08.012.

Kwanele Phinzi and Njoya Silas Ngetar. The assessment of water-borne erosion at catchment level using GIS-based RUSLE and remote sensing: A review. *International Soil and Water Conservation Research*, 7(1):27–46, 2019. doi: 10.1016/j.iswcr.2018.12.002. URL https://doi.org/10.1016/j.iswcr.2018.12.002.

Silvia L. Pintea, Yancong Lin, Jouke Dijkstra, and Jan C. van Gemert. A step towards understanding why classification helps regression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19915–19924, 2023.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045. URL https://doi.org/10.1016/j.jcp.2018.10.045.

Sarah Elyane Rashiwa and Paula Branco. Segmentation expert-mixture regularization: An adaptive learning method for imbalanced regression problems. In *Proceedings of the 39th Canadian Conference on Artificial Intelligence*, volume 318, pp. 612–623, 2026. URL https://proceedings.mlr.press/v318/rashiwa26a.html.

Kenneth G. Renard, George R. Foster, Glenn A. Weesies, Donald K. McCool, and Douglas C. Yoder. Predicting soil erosion by water: A guide to conservation planning with the revised universal soil loss equation (RUSLE). Agriculture Handbook 703, U.S. Department of Agriculture, Washington, DC, USA, 1997. URL https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB97153704.xhtml.

Rita P. Ribeiro and Nuno Moniz. Imbalanced regression and extreme value prediction. *Machine Learning*, 109:1803–1835, 2020. doi: 10.1007/s10994-020-05900-9. URL https://doi.org/10.1007/s10994-020-05900-9.

Nikiforos Samarinas, Nikolaos L. Tsakiridis, Eleni Kalopesa, and George C. Zalidis. Soil loss estimation by water erosion in agricultural areas introducing artificial intelligence geospatial layers into the RUSLE model. *Land*, 13(2):174, 2024. doi: 10.3390/land13020174. URL https://doi.org/10.3390/land13020174.

Leo Stroosnijder. Measurement of erosion: Is it possible? *CATENA*, 64(2–3):162–173, 2005. doi: 10.1016/j.catena.2005.08.004. URL https://doi.org/10.1016/j.catena.2005.08.004.

U.S. Department of Agriculture, Natural Resources Conservation Service. Soil survey geographic (ssurgo) database. https://www.nrcs.usda.gov/resources/data-and-reports/soil-survey-geographic-database-ssurgo, 2023. Accessed 2025-12-14.

U.S. Geological Survey. USGS 1/3 Arc Second n39w108 20210312. U.S. Geological Survey, The National Map, 2021. Accessed 2025-12-20.

Laurens van der Maaten, Minmin Chen, Stephen Tyree, and Kilian Weinberger. Learning with marginalized corrupted features. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 410–418. PMLR, 2013. URL https://proceedings.mlr.press/v28/vandermaaten13.html.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017.

Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, 55(4):66:1–66:37, 2022. doi: 10.1145/3514228. URL https://doi.org/10.1145/3514228.

Yuzhe Yang, Kaiwen Zha, Yingcong Chen, Hao Wang, and Dina Katabi. Delving into deep imbalanced regression. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11842–11851. PMLR, 2021. URL https://proceedings.mlr.press/v139/yang21m.html.

Jinsung Yoon, James Jordon, and Mihaela van der Schaar. GAIN: Missing data imputation using generative adversarial nets. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 5689–5698. PMLR, 2018. URL https://proceedings.mlr.press/v80/yoon18a.html.

Ahmet Faruk İpek and Ercan Kahya. Spatiotemporal prioritization of soil erosion risk using the rusle model and cmip6 projections under future climate scenarios in a mediterranean watershed. *Frontiers in Environmental Science*, 14, 2026. doi: 10.3389/fenvs.2026.1760569.