CONDITION-ONCE FLOW MATCHING FOR CONDITIONAL MULTIVARIATE TABULAR GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Tabular data often pair observed covariates with multiple continuous responses whose joint distribution matters beyond a point prediction. We study conditional multivariate tabular generation, where covariates remain fixed while a designated response block is generated jointly, preserving conditional location, uncertainty, and cross-response dependence. Conventional conditional flow models repeatedly supply the same context to the velocity throughout transport. With a learnable conditional source, however, context information can already be encoded into the evolving state, leaving open how much direct velocity conditioning remains useful. We characterize this residual predictive value and show that it vanishes as the state becomes sufficient for velocity prediction, yielding condition saturation. Guided by this result, we introduce Condition-Once Flow Matching (CoFM), which uses a learnable conditional source and a velocity shared across contexts, together with Conditional Source Forcing (CS-Forcing) to control local source expansion over heterogeneous tabular fields. Controlled studies support the predicted saturation behavior. Across four real-world benchmarks covering Beijing air quality, Superconductivity, GTEx, and UK Biobank proteomics, CoFM-CSF consistently surpasses ten competitive probabilistic and generative baselines in overall predictive and distributional performance.

1 INTRODUCTION

Tabular data is one of the most pervasive forms of real-world data, underlying relational databases and spreadsheets across healthcare, finance, engineering, commerce, and scientific applications (Hollmann et al., 2025; Wang & Sun, 2022; Grinsztajn et al., 2022). In many of these datasets, only part of a record is available or treated as observed, while another set of variables must be inferred from the available fields (Hollmann et al., 2025; Shih et al., 2022; Zhang et al., 2025; Mitra et al., 2023; Viñas et al., 2023). This need underlies applications such as multiple imputation and conditional completion (Li et al., 2021; Mainzer et al., 2024; Shih et al., 2022; Zhang et al., 2025; Yoon et al., 2018; Mattei & Frellsen, 2019).

We focus on conditional multivariate tabular generation, which keeps the supplied context fixed and generates joint samples of a predefined block of continuous responses. Figure 1 contrasts this task with multi-output regression and full-table generation. Point predictions alone do not specify response uncertainty or dependence (Borchani et al., 2015; Duan et al., 2020). Full-table generation instead produces new records, including newly sampled context values (Xu et al., 2019; Shi et al., 2025; Yan et al., 2022; Kim et al., 2023). These records may be realistic, but they do not answer the original query. Our task requires learning how response location, uncertainty, and dependence vary with context. Predicting the mean of each response is not enough. The model must also capture how widely each response can vary and which response values are likely to occur together under the supplied context. These requirements concern the joint distribution for each context, not merely statistics pooled across the table.

Flow matching provides a natural generative framework for this setting by transporting samples from a simple source distribution to the target response distribution (Lipman et al., 2023; Pooladian et al., 2023). A common design starts from context-independent noise and supplies context to the velocity field throughout transport (Chen et al., 2025). Unlike text prompts in image generation, the context here consists of predefined numerical and categorical fields rather than rich, open-ended semantic descriptions (Yoon et al., 2020). A learnable conditional source can use these fields to

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.

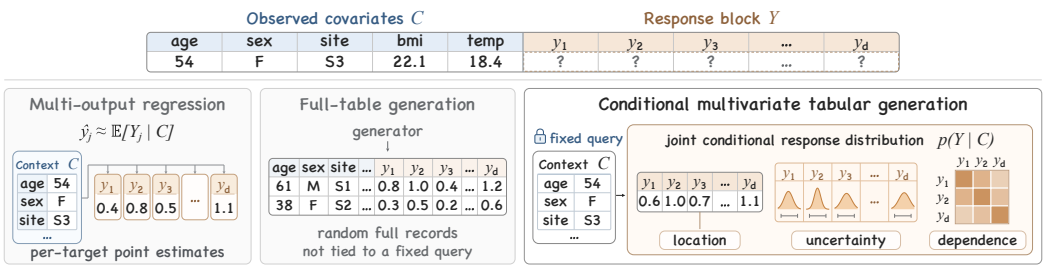

Figure 1: **Conditional multivariate tabular generation.** Multi-output regression produces per-target point estimates, while full-table generation synthesizes complete records without anchoring generation to a fixed query. Our setting keeps the observed covariates fixed and models the joint conditional distribution of a designated multivariate response block.

set the location and spread of the initial samples before transport begins (Issachar et al., 2025; Kim et al., 2026). Transport then starts from samples already adapted to the supplied context, rather than generic noise. This gives the velocity a way to use context information through the state itself, without reading the original fields again. We therefore study condition allocation between source and velocity. The key question is whether the evolving state carries enough context information to make repeated direct access unnecessary.

We quantify the remaining value of direct context access by comparing the best achievable velocity prediction errors with and without it. Both predictors observe the evolving state and time. Our analysis bounds this gap using the context information the state still lacks for velocity prediction. When the state provides all the information needed for this prediction, the gap vanishes. We call this regime condition saturation.

Guided by this principle, we introduce Condition-Once Flow Matching (CoFM). CoFM encodes context once through a learnable conditional source. Its velocity is shared across contexts and receives only the evolving response state and time. To control how this source responds to context changes, we introduce Conditional Source Forcing (CS-Forcing). It controls local source expansion over type-respecting tabular neighborhoods while preserving context-specific variation. Controlled studies verify the predicted link between the context information carried by the source and the remaining value of direct velocity conditioning. Across four real-world benchmarks, CoFM-CSF outperforms ten probabilistic and generative baselines in overall predictive and distributional performance. Our contributions are summarized below.

- **Condition saturation.** We establish condition saturation in learnable-source flow matching and bound conditional generation error through residual conditioning value. CoFM provides a tabular instantiation of this principle.
- **Conditional Source Forcing.** We introduce CS-Forcing to control local source expansion over type-respecting tabular neighborhoods through a finite budget, and relate this control to output sensitivity under shared transport.
- **Task and benchmarks.** We formulate conditional multivariate tabular generation and derive four real-world benchmarks, with ten baselines and a common protocol evaluating location, uncertainty, and cross-response dependence.

2 Related Work

Conditional tabular generation and multivariate responses Multi-output and distributional regression directly target responses conditioned on covariates, with Distributional Random Forests estimating multivariate conditional distributions and TreeFlow learning flexible conditional densities through tree-informed flows (Cevid et al., 2022; Wielopolski & Zieba, 2023). General-purpose tabular generators instead model complete-record distributions, including CTGAN and TabDDPM (Xu et al., 2019; Kotelnikov et al., 2023), while more recent models such as TabDiff and TabNAT support conditional completion from observed columns (Shi et al., 2025; Zhang et al., 2025). In our fixed-role setting, we jointly generate a designated multivariate continuous response block given an observed covariate block (the context). We target the conditional response distribution, including

location, uncertainty, and cross-response dependence, rather than point prediction, arbitrary-subset completion, or complete-record generation.

Conditioning design in flow matching Flow matching learns continuous transport through simulation-free vector-field regression, while subsequent work has improved probability paths, endpoint couplings, and source parameterization (Lipman et al., 2023; Tong et al., 2024). Conditioning design has increasingly moved beyond supplying context only to the velocity. CAR-Flow learns context-dependent source and target reparameterizations while retaining a context-conditioned velocity (Chen et al., 2025). CrossFlow uses a representation of the conditioning modality as its source (Liu et al., 2025). CSFM jointly learns a conditional Gaussian source and a velocity conditioned on context throughout transport (Kim et al., 2026), whereas Source-Guided Flow Matching conditions a fixed pretrained flow through inference-time source modification (Wang et al., 2026). Together, these works show that the source can carry substantial conditioning information and reshape the transport problem. They do not, however, characterize when direct context access by the velocity remains necessary once that information is already present in the evolving state. We study this residual need through condition allocation and condition saturation.

3 Condition Allocation in Learnable-Source Flow Matching

Learnable-source flow matching provides two routes for context: through the source-induced state and through direct velocity access. We analyze their allocation through the residual predictive value of direct context access, leading to condition saturation and its implications for shared transport. Proofs and technical details are in Appendix A.

3.1 Preliminaries

Flow matching with a learnable source. For a tabular record with observed covariates $C \in \mathcal{C}$ and continuous responses $Y \in \mathbb{R}^{d_y}$, conditional generation aims to reproduce $P_c^\star := \mathcal{L}(Y \mid C = c)$. Generation begins with $X_0 \sim Q_{\phi,c}$, where the learnable source $Q_{\phi,c}$ allows the initial distribution to carry covariate information. A velocity field $v(x, t, c)$ then transports this initial state. Assuming its ODE flow is well defined, the state $\Phi_t(x; c)$ reached from x satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t(x; c) = v(\Phi_t(x; c), t, c), \qquad \Phi_0(x; c) = x. \tag{1}$$

The output is $\widehat{Y} = \Phi_1(X_0; c)$, so covariates can shape generation through the source, the velocity, or both. When the velocity receives only state and time, the same flow map $\Phi_t(x)$ applies across contexts; conditional variation then enters entirely through $Q_{\phi,c}$.

To learn this transport, we pair source and target samples independently conditional on C, assuming finite second moments at both endpoints. The linear interpolation provides a training state and its velocity target:

$$X_t = (1 - t)X_0 + tY, \qquad U_t := \dot{X}_t = Y - X_0, \qquad t \sim \mathcal{U}(0, 1). \tag{2}$$

Here X_t denotes the training interpolation; the generated state $\Phi_t(X_0; C)$ generally follows a different trajectory. Flow matching trains the velocity to predict U_t from the interpolated state and time. Since X_t may already carry information about C, we compare what can be predicted with and without direct access to the covariates. For a fixed source and time t, the optimal squared-error predictors are the conditional and shared Bayes velocities:

$$v_t^\star(x, c) := \mathbb{E}[U_t \mid X_t = x, C = c], \tag{3}$$

$$\bar{v}_t^\star(x) := \mathbb{E}[U_t \mid X_t = x]. \tag{4}$$

Residual conditioning value. The discrepancy between these velocities measures what direct covariate access adds to the prediction:

$$\Gamma_t := \mathbb{E}\|v_t^\star(X_t, C) - \bar{v}_t^\star(X_t)\|_2^2.$$

For the same source and path, let $\mathcal{R}^\star_{\mathrm{cond},t}$ and $\mathcal{R}^\star_{\mathrm{sh},t}$ denote the minimum expected squared errors for predicting U_t from (X_t, C) and X_t, respectively. The L^2 projection identity relates this discrepancy to the reduction in prediction risk:

$$\mathcal{R}^\star_{\mathrm{sh},t} - \mathcal{R}^\star_{\mathrm{cond},t} = \Gamma_t = \mathbb{E}\big[\operatorname{tr}\operatorname{Cov}\big(v_t^\star(X_t, C) \mid X_t\big)\big]. \tag{5}$$

Thus, Γ_t quantifies the predictive value left in the covariates after the state is observed. It vanishes exactly when the state determines the conditional Bayes velocity almost surely. We use this remaining predictive value as the population criterion for condition allocation.

3.2 CONDITION SATURATION

For a conditionally independent copy C'_t of C given X_t, Eq. (5) becomes (Appendix A.5)

$$\Gamma_t = \frac{1}{2}\mathbb{E}\left\|v_t^\star(X_t, C) - v_t^\star(X_t, C'_t)\right\|_2^2.$$

Zero posterior velocity disagreement, $\Gamma_t = 0$, is *condition saturation*. A sufficient decoder-based route uses the *transport-relevant population distortion*

$$\Delta_t(c, c') := \operatorname*{ess\,sup}_{x\sim\mathcal{L}(X_t)} \left\|v_t^\star(x, c) - v_t^\star(x, c')\right\|_2. \tag{6}$$

It describes population velocity sensitivity, rather than a prescribed training metric. Two contexts can therefore have zero distortion even when their raw tabular values differ, provided those differences do not change the expected transport. Conversely, a small change in context can have large distortion when it substantially changes the required velocity. In particular, Δ_t is distinct from the designed table-aware metric d_{tab} used later by CS-Forcing.

Theorem 1 (Decoder-based condition-saturation bound)**.** *Under the simultaneous comparison condition in Appendix A.1, for any measurable decoder $\widehat{C}_t = \mathcal{D}_t(X_t) \in \mathcal{C}$ with finite right-hand side,*

$$\Gamma_t \le \mathbb{E}\left\|v_t^\star(X_t, C) - v_t^\star(X_t, \widehat{C}_t)\right\|_2^2 \le \mathbb{E}\Delta_t^2(C, \widehat{C}_t). \tag{7}$$

Thus, vanishing transport-relevant decoding error suffices for $\Gamma_t \to 0$.

The decoder is used for analysis, not generation. Only information that changes the expected velocity needs to be recovered. For example, if $R_t = r_t(C)$ and $v_t^\star(x, C) = q_t^\star(x, R_t)$, exact recovery of R_t from X_t already gives $\Gamma_t = 0$. Approximate recovery of such representations and fieldwise upper envelopes for Δ_t are developed in Appendix A.5.

3.3 CONSEQUENCES FOR CONDITIONAL GENERATION

Let Z_t^{cond} and Z_t^{sh} solve the population ODEs with velocities $v_t^\star(\cdot, C)$ and $\bar{v}_t^\star$, respectively, from the same draw $Z_0 \mid C = c \sim Q_{\phi,c}$. Denote their conditional endpoint laws by P_c^{cond} and P_c^{sh}. These generation trajectories are distinct from the training interpolation X_t.

Theorem 2 (Conditional endpoint error)**.** *Suppose the ODEs have unique solutions on $[0, 1]$, measurable in context and initialization, with finite second moments. Assume $\mathcal{L}(Z_t^{\mathrm{cond}} \mid C = c) = \mathcal{L}(X_t \mid C = c)$ for all t and almost every c, and that $\bar{v}_t^\star$ is $L_x(t)$-Lipschitz in state, with $L_x \ge 0$ and $L_x \in L_1[0, 1]$. Then*

$$\begin{aligned}\left(\mathbb{E}_C W_2^2(P_C^{\mathrm{cond}}, P_C^{\mathrm{sh}})\right)^{1/2} &\le \int_0^1 e^{\int_t^1 L_x(r)\,dr}\sqrt{\Gamma_t}\,dt \\ &\le \int_0^1 e^{\int_t^1 L_x(r)\,dr}\sqrt{\mathbb{E}\Delta_t^2(C, \widehat{C}_t)}\,dt.\end{aligned} \tag{8}$$

The first bound uses Γ_t directly; the second assumes Theorem 1 a.e. in t, jointly measurable decoder and distortion families, and a finite right-hand side. Path consistency gives $P_c^{\mathrm{cond}} = P_c^\star$ a.e. in c.

The bound connects residual conditioning value to conditional-distribution error, accounting for amplification by the shared dynamics. Exact saturation for almost every t yields $P_c^{\mathrm{sh}} = P_c^\star$ for almost every c. The relevant criterion is the accumulated residual value along the training path, not recoverability at initialization alone; the crossing construction in Appendix B.4 illustrates this distinction. These population bounds do not include finite-network fitting error.

3.4 ILLUSTRATING CONDITION ALLOCATION

Figure 2 visualizes the mechanism behind condition saturation in the Gaussian-ring construction. As the source strength α increases, the source-induced states become more informative about the

transport-relevant condition. The shared velocity can therefore infer more of the required transport directly from the evolving state, bringing its trajectories closer to the condition-specific references while reducing the residual conditioning value Γ_t. The corresponding improvement in condition-to-target assignment follows the same transition.

The relevant quantity is thus not source separation by itself, but whether the evolving state preserves the context information needed for velocity prediction along the transport path. Appendix B provides the complete source-strength sweep, finite-network replication, and a crossing-path construction that further isolates this pathwise requirement.

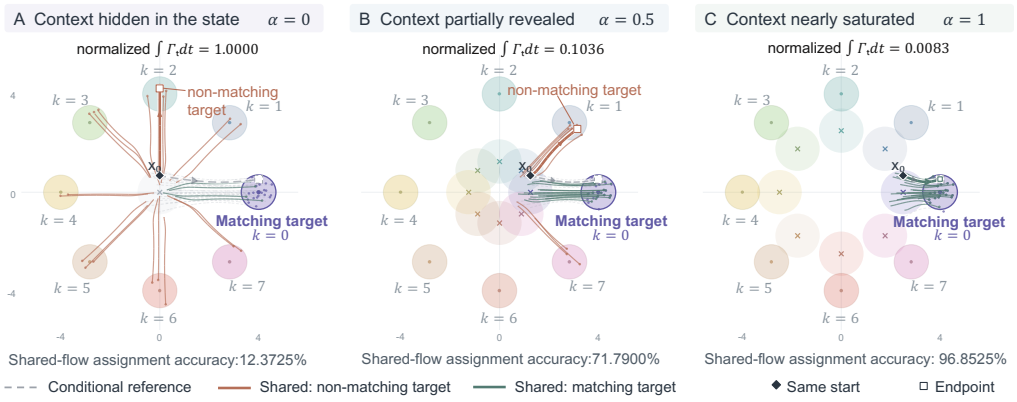

Figure 2: **Condition Saturation in a Gaussian Ring.** Panels A–C increase source strength α while fixing target distributions and all covariances. For condition $k = 0$, shared trajectories (solid; green for matching targets, orange otherwise) are paired with conditional references (gray dashed) from identical initial states. Crosses ($\times$) and dots ($\cdot$) mark starts and endpoints; diamonds and open squares highlight one pair. Assignment accuracy is evaluated across all eight conditions. Stronger source encoding brings shared transport closer to the conditional reference as the residual value of direct velocity conditioning decreases.

4 Condition-Once Flow Matching

Following the condition allocation principle in Section 3, CoFM, as shown in Figure 3, concentrates context dependence in a conditional source and uses a velocity shared across contexts. CS-Forcing further controls the local geometry of this source map over valid tabular neighborhoods.

4.1 Conditional Source and Shared Transport

Given context c, the source network g_ϕ defines a diagonal Gaussian source,

$$g_\phi(c) = \big(\mu_\phi(c), \sigma_\phi(c)\big), \qquad Q_{\phi,c} = \mathcal{N}\big(\mu_\phi(c), \mathrm{Diag}[\sigma_\phi^2(c)]\big)\,.$$

Samples from this source are reparameterized as

$$X_0 = \mu_\phi(c) + \sigma_\phi(c) \odot \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I_{d_y}). \tag{9}$$

CoFM uses a velocity $v_\theta(x, t)$ that receives only the response state and time. Its flow map Φ_t, defined in Section 3.1, is shared across contexts, and generation at fixed c returns $\widehat{Y} = \Phi_1(X_0)$. The conditional source supplies covariate dependence, while shared transport shapes the joint response distribution.

4.2 Conditional Source Forcing

In CoFM, the conditional source is the direct interface through which context enters generation. Condition saturation concerns whether transport-relevant context is available in the evolving state; CS-Forcing addresses the complementary question of how that information is encoded into the source. We therefore control local source expansion over valid tabular neighborhoods. For a context $c = (c_1, \ldots, c_{d_f})$, we construct a valid neighbor $\tilde{c}$ by changing one logical field according to its type:

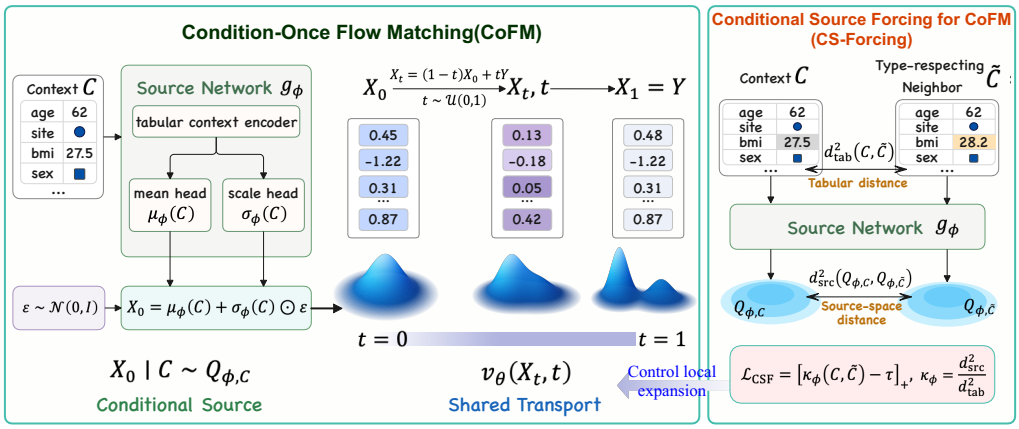

Figure 3: **CoFM architecture for conditional multivariate tabular generation.** Observed covariates enter through the conditional source, followed by a shared velocity. CS-Forcing acts during training on the local geometry of the source map over type-respecting tabular neighborhoods.

continuous fields use scale-normalized local changes, ordinal fields use normalized rank moves, and binary or categorical fields use valid state changes. The corresponding fieldwise distances define

$$d^2_{\text{tab}}(c, \tilde{c}) = \sum_{j=1}^{d_f} d^2_j(c_j, \tilde{c}_j) > 0. \tag{10}$$

This places heterogeneous field changes on a common local scale; the complete neighborhood construction is given in Appendix C.1.

The source map $c \mapsto Q_{\phi,c}$ carries these local neighborhoods into the space of source distributions. We measure the induced displacement by the response-dimension-normalized Wasserstein distance

$$d^2_{\text{src}}(Q_{\phi,c}, Q_{\phi,\tilde{c}}) := \frac{1}{d_y} W_2^2(Q_{\phi,c}, Q_{\phi,\tilde{c}}).$$

For the diagonal Gaussian source in Section 4.1,

$$d^2_{\text{src}}(Q_{\phi,c}, Q_{\phi,\tilde{c}}) = \frac{\|\mu_\phi(c) - \mu_\phi(\tilde{c})\|_2^2 + \|\sigma_\phi(c) - \sigma_\phi(\tilde{c})\|_2^2}{d_y},$$

and we define the local source expansion

$$\kappa_\phi(c, \tilde{c}) := \frac{d^2_{\text{src}}(Q_{\phi,c}, Q_{\phi,\tilde{c}})}{d^2_{\text{tab}}(c, \tilde{c})}. \tag{11}$$

Thus, κ_ϕ measures source displacement relative to the corresponding local tabular change. CS-Forcing places a finite upper budget τ on this expansion:

$$\mathcal{L}_{\text{CSF}} = \mathbb{E}_{C,\tilde{C}}\left[\kappa_\phi(C, \tilde{C}) - \tau\right]_+, \qquad [a]_+ = \max(a, 0). \tag{12}$$

The one-sided objective is inactive within the budget, leaving task learning free to preserve context-dependent source differences that are useful for transport, and acts only when a local context change is expanded disproportionately in source space. Under shared transport, controlling source expansion also bounds local output sensitivity; Appendix C gives the theoretical interpretation and controlled analysis.

4.3 Joint Training and Inference

Using the conditional source in Equation (9) and the linear probability path in Equation (2), we jointly optimize the source network g_ϕ and the shared velocity v_θ with the flow-matching objective

$$\mathcal{L}_{\text{FM}}(\phi, \theta) = \mathbb{E}\left[\frac{1}{d_y} \|v_\theta(X_t, t) - U_t\|_2^2\right]. \tag{13}$$

The velocity is trained from the response state and time alone; context influences transport through the source-induced state rather than through a separate velocity input. We retain the source-variance control and source-target alignment terms in CSFM, and combine them with CS-Forcing as

$$\mathcal{L}_{\text{total}}(s) = \mathcal{L}_{\text{FM}} + \lambda_{\text{var}}\mathcal{L}_{\text{var}} + \lambda_{\text{align}}\mathcal{L}_{\text{align}} + \lambda_{\text{CSF}}(s)\mathcal{L}_{\text{CSF}}, \tag{14}$$

where s denotes the optimization step. CS-Forcing acts only on the conditional source map and is activated after the initial 10% of optimization. Definitions and coefficients of the auxiliary source terms are given in Appendix D.3.

In our tabular instantiation, g_ϕ maps the processed context C through a tabular encoder, followed by separate mean and log-scale heads that parameterize $Q_{\phi,c}$. We use a residual MLP for the source encoder. For the shared velocity v_θ, we consider two backbones, an MLP and a diffusion transformer (DiT). Both receive only the current response state and flow time and output a d_y-dimensional velocity. Detailed architectural configurations are provided in Appendix D.3. At inference, the fixed context c parameterizes $Q_{\phi,c}$, from which X_0 is sampled once; generation then follows the shared velocity $v_\theta(x,t)$ without further context access. CS-Forcing is used only during training, with solver choices and dataset-specific configurations deferred to Appendix D.3.

Algorithm 1: Training CoFM with CS-Forcing

Require: Data $\mathcal{D}_{\text{train}}$; networks g_ϕ, v_θ; batch size B; steps S

for $s = 1, \dots, S$ **do**
  Draw minibatch $\{(c_b, y_b)\}_{b=1}^{B}$ from $\mathcal{D}_{\text{train}}$
  Reparameterize $x_{0,b}$ (9); $t_b \sim \mathcal{U}(0,1)$
  Form $x_{t_b,b}$ and u_b via Eq. (2)
  Compute $\mathcal{L}_{\text{FM}}$ [Eq. (13)] and $\mathcal{L}_{\text{var}}, \mathcal{L}_{\text{align}}$ [Appendix D.3]
  $\mathcal{L}_{\text{CSF}} \leftarrow 0$
  if $\lambda_{\text{CSF}}(s) > 0$ **then**
    $\mathcal{I}_{\text{valid}} \leftarrow \{b : c_b \text{ admits a valid local move}\}$
    Sample type-respecting $\tilde{c}_b$, $b \in \mathcal{I}_{\text{valid}}$
    $\mathcal{L}_{\text{CSF}} \leftarrow \frac{1}{B} \sum_{b \in \mathcal{I}_{\text{valid}}} [\, \kappa_\phi(c_b, \tilde{c}_b) - \tau \,]_+$
  end if
  Update (ϕ, θ) using $\nabla \mathcal{L}_{\text{total}}$ [Eq. (14)]
end for
return (ϕ, θ)

5 Experiments

5.1 Experimental Setup

Benchmarks. We derive four fixed-role conditional multivariate generation benchmarks from Beijing, Superconductivity, GTEx, and UKB-PPP (Zhang et al., 2017; Hamidieh, 2018; GTEx Consortium, 2020; Sun et al., 2023; Sudlow et al., 2015). In every task, the covariate block C is observed and held fixed, while a continuous multivariate response block Y is generated jointly. The response blocks comprise 72 pollutant measurements conditioned on weather and time variables, 37 material descriptors conditioned on composition summaries, 37 gene-expression responses conditioned on landmark expression and tissue/donor attributes, and 150 protein measurements conditioned on clinical variables, respectively. The corresponding splits are chronological, fixed random, donor-grouped, and fixed random.

Baselines. Our comparisons cover direct parametric generators (Gaussian and MDN (Bishop, 1994)), latent-variable and adversarial models (CVAE (Sohn et al., 2015) and CTGAN (Xu et al., 2019)), normalizing flows (RQ-NSF (Durkan et al., 2019) and TreeFlow (Wielopolski & Zieba, 2023)), tabular diffusion models (TabDDPM (Kotelnikov et al., 2023) and TabDiff (Shi et al., 2025)), and reparameterized or learnable-source flow matching (CAR-Flow (Chen et al., 2025) and CSFM (Kim et al., 2026)). Gaussian and MDN use the same tabular encoder family as our source network and test direct conditional generation without transport. Methods originally designed for full-table generation or completion are adapted to generate Y while keeping C fixed; Appendix D.2 gives the task-specific adaptations.

Metrics and protocol. We assess conditional location with R^2, marginal predictive distributions with CRPS, overall joint distributional fidelity with Energy Score, and cross-response dependence with Variogram Score. All methods use the same test partitions and evaluation protocol; table entries report means and sample standard deviations over three training seeds. Appendices D.1–D.3 and D.5 give task, model, and evaluation details.

5.2 Main Results

We first evaluate whether source-concentrated conditioning can recover the conditional multivariate response distribution across its complementary aspects: location, marginal predictive distributions, and cross-response structure.

Table 1: **Conditional multivariate tabular generation results on four benchmark datasets.** Bold indicates the best, underline the second best result.

Method	Beijing				Superconductivity			
	$R^2\uparrow$	CRPS↓	Variogram↓	Energy↓	$R^2\uparrow$	CRPS↓	Variogram↓	Energy↓
Gaussian	0.3386±0.0117	0.4241±0.0042	0.1900±0.0019	4.6317±0.0359	0.8642±0.0021	0.1337±0.0013	0.0431±0.0005	1.2220±0.0118
MDN	0.3649±0.0456	0.3933±0.0168	0.1427±0.0051	4.2764±0.1612	0.8673±0.0092	0.1420±0.0038	0.0478±0.0017	1.2679±0.0353
CVAE	0.3720±0.0048	0.3954±0.0021	0.1350±0.0006	4.2552±0.0206	0.8768±0.0014	0.1408±0.0005	0.0438±0.0002	1.1796±0.0068
CTGAN	0.4155±0.0540	0.3892±0.0214	0.1336±0.0054	4.1967±0.1847	0.9109±0.0061	0.1557±0.0081	0.0431±0.0022	1.2144±0.0561
RQ-NSF	0.4299±0.0369	0.3781±0.0143	0.1256±0.0040	4.1706±0.1298	0.9878±0.0024	0.0274±0.0014	0.0058±0.0004	0.2738±0.0138
TreeFlow	0.4433±0.0350	0.3719±0.0145	0.1263±0.0057	4.0654±0.1548	0.8797±0.0065	0.1460±0.0036	0.0454±0.0020	1.1865±0.0304
TabDDPM	0.3807±0.1126	0.3741±0.0327	0.1250±0.0087	4.1131±0.3098	0.9772±0.0014	0.0639±0.0021	0.0135±0.0009	0.5212±0.0159
TabDiff	0.1089±0.0051	0.4780±0.0031	0.1607±0.0012	5.0988±0.0276	0.6569±0.0061	0.2509±0.0020	0.0835±0.0015	2.0389±0.0090
CAR-Flow	0.3789±0.0325	0.4010±0.0109	0.1327±0.0045	4.3754±0.1133	0.9912±0.0003	0.0390±0.0002	0.0063±0.0001	0.3196±0.0010
CSFM	0.4727±0.0163	0.3603±0.0079	0.1190±0.0027	3.9854±0.0684	0.9882±0.0002	0.0478±0.0008	0.0085±0.0000	0.3903±0.0033
CoFM-CSF (MLP)	**0.5390**±0.0049	0.3335±0.0040	**0.1117**±0.0005	3.6878±0.0323	0.9952±0.0000	0.0183±0.0002	**0.0025**±0.0001	**0.1528**±0.0020
CoFM-CSF (DiT)	0.5343±0.0096	**0.3285**±0.0038	0.1123±0.0020	**3.6758**±0.0305	**0.9953**±0.0001	**0.0182**±0.0003	0.0025±0.0000	0.1536±0.0026

Method	GTEx				UKB-PPP			
	$R^2\uparrow$	CRPS↓	Variogram↓	Energy↓	$R^2\uparrow$	CRPS↓	Variogram↓	Energy↓
Gaussian	0.5920±0.0002	0.3252±0.0001	0.1176±0.0001	2.6830±0.0009	0.1160±0.0001	0.5224±0.0006	0.1771±0.0005	8.1236±0.0092
MDN	0.5384±0.0083	0.3506±0.0039	0.1208±0.0007	2.8416±0.0213	0.1925±0.0139	0.4892±0.0060	0.1637±0.0005	7.6902±0.0797
CVAE	0.6364±0.0024	0.2977±0.0005	0.1041±0.0003	2.5131±0.0043	0.2652±0.0033	0.4589±0.0006	0.1560±0.0002	7.3215±0.0109
CTGAN	0.3455±0.0275	0.4807±0.0142	0.1691±0.0033	3.7004±0.0552	0.2527±0.0065	0.5611±0.0036	0.1932±0.0012	8.8470±0.0561
RQ-NSF	0.6457±0.0022	0.2962±0.0021	0.1036±0.0006	2.5064±0.0099	0.2713±0.0018	0.4600±0.0008	0.1592±0.0002	7.3339±0.0116
TreeFlow	0.5911±0.0021	0.3275±0.0010	0.1144±0.0004	2.6756±0.0064	0.2659±0.0006	0.4628±0.0007	0.1591±0.0002	7.3395±0.0050
TabDDPM	0.6426±0.0008	0.2916±0.0004	0.1021±0.0007	2.4970±0.0031	0.2911±0.0010	0.4569±0.0003	0.1591±0.0001	7.2379±0.0032
TabDiff	0.3257±0.0103	0.4472±0.0049	0.1678±0.0022	3.4148±0.0300	0.0922±0.0831	0.5304±0.0234	0.1726±0.0017	8.2222±0.3450
CAR-Flow	0.3992±0.0041	0.4607±0.0024	0.1677±0.0010	3.9073±0.0251	0.2130±0.0037	0.4767±0.0021	0.1591±0.0005	7.5867±0.0261
CSFM	0.6120±0.0011	0.3127±0.0009	0.1090±0.0005	2.6736±0.0069	0.2831±0.0034	0.4550±0.0011	**0.1560**±0.0003	7.2583±0.0152
CoFM-CSF (MLP)	0.6509±0.0007	0.2933±0.0012	0.1014±0.0002	2.4721±0.0040	0.3029±0.0010	0.4530±0.0003	0.1565±0.0001	7.1959±0.0047
CoFM-CSF (DiT)	**0.6519**±0.0015	**0.2874**±0.0012	**0.1001**±0.0004	**2.4652**±0.0046	**0.3030**±0.0002	**0.4504**±0.0009	0.1564±0.0008	**7.1887**±0.0127

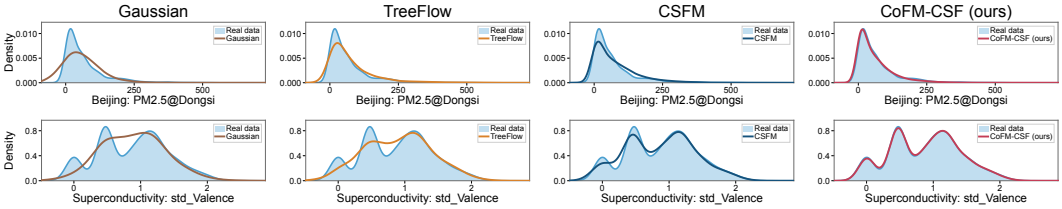

Figure 4: **Marginal density comparison on Beijing and Superconductivity.** Pooled densities are shown for $PM_{2.5}$ and std_Valence, respectively. Blue shading denotes observed data.

Overall performance. CoFM-CSF performs strongly across these aspects even though its velocity never reads the covariates directly. Both the MLP and DiT variants outperform all comparison methods in R^2 and Energy Score on all four datasets. The DiT variant achieves the lowest CRPS on all four datasets. For Variogram Score, CoFM-CSF gives the best results on Beijing, Superconductivity, and GTEx, while CSFM is slightly lower on UKB-PPP. The same condition allocation therefore supports accurate conditional location and joint distributional modeling across two substantially different transport backbones.

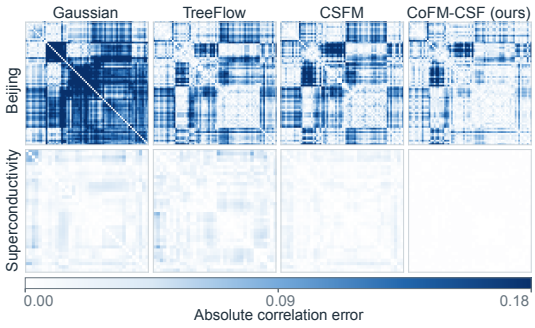

Figure 5: **Pooled response correlation errors on Beijing and Superconductivity.** Absolute errors between generated and observed pooled response correlations are shown on a common scale.

Role of transport. Gaussian and MDN directly parameterize $p(Y \mid C)$ using the same tabular encoder family as our source network but without a learned transport stage. Their weaker scores across all four tasks indicate that CoFM-CSF's gains are not explained by the conditional source alone. Shared transport contributes substantially to shaping non-Gaussian marginals and multivariate response structure. A diagonal Gaussian source therefore does not restrict the generated responses to a Gaussian or independent output family. Figures 4–5 provide pooled distributional diagnostics that complement the conditional proper scores in Table 1. Appendix E.1–E.2 extends these analyses across all four benchmarks.

Performance across datasets. The larger gains on Beijing and Superconductivity span marginal and joint scores, whereas the narrower margins on GTEx and UKB-PPP indicate more limited benefits over the leading alternatives. On UKB-PPP, CSFM retains the best Variogram Score despite CoFM-CSF's stronger R^2 and Energy Score. This separates overall joint fidelity from dependence-specific recovery and identifies the latter as a remaining challenge on this dataset.

5.3 Condition Saturation on Real Data

The allocation argument predicts that encoding the covariates in the source should reduce the benefit of direct velocity conditioning. We test this by freezing each learned source and restoring covariate dependence in its mean and scale from fixed averages ($\alpha = 0$) to its learned outputs ($\alpha = 1$). At each level, we refit state-only and conditioned velocity probes on paired interpolation samples. The normalized residual value decreases by 73–98% across the four datasets (Figure 6). Since this gap measures what conditioning adds after observing the state, its contraction supports the proposed substitution: covariates encoded in the source make direct velocity conditioning less useful. These finite-probe results connect the predictive success above to the condition-saturation principle in Section 3. Probe construction and capacity controls are in Appendix D.4.

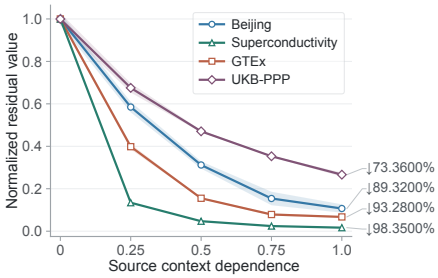

Figure 6: **Residual-conditioning probes on four datasets.** Normalized, time-integrated held-out risk gap between state-only and conditioned probes as source dependence on covariates varies with α.

5.4 Condition Allocation and CS-Forcing Ablations

Table 2: **Matched 2×2 ablation of condition allocation and CS-Forcing with the MLP backbone**

Method	Beijing		Superconductivity		GTEx		UKB-PPP	
	CRPS↓	Variogram↓	CRPS↓	Variogram↓	CRPS↓	Variogram↓	CRPS↓	Variogram↓
Dual	0.4330±0.0033	0.1422±0.0027	0.0288±0.0002	0.0037±0.0001	0.2948±0.0013	0.1020±0.0004	0.4553±0.0005	0.1583±0.0001
Dual + CSF	0.3584±0.0179	0.1477±0.0057	0.0239±0.0001	0.0031±0.0000	0.2945±0.0002	0.1020±0.0001	0.4559±0.0005	0.1583±0.0003
Once	0.4224±0.0114	0.1391±0.0042	0.0280±0.0002	0.0037±0.0001	0.2935±0.0022	**0.1014**±0.0004	0.4551±0.0006	0.1580±0.0001
CoFM-CSF	**0.3335**±0.0040	**0.1117**±0.0005	**0.0183**±0.0002	**0.0025**±0.0001	**0.2933**±0.0012	0.1014±0.0002	**0.4530**±0.0003	**0.1565**±0.0001

To isolate condition allocation within the CoFM design, we construct matched Dual and Once variants that share the same source architecture, velocity backbone, training objective, and optimization protocol. Dual additionally supplies the context to the velocity, whereas Once removes this direct input and conditions generation only through the source. These variants are controlled ablations of CoFM rather than reimplementations of the standalone CSFM baseline in Section 5.2. Adding CS-Forcing gives Dual + CSF and CoFM-CSF, respectively.

Without CS-Forcing, Once remains comparable to or modestly improves over Dual across the four datasets. Within this matched design, removing direct velocity access therefore does not produce a systematic loss, consistent with the residual-conditioning results in Section 5.3. Adding CS-Forcing gives its clearest gains under Once, particularly on Beijing and Superconductivity. On Beijing, CS-Forcing reduces CRPS from 0.4224 to 0.3335 and Variogram Score from 0.1391 to 0.1117 under Once, improving marginal prediction and dependence recovery together. This interaction is consistent with the intended role of CS-Forcing, as the source is the model's direct conditioning interface under Condition-Once, while Dual retains an additional condition route outside source-side control.

6 Conclusion

We studied conditional multivariate tabular generation with fixed observed covariates. Our analysis of learnable-source flow matching establishes the condition saturation principle. The value of direct context access vanishes as the evolving state becomes sufficient for velocity prediction. Guided by this principle, CoFM encodes context once through a learnable conditional source and uses a velocity field shared across contexts. CS-Forcing controls local source expansion over type-respecting tabular neighborhoods. Controlled studies and real-data probes support the predicted saturation behavior. Across four benchmarks, CoFM-CSF achieves strong predictive and distributional performance against ten baselines. These results support encoding context at the source while using shared transport to model response shape and cross-response dependence.

AI USE STATEMENT

We used AI-assisted tools to support several aspects of this work, including identifying potentially relevant literature, improving the clarity and presentation of the manuscript, assisting with document formatting, and supporting the implementation of experimental code and the aggregation of experimental results. All retrieved references, generated code, experimental outputs, and manuscript content were reviewed and verified by the authors. The research design, methodological decisions, interpretation of results, and scientific conclusions were made by the authors.

ETHICS STATEMENT

Synthetic outcomes are model samples and do not replace measurements. The models provide no formal privacy guarantee and can reproduce biases in their source data. The biomedical evaluations concern statistical generation and support no clinical-use claim. UK Biobank data were accessed under an approved application. Access and redistribution are governed by UK Biobank's approved-use conditions.

REPRODUCIBILITY STATEMENT

The appendices specify the theoretical assumptions, dataset construction, model configurations, and training and evaluation protocols. Reported aggregates use three training seeds. The accompanying scripts regenerate the benchmark tables and figures; the type-aware objective is documented in Appendix C.2, and the ablation is reported in Section 5.4.

REFERENCES

Christopher M. Bishop. Mixture density networks. Technical Report NCRG/94/004, Aston University, 1994.

Hanen Borchani, Gherardo Varando, Concha Bielza, and Pedro Larrañaga. A survey on multi-output regression. *WIREs Data Mining and Knowledge Discovery*, 5(5):216–233, 2015. doi: 10.1002/widm.1157.

Domagoj Cevid, Loris Michel, Jeffrey Näf, Peter Bühlmann, and Nicolai Meinshausen. Distributional random forests: Heterogeneity adjustment and multivariate distributional regression. *Journal of Machine Learning Research*, 23(333):1–79, 2022.

Chen Chen, Pengsheng Guo, Liangchen Song, Jiasen Lu, Rui Qian, Tsu-Jui Fu, Xinze Wang, Wei Liu, Yinfei Yang, and Alex Schwing. CAR-Flow: Condition-aware reparameterization aligns source and target for better flow matching. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-3173.

Tony Duan, Anand Avati, Daisy Yi Ding, Khanh K. Thai, Sanjay Basu, Andrew Ng, and Alejandro Schuler. NGBoost: Natural gradient boosting for probabilistic prediction. In *International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 2690–2700. PMLR, 2020.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-0037.

GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509):1318–1330, 2020. doi: 10.1126/science.aaz1776.

Kam Hamidieh. A data-driven statistical model for predicting the critical temperature of a superconductor. *Computational Materials Science*, 154:346–354, 2018. doi: 10.1016/j.commatsci.2018.07.052.

Noah Hollmann, Samuel Müller, Lennart Purucker, Arjun Krishnakumar, Max Körfer, Shi Bin Hoo, Robin Tibor Schirrmeister, and Frank Hutter. Accurate predictions on small data with a tabular foundation model. *Nature*, 637:319–326, 2025. doi: 10.1038/s41586-024-08328-6.

Noam Issachar, Mohammad Salama, Raanan Fattal, and Sagi Benaim. Designing a conditional prior distribution for flow-based generative models. *Transactions on Machine Learning Research*, 2025.

Jayoung Kim, Chaejeong Lee, and Noseong Park. STaSy: Score-based tabular data synthesis. In *International Conference on Learning Representations*, 2023.

Junwan Kim, Jiho Park, Seonghu Jeon, and Seungryong Kim. Better source, better flow: Learning condition-dependent source distribution for flow matching. arXiv preprint arXiv:2602.05951, 2026.

Akim Kotelnikov, Dmitry Baranchuk, Ivan Rubachev, and Artem Babenko. TabDDPM: Modelling tabular data with diffusion models. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 17564–17579. PMLR, 2023.

Jiang Li, Xiaowei S. Yan, Durgesh Chaudhary, Venkatesh Avula, Satish Mudiganti, Hannah Husby, Shima Shahjouei, Ardavan Afshar, Walter F. Stewart, Mohammed Yeasin, Ramin Zand, and Vida Abedi. Imputation of missing values for electronic health record laboratory data. *npj Digital Medicine*, 4:147, 2021. doi: 10.1038/s41746-021-00518-0.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Qihao Liu, Xi Yin, Alan Yuille, Andrew Brown, and Mannat Singh. Flowing from words to pixels: A noise-free framework for cross-modality evolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2755–2765, 2025.

Rheanna M. Mainzer, Margarita Moreno-Betancur, Cattram D. Nguyen, Julie A. Simpson, John B. Carlin, and Katherine J. Lee. Gaps in the usage and reporting of multiple imputation for incomplete data: Findings from a scoping review of observational studies addressing causal questions. *BMC Medical Research Methodology*, 24:193, 2024. doi: 10.1186/s12874-024-02302-6.

Pierre-Alexandre Mattei and Jes Frellsen. MIWAE: Deep generative modelling and imputation of incomplete data sets. In *International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4413–4423. PMLR, 2019.

Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.

Robin Mitra, Sarah F. McGough, Tapabrata Chakraborti, Chris Holmes, Ryan Copping, Niels Hagenbuch, Stefanie Biedermann, Jack Noonan, Brieuc Lehmann, Aditi Shenvi, Xuan Vinh Doan, David Leslie, Ginestra Bianconi, Ruben Sanchez-Garcia, Alisha Davies, Maxine Mackintosh, Eleni-Rosalina Andrinopoulou, Anahid Basiri, Chris Harbron, and Ben D. MacArthur. Learning from data with structured missingness. *Nature Machine Intelligence*, 5:13–23, 2023. doi: 10.1038/s42256-022-00596-z.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample flow matching: Straightening flows with minibatch couplings. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 28100–28127. PMLR, 2023.

Juntong Shi, Minkai Xu, Harper Hua, Hengrui Zhang, Stefano Ermon, and Jure Leskovec. TabDiff: A mixed-type diffusion model for tabular data generation. In *International Conference on Learning Representations*, 2025.

Andy Shih, Dorsa Sadigh, and Stefano Ermon. Training and inference on any-order autoregressive models the right way. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-0200.

Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, 12(3):e1001779, 2015. doi: 10.1371/journal.pmed.1001779.

Benjamin B. Sun, Joshua Chiou, Matthew Traylor, Christian Benner, Yi-Hsiang Hsu, Tom G. Richardson, Praveen Surendran, Anubha Mahajan, Chloe Robins, Steven G. Vasquez-Grinnell, Liping Hou, Erika M. Kvikstad, Oliver S. Burren, Jonathan Davitte, Kyle L. Ferber, Christopher E. Gillies, Åsa K. Hedman, Sile Hu, Tinchi Lin, Rajesh Mikkilineni, Rion K. Pendergrass, Corran Pickering, Bram Prins, Denis Baird, Chia-Yen Chen, Lucas D. Ward, Aimee M. Deaton, Samantha Welsh, Carissa M. Willis, Nick Lehner, Matthias Arnold, Maria A. Wörheide, Karsten Suhre, Gabi Kastenmüller, Anurag Sethi, Madeleine Cule, Anil Raj, Alnylam Human Genetics, AstraZeneca Genomics Initiative, Biogen Biobank Team, Bristol Myers Squibb, Genentech Human Genetics, GlaxoSmithKline Genomic Sciences, Pfizer Integrative Biology, Population Analytics of Janssen Data Sciences, Regeneron Genetics Center, Lucy Burkitt-Gray, Eugene Melamud, Mary Helen Black, Eric B. Fauman, Joanna M. M. Howson, Hyun Min Kang, Mark I. McCarthy, Paul Nioi, Slavé Petrovski, Robert A. Scott, Erin N. Smith, Sándor Szalma, Dawn M. Waterworth, Lyndon J. Mitnaul, Joseph D. Szustakowski, Bradford W. Gibson, Melissa R. Miller, and Christopher D. Whelan. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature*, 622:329–338, 2023. doi: 10.1038/s41586-023-06592-6.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

Ramon Viñas, Chaitanya K. Joshi, Dobrik Georgiev, Phillip Lin, Bianca Dumitrascu, Eric R. Gamazon, and Pietro Liò. Hypergraph factorization for multi-tissue gene expression imputation. *Nature Machine Intelligence*, 5:739–753, 2023. doi: 10.1038/s42256-023-00684-8.

Zifan Wang, Alice Harting, Matthieu Barreau, Michael M. Zavlanos, and Karl H. Johansson. Source-guided flow matching. In *International Conference on Learning Representations*, 2026.

Zifeng Wang and Jimeng Sun. TransTab: Learning transferable tabular transformers across tables. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-0210.

Patryk Wielopolski and Maciej Zieba. TreeFlow: Going beyond tree-based parametric probabilistic regression. In *European Conference on Artificial Intelligence*, volume 372 of *Frontiers in Artificial Intelligence and Applications*, pp. 2631–2638, 2023. doi: 10.3233/FAIA230570.

Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. Modeling tabular data using conditional GAN. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Chao Yan, Yao Yan, Zhiyu Wan, Ziqi Zhang, Larsson Omberg, Justin Guinney, Sean D. Mooney, and Bradley A. Malin. A multifaceted benchmarking of synthetic electronic health record generation models. *Nature Communications*, 13:7609, 2022. doi: 10.1038/s41467-022-35295-1.

Jinsung Yoon, James Jordon, and Mihaela van der Schaar. GAIN: Missing data imputation using generative adversarial nets. In *International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 5689–5698. PMLR, 2018.

Jinsung Yoon, Yao Zhang, James Jordon, and Mihaela van der Schaar. VIME: Extending the success of self- and semi-supervised learning to tabular domain. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Hengrui Zhang, Liancheng Fang, Qitian Wu, and Philip S. Yu. TabNAT: A continuous-discrete joint generative framework for tabular data. In *International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 74858–74881. PMLR, 2025.

Shuyi Zhang, Bin Guo, Anlan Dong, Jing He, Ziping Xu, and Song Xi Chen. Cautionary tales on air-quality improvement in Beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2205):20170457, 2017. doi: 10.1098/rspa.2017.0457.

A PROOFS

We derive the residual-conditioning identity in Eq. (5) and prove Theorems 1–2, then add results on decoding and shared transport.

A.1 NOTATION AND ASSUMPTIONS

Let the conditioning variables C (observed covariates in our tabular task) take values in a standard Borel space $\mathcal{C}$, with $X_0, Y \in \mathbb{R}^{d_y}$. Write $P_c^\star = \mathcal{L}(Y \mid C = c)$ and draw $X_0 \mid C = c \sim Q_{\phi,c}$ with $X_0 \perp Y \mid C$. We fix the source, endpoint coupling, and training path throughout each comparison and assume

$$\mathbb{E}\|X_0\|_2^2 + \mathbb{E}\|Y\|_2^2 < \infty, \qquad X_t = (1-t)X_0 + tY, \qquad U_t = Y - X_0. \tag{A1}$$

Thus $X_t, U_t \in L^2$. Time is fixed in the regression identities and integrated over $[0,1]$ for the uniform-time objective; squared risks sum over response coordinates. Random-variable identities hold almost surely, and sigma-field equalities are modulo null sets. Almost-everywhere statements in context and time refer to the law of C and Lebesgue measure, respectively.

The conditional and shared Bayes velocities are

$$v_t^\star(x,c) := \mathbb{E}[U_t \mid X_t = x, C = c], \qquad \bar{v}_t^\star(x) := \mathbb{E}[U_t \mid X_t = x]. \tag{A2}$$

We fix versions on the full state and context domains that are jointly measurable, including in time. Define the transport-relevant population distortion by

$$\Delta_t(c,c') := \operatorname*{ess\,sup}_{x \sim \mathcal{L}(X_t)} \|v_t^\star(x,c) - v_t^\star(x,c')\|_2, \qquad c, c' \in \mathcal{C}. \tag{A3}$$

Simultaneous comparison condition. For decoding bounds, assume Δ_t is measurable and

$$\|v_t^\star(x,c) - v_t^\star(x,c')\|_2 \le \Delta_t(c,c') \qquad \text{for all } c, c' \in \mathcal{C} \tag{A4}$$

holds on a single $\mathcal{L}(X_t)$-full-measure state set. A decoder is a measurable map $\mathcal{D}_t$ with $\widehat{C}_t = \mathcal{D}_t(X_t) \in \mathcal{C}$.

Table S1 summarizes the notation. We distinguish the training interpolation X_t from generation trajectories Z_t; the ODE assumptions are stated in Appendix A.4.

Table S1: **Notation used in the proofs.**

Symbol	Meaning
d_c, d_f, d_y	Encoded conditioning dimension, logical-field count, and response dimension
$Q_{\phi,c}, P_c^\star$	Conditional source and target laws
X_0, X_t, U_t	Training source sample, interpolation state, and velocity target
$v_t^\star, \bar{v}_t^\star$	Conditional and shared Bayes velocities
$\mathcal{R}^\star_{\mathrm{cond},t}, \mathcal{R}^\star_{\mathrm{sh},t}$	Optimal conditional and shared squared risks
Γ_t	Residual value of direct velocity conditioning
$\mathcal{D}_t, \widehat{C}_t, \Delta_t$	Decoder, recovered conditioning values, and population distortion
$R_t = r_t(C), q_t^\star$	Transport-relevant representation and its velocity factorization
$Z_t^{\mathrm{cond}}, Z_t^{\mathrm{sh}}$	Population generation trajectories from a common initialization
$P_c^{\mathrm{cond}}, P_c^{\mathrm{sh}}$	Conditional endpoint laws of the two population flows
v_θ, Φ_t	Learned shared velocity and its deterministic flow map

A.2 DERIVATION OF THE RESIDUAL-CONDITIONING IDENTITY

Proof. For $\mathcal{G} \in \{\sigma(X_t), \sigma(X_t, C)\}$, let $m_{\mathcal{G}} = \mathbb{E}[U_t \mid \mathcal{G}]$. Conditional expectation is an L^2 orthogonal projection, so every square-integrable $\mathcal{G}$-measurable predictor f satisfies

$$\mathbb{E}\|U_t - f\|_2^2 = \mathbb{E}\|U_t - m_{\mathcal{G}}\|_2^2 + \mathbb{E}\|m_{\mathcal{G}} - f\|_2^2. \tag{A5}$$

The first term is the minimum risk, namely $\mathcal{R}^\star_{\mathrm{sh},t}$ or $\mathcal{R}^\star_{\mathrm{cond},t}$. This covers every finite-risk predictor, since $U_t \in L^2$ implies such predictors lie in L^2. The tower property gives

$$\mathbb{E}[v_t^\star(X_t, C) \mid X_t] = \bar{v}_t^\star(X_t). \tag{A6}$$

Apply (A5) with $\mathcal{G} = \sigma(X_t, C)$ and $f = \bar{v}_t^\star(X_t)$ to obtain the risk difference below. By (A6), the squared deviation is from the conditional mean given X_t, which gives the covariance expression:

$$\mathcal{R}^\star_{\mathrm{sh},t} - \mathcal{R}^\star_{\mathrm{cond},t} = \Gamma_t := \mathbb{E}\|v_t^\star(X_t, C) - \bar{v}_t^\star(X_t)\|_2^2 = \mathbb{E}\big[\operatorname{tr}\operatorname{Cov}\big(v_t^\star(X_t, C) \mid X_t\big)\big]. \tag{A7}$$

In particular, $\Gamma_t = 0$ exactly when $v_t^\star(X_t, C) = \bar{v}_t^\star(X_t)$ almost surely. □

For a fixed learned shared field $v_\theta(X_t, t) \in L^2$, the same projection identity gives

$$\mathbb{E}\|U_t - v_\theta(X_t, t)\|_2^2 = \mathcal{R}^\star_{\mathrm{cond},t} + \Gamma_t + \mathbb{E}\|\bar{v}_t^\star(X_t) - v_\theta(X_t, t)\|_2^2. \tag{A8}$$

The terms are the conditional Bayes risk, residual conditioning value, and shared-field fitting error.

A.3 Proof of Theorem 1

Proof. Let $\widehat{C}_t = \mathcal{D}_t(X_t)$ satisfy $\mathbb{E}\Delta_t^2(C, \widehat{C}_t) < \infty$. Under (A4), the X_t-measurable predictor $v_t^\star(X_t, \widehat{C}_t)$ is square-integrable: the squared triangle inequality bounds its second moment by $2\mathbb{E}\|v_t^\star(X_t, C)\|_2^2 + 2\mathbb{E}\Delta_t^2(C, \widehat{C}_t) < \infty$. Orthogonality using (A6) therefore gives

$$\mathbb{E}\|v_t^\star(X_t, C) - v_t^\star(X_t, \widehat{C}_t)\|_2^2 = \Gamma_t + \mathbb{E}\|\bar{v}_t^\star(X_t) - v_t^\star(X_t, \widehat{C}_t)\|_2^2. \tag{A9}$$

Dropping the nonnegative final term proves the first inequality in (7). For the second, (A4) applies to the random pair $(C, \widehat{C}_t)$ on the common full-measure state set; squaring and taking expectations gives the claimed distortion bound. □

Equality in the first bound holds exactly when $v_t^\star(X_t, \widehat{C}_t) = \bar{v}_t^\star(X_t)$. Zero distortion almost surely implies $\Gamma_t = 0$; exact recovery $\widehat{C}_t = C$ is a sufficient case because $\Delta_t(c, c) = 0$.

Sufficient conditions for simultaneous comparison. If $\mathcal{C}$ is countable, intersect the full-measure sets for all context pairs to obtain (A4). Alternatively, if $x \mapsto v_t^\star(x, c)$ is continuous on $\operatorname{supp}\mathcal{L}(X_t)$ for every c, a violation at a support point would persist on a neighborhood of positive probability, contradicting (A3). The support is then a common full-measure set.

A.4 Proof of Theorem 2

Draw $Z_0 \mid C = c \sim Q_{\phi,c}$ and consider the coupled population ODEs

$$\dot{Z}_t^{\mathrm{cond}} = v_t^\star(Z_t^{\mathrm{cond}}, C), \qquad \dot{Z}_t^{\mathrm{sh}} = \bar{v}_t^\star(Z_t^{\mathrm{sh}}), \qquad Z_0^{\mathrm{cond}} = Z_0^{\mathrm{sh}} = Z_0. \tag{A10}$$

Assume unique, absolutely continuous solutions on $[0, 1]$, jointly measurable in time, context, and initialization, with finite second moments under (C, Z_0). Require conditional path consistency: for almost every c,

$$\mathcal{L}(Z_t^{\mathrm{cond}} \mid C = c) = \mathcal{L}(X_t \mid C = c) \qquad \text{for every } t \in [0, 1]. \tag{A11}$$

For almost every t, assume

$$\|\bar{v}_t^\star(x) - \bar{v}_t^\star(x')\|_2 \le L_x(t)\|x - x'\|_2 \quad \text{for all } x, x' \in \mathbb{R}^{d_y},\ L_x \ge 0,\ \int_0^1 L_x(t)\,dt < \infty. \tag{A12}$$

Let $P_c^{\mathrm{cond}} = \mathcal{L}(Z_1^{\mathrm{cond}} \mid C = c)$ and $P_c^{\mathrm{sh}} = \mathcal{L}(Z_1^{\mathrm{sh}} \mid C = c)$.

Proof. Write $e_t = Z_t^{\mathrm{sh}} - Z_t^{\mathrm{cond}}$ and $\varepsilon_t = \bar{v}_t^\star(Z_t^{\mathrm{cond}}) - v_t^\star(Z_t^{\mathrm{cond}}, C)$. Path consistency identifies the joint laws of (Z_t^{cond}, C) and (X_t, C), so Eq. (5) gives

$$\mathbb{E}\|\varepsilon_t\|_2^2 = \Gamma_t \le \mathcal{R}^\star_{\mathrm{sh},t} \le \mathbb{E}\|U_t\|_2^2 < \infty. \tag{A13}$$

Tonelli and Cauchy–Schwarz imply $\int_0^1 \|\varepsilon_t\|_2 \, dt < \infty$ almost surely and $\int_0^1 (\mathbb{E}\|\varepsilon_t\|_2^2)^{1/2} \, dt < \infty$. On this probability-one event, the common initialization gives $e_0 = 0$. Subtracting the integral forms of (A10) and using (A12) gives

$$\|e_t\|_2 \le \int_0^t L_x(s)\|e_s\|_2 \, ds + \int_0^t \|\varepsilon_s\|_2 \, ds.$$

Grönwall's inequality therefore yields

$$\|e_t\|_2 \le \int_0^t \exp\left(\int_s^t L_x(r) \, dr\right) \|\varepsilon_s\|_2 \, ds \qquad \text{for every } t \in [0,1]. \tag{A14}$$

For almost every c, the joint endpoints form a coupling of P_c^{cond} and P_c^{sh}, so $W_2^2(P_c^{\mathrm{cond}}, P_c^{\mathrm{sh}}) \le \mathbb{E}[\|e_1\|_2^2 \mid C = c]$. Averaging over C gives the first inequality below; Minkowski's integral inequality and (A13) give the second:

$$\left(\mathbb{E}_C W_2^2(P_C^{\mathrm{cond}}, P_C^{\mathrm{sh}})\right)^{1/2} \le \left(\mathbb{E}\|e_1\|_2^2\right)^{1/2} \le \int_0^1 \exp\left(\int_t^1 L_x(r) \, dr\right) \sqrt{\Gamma_t} \, dt. \tag{A15}$$

The exponential weight is bounded by (A12), so all are finite. Applying Theorem 1 for almost every t gives the second bound in (8), provided the decoder and distortion families are jointly measurable and its right-hand side is finite. Finally, (A11) at $t = 1$ gives $P_c^{\mathrm{cond}} = P_c^\star$ for almost every c. □

Time-integrated residual value. Bounding the exponential weight in (A15) and applying Cauchy–Schwarz yields

$$\left(\mathbb{E}_C W_2^2(P_C^\star, P_C^{\mathrm{sh}})\right)^{1/2} \le \exp\left(\int_0^1 L_x(r) \, dr\right) \left(\int_0^1 \Gamma_t \, dt\right)^{1/2}. \tag{A16}$$

If $\Gamma_t = 0$ for almost every t, Tonelli and (A13) give $\varepsilon_t = 0$ for almost every time on a common probability-one event. By (A14), the trajectories coincide for every t on that event, and $P_c^{\mathrm{sh}} = P_c^\star$ for almost every c.

A.5 Additional Theoretical Results

Decoder bounds and representations. Write $c = (c_1, \ldots, c_{d_f})$. For measurable, nonnegative, finite-valued field distortions δ_j and deterministic finite weights $\lambda_{t,j} \ge 0$, suppose

$$\Delta_t^2(c, c') \le \sum_{j=1}^{d_f} \lambda_{t,j} \delta_j^2(c_j, c'_j) \qquad \text{for all } c, c' \in \mathcal{C}. \tag{A17}$$

Under the simultaneous comparison condition (A4), Theorem 1 then gives

$$\Gamma_t \le \mathbb{E}\left[\sum_{j=1}^{d_f} \lambda_{t,j} \delta_j^2(C_j, \widehat{C}_{t,j})\right]$$

for every measurable decoder of the conditioning variables with finite expected weighted distortion; $\widehat{C}_{t,j}$ denotes its jth logical field. This follows by substituting $(C, \widehat{C}_t)$ into (A17). The δ_j may encode scale-normalized continuous differences, ordinal rank differences, or categorical mismatches. The weights bound population velocity sensitivity, with zero-weight fields contributing no cost. Such an envelope follows from a uniform squared-velocity bound on the common full-measure state set by taking the essential supremum in (A3).

For countable $\mathcal{C}$ with bounded Δ_t-diameter, every measurable decoder satisfies

$$\Gamma_t \le \left(\sup_{c,c' \in \mathcal{C}} \Delta_t^2(c, c')\right) \Pr(\widehat{C}_t \ne C), \tag{A18}$$

because distortion vanishes on correct decoding and is otherwise bounded by the squared diameter.

Recovering all conditioning variables is unnecessary when measurable maps r_t and $q_t^\star$ factor the velocity as $R_t = r_t(C)$ and $v_t^\star(X_t, C) = q_t^\star(X_t, R_t)$. Assume a measurable, nonnegative distortion

$\Delta_{R,t}$ bounds $\|q_t^\star(x,r) - q_t^\star(x,r')\|_2^2$ by $\Delta_{R,t}^2(r,r')$ simultaneously for all representation pairs on a common full-measure state set. Every X_t-measurable decoder $\widehat{R}_t$ with $\mathbb{E}\Delta_{R,t}^2(R_t, \widehat{R}_t) < \infty$ then satisfies

$$\Gamma_t \le \mathbb{E}\|q_t^\star(X_t, R_t) - q_t^\star(X_t, \widehat{R}_t)\|_2^2 \le \mathbb{E}\Delta_{R,t}^2(R_t, \widehat{R}_t). \tag{A19}$$

Indeed, finite distortion makes the decoded predictor square-integrable, and (A6) gives $\mathbb{E}[q_t^\star(X_t, R_t) \mid X_t] = \bar{v}_t^\star(X_t)$. Conditional-mean projection gives the first inequality, and the simultaneous bound gives the second. Exact recovery of R_t makes the conditional velocity state-measurable, so $\Gamma_t = 0$.

Posterior variation. On a probability-space extension, let C_t' be a conditionally independent copy of C given X_t. The conditional variance identity and Eq. (5) give

$$\Gamma_t = \frac{1}{2}\mathbb{E}\|v_t^\star(X_t, C) - v_t^\star(X_t, C_t')\|_2^2, \tag{A20}$$

since the two velocity vectors have the same conditional mean and their centered cross term is zero. Under (A4), this also yields

$$\Gamma_t \le \frac{1}{2}\mathbb{E}\Delta_t^2(C, C_t'). \tag{A21}$$

Thus Γ_t measures velocity variation across conditioning values in the state posterior.

If $C \in \mathbb{R}^{d_c}$ has finite second moment and the Bayes field is affine in the conditioning variables with deterministic linear part $B_t \in \mathbb{R}^{d_y \times d_c}$, so that $v_t^\star(x,c) - v_t^\star(x,c') = B_t(c - c')$ on the comparison domain, then

$$\Gamma_t = \mathbb{E}\,\mathrm{tr}\big[B_t\,\mathrm{Cov}(C \mid X_t)B_t^\top\big]. \tag{A22}$$

To see this, center the affine field using (A6) to obtain $v_t^\star(X_t, C) - \bar{v}_t^\star(X_t) = B_t(C - \mathbb{E}[C \mid X_t])$, then take its expected squared norm. For fixed B_t, decreasing the expected posterior covariance in the positive semidefinite order cannot increase Γ_t.

Information under shared transport. Fix t and let $Z_t = \Phi_t(Z_0)$, where Φ_t is a deterministic shared measurable bijection from its initial-state domain to its image, with measurable inverse. Each state is then a measurable function of the other, so

$$\sigma(Z_t) = \sigma(Z_0). \tag{A23}$$

Composing decoders with Φ_t or Φ_t^{-1} preserves their predictions. Hence, for every square-integrable context function $h(C)$,

$$\inf_f \mathbb{E}\|h(C) - f(Z_t)\|_2^2 = \inf_f \mathbb{E}\|h(C) - f(Z_0)\|_2^2, \tag{A24}$$

where both infima range over measurable decoders with finite risk. Data processing in both directions likewise gives

$$I(C; Z_t) = I(C; Z_0). \tag{A25}$$

Independence requires no invertibility: for any measurable deterministic shared map Φ_t, $Z_0 \perp C$ implies, for almost every c,

$$\mathcal{L}(Z_t \mid C = c) = (\Phi_t)_\# \mathcal{L}(Z_0) = \mathcal{L}(Z_t), \tag{A26}$$

and therefore $Z_t \perp C$. A shared-flow output therefore inherits its dependence on the conditioning variables from its conditional source laws.

Population source objective. Integrating Eq. (5) over flow time gives

$$\int_0^1 \mathcal{R}_{\mathrm{sh},t}^\star\,dt = \int_0^1 \mathcal{R}_{\mathrm{cond},t}^\star\,dt + \int_0^1 \Gamma_t\,dt. \tag{A27}$$

All terms are finite by (A1). Source optimization changes the source-induced training law, so the ideal shared objective minimizes the sum of both right-hand terms under that law. For a learned shared velocity with finite integrated squared risk, integrating (A8) additionally contributes the fitting term $\int_0^1 \mathbb{E}\|\bar{v}_t^\star(X_t) - v_\theta(X_t, t)\|_2^2\,dt$.

B CONTROLLED STUDY OF CONDITION SATURATION

This section isolates the mechanism behind condition saturation in a setting where the state posterior, the conditional and shared Bayes velocities, and the residual conditioning value are available in closed form. We use the same eight-condition Gaussian ring introduced in Section 3, but expose the full source-strength sweep and its endpoint consequences. The construction changes only how strongly the source encodes the conditioning variable while keeping the target family and covariance structure fixed. This makes the two conditioning routes directly comparable: conditioning information can reach transport through the source-induced state, or be supplied again to the velocity.

We then move beyond the analytic Bayes fields in two directions. First, matched finite neural velocity models reproduce the same transition, showing that the residual-value prediction remains visible under function approximation. Second, a crossing construction keeps the source informative about the condition index but makes the prescribed probability path temporarily erase that information. Together, these experiments show that the relevant quantity for condition allocation is not source separation by itself, but whether the evolving state carries the conditioning information needed for velocity prediction along the transport path.

B.1 ANALYTIC GAUSSIAN-RING CONSTRUCTION

Let $K = 8$, $C \sim \mathrm{Unif}\{0, \dots, K-1\}$, and

$$u_k = \left(\cos \frac{2\pi k}{K}, \sin \frac{2\pi k}{K}\right).$$

For condition index $C = k$, the source and target are

$$X_0 \mid C = k \sim \mathcal{N}(\alpha r_0 u_k, \sigma_0^2 I), \qquad Y \mid C = k \sim \mathcal{N}(r_1 u_k, \sigma_1^2 I), \tag{B1}$$

with $r_0 = 2.5$, $r_1 = 4$, $\sigma_0 = 0.45$, and $\sigma_1 = 0.35$. The source strength

$$\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$$

controls how much conditioning information is present at initialization. At $\alpha = 0$, all source distributions coincide; increasing α separates them along the condition-specific radial directions. The target distributions are unchanged throughout the sweep.

We sample X_0 and Y independently given C. Under the linear probability path

$$X_t = (1-t)X_0 + tY,$$

the conditional state distribution remains Gaussian,

$$X_t \mid C = k \sim \mathcal{N}(\rho_t u_k, s_t^2 I), \qquad \rho_t = (1-t)\alpha r_0 + t r_1, \tag{B2}$$

$$s_t^2 = (1-t)^2 \sigma_0^2 + t^2 \sigma_1^2. \tag{B3}$$

The posterior over the condition index is therefore

$$\pi_{t,k}(x) := \Pr(C = k \mid X_t = x) = \frac{\exp\left[-\|x - \rho_t u_k\|_2^2 / (2s_t^2)\right]}{\sum_{\ell=0}^{K-1} \exp[-\|x - \rho_t u_\ell\|_2^2 / (2s_t^2)]}. \tag{B4}$$

Gaussian conditioning gives the conditional Bayes velocity in the affine form

$$v_t^\star(x, k) = a_t x + b_t u_k, \tag{B5}$$

where

$$a_t = \frac{t\sigma_1^2 - (1-t)\sigma_0^2}{s_t^2}, \qquad b_t = (r_1 - \alpha r_0) - a_t \rho_t. \tag{B6}$$

Marginalizing over the state posterior yields the shared Bayes velocity,

$$\bar{v}_t^\star(x) = a_t x + b_t \sum_{k=0}^{K-1} \pi_{t,k}(x) u_k. \tag{B7}$$

This construction makes the condition-saturation mechanism explicit. Define the posterior recovery error of the direction associated with the condition index as

$$E_{\mathrm{rec},t} = \mathbb{E}\left\| u_C - \sum_{k=0}^{K-1} \pi_{t,k}(X_t) u_k \right\|_2^2. \tag{B8}$$

Substituting Equations (B5)–(B7) into the residual-value identity gives

$$\Gamma_t(\alpha) = b_t^2 E_{\mathrm{rec},t}. \tag{B9}$$

Equation (B9) is the central relation in this controlled study. The posterior term measures how much transport-relevant conditioning information remains unresolved by the state, while b_t^2 measures how strongly that unresolved distinction changes the required velocity. Direct conditioning is valuable only when both are present.

B.2 Source Encoding and Condition Saturation

Figure S1 follows the complete source-strength sweep. When $\alpha = 0$, the source does not distinguish conditions, so the early state is highly ambiguous and $E_{\mathrm{rec},t}$ is close to its maximum. As α increases, the state reveals the condition index earlier in the path, and the recovery error contracts accordingly. The residual conditioning value follows the same transition through Equation (B9): the large gap between shared and conditioned velocity prediction at weak source encoding progressively disappears as the source-induced state becomes more informative.

The endpoint behavior mirrors this pathwise change. The conditional generator remains close to the target distribution across the entire sweep, whereas the shared generator improves sharply with source strength. Its conditional energy distance drops from approximately 4.4997 at $\alpha = 0$ to 0.0017 at $\alpha = 1$. The same transition is visible in condition assignment: the shared generator moves from 12.3725%, essentially chance for eight conditions, to 71.7900% at $\alpha = 0.5$ and 96.8525% at $\alpha = 1$.

This result isolates the allocation effect. Neither the target family nor the response covariance changes across the sweep; only the strength of the source-side conditioning route changes. The corresponding loss of value from conditioning the velocity directly therefore reflects substitution between the two routes. Once the state already identifies the transport-relevant condition, direct velocity conditioning has little remaining role. This is the empirical form of the condition-saturation principle that motivates source-concentrated conditioning in CoFM.

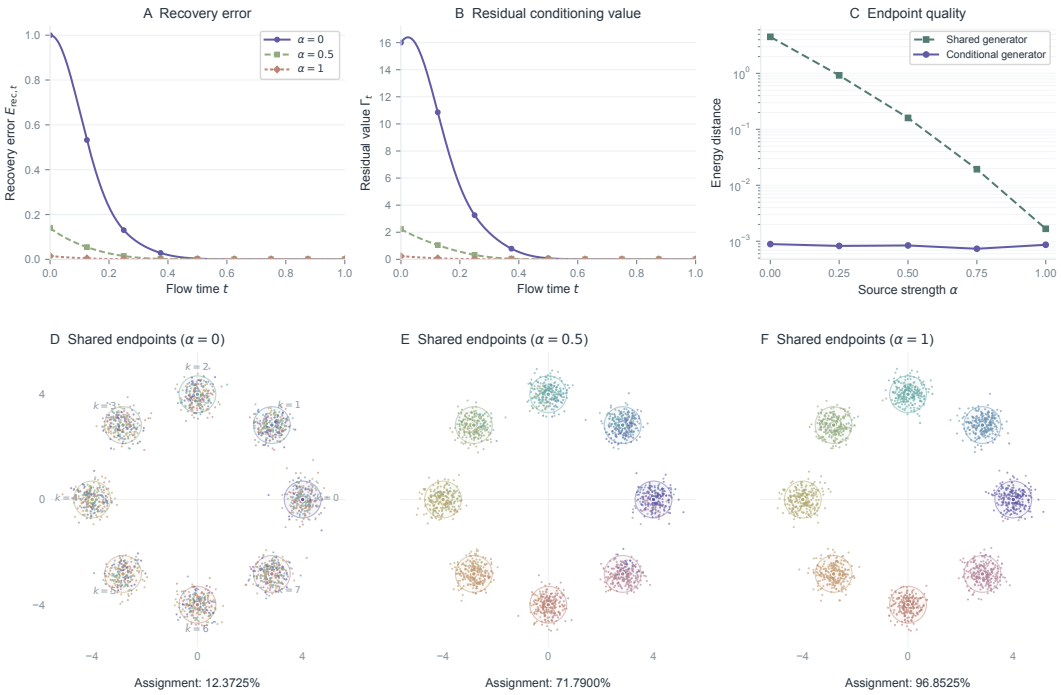

Figure S1: **Experimental results on the Gaussian-ring construction.** Panels A and B show the recovery error $E_{\mathrm{rec},t}$ and residual conditioning value Γ_t along the probability path as source strength increases. Panel C reports conditional energy distance to the target distribution for shared and conditioned Bayes generators. Panels D–F show shared-generator endpoints for $\alpha = 0$, 0.5, and 1; colors denote the originating condition, and assignment is evaluated against the corresponding target component. Increasing source encoding improves condition recovery from the state, reduces the residual value of direct velocity conditioning, and restores the condition-to-target correspondence.

Table S2: **Experimental results across source strengths in the Gaussian-ring construction.** The target family and covariance structure are fixed across source strengths. Lower recovery error indicates that the evolving state more clearly identifies the transport-relevant condition.

α	$\int_0^1 E_{\mathrm{rec},t}\, dt$	$\int_0^1 \Gamma_t\, dt$	Shared energy distance	Conditional energy distance
0.00	0.1494	2.9200	4.4997	0.0009
0.25	0.0517	0.9732	0.9227	0.0008
0.50	0.0164	0.3026	0.1601	0.0008
0.75	0.0053	0.0942	0.0194	0.0007
1.00	0.0014	0.0242	0.0017	0.0009

Table S2 provides the complete sweep underlying Figure S1. The reduction in residual conditioning value is not confined to the three visualized endpoint settings: it follows the same progression across all intermediate source strengths, together with the contraction of the shared generator's endpoint discrepancy. The conditional reference remains essentially unchanged, isolating the transition to the information carried by the source-induced state.

B.3 FINITE NEURAL REPLICATION

The previous analysis uses population Bayes velocities. We next fit matched-capacity shared and conditioned MLP velocities to the same family of flow-matching problems and evaluate their held-out flow-matching risks and generated endpoints.

Figure S2 compares held-out risks and conditional endpoint energy distances across source strengths. Its risk-gap panel tests whether the analytic residual value anticipates the benefit of directly conditioning the learned velocity.

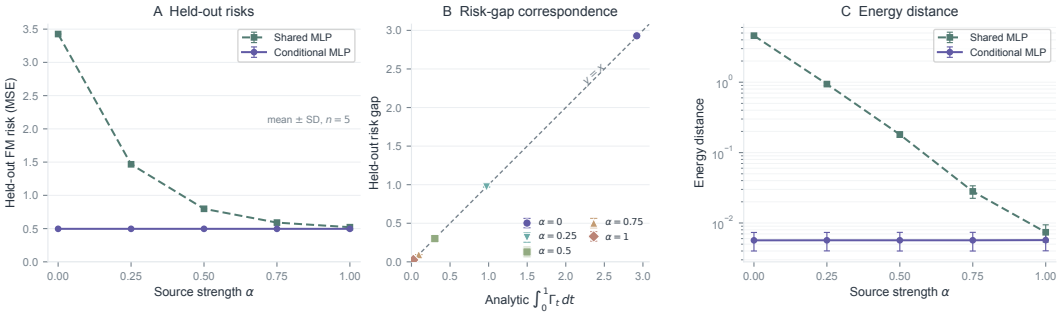

Figure S2: **Experimental results with finite neural velocity models.** Panel A compares held-out flow-matching risks of matched shared and conditioned MLPs across source strengths. Panel B plots the observed held-out risk gap against the analytic $\int_0^1 \Gamma_t \, dt$, with the identity line indicating exact correspondence. Panel C shows the resulting conditional energy distance. As source encoding strengthens, the finite-model benefit of direct conditioning contracts in the same way predicted by the analytic residual value.

Table S3: **Experimental results with finite neural velocity models on the Gaussian ring.** Both gaps are shared-minus-conditional. Held-out flow-matching squared risk is summed over coordinates and averaged over examples; the generation gap uses conditional energy distance to the target law. Neural results are mean ± sample standard deviation across five runs.

α	Analytic $\int_0^1 \Gamma_t \, dt$	Neural held-out risk gap	Neural generation gap
0.00	2.9200	2.9299 ± 0.0011	4.5821 ± 0.0949
0.25	0.9732	0.9706 ± 0.0003	0.9350 ± 0.0400
0.50	0.3026	0.3013 ± 0.0001	0.1749 ± 0.0156
0.75	0.0942	0.0939 ± 0.0001	0.0224 ± 0.0048
1.00	0.0242	0.0248 ± 0.0001	0.0017 ± 0.0006

Table S3 shows that the agreement is quantitative rather than merely directional. Across the full source-strength sweep, the held-out advantage of the conditioned network follows the Bayes residual value predicted before fitting either network. Thus the population allocation criterion remains visible after replacing the analytic velocity fields with learned function approximators. The accompanying generation gaps connect this regression-level substitution to conditional endpoint quality.

B.4 Crossing Paths and Pathwise State Sufficiency

Source informativeness at $t = 0$ is not sufficient by itself. To separate initialization from pathwise state sufficiency, we keep the strongly separated source

$$X_0 \mid C = k \sim \mathcal{N}(r_0 u_k, \sigma_0^2 I)$$

but reverse the condition-to-target assignment,

$$Y \mid C = k \sim \mathcal{N}(-r_1 u_k, \sigma_1^2 I). \tag{B10}$$

The conditional path mean becomes

$$\big[(1-t)r_0 - t r_1\big] u_k. \tag{B11}$$

All eight conditional means therefore meet at

$$t^\star = \frac{r_0}{r_0 + r_1} = 0.384615. \tag{B12}$$

At this time, the conditional path laws coincide. The state posterior over conditions becomes uniform even though the source itself clearly distinguished them. Consequently,

$$E_{\mathrm{rec},t^\star} = 1, \qquad \Gamma_{t^\star} = 42.2500. \tag{B13}$$

Figure S3 contrasts this with the aligned construction. In the aligned mapping, the initially separated conditions remain distinguishable along the path and the residual value stays small. In the crossing mapping, recoverability collapses precisely around $t^\star$, and the residual conditioning value rises sharply at the same location. The shared endpoint samples still occupy the ring-shaped target geometry, but the condition-to-target correspondence is lost, yielding 0.0000% assignment accuracy.

This case sharpens the condition-allocation insight. What matters for a shared velocity is not how much conditioning information the source alone encodes, but whether the evolving state preserves the distinctions that determine the required transport. A source can begin with nearly perfect separation between conditions and still fail to support shared transport if the prescribed path later aliases them. Condition saturation is therefore a property of the source-induced **state along the probability path**. This pathwise view is the criterion behind CoFM: source-side conditioning is effective when it makes the state transport-sufficient throughout the region where the velocity is learned.

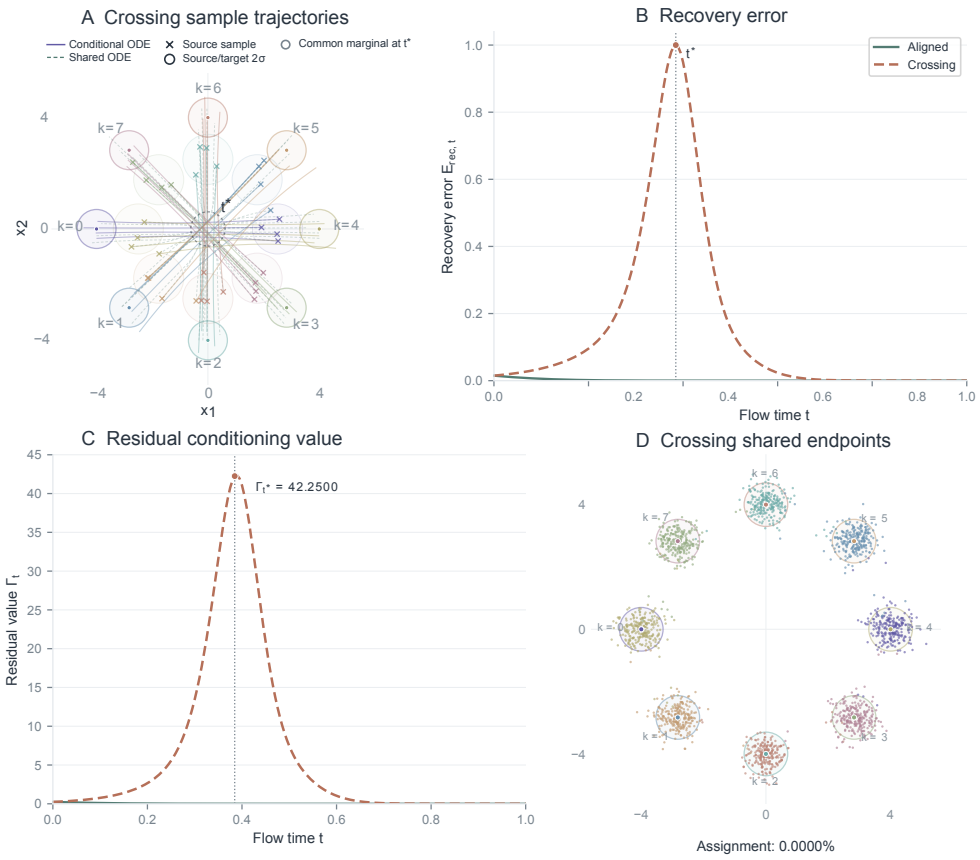

Figure S3: **Experimental results with crossing probability paths.** Panel A shows the reversed condition-to-target construction, in which initially separated source states pass through a common intermediate marginal. Panels B and C compare recovery error and residual conditioning value for aligned and crossing paths. The loss of condition recoverability at $t^\star$ coincides with a sharp rise in Γ_t. Panel D shows the shared-generator endpoints: the global ring geometry is retained, while the conditional assignment is lost.

Taken together, the three studies establish a single mechanism across analytic and learned settings. Strengthening conditioning information in the source reduces the additional predictive value of direct velocity conditioning when that information remains available in the state. The finite-network experiment reproduces this substitution quantitatively, while the crossing construction shows that the relevant notion of sufficiency is pathwise. These results connect the theoretical residual-value criterion directly to the architectural choice made by Condition-Once Flow Matching.

C CONDITIONAL SOURCE FORCING: THEORY AND CONTROLLED STUDY

Section 4 introduces Conditional Source Forcing (CS-Forcing) as a finite budget on source expansion under valid changes to observed tabular fields. This appendix makes that geometry explicit, gives two complementary theoretical interpretations, and uses a controlled source-radius study to separate condition recoverability from excessive source expansion.

C.1 TYPE-RESPECTING TABULAR NEIGHBORHOODS

CS-Forcing operates on the logical conditioning fields rather than directly on the encoded network input. Write a covariate configuration as

$$c = (c_1, \dots, c_{d_f}),$$

where one logical field may correspond to one or several encoded coordinates. Let

$$\mathcal{M}(c) \subseteq \{1, \dots, d_f\}$$

denote the fields that admit a valid nonzero move at c. We sample

$$J \sim p(J \mid c), \qquad J \in \mathcal{M}(c),$$

and construct a neighboring configuration by

$$\widetilde{c} = \mathcal{T}_J(c; \xi),$$

where ξ contains only the randomness required by the transition for field J. Unless a task-specific proposal is specified, $p(J \mid c)$ is uniform over $\mathcal{M}(c)$. The transition changes one logical field at a time and leaves all remaining fields fixed.

The corresponding table-aware distance is

$$d^2_{\mathrm{tab}}(c, c') = \sum_{j=1}^{d_f} d^2_j(c_j, c'_j), \tag{C1}$$

with field-specific components

$$d^2_j(c_j, c'_j) = \begin{cases} \left(\dfrac{c_j - c'_j}{a_j}\right)^2, & \text{continuous,} \\ \left(\dfrac{\mathrm{rank}_j(c_j) - \mathrm{rank}_j(c'_j)}{K_j - 1}\right)^2, & \text{ordinal,} \\ \mathbf{1}[c_j \neq c'_j], & \text{binary or categorical.} \end{cases}$$

Here $a_j > 0$ is a scale estimated from the training data and K_j is the number of valid levels of an ordinal field. The ordinal formula applies when $K_j > 1$; a constant field contributes zero to the distance and is excluded from $\mathcal{M}(c)$. The fieldwise normalization makes changes of heterogeneous types comparable without treating the encoded covariates as a homogeneous Euclidean vector.

Continuous fields. For a selected continuous field j, we apply a bounded nonzero displacement on its training-derived scale,

$$\widetilde{c}_j = c_j + a_j h,$$

for a sampled signed step h. When the proposed value crosses a declared domain boundary, it is rejected or truncated, and the realized displacement is used in d_j.

Ordinal fields. Ordinal variables move only to an adjacent valid rank. Interior levels therefore admit two local directions, whereas a boundary level admits only the inward move. The normalized rank distance above preserves the ordering while accounting for the number of levels.

Binary and categorical fields. A binary field is flipped to its other valid state. For an unordered categorical field, let

$$\mathcal{A}_j(c_j) \subseteq \mathcal{C}_j \setminus \{c_j\}$$

denote the admissible replacement set. We draw the new category from a fixed proposal $q_j(\cdot \mid c_j)$ supported on $\mathcal{A}_j(c_j)$; in the absence of a task-specific proposal, this distribution is uniform over the admissible alternatives. We do not impose an artificial ordering among unordered categories, and every valid replacement therefore has $d_j^2 = 1$. Locality in this case refers to a valid change of one logical field, rather than to a metric ordering among category labels.

Grouped encodings are handled at the logical-field level. A one-hot group is changed as one categorical variable rather than by perturbing individual bits. Likewise, a cyclic field represented by sine and cosine coordinates is changed through a small angular move and then re-encoded on the unit circle. Missingness indicators remain attached to their logical fields, so a missing value is not perturbed as though its imputed numerical placeholder were an observed continuous value. Application-specific coupled constraints can be incorporated by restricting the admissible transition set.

Every evaluated CS-Forcing pair satisfies

$$d_{\mathrm{tab}}(c, \widetilde{c}) > 0.$$

If no conditioning field is mutable for a row, it contributes zero to the CS-Forcing pair expectation.

C.2 Local Source Expansion and CS-Forcing

For the diagonal-Gaussian conditional source

$$Q_{\phi,c} = \mathcal{N}\left(\mu_\phi(c), \mathrm{Diag}\left[\sigma_\phi^2(c)\right]\right),$$

we measure source displacement using the normalized Gaussian Wasserstein distance

$$d_{\mathrm{src}}^2\left(Q_{\phi,c}, Q_{\phi,c'}\right) := \frac{1}{d_y} W_2^2\left(Q_{\phi,c}, Q_{\phi,c'}\right).$$

For diagonal Gaussians this has the closed form

$$d_{\mathrm{src}}^2\left(Q_{\phi,c}, Q_{\phi,c'}\right) = \frac{\|\mu_\phi(c) - \mu_\phi(c')\|_2^2 + \|\sigma_\phi(c) - \sigma_\phi(c')\|_2^2}{d_y}. \tag{C2}$$

The local source expansion is then

$$\kappa_\phi(c, c') := \frac{d_{\mathrm{src}}^2\left(Q_{\phi,c}, Q_{\phi,c'}\right)}{d_{\mathrm{tab}}^2(c, c')}. \tag{C3}$$

Thus κ_ϕ measures source displacement per unit type-aware covariate change. CS-Forcing applies a hinge only when this expansion exceeds a finite budget:

$$\mathcal{L}_{\mathrm{CSF}} = \mathbb{E}_{C,J,\xi}\left[\kappa_\phi\left(C, \mathcal{T}_J(C;\xi)\right) - \tau\right]_+, \qquad [a]_+ := \max(a, 0). \tag{C4}$$

The budget τ is therefore an upper target for sampled local expansion, not an invariance target. In particular, CS-Forcing does not drive $Q_{\phi,c} \approx Q_{\phi,c'}$ for neighboring covariate configurations. Source distributions may remain distinct whenever that distinction is useful for transport; the objective acts

only on separation that is large relative to the corresponding type-aware field change. This is the sense in which CS-Forcing controls **bounded local expansion rather than invariance**.

The designed geometry d_{tab} is distinct from the transport-relevant population distortion Δ_t used in the condition-saturation analysis. The former defines which source changes are locally disproportionate during training, whereas the latter measures how differences in conditioning variables change the population Bayes velocity.

C.3 Theoretical Interpretation

C.3.1 Continuous-field local limit

The finite-difference definition of κ_ϕ has a direct Jacobian interpretation along continuous fields.

Proposition C.1 (Continuous-field local expansion)**.** *Consider a continuous logical field j with scale a_j, and let*

$$c_h = c + a_j h e_j,$$

where e_j is the corresponding logical-field direction. Suppose the source mean and scale maps are differentiable at c. Then, as $h \to 0$,

$$\kappa_\phi(c, c_h) \longrightarrow \frac{\left\|J_{\mu_\phi}(c) a_j e_j\right\|_2^2 + \left\|J_{\sigma_\phi}(c) a_j e_j\right\|_2^2}{d_y}. \tag{C5}$$

Proof. By construction,

$$d_{\mathrm{tab}}^2(c, c_h) = h^2.$$

A first-order expansion of the source parameters gives

$$\mu_\phi(c_h) - \mu_\phi(c) = h\, J_{\mu_\phi}(c) a_j e_j + o(h),$$

and

$$\sigma_\phi(c_h) - \sigma_\phi(c) = h\, J_{\sigma_\phi}(c) a_j e_j + o(h).$$

Substituting these expansions into the diagonal-Gaussian Wasserstein identity yields

$$d_{\mathrm{src}}^2\left(Q_{\phi,c}, Q_{\phi,c_h}\right) = \frac{h^2}{d_y}\left(\left\|J_{\mu_\phi}(c) a_j e_j\right\|_2^2 + \left\|J_{\sigma_\phi}(c) a_j e_j\right\|_2^2\right) + o(h^2). \tag{C6}$$

Dividing by $d_{\mathrm{tab}}^2(c, c_h) = h^2$ and taking $h \to 0$ gives the result. □

This limit connects CS-Forcing to directional control of the source map while retaining the finite-change definition used in training. For discrete fields, projected boundaries, and nondifferentiable points, the finite-difference quantity κ_ϕ remains the operative definition.

C.3.2 Endpoint sensitivity under shared transport

The source geometry also propagates to the joint response distribution because CoFM uses the same transport map for all covariate configurations. Let $\Phi_{0\to1}$ denote the shared endpoint map and define the learned conditional output law

$$\widehat{P}_{\phi,\theta,c} = (\Phi_{0\to1})_{\#} Q_{\phi,c}.$$

Proposition C.2 (Output sensitivity under shared transport)**.** *Suppose $\Phi_{0\to1}$ is L_Φ-Lipschitz. Then for any two conditioning values c, c',*

$$W_2\left(\widehat{P}_{\phi,\theta,c}, \widehat{P}_{\phi,\theta,c'}\right) \le L_\Phi\, W_2\left(Q_{\phi,c}, Q_{\phi,c'}\right).$$

Consequently, for a valid neighboring pair satisfying $\kappa_\phi(c, c') \le \tau$, we have

$$W_2\left(\widehat{P}_{\phi,\theta,c}, \widehat{P}_{\phi,\theta,c'}\right) \le L_\Phi \sqrt{d_y \tau}\, d_{\mathrm{tab}}(c, c'). \tag{C7}$$

Proof. Let $\pi^\star$ be an optimal coupling of $Q_{\phi,c}$ and $Q_{\phi,c'}$. Pushing both components of this coupling through the same endpoint map gives a valid coupling of $\widehat{P}_{\phi,\theta,c}$ and $\widehat{P}_{\phi,\theta,c'}$. By Lipschitz continuity,

$$\begin{aligned} W_2^2\left(\widehat{P}_{\phi,\theta,c}, \widehat{P}_{\phi,\theta,c'}\right) &\le \int \|\Phi_{0\to1}(z) - \Phi_{0\to1}(z')\|_2^2 \, d\pi^\star(z, z') \\ &\le L_\Phi^2 W_2^2\left(Q_{\phi,c}, Q_{\phi,c'}\right). \end{aligned}$$

Taking square roots gives the first inequality. For the second, $\kappa_\phi(c, c') \le \tau$ implies

$$W_2^2\left(Q_{\phi,c}, Q_{\phi,c'}\right) = d_y\, d_{\mathrm{src}}^2\left(Q_{\phi,c}, Q_{\phi,c'}\right) \le d_y \tau\, d_{\mathrm{tab}}^2(c, c'),$$

which completes the proof. □

When the shared ODE velocity is $L_v(t)$-Lipschitz in state, a sufficient endpoint constant is

$$L_\Phi \le \exp\left(\int_0^1 L_v(t)\, dt\right).$$

The two propositions describe complementary aspects of CS-Forcing. Proposition C.1 identifies κ_ϕ with directional source sensitivity in the local continuous limit, while Proposition C.2 shows how bounded source expansion limits the amplification of local covariate changes by a shared transport map. The training loss estimates this geometry on sampled valid neighbors rather than imposing a global constraint over the complete covariate space.

C.4 Controlled Study of Recovery and Source Expansion

We use a radius-controlled source family to isolate the distinction between making conditioning information recoverable and controlling the geometry used to encode it. The source mean places the condition codes at radius $R \in \{0.25, 0.5, 1, 2, 4, 8\}$, while the observation noise and target construction are held fixed. The construction is normalized to give local source expansion $\kappa = R^2$.

Figure S4 and Table S4 show that increasing R makes the condition progressively easier to recover from the source state: the posterior embedding error decreases from 0.8652 at $R = 0.25$ to 0.0032 at $R = 8$. The same change increases local source expansion from 0.0625 to 64.0000. Thus increasing source separation continues to improve recoverability even after the source has become highly informative, but it does so through increasingly large local expansion.

To visualize the role of a finite expansion budget, we set $\tau_{\mathrm{toy}} = 4$ for this controlled construction. This places $R = 2$ exactly on the budget boundary. At that point the recovery error is already 0.0520, whereas increasing the radius to $R = 4$ and $R = 8$ reduces the error further to 0.0127 and 0.0032 while increasing κ to 16.0000 and 64.0000, respectively. The associated hinge is zero through $R = 2$, then rises to 12.0000 and 60.0000. The symbol τ_{toy} is used only for this controlled study and is distinct from the training budget τ used in the real-data models.

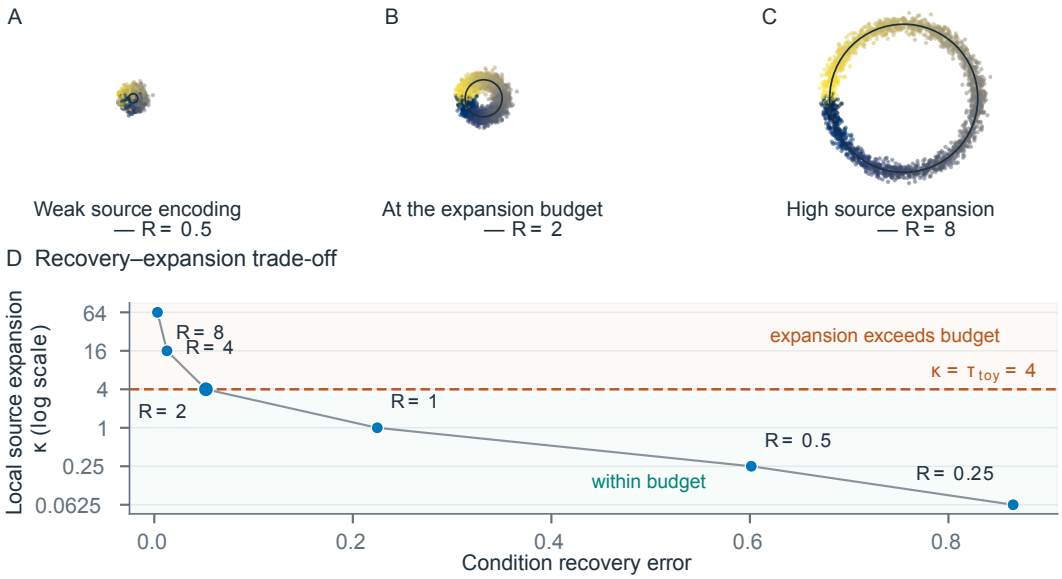

Figure S4: **Recovery and local source expansion in the radius-controlled study.** Panels A–C visualize the same source-code geometry at three radii: weak source encoding at $R = 0.5$, the expansion-budget boundary at $R = 2$, and high source expansion at $R = 8$. Panel D shows the relationship between posterior condition-recovery error and local source expansion $\kappa = R^2$ for the full radius sweep. The dashed line marks the illustrative budget $\tau_{\text{toy}} = 4$: $R = 2$ lies on the boundary, while $R = 4$ and $R = 8$ obtain progressively smaller recovery error at increasingly larger expansion.

Table S4: **Experimental results for radius-controlled source expansion.** Recovery error is posterior embedding MMSE, local expansion is $\kappa = R^2$, and the hinge uses the illustrative budget $\tau_{\text{toy}} = 4$.

R	Condition recovery error	Local source expansion κ	$[\kappa - \tau_{\text{toy}}]_+$
0.25	0.8652	0.0625	0.0000
0.50	0.6013	0.2500	0.0000
1.00	0.2246	1.0000	0.0000
2.00	0.0520	4.0000	0.0000
4.00	0.0127	16.0000	12.0000
8.00	0.0032	64.0000	60.0000

The study makes the role of CS-Forcing explicit. Stronger source encoding can improve condition recoverability, but recoverability alone does not determine whether the source geometry is locally controlled. A finite expansion budget preserves the ability of the source to distinguish conditions while separating further information gain from disproportionately large source displacement. This complements condition saturation: the saturation analysis concerns how much transport-relevant conditioning information remains unresolved in the evolving state, whereas CS-Forcing governs how that information is written into the conditional source.

D EXPERIMENTAL DETAILS

D.1 BENCHMARK CONSTRUCTION AND DATA SPLITS

Each benchmark defines a conditional multivariate generation task with observed covariates C and a continuous response vector Y. The final test partition is fixed across methods. Table S5 summarizes the data partitions and encoded dimensions used in the benchmark.

Table S5: **Benchmark partitions and encoded dimensions.**

Dataset	Outer training pool	Fit	Internal validation	Test	d_c	d_y
Beijing	24,864	22,378	2,486	5,832	11	72
Superconductivity	17,011	15,310	1,701	4,252	33	37
GTEx	4,348	3,914	434	1,113	110	37
UKB-PPP	36,328	32,696	3,632	9,082	49	150

Beijing. The Beijing benchmark is derived from the UCI Beijing Multi-Site Air Quality data (Zhang et al., 2017). Each example corresponds to a common hourly timestamp across 12 monitoring stations. The response concatenates six pollutant measurements across the stations, giving $d_Y = 72$. The covariates contain cross-station summaries of five meteorological variables together with circular encodings of wind direction, hour, and day of year, giving $d_C = 11$. The outer training period spans March 2013 through June 2016 and a fixed test period from July 2016 through February 2017. Within the outer training pool, 22,378 rows are used for fitting and 2,486 for internal validation. Missing pollutant values are filled with pollutant-by-station medians estimated from the training period. Meteorological variables and circular wind coordinates are aggregated across available stations, and remaining missing covariate values are filled with training-period medians. Covariate and response coordinates are standardized using training statistics.

Superconductivity. The Superconductivity benchmark derives from the UCI resource of Hamidieh (2018). The 33 covariate coordinates summarize element count and elemental-property locations, while the 37 response coordinates describe dispersion and shape statistics of those properties. Neither block includes critical temperature. The 21,263 complete rows are split into 17,011 outer-training and 4,252 test examples; within the outer training pool, 15,310 rows are used for fitting and 1,701 for internal validation. Continuous coordinates are standardized using outer-training statistics.

GTEx. The GTEx benchmark uses v8 tissue RNA-seq samples (GTEx Consortium, 2020). The covariates contain 100 landmark genes plus tissue, sex, and age bracket, encoded into $d_C = 110$ coordinates. The response is a fixed set of 37 genes distinct from the landmark set. Donors are separated between the outer training and test partitions so that samples from the same donor do not occur in both. The outer training pool has 4,348 samples, of which 3,914 are used for fitting and 434 for internal validation; 1,113 samples are held out for testing. Gene expression is transformed as $\log_2(\mathrm{TPM} + 1)$, and continuous coordinates are standardized using outer-training statistics.

UKB-PPP. The UKB-PPP benchmark combines UK Biobank clinical variables with Olink plasma protein measurements (Sun et al., 2023; Sudlow et al., 2015). After cohort construction, the benchmark contains 45,410 participants. The covariates comprise 48 numerical clinical variables and sex, giving $d_C = 49$, and the response is a fixed panel of 150 proteins. The 36,328-participant outer training pool is divided into 32,696 fitting and 3,632 validation participants, with 9,082 participants held out for testing. Clinical variables are transformed, and missing clinical and protein values are imputed using statistics estimated from the outer training pool, comprising the fitting and validation partitions. These statistics are then applied unchanged to the held-out test participants. Covariate and protein coordinates are standardized using outer-training statistics.

D.2 BASELINES AND TASK ADAPTATION

Ten baselines cover direct conditional density models, latent-variable and adversarial generators, normalizing flows, tabular diffusion, and reparameterized or learnable-source flow matching. We

describe each method's generative mechanism and its adaptation to jointly generating the response vector Y at fixed covariates C. The held-out response is used only for evaluation; all methods follow the sampling protocol in Appendix D.5.

Gaussian. The conditional Gaussian directly parameterizes $p(Y \mid C)$ and tests whether a source-sized distribution can model the responses without learned transport. The processed covariates pass through the residual-GELU encoder family used by CoFM, matched in parameter count to its source, followed by mean and log-scale heads defining a diagonal Gaussian. Training minimizes exact Gaussian negative log likelihood. For fixed C, generation evaluates these parameters once and draws $Y = \mu(C) + \sigma(C) \odot \epsilon$, with $\epsilon \sim \mathcal{N}(0, I)$. This control concentrates conditional modeling capacity in the output distribution, separating direct parameterization from learned transport.

MDN. The mixture density network (MDN; Bishop, 1994) extends the Gaussian control to a five-component conditional Gaussian mixture. A residual encoder from CoFM's source family predicts covariate-dependent mixture weights, means, and diagonal scales, with widths chosen to closely match the source's parameter budget. A single component index is shared across the full response vector, allowing cross-response dependence through mixture selection. Training uses exact joint mixture negative log likelihood; generation samples a component and then the response vector for fixed C. Together with Gaussian, MDN tests expressive direct conditional modeling without a learned transport stage.

CVAE. The conditional variational autoencoder (CVAE) models structured outputs through a conditional latent variable (Sohn et al., 2015). Its recognition network $q_\phi(z \mid C, Y)$ approximates the latent posterior, while a learned prior $p_\theta(z \mid C)$ uses the covariates and an initial response prediction. The generator maps $[C, z]$ to a response mean with learned global per-response scales. Training combines Gaussian reconstruction negative log likelihood with the analytic posterior–prior KL divergence. We use a 200-dimensional latent variable and, for fixed C, sample from the learned prior and decoder to generate the complete response vector jointly.

CTGAN. CTGAN is a tabular generative adversarial network built on mode-specific normalization, conditional adversarial training, and PacGAN (Xu et al., 2019). We replace its original single-discrete-field conditioning vector with the full observed covariates, following the conditional adversarial design (Mirza & Osindero, 2014). Separate data transformers encode the covariate and response blocks. The residual generator receives Gaussian noise together with the transformed covariates and outputs transformed responses; the discriminator scores real and generated responses paired with those same covariates. Training uses the WGAN-GP objective. At inference, new noise produces joint response samples for fixed covariates, followed by the response inverse transform.

RQ-NSF. Rational-quadratic neural spline flows (RQ-NSF) learn flexible densities through invertible monotonic spline transformations (Durkan et al., 2019). We use a conditional coupling-flow formulation: the flow transforms only the response vector Y, while the observed covariates C enter every coupling network. The model directly parameterizes $p(Y \mid C)$ and is trained by exact change-of-variables likelihood. Joint responses are generated by transforming a standard Gaussian sample through the inverse flow with C fixed. We use ten coupling steps with eight-bin splines.

TreeFlow. TreeFlow (Wielopolski & Zieba, 2023) combines tree-based representations with a conditional continuous normalizing flow for probabilistic regression. A CatBoost model trained with MultiRMSE supplies leaf-index features of C; these are encoded into a 32-dimensional representation that conditions a two-block flow over the full response vector. Conditional change-of-variables likelihood trains the flow with a standard Gaussian base. For fixed covariates, generation integrates the conditional flow in reverse to produce a joint response sample. This provides a comparison based on tree-derived conditioning representations while matching the multivariate regression setting.

TabDDPM. TabDDPM models mixed-type tables with Gaussian diffusion for numerical variables and multinomial diffusion for categorical variables (Kotelnikov et al., 2023). We adapt it to block-conditional generation by diffusing the joint $[Y, C]$ representation while treating C as observed and Y as the generated block. During reverse diffusion, the response proposal is updated while covariate coordinates are refreshed from the exact forward-process distribution for the fixed observed

C. The numerical branch uses v-prediction, and the categorical branch retains its variational-bound objective. Generation uses 100 ancestral reverse steps.

TabDiff. TabDiff is a continuous-time diffusion model for mixed-type tables that supports conditional generation through missing-column imputation (Shi et al., 2025). We treat the complete response block Y as missing and the covariates C as observed, so reverse diffusion jointly generates responses while re-imposing the fixed covariates at their corresponding noised states. The model retains mixed numerical/categorical diffusion and learnable feature-wise noise schedules. A target-only model over Y supplies classifier-free guidance, extending the original imputation mask and guidance from a single target to the full multivariate response block. The guidance weight is 0.6.

CAR-Flow. CAR-Flow (Chen et al., 2025) learns condition-aware reparameterizations of flow-matching endpoints. We use its joint source-and-target variant: two covariate-dependent networks shift Gaussian source noise and the observed response, giving $Z_0 = \epsilon + \mu_0(C)$ and $Z_1 = Y + \mu_1(C)$. A velocity conditioned on C is trained by flow-matching regression along the path between these endpoints. Generation starts at the shifted source, follows this conditional velocity, and applies $\widehat{Y} = Z_1 - \mu_1(C)$ to recover the response. The model therefore represents endpoint reparameterization while retaining direct conditioning throughout transport.

CSFM. CSFM (Kim et al., 2026) combines a learnable conditional source with a condition-dependent velocity field. A source network maps C to the mean and scale of a diagonal Gaussian, yielding $X_0 = \mu_\phi(C) + \sigma_\phi(C) \odot \epsilon$. Its velocity reads the evolving response state, time, and C at every evaluation, so conditioning enters both initialization and transport. The networks are trained jointly using linear-path flow matching, source-variance control, and source-target alignment. We instantiate both networks in the tabular response space and generate by drawing from the conditional source and integrating the conditioned velocity. CSFM is a direct comparator for dual conditioning. This standalone implementation serves as an external baseline. The Dual variant in Section 5.4 instead uses the CoFM configuration as a matched control for isolating direct velocity conditioning.

D.3 CoFM-CSF Configuration

CoFM uses the conditional diagonal-Gaussian source defined in Section 4.1 and a shared velocity that receives only the current response state and flow time. The source network is a 3×256 residual GELU MLP with separate mean and log-scale heads. We evaluate two shared-velocity backbones on all four benchmarks. The MLP velocity uses five 512-unit SiLU layers with a 64-dimensional time embedding. The DiT velocity uses four transformer blocks with token dimension 128, one attention head, and expansion factor 2. Both backbones output a d_Y-dimensional velocity and receive no direct context input.

The source and velocity are optimized jointly with

$$\mathcal{L}_{\text{total}}(s) = \mathcal{L}_{\text{FM}} + \lambda_{\text{var}}\mathcal{L}_{\text{var}} + \lambda_{\text{align}}\mathcal{L}_{\text{align}} + \lambda_{\text{CSF}}(s)\mathcal{L}_{\text{CSF}}.$$

For source-variance control, we use a CSFM-style variance-only Gaussian KL with a dataset-specific reference scale. Writing

$$r_i(C) = \frac{\sigma^2_{\phi,i}(C)}{\sigma^2_{\text{ref}}},$$

the normalized variance term is

$$\mathcal{L}_{\text{var}} = \frac{1}{2d_Y}\mathbb{E}_C \sum_{i=1}^{d_Y} \left[r_i(C) - \log r_i(C) - 1 \right].$$

Equivalently, this is the KL divergence from $Q_{\phi,C}$ to the same-mean isotropic Gaussian $\mathcal{N}(\mu_\phi(C), \sigma^2_{\text{ref}} I)$, averaged over response dimensions. The alignment term is the cosine loss on the source mean,

$$\mathcal{L}_{\text{align}} = \mathbb{E}[1 - \cos(\mu_\phi(C), Y)],$$

with $\lambda_{\text{align}} = 0.1$ for all datasets.

CS-Forcing uses the type-respecting neighborhoods and local expansion defined in Section 4.2 and Appendix C. Independent training-only calibration runs estimate M_d and $H_{75,d}$ for each dataset and

neighborhood geometry. Let $\mathcal{B}_d^{\text{cal}}$ be the finite, positive batch base-loss values recorded at calibration probe steps with forcing inactive, and let κ_d^{cal} denote the empirical distribution of saved valid-pair expansion values. Using the base objective and reductions defined above, we set

$$M_d = \text{median}\, \mathcal{B}_d^{\text{cal}}, \qquad H_{75,d} = \mathbb{E}_{\kappa \sim \kappa_d^{\text{cal}}}\big[(\kappa - Q_{0.75}(\kappa_d^{\text{cal}}))_+\big],$$

where the expectation is the empirical mean over valid calibration pairs. The reference coefficient, normalized strength, and expansion budget are

$$\lambda_{\text{anchor},d} = \frac{0.1 M_d}{H_{75,d}}, \qquad \rho_d = \frac{\lambda_{\text{CSF},d} H_{75,d}}{M_d} = 0.1m, \qquad \tau_d = Q_q(\kappa_d^{\text{cal}}),$$

with $\lambda_{\text{CSF},d} = m\lambda_{\text{anchor},d}$. Thus $m \in \{0.25, 0.5, 1\}$ gives the common search grid

$$\rho \in \{0.025, 0.05, 0.10\}, \qquad q \in \{0.50, 0.75, 0.90\}$$

for all four datasets. The factor 0.1, reference quantile 0.75, and grid are fixed design choices; calibration scales and budgets use no validation or test data and remain frozen throughout search and confirmation runs. In particular, $H_{75,d}$ stays fixed when q changes. Thus ρ_d calibrates strength against the reference hinge; it need not equal the realized forcing-to-base-loss ratio during training.

We select the nonzero-forcing candidate with the lowest unweighted mean validation rank across R^2, CRPS, Energy Score, and Variogram Score. Each metric ranks the nine grid candidates together with the zero-forcing control, which provides a common no-forcing reference. Selection is restricted to the nonzero candidates because this search tunes active CS-Forcing; the zero-forcing setting is evaluated separately as an ablation (Appendix E.3). Selection uses a single search run per candidate, followed by evaluation across training seeds; test data never enter selection. Table S6 reports the selected configurations. Their raw coefficients in Table S7 follow from $\lambda_{\text{CSF},d} = \rho_d M_d / H_{75,d}$, so their magnitudes reflect dataset-specific calibration under a common normalized grid.

For the matched 2×2 ablation, the same dataset-specific CS-Forcing configuration is used for the Dual + CSF and Once + CSF arms; we do not retune the forcing parameters separately by allocation, so that context placement remains the only changed factor.

Table S6: **Normalized CoFM-CSF configurations.**

Dataset	σ_{ref}	λ_{var}	λ_{align}	ρ	q
Beijing	2.0	5.0	0.1	0.025	0.90
Superconductivity	0.75	2.0	0.1	0.025	0.90
GTEx	1.5	2.0	0.1	0.025	0.90
UKB-PPP	2.0	5.0	0.1	0.050	0.50

Table S7: **Raw CS-Forcing coefficients induced by the normalized calibration in Table S6.** Their magnitudes are dataset-specific and are not comparable across datasets.

Dataset	λ_{CSF}	τ
Beijing	0.1061	1.3585
Superconductivity	0.001130	5.9044
GTEx	18.2753	0.005362
UKB-PPP	505.5546	0.001112

CS-Forcing is activated after the first 10% of optimization. CoFM-CSF uses AdamW with learning rate 3×10^{-4} and batch size 256. Training runs for 500 epochs on Beijing, 2,000 epochs on Superconductivity, 250 epochs on GTEx, and 15 epochs on UKB-PPP. Generation draws one sample from $Q_{\phi,c}$ and integrates the shared velocity with 25 Euler steps.

D.4 Residual-Conditioning Probe Protocol

The real-data probe in Section 5.3 mirrors the condition-allocation quantity studied in Section 3. For each dataset, we freeze a trained conditional source and vary only how strongly it depends on the context. Let $\bar{\mu}$ and $\bar{\sigma}$ denote the context-averaged source outputs. We interpolate between this context-independent reference and the learned source as

$$\mu_\alpha(c) = \bar{\mu} + \alpha\big(\mu(c) - \bar{\mu}\big), \qquad \sigma_\alpha(c) = \bar{\sigma} + \alpha\big(\sigma(c) - \bar{\sigma}\big),$$

with $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$. Thus $\alpha = 0$ removes source-side context dependence, whereas $\alpha = 1$ recovers the learned source.

For each source level, we construct interpolation states using the same probability path as in the main model and fit two matched velocity probes: a state-only predictor $f_{\mathrm{sh}}(X_t, t)$ and a context-reading predictor $f_{\mathrm{cond}}(X_t, t, C)$. Their held-out squared-risk difference measures the predictive value of context beyond the evolving state. We evaluate this difference across flow time and use its uniform-time average as the residual-conditioning quantity reported in Figure 6. Curves are normalized by the corresponding value at $\alpha = 0$ to compare the contraction of residual conditioning across datasets. The two probes use matched-capacity architectures, so their held-out risk gap isolates informative context rather than additional predictor capacity.

D.5 Evaluation Protocol and Metrics

For each test row, every method generates $M_0 = 256$ joint response samples. Predictive-mean accuracy uses all generated samples and all test examples. Distributional scores use the same fixed subset of $M = 64$ samples and at most $n_s = 1{,}024$ test rows for every method. Before evaluation, observed and generated responses are standardized coordinate-wise using response statistics estimated from the outer training data.

Let $z_i \in \mathbb{R}^{d_Y}$ denote the standardized observed response for test row i, and let $\tilde{z}_{im}$ be its m-th generated standardized response. The generated conditional mean is

$$\bar{z}_i = \frac{1}{M_0} \sum_{m=1}^{M_0} \tilde{z}_{im}.$$

Coefficient of determination (R^2). R^2 measures the accuracy of the generated conditional mean. We use the pooled multivariate form

$$R^2 = 1 - \frac{\sum_i \|z_i - \bar{z}_i\|_2^2}{\sum_i \|z_i\|_2^2}.$$

Because responses are standardized with outer-training statistics, the zero vector corresponds to the outer-training response mean. Larger R^2 therefore indicates that the generated conditional means explain more response variation beyond this reference.

Continuous Ranked Probability Score (CRPS). CRPS evaluates each response coordinate as a full predictive distribution rather than only through its mean. For response coordinate j, it compares generated samples with the observation while subtracting the expected pairwise distance among generated samples. The reported score averages this empirical quantity over contexts and response coordinates:

$$\mathrm{CRPS} = \frac{1}{n_s d_Y} \sum_{i=1}^{n_s} \sum_{j=1}^{d_Y} \left[\frac{1}{M} \sum_{m=1}^{M} |\tilde{z}_{imj} - z_{ij}| - \frac{1}{2M^2} \sum_{m=1}^{M} \sum_{m'=1}^{M} |\tilde{z}_{imj} - \tilde{z}_{im'j}| \right].$$

Lower CRPS indicates better marginal probabilistic accuracy.

Energy Score (ES). Energy Score extends the same sample-based comparison to the complete multivariate response vector. It evaluates whether generated joint samples lie close to the observed response while accounting for the dispersion of the generated distribution:

$$\mathrm{ES} = \frac{1}{n_s} \sum_{i=1}^{n_s} \left[\frac{1}{M} \sum_{m=1}^{M} \|\tilde{z}_{im} - z_i\|_2 - \frac{1}{2M^2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \|\tilde{z}_{im} - \tilde{z}_{im'}\|_2 \right].$$

Lower Energy Score indicates better overall multivariate distributional fidelity.

Variogram Score (VS). Variogram Score focuses on the relative variation between response coordinates and is therefore particularly informative about cross-response dependence. Using exponent $p = \frac{1}{2}$ and equal weight for every unordered pair of response coordinates, we compute

$$\mathrm{VS} = \frac{1}{n_s \binom{d_Y}{2}} \sum_{i=1}^{n_s} \sum_{j<k} \left[|z_{ij} - z_{ik}|^{1/2} - \frac{1}{M} \sum_{m=1}^{M} |\tilde{z}_{imj} - \tilde{z}_{imk}|^{1/2} \right]^2.$$

Lower Variogram Score indicates that the generated samples more closely reproduce the observed pairwise response structure.

The distributional metrics are proper scoring rules (Gneiting & Raftery, 2007). The four metrics therefore evaluate complementary aspects of conditional generation: R^2 measures conditional location, CRPS measures marginal predictive distributions, Energy Score evaluates the joint response distribution, and Variogram Score emphasizes cross-response dependence. Main benchmark entries are reported as the mean and sample standard deviation over three independent training runs.

E Additional Experimental Results

This section extends the main experiments with distributional diagnostics, controlled comparisons, and sensitivity and sampling-time analyses of CoFM.

E.1 Extended Marginal Distribution Recovery

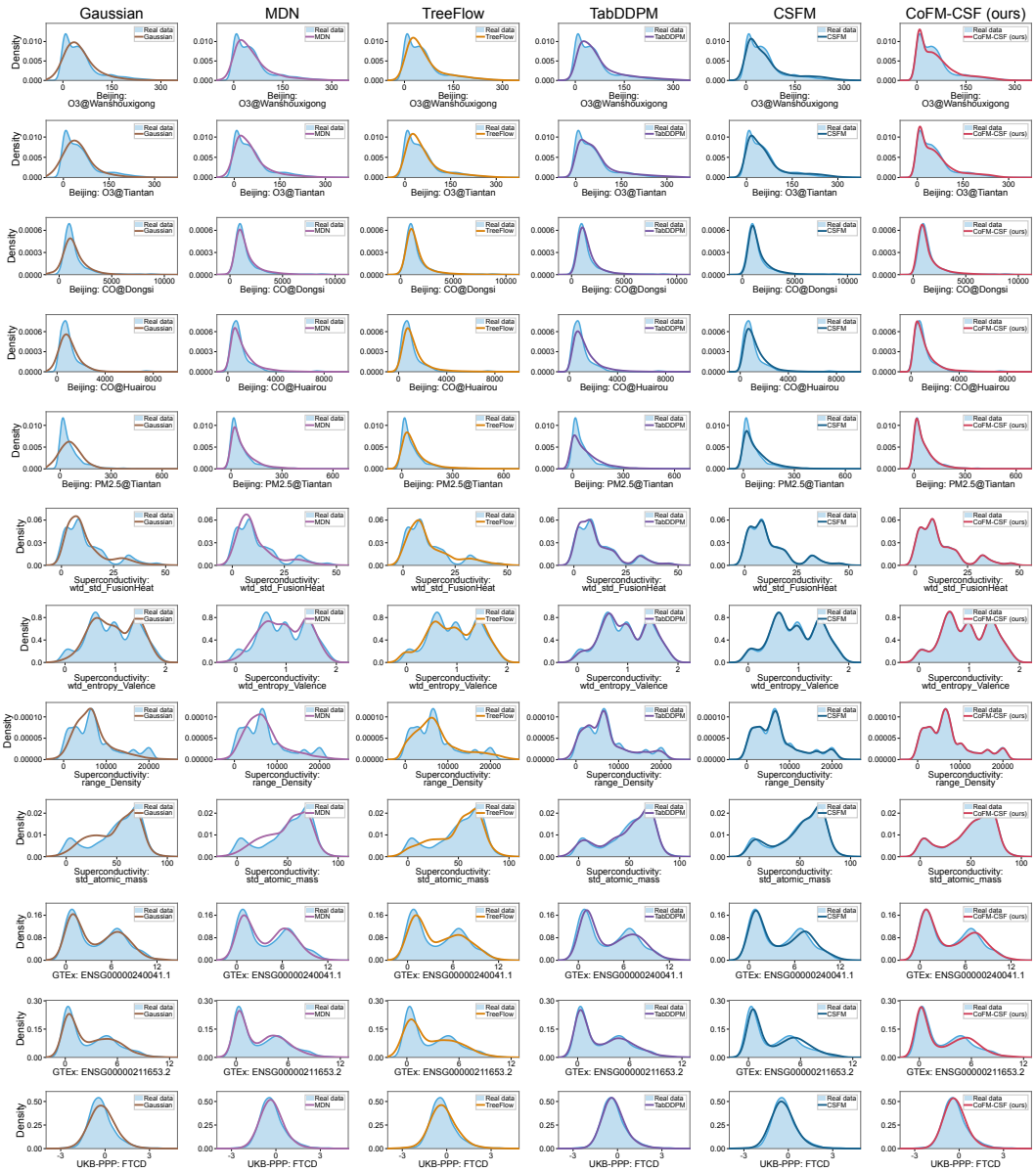

Figure S5: **Extended marginal distribution recovery across all four benchmarks.** Pooled response densities are shown for representative response coordinates from Beijing, Superconductivity, GTEx, and UKB-PPP. Each row compares the observed test distribution with samples from Gaussian, MDN, TreeFlow, TabDDPM, CSFM, and CoFM-CSF under the evaluation protocol in Appendix D.5.

The expanded density atlas in Figure S5 examines the source–transport division across a broader range of response geometries. Beijing contains strongly skewed pollutant marginals with long upper tails, the Superconductivity responses exhibit irregular multi-peak structure at different numerical

scales, and the GTEx examples contain separated modes. CoFM-CSF follows these qualitatively different shapes while using the same diagonal Gaussian source family across tasks. This flexibility is consistent with the intended division of labor: the source organizes where each conditioning input enters the flow, while the shared multivariate dynamics can reshape those initial distributions into the required response geometry.

A source that is simple in distributional form can still carry the covariate information needed to position the state, leaving higher-order shape to the transport. The atlas extends the pooled-density diagnostics beyond the two examples in Figure 4. Differences among methods are visible in mode placement, tail mass, and local density structure, complementing the scalar benchmark scores with evidence of recovered pooled distributional shape. The same shared velocity can act differently on different source-induced regions of state space, allowing conditional initialization to support heterogeneous output marginals without repeatedly supplying covariates to the velocity.

E.2 Response Dependence Across Datasets

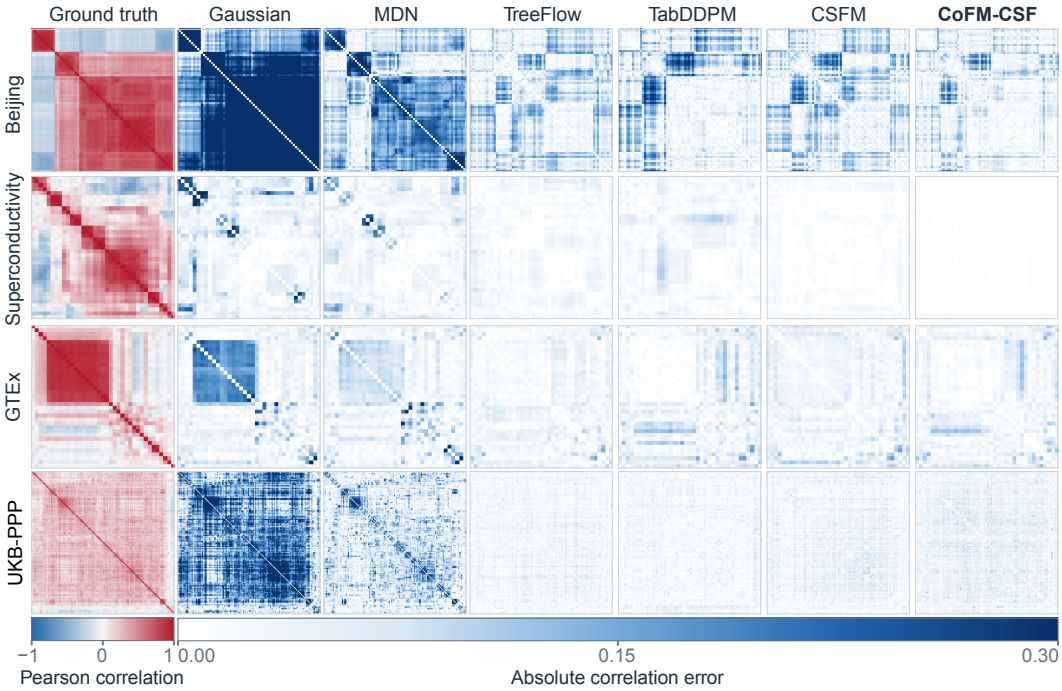

Figure S6: **Response correlation recovery across all four benchmarks.** Rows follow Beijing, Superconductivity, GTEx, and UKB-PPP. The first column shows the observed Pearson correlation matrix for each response block. The remaining columns show absolute correlation error for Gaussian, MDN, TreeFlow, TabDDPM, CSFM, and CoFM-CSF. Correlations are pooled across test conditions; all error maps use a common scale.

The full correlation maps in Figure S6 probe a complementary aspect of conditional generation: whether the generated response block recovers the multivariate organization of the data. The benchmarks contain markedly different dependence geometries: broad blocks in Beijing, localized patterns in Superconductivity, structured blocks in GTEx, and dense high-dimensional structure in UKB-PPP. CoFM-CSF produces low-error correlation maps across these regimes, supporting structured multivariate outputs in addition to accurate marginals.

The dependence results also inform the role of shared transport. At fixed covariates, CoFM's diagonal source has no cross-response covariance, whereas the multivariate flow can introduce conditional coupling. The plotted pooled correlations also include dependence induced by variation across conditions, so they assess overall recovery rather than isolate the origin of coupling. The smaller errors of transported models relative to direct parametric generators are consistent with transport improving multivariate structure. In CoFM-CSF, transport is shared across conditions and receives covariate

information only through the evolving state. Recovering dataset-specific dependence under this design supports the allocation view: the source can encode differences between conditioning inputs, while the shared flow learns reusable multivariate transformations that turn those state differences into response dependence.

E.3 Complete Ablation Across Velocity Backbones

We expand the 2×2 ablation in Section 5.4 to both velocity backbones and all four benchmark metrics (Tables S8 and S9). All four variants share the same CoFM source architecture, velocity backbone, non-CSF training components, and optimization protocol. The first factor controls context allocation: Dual supplies context to both the source and velocity, whereas Once supplies it only to the source. The second factor toggles CS-Forcing, yielding Dual + CSF and CoFM-CSF. This design isolates the two factors within a common implementation and is separate from the independently configured CSFM baseline in Section 5.2.

Table S8: **Complete 2×2 ablation with the MLP velocity backbone.** Values are means $\pm$ sample SDs over three seeds; bold marks the best unrounded mean within each dataset and metric.

Method	Beijing				Superconductivity			
	$R^2 \uparrow$	CRPS↓	Variogram↓	Energy↓	$R^2 \uparrow$	CRPS↓	Variogram↓	Energy↓
Dual	0.3635±0.0083	0.4330±0.0033	0.1422±0.0027	4.7296±0.0459	0.9945±0.0000	0.0288±0.0002	0.0037±0.0001	0.2321±0.0020
Dual + CSF	0.5038±0.0361	0.3584±0.0179	0.1477±0.0057	3.8484±0.1682	**0.9955**±0.0001	0.0239±0.0001	0.0031±0.0000	0.1985±0.0003
Once	0.3826±0.0140	0.4224±0.0114	0.1391±0.0042	4.6180±0.1139	0.9946±0.0000	0.0280±0.0002	0.0037±0.0001	0.2261±0.0021
CoFM-CSF	**0.5390**±0.0049	**0.3335**±0.0040	**0.1117**±0.0005	**3.6878**±0.0323	0.9952±0.0000	**0.0183**±0.0002	**0.0025**±0.0001	**0.1528**±0.0020

Method	GTEx				UKB-PPP			
	$R^2 \uparrow$	CRPS↓	Variogram↓	Energy↓	$R^2 \uparrow$	CRPS↓	Variogram↓	Energy↓
Dual	0.6494±0.0028	0.2948±0.0013	0.1020±0.0004	2.4764±0.0093	0.2990±0.0006	0.4553±0.0005	0.1583±0.0001	7.2073±0.0076
Dual + CSF	0.6498±0.0008	0.2945±0.0002	0.1020±0.0001	2.4751±0.0026	0.2986±0.0014	0.4559±0.0005	0.1583±0.0003	7.2148±0.0037
Once	0.6502±0.0025	0.2935±0.0022	**0.1014**±0.0004	2.4747±0.0092	0.2994±0.0009	0.4551±0.0006	0.1580±0.0001	7.2077±0.0082
CoFM-CSF	**0.6509**±0.0007	**0.2933**±0.0012	0.1014±0.0002	**2.4721**±0.0040	**0.3029**±0.0010	**0.4530**±0.0003	**0.1565**±0.0001	**7.1959**±0.0047

Table S9: **Complete 2×2 ablation with the DiT velocity backbone.** Values are means $\pm$ sample SDs over three seeds; bold marks the best unrounded mean within each dataset and metric.

Method	Beijing				Superconductivity			
	$R^2 \uparrow$	CRPS↓	Variogram↓	Energy↓	$R^2 \uparrow$	CRPS↓	Variogram↓	Energy↓
Dual	0.3832±0.0262	0.4246±0.0109	0.1395±0.0038	4.6340±0.1021	0.9952±0.0003	0.0260±0.0018	0.0033±0.0002	0.2147±0.0122
Dual + CSF	0.4521±0.0193	0.3794±0.0061	0.1231±0.0023	4.1611±0.0664	0.9951±0.0003	0.0261±0.0019	0.0033±0.0003	0.2155±0.0125
Once	0.4019±0.0464	0.3899±0.0141	0.1266±0.0034	4.2760±0.1356	0.9946±0.0001	0.0266±0.0000	0.0036±0.0001	0.2226±0.0007
CoFM-CSF	**0.5343**±0.0096	**0.3285**±0.0038	**0.1123**±0.0020	**3.6758**±0.0305	**0.9953**±0.0001	**0.0182**±0.0003	**0.0025**±0.0000	**0.1536**±0.0026

Method	GTEx				UKB-PPP			
	$R^2 \uparrow$	CRPS↓	Variogram↓	Energy↓	$R^2 \uparrow$	CRPS↓	Variogram↓	Energy↓
Dual	0.6435±0.0075	0.2906±0.0039	0.1015±0.0013	2.4951±0.0239	0.2897±0.0053	0.4655±0.0022	0.1734±0.0018	7.5131±0.0337
Dual + CSF	0.6427±0.0075	0.2910±0.0040	0.1017±0.0013	2.4977±0.0240	0.2925±0.0034	0.4646±0.0023	0.1734±0.0037	7.4995±0.0441
Once	0.6516±0.0008	0.2875±0.0009	0.1002±0.0003	2.4667±0.0023	0.2945±0.0018	0.4657±0.0004	0.1751±0.0010	7.5146±0.0045
CoFM-CSF	**0.6519**±0.0015	**0.2874**±0.0012	**0.1001**±0.0004	**2.4652**±0.0046	**0.3030**±0.0002	**0.4504**±0.0009	**0.1564**±0.0008	**7.1887**±0.0127

The complete factorial results sharpen the condition-allocation finding from the main paper. Before source forcing is introduced, concentrating covariates at the source is already competitive with dual conditioning across both velocity families. Before source forcing, Once remains competitive with Dual across both velocity families, with no systematic degradation across the four metrics. In these matched settings, this is consistent with the residual-conditioning analysis: direct velocity access contributes little once the source-induced state carries the relevant context.

The stronger result is the interaction between allocation and source geometry. Adding CS-Forcing to Once improves 31 of 32 mean comparisons, while adding the same source-side intervention to Dual produces a more mixed pattern. Across the complete 2×2 design, CoFM-CSF gives the best mean in 30 of the 32 dataset–backbone–metric cells. This interaction matches the architectural role of CS-Forcing. Under Condition-Once, the source is the model's direct conditioning interface, so controlling its local expansion acts on the conditioning geometry used throughout generation. The gains are especially pronounced on Beijing and Superconductivity and remain visible on the higher-dimensional UKB-PPP task.

The backbone comparison adds a second piece of evidence. The same allocation–forcing interaction appears with a residual MLP and with DiT, despite their different inductive biases and transport capacity. Improvements also occur jointly in R^2, CRPS, Energy Score, and Variogram Score, linking the design to response location, marginal uncertainty, and cross-response structure. Together, the ablation supports CoFM-CSF as a coupled design: source-concentrated condition allocation defines where covariates enter, and CS-Forcing shapes the geometry of that interface so a shared transport can use it effectively.

E.4 Effect of Source-Family Expressivity

CoFM uses a conditional diagonal Gaussian to initialize covariate-dependent states before shared transport. We examine whether a richer conditional source systematically improves generation under a common auxiliary Condition-Once configuration, where covariates enter only through the source. We compare the default diagonal Gaussian, a wider diagonal Gaussian, a rank-8 low-rank Gaussian, and an equal-weight four-component diagonal Gaussian mixture. Within each dataset, the shared velocity architecture, initialization, optimization budget, loss coefficients, and training schedule are kept fixed. The wider diagonal model is closely matched to the mixture source in source-network parameter count, providing a capacity control for changes in the source distribution family.

To make the source-side objectives comparable across these families, we express them through the conditional mean and marginal variance induced by each source,

$$m_\phi(c) = \mathbb{E}[X_0 \mid C = c], \qquad v_\phi(c) = \operatorname{diag}\operatorname{Cov}(X_0 \mid C = c).$$

For a diagonal Gaussian, $m_\phi = \mu_\phi$ and $v_\phi = \sigma_\phi^2$. For the low-rank source, the initialization and induced marginal variance are

$$\begin{aligned} X_0 &= \mu_\phi(c) + \sigma_\phi(c) \odot \epsilon + B(a_\phi(c) \odot z)\,, \\ v_\phi(c) &= \sigma_\phi^2(c) + (B \odot B) a_\phi^2(c). \end{aligned}$$

For the equal-weight K-component mixture,

$$\begin{aligned} m_\phi(c) &= \frac{1}{K} \sum_{k=1}^{K} \mu_{\phi,k}(c), \\ v_{\phi,i}(c) &= \frac{1}{K} \sum_{k=1}^{K} \left[\sigma_{\phi,k,i}^2(c) + \big(\mu_{\phi,k,i}(c) - m_{\phi,i}(c)\big)^2 \right]. \end{aligned}$$

The variance-control term uses these marginal variances, while mean-based source objectives use $m_\phi(c)$. Flow matching always draws X_0 from the actual source family, so low-rank correlation and mixture structure remain present during transport training. The fixed source-side local control used in this auxiliary comparison is likewise evaluated from the induced mean and marginal scale, and therefore does not require an exact Wasserstein distance between low-rank or mixture sources. This source-family study uses a separate fixed auxiliary configuration from the final dataset-specific CoFM-CSF models in the main benchmark. Accordingly, Table S10 is intended for within-table source-family comparisons rather than direct numerical comparison with Table 1.

On Beijing, increasing source-family expressivity does not improve conditional generation. The diagonal Gaussian gives the strongest results across all four metrics, while widening the same family leaves performance nearly unchanged. Since the wider diagonal source and the mixture source have closely matched source-network parameter counts, their different behavior cannot be attributed to network size alone. These results indicate that a comparatively simple covariate-dependent initialization can already provide shared transport with a useful conditional representation.

Superconductivity exhibits a different regime. The four-component mixture achieves the best mean R^2, CRPS, Variogram, and Energy, showing that additional source structure can be beneficial when it matches the geometry of the task. GTEx follows the Beijing pattern: the diagonal Gaussian leads all four metrics, with negligible change from widening the same family. UKB-PPP shows a milder pattern: the diagonal, wider diagonal, and low-rank Gaussian sources remain closely grouped, whereas the mixture source degrades across all four metrics. The value of additional source expressivity is therefore strongly dataset dependent.

Table S10: **Effect of conditional source-family expressivity.** All source variants are trained under the same auxiliary Condition-Once configuration within each dataset, with the shared velocity architecture and training budget held fixed. Results are mean ± sample standard deviation over three training seeds. Bold denotes the best unrounded mean within each dataset and metric.

Source family	Beijing				Superconductivity			
	$R^2\uparrow$	CRPS↓	Variogram↓	Energy↓	$R^2\uparrow$	CRPS↓	Variogram↓	Energy↓
Diag Gaussian	**0.5249**±0.0064	**0.3343**±0.0036	**0.1138**±0.0005	**3.7275**±0.0248	0.9944±0.0002	0.0292±0.0009	0.0037±0.0001	0.2344±0.0065
Wide Diag Gaussian	0.5234±0.0070	0.3353±0.0028	0.1141±0.0007	3.7372±0.0219	0.9945±0.0001	0.0288±0.0003	0.0037±0.0001	0.2318±0.0008
Low-Rank Gaussian	0.0397±0.1580	0.5806±0.0936	0.1800±0.0048	6.2622±0.7940	0.9949±0.0005	0.0315±0.0038	0.0036±0.0005	0.2561±0.0286
4-component GMM	0.3054±0.1490	0.4964±0.0757	0.1535±0.0123	5.3397±0.6760	**0.9951**±0.0001	**0.0283**±0.0011	**0.0033**±0.0001	**0.2303**±0.0078
Source family	GTEx				UKB-PPP			
	$R^2\uparrow$	CRPS↓	Variogram↓	Energy↓	$R^2\uparrow$	CRPS↓	Variogram↓	Energy↓
Diag Gaussian	**0.6508**±0.0027	**0.2935**±0.0014	**0.1013**±0.0006	**2.4715**±0.0106	0.2849±0.0002	0.4767±0.0006	0.1794±0.0007	7.4659±0.0073
Wide Diag Gaussian	0.6495±0.0027	0.2942±0.0021	0.1015±0.0007	2.4766±0.0113	**0.2876**±0.0003	**0.4754**±0.0007	**0.1790**±0.0005	**7.4485**±0.0079
Low-Rank Gaussian	0.5173±0.0101	0.4152±0.0090	0.1476±0.0033	3.5039±0.0820	0.2849±0.0005	0.4765±0.0006	0.1794±0.0007	7.4634±0.0069
4-component GMM	0.5366±0.0193	0.3845±0.0067	0.1385±0.0007	3.3028±0.0413	0.2639±0.0082	0.5444±0.0111	0.2413±0.0085	8.8323±0.1760

Taken together, this controlled comparison suggests that source expressivity and final generation quality are not monotonically related. The source establishes the covariate-dependent starting distribution, while the shared transport continues to shape substantial aspects of the response distribution. The diagonal Gaussian consequently provides a compact and consistently effective source representation in this study, while richer source families remain useful extensions in regimes where their additional structure translates into downstream gains.

E.5 Validation Sensitivity to CS-Forcing Strength and Budget

We examine the normalized grid in Appendix D.3 using $\lambda_{\mathrm{CSF},d} = m\lambda_{\mathrm{anchor},d}$, with the training-calibrated reference $\lambda_{\mathrm{anchor},d} = 0.1M_d/H_{75,d}$. Thus $m \in \{0.25, 0.5, 1\}$ gives $\rho \in \{0.025, 0.05, 0.10\}$, while $\tau_d = Q_q(\kappa_d^{\mathrm{cal}})$ with $q \in \{0.50, 0.75, 0.90\}$. Figure S7 reports validation R^2, CRPS, Variogram Score, and Energy Score; the same four metrics define the mean-rank selection rule described in Appendix D.3.

Across Beijing, Superconductivity, and GTEx, the four metrics respond coherently as the source constraint is strengthened. Increasing m increases the forcing term's weight on expansion beyond the budget, whereas increasing q raises the budget and preserves a larger range of covariate-dependent source variation. The interaction between the two controls is particularly visible on Beijing: a more permissive budget substantially attenuates the effect of increasing the forcing strength. This behavior reflects the role of the source in CoFM: its geometry should retain distinctions between conditions that support shared transport while controlling excessive local expansion.

UKB-PPP exhibits a substantially flatter response over the same grid, indicating that the influence of the forcing controls is dataset dependent. Together with the no-forcing ablation in Appendix E.3, these results support a conservative operating regime for CS-Forcing: finite source control is beneficial, while increasingly aggressive constraints can suppress useful covariate-dependent variation. The coherent response of all four metrics on the first three datasets shows that this trade-off spans conditional location, marginal distribution quality, and multivariate dependence.

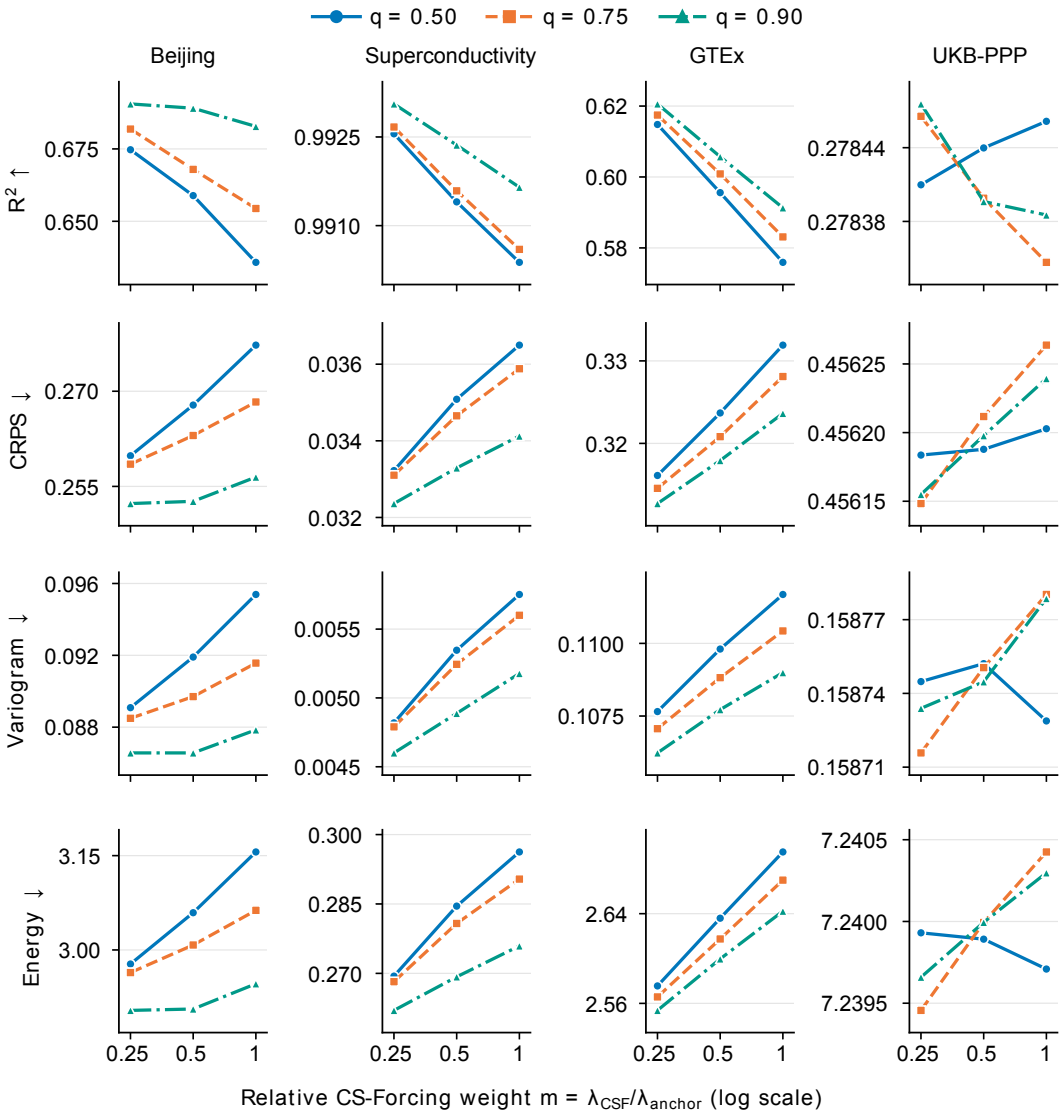

Figure S7: **Validation sensitivity to CS-Forcing strength and expansion budget.** Columns correspond to Beijing, Superconductivity, GTEx, and UKB-PPP; rows report $R^2\uparrow$, CRPS↓, Variogram↓, and Energy↓. The horizontal axis gives the relative forcing strength $m = \lambda_{\mathrm{CSF},d}/\lambda_{\mathrm{anchor},d} = 10\rho_d$, with $\lambda_{\mathrm{anchor},d} = 0.1M_d/H_{75,d}$ as defined in Appendix D.3. Each curve fixes an expansion-budget quantile q, with $\tau_d = Q_q(\kappa_d^{\mathrm{cal}})$. Larger q corresponds to a more permissive source-expansion budget. All values are computed on the validation partitions; final held-out test performance is reported in the main benchmark (Table 1).

E.6 Sampling-Time Trade-off on Beijing

We further examine the sampling-time trade-off on Beijing, using the same trained checkpoints as in the main benchmark (Table 1). For each method, we generate 256 joint responses per row for a fixed subset of 512 test rows while varying the sampling budget from 5 to 50 steps. For this budget sweep, CoFM-CSF, CSFM, and TreeFlow use fixed-step Euler integration, while TabDDPM uses deterministic DDIM sampling over the same step grid. Figure S8 reports CRPS against generation time; each point averages CRPS over three trained models, with shaded regions indicating one sample standard deviation across training seeds. The horizontal coordinate is the median generation time across the three runs, all measured on an NVIDIA L40S GPU.

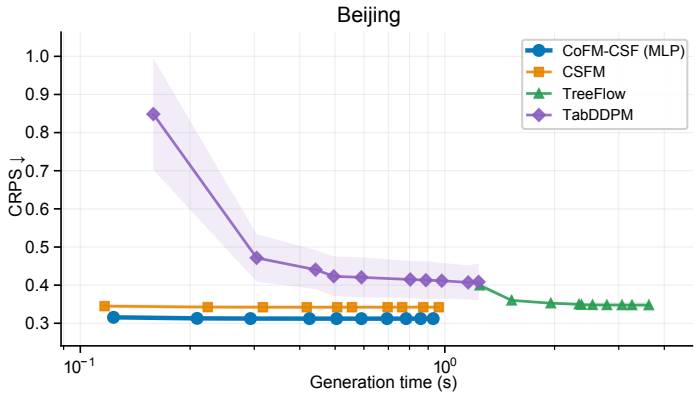

Figure S8: **Sampling-time trade-off on Beijing.** CRPS is plotted against generation time for CoFM-CSF (MLP), CSFM, TreeFlow, and TabDDPM over sampling budgets from 5 to 50 steps. CoFM-CSF, CSFM, and TreeFlow use fixed-step Euler integration, while TabDDPM uses deterministic DDIM sampling. Each point reports the mean CRPS over three trained models on the same 512 test rows with 256 generated responses per row; shaded regions show one sample standard deviation across training seeds. Generation time is the median of three runs on an NVIDIA L40S GPU.

On Beijing, CoFM-CSF reaches a stable low-CRPS regime with few integration steps and remains insensitive to larger ODE budgets over the tested range.

F LIMITATIONS AND FUTURE WORK

Response types. This work focuses on jointly generating a predefined block of continuous responses from observed context. Although context fields can have heterogeneous types, the response space considered here is continuous. A natural extension is to generate mixed-type responses that combine continuous, categorical, and count-valued variables. This would require source and transport formulations that respect the support of each response type while preserving conditional dependence across them. Extending the benchmarks accordingly would broaden the task while retaining the fixed-context, joint-response formulation.

Beyond tabular generation. The condition-saturation analysis characterizes the residual value of direct velocity conditioning for a given source and flow path, and CoFM instantiates this principle for tabular generation. Our empirical evidence comes from controlled studies and real-world tabular benchmarks. Future work could investigate condition saturation on real-world data beyond tabular settings, where contexts and generated states have different structures. Such studies would examine how source representations and transport paths affect the context information retained in the state and the remaining value of direct velocity conditioning. Developing corresponding theoretical extensions and empirical diagnostics could guide condition allocation across a broader range of generation tasks.