Recovering Governing Dynamics from Fragmented Observations via Exact Distributed Spline Merging

Anonymous authors
Paper under double-blind review

Abstract

Scientific observations are often fragmented across space, time, and institutions. If such fragments can be combined into a continuous, differentiable field, its derivatives yield governing physical parameters through regression. This paper makes two contributions in this setting. First, the established additive structure of fixed-basis ridge-regression statistics is applied to tensor-product spline fields: each data holder computes a local Gram matrix and moment vector, and the merged solution is mathematically identical to centralized fitting, with no raw data shared and no iterative synchronization. This property is specific to the fixed-feature squared-error setting; the present derivation does not establish an analogous guarantee for general jointly trained multilayer networks. Second, a complete pipeline connects distributed observations to physical parameter inference through field reconstruction, derivative extraction, and linear regression. The pipeline is validated on four PDEs: diffusion, wave, heat-with-source, and the nonlinear viscous Burgers equation, recovering governing parameters to sub-percent accuracy in the linear cases and 5% for Burgers. In all cases, distributed merging introduces zero degradation relative to centralized fitting. Synthetic experiments validate parameter recovery; application to 41 years of NOAA sea-surface temperature data validates field reconstruction and aggregation equivalence on real spatiotemporal observations. Source code to reproduce all experiments is available at https://anonymous.4open.science/r/splinemerge-F887.

1 Introduction

Scientific observations are often distributed across space, time, and institutions. No single observer has access to the full field, yet recovering governing dynamics benefits from a continuous, differentiable field representation. For models linear in their parameters, specifically fixed-basis spline expansions, the distributed merge problem has an exact, one-shot solution: each data holder computes a local Gram matrix $G_s = \Phi_s^\top \Phi_s$ and moment vector $h_s = \Phi_s^\top \mathbf{y}_s$, and the merged solution $c^* = (\sum_s G_s + \lambda I)^{-1} \sum_s h_s$ is mathematically identical to centralized fitting. No raw data is shared. No iterative synchronization is needed. Figure 1 illustrates the pipeline and demonstrates it on MNIST images.

This additive structure of fixed-feature ridge-regression statistics is classical (Hoerl & Kennard, 1970) and has been explicitly formulated as one-shot federated aggregation by Alsulaimawi (2026). Related model merging methods (McMahan et al., 2017; Wortsman et al., 2022; Ainsworth et al., 2023; Jin et al., 2023) and KAN merging (Polar & Poluektov, 2025) of Kolmogorov–Arnold Networks (Liu et al., 2025) address different settings. Federated scientific machine learning (Zhang et al., 2025; Shukla et al., 2021) requires iterative training. Field reconstruction methods (Raissi et al., 2019; Santos et al., 2023) and equation discovery (Brunton et al., 2016; Rudy et al., 2017; Messenger & Bortz, 2021) typically assume centralized data access.

Two distinct estimation problems are involved. The first recovers field coefficients from distributed observations; the exact aggregation guarantee applies here. The second estimates

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.

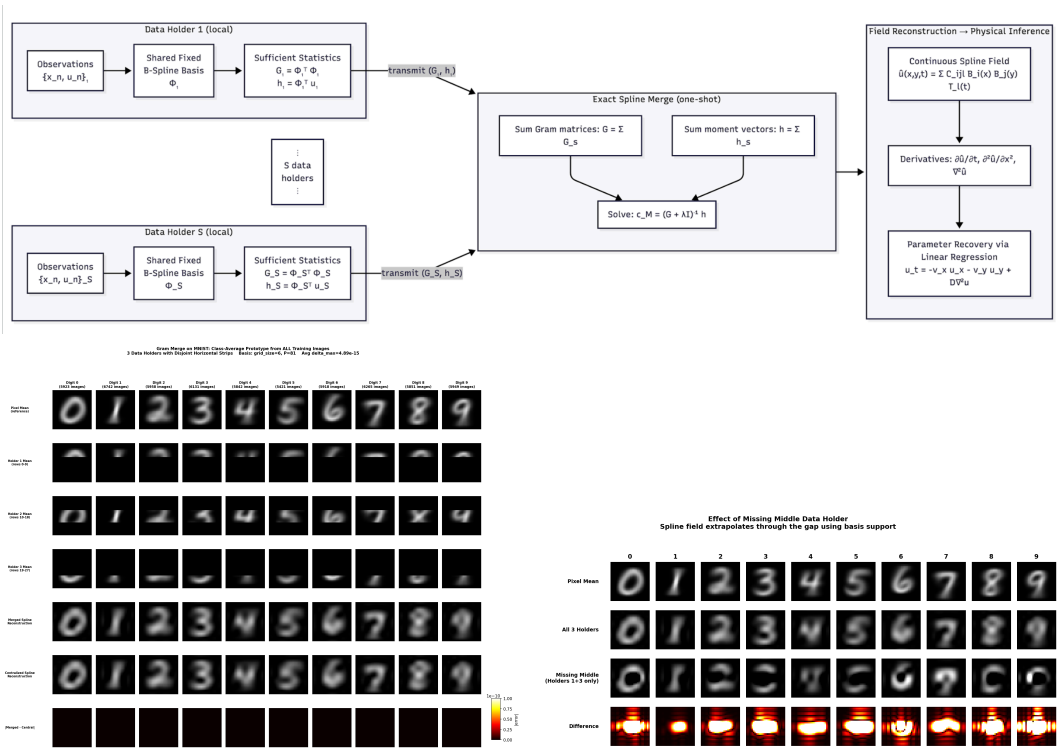


Figure 1: Top: End-to-end pipeline. Each data holder computes sufficient statistics (G_s, h_s) from local observations using a shared fixed B-spline basis and transmits only these (not the raw data). A single matrix solve produces the exact centralized solution; the resulting continuous field yields derivatives for parameter estimation. Bottom left: Demonstration on MNIST. Each 28×28 image is a 2D field $u(x, y)$ fitted with tensor-product B-splines (P=81). Three data holders each observe a disjoint horizontal strip (rows 0–9, 10–18, 19–27) across all ∼6,000 training images per digit class. The merged spline reconstruction is identical to the centralized fit (worst-case $\delta_{\max} < 10^{-13}$ across all ten digits), reconstructing class-average prototypes from ∼1.5M pixels per holder per digit class. Bottom right: When the middle holder is absent, the spline extrapolates through the gap. Digits with distinctive middle features (3, 4, 6, 8) degrade; simpler shapes (0, 1) remain recognizable. This demonstrates exact merging when all holders contribute and graceful degradation when observations are incomplete.

physical parameters (diffusivity, wave speed, viscosity) from derivatives of the reconstructed field; this step is approximate. Section 2 develops both constructions from first principles.

This paper makes two contributions: (1) Distributed field reconstruction: the merge reproduces the centralized estimator on synthetic PDEs, MNIST images, and real climate data. (2) End-to-end physics recovery: a complete pipeline recovers governing parameters across linear and nonlinear PDEs to sub-percent accuracy (5% for nonlinear Burgers), with a systematic noise sensitivity analysis.

2 Background

This section builds the core framework from first principles: basis function representations, the least-squares fitting problem, sufficient statistics, and why they compose exactly under distributed merging.

Problem setup. Multiple data holders observe the same scalar field in a common coordinate system. Each holder retains its own observation locations and measured values. Before

fitting, all holders agree on the same fixed basis functions, coefficient ordering, and regularized fitting objective. The first goal is to recover the coefficients that would be obtained by pooling all observations and fitting one model. The second goal is to use the reconstructed field to estimate parameters of a specified governing equation. The term "data holder" is used throughout; in the experiments, a data holder may represent a spatial region, an institutional partition, or a class-conditional subset.

2.1 B-Spline Basis Functions

A B-spline basis function $B_k(x)$ is a piecewise polynomial defined over a sequence of knot points (de Boor, 1978). For cubic B-splines (degree 3) on a uniform grid with simple interior knots, each basis function has support over four adjacent knot intervals and is zero outside this region. This property, called compact support, means that each B_k responds only to nearby data. The resulting spline is C^2-continuous (twice continuously differentiable) across interior knots, and the complete basis forms a partition of unity on its base interval.

A scalar field can then be represented as a weighted sum of these basis functions:

$$u(x) = \sum_{k=1}^{K} c_k \, B_k(x). \tag{1}$$

The representation can express nonlinear dependence on x, while remaining linear in its coefficients $c = [c_1, \ldots, c_K]^\top$. This distinction is the foundation of everything that follows.

2.2 Fitting as Linear Regression

Given N observations $\{(x_n, y_n)\}_{n=1}^{N}$, where y_n denotes the measured value at location x_n, the goal is to find coefficients c such that $u(x_n) \approx y_n$. Define the feature matrix $\Phi \in \mathbb{R}^{N \times K}$ with entries $\Phi_{nk} = B_k(x_n)$, so that the model predictions are $\hat{\mathbf{y}} = \Phi\, c$. The optimal coefficients minimize the regularized squared error:

$$\mathcal{L}(c) = \| \mathbf{y} - \Phi\, c \|^2 + \lambda \|c\|^2, \tag{2}$$

where $\lambda > 0$ is a ridge regularization parameter (Hoerl & Kennard, 1970). With $\lambda > 0$, the objective is strictly convex (the Hessian $2(\Phi^\top \Phi + \lambda I)$ is positive definite even when Φ is rank-deficient), ensuring a unique global minimum found by setting the gradient to zero:

$$\frac{\partial \mathcal{L}}{\partial c} = -2\, \Phi^\top (\mathbf{y} - \Phi\, c) + 2\lambda\, c = 0. \tag{3}$$

Rearranging yields the normal equations:

$$(\Phi^\top \Phi + \lambda I)\, c^* = \Phi^\top \mathbf{y}. \tag{4}$$

2.3 Gram Matrix, Moment Vector, and Sufficient Statistics

The solution in Eq. (4) depends on the observations only through two quantities:

$$G = \Phi^\top \Phi \in \mathbb{R}^{K \times K}, \qquad h = \Phi^\top \mathbf{y} \in \mathbb{R}^{K}. \tag{5}$$

The Gram matrix G captures how the basis functions co-activate across the observation points: entry $G_{ij} = \sum_n B_i(x_n)\, B_j(x_n)$. The moment vector h captures how each basis function correlates with the observed values: $h_k = \sum_n B_k(x_n)\, y_n$.

For the fixed-basis ridge-regression objective in Eq. (2), the pair (G, h) is sufficient to compute c^*: the coefficient estimate depends on the observations only through these two summaries. Once computed, the raw observations $\{(x_n, y_n)\}$ can be discarded without affecting the estimate. This sufficiency is specific to the chosen fitting objective; changing the basis, the loss function, or adding new analyses (such as computing the residual sum of squares, which requires $\mathbf{y}^\top \mathbf{y}$) may require information not preserved in (G, h). For a fixed basis, N observations reduce to a $K \times K$ matrix and a K-vector, with the summary size independent of observation count. The term "sufficient statistics" is used here in the computational sense of sufficient for the specified ridge estimator; under a Gaussian linear observation model with known noise variance, the same quantities also arise as classical statistical sufficient statistics (joint inference with unknown variance additionally requires $\mathbf{y}^\top \mathbf{y}$).

2.4 Composability: From Sufficiency to Exact Distributed Merging

Sufficiency alone is useful for compression. Composability makes it useful for distributed computing.

Suppose observations come from two data holders, A and B, at different locations. If all data were pooled, the stacked feature matrix and observation vector would give normal equations:

$$\left(\begin{bmatrix} \Phi_A \\ \Phi_B \end{bmatrix}^\top \begin{bmatrix} \Phi_A \\ \Phi_B \end{bmatrix} + \lambda I \right) c^* = \begin{bmatrix} \Phi_A \\ \Phi_B \end{bmatrix}^\top \begin{bmatrix} \mathbf{y}_A \\ \mathbf{y}_B \end{bmatrix}. \tag{6}$$

Expanding the block products:

$$(\underbrace{\Phi_A^\top \Phi_A}_{G_A} + \underbrace{\Phi_B^\top \Phi_B}_{G_B} + \lambda I)\, c^* = \underbrace{\Phi_A^\top \mathbf{y}_A}_{h_A} + \underbrace{\Phi_B^\top \mathbf{y}_B}_{h_B}. \tag{7}$$

The sufficient statistics are additive. Each data holder computes its local (G_s, h_s) independently and transmits only these compact summaries. The merged solution

$$c_M^* = (G_A + G_B + \lambda I)^{-1}(h_A + h_B) \tag{8}$$

is identical to the solution that would be obtained by fitting on all observations centrally. No raw data is shared. No iterative synchronization is needed. The merge requires a single matrix solve.

This extends to any number of data holders S by induction: $c_M^* = (\sum_{s=1}^S G_s + \lambda I)^{-1} \sum_{s=1}^S h_s$.

Two merge variants. The formula above reconstructs the centralized data fit, using the moment vectors $h_s = \Phi_s^\top \mathbf{y}_s$ that summarize raw observations. A related variant preserves trained model predictions instead: if data holder s has already fitted local coefficients c_s, the prediction-matching merge is $c_M = (\sum_s G_s + \lambda I)^{-1} \sum_s G_s c_s$. The two coincide when each local fit is at its unregularized optimum ($G_s c_s = h_s$). The experiments in this paper use the data merge.

The summary size per data holder is $O(K^2)$ for the symmetric Gram matrix plus $O(K)$ for the moment vector, independent of observation count. For an ordered univariate degree-p B-spline basis, compact support yields a Gram matrix with half-bandwidth at most p. Tensor-product representations induce different sparsity patterns (Section 3). Note that dense summaries can exceed the raw data size when K is large relative to N; sparsity exploitation or structured transmission would be needed for a practical bandwidth advantage.

Fixed versus learned features. The aggregation derived above applies whenever the feature map is fixed and the coefficients enter a quadratic objective. This includes spline expansions, Fourier bases, and linear readouts on frozen neural representations. A fully trainable multi-layer network generally does not satisfy this condition: changing an earlier layer changes the features supplied to subsequent layers, so the Gram matrix from one parameter setting is no longer valid at another. The present construction therefore does not extend to exact whole-network merging in that setting.

From aggregation to physical inference. The preceding derivation concerns the first estimation problem: recovering field coefficients from distributed observations. Exact aggregation removes any discrepancy between distributed and centralized fitting. It does not, by itself, remove field-approximation error (the basis may not perfectly represent the true field), derivative-estimation error (finite differences or analytic derivatives of an approximate field), or identifiability issues in the physical regression (ill-conditioned derivative ratios). Section 3 specifies the field representation and the parameter-recovery procedure; the experiments distinguish aggregation discrepancy from reconstruction and parameter-estimation errors, although a complete decomposition of individual error contributions has not been performed.

3 Method

The dimensionality of the input does not change the aggregation algebra. A 1D spline curve $u(x) = \sum_i c_i B_i(x)$, a 2D surface $u(x, y) = \sum_{i,j} C_{ij} B_i(x) B_j(y)$, and a 3D spatiotemporal field all reduce to the same object: $f(\mathbf{x}) = \phi(\mathbf{x})^\top c$, where ϕ is a fixed feature vector and c is the coefficient vector to be estimated. The Gram merge applies identically in each case (Figure 1). This section specifies the field representation, basis geometry, and parameter-recovery procedure.

3.1 Tensor-Product Spatiotemporal Fields

A scalar spatiotemporal field $u : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ is represented using a tensor-product expansion. For a 2D spatial domain with one temporal dimension:

$$u(x, y, t) = \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \sum_{l=1}^{K_t} C_{ijl}\, B_i(x)\, B_j(y)\, T_l(t), \tag{9}$$

where B_i are cubic B-spline basis functions and T_l are temporal basis functions (either B-splines or Fourier harmonics). Defining the tensor-product feature $\phi_{ijl}(x, y, t) = B_i(x)\, B_j(y)\, T_l(t)$ and flattening the coefficients into a vector $c = \mathrm{vec}(C) \in \mathbb{R}^P$ where $P = K_x \cdot K_y \cdot K_t$, the field becomes $u = \phi^\top c$. For N observations, the feature matrix $\Phi \in \mathbb{R}^{N \times P}$ has entries $\Phi_{n,(i,j,l)} = B_i(x_n)\, B_j(y_n)\, T_l(t_n)$, and the Gram merge from Section 2 applies directly.

B-spline basis functions are evaluated using the piecewise polynomial evaluation described by Mysore (2026), implemented in the publicly available InKAN package (v0.4.2). All experiments are implemented in PyTorch (Paszke et al., 2019) with NumPy (Harris et al., 2020) and SciPy (Virtanen et al., 2020) for numerical routines.

3.2 Basis Geometry and Merge Statistics

Basis support constrains the possible sparsity of the empirical feature-interaction statistics. For local B-splines (degree p), each $B_i(x)$ has compact support spanning at most $p+1$ knot intervals, and the resulting Gram matrix is banded within each univariate block. For tensor-product representations, an interaction is zero when the basis supports fail to overlap in any coordinate. Global features such as $\cos(m\omega x)$ produce generally non-sparse Gram matrices whose conditioning depends on frequency selection, observation coverage, and weighting.

The actual entries, conditioning, and behavior of approximate merging methods additionally depend on the represented function space, regularization, and local fitting accuracy. Basis support alone is insufficient as a causal explanation for averaging performance. The present experiments demonstrate exact statistic aggregation for spline and hybrid representations; a controlled basis-family comparison is left to future work.

The NOAA experiment (Section 4.6) uses a hybrid representation: spatial B-splines crossed with temporal Fourier harmonics and polynomial trend terms, producing a tensor-product model with spatially varying seasonal amplitudes.

3.3 Physical Parameter Recovery

Given the fitted continuous field $u(x, y, t)$, spatial and temporal derivatives are computed via central finite differences on a regular evaluation grid (interior points only). Physical parameters that appear linearly in a specified PDE are then recovered by ordinary least squares. Let ℓ_i denote the computed Laplacian $\nabla^2 u$ at interior grid point i. For the diffusion equation $u_t = D\nabla^2 u$, the recovered diffusion coefficient is:

$$D_{\mathrm{rec}} = \frac{\sum_i \ell_i\, u_{t,i}}{\sum_i \ell_i^2}. \tag{10}$$

For the wave equation $u_{tt} = c_w^2 \nabla^2 u$, the analogous formula yields $c^2_{w,\mathrm{rec}} = \sum_i \ell_i\, u_{tt,i} \,/\, \sum_i \ell_i^2$, with $c_{w,\mathrm{rec}} = \sqrt{c^2_{w,\mathrm{rec}}}$ when the recovered squared value is non-negative and treated as

invalid otherwise. For the inhomogeneous heat equation $u_t = D\nabla^2 u + S$, the known source is subtracted: $D_{\text{rec}} = \sum_i \ell_i \, (u_{t,i} - S_i) \,/\, \sum_i \ell_i^2$. For Burgers $u_t + u\, u_x = \kappa\, u_{xx}$, the viscosity is recovered as $\kappa_{\text{rec}} = \sum_i u_{xx,i} \, (u_{t,i} + u_i \, u_{x,i}) \,/\, \sum_i u_{xx,i}^2$.

This is the second problem described in Section 2: the physical parameters are inferred from the reconstructed field, not from raw observations. The accuracy depends on field-approximation quality, derivative computation, and conditioning of $\sum_i \ell_i^2$ (a numerical stabilizer of 10^{-8} is added to the denominator in the implementation). When the Laplacian is nearly zero everywhere, the denominator is small and the recovery is unreliable. Because the computed Laplacian contains finite-difference truncation error, the OLS regression operates with noisy features. The resulting bias direction is not fixed: errors in the response (u_t) and features (ℓ) are derived from the same fitted field and can be correlated, so the bias may have either sign.

Remark on local rank deficiency. Under spatial fragmentation, a data holder's local Gram matrix G_s is typically rank-deficient for compactly supported B-splines (basis functions outside the observed region contribute zero columns). However, all exact unregularized local least-squares minimizers produce the same product $G_s c_s = h_s$, so local uniqueness of c_s is unnecessary for the data merge. The composability guarantee extends to any globally shared quadratic penalty $\lambda\, c^\top R\, c$ (e.g., derivative-based smoothing penalties), provided $\sum_s G_s + \lambda R$ is positive definite.

4 Experiments

The pipeline is validated on four synthetic PDEs (three linear, one nonlinear) with known ground truth and on real climate data. All PDE experiments use cubic B-splines (p=3) with ridge regularization $\lambda = 10^{-4}$ and float64 arithmetic. Observations are sampled uniformly at random; distributed merge experiments split the domain spatially, with equal observations per data holder. Synthetic reference solutions use analytic Fourier-mode evolution (diffusion, wave, heat-with-source) or pseudospectral integration (Burgers). Full experimental configuration is in Appendix Table 2.

4.1 Diffusion Equation

The diffusion equation $u_t = D(u_{xx} + u_{yy})$ with true diffusivity $D = 0.05$ is the simplest PDE in the pipeline: one governing parameter, first-order in time. Each Fourier mode evolves analytically, decaying as $\hat{u}(\mathbf{k}, t) = \hat{u}(\mathbf{k}, 0)\, e^{-D|\mathbf{k}|^2 t}$. From 10,000 random spatiotemporal observations, a 3D tensor-product B-spline field (K_x=11, K_y=11, K_t=11; P=1,331) is fitted and the diffusion coefficient is recovered via single-parameter regression: $D_{\text{rec}} = 0.04994$ (0.12% relative error from unrounded value).

Figure 2 shows both 2D PDE experiments. The diffusion field (Figure 2a) smooths progressively; the wave field (Figure 2b) conserves energy. In both cases, centralized and merged reconstructions are visually indistinguishable; the error panels reflect field-approximation error, not the merge.

4.2 Wave Equation

The wave equation $u_{tt} = c_w^2(u_{xx} + u_{yy})$ with $c_w = 1$ is solved by analytic evolution of its Fourier modes. From 10,000 observations (K_x=11, K_y=11, K_t=13; P=1,573), the recovered wave speed is $c_{w,\text{rec}} = 0.9991$ (0.09% error). The distributed merge matches centralized to $\delta_{\max} < 10^{-12}$.

4.3 Viscous Burgers Equation (Nonlinear)

The 1D viscous Burgers equation $u_t + u\, u_x = \kappa\, u_{xx}$ with $\kappa = 0.1$ is solved via pseudospectral methods. The domain is $[0, 2\pi] \times [0, 1]$, fitted with a 2D basis (K_x=15, K_t=13; P=195). With 5,000 observations, $\kappa_{\text{rec}} = 0.1050$ (5.0% error). The Burgers experiment introduces the nonlinear feature $u\, u_x$ and has higher parameter error in this configuration; the contributions of

Table 1: Distributed merge: Gram merge versus centralized fitting on balanced regional draws (5,000 per holder for 2D PDEs, 2,500 for Burgers). $\delta_{\max}$: maximum absolute prediction difference on the evaluation grid. Main-text parameter estimates use separate full-domain draws, explaining slight differences.

PDE	Method	RMSE	Recovered Parameter	$\delta_{\max}$
Diffusion	Centralized	0.00252	$D_{\text{rec}} = 0.04989$	
	Gram merge	0.00252	$D_{\text{rec}} = 0.04989$	$< 10^{-12}$
Wave	Centralized	0.00178	$c_{w,\text{rec}} = 0.9991$	
	Gram merge	0.00178	$c_{w,\text{rec}} = 0.9991$	$< 10^{-12}$
Burgers	Centralized	0.01096	$\kappa_{\text{rec}} = 0.1050$	
	Gram merge	0.01096	$\kappa_{\text{rec}} = 0.1050$	$< 2 \times 10^{-11}$
Heat+source	Centralized	0.00095	$D_{\text{rec}} = 0.02995$	
	Gram merge	0.00095	$D_{\text{rec}} = 0.02995$	$< 10^{-12}$

representation and derivative-estimation errors have not been isolated. A grid-convergence check on the reference field shows that refining from 80×60 to 320×240 reduces reference-field recovery error from 0.89% to 0.06%, indicating finite-difference discretization error in the reference evaluation. The merge reproduces centralized exactly ($\delta_{\max} < 2 \times 10^{-11}$).

4.4 Heat Equation with Spatially Varying Source

The inhomogeneous heat equation $u_t = D(u_{xx} + u_{yy}) + S(x, y)$ with $D = 0.03$ and a periodic source $S(x, y) = 0.3 + 0.2 \cos x + 0.1 \sin 2y$ is solved exactly via the modal formula $\hat{u}_k(t) = e^{-a_k t}\hat{u}_k(0) + (1 - e^{-a_k t})\hat{S}_k/a_k$ for $|k| > 0$ (where $a_k = D|k|^2$) and $\hat{u}_0(t) = \hat{u}_0(0) + t\hat{S}_0$ for the zero mode. With 10,000 observations, $D_{\text{rec}} = 0.02977$ (0.76% relative error from unrounded value). The distributed merge matches centralized to $\delta_{\max} < 10^{-12}$.

4.5 Noise Sensitivity and Robustness

A systematic noise sensitivity study (Appendix Table 3) crosses three observation densities (1K, 3K, 10K), three noise levels (0%, 1%, 5%), and five independent seeds for both diffusion and wave. Sub-percent parameter recovery is achieved at 3,000+ observations with up to 1% noise. The distributed merge reproduces the centralized estimator in all 90 conditions ($\delta_{\max} < 10^{-11}$). Signed-bias analysis reveals systematic attenuation of diffusion estimates under high noise, consistent with errors-in-variables effects. Note that with a fixed unnormalized ridge penalty λ, increasing N changes both data density and effective regularization strength; this confound has not been controlled for.

Table 1 summarizes the distributed merge results across all four PDEs. The dense reconstruction experiments use a full-domain random draw; the distributed experiments use a balanced regional draw with a different sampling procedure, which is why parameter estimates differ slightly between the main text and the table. Individual data holders fail as extrapolators outside their observed regions, but the merged field recovers the full domain. An independent verification applying the same finite-difference stencils to exact analytic field values (without spline fitting) yields 0.13% wave-speed error and 0.45% diffusion error from the derivative grid alone. These controls quantify a nonzero discretization contribution; they are not lower bounds on full-pipeline parameter error, because reconstruction and differentiation errors may reinforce or partially cancel.

4.6 Real Data: NOAA Sea-Surface Temperature

To test the framework beyond synthetic settings, it is applied to National Oceanic and Atmospheric Administration (NOAA) OI SST V2 monthly sea-surface temperature data (Reynolds et al., 2002), covering December 1981 to January 2023 over the tropical Pacific

(30°S–30°N, 120°E–280°E). The representation uses 2D spatial B-splines ($K_{\text{lat}} = 13$, $K_{\text{lon}} = 23$) combined with temporal Fourier harmonics and polynomial trend terms, totaling 2,093 parameters fitted to approximately 4.41 million ocean grid-month values (after excluding land cells using the product's land–sea mask).

Figure 3 shows the merged reconstruction for two months. Partitioning into west (120–200°E) and east (200–280°E) data holders, the Gram merge matches centralized fitting (in-sample RMSE 0.639°C both). Held-out evaluation withholds 20% of ocean points (spatial, seed 42) or every 6th month (temporal); held-out RMSE is 0.643°C and 0.663°C respectively (using λ=10^{-4}; the full-data experiment uses λ=10^{-1}), supporting spatial and temporal interpolation on this product. The merge reproduces centralized reconstruction RMSEs in both held-out tests. Note that this partitions a gridded analysis product (NOAA Physical Sciences Laboratory), not raw sensor records.

5 Conclusion

Tensor-product spline fields admit composable sufficient statistics, enabling exact distributed merging and physical inference from fragmented observations. Validated on four PDEs and real climate data, the merge introduces zero degradation: prediction-space discrepancies remain below 2×10^{-11} across all synthetic experiments, with sub-percent parameter recovery for linear PDEs and 5% for nonlinear Burgers. For NOAA SST, merged and centralized reconstruction RMSEs match to displayed precision. For fixed-feature squared-error fitting, the protocol requires communicating only Gram matrices and moment vectors. Limitations include tensor-product scaling ($P = k^q$), finite-difference derivatives, and restriction to known PDE structure. See Appendix C for further discussion.

Reproducibility Statement

All experiments use fixed random seeds. Diffusion, wave, and heat-with-source solutions are generated by analytic evolution of represented Fourier modes. Burgers solutions use pseudospectral integration (dealiased 3/2 rule, RK4, N=256, Δt=0.001). Observation evaluation and derivative estimation introduce additional numerical approximations, assessed through reference tests described in Section 4. The NOAA SST data is publicly available from NOAA/OAR/ESRL PSL. The InKAN B-spline package is publicly available (Mysore, 2026). Experiment code is available at the anonymous repository linked in the abstract. Key hyperparameters (grid sizes, regularization strength, observation counts) are reported in each experiment description. Complete specifications, including initial conditions, coordinate normalization, spline knot construction, NOAA temporal basis definitions, noise distributions, and random seeds, are provided in the supplementary code.

Ethics Statement

This work involves no human subjects or private data. The NOAA SST dataset is publicly available. The distributed merging framework communicates sufficient statistics (Gram matrices and moment vectors) rather than individual observation records. Privacy leakage through the shared statistics is not analyzed in this work and no formal privacy guarantee is claimed.

AI Use Statement

The research concept, experimental design, and scientific analysis are the author's own work. The author wrote the core implementation, ran all experiments, and performed the scholarly investigation including primary-source verification of all citations. Claude (Anthropic) and ChatGPT (OpenAI) were used as assistive tools for code debugging, literature search, prose editing, and mathematical cross-checks. The author takes responsibility for the final methods, results, and text.

References

Samuel K. Ainsworth, Jonathan Hayase, and Siddhartha Srinivasa. Git Re-Basin: Merging models modulo permutation symmetries. In International Conference on Learning Representations, 2023. URL https://openreview.net/forum?id=CQsmMYmlP5T.

Zahir Alsulaimawi. One-shot federated ridge regression: Exact recovery via sufficient statistic aggregation, 2026. URL https://arxiv.org/abs/2601.08216. arXiv:2601.08216.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences, 113(15):3932–3937, 2016. doi: 10.1073/pnas.1517384113.

Carl de Boor. A Practical Guide to Splines, volume 27. Springer, 1978.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, et al. Array programming with NumPy. Nature, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.

Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1):55–67, 1970. doi: 10.1080/00401706.1970.10488634.

Xisen Jin, Xiang Ren, Daniel Preotiuc-Pietro, and Pengxiang Cheng. Dataless knowledge fusion by merging weights of language models. In International Conference on Learning Representations, 2023. URL https://openreview.net/forum?id=FCnohuR6AnM.

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljacic, Thomas Hou, and Max Tegmark. KAN: Kolmogorov–Arnold networks. In International Conference on Learning Representations, volume 2025, pp. 70367–70413, 2025.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, volume 54 of Proceedings of Machine Learning Research, pp. 1273–1282. PMLR, 2017.

Daniel A. Messenger and David M. Bortz. Weak SINDy for partial differential equations. Journal of Computational Physics, 443:110525, 2021. doi: 10.1016/j.jcp.2021.110525.

Naveen Mysore. InKAN: B-spline KANs via truncated power form, 2026. URL https://arxiv.org/abs/2609.01956. arXiv:2609.01956. Software: https://pypi.org/project/inkan/.

NOAA Physical Sciences Laboratory. NOAA optimum interpolation (OI) SST V2: Monthly mean data. https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html. URL https://psl.noaa.gov/thredds/dodsC/Datasets/noaa.oisst.v2/sst.mnmean.nc.html. Data file: sst.mnmean.nc. Publication date and data retrieval date not specified.

Adam Paszke, Sam Gross, Francisco Massa, et al. PyTorch: An imperative style, high-performance deep learning library. In Advances in Neural Information Processing Systems, volume 32, 2019.

Andrew Polar and Michael Poluektov. Merging of Kolmogorov–Arnold networks trained on disjoint datasets, 2025. URL https://arxiv.org/abs/2512.18921v1. arXiv:2512.18921v1.

Maziar Raissi, Paris Perdikaris, and George E. 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.

Richard W. Reynolds, Nick A. Rayner, Thomas M. Smith, Diane C. Stokes, and Wanqiu Wang. An improved in situ and satellite SST analysis for climate. Journal of Climate, 15 (13):1609–1625, 2002. doi: 10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

Samuel H. Rudy, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Data-driven discovery of partial differential equations. Science Advances, 3(4):e1602614, 2017. doi: 10.1126/sciadv.1602614.

Javier E. Santos, Zachary R. Fox, Arvind Mohan, Daniel O'Malley, Hari Viswanathan, and Nicholas Lubbers. Development of the Senseiver for efficient field reconstruction from sparse observations. Nature Machine Intelligence, 5(11):1317–1325, 2023. doi: 10.1038/s42256-023-00746-x.

Khemraj Shukla, Ameya D. Jagtap, and George Em. Karniadakis. Parallel physics-informed neural networks via domain decomposition. Journal of Computational Physics, 447: 110683, 2021.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S. Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, and Ludwig Schmidt. Model soups: Averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In Proceedings of the 39th International Conference on Machine Learning, volume 162 of Proceedings of Machine Learning Research, pp. 23965–23998. PMLR, 2022.

Handi Zhang, Langchen Liu, Kangyu Weng, and Lu Lu. Federated scientific machine learning for approximating functions and solving differential equations with data heterogeneity. IEEE Transactions on Neural Networks and Learning Systems, 36(10):18104–18117, 2025. doi: 10.1109/TNNLS.2025.3580409.

A Experimental Configuration and Additional Results

All experiments use InKAN v0.4.2 (piecewise polynomial basis), PyTorch (float64), and were developed on NVIDIA H100 80GB GPUs; the released code runs on CPU. Table 2 summarizes the configuration.

Table 2: Experimental configuration. P: total coefficients. N_{eval}: FD derivative grid.

	Diffusion	Wave	Burgers	Heat+source	NOAA SST
PDE	$u_t{=}D\nabla^2 u$	$u_{tt}{=}c_w^2\nabla^2 u$	$u_t{+}uu_x{=}\kappa u_{xx}$	$u_t{=}D\nabla^2 u{+}S$	(fitting)
Domain	$[0,2\pi]^2{\times}[0,2]$	$[0,2\pi]^2{\times}[0,3]$	$[0,2\pi]{\times}[0,1]$	$[0,2\pi]^2{\times}[0,2]$	Pacific, 41yr
True param.	$D{=}0.05$	$c_w{=}1$	$\kappa{=}0.1$	$D{=}0.03$	–
Basis	11^3	$11^2{\times}13$	$15{\times}13$	11^3	$13{\times}23{\times}7$
P	1331	1573	195	1331	2093
N_{eval}	$40^2{\times}30$	$40^2{\times}40$	$80{\times}60$	$40^2{\times}30$	–
N_{obs}	10K	10K	5K	10K	4.41M

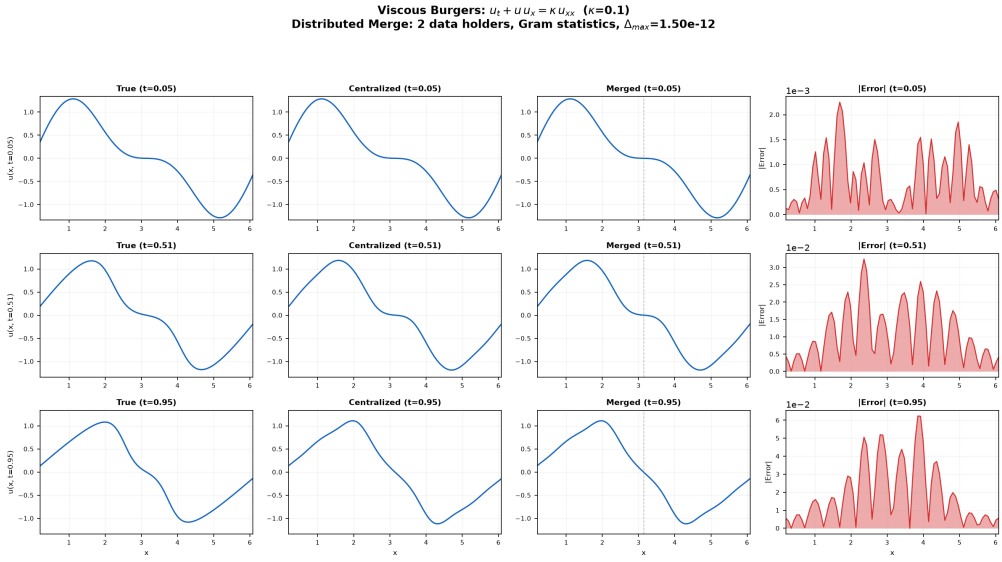


Figure 4: Viscous Burgers $u_t + u\,u_x = \kappa\,u_{xx}$: ground truth, centralized fit, Gram merge, and absolute error. The merged field matches centralized ($\delta_{\max} < 2 \times 10^{-11}$).

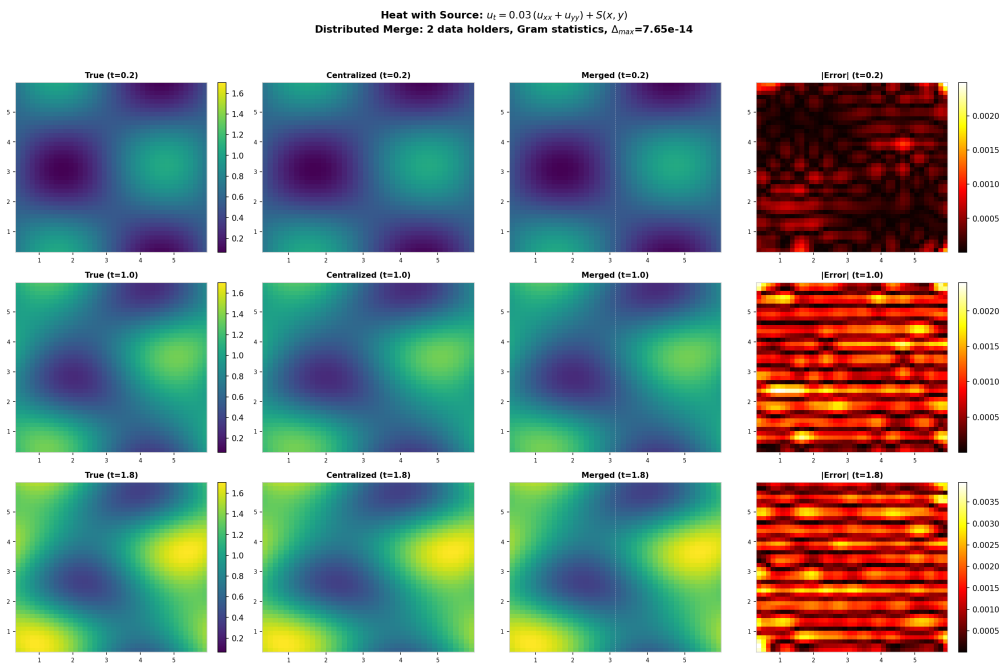


Figure 5: Heat with source $u_t = D\nabla^2 u + S(x, y)$: ground truth, centralized fit, Gram merge, and absolute error. The merged field matches centralized ($\delta_{\max} < 10^{-12}$).

B Noise Sensitivity Study

Table 3: Parameter recovery: mean ± std of relative error (%) over 5 independent seeds. Noise levels are defined relative to a fixed reference-grid field standard deviation. Sample standard deviations use ddof=1. The distributed merge reproduces the centralized estimator in all 90 conditions ($\delta_{\max} < 10^{-11}$). All wave conditions produced valid $\widehat{c_w^2} \geq 0$.

	Diffusion (D=0.05)			Wave (c_w=1)		
N_{obs}	0% noise	1% noise	5% noise	0% noise	1% noise	5% noise
1,000	2.3 ± 2.3	2.0 ± 2.6	8.1 ± 5.3	0.9 ± 0.6	0.6 ± 0.5	8.4 ± 0.8
3,000	0.3 ± 0.2	0.4 ± 0.2	2.6 ± 1.7	0.4 ± 0.04	0.3 ± 0.2	1.2 ± 0.9
10,000	0.1 ± 0.08	0.4 ± 0.4	1.8 ± 1.9	0.04 ± 0.03	0.06 ± 0.03	0.5 ± 0.4

Signed-bias analysis shows diffusion estimates are systematically attenuated under high noise (mean bias -0.003 at 5% noise, 1K observations), consistent with errors-in-variables effects. Wave-speed recovery is more robust at high observation counts because the estimator depends on u_{tt} and the Laplacian, both of which benefit from temporal smoothing in the spline fit.

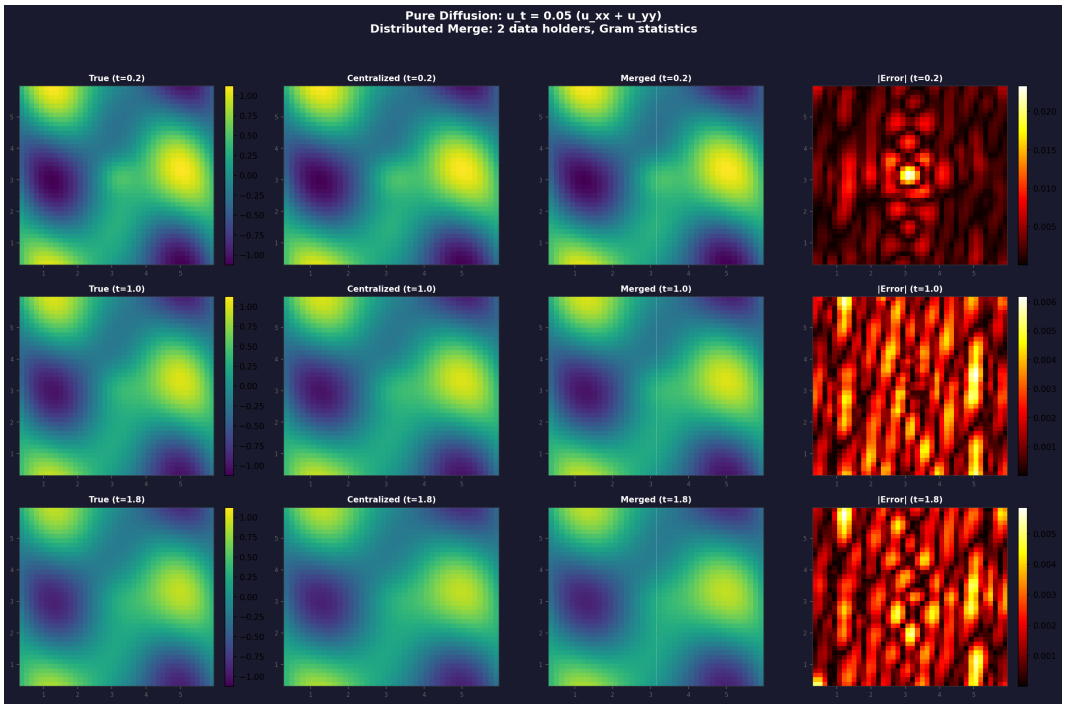
C Discussion

Scope of exactness. The exact merge holds for fixed-feature models optimized under the quadratic objective in Eq. (2). It does not hold for multi-layer networks where independently trained hidden representations diverge, because changing an earlier layer invalidates the Gram matrices computed at later layers. The composability guarantee extends to any globally shared quadratic penalty $\lambda\, c^\top R\, c$ (e.g., derivative-based smoothing penalties), provided $\sum_s G_s + \lambda R$ is positive definite.

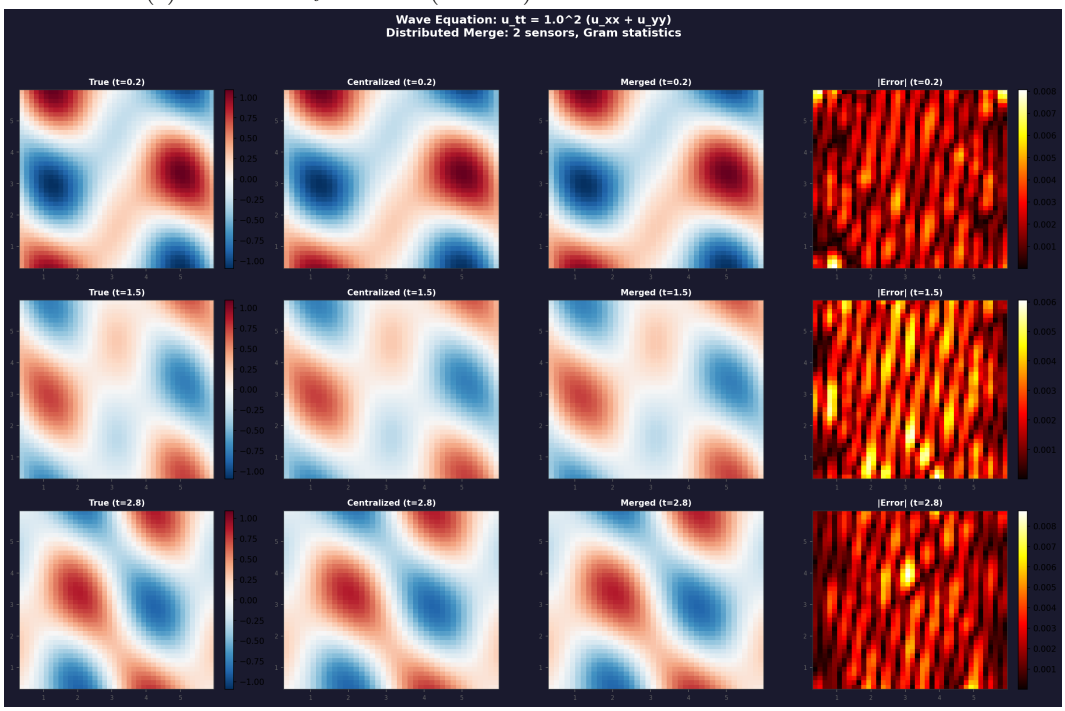
Basis geometry. Basis support constrains Gram sparsity, but averaging performance additionally depends on function space, coverage, regularization, and conditioning. A controlled basis-family comparison is left to future work.

Scaling and limitations. The coefficient count $P = k^q$ grows exponentially with input dimension; dense Gram storage is $O(P^2)$ and the solve is $O(P^3)$. In the synthetic experiments, dense summaries exceed raw-data size; communication savings require sparsity exploitation not implemented here. Finite-difference derivative estimation limits physics recovery accuracy. The NOAA SST experiment uses an already-interpolated product; real-data physical-parameter inference remains outside the validated scope. Privacy leakage through the shared statistics is not analyzed.

Future directions. Analytic B-spline derivatives, periodic basis functions, vector-field extensions, and quantitative comparison with federated scientific machine learning methods (Zhang et al., 2025) are natural next steps.



(a) Diffusion: $u_t = D\nabla^2 u$ (D=0.05). The field smooths as heat diffuses.



(b) Wave: $u_{tt} = c_w^2 \nabla^2 u$ (c_w=1). Energy is conserved; the field persists.

Figure 2: Synthetic 2D PDE experiments. Each row shows ground truth, centralized fit, Gram merge (two data holders, left/right split at dashed line), and absolute error at three time steps. Centralized and merged reconstructions are identical ($\delta_{\max} < 10^{-12}$). The Burgers experiment (1D, nonlinear) is in Appendix A.

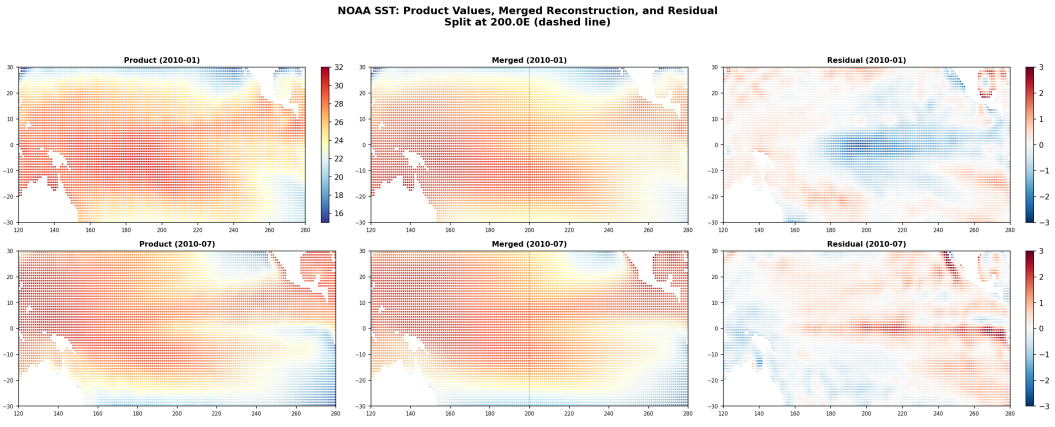


Figure 3: NOAA SST: product values (left), merged reconstruction from two data holders split at 200°E (center, dashed line), and reconstruction residual (reconstruction minus product, in °C, right) for January and July 2010. This is reconstruction error, not merged-versus-centralized disagreement.