MEALGLUCO: PHYSIOLOGICALLY STRUCTURED POSTPRANDIAL GLUCOSE TRAJECTORY FORECASTING WITH ONLINE PERSONALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Predicting postprandial glucose trajectories is important for proactive metabolic health management but remains underexplored. This task is challenging because a model must capture long-horizon glucose dynamics, integrate the effects of the current meal with the evolving pre-meal glucose state, and account for substantial inter-individual variability. To address these challenges, we propose MealGluco, a physiologically structured model that predicts the complete, high-resolution postprandial glucose trajectory from glucose history and meal composition. MealGluco represents the meal-induced response using a peak-normalized Gamma function, captures background glucose dynamics through a separate state pathway, and recursively updates two participant-specific coordinates to enable lightweight online personalization without retraining the shared network. Across three datasets and five evaluation metrics, MealGluco achieves a mean rank of 1.33 among seven evaluated models. Its online personalization reduces trajectory prediction error relative to the non-personalized model while matching the performance of participant-specific fine-tuning at orders-of-magnitude lower computational and storage costs. MealGluco also consistently outperforms the strongest baseline in cross-dataset evaluation, demonstrating improved generalization across populations and devices.

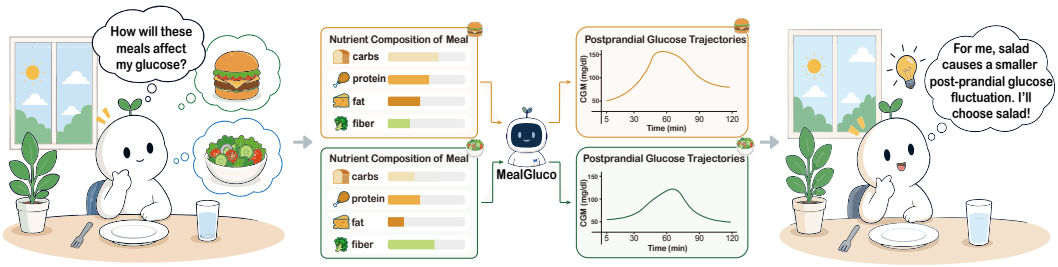

Figure 1: Postprandial glucose trajectory forecasting task.

1 INTRODUCTION

Glucose fluctuations have important implications for metabolic health, and diet is one of their major modifiable drivers. As illustrated in Figure 1, predicting postprandial glucose responses can therefore support precision nutrition and proactive, personalized health management (Zeevi et al., 2015; Berry et al., 2020).

Existing AI-enabled glucose prediction research has primarily focused on two tasks, each leaving an important gap. General glucose forecasting methods learn temporal patterns from historical continuous glucose monitoring (CGM) data, but they largely target short prediction horizons, often around 30 minutes, and do not explicitly model the prolonged and complex effects of the current meal (Sergazinov et al., 2023; Muhammad et al., 2026). In contrast, postprandial glucose prediction methods incorporate meal composition, clinical phenotypes, and lifestyle factors, but typically reduce the response to scalar outcomes, such as incremental area under the curve (iAUC) or peak

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.

glucose, or to glucose values at a few predefined time points (Zeevi et al., 2015; Mendes-Soares et al., 2019; Berry et al., 2020; Indriani et al., 2026). These sparse summaries cannot fully characterize how glucose rises, peaks, and recovers, limiting their utility for precision nutrition. To bridge this gap, we formulate the task of predicting the complete, high-resolution CGM trajectory over the hours following a meal, explicitly conditioned on the current meal and individual context.

This task presents three key challenges. First, postprandial trajectory forecasting must capture long-horizon glucose dynamics, including the delayed rise, peak, and recovery, rather than relying only on temporal continuity between successive glucose measurements (Dalla Man et al., 2007). Second, accurate prediction requires integrating the nutritional perturbation introduced by the current meal with the participant's evolving pre-meal glucose state. Third, the same meal can elicit markedly different responses across individuals, requiring scalable personalization beyond a shared population-level model (Zeevi et al., 2015; Berry et al., 2020).

To address these challenges, we propose **MealGluco**, a physiologically structured framework for complete postprandial glucose trajectory forecasting. MealGluco decomposes the forecast into a history-dependent state response and a meal-induced response. The state pathway uses pre-meal CGM and temporal context to model background glucose dynamics. In parallel, the meal pathway maps nutrient composition and pre-meal context to a low-dimensional kinetic kernel parameterized by response amplitude, peak time, and shape, thereby capturing the characteristic rise, peak, and recovery of postprandial glucose (Thompson Jr et al., 1964; Glover, 1999; Savic et al., 2007). The representations produced by the two pathways are then fused in the latent space to predict the complete trajectory. To accommodate inter-individual variability, MealGluco introduces two participant-specific coordinates that adjust response magnitude and post-peak recovery. These coordinates are recursively updated from completed meals through a local Gaussian information update (Eberle & Ament, 2012; Albers et al., 2017), while the shared network remains fixed. MealGluco can therefore generate an initial population-level forecast for a participant with no prior meal history and progressively personalize subsequent forecasts as completed meals accumulate, without participant-specific model fine-tuning.

We evaluate MealGluco on three independent datasets spanning diverse populations, devices, and study settings, and compare it with a broad range of state-of-the-art glucose forecasting baselines. Our evaluation examines five complementary aspects of the predicted trajectory: overall trajectory accuracy, peak excursion, cumulative glycemic exposure, peak timing, and post-peak recovery. Component-wise ablations quantify the contributions of the structured meal response, history-dependent state modeling, and online personalization. We further conduct cross-dataset evaluation to assess generalization across devices and populations, reflecting a practical deployment scenario in which a model is applied to data collected under previously unseen conditions.

Our main contributions are as follows:

- We formulate **meal-conditioned complete postprandial glucose trajectory prediction**: predicting a high-resolution, long-horizon glucose trajectory from pre-meal CGM and current meal composition in a single forward pass.
- We propose **MealGluco**, which combines physiologically motivated kinetic structure with data-driven modeling and supports efficient online personalization by updating only two low-dimensional response coordinates, without model fine-tuning.
- Across three independent datasets and five evaluation metrics, MealGluco ranks first or second in all **15 dataset–metric comparisons**, achieving a mean rank of **1.33** among seven evaluated models. It also demonstrates the strongest cross-dataset generalization, achieving the lowest *Trajectory-MAE* in all six transfer settings.

2 Related Work

CGM forecasting and glucose representation learning. CGM forecasting has largely borrowed architectures from general sequence modeling, including recurrent networks, temporal convolutional networks, and Transformers (Hochreiter & Schmidhuber, 1997; Bai et al., 2018; Vaswani et al., 2017). Gluformer applies a Transformer to individual glucose series and models predictive uncertainty (Sergazinov et al., 2023); GlucoBench curates public CGM datasets and standardizes

prediction benchmarks (Sergazinov et al., 2024); and recent work learns transferable glucose representations through self-supervised pretraining on large CGM corpora (Lu et al., 2025; Muhammad et al., 2026; Li et al., 2026). These models continue an observed glucose history with high fidelity, but the current meal enters only through that history: the learning problem is a history-to-future mapping rather than a meal-level response whose amplitude, peak timing, and shape are explicit model variables.

Meal-conditioned postprandial prediction. Conditioning a forecast on the current meal requires fusing nutritional inputs with the glucose history. *TFT* selects and gates exogenous variables (Lim et al., 2021), *TimeXer* models interactions between an endogenous series and exogenous covariates (Wang et al., 2024), and *UniCA* maps heterogeneous covariates into a shared sequence space before fusion (Han et al., 2026). A parallel literature predicts postprandial responses from diet composition, clinical phenotypes, and lifestyle information, typically as a postprandial glucose response score or an incremental area under the curve (Zeevi et al., 2015; Mendes-Soares et al., 2019; Berry et al., 2020), and relates nutrient composition to glycemia in free-living datasets (Liang, 2026; Das et al., 2025; Cho et al., 2023; Alsuhaymi et al., 2025). Diagnosis-aware stacked learning trains separate predictors for a few fixed postprandial endpoints and combines them in a second stage (Indriani et al., 2026; Prokhorenkova et al., 2018). Across these works, meal information is either treated as a generic covariate or compressed into an aggregate endpoint, so the nutritional response is not represented as a structured object inside the model. Our work instead maps the current meal to a small set of response coordinates and decodes the complete postprandial trajectory with one shared predictor in a single forward pass.

Structured response modeling. Low-dimensional kernels have a long history in describing delayed, unimodal, and decaying exogenous responses, from indicator-dilution and hemodynamic analysis to drug-absorption modeling (Thompson Jr et al., 1964; Glover, 1999; Savic et al., 2007; Hentze et al., 2017). Meal simulation models similarly treat ingestion as an exogenous event that drives glucose appearance through gastrointestinal absorption (Dalla Man et al., 2007; Cobelli et al., 2014). Such mechanistic descriptions are physiologically informative but rely on variables and controlled conditions that free-living CGM settings rarely provide. A complementary line couples structural priors with data-driven components through Bayesian calibration and hybrid or universal differential equations (Kennedy & O'Hagan, 2001; Pathak et al., 2018; Karniadakis et al., 2021; Rackauckas et al., 2020). Our work adopts a *mechanism–data dual-driven approach*, using a peak-normalized Gamma function to preserve the physiological pattern of delayed rise, peak, and recovery. Neural networks learn the function parameters from free-living data, while a state component captures background glucose variation.

Online personalization. Because individuals differ in insulin sensitivity and in the magnitude and duration of their postprandial responses, population models are often complemented with participant-specific adaptation. State estimation and data assimilation provide classical tools for updating latent states from sequential observations (Kalman, 1960; Eberle & Ament, 2012; Albers et al., 2017), while recent neural approaches explore incremental learning, meta-learning, and patient-level parameter updates (Li et al., 2020; Zhu et al., 2022). Although participant-level adaptation can improve accuracy, model fine-tuning requires repeated optimization and participant-specific model states. Our proposed MealGluco instead freezes the shared network and recursively updates only two participant-level response coordinates, therefore enabling lightweight online personalization without backpropagation or optimizer states.

3 METHODS

Problem formulation. For participant i's m-th meal at time t_0, we observe an H-point pre-meal CGM window

$$\mathbf{G}_{i,m}^{\mathrm{pre}} = \{G_i(t_0 + h\Delta t)\}_{h=-(H-1)}^{0},$$

where Δt is the sampling interval. The meal is represented by its nutrient doses $\mathbf{n}_{i,m} = [n_{i,m}^{\mathrm{c}}, n_{i,m}^{\mathrm{p}}, n_{i,m}^{\mathrm{f}}, n_{i,m}^{\mathrm{fi}}]$, corresponding to carbohydrate, protein, fat, and fiber, and the participant's fasting glucose is denoted by G_i^{fast}. Our goal is to predict the complete postprandial trajectory

$$\mathbf{G}_{i,m}^{\mathrm{post}} = \{G_i(t_0 + h\Delta t)\}_{h=1}^{H}.$$

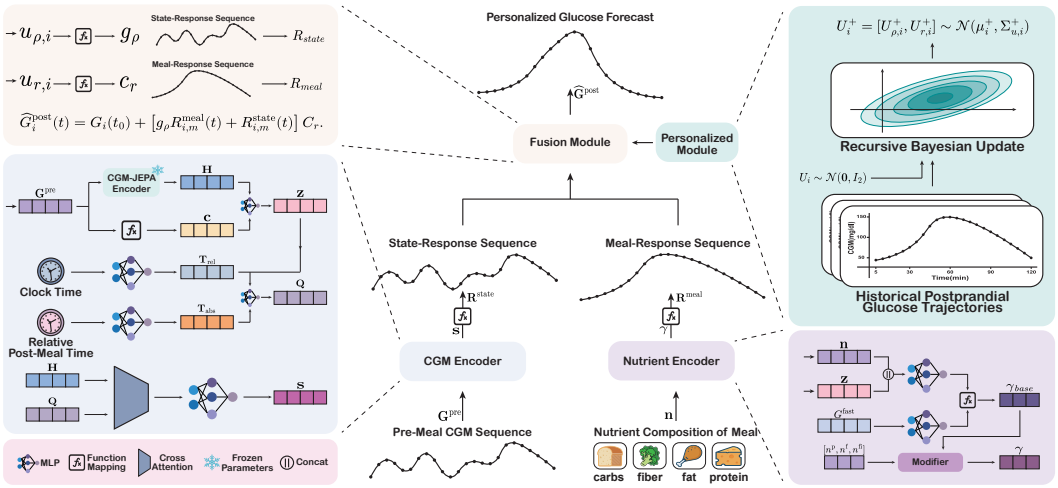

Figure 2: Overview of MealGluco. The model separates the structured response to the current meal from the continuation of the pre-meal glucose state, and adapts the resulting forecast through two recursively updated participant-specific coordinates.

This task differs from conventional history-to-future forecasting because the future trajectory is jointly driven by two sources: the evolving physiological state before the meal and the exogenous perturbation introduced by the current meal. It also requires participant-specific adaptation, as similar meals can elicit substantially different responses across individuals. A useful model should therefore (i) represent the characteristic rise, peak, and recovery induced by a meal, (ii) preserve background glucose dynamics that are not explained by the meal, and (iii) personalize predictions from a small and growing number of completed meals.

3.1 MealGluco: model overview

MealGluco addresses these requirements by decomposing the postprandial trajectory into a meal-time anchor, a structured meal-induced response, and a history-dependent state response. Let $t_h = h\Delta t$ denote time since the meal. The population forecast at horizon t_h is

$$\widehat{G}^{\mathrm{pop}}_{i,m}(t_h) = G_i(t_0) + R^{\mathrm{meal}}_{i,m}(t_h) + R^{\mathrm{state}}_{i,m}(t_h). \tag{1}$$

The meal response $R^{\mathrm{meal}}_{i,m}$ is generated by a low-dimensional kinetic kernel whose amplitude, peak time, and shape are inferred from meal composition and the pre-meal state. The state response $R^{\mathrm{state}}_{i,m}$ carries forward glucose dynamics already present before the meal and acts as a bounded correction to the structured meal response.

To personalize the forecast, MealGluco maintains two participant-specific coordinates $\mathbf{u}_i = [u_{\rho,i}, u_{r,i}]^\top$, which adjust response magnitude and post-peak recovery, respectively. These coordinates are initialized from a population prior and updated recursively after each completed meal; all shared model parameters remain fixed. The resulting model provides a population forecast when no prior meal history is available for personalization, obtained by setting $\mathbf{u}_i = \mathbf{0}$, and progressively personalized the forecast as individual meal history accumulates. Figure 2 summarizes the architecture. We next describe the state response, structured meal response, trajectory composition, and online update. The main text presents the core formulation, and detailed derivations and implementation choices for the components described below are provided in Appendices A.3–A.7.

3.2 History-dependent state response

The state pathway summarizes the pre-meal CGM history and decodes its continuation over the forecast horizon. We first apply a pretrained and frozen CGM-JEPA encoder E_ψ (Muhammad et al., 2026) to obtain temporal representations. Because pretrained sequence representations may not preserve the latest glucose level and short-term slope with sufficient precision, we augment them with recent-history statistics $\mathbf{c}(\mathbf{G}^{\mathrm{pre}}_{i,m})$. A trainable state encoder E_z then produces

$$\mathbf{z}_{i,m} = E_z\big(\mathrm{Pool}\big(E_\psi(\mathbf{G}^{\mathrm{pre}}_{i,m})\big) \,\|\, \mathbf{c}(\mathbf{G}^{\mathrm{pre}}_{i,m})\big)\,, \tag{2}$$

where $\|$ denotes concatenation.

The effect of a pre-meal state need not be constant across the forecast horizon. We therefore construct a query $\mathbf{q}_{i,m,h}$ for each future point from relative post-meal time and clock time. Relative time locates the point within the postprandial response, whereas clock time captures time-of-day variation. A shared future decoder F_Q maps the state and queries to pointwise representations,

$$\mathbf{Q}_{i,m} = F_Q\big(\mathbf{z}_{i,m}, \{\mathbf{q}_{i,m,h}\}_{h=1}^{H}\big)\,. \tag{3}$$

A prediction head F_s converts these representations into a bounded residual sequence:

$$\mathbf{R}_{i,m}^{\text{state}} = R_{\max}\tanh(F_s(\mathbf{Q}_{i,m}))\,. \tag{4}$$

Bounding this pathway prevents the flexible decoder from absorbing the entire meal response and preserves the intended separation between background dynamics and the nutritional perturbation.

3.3 Structured meal-induced response

Rather than predicting the nutritional effect as an unconstrained sequence, MealGluco represents it using a peak-normalized Gamma kernel. The nutrient encoder E_γ first maps the pre-meal state $\mathbf{z}_{i,m}$, nutrient doses $\mathbf{n}_{i,m}$, and fasting glucose G_i^{fast} to a base response:

$$\boldsymbol{\gamma}_{i,m}^{\text{base}} = E_\gamma(\mathbf{z}_{i,m}, \mathbf{n}_{i,m}, G_i^{\text{fast}}) = [\rho_{i,m}^{\text{base}}, \tau_{i,m}^{\text{base}}, \kappa_{i,m}^{\text{base}}]^\top, \tag{5}$$

where $\rho_{i,m}^{\text{base}}$ controls the base response amplitude, $\tau_{i,m}^{\text{base}}$ is the base time to peak, and $\kappa_{i,m}^{\text{base}}$ controls the base response shape. A modifier M_ϕ then combines this base response with the protein, fat, and fiber doses, collectively denoted $\mathbf{n}_{i,m}^{\text{mix}}$, to produce the final response parameters:

$$\boldsymbol{\gamma}_{i,m} = M_\phi\big(\boldsymbol{\gamma}_{i,m}^{\text{base}}, \mathbf{n}_{i,m}^{\text{mix}}\big) = [\rho_{i,m}, \tau_{i,m}, \kappa_{i,m}]^\top. \tag{6}$$

This two-stage design first forms a base excursion from the pre-meal state and the full nutrient composition, and the modifier then uses protein, fat, and fiber to refine its amplitude and peak timing. The modifier corresponds to the Modifier block in Figure 2. These final parameters generate

$$R_{i,m}^{\text{meal}}(t) = \rho_{i,m} K(t; \tau_{i,m}, \kappa_{i,m}), \tag{7}$$

$$K(t; \tau, \kappa) = \left(\frac{t}{\tau}\right)^{\kappa} \exp\left[\kappa\left(1 - \frac{t}{\tau}\right)\right]. \tag{8}$$

The kernel satisfies $K(\tau; \tau, \kappa) = 1$, so ρ directly specifies the peak amplitude, while τ and κ separately control its timing and shape. This parameterization provides a compact inductive bias for delayed, unimodal responses (Thompson Jr et al., 1964; Savic et al., 2007) while allowing the response parameters to be learned from free-living data.

3.4 Trajectory composition and online personalization

Eq. 1 combines the two response pathways with the observed mealtime glucose. To capture systematic participant-level variation, we introduce latent coordinates

$$\mathbf{U}_i = [U_{\rho,i}, U_{r,i}]^\top \sim \mathcal{N}(\mathbf{0}, I_2), \tag{9}$$

shared across participant i's meals. The first coordinate changes the overall response gain, and the second changes post-peak recovery. We map them to bounded multiplicative factors,

$$g_\rho(u_{\rho,i}) = \exp[\log(2)\tanh(u_{\rho,i})]\,, \tag{10}$$

$$C_r(t; u_{r,i}, \tau) = \exp[-\log(2)\tanh(u_{r,i}) q(t; \tau)]\,, \tag{11}$$

where $q(t; \tau)$ is a post-peak ramp. Both factors lie in $[1/2, 2]$, preventing a small amount of participant history from producing arbitrarily large corrections. The personalized forecast is

$$\widehat{G}_{i,m}^{\text{post}}(t) = G_i(t_0) + \big[g_\rho(u_{\rho,i}) R_{i,m}^{\text{meal}}(t) + R_{i,m}^{\text{state}}(t)\big]\, C_r(t; u_{r,i}, \tau_{i,m}). \tag{12}$$

At the prior mean $\mathbf{u}_i = \mathbf{0}$, both modifiers equal one and Eq. 12 reduces to the population forecast.

After a meal's outcome window has closed, its observed trajectory provides evidence about whether the participant tends to exhibit a stronger or weaker response and faster or slower recovery than

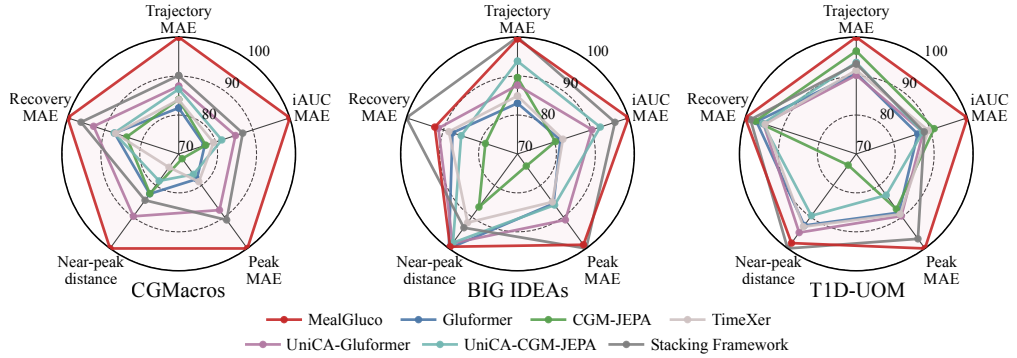

Figure 3: Normalized performance across the five metrics on the three datasets. Higher is better on all axes. Raw errors are reported in Table 1 and Appendix Tables 7–8.

predicted by the population model. We assimilate this evidence through a local Gaussian update. Let the current posterior be represented in information form by precision Λ_i, natural parameter $\boldsymbol{\eta}_i$, and mean $\boldsymbol{\mu}_i = \Lambda_i^{-1}\boldsymbol{\eta}_i$. Linearizing Eq. 12 around $\boldsymbol{\mu}_i$ for a completed meal m' yields Jacobian $J_{m'}$ and residual observation $\widetilde{\mathbf{G}}_{m'}^{\text{post}}$. The posterior update is

$$\Lambda_i^+ = \Lambda_i + J_{m'}^\top W_\varepsilon J_{m'}, \tag{13}$$

$$\boldsymbol{\eta}_i^+ = \boldsymbol{\eta}_i + J_{m'}^\top W_\varepsilon \widetilde{\mathbf{G}}_{m'}^{\text{post}}, \tag{14}$$

$$\boldsymbol{\mu}_i^+ = (\Lambda_i^+)^{-1}\boldsymbol{\eta}_i^+, \tag{15}$$

where W_ε is the residual precision. This is the standard Gaussian information-form update (Bishop, 2006, Section 3.3). Before predicting meal m, MealGluco assimilates only earlier meals whose full outcome windows are available, each exactly once. The update is therefore causal and requires neither backpropagation nor participant-specific optimizer states. The shared network is unchanged, and the persistent participant state remains two-dimensional.

3.5 Training objective

We train the shared population model with participant coordinates fixed at their prior mean, $\mathbf{u}_i = \mathbf{0}$, and apply online personalization only after training. Let $e_{i,m,h} = \widehat{G}_{i,m}^{\text{pop}}(t_h) - G_{i,m}^{\text{post}}(t_h)$, and let $a_{i,m,h} \in \{0,1\}$ indicate whether the target is observed. The pointwise trajectory loss is

$$\mathcal{L}_{\text{point}} = \frac{1}{M}\sum_{i,m}\frac{\sum_{h=1}^{H} a_{i,m,h} w_h \ell_{i,m,h}^{\text{point}}}{\sum_{h=1}^{H} a_{i,m,h} w_h}, \tag{16}$$

where $\ell_{i,m,h}^{\text{point}}$ is a Huber–squared-error mixture and w_h is a fixed temporal weight. The robust component reduces sensitivity to isolated CGM fluctuations, while the squared-error component and temporal weighting preserve accuracy across the full forecast horizon.

Pointwise supervision alone may underestimate the magnitude of the postprandial excursion. We therefore add a peak loss. Let $\bar{G}_{i,m}^{\text{post}}(t_h)$ and $\widehat{\bar{G}}_{i,m}^{\text{post}}(t_h)$ denote smoothed observed and predicted trajectories, including the mealtime anchor at $h = 0$. We define

$$P_{i,m} = \max_{0\le h\le H} \bar{G}_{i,m}^{\text{post}}(t_h) - G_i(t_0), \tag{17}$$

$$\widehat{P}_{i,m} = \max_{0\le h\le H} \widehat{\bar{G}}_{i,m}^{\text{post}}(t_h) - G_i(t_0), \tag{18}$$

$$\mathcal{L}_{\text{peak}} = \frac{1}{M}\sum_{i,m} \ell_{\text{H}}(\widehat{P}_{i,m} - P_{i,m}). \tag{19}$$

The final objective is $\mathcal{L} = \mathcal{L}_{\text{point}} + \lambda_{\text{peak}}\mathcal{L}_{\text{peak}}$. The pretrained encoder E_ψ remains frozen, and all other shared parameters are optimized jointly with AdamW.

4 Results

Datasets. We evaluate MealGluco on three public datasets that pair 5-minute CGM with meal-level macronutrient records: CGMacros (Das et al., 2025) (45 participants and 1,650 qualifying

Table 1: Main comparison on CGMacros. Entries are mean ± sample SD over three independent seeds; lower is better in every column, the best mean is bold and the second best is underlined.

Model	Trajectory-MAE	iAUC-MAE	Peak-MAE	Near-Peak Distance	Recovery-MAE
Gluformer (CGM-only)	25.87 ± 0.32	2209.4 ± 30.4	32.07 ± 0.17	21.22 ± 0.46	21.97 ± 0.39
CGM-JEPA (CGM-only)	25.19 ± 0.15	2201.2 ± 20.4	34.97 ± 0.69	21.24 ± 0.99	22.82 ± 0.11
UniCA-Gluformer	24.21 ± 0.57	1994.3 ± 47.1	28.47 ± 0.33	<u>19.53 ± 0.23</u>	20.63 ± 0.23
UniCA-CGM-JEPA	24.43 ± 0.18	2084.4 ± 20.3	32.70 ± 0.54	22.33 ± 1.14	21.95 ± 0.41
TimeXer	25.22 ± 0.05	2143.2 ± 18.6	31.73 ± 1.04	23.63 ± 2.42	22.04 ± 0.99
Stacking Framework	<u>23.48 ± 0.01</u>	<u>1949.4 ± 1.4</u>	<u>27.49 ± 0.02</u>	20.71 ± 0.36	<u>19.91 ± 0.04</u>
MealGluco	**21.16 ± 0.17**	**1701.7 ± 21.3**	**25.00 ± 0.32**	**17.54 ± 0.33**	**19.18 ± 0.32**
MealGluco vs. second best	-9.9%	-12.7%	-9.1%	-10.2%	-3.6%

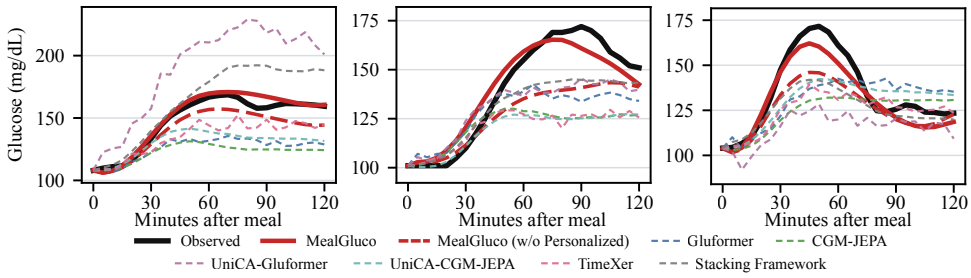

Figure 4: Representative validation-set cases.

meals), BIG IDEAs (Cho et al., 2023) (14 participants and 534 qualifying meals), and T1D-UOM (Alsuhaymi et al., 2025) (7 participants and 1,487 qualifying meals). Each prediction is conditioned on the preceding two-hour pre-meal CGM window and targets the subsequent two-hour postprandial trajectory.

Baselines. We compare MealGluco with six recent models that cover general glucose forecasting method *Gluformer* (Sergazinov et al., 2023) and *CGM-JEPA* (Muhammad et al., 2026), covariate fusion method *UniCA-Gluformer* and *UniCA-CGM-JEPA* (Han et al., 2026), postprandial endpoint prediction model *TimeXer* (Wang et al., 2024), and the *Stacking Framework* for postprandial glucose prediction (Indriani et al., 2026; Prokhorenkova et al., 2018). Gluformer and CGM-JEPA use only the pre-meal CGM window and predict directly from the CGM history without meal covariates. UniCA-Gluformer, UniCA-CGM-JEPA, and TimeXer use the meal-level nutrient, fasting-glucose, and clock-time inputs and directly predict the complete 24-point, two-hour CGM trajectory. The Stacking Framework instead trains a separate predictor for each forecast horizon and concatenates the 24 point predictions, so it does not directly generate the full trajectory.

Metrics. We score every model with five complementary errors, all averaged over meals and all lower-is-better. (i) *Trajectory-MAE* (mg/dL) measures the overall error across the 24 postprandial points. (ii) *iAUC-MAE* (mg/dL·min) measures the error in cumulative glycemic exposure above the meal-time anchor. (iii) *Peak-MAE* (mg/dL) measures the error in the postprandial peak CGM level. (iv) *Near-Peak Distance* (min) measures the error in the timing of the peak. (v) *Recovery-MAE* (mg/dL) measures the error in the post-peak decline by 120 min.

Experimental design. Every model is trained and evaluated separately within each dataset under participant-disjoint outer validation: 5-fold grouped cross-validation on CGMacros and BIG IDEAs, and leave-one-participant-out on T1D-UOM, which contains seven participants. We repeat the full procedure with 3 seeds and report the mean ± sample SD over the three seeds. Participant-clustered paired bootstrap intervals are reported for the online-versus-fine-tuning comparison and the locked personalization analysis. Online personalization is strictly causal: before predicting a meal, only that participant's earlier meals whose 120-minute outcome windows have closed are assimilated, each exactly once; with no such history, the prediction reduces to the population model. All dataset construction, baseline adaptation, metric definitions, and validation details are provided in Appendices B.1–B.4.

Overall performance. Extensive comparisons against six recent baselines demonstrate that MealGluco achieves the strongest overall performance across the three datasets and five complementary metrics: it ranks first or second in all 15 dataset–metric comparisons, with a mean rank of 1.33 among the seven evaluated models. Table 1 and Appendix Tables 7–8 report the complete quanti-

tative comparison, and Figure 3 summarizes normalized performance across all five metrics. Consistent results across the three datasets support this overall advantage: MealGluco is best on all five CGMacros metrics and on *iAUC-MAE* in every dataset, while its remaining per-metric results stay within the top two. Figure 4 further shows that similar aggregate errors can mask different trajectory shapes. Several baselines attain comparable metric values but still deviate from the observed curves, whereas MealGluco uses its participant-specific coordinates to better adapt amplitude, timing, and recovery to each participant.

Meal-conditioned response. We next examine whether the learned Gamma parameters are consistent with the intended physiologically structured parameterization and respond to changes in nutrient intake. In Figure 5(a), the generated (ρ, τ, κ) correlate positively with peak amplitude, response timing, and peak width across datasets, with the strongest association between ρ and peak amplitude. The τ-timing and κ-width associations are weaker but consistently positive. Figure 5(b) fixes the participant and varies meal composition. Across meals, the observed trajectory, Gamma parameters, and forecast change together, confirming that the meal pathway carries meal-specific information beyond participant identity. Figure 5(c) reports local Jacobian sensitivities to the four nutrient inputs. Peak amplitude ρ is dominated by carbohydrate and secondarily by fiber, while response timing τ and shape κ receive substantial contributions from protein, fat, and fiber. Together, these results support the physiologically structured design of MealGluco: the learned Gamma parameters align with observable trajectory morphology, vary systematically with meal composition, and expose nutrient-specific control over response amplitude, timing, and shape.

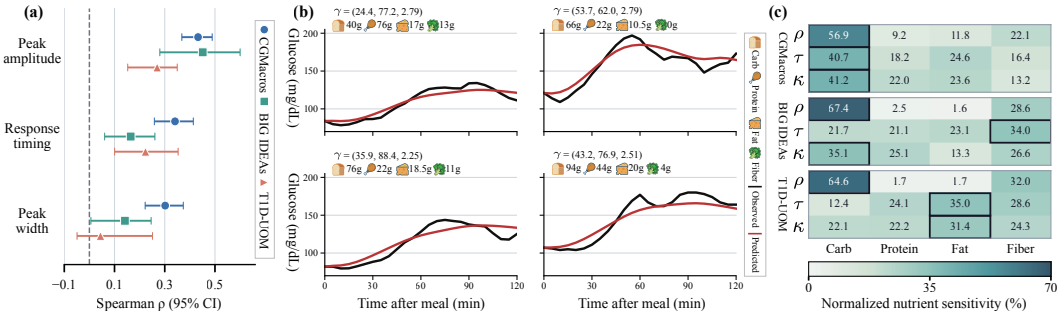

Figure 5: Nutrient-conditioned meal response. (a) Gamma–trajectory correlations. (b) Representative meals from one participant. (c) Nutrient sensitivity of the Gamma parameters.

Online personalization. To compare our proposed online personalization method with direct per-participant fine-tuning, we fine-tune MealGluco separately for each participant on the same causal completed-meal histories. Figure 6(a) shows that the two methods achieve comparable *Trajectory-MAE*, with complementary strengths across datasets: our personalized method has lower error on CGMacros, per-participant fine-tuning has a small advantage on T1D-UOM, and the difference is close to zero on BIG IDEAs. Figure 6(b) compares the cost of each personalization update. Our method updates only the two personalization parameters without backpropagation and requires 583–982× fewer FLOPs per completed meal and 277–4,034× less persistent participant-specific state than fine-tuning. Figure 6(c) then shows how performance changes on CGMacros as completed meals accumulate. Once four to six historical meals are available, all five errors fall below the population reference, and the improvement persists as more meals are assimilated.

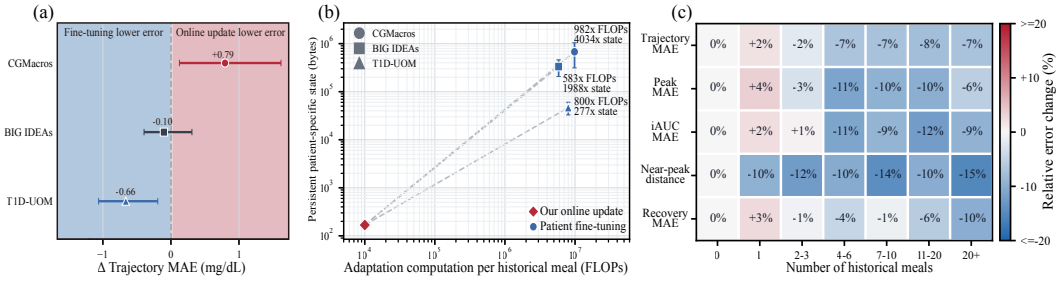

Figure 6: Online personalization versus per-participant fine-tuning. (a) Trajectory-MAE difference. (b) Cost per personalization update. (c) Error change with accumulated history on CGMacros.

Component ablations. The component ablations confirm that every proposed module contributes to the model's overall forecasting performance. Table 2 reports selected ablations on CGMacros, while the complete dataset-specific results are reported in Appendix Tables 12–14. Across all three datasets, removing any proposed module leads to a clear performance drop. The state-response and short-horizon-statistics ablations indicate that the history-dependent pathway and recent-history descriptors capture background and short-term glucose trends that the meal branch alone cannot represent. The unstructured meal-response head and nutrient-modifier ablations show that accurate forecasts rely on both the structured Gamma parameterization and nutrient-specific modulation. Finally, removing the personalized module increases prediction error, suggesting that the two participant-level coordinates effectively capture individual differences in postprandial glucose responses.

Table 2: Selected component ablations on CGMacros.

Configuration	Trajectory-MAE	iAUC-MAE	Peak-MAE	Near-Peak Distance	Recovery-MAE
MealGluco	**21.16 ± 0.17**	**1701.7 ± 21.3**	**25.00 ± 0.32**	17.54 ± 0.33	19.18 ± 0.32
w/o short-horizon statistics	23.35 ± 0.04	1934.0 ± 4.0	27.69 ± 0.22	19.72 ± 0.69	20.45 ± 0.20
w/o state response	23.54 ± 0.11	1951.0 ± 16.4	28.58 ± 0.28	18.01 ± 0.09	**19.04 ± 0.11**
w/o meal response	25.18 ± 0.19	2177.7 ± 13.5	36.45 ± 0.12	26.54 ± 0.25	24.20 ± 0.14
Unstructured meal-response head	23.04 ± 0.24	1901.6 ± 14.7	26.65 ± 0.27	20.25 ± 1.01	20.11 ± 0.54
w/o nutrient modifiers	22.75 ± 0.16	1887.4 ± 20.7	26.95 ± 0.40	18.91 ± 0.46	19.99 ± 0.19
w/o personalized module	22.63 ± 0.23	1873.1 ± 20.0	26.96 ± 0.08	20.08 ± 0.29	20.48 ± 0.21
MAE point loss	23.76 ± 0.20	1954.7 ± 18.7	26.63 ± 0.20	17.69 ± 0.62	19.31 ± 0.35

Cross-dataset generalization. We further test whether MealGluco transfers across datasets with different CGM devices and participant populations without target-specific parameter training or hyperparameter tuning. Each model is trained on one dataset and evaluated directly on the other two; for MealGluco, the two participant-specific coordinates may still be updated causally from completed target-participant meals, while the *w/o pers.* variant and all other baselines use fixed source-trained parameters. MealGluco achieves the best overall performance across the six cross-dataset transfers. Table 3 shows that *Trajectory-MAE* is reduced by 1.8–7.9% relative to the strongest baseline in every transfer. Overall, these results demonstrate that MealGluco generalizes well across datasets: its design models nutrient-induced response dynamics separately from the pre-meal glucose state, while the personalization module captures differences in glucose responses across participants. The remaining four metrics are reported in Appendix Tables 16–19.

Table 3: Trajectory-MAE for cross-dataset generalization. *MealGluco w/o pers.* disables the personalization module and uses no historical meal records.

Trained on	CGMacros		BIG IDEAs		T1D-UOM	
Evaluated on	BIG IDEAs	T1D-UOM	CGMacros	T1D-UOM	CGMacros	BIG IDEAs
MealGluco	**17.28 ± 0.46**	**30.48 ± 0.10**	**24.73 ± 0.75**	**30.54 ± 0.35**	**24.43 ± 0.64**	**19.87 ± 0.81**
MealGluco w/o pers.	18.44 ± 0.97	31.07 ± 0.33	26.64 ± 0.44	31.11 ± 0.23	26.58 ± 0.33	21.30 ± 0.81
UniCA-Gluformer	18.99 ± 1.06	34.89 ± 0.83	25.36 ± 0.64	35.25 ± 0.26	27.57 ± 0.47	21.76 ± 1.06
Stacking Framework	19.84 ± 0.02	31.17 ± 0.02	30.17 ± 0.05	38.75 ± 0.03	26.52 ± 0.03	21.37 ± 0.04

5 Conclusion

MealGluco predicts complete meal-conditioned postprandial trajectories by combining a nutrient-conditioned Gamma response with a bounded pre-meal state pathway. Two participant-level coordinates are updated online from completed meals, enabling progressively personalized forecasts without retraining the shared model. Across three public datasets and five metrics, MealGluco ranks first or second in all 15 dataset–metric comparisons, with a mean rank of 1.33 among the seven evaluated models. Online personalization reduces trajectory error relative to the population model while matching participant-specific fine-tuning at substantially lower update and storage cost. In cross-dataset evaluation, MealGluco achieves the strongest overall generalization, obtaining the lowest trajectory error in all six transfers and reducing MAE by 1.8–7.9% relative to the strongest baseline. Component ablations further show that removing any proposed module degrades overall forecasting performance. Together, these results support MealGluco as a practical foundation for personalized postprandial glucose modeling and nutrient–response analysis from CGM and meal records.

AI USE STATEMENT

We used generative AI tools to assist with language polishing, LaTeX formatting and table preparation, and selected coding tasks. The core research ideas, methodological contributions, and scientific claims were developed by the authors without relying on generative AI. The majority of the writing and coding was completed by the authors. All AI-assisted outputs were reviewed, verified, and tested by the authors, who take responsibility for the final content of this work, including its text, code, analyses, and other artifacts.

ETHICS STATEMENT

This study uses publicly available, de-identified continuous glucose monitoring datasets and does not involve new human-subject data collection or the release of personally identifiable information. The model is intended for research on postprandial glucose trajectory prediction and is not intended to replace clinical judgment.

REPRODUCIBILITY STATEMENT

The complete model formulation and training objective are presented in Section 3 and Appendix A. Optimization settings, checkpoint selection, covariance fitting, and validation procedures are specified in Appendix B.4. Dataset construction, baseline details, evaluation metrics, and the participant-disjoint validation protocol are described in Appendix B. Additional results and controls are reported in Appendix C. All datasets are publicly available, with access links provided in Appendix B.1. Preprocessing, model selection, and covariance estimation are fold-local, and all reported predictions are out-of-fold. Code is available at the anonymous repository `https://anonymous.4open.science/r/MealGluco`.

REFERENCES

David J Albers, Matthew Levine, Bruce Gluckman, Henry Ginsberg, George Hripcsak, and Lena Mamykina. Personalized glucose forecasting for type 2 diabetes using data assimilation. *PLoS computational biology*, 13(4):e1005232, 2017.

Ashwaq Alsuhaymi, Ahmad Bilal, Daniel Gasca García, Rujiravee Kongdee, Nicole Lubasinski, Hood Thabit, Paul W Nutter, and Simon Harper. A longitudinal multimodal dataset of type 1 diabetes. *Scientific Data*, 12(1):1379, 2025.

Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

Sarah E Berry, Ana M Valdes, David A Drew, Francesco Asnicar, Mohsen Mazidi, Jonathan Wolf, Joan Capdevila, George Hadjigeorgiou, Richard Davies, Haya Al Khatib, et al. Human postprandial responses to food and potential for precision nutrition. *Nature medicine*, 26(6):964–973, 2020.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006. ISBN 978-0-387-31073-2.

Peter Cho, Juseong Kim, Brinnae Bent, and Jessilyn Dunn. BIG IDEAs Lab Glycemic Variability and Wearable Device Data, September 2023. URL `https://doi.org/10.13026/zthx-5212`. Version 1.1.2.

Claudio Cobelli, Chiara Dalla Man, Gianna Toffolo, Rita Basu, Adrian Vella, and Robert Rizza. The oral minimal model method. *Diabetes*, 63(4):1203–1213, 2014.

Chiara Dalla Man, Robert A Rizza, and Claudio Cobelli. Meal simulation model of the glucose-insulin system. *IEEE Transactions on biomedical engineering*, 54(10):1740–1749, 2007.

Anurag Das, David Kerr, Namino Glantz, Wendy Bevier, Rony Santiago, Ricardo Gutierrez-Osuna, and Bobak J Mortazavi. Cgmacros: a pilot scientific dataset for personalized nutrition and diet monitoring. *Scientific Data*, 12(1):1557, 2025.

Claudia Eberle and Christoph Ament. Real-time state estimation and long-term model adaptation: a two-sided approach toward personalized diagnosis of glucose and insulin levels. *Journal of diabetes science and technology*, 6(5):1148–1158, 2012.

Gary H Glover. Deconvolution of impulse response in event-related bold fmri1. *Neuroimage*, 9(4): 416–429, 1999.

Lu Han, Yu Liu, Lan Li, Qiwen Deng, Jian Jiang, Yinbo sun, Zhe Yu, Binfeng Wang, Xingyu Lu, Lintao Ma, Han-Jia Ye, and De-Chuan Zhan. Unica: Unified covariate adaptation for time series foundation model. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 6403–6443, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/0b5eb45a22ff33956c043dd271f244ea-Paper-Conference.pdf.

Benjamin Hentze, Thomas Muders, Henning Luepschen, Steffen Leonhardt, Christian Putensen, and Marian Walter. Gamma-variate modeling of indicator dilution curves in electrical impedance tomography. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3596–3599. IEEE, 2017.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*, pp. 492–518. Springer, 1992.

Fatma Indriani, Mohammad Reza Faisal, and Naufal Said. Improving postprandial glucose forecasting using diagnosis-aware stacked learning. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, pp. 349–362, 2026.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Marc C Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

Ning Li, Jianyong Tuo, Youqing Wang, and Menghui Wang. Prediction of blood glucose concentration for type 1 diabetes based on echo state networks embedded with incremental learning. *Neurocomputing*, 378:248–259, 2020.

Zechen Li, Keerthana Natarajan, Weizhi Zhang, Menglian Zhou, Simon A Lee, Yuwei Zhang, Maxwell A Xu, Zeinab Esmaeilpour, Flora D Salim, Mark Malhotra, et al. Glucofm: A dual-stream foundation model for continuous glucose monitoring. *arXiv preprint arXiv:2605.30865*, 2026.

Zilu Liang. Within-person modeling of postprandial glucose using multimodal wearable data. *Frontiers in Digital Health*, 8:1847884, 2026.

Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, 37(4): 1748–1764, 2021.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Yurun Lu, Dan Liu, Zhongming Liang, Rui Liu, Pei Chen, Yitong Liu, Jiachen Li, Zhanying Feng, Lei M Li, Bin Sheng, et al. A pretrained transformer model for decoding individual glucose dynamics from continuous glucose monitoring data. *National Science Review*, 12(5):nwaf039, 2025.

Helena Mendes-Soares, Tali Raveh-Sadka, Shahar Azulay, Kim Edens, Yatir Ben-Shlomo, Yossi Cohen, Tal Ofek, Davidi Bachrach, Josh Stevens, Dorin Colibaseanu, et al. Assessment of a personalized approach to predicting postprandial glycemic responses to food among individuals without diabetes. *JAMA network open*, 2(2):e188102, 2019.

Hada Melino Muhammad, Zechen Li, Flora Salim, and Ahmed A Metwally. Cgm-jepa: Learning consistent continuous glucose monitor representations via predictive self-supervised pretraining. *arXiv preprint arXiv:2605.00933*, 2026.

Jaideep Pathak, Alexander Wikner, Rebeckah Fussell, Sarthak Chandra, Brian R Hunt, Michelle Girvan, and Edward Ott. Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos: An interdisciplinary journal of nonlinear science*, 28(4), 2018.

Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. volume 31, 2018.

Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

Radojka M Savic, Daniël M Jonker, Thomas Kerbusch, and Mats O Karlsson. Implementation of a transit compartment model for describing drug absorption in pharmacokinetic studies. *Journal of pharmacokinetics and pharmacodynamics*, 34(5):711–726, 2007.

Renat Sergazinov, Mohammadreza Armandpour, and Irina Gaynanova. Gluformer: Transformer-based personalized glucose forecasting with uncertainty quantification. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

Renat Sergazinov, Elizabeth Chun, Valeriya Rogovchenko, Nathaniel Fernandes, Nicholas Kasman, and Irina Gaynanova. Glucobench: Curated list of continuous glucose monitoring datasets with prediction benchmarks. In *International Conference on Learning Representations*, volume 2024, pp. 17604–17631, 2024.

Howard K Thompson Jr, C Frank Starmer, Robert E Whalen, and Henry D McIntosh. Indicator transit time considered as a gamma variate. *Circulation research*, 14(6):502–515, 1964.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. volume 30, 2017.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, and Mingsheng Long. Timexer: Empowering transformers for time series forecasting with exogenous variables. volume 37, pp. 469–498, 2024.

David Zeevi, Tal Korem, Niv Zmora, David Israeli, Daphna Rothschild, Adina Weinberger, Orly Ben-Yacov, Dar Lador, Tali Avnit-Sagi, Maya Lotan-Pompan, et al. Personalized nutrition by prediction of glycemic responses. *Cell*, 163(5):1079–1094, 2015.

Taiyu Zhu, Kezhi Li, Pau Herrero, and Pantelis Georgiou. Personalized blood glucose prediction for type 1 diabetes using evidential deep learning and meta-learning. *IEEE Transactions on Biomedical Engineering*, 70(1):193–204, 2022.

APPENDIX

APPENDIX CONTENTS

- A Methods
  - A.1 Notation and Data Interface
  - A.2 Symbol Table
  - A.3 CGM Encoder Details
  - A.4 Nutrient Encoder Details
  - A.5 Fusion Module Details
  - A.6 Personalized Module Details
  - A.7 Training Objective Details
- B Experimental Design
  - B.1 Datasets
  - B.2 Baselines
  - B.3 Evaluation Metrics
  - B.4 Experimental Design, Optimization, and Validation Protocol
  - B.5 Ablation Definitions
- C Additional Results
  - C.1 Additional Main Results
  - C.2 Online Personalization
  - C.3 Component Ablations
  - C.4 Additional Controls
  - C.5 Cross-Dataset Generalization
  - C.6 Statistical Tests
    - C.6.1 Main Comparisons
    - C.6.2 Online Personalization
    - C.6.3 Cross-Dataset Generalization
    - C.6.4 Component Ablations

A METHODS

A.1 NOTATION AND DATA INTERFACE

Feature and trajectory vectors are columns; the forecast horizon is $H = 24$; sequence matrices store one transposed feature vector per row. Thus $\mathbf{h}_\ell = \mathbf{H}[\ell, :]^\top$ and $\mathbf{q}_h^{(v)} = \mathbf{Q}^{(v)}[h, :]^\top$ for decoder stage $v \in \{0, 1, 2\}$. The indices i, m, m', and h denote participant, the meal being predicted, a completed historical meal with $0 \le m' < m$, and the relative sample index, respectively; as in the main text, $h = -(H-1), \ldots, 0$ for the pre-meal history and $h = 1, \ldots, H$ for the post-meal forecast. Concatenation $\|$ stacks column vectors; Pool averages token rows and returns a column vector; LN normalizes each feature vector. We write $\sigma(a) = (1 + e^{-a})^{-1}$ for the sigmoid, with scalar nonlinearities applied elementwise to vectors. The subscript H in the Huber loss denotes its name. I_d and $\mathbf{1}_d$ denote the d-dimensional identity matrix and all-ones vector.

Meal clock time $t_{i,m}$ is an observed date-and-time value, and $\phi_{i,m} = 2\pi t_{i,m}/1440$ is its periodic coordinate in radians; on the post-meal grid, $t_h = t_0 + h\Delta t$ is the sampled time after the current meal. We write $\mathbf{G}^{\mathrm{pre}}$ and $\mathbf{G}^{\mathrm{post}}$ for the pre- and post-meal CGM vectors, $G(t_0)$ for the mealtime anchor, and $\widehat{\mathbf{G}}^{\mathrm{post}}$ for the forecast vector. F_Θ denotes the full predictor, Θ its shared parameters (including frozen encoder parameters), and $\mathcal{F}_{m'}(\mathbf{u}; \Theta)$ the same predictor with the inputs of a completed historical meal m' fixed. During online adaptation, only the participant posterior changes: in the personalization update of Appendix A.6, m' indexes the completed historical meal being assimilated and m the meal being predicted.

For each meal, the two-hour history contains 24 finite CGM samples at 5-min intervals from -115 to 0 min:

$$\mathbf{G}^{\mathrm{pre}} = [G(-115), G(-110), \ldots, G(0)]^\top \in \mathbb{R}^{24},$$

and the future trajectory is $\mathbf{G}^{\mathrm{post}} = [G(5), \ldots, G(120)]^\top \in \mathbb{R}^H$ with $H = 24$. The generic interface carries a target-availability mask $\mathbf{a} \in \{0,1\}^H$; unavailable values are excluded from losses and metrics rather than set to zero. The locked main-paper datasets use complete 24-point postprandial trajectories for training and primary evaluation.

Nutrient doses are measured in grams. The nonnegative nutrient vector is $\mathbf{n} = [n^{\mathrm{c}}, n^{\mathrm{p}}, n^{\mathrm{f}}, n^{\mathrm{fi}}]^\top$. Before the logarithm, each reported nutrient dose is clipped to the nonnegative domain. For nutrient type $\chi \in \{\mathrm{c}, \mathrm{p}, \mathrm{f}, \mathrm{fi}\}$, fixed nutrient scales s_χ define

$$\widetilde{n}^{\chi} = \frac{\log(1 + n^{\chi})}{\log(1 + s_{\chi})}, \qquad (s_{\mathrm{c}}, s_{\mathrm{p}}, s_{\mathrm{f}}, s_{\mathrm{fi}}) = (100, 80, 80, 30).$$

The transformation preserves $n^{\chi} = 0 \Rightarrow \widetilde{n}^{\chi} = 0$ and does not clip values above one.

A.2 Symbol Table

Symbol	Meaning
i	Participant index.
m	Index of the meal being predicted.
m'	Index of a completed historical meal, with $0 \le m' < m$, used for personalization.
t	Time relative to the current meal; $t_0 = 0$ is the mealtime anchor.
t_h	Post-meal time grid for $h = 1, \ldots, H$.
$\mathcal{T}_{\mathrm{pre}}, \mathcal{T}_{\mathrm{post}}$	Discrete pre- and post-meal time grids; t_{pre} is the earliest pre-meal time.
H, T	Number of forecast points and forecast horizon.
$G_{i,m}(t)$	CGM value for participant i at time t relative to meal m.
$\mathbf{G}^{\mathrm{pre}}_{i,m}, \mathbf{G}^{\mathrm{post}}_{i,m}$	Pre- and post-meal CGM vectors.
$\widehat{\mathbf{G}}^{\mathrm{post}}_{i,m}$	Predicted postprandial CGM trajectory.
$G_{i,m}(0)$	Observed mealtime anchor.
$G^{\mathrm{fast}}_{i,m}$	Fasting-glucose input (measured or proxy).
$t_{0,i,m}$	Clock time of the meal.
$\mathbf{n}_{i,m}$	Nutrient-dose vector for carbohydrate, protein, fat, and fiber.
$\mathbf{z}_{i,m}$	Pre-meal state produced by the CGM encoder.
$\mathbf{H}_{i,m}$	Encoded temporal-token matrix.
E_ψ, E_z, F_Q	Frozen CGM encoder and trainable state encoder, future decoder, state
F_s, E_γ	head, and nutrient encoder.
$\rho_{i,m}, \tau_{i,m}, \kappa_{i,m}$	Gamma amplitude, peak time, and shape parameters.
$K(t; \tau, \kappa)$	Peak-normalized Gamma kernel.
$\mathbf{R}^{\mathrm{meal}}_{i,m}$	Sampled meal-response sequence.
$\mathbf{R}^{\mathrm{state}}_{i,m}$	Bounded state-response sequence.
$R_{\max}$	Residual bound of the state pathway.
$\mathbf{U}_i, \mathbf{u}_i$	Prior and realization of the participant personalization state.
$u_{\rho,i}, u_{r,i}$	Personalization parameters controlling response gain and post-peak clearance.
g_ρ, C_r, q	Bounded gain, clearance factor, and post-peak ramp.
ϵ_t	Guard interval used in the post-peak ramp.
$\mathcal{F}_{m'}(\mathbf{u})$	Personalized trajectory for historical meal m'.
$J_{m'}$	Jacobian of the personalized trajectory with respect to the personalization state.
$\boldsymbol{\varepsilon}_{m'}$	Observation residual for historical meal m'.

Symbol	Meaning
$\boldsymbol{\mu}_i, \Sigma_{u,i}, \Lambda_i$	Participant posterior mean, covariance, and precision.
$\mathbf{a}_{m'}, \widetilde{\mathbf{G}}^{\text{post}}_{m'}$	Linearization intercept and residual observation for historical meal m'.
$W_\varepsilon, \Sigma_\varepsilon$	Residual precision and covariance.
$\ell_{\text{H}}, w_h, \lambda_{\text{peak}}$	Huber loss, temporal weight, and peak-loss weight.
$\mathcal{L}_{\text{point}}, \mathcal{L}_{\text{peak}}, \mathcal{L}_{\text{final}}$	Point objective, peak objective, and final training objective.
$e_{i,m,h}, a_{i,m,h}$	Pointwise prediction error and target-availability mask.
$\mathbf{n}^\chi$	Nutrient dose for nutrient type $\chi \in \{\text{c}, \text{p}, \text{f}, \text{fi}\}$.
$\mathcal{M}, \mathcal{M}_d$	Qualifying meals and meals with a discernible excursion.
$A_{i,m}$	Positive incremental area under the curve for meal (i, m).
$D_{i,m}, \widehat{D}_{i,m}$	True and predicted post-peak decline for meal (i, m).
$\text{MAE}_{\text{traj}}, \text{MAE}_{\text{iAUC}}, \text{MAE}_{\text{peak}}, \text{MAE}_{\text{rec}}$	Main evaluation errors.
$d_{i,m}$	Near-peak timing distance.

A.3 CGM Encoder Details

We use the public CGM-JEPA checkpoint (Muhammad et al., 2026) from `https://huggingface.co/CRUISEResearchGroup/CGM-JEPA`, subfolder `cgm_jepa`. Its encoder has three transformer layers, embedding width 96, six attention heads, MLP expansion ratio 4, convolutional embedding kernel size 3, and zero encoder dropout. The six encoder heads are distinct from the four heads in MealGluco's trainable decoder. The upstream release describes pretraining on 228 participants from the Stanford and Colas datasets; we use those released weights without additional encoder pretraining.

The trainable components in this subsection correspond to E_z, F_Q, and F_s in the main text.

CGM values in mg/dL are reshaped into two nonoverlapping 12-sample patches, in chronological order. The encoder receives these raw-valued patches without time marks; its two output tokens are retained. The encoder is kept in evaluation mode with gradient tracking disabled:

$$\mathbf{H} = E_\psi(\mathbf{G}^{\text{pre}}) \in \mathbb{R}^{L \times D}, \qquad L = 2, \quad D = 96.$$

We compute the four recent-CGM anchors used in the main paper as

$$c_{\text{last}} = \frac{G(0) - 120}{50}, \qquad c_{\text{mean}} = \frac{\frac{1}{6}\sum_{r=0}^{5} G(-5r) - 120}{50},$$

$$c_{\text{slope}} = \frac{G(0) - G(-25)}{50}, \qquad c_{\text{std}} = \frac{\text{std}(G(-25), \ldots, G(0))}{25}.$$

Here std is the population standard deviation of the six samples. These constants are fixed engineering scales. Let $\mathbf{c}(\mathbf{G}^{\text{pre}}) = [c_{\text{last}}, c_{\text{mean}}, c_{\text{slope}}, c_{\text{std}}]^\top$. The state encoder receives

$$\mathbf{h}_x = \left[\frac{\mathbf{h}_1 + \mathbf{h}_2}{2} \parallel \mathbf{c}(\mathbf{G}^{\text{pre}})\right] \in \mathbb{R}^{D+4}$$

and produces the state vector

$$\mathbf{z} = W_{\text{state},2}\,\text{LN}\left[\text{GELU}(W_{\text{state},1}\mathbf{h}_x + \mathbf{b}_{\text{state},1})\right] + \mathbf{b}_{\text{state},2} \in \mathbb{R}^D,$$

with $W_{\text{state},1} \in \mathbb{R}^{D \times (D+4)}$ and $W_{\text{state},2} \in \mathbb{R}^{D \times D}$. The token matrix $\mathbf{H}$ is retained as the key–value sequence for the decoder cross-attention: each future query can assign weights to the two encoded history tokens and retrieve their content. The vector $\mathbf{z}$ is the global pre-meal-state summary produced by the state encoder; it is sent to the nutrient parameter head and added to every future

query, whereas the relative-time and clock-time projections vary with h. The CGM-JEPA parameters ψ are frozen, whereas the state-encoder parameters ω are optimized.

A.3.1 TIME QUERIES AND STATE-RESPONSE DECODER

For each $h \in \{1, \dots, H\}$, relative post-meal time uses $\bar{t}_h = t_h/T = h/H$ and $\zeta_h = 2\bar{t}_h - 1$:

$$P_1(\zeta_h) = \zeta_h, \qquad P_2(\zeta_h) = \frac{1}{2}(3\zeta_h^2 - 1), \qquad \mathbf{b}_h^{\text{time}} = [P_1(\zeta_h), P_2(\zeta_h)]^\top \in \mathbb{R}^2.$$

The relative post-meal time projection uses $W_t \in \mathbb{R}^{D\times 2}$ and bias $\mathbf{b}_t \in \mathbb{R}^D$. Absolute Clock Time (Clock Time hereafter) represents the time of day with a 24-hour sine–cosine pair. We use the effective projection $W_c = \sigma(\gamma_{\text{abs}})\widetilde{W}_c$, where the learned scalar gate is initialized at $\gamma_{\text{abs}} = -1$ and the learned matrix $\widetilde{W}_c \in \mathbb{R}^{D\times 2}$ at zero:

$$\mathbf{c}_h^{\text{abs}} = \left[\sin\left(\phi_i + 2\pi\frac{t_h}{1440}\right), \cos\left(\phi_i + 2\pi\frac{t_h}{1440}\right)\right]^\top \in \mathbb{R}^2, \qquad W_c \in \mathbb{R}^{D\times 2}.$$

The two time embeddings and initial future queries are

$$\mathbf{T}_{\text{rel}}[h,:]^\top = W_t\mathbf{b}_h^{\text{time}} + \mathbf{b}_t, \qquad \mathbf{T}_{\text{abs}}[h,:]^\top = W_c\mathbf{c}_h^{\text{abs}}, \qquad \mathbf{T}_{\text{rel}}, \mathbf{T}_{\text{abs}} \in \mathbb{R}^{H\times D},$$

$$\mathbf{q}_h^{(0)} = \mathbf{T}_{\text{rel}}[h,:]^\top + \mathbf{z} + \mathbf{T}_{\text{abs}}[h,:]^\top, \qquad \mathbf{Q}^{(0)} \in \mathbb{R}^{H\times D} = \mathbb{R}^{24\times 96}.$$

Multi-head cross-attention uses $N_{\text{head}} = 4$ heads, each of dimension $D_{\text{head}} = D/N_{\text{head}} = 24$, with

$$\mathbf{Q} = \mathbf{Q}^{(0)} \in \mathbb{R}^{24\times 96}, \qquad \mathbf{K} = \mathbf{V} = \mathbf{H} \in \mathbb{R}^{2\times 96}.$$

For attention head $a \in \{1, \dots, N_{\text{head}}\}$, $\mathbf{Q}_a$, $\mathbf{K}_a$, and $\mathbf{V}_a$ are separate affine projections of $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ onto D_{head} features. Thus $\mathbf{Q}_a \in \mathbb{R}^{H\times D_{\text{head}}}$ and $\mathbf{K}_a, \mathbf{V}_a \in \mathbb{R}^{L\times D_{\text{head}}}$. The scores and token-wise weights are

$$\mathbf{S}_a = \frac{\mathbf{Q}_a\mathbf{K}_a^\top}{\sqrt{D_{\text{head}}}} \in \mathbb{R}^{24\times 2}, \qquad \mathbf{A}_a = \text{softmax}_{\text{token}}(\mathbf{S}_a).$$

During training, dropout is applied to $\mathbf{A}_a$ before multiplication by $\mathbf{V}_a$. Concatenating the head outputs and applying an affine output projection defines MHA. The decoder then performs

$$\mathbf{Q}^{(1)} = \text{LN}_1\left[\mathbf{Q}^{(0)} + \text{MHA}(\mathbf{Q}^{(0)}, \mathbf{H}, \mathbf{H})\right].$$

At each position, the shared feed-forward network has hidden dimension 192:

$$\text{FFN}(\mathbf{q}_h^{(1)}) = W_{\text{ff},2}\,\text{Dropout}\left(\text{GELU}(W_{\text{ff},1}\mathbf{q}_h^{(1)} + \mathbf{b}_{\text{ff},1})\right) + \mathbf{b}_{\text{ff},2},$$

$$W_{\text{ff},1} \in \mathbb{R}^{192\times 96}, \qquad W_{\text{ff},2} \in \mathbb{R}^{96\times 192}, \qquad \mathbf{q}_h^{(2)} = \text{LN}_2\left[\mathbf{q}_h^{(1)} + \text{FFN}(\mathbf{q}_h^{(1)})\right].$$

The scalar residual head is

$$s_h = \mathbf{w}_\vartheta^\top\mathbf{q}_h^{(2)} + b_\vartheta, \qquad R^{\text{state}}(t_h) = R_{\max}\tanh(s_h).$$

The residual-head weights and bias are initialized to zero. Thus, optimization starts from $G(0) + R^{\text{meal}}$ and subsequently learns bounded corrections, while the residual branch remains nutrient-independent by construction. We set $R_{\max} = 20$ mg/dL because a sweep on CGMacros over $\{10, 15, 20, 25, 30\}$ under the same evaluation protocol as the main experiments gave the lowest mean *Trajectory-MAE*, *iAUC-MAE*, and *Peak-MAE* at this value (Appendix B.4.2).

A.4 NUTRIENT ENCODER DETAILS

The main paper summarizes the full nutrient-to-response mapping as one Gamma generator

$$\boldsymbol{\gamma} = E_\gamma(\mathbf{z}, \widetilde{\mathbf{n}}, \widetilde{G}^{\text{fast}}) = [\rho, \tau, \kappa]^\top,$$

where ρ, τ, and κ are the response amplitude, peak time, and shape. This generator corresponds to E_γ in the main text. Concretely, E_γ composes (i) the nutrient head H_θ mapping $[\mathbf{z}\|\widetilde{\mathbf{n}}]$ to logits $\mathbf{r}^{(0)}$, (ii) a bounded fasting-glucose offset followed by saturation, (iii) the base-coordinate maps of the next subsubsections, and (iv) the explicit protein/fat/fiber modifiers. The following subsubsections give the exact layers, bounds, and initialization; the guarantee $n^{\text{c}} = 0 \Rightarrow \rho = 0$ follows from $\rho^{\text{base}} \propto \widetilde{n}^{\text{c}}$.

A.4.1 PARAMETER HEAD AND FASTING CONDITIONING

The nutrient parameter head H_θ operates on

$$\mathbf{h} = [\mathbf{z} \parallel \widetilde{\mathbf{n}}] \in \mathbb{R}^{D+4} = \mathbb{R}^{100}.$$

It is a two-layer MLP with hidden width 68:

$$\mathbf{v} = \text{LN}_{68}\left[\text{GELU}(W_{\text{nut},1}\mathbf{h} + \mathbf{b}_{\text{nut},1})\right] \in \mathbb{R}^{68},$$

$$W_{\text{nut},1} \in \mathbb{R}^{68\times 100}, \quad \mathbf{b}_{\text{nut},1} \in \mathbb{R}^{68}, \quad \mathbf{r}^{(0)} = W_{\text{nut},2}\mathbf{v} + \mathbf{b}_{\text{nut},2} \in \mathbb{R}^3,$$

$$W_{\text{nut},2} \in \mathbb{R}^{3\times 68}, \qquad \mathbf{b}_{\text{nut},2} \in \mathbb{R}^3, \qquad \mathbf{r}^{(0)} = [r_\rho, r_\tau, r_\kappa]^\top.$$

The nutrient output layer is initialized to zero. The independent fasting map is

$$F_{\text{fast}}(\widetilde{G}^{\text{fast}}) = W_{\text{fast},2}\,\text{LN}_{16}\Big[\text{GELU}(W_{\text{fast},1}\widetilde{G}^{\text{fast}} + \mathbf{b}_{\text{fast},1})\Big] + \mathbf{b}_{\text{fast},2},$$

where $W_{\text{fast},1} \in \mathbb{R}^{16\times 1}$, $W_{\text{fast},2} \in \mathbb{R}^{3\times 16}$, and the biases have dimensions 16 and 3. Its output weight and bias are also initialized to zero, so fasting conditioning initially contributes no offset. It produces

$$\Delta\mathbf{r}_{\text{fast}} = \sigma(\gamma_{\text{fast}})\tanh(F_{\text{fast}}(\widetilde{G}^{\text{fast}})), \qquad \gamma_{\text{fast}} \text{ initialized to } -1.$$

The conditionally shifted logits are bounded using

$$\mathbf{r}^{(1)} = \mathbf{r}^{(0)} + \Delta\mathbf{r}_{\text{fast}}, \qquad \bar{\mathbf{r}} = d_{\text{sat}}\tanh\left(\frac{\mathbf{r}^{(1)}}{d_{\text{sat}}}\right), \qquad d_{\text{sat}} = 2.$$

A.4.2 GAMMA PARAMETERIZATION AND MIXED-MEAL MODIFIERS

The base amplitude, peak time, and shape are

$$\widetilde{\rho}_{\max} = 75[1 + 0.5\tanh(\theta_\rho)], \qquad \rho^{\text{base}} = \widetilde{\rho}_{\max}\sigma(\bar{r}_\rho)\widetilde{n}^{\text{c}},$$

$$\tau^{\text{base}}_{\text{step}} = 1 + (H-1)\sigma(\bar{r}_\tau) \in (1, H), \qquad \tau^{\text{base}} = 5\tau^{\text{base}}_{\text{step}},$$

$$\kappa = 1 + 4\sigma(\bar{r}_\kappa) \in (1, 5).$$

$\widetilde{\rho}_{\max}$ is the amplitude scale for one normalized carbohydrate unit, and the meal amplitude scales with the normalized carbohydrate dose $\widetilde{n}^{\mathrm{c}}$. For $\mathcal{M} = \{\mathrm{p}, \mathrm{f}, \mathrm{fi}\}$, mixed-meal modifiers are

$$\alpha = \exp\left[-\sum_{\nu\in\mathcal{M}} \beta_\nu \widetilde{n}^\nu\right], \qquad \delta_{\mathrm{meal}} = 1 - \exp\left[-\sum_{\nu\in\mathcal{M}} \xi_\nu \widetilde{n}^\nu\right],$$

$$\beta_\nu = 1.5\sigma(b_\nu), \qquad \xi_\nu = 1.5\sigma(d_\nu),$$

$$\rho = \alpha\rho^{\mathrm{base}}, \qquad \tau = \tau^{\mathrm{base}} + (T - \tau^{\mathrm{base}})\delta_{\mathrm{meal}}, \qquad T = 120\,\mathrm{min}.$$

Thus, $\alpha \in (0, 1]$ decreases the explicit amplitude path and $\delta_{\mathrm{meal}} \in [0, 1)$ delays the explicit peak-time path. These explicit modifiers provide monotone directional control of the amplitude and peak-time paths, while the unconstrained parameter head adds a data-driven correction on top of them.

Following the Gamma-variate and transit-time parameterization (Thompson Jr et al., 1964; Glover, 1999; Savic et al., 2007; Hentze et al., 2017), we start from

$$f(t) \propto t^\kappa e^{-\kappa t/\tau},$$

whose global maximum is at $t = \tau$. Dividing $f(t)$ by $f(\tau)$ gives the peak-normalized kernel, which is evaluated at $t_h = 5h$:

$$K(t; \tau, \kappa) = \left(\frac{t}{\tau}\right)^\kappa \exp\left[\kappa\left(1 - \frac{t}{\tau}\right)\right], \qquad \mathbf{R}^{\mathrm{meal}} = [\rho K(t_h; \tau, \kappa)]_{h=1}^{H}.$$

The kernel is peak-normalized: it satisfies $K(\tau; \tau, \kappa) = 1$, and its log derivative

$$\frac{\partial}{\partial t}\log K(t; \tau, \kappa) = \kappa\left(\frac{1}{t} - \frac{1}{\tau}\right)$$

has the unique interior stationary point $t = \tau$, with second derivative $-\kappa/\tau^2 < 0$. Hence ρ and τ describe the continuous-time meal component: ρ is its maximum and τ its peak time. Sampling on the 5-min grid can place the observed maximum below ρ. When $n^{\mathrm{c}} = 0$, the meal component is identically zero.

A.4.3 Initialization and structural interpretation

All parameters of F_Θ are shared across meals and participants. We initialize the calibration logit at $\theta_\rho = 0$. Modifier logits are chosen by $b_\nu = \mathrm{logit}(\beta_\nu/1.5)$ and $d_\nu = \mathrm{logit}(\xi_\nu/1.5)$, where $\mathrm{logit}(p) = \log[p/(1-p)]$.

The remaining initial values are $(\beta_p, \beta_f, \beta_{\mathrm{fi}}) = (0.05, 0.05, 0.80)$, $(\xi_p, \xi_f, \xi_{\mathrm{fi}}) = (0.02, 0.04, 0.05)$, $\widetilde{\rho}_{\max} = 75$ mg/dL, and $\gamma_{\mathrm{fast}} = -1$, with the output weights and biases of the nutrient and fasting branches initialized at zero.

These are learned starting values. The constrained maps enforce a smooth nonnegative event, a carbohydrate zero gate, and bounded explicit nutrient modifiers. They specify an inductive bias for predicting observed trajectories; the coordinates are learned response parameters that control amplitude, peak timing, and clearance. At $t = 0$, the kernel extends continuously as $K(0; \tau, \kappa) = 0$. For $t > 0$ it is smooth and unimodal when $\rho > 0$; the sum with the state response can have more than one local maximum.

A.5 Fusion Module Details

A.5.1 Population fusion

The Fusion Module combines the observed anchor and both response sequences:

$$\widehat{\mathbf{G}}^{\mathrm{pop}} = G(0)\mathbf{1}_H + \mathbf{R}^{\mathrm{meal}} + \mathbf{R}^{\mathrm{state}}.$$

The sum yields the 24-point population forecast. The meal branch depends on $(\mathbf{G}^{\mathrm{pre}}, \mathbf{n}, G^{\mathrm{fast}})$ through its inputs and the state encoder, whereas the state branch depends on $(\mathbf{G}^{\mathrm{pre}}, \phi_i)$. All future queries are evaluated together using the observed history. The forecast contains no autoregressive dependence on future CGM labels.

Training supervises the sum of the branches, and the input restrictions of each pathway make the current-meal conditioning path explicit in the computation graph. The 20 mg/dL bound applies to the population state component; subsequent personalized clearance scales the combined offset. The observed anchor is used at $t = 0$; the learned state residual is defined on the forecast grid starting at 5 min.

A.5.2 Personalized fusion

The personalized forecast is

$$\widehat{\mathbf{G}}_i^{\mathrm{post}}(t) = G(0) + \left[g_\rho(u_{\rho,i})R^{\mathrm{meal}}(t) + R^{\mathrm{state}}(t)\right] C_r(t; u_{r,i}, \tau).$$

A.6 Personalized Module Details

A.6.1 Individualized transform

For participant i, the latent state is $\mathbf{U}_i = [U_{\rho,i}, U_{r,i}]^\top \sim \mathcal{N}(\mathbf{0}, I_2)$. Given realization $\mathbf{u}_i$, define

$$g_\rho(u_{\rho,i}) = \exp[\log(2)\tanh(u_{\rho,i})],$$

$$q(t;\tau) = \mathrm{clip}\left(\frac{t-\tau}{\max(T-\tau, \epsilon_t)}, 0, 1\right),$$

$$C_r(t; u_{r,i}, \tau) = \exp[-\log(2)\tanh(u_{r,i})q(t;\tau)].$$

We set $\epsilon_t = 5$ min; this guard interval prevents an almost-zero post-peak interval from magnifying the clearance effect when τ approaches the terminal horizon. Positive $u_{r,i}$ contracts the post-peak deviation toward $G(0)$; negative values expand it. Positive $u_{\rho,i}$ increases only the meal-component gain. These coordinates are learned response parameters that modulate the meal component and the post-peak clearance. Neither latent variable changes the frozen population weights.

A.6.2 Residual covariance and local likelihood

For a completed historical meal m', $\mathcal{F}_{m'}(\mathbf{u}; \Theta) \in \mathbb{R}^H$ denotes the personalized trajectory with its meal inputs and shared parameters fixed. The two-dimensional latent state is shared across a participant's meals. We use conditionally independent meal residuals given this state and the observed inputs, with within-meal temporal dependence represented by Σ_ε. The working observation model is

$$\mathbf{G}_{m'}^{\mathrm{post}} = \mathcal{F}_{m'}(\mathbf{U}_i) + \boldsymbol{\varepsilon}_{m'}, \qquad \boldsymbol{\varepsilon}_{m'} \sim \mathcal{N}(\mathbf{0}, \Sigma_\varepsilon).$$

Within each outer fold and seed, Σ_ε is fitted from M inner-validation residuals under zero individual state:

$$\mathbf{e}_b^{\mathrm{val}} = \mathbf{G}_b^{\mathrm{post}} - \mathcal{F}_b(\mathbf{0}), \qquad b = 1, \ldots, M.$$

Center these validation residuals to estimate their empirical covariance S_ε with divisor M:

$$\bar{\mathbf{e}}^{\mathrm{val}} = \frac{1}{M}\sum_{b=1}^{M}\mathbf{e}_b^{\mathrm{val}}, \qquad S_\varepsilon = \frac{1}{M}\sum_{b=1}^{M}(\mathbf{e}_b^{\mathrm{val}} - \bar{\mathbf{e}}^{\mathrm{val}})(\mathbf{e}_b^{\mathrm{val}} - \bar{\mathbf{e}}^{\mathrm{val}})^\top.$$

We apply Ledoit–Wolf shrinkage (Ledoit & Wolf, 2004), with $\lambda_{\mathrm{LW}} \in [0,1]$ estimated analytically from these centered residuals by the standard Ledoit–Wolf criterion:

$$\bar{v}_\varepsilon = \frac{\mathrm{tr}(S_\varepsilon)}{H}, \qquad \Sigma_\varepsilon = (1-\lambda_{\mathrm{LW}})S_\varepsilon + \lambda_{\mathrm{LW}}\bar{v}_\varepsilon I_H.$$

We use the centered estimator implemented by scikit-learn's `LedoitWolf`. The covariance is shared across participants within each dataset, fold, and seed. It captures residual variation after the population fit and is held fixed during adaptation. Centering is used for covariance fitting only: the mean is removed before estimating S_ε, while the update uses the un-centered residual observation through $\widetilde{\mathbf{G}}_{m'}^{\mathrm{post}}$. The local likelihood absorbs linearization error into this fitted residual covariance. Outer-test outcomes never enter its estimate.

We linearize the observation model at the current posterior mean $\boldsymbol{\mu}_i$ and update the posterior sequentially in information form:

$$\mathcal{F}_{m'}(\mathbf{u}) \approx \mathcal{F}_{m'}(\boldsymbol{\mu}_i) + J_{m'}(\mathbf{u}-\boldsymbol{\mu}_i), \qquad J_{m'} = \left.\frac{\partial \mathcal{F}_{m'}(\mathbf{u})}{\partial \mathbf{u}}\right|_{\mathbf{u}=\boldsymbol{\mu}_i} \in \mathbb{R}^{H\times 2}.$$

Equivalently, with $\mathbf{a}_{m'} = \mathcal{F}_{m'}(\boldsymbol{\mu}_i) - J_{m'}\boldsymbol{\mu}_i$, define

$$\widetilde{\mathbf{G}}_{m'}^{\mathrm{post}} = \mathbf{G}_{m'}^{\mathrm{post}} - \mathbf{a}_{m'} = \mathbf{G}_{m'}^{\mathrm{post}} - \mathcal{F}_{m'}(\boldsymbol{\mu}_i) + J_{m'}\boldsymbol{\mu}_i.$$

Then $\widetilde{\mathbf{G}}_{m'}^{\mathrm{post}} \approx J_{m'}\mathbf{u}_i + \boldsymbol{\varepsilon}_{m'}$. The approximate likelihood for a candidate $\mathbf{u}$ is

$$p(\widetilde{\mathbf{G}}_{m'}^{\mathrm{post}} \mid \mathbf{u}) \propto \exp\left[-\frac{1}{2}(\widetilde{\mathbf{G}}_{m'}^{\mathrm{post}} - J_{m'}\mathbf{u})^\top W_\varepsilon (\widetilde{\mathbf{G}}_{m'}^{\mathrm{post}} - J_{m'}\mathbf{u})\right], \qquad W_\varepsilon = \Sigma_\varepsilon^{-1}.$$

A.6.3 Bayesian Posterior Update in Information Form

Given the current prior $\mathbf{U}_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \Sigma_{u,i})$, define the precision and natural parameter

$$\Lambda_i = \Sigma_{u,i}^{-1}, \qquad \boldsymbol{\eta}_i = \Lambda_i\boldsymbol{\mu}_i.$$

Combining the Gaussian prior with the local likelihood yields

$$\Lambda_i^+ = \Lambda_i + J_{m'}^\top W_\varepsilon J_{m'}, \qquad \boldsymbol{\eta}_i^+ = \boldsymbol{\eta}_i + J_{m'}^\top W_\varepsilon \widetilde{\mathbf{G}}_{m'}^{\mathrm{post}},$$

$$\Sigma_{u,i}^+ = (\Lambda_i^+)^{-1}, \qquad \boldsymbol{\mu}_i^+ = \Sigma_{u,i}^+\boldsymbol{\eta}_i^+.$$

Equivalently, the new posterior mean solves the two-dimensional regularized weighted least-squares problem

$$\boldsymbol{\mu}_i^+ = \arg\min_{\mathbf{u}} \left\{\frac{1}{2}(\mathbf{u}-\boldsymbol{\mu}_i)^\top\Sigma_{u,i}^{-1}(\mathbf{u}-\boldsymbol{\mu}_i) + \frac{1}{2}(\widetilde{\mathbf{G}}_{m'}^{\mathrm{post}} - J_{m'}\mathbf{u})^\top W_\varepsilon(\widetilde{\mathbf{G}}_{m'}^{\mathrm{post}} - J_{m'}\mathbf{u})\right\}.$$

For completeness, discarding terms independent of $\mathbf{u}$, the negative log posterior is

$$\begin{aligned}-\log p(\mathbf{u} \mid \widetilde{\mathbf{G}}_{m'}^{\mathrm{post}}) = {}& \tfrac{1}{2}\mathbf{u}^\top(\Lambda_i + J_{m'}^\top W_\varepsilon J_{m'})\mathbf{u} \\ & - \mathbf{u}^\top(\boldsymbol{\eta}_i + J_{m'}^\top W_\varepsilon \widetilde{\mathbf{G}}_{m'}^{\mathrm{post}}) + \mathrm{const}.\end{aligned}$$

Completing the square gives the information update above. Since Λ_i is positive definite and $J_{m'}^\top W_\varepsilon J_{m'}$ is positive semidefinite, the updated precision is positive definite. The posterior mean is obtained by solving the 2×2 system $\Lambda_i^+ \boldsymbol{\mu}_i^+ = \boldsymbol{\eta}_i^+$. The implementation precomputes the observation precision, symmetrizes the updated information matrix, and forms the posterior covariance only for uncertainty diagnostics.

A.6.4 No-lookahead replay

Initialize $\Lambda_i = I_2$, $\boldsymbol{\eta}_i = \mathbf{0}$, and the set of processed meals $\mathcal{C}_i = \varnothing$. With the shared parameters frozen, process participant i's meals in chronological order, where m denotes the meal being predicted and m' a historical meal being assimilated. Before predicting meal m:

1. Collect the completed historical meals that have not yet been incorporated:

$$\mathcal{A}_{i,m} = \{m' : t_{i,m'} + T \le t_{i,m},\ m' \notin \mathcal{C}_i\}.$$

2. For each $m' \in \mathcal{A}_{i,m}$ in chronological order, solve $\Lambda_i \boldsymbol{\mu}_i = \boldsymbol{\eta}_i$. Evaluate $J_{m'}$ and $\widetilde{\mathbf{G}}_{m'}^{\mathrm{post}}$ at this current mean, apply the information updates above, and set $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{m'\}$.
3. Solve for the updated posterior mean and emit the current prediction:

$$\widehat{\mathbf{G}}_{i,m}^{\mathrm{post}} = F_\Theta\big(\mathbf{G}_{i,m}^{\mathrm{pre}}, \mathbf{n}_{i,m}, G_{i,m}^{\mathrm{fast}}, \phi_{i,m}; \boldsymbol{\mu}_i\big).$$

Meal m becomes eligible for a subsequent update only at $t_{i,m} + 120$ min. Each historical trajectory is incorporated once, using the posterior mean available at the time of its update. Thus, overlapping meals whose outcomes remain incomplete cannot affect the current prediction; all updates use information available at prediction time.

A.6.5 Numerical Jacobian and computational cost

The implementation caches each meal's population components $(G(0), \mathbf{R}^{\mathrm{meal}}, \mathbf{R}^{\mathrm{state}}, \tau)$. At a historical update, only the two-coordinate transform is reevaluated. For coordinate $\ell \in \{1, 2\}$ and unit vector $\mathbf{e}_\ell \in \mathbb{R}^2$, the implemented Jacobian is the central difference

$$J_{m'}[:, \ell] \approx \frac{\mathcal{F}_{m'}(\boldsymbol{\mu}_i + \epsilon_{\mathrm{FD}}\mathbf{e}_\ell) - \mathcal{F}_{m'}(\boldsymbol{\mu}_i - \epsilon_{\mathrm{FD}}\mathbf{e}_\ell)}{2\epsilon_{\mathrm{FD}}}, \qquad \epsilon_{\mathrm{FD}} = 10^{-3}.$$

Thus one center evaluation and four perturbed evaluations suffice; all use cached components, with FP64 arithmetic for the transforms and posterior update. Adaptation differentiates only the two-coordinate transform; the population network stays frozen. Differentiating the transform gives the analytic sensitivities

$$\begin{aligned}
\frac{\partial \widehat{\mathbf{G}}_i^{\mathrm{post}}(t)}{\partial u_{\rho,i}} &= \log(2)\big[1 - \tanh^2(u_{\rho,i})\big]\, g_\rho(u_{\rho,i}) R^{\mathrm{meal}}(t) C_r(t; u_{r,i}, \tau), \\
\frac{\partial \widehat{\mathbf{G}}_i^{\mathrm{post}}(t)}{\partial u_{r,i}} &= -\log(2)\big[1 - \tanh^2(u_{r,i})\big] q(t; \tau) \\
&\qquad \times \big[g_\rho(u_{\rho,i}) R^{\mathrm{meal}}(t) + R^{\mathrm{state}}(t)\big] C_r(t; u_{r,i}, \tau).
\end{aligned}$$

These identities describe sensitivity directions; replay uses the central differences above. A zero meal component supplies no gain information, and clearance is uninformative before τ; saturated transforms can also reduce sensitivity. The Gaussian prior regularizes weakly informed coordinates.

With latent dimension $q = 2$, the numerical Jacobian costs $O(Hq)$. With shared dense observation precision precomputed, the information update costs $O(H^2 q + Hq^2 + q^3)$. Current-meal prediction adds $O(Hq)$ to the shared forward pass. The minimal posterior representation is a two-vector and a 2×2 precision matrix, six FP64 values (48 bytes). Beyond this six-value state, the full system also stores metadata, pending-meal caches, diagnostic records, shared weights, and covariance. Completed meals are assimilated once; no optimizer state or forgetting factor is used.

A.7 Training Objective Details

A.7.1 Masked trajectory objective

The population backbone is trained before online adaptation. Let M be the number of meals averaged in the loss, each with at least one available target. Let $a_{mh} \in \{0, 1\}$ indicate the availability of target $G_m^{\text{post}}(t_h)$, and define $e_{mh} = \widehat{G}_m^{\text{post}}(t_h) - G_m^{\text{post}}(t_h)$. We use a Huber loss (Huber, 1992) with $\delta_{\text{H}} = 10$ mg/dL:

$$\ell_{\text{H}}(e) = \begin{cases} \frac{1}{2}e^2, & |e| \le \delta_{\text{H}}, \\ \delta_{\text{H}}\left(|e| - \frac{1}{2}\delta_{\text{H}}\right), & |e| > \delta_{\text{H}}. \end{cases}$$

Late-horizon errors receive a smoothly increasing MSE component,

$$r_h = \left(\frac{h}{H}\right)^{\gamma_{\text{late}}}, \qquad m_h^{\text{late}} = \eta_{\text{late}} r_h,$$

$$\ell_{mh}^{\text{point}} = (1 - m_h^{\text{late}})\ell_{\text{H}}(e_{mh}) + m_h^{\text{late}} \frac{1}{2} e_{mh}^2, \qquad \gamma_{\text{late}} = 4, \quad \eta_{\text{late}} = 0.25.$$

Here r_h is a scalar time ramp, $m_h^{\text{late}} \in [0, 0.25]$ is the late-horizon mixture coefficient of the main-text objective, and $a_{mh} \in \{0, 1\}$ is the target-availability mask. We additionally emphasize the early postprandial region with a Gaussian window centered at 30 min:

$$a_{30,h} = \exp\left[-\frac{1}{2}\left(\frac{t_h - 30}{10}\right)^2\right], \qquad w_h = 1 + a_{30,h}.$$

The masked trajectory objective is

$$\mathcal{L}_{\text{point}} = \frac{1}{M} \sum_{m=1}^{M} \frac{\sum_{h=1}^{H} a_{mh} w_h \ell_{mh}^{\text{point}}}{\sum_{h=1}^{H} a_{mh} w_h}.$$

A.7.2 Smoothed peak supervision

We prepend $G_m(0)$ to both trajectories and use a three-point centered moving average at 5-min spacing. Its endpoint convention is replication padding:

$$g_{m,-1} = g_{m,0}, \qquad g_{m,H+1} = g_{m,H},$$

with the same convention for $\widehat{g}$. Thus the 3-point average used in $\mathcal{L}_{\text{peak}}$ is defined for every index $h \in \{0, \ldots, H\}$. The three study datasets use complete target trajectories, so every meal has a defined smoothed peak target.

$$\bar{g}_{m,h} = \frac{g_{m,h-1} + g_{m,h} + g_{m,h+1}}{3}, \qquad \bar{\widehat{g}}_{m,h} = \frac{\widehat{g}_{m,h-1} + \widehat{g}_{m,h} + \widehat{g}_{m,h+1}}{3}.$$

With $g_{m,0} = \widehat{g}_{m,0} = G_m(0)$, $g_{m,h} = G_m^{\text{post}}(t_h)$, and $\widehat{g}_{m,h} = \widehat{G}_m^{\text{post}}(t_h)$, define the smoothed peak excursions as

$$P_m = \max_{0 \le h \le H} \bar{g}_{m,h} - G_m(0), \qquad \widehat{P}_m = \max_{0 \le h \le H} \bar{\widehat{g}}_{m,h} - G_m(0).$$

The peak objective and final training objective are

$$\mathcal{L}_{\text{peak}} = \frac{1}{M}\sum_{m=1}^{M} \ell_{\text{H}}(\widehat{P}_m - P_m), \qquad \lambda_{\text{peak}} = 0.20, \qquad \boxed{\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{point}} + \lambda_{\text{peak}}\mathcal{L}_{\text{peak}}.}$$

The same smoothed peak definition is used for the auxiliary loss and the corresponding *Peak-MAE* evaluation. The frozen CGM-JEPA encoder is excluded from optimization; all remaining shared parameters are optimized jointly, while participant-specific latent states are updated only after historical outcomes are released.

B Experimental Design

B.1 Datasets

All datasets are public and are used as released, after the common meal-window and CGM-quality filtering described below. The number of qualifying meals differs from the raw record count.

CGMacros. CGMacros (Das et al., 2025) contains paired 5-minute CGM and meal-level macronutrient records from 45 participants, yielding 1,650 qualifying meals for the locked analysis. It also provides measured laboratory fasting plasma glucose at mealtime. The dataset is available through the Scientific Data record at `https://doi.org/10.1038/s41597-025-05851-7`.

BIG IDEAs. BIG IDEAs (Cho et al., 2023) contains CGM and wearable data from 14 participants; 534 meals satisfy the complete-window criterion used in our evaluation. The dataset is available on PhysioNet at `https://physionet.org/content/big-ideas-glycemic-wearable/1.1.3/`.

T1D-UOM. T1D-UOM (Alsuhaymi et al., 2025) contains longitudinal type 1 diabetes data from seven participants, with 1,487 qualifying meals. It does not provide insulin records for the meals used here. The dataset is available through the Scientific Data record at `https://doi.org/10.1038/s41597-025-05695-1`.

Data preprocessing and quality control. CGM timestamps were parsed, chronologically sorted, and restricted to finite observations. Duplicate timestamps were resolved by retaining the last record for CGMacros and by averaging repeated CGM observations for BIG IDEAs and T1D-UOM. All meals were represented on a common 5-minute grid from -115 to $+120$ min. For BIG IDEAs and T1D-UOM, a meal was retained only if every target point could be obtained by interpolation between observed samples separated by at most 12 min; no extrapolation was used. T1D-UOM glucose values were converted from mmol/L to mg/dL by multiplying by 18. For CGMacros, missing samples were permitted only when at most 10% of the samples in a window were missing, and only runs of at most two missing grid points were linearly interpolated. The mealtime anchor and all post-meal values were required to lie within $[40, 400]$ mg/dL. Consecutive records of the same meal type within 30 min were merged, all four macronutrient values were required, and meals were excluded when another logged meal occurred less than 180 min before or less than 120 min after the target meal. BIG IDEAs and T1D-UOM additionally required complete carbohydrate, protein, fat, and fiber records. Meals failing any criterion were removed before fold construction and were not reintroduced during model fitting or evaluation.

Table 4: Locked analysis sets and evaluation protocol. Each dataset is trained and validated separately; outer folds are participant-disjoint and every model/seed is evaluated on all outer folds. CGMacros carries measured fasting plasma glucose; the other datasets use a strictly causal morning CGM-derived fasting proxy (running median of qualified anchors with a 120-minute outcome-release rule), which is available for the stated number of meals and otherwise imputed within the training partition.

Dataset	Meals	Participants	Outer folds	Fasting covariate
CGMacros	1,650	45	5	measured FPG
BIG IDEAs	534	14	5	proxy, 424/534 meals
T1D-UOM	1,487	7	7	proxy, 1,424/1,487 meals
Total	3,671	66	51	

B.1.1 Fasting-Glucose Input and Dataset-Specific Proxies

For a training partition, fasting glucose is standardized as

$$\widetilde{G}^{\text{fast}} = \frac{G^{\text{fast}} - m^{\text{tr}}_{\text{fast}}}{\sigma^{\text{tr}}_{\text{fast}}},$$

The scaler gives each training participant equal weight. For the set $\mathcal{S}^{\text{tr}}_{\text{fast}}$ of inner-training participants with at least one available fasting value, define one representative value per participant:

$$a_i^{\text{fast}} = \text{median}_{m \in \mathcal{M}_i^{\text{tr}}} G^{\text{fast}}_{i,m}, \qquad m^{\text{tr}}_{\text{fast}} = \text{median}_{i \in \mathcal{S}^{\text{tr}}_{\text{fast}}} a_i^{\text{fast}},$$

$$\bar{a}^{\text{fast}} = \frac{1}{|\mathcal{S}^{\text{tr}}_{\text{fast}}|} \sum_{i \in \mathcal{S}^{\text{tr}}_{\text{fast}}} a_i^{\text{fast}}, \qquad \sigma^{\text{tr}}_{\text{fast}} = \sqrt{\frac{\sum_{i \in \mathcal{S}^{\text{tr}}_{\text{fast}}} (a_i^{\text{fast}} - \bar{a}^{\text{fast}})^2}{|\mathcal{S}^{\text{tr}}_{\text{fast}}|}}.$$

Here $\mathcal{M}_i^{\text{tr}}$ contains that participant's inner-training meals with available values. Repeated laboratory measurements therefore do not weight a participant by meal count; for a time-varying proxy, the representative value is its training-meal median. Missing meal-level inputs are imputed by $m^{\text{tr}}_{\text{fast}}$. If the set is empty, the center and scale are 100 mg/dL and 1; a scale at most 10^{-6} is replaced by 1. The fitted scaler is fixed for validation and testing.

CGMacros uses its available measured fasting-plasma-glucose covariate. For BIG IDEAs and T1D-UOM, the corresponding feature is a past-only CGM-derived fasting proxy: a qualified anchor occurs from 04:00 to 11:00, after at least 8 h without recorded food, when the median CGM in $[-30, 0]$ min lies in $[40, 400]$ mg/dL. A meal uses the running median of qualified anchors observed no later than that meal; no historical anchor leaves the feature missing until the training-partition imputation above. In these two datasets the proxy supplies the fasting-glucose input G^{fast}.

B.2 Baseline Adaptation Details

The baselines are intentionally reported in two input groups. The CGM-only rows use the 24 pre-meal CGM samples and no current-meal covariates; *Gluformer* uses the GlucoBench-maintained core and predicts a normalized postprandial residual that is rescaled to mg/dL, while *CGM-JEPA (CGM-only)* freezes the same encoder used by MealGluco, pools its two 12-sample tokens, and decodes the trajectory with a 128-dimensional MLP head. The covariate-aware rows receive the four nutrient doses and the observed fasting/clock features where available. *UniCA-Gluformer* and *UniCA-CGM-JEPA* map each nutrient to a token that combines its scaled value, an observation mask, and a nutrient-type embedding, then fuse these tokens with the glucose representation by cross-attention. *TimeXer* is the official core adapted to the causal meal protocol with patch length 6; its default normalization is disabled so time-constant meal covariates are not normalized to zero. The *Stacking Framework* is a phenotype-stacked Ridge→CatBoost pipeline: the first stage fits an independent Ridge model at each horizon, and the second stage fits one CatBoost model per glycemic

phenotype and horizon on inner-fold out-of-fold first-stage predictions. Because two datasets do not provide the diagnosis labels used in the original stacked formulation, the phenotype is defined without labels by training-fold tertiles of the pre-meal glucose level; fold-specific thresholds are then applied to validation meals.

B.3 Evaluation Metrics

For meal (i, m), let $G^{\text{post}}_{i,m}(t_h)$ and $\widehat{G}^{\text{post}}_{i,m}(t_h)$ denote the observed and predicted postprandial CGM values at $t_h = 5h$ min for $h = 1, \dots, H$, and let $G_{i,m}(0)$ be the observed mealtime anchor. Let $\mathcal{M}$ denote the set of qualifying meals.

Trajectory-MAE.

$$\text{MAE}_{\text{traj}} = \frac{1}{|\mathcal{M}|H} \sum_{(i,m)\in\mathcal{M}} \sum_{h=1}^{H} \left| \widehat{G}^{\text{post}}_{i,m}(t_h) - G^{\text{post}}_{i,m}(t_h) \right|.$$

iAUC-MAE. Let

$$A_{i,m} = \int_0^{120} [G_{i,m}(t) - G_{i,m}(0)]_+ \, dt, \qquad \widehat{A}_{i,m} = \int_0^{120} [\widehat{G}^{\text{post}}_{i,m}(t) - G_{i,m}(0)]_+ \, dt.$$

We approximate each integral with the trapezoidal rule on the 5-minute grid after clipping each ordinate at zero. The metric is

$$\text{MAE}_{\text{iAUC}} = \frac{1}{|\mathcal{M}|} \sum_{(i,m)\in\mathcal{M}} \left| \widehat{A}_{i,m} - A_{i,m} \right|.$$

Peak-MAE. We prepend $G_{i,m}(0)$ to both trajectories and form a three-point centered moving average at 5-min spacing, corresponding to a 15-minute centered window. Endpoint values use replication padding:

$$g_{i,m,-1} = g_{i,m,0}, \qquad g_{i,m,H+1} = g_{i,m,H},$$

with the same convention for $\widehat{g}$. The smoothed trajectories are

$$\bar{g}_{i,m,h} = \frac{g_{i,m,h-1} + g_{i,m,h} + g_{i,m,h+1}}{3}, \qquad \bar{\widehat{g}}_{i,m,h} = \frac{\widehat{g}_{i,m,h-1} + \widehat{g}_{i,m,h} + \widehat{g}_{i,m,h+1}}{3},$$

where $g_{i,m,0} = \widehat{g}_{i,m,0} = G_{i,m}(0)$, $g_{i,m,h} = G^{\text{post}}_{i,m}(t_h)$, and $\widehat{g}_{i,m,h} = \widehat{G}^{\text{post}}_{i,m}(t_h)$. The metric is

$$\text{MAE}_{\text{peak}} = \frac{1}{|\mathcal{M}|} \sum_{(i,m)\in\mathcal{M}} \left| \max_h \bar{\widehat{g}}_{i,m,h} - \max_h \bar{g}_{i,m,h} \right|.$$

Since both trajectories use the same observed anchor, this is equivalent to comparing their smoothed peak excursions above $G_{i,m}(0)$.

Near-Peak Distance. Let $\tau^{+}_{i,m} = \min \arg\max_h \bar{\widehat{g}}_{i,m,h}$ be the earliest predicted smoothed-peak time, and let

$$\mathcal{T}_{i,m} = \left\{ t_h : \left| \bar{g}_{i,m,h} - \max_{h'} \bar{g}_{i,m,h'} \right| \le 5 \text{ mg/dL} \right\}$$

be the observed near-peak grid points. The distance is

$$d_{i,m} = \min_{t\in\mathcal{T}_{i,m}} \left| \tau^{+}_{i,m} - t \right|.$$

Recovery-MAE. Let

$$D_{i,m} = \max_h \bar{g}_{i,m,h} - \bar{g}_{i,m,H}, \qquad \widehat{D}_{i,m} = \max_h \bar{\widehat{g}}_{i,m,h} - \bar{\widehat{g}}_{i,m,H},$$

and define

$$\text{MAE}_{\text{rec}} = \frac{1}{|\mathcal{M}_d|} \sum_{(i,m)\in\mathcal{M}_d} \left| \widehat{D}_{i,m} - D_{i,m} \right|.$$

Discernible meals. *Near-Peak Distance* and *Recovery-MAE* require a well-defined peak and are therefore computed on

$$\mathcal{M}_d = \left\{ (i,m) \in \mathcal{M} : \max_h \bar{g}_{i,m,h} - G_{i,m}(0) \geq 10 \text{ mg/dL} \right\}.$$

The remaining three metrics use all meals in $\mathcal{M}$.

B.4 Experimental Design, Optimization, and Validation Protocol

Every model is trained and evaluated separately within each dataset under participant-disjoint outer validation. CGMacros and BIG IDEAs use 5-fold grouped cross-validation, and T1D-UOM uses leave-one-participant-out because it contains seven participants. We repeat the full procedure with seeds 2026, 2027, and 2028 and report the mean and sample SD over the three seeds. Preprocessing statistics, model selection, and residual-covariance fitting are fold-local; all reported predictions are out-of-fold. Online personalization follows the causal replay rule described in Appendix A.6: before predicting meal m, only earlier completed meals $m' \in \{0, \ldots, m-1\}$ whose outcome windows have closed are assimilated, each exactly once.

B.4.1 Optimization and validation

The population network is trained with AdamW (Loshchilov & Hutter, 2019) using batch size 64, constant learning rate 10^{-3}, weight decay 10^{-4}, gradient-norm clipping at 1.0, at most 100 epochs, and early stopping patience 15. Dropout 0.20 is applied to the decoder attention weights and within its FFN; the frozen encoder remains in evaluation mode. The inner-validation checkpoint minimizes $\mathcal{L}_{\text{final}}$, with an improvement threshold of 10^{-4}. All parameters outside the frozen encoder are jointly optimized; participant coordinates are fixed at zero during population training.

MealGluco has 101,896 effective trainable parameters, excluding the frozen CGM-JEPA encoder. The stated architecture, bounds, and loss coefficients are shared across the three datasets and folds; weights, fasting-scaling statistics, and observation covariance are fitted separately within each fold.

Each outer fold uses participant-disjoint training and test partitions. Within its training participants, a grouped split assigns 18% to inner validation, using random state equal to the training seed plus the fold index. Preprocessing statistics and imputations are fitted on inner training; checkpoint selection and residual-covariance fitting use inner validation. The selected checkpoint is restored before covariance estimation. All fold-level fits use training and inner-validation data only. Seeds 2026, 2027, and 2028 independently repeat fitting and evaluation; seed variation is computed from separate out-of-fold predictions.

B.4.2 State-residual bound selection

We select the state-residual bound on CGMacros before applying the fixed architecture to the remaining datasets. Using the same participant-disjoint five-fold splits and random seeds as the main experiments, we sweep $R_{\max} \in \{10, 15, 20, 25, 30\}$ mg/dL. This is a one-dimensional sensitivity analysis, so only $R_{\max}$ changes. As shown in Table 5, 20 mg/dL gives the lowest mean *Trajectory-MAE*, *iAUC-MAE*, and *Peak-MAE* among the five values. The paired comparison in Table 6 shows that 10, 15, and 25 mg/dL have higher mean *Trajectory-MAE* than 20 mg/dL, whereas 30 mg/dL is not statistically distinguishable but has no accuracy advantage. We therefore keep $R_{\max} = 20$ mg/dL fixed across datasets.

Table 5: State-residual bound sensitivity on CGMacros. Values are means over three random seeds under the same participant-disjoint five-fold protocol. Bold entries mark the best mean in each column.

$R_{\max}$ (mg/dL)	Trajectory-MAE	iAUC-MAE	Peak-MAE	Near-Peak Distance	Recovery-MAE
10	22.98	1924.38	28.89	**17.19**	**20.02**
15	23.05	1936.15	28.89	18.37	20.33
20	**22.68**	**1890.45**	**28.16**	19.77	20.77
25	22.97	1923.26	28.64	20.30	20.75
30	22.90	1922.93	28.85	21.00	21.12

Table 6: Paired participant-cluster bootstrap comparison with $R_{\max} = 20$ on CGMacros for *Trajectory-MAE*. Δ is the candidate setting minus $R_{\max} = 20$, so negative values favor the candidate.

$R_{\max}$	Δ	95% CI	p
10	+0.30	[0.001, 0.64]	0.0500
15	+0.37	[0.07, 0.73]	0.0122
25	+0.29	[0.09, 0.51]	0.0046
30	+0.23	[-0.09, 0.61]	0.1796

B.5 Ablation Definitions

Each ablation variant is trained from scratch under the same outer folds, seeds, preprocessing, optimization settings, and evaluation protocol as the full model. The *w/o personalized module* row removes the two participant coordinates and predicts with the population model. *w/o frozen CGM-JEPA tokens* retains only the four short-horizon statistics in the state encoder, and *w/o short-horizon statistics* retains only the frozen encoder tokens. *w/o state residual* removes the bounded state-response sequence. *Unstructured meal-response head* replaces the Gamma event with a parameter-matched unconstrained head, *w/o meal response* removes the meal branch, and *w/o nutrient modifiers* disables the protein, fat, and fiber modulation of the Gamma amplitude and peak time while keeping the carbohydrate-driven base response. *Raw polynomial post-meal time* replaces the shifted-Legendre features by a raw polynomial of the same degree, *w/o post-meal time* removes the relative-time branch, and *w/o clock time* removes the clock-time branch. *MAE point loss* replaces the Huber point objective with a mean absolute error while holding the peak-amplitude term and all other settings fixed.

C Additional Results

C.1 Additional Main Results

Table 4 lists the locked analysis sets and validation protocol. Tables 7 and 8 report the full five-metric comparison on BIG IDEAs and T1D-UOM, and Tables 13 and 14 report the corresponding component ablations, in which each row changes one component relative to the information-matched population reference (*w/o personalized module*); the unstructured meal-response head is parameter-matched to the Gamma event (102,247 vs. 102,192 trainable parameters). Table 10 isolates the two online coordinates, and Table 11 compares their accuracy and deployment cost with patient-level fine-tuning. Table 9 collects *Trajectory-MAE* on all three datasets in compact form, and Table 15 reports the wrong-history negative control.

Table 7: Main comparison on BIG IDEAs.

Model	Trajectory-MAE	iAUC-MAE	Peak-MAE	Near-Peak Distance	Recovery-MAE
Gluformer (CGM-only)	19.86 ± 0.59	1460.8 ± 95.4	21.62 ± 1.07	$\mathbf{16.55 \pm 0.03}$	19.68 ± 0.50
CGM-JEPA (CGM-only)	18.40 ± 0.21	1485.1 ± 1.0	25.11 ± 0.27	19.06 ± 0.42	21.91 ± 0.47
UniCA-Gluformer	18.81 ± 0.67	1322.4 ± 76.7	20.39 ± 1.26	16.84 ± 0.49	18.86 ± 0.58
UniCA-CGM-JEPA	17.57 ± 0.11	1290.3 ± 23.5	21.50 ± 0.37	16.86 ± 0.51	20.23 ± 0.28
TimeXer	19.42 ± 0.37	1448.7 ± 42.7	21.75 ± 0.47	18.03 ± 0.53	19.47 ± 0.98
Stacking Framework	$\mathbf{16.49 \pm 0.01}$	1237.2 ± 1.7	$\mathbf{18.54 \pm 0.03}$	17.72 ± 0.35	$\mathbf{17.23 \pm 0.04}$
MealGluco	16.56 ± 0.18	$\mathbf{1191.7 \pm 16.4}$	18.76 ± 0.10	16.65 ± 0.39	18.66 ± 0.40

Table 8: Main comparison on T1D-UOM.

Model	Trajectory-MAE	iAUC-MAE	Peak-MAE	Near-Peak Distance	Recovery-MAE
Gluformer (CGM-only)	33.17 ± 1.19	2294.7 ± 83.0	35.96 ± 0.42	26.33 ± 0.74	24.82 ± 0.48
CGM-JEPA (CGM-only)	31.19 ± 0.15	2182.4 ± 13.9	36.42 ± 0.51	33.18 ± 0.60	24.73 ± 0.25
UniCA-Gluformer	33.34 ± 0.46	2256.1 ± 25.5	35.53 ± 0.41	25.70 ± 2.86	25.31 ± 0.33
UniCA-CGM-JEPA	32.17 ± 0.02	2259.9 ± 11.1	38.36 ± 0.66	27.24 ± 2.66	25.27 ± 0.10
TimeXer	32.86 ± 0.53	2221.0 ± 39.6	35.75 ± 0.85	26.21 ± 1.74	25.43 ± 0.43
Stacking Framework	32.30 ± 0.02	2248.5 ± 1.5	32.88 ± 0.05	$\mathbf{24.40 \pm 0.40}$	24.41 ± 0.03
MealGluco	$\mathbf{30.08 \pm 0.55}$	$\mathbf{1986.8 \pm 45.3}$	$\mathbf{31.87 \pm 0.53}$	24.83 ± 1.54	$\mathbf{24.06 \pm 0.55}$

Table 9: Trajectory-MAE on the three formal datasets. Each model is trained and evaluated independently per dataset with participant-disjoint outer folds; no prediction ensemble is used. BIG IDEAs is led by the strong tabular Stacking Framework baseline; MealGluco leads on CGMacros and T1D-UOM.

Model	CGMacros	BIG IDEAs	T1D-UOM
Gluformer (CGM-only)	25.87 ± 0.32	19.86 ± 0.59	33.17 ± 1.19
CGM-JEPA (CGM-only)	25.19 ± 0.15	18.40 ± 0.21	31.19 ± 0.15
UniCA-Gluformer	24.21 ± 0.57	18.81 ± 0.67	33.34 ± 0.46
UniCA-CGM-JEPA	24.43 ± 0.18	17.57 ± 0.11	32.17 ± 0.02
TimeXer	25.22 ± 0.05	19.42 ± 0.37	32.86 ± 0.53
Stacking Framework	23.48 ± 0.01	$\mathbf{16.49 \pm 0.01}$	32.30 ± 0.02
MealGluco	$\mathbf{21.16 \pm 0.17}$	16.56 ± 0.18	$\mathbf{30.08 \pm 0.55}$

C.2 Online Personalization

We report the two-personalization-parameter variants and the comparison with per-participant fine-tuning.

Table 10: Online physiological personalization axes. The Population row is the model before any completed meal is assimilated; the availability of completed historical meals is otherwise held fixed, and only the active latent coordinates change. The amplitude axis carries most of the gain in every dataset, and adding peak clearance lowers the error further on all three datasets, which is why the final state has two coordinates.

Personalization state	CGMacros	BIG IDEAs	T1D-UOM
Population	22.63 ± 0.23	16.87 ± 0.46	31.01 ± 0.79
Peak-clearance only	22.37 ± 0.19	16.78 ± 0.42	30.91 ± 0.76
Amplitude only	21.47 ± 0.23	16.61 ± 0.15	30.17 ± 0.58
Amplitude + peak clearance	$\mathbf{21.16 \pm 0.17}$	$\mathbf{16.56 \pm 0.18}$	$\mathbf{30.08 \pm 0.55}$

Table 11: Online personalization: the two-coordinate online state versus patient-level fine-tuning. Top panel: Trajectory-MAE; Δ is fine-tuning minus the online state with a participant-cluster paired 95% bootstrap CI, so positive values favor the online state. Bottom panel: per completed historical meal, the adaptation cost and the persistent patient-specific state. The online update runs on CPU without backpropagation; fine-tuning FLOPs are cache-adjusted per-gradient-step lower bounds and its elapsed time is a deployment measurement, not a same-kernel benchmark. The fine-tuning scope and optimizer are selected on inner validation.

Accuracy (Trajectory-MAE, mg/dL)

Dataset	Online	Patient FT	ΔMAE	95% paired CI
CGMacros	21.16 ± 0.17	21.96 ± 0.23	+0.80	[0.13, 1.62]
BIG IDEAs	16.56 ± 0.18	16.46 ± 0.37	-0.10	[-0.40, 0.31]
T1D-UOM	30.08 ± 0.55	29.41 ± 0.47	-0.66	[-1.06, -0.19]

Cost per completed historical meal

Dataset	Adaptation FLOPs	Ratio	Persistent state	Ratio
CGMacros	10,005 vs 9,829,631	982×	168 B vs 677,651 B	4,034×
BIG IDEAs	10,005 vs 5,836,127	583×	168 B vs 333,975 B	1,988×
T1D-UOM	10,005 vs 8,005,005	800×	168 B vs 46,617 B	277×

C.3 Component Ablations

We report the complete dataset-specific component ablations under the same folds and seeds as the full model.

Table 12: Complete component ablation on CGMacros.

Configuration	Trajectory-MAE	iAUC-MAE	Peak-MAE	Near-Peak Distance	Recovery-MAE
MealGluco	$\mathbf{21.16 \pm 0.17}$	$\mathbf{1701.7 \pm 21.3}$	$\mathbf{25.00 \pm 0.32}$	17.54 ± 0.33	19.18 ± 0.32
w/o personalized module	22.63 ± 0.23	1873.1 ± 20.0	26.96 ± 0.08	20.08 ± 0.29	20.48 ± 0.21
w/o frozen CGM-JEPA tokens	23.42 ± 0.20	1960.5 ± 33.9	27.77 ± 0.68	19.91 ± 0.63	20.11 ± 0.35
w/o short-horizon statistics	23.35 ± 0.04	1934.0 ± 4.0	27.69 ± 0.22	19.72 ± 0.69	20.45 ± 0.20
w/o state response	23.54 ± 0.11	1951.0 ± 16.4	28.58 ± 0.28	18.01 ± 0.09	19.04 ± 0.11
Unstructured meal-response head	23.04 ± 0.24	1901.6 ± 14.7	26.65 ± 0.27	20.25 ± 1.01	20.11 ± 0.54
w/o meal response	25.18 ± 0.19	2177.7 ± 13.5	36.45 ± 0.12	26.54 ± 0.25	24.20 ± 0.14
w/o nutrient modifiers	22.75 ± 0.16	1887.4 ± 20.7	26.95 ± 0.40	18.91 ± 0.46	19.99 ± 0.19
Raw polynomial post-meal time	23.16 ± 0.14	1925.2 ± 13.7	27.91 ± 0.30	19.02 ± 0.88	20.41 ± 0.55
w/o post-meal time	23.47 ± 0.19	1930.8 ± 25.4	28.16 ± 0.32	$\mathbf{17.34 \pm 0.06}$	$\mathbf{18.72 \pm 0.15}$
w/o clock time	23.32 ± 0.19	1951.5 ± 27.1	27.70 ± 0.10	19.58 ± 0.77	20.38 ± 0.22
MAE point loss	23.76 ± 0.20	1954.7 ± 18.7	26.63 ± 0.20	17.69 ± 0.62	19.31 ± 0.35

Table 13: Component ablation on BIG IDEAs.

Configuration	Trajectory-MAE	iAUC-MAE	Peak-MAE	Near-Peak Distance	Recovery-MAE
MealGluco	$\mathbf{16.56 \pm 0.18}$	$\mathbf{1191.7 \pm 16.4}$	18.76 ± 0.10	16.65 ± 0.39	18.66 ± 0.40
w/o personalized module	16.87 ± 0.46	1247.3 ± 58.8	19.03 ± 0.59	15.21 ± 0.15	18.02 ± 0.13
w/o frozen CGM-JEPA tokens	16.67 ± 0.05	1227.3 ± 8.8	18.47 ± 0.05	16.21 ± 0.16	18.36 ± 0.19
w/o short-horizon statistics	17.15 ± 0.09	1232.5 ± 1.9	18.91 ± 0.27	14.85 ± 0.42	18.48 ± 0.10
w/o state response	17.36 ± 0.21	1244.6 ± 43.5	18.95 ± 0.81	14.69 ± 0.40	18.40 ± 0.38
Unstructured meal-response head	16.82 ± 0.14	1235.0 ± 42.6	18.29 ± 0.49	15.83 ± 0.58	$\mathbf{16.81 \pm 0.25}$
w/o meal response	18.32 ± 0.18	1426.4 ± 43.2	22.94 ± 0.55	15.54 ± 1.03	19.43 ± 0.41
w/o nutrient modifiers	16.82 ± 0.27	1216.2 ± 43.1	18.71 ± 0.34	15.65 ± 0.14	18.14 ± 0.38
Raw polynomial post-meal time	17.10 ± 0.36	1265.3 ± 65.9	19.27 ± 0.56	14.88 ± 0.23	17.24 ± 0.11
w/o post-meal time	16.98 ± 0.26	1242.8 ± 46.8	19.26 ± 0.41	15.18 ± 0.09	17.48 ± 0.21
w/o clock time	16.86 ± 0.16	1239.5 ± 32.0	19.04 ± 0.23	15.16 ± 0.71	18.36 ± 0.14
MAE point loss	17.02 ± 0.09	1211.7 ± 9.5	$\mathbf{17.84 \pm 0.15}$	$\mathbf{14.36 \pm 0.11}$	16.92 ± 0.59

Table 14: Component ablation on T1D-UOM.

Configuration	Trajectory-MAE	iAUC-MAE	Peak-MAE	Near-Peak Distance	Recovery-MAE
MealGluco	**30.08 ± 0.55**	**1986.8 ± 45.3**	**31.87 ± 0.53**	24.83 ± 1.54	**24.06 ± 0.55**
w/o personalized module	31.01 ± 0.79	2093.1 ± 84.1	33.15 ± 0.98	25.82 ± 1.96	24.54 ± 0.47
w/o frozen CGM-JEPA tokens	30.98 ± 0.46	2087.7 ± 38.2	32.67 ± 0.52	26.71 ± 0.71	24.36 ± 0.60
w/o short-horizon statistics	31.18 ± 0.63	2082.3 ± 88.0	33.04 ± 1.10	26.66 ± 2.44	25.64 ± 0.56
w/o state response	31.14 ± 0.40	2100.7 ± 51.2	33.99 ± 0.38	**24.21 ± 0.76**	24.59 ± 0.31
Unstructured meal-response head	30.96 ± 0.45	2081.4 ± 66.2	32.29 ± 0.84	27.01 ± 1.24	25.13 ± 0.97
w/o meal response	31.14 ± 0.31	2130.2 ± 35.6	36.67 ± 0.36	26.93 ± 0.79	25.06 ± 0.55
w/o nutrient modifiers	31.50 ± 0.49	2164.1 ± 62.8	33.62 ± 1.06	25.03 ± 1.11	24.47 ± 0.50
Raw polynomial post-meal time	30.82 ± 0.42	2072.3 ± 38.9	33.46 ± 0.40	25.64 ± 0.76	24.63 ± 0.23
w/o post-meal time	31.10 ± 0.53	2074.5 ± 56.6	33.22 ± 0.77	24.81 ± 0.47	24.50 ± 0.19
w/o clock time	31.05 ± 0.40	2101.5 ± 40.2	33.38 ± 1.08	26.64 ± 1.45	24.92 ± 0.47
MAE point loss	31.38 ± 0.21	2064.4 ± 37.3	32.18 ± 0.57	24.89 ± 0.39	25.60 ± 0.55

C.4 Additional Controls

The wrong-history control replaces each participant's historical meals with another participant's released histories while holding the current-meal inputs, model parameters, recipient timeline, and outcome-release delay fixed.

Table 15: Negative control: replacing each participant's historical meals with another participant's released histories while holding the current-meal inputs, model parameters, recipient timeline, and the 120-minute outcome-release delay fixed. Δ is wrong-history minus correct-history, so positive values mean that identity mismatch degrades the metric. All fifteen dataset-metric combinations degrade on average.

Dataset	Metric	Correct history	Wrong subject	Δ
CGMacros	Trajectory-MAE	21.16 ± 0.17	24.43 ± 0.68	+3.27
CGMacros	iAUC-MAE	1701.7 ± 21.3	2115.2 ± 89.4	+413.5
CGMacros	Peak-MAE	25.00 ± 0.32	30.72 ± 0.54	+5.72
CGMacros	Near-Peak Distance	17.54 ± 0.33	21.88 ± 0.94	+4.34
CGMacros	Recovery-MAE	19.18 ± 0.32	21.02 ± 0.34	+1.84
BIG IDEAs	Trajectory-MAE	16.56 ± 0.18	17.31 ± 0.21	+0.75
BIG IDEAs	iAUC-MAE	1191.7 ± 16.4	1302.4 ± 56.3	+110.7
BIG IDEAs	Peak-MAE	18.76 ± 0.10	20.33 ± 0.46	+1.56
BIG IDEAs	Near-Peak Distance	16.65 ± 0.39	17.21 ± 0.71	+0.56
BIG IDEAs	Recovery-MAE	18.66 ± 0.40	19.42 ± 0.29	+0.76
T1D-UOM	Trajectory-MAE	30.08 ± 0.55	33.47 ± 2.33	+3.39
T1D-UOM	iAUC-MAE	1986.8 ± 45.3	2408.3 ± 267.5	+421.4
T1D-UOM	Peak-MAE	31.87 ± 0.53	37.01 ± 3.44	+5.14
T1D-UOM	Near-Peak Distance	24.83 ± 1.54	26.05 ± 1.91	+1.22
T1D-UOM	Recovery-MAE	24.06 ± 0.55	24.56 ± 0.40	+0.50

C.5 Cross-Dataset Generalization

We evaluate source-to-target transfer without fitting or tuning on the target dataset. For each source dataset, all models are trained on source fold 0 and evaluated directly on all participants in each of the two remaining datasets. Target preprocessing uses only statistics fitted on the source training partition. MealGluco online personalization uses only completed target histories whose 120-minute outcome windows have closed. Tables 16–19 report *iAUC-MAE*, *Peak-MAE*, *Near-Peak Distance*, and *Recovery-MAE* under this protocol.

Table 16: iAUC-MAE for cross-dataset generalization. *MealGluco w/o pers.* disables the personalization module and uses no historical meal records.

Trained on	CGMacros		BIG IDEAs		T1D-UOM	
Evaluated on	BIG IDEAs	T1D-UOM	CGMacros	T1D-UOM	CGMacros	BIG IDEAs
MealGluco	**1241.8 ± 50.0**	**2045.9 ± 19.8**	2094.7 ± 106.5	**2050.3 ± 73.2**	**1975.8 ± 33.2**	**1411.4 ± 35.8**
MealGluco w/o pers.	1349.1 ± 119.9	2108.0 ± 24.2	2315.7 ± 58.4	2137.0 ± 30.1	2244.7 ± 45.1	1548.7 ± 59.1
UniCA-Gluformer	1431.2 ± 124.9	2444.6 ± 65.9	**2079.1 ± 81.2**	2281.7 ± 14.4	2312.2 ± 146.9	1665.5 ± 96.0
Stacking Framework	1516.2 ± 2.1	2092.6 ± 2.5	2354.8 ± 1.7	2079.4 ± 1.7	2216.0 ± 2.7	1559.2 ± 3.2

Table 17: Peak-MAE for cross-dataset generalization. *MealGluco w/o pers.* disables the personalization module and uses no historical meal records.

Trained on	CGMacros		BIG IDEAs		T1D-UOM	
Evaluated on	BIG IDEAs	T1D-UOM	CGMacros	T1D-UOM	CGMacros	BIG IDEAs
MealGluco	18.38 ± 0.25	**32.62 ± 0.12**	**31.86 ± 2.07**	**35.27 ± 2.55**	**28.65 ± 0.68**	**19.55 ± 1.06**
MealGluco w/o pers.	**18.05 ± 0.36**	33.37 ± 0.19	35.87 ± 1.57	37.79 ± 1.58	31.94 ± 1.15	20.62 ± 0.52
UniCA-Gluformer	19.19 ± 0.21	37.01 ± 1.33	32.40 ± 1.57	38.01 ± 0.69	33.89 ± 4.09	23.33 ± 1.80
Stacking Framework	20.04 ± 0.04	34.06 ± 0.08	37.33 ± 0.07	39.78 ± 0.07	32.58 ± 0.10	23.74 ± 0.09

Table 18: Near-Peak Distance for cross-dataset generalization. *MealGluco w/o pers.* disables the personalization module and uses no historical meal records.

Trained on	CGMacros		BIG IDEAs		T1D-UOM	
Evaluated on	BIG IDEAs	T1D-UOM	CGMacros	T1D-UOM	CGMacros	BIG IDEAs
MealGluco	17.24 ± 0.94	**23.95 ± 1.11**	22.80 ± 1.73	33.51 ± 2.17	26.89 ± 2.69	27.90 ± 3.51
MealGluco w/o pers.	21.20 ± 3.27	25.59 ± 0.35	24.33 ± 1.66	37.23 ± 0.72	30.40 ± 2.36	31.34 ± 3.32
UniCA-Gluformer	**15.96 ± 0.66**	32.46 ± 2.49	**20.64 ± 0.64**	**33.46 ± 0.61**	**26.75 ± 1.59**	**26.35 ± 2.86**
Stacking Framework	20.58 ± 0.21	26.05 ± 0.16	28.60 ± 0.04	38.98 ± 0.12	30.80 ± 0.22	28.16 ± 0.31

Table 19: Recovery-MAE for cross-dataset generalization. *MealGluco w/o pers.* disables the personalization module and uses no historical meal records.

Trained on	CGMacros		BIG IDEAs		T1D-UOM	
Evaluated on	BIG IDEAs	T1D-UOM	CGMacros	T1D-UOM	CGMacros	BIG IDEAs
MealGluco	19.50 ± 0.59	23.63 ± 0.25	18.70 ± 0.33	**24.81 ± 1.27**	**21.99 ± 0.97**	22.90 ± 0.77
MealGluco w/o pers.	21.25 ± 1.34	24.44 ± 0.28	**18.33 ± 0.13**	25.05 ± 0.67	23.56 ± 0.70	24.18 ± 0.73
UniCA-Gluformer	**18.69 ± 0.45**	25.20 ± 0.30	19.52 ± 0.26	26.47 ± 0.71	22.36 ± 0.37	**21.78 ± 0.78**
Stacking Framework	20.28 ± 0.14	**23.18 ± 0.06**	20.25 ± 0.08	27.66 ± 0.15	23.67 ± 0.08	22.68 ± 0.16

C.6 Statistical Tests

Unless otherwise stated, all intervals are participant-clustered paired 95% bootstrap intervals over 10,000 resamples. The cluster unit is the participant, and all eligible meals from each resampled participant are retained. The main-comparison and personalization intervals are paired within the same participants and meals. Cross-dataset intervals are computed for each source-trained model and target participant population, and the reported Holm correction is applied within each metric across the six source–target transfers. Differences are reported as candidate minus reference, so negative values favor the candidate in the main and cross-dataset comparisons. For component ablations, differences are reported as the ablated population model minus the full population model, so positive values indicate degradation from removing the component.

C.6.1 Main Comparisons

Table 20: Participant-clustered paired bootstrap intervals for the main comparisons. For each dataset and metric, the reference is the strongest non-MealGluco baseline by mean error. Δ is MealGluco minus the reference, so negative values favor MealGluco. Intervals are paired 95% bootstrap intervals over participants, retaining all eligible meals per participant. Bold entries have an interval excluding zero.

Dataset	Metric	Strongest baseline	Δ	95% CI
CGMacros	*Trajectory-MAE*	Stacking Framework	**-2.31**	**[-3.13, -1.52]**
CGMacros	*iAUC-MAE*	Stacking Framework	**-247.72**	**[-347.54, -155.01]**
CGMacros	*Peak-MAE*	Stacking Framework	**-2.49**	**[-3.99, -1.12]**
CGMacros	*Near-Peak Distance*	UniCA-Gluformer	**-1.99**	**[-3.75, -0.26]**
CGMacros	*Recovery-MAE*	Stacking Framework	-0.72	[-1.75, 0.26]
BIG IDEAs	*Trajectory-MAE*	Stacking Framework	+0.07	[-1.02, 1.06]
BIG IDEAs	*iAUC-MAE*	Stacking Framework	-45.52	[-131.29, 29.19]
BIG IDEAs	*Peak-MAE*	Stacking Framework	+0.22	[-0.79, 1.20]
BIG IDEAs	*Near-Peak Distance*	Gluformer (CGM-only)	+0.09	[-2.63, 3.12]
BIG IDEAs	*Recovery-MAE*	Stacking Framework	**+1.43**	**[0.21, 2.65]**
T1D-UOM	*Trajectory-MAE*	CGM-JEPA (CGM-only)	**-1.12**	**[-1.47, -0.73]**
T1D-UOM	*iAUC-MAE*	CGM-JEPA (CGM-only)	**-195.55**	**[-284.41, -95.16]**
T1D-UOM	*Peak-MAE*	Stacking Framework	-1.00	[-3.34, 1.44]
T1D-UOM	*Near-Peak Distance*	Stacking Framework	+0.42	[-2.39, 2.84]
T1D-UOM	*Recovery-MAE*	Stacking Framework	-0.35	[-1.65, 1.08]

C.6.2 Online Personalization

Table 21: Participant-clustered paired bootstrap intervals for online personalization. The reference is the population prediction before any historical meal is assimilated. Δ is online minus population, so negative values favor online personalization. Bold entries have an interval excluding zero.

Dataset	Metric	Δ	95% CI
CGMacros	*Trajectory-MAE*	**-1.47**	**[-2.25, -0.81]**
CGMacros	*iAUC-MAE*	**-171.40**	**[-276.42, -78.69]**
CGMacros	*Peak-MAE*	**-1.96**	**[-3.59, -0.63]**
CGMacros	*Near-Peak Distance*	**-2.54**	**[-3.78, -1.39]**
CGMacros	*Recovery-MAE*	**-1.30**	**[-2.15, -0.53]**
BIG IDEAs	*Trajectory-MAE*	-0.31	[-0.91, 0.09]
BIG IDEAs	*iAUC-MAE*	**-55.59**	**[-121.91, -8.73]**
BIG IDEAs	*Peak-MAE*	-0.27	[-1.12, 0.53]
BIG IDEAs	*Near-Peak Distance*	+1.43	[-0.10, 3.23]
BIG IDEAs	*Recovery-MAE*	**+0.64**	**[0.09, 1.13]**
T1D-UOM	*Trajectory-MAE*	**-0.93**	**[-1.62, -0.09]**
T1D-UOM	*iAUC-MAE*	**-106.25**	**[-190.20, -8.59]**
T1D-UOM	*Peak-MAE*	-1.28	[-3.23, 0.40]
T1D-UOM	*Near-Peak Distance*	-0.99	[-2.15, 0.30]
T1D-UOM	*Recovery-MAE*	**-0.48**	**[-1.10, -0.04]**

Table 22: Participant-clustered paired bootstrap intervals for two-coordinate online personalization versus participant-specific fine-tuning. Δ is fine-tuning minus the online state, so positive values favor the online state and negative values favor fine-tuning. Bold entries have an interval excluding zero.

Dataset	Online	Fine-tuning	Δ	95% CI
CGMacros	21.16	21.96	**+0.79**	**[0.13, 1.62]**
BIG IDEAs	16.56	16.46	-0.10	[-0.40, 0.31]
T1D-UOM	30.08	29.41	**-0.66**	**[-1.06, -0.19]**

C.6.3 Cross-Dataset Generalization

Table 23: Participant-clustered paired bootstrap intervals for cross-dataset *Trajectory-MAE*. The reference is the stronger of UniCA-Gluformer and the Stacking Framework for each transfer. Δ is MealGluco minus the reference, so negative values favor MealGluco. The Holm column adjusts the six bootstrap p-values within this metric. Bold entries have an interval excluding zero.

Transfer	Δ	95% CI	p	Holm p
CGMacros $\rightarrow$ BIG IDEAs	**-1.72**	**[-2.98, -0.43]**	0.0052	0.0208
CGMacros $\rightarrow$ T1D-UOM	-0.69	[-1.71, 0.34]	0.1804	0.3608
BIG IDEAs $\rightarrow$ CGMacros	-0.64	[-1.77, 0.59]	0.2954	0.3608
BIG IDEAs $\rightarrow$ T1D-UOM	**-4.71**	**[-7.07, -2.93]**	< 0.0001	< 0.0001
T1D-UOM $\rightarrow$ CGMacros	**-2.09**	**[-3.01, -1.23]**	< 0.0001	< 0.0001
T1D-UOM $\rightarrow$ BIG IDEAs	**-1.51**	**[-3.06, -0.35]**	0.0068	0.0208

Table 24: Participant-clustered paired bootstrap intervals for cross-dataset *iAUC-MAE*. The reference is the stronger of UniCA-Gluformer and the Stacking Framework for each transfer. Δ is MealGluco minus the reference, so negative values favor MealGluco. The Holm column adjusts the six bootstrap p-values within this metric. Bold entries have an interval excluding zero.

Transfer	Δ	95% CI	p	Holm p
CGMacros → BIG IDEAs	**-189.39**	**[-332.59, -50.52]**	0.0034	0.0170
CGMacros → T1D-UOM	-46.69	[-171.76, 41.14]	0.3448	1.0000
BIG IDEAs → CGMacros	+15.59	[-108.87, 149.75]	0.8250	1.0000
BIG IDEAs → T1D-UOM	-29.16	[-188.22, 77.38]	0.6662	1.0000
T1D-UOM → CGMacros	**-240.23**	**[-350.41, -138.27]**	< 0.0001	< 0.0001
T1D-UOM → BIG IDEAs	**-147.74**	**[-327.68, -2.94]**	0.0456	0.1824

Table 25: Participant-clustered paired bootstrap intervals for cross-dataset *Peak-MAE*. The reference is the stronger of UniCA-Gluformer and the Stacking Framework for each transfer. Δ is MealGluco minus the reference, so negative values favor MealGluco. The Holm column adjusts the six bootstrap p-values within this metric. Bold entries have an interval excluding zero.

Transfer	Δ	95% CI	p	Holm p
CGMacros → BIG IDEAs	-0.81	[-2.44, 0.80]	0.3468	1.0000
CGMacros → T1D-UOM	-1.44	[-6.78, 0.87]	0.3926	1.0000
BIG IDEAs → CGMacros	-0.54	[-2.41, 1.40]	0.5760	1.0000
BIG IDEAs → T1D-UOM	**-2.75**	**[-5.48, -1.13]**	< 0.0001	< 0.0001
T1D-UOM → CGMacros	**-3.92**	**[-5.32, -2.58]**	< 0.0001	< 0.0001
T1D-UOM → BIG IDEAs	**-3.78**	**[-4.72, -2.71]**	< 0.0001	< 0.0001

Table 26: Participant-clustered paired bootstrap intervals for cross-dataset *Near-Peak Distance*. The reference is the stronger of UniCA-Gluformer and the Stacking Framework for each transfer. Δ is MealGluco minus the reference, so negative values favor MealGluco. The Holm column adjusts the six bootstrap p-values within this metric. Bold entries have an interval excluding zero.

Transfer	Δ	95% CI	p	Holm p
CGMacros → BIG IDEAs	+1.28	[-0.42, 3.05]	0.1338	0.5352
CGMacros → T1D-UOM	**-2.10**	**[-4.61, -0.42]**	0.0024	0.0144
BIG IDEAs → CGMacros	**+2.16**	**[0.38, 4.15]**	0.0162	0.0810
BIG IDEAs → T1D-UOM	+0.05	[-2.57, 2.94]	0.9876	1.0000
T1D-UOM → CGMacros	+0.15	[-1.27, 1.63]	0.8658	1.0000
T1D-UOM → BIG IDEAs	+1.56	[-0.92, 3.83]	0.2214	0.6642

Table 27: Participant-clustered paired bootstrap intervals for cross-dataset *Recovery-MAE*. The reference is the stronger of UniCA-Gluformer and the Stacking Framework for each transfer. Δ is MealGluco minus the reference, so negative values favor MealGluco. The Holm column adjusts the six bootstrap p-values within this metric. Bold entries have an interval excluding zero.

Transfer	Δ	95% CI	p	Holm p
CGMacros → BIG IDEAs	**+0.81**	**[0.06, 1.54]**	0.0342	0.1710
CGMacros → T1D-UOM	+0.45	[-1.13, 2.38]	0.6222	0.7004
BIG IDEAs → CGMacros	-0.82	[-1.76, 0.13]	0.0906	0.3624
BIG IDEAs → T1D-UOM	-1.66	[-3.35, 0.77]	0.1532	0.4596
T1D-UOM → CGMacros	-0.37	[-1.20, 0.39]	0.3502	0.7004
T1D-UOM → BIG IDEAs	**+1.12**	**[0.44, 1.78]**	0.0018	0.0108

C.6.4 COMPONENT ABLATIONS

Table 28: Participant-clustered paired bootstrap intervals for the component ablations on *Trajectory-MAE*. Δ is the ablated population model minus the full population model, so positive values indicate degradation from removing the component. Bold entries have an interval excluding zero.

Dataset	Ablation	Δ	95% CI
CGMacros	MAE point loss	**+1.13**	**[0.74, 1.52]**
CGMacros	Raw polynomial post-meal time	**+0.44**	**[0.22, 0.70]**
CGMacros	Unstructured meal-response head	**+0.43**	**[0.14, 0.77]**
CGMacros	w/o clock time	**+0.69**	**[0.39, 1.05]**
CGMacros	w/o frozen CGM-JEPA tokens	**+0.78**	**[0.49, 1.09]**
CGMacros	w/o meal response	**+2.42**	**[1.19, 3.76]**
CGMacros	w/o nutrient modifiers	**+0.36**	**[0.08, 0.74]**
CGMacros	w/o post-meal time	**+0.63**	**[0.35, 0.91]**
CGMacros	w/o short-horizon statistics	**+0.83**	**[0.42, 1.25]**
CGMacros	w/o state response	**+0.73**	**[0.33, 1.16]**
BIG IDEAs	MAE point loss	+0.16	[-0.12, 0.39]
BIG IDEAs	Raw polynomial post-meal time	**+0.23**	**[0.15, 0.33]**
BIG IDEAs	Unstructured meal-response head	-0.05	[-0.31, 0.18]
BIG IDEAs	w/o clock time	-0.00	[-0.32, 0.30]
BIG IDEAs	w/o frozen CGM-JEPA tokens	-0.20	[-0.58, 0.13]
BIG IDEAs	w/o meal response	**+1.45**	**[0.35, 2.39]**
BIG IDEAs	w/o nutrient modifiers	-0.05	[-0.23, 0.11]
BIG IDEAs	w/o post-meal time	+0.11	[-0.00, 0.23]
BIG IDEAs	w/o short-horizon statistics	+0.29	[-0.25, 0.75]
BIG IDEAs	w/o state response	**+0.50**	**[0.19, 0.86]**
T1D-UOM	MAE point loss	+0.37	[-0.06, 1.15]
T1D-UOM	Raw polynomial post-meal time	-0.19	[-0.43, 0.16]
T1D-UOM	Unstructured meal-response head	-0.05	[-0.43, 0.44]
T1D-UOM	w/o clock time	+0.04	[-0.12, 0.28]
T1D-UOM	w/o frozen CGM-JEPA tokens	-0.03	[-0.70, 0.81]
T1D-UOM	w/o meal response	+0.13	[-0.78, 1.40]
T1D-UOM	w/o nutrient modifiers	+0.49	[-0.06, 1.56]
T1D-UOM	w/o post-meal time	+0.09	[-0.30, 0.64]
T1D-UOM	w/o short-horizon statistics	+0.17	[-1.45, 2.35]
T1D-UOM	w/o state response	+0.14	[-0.37, 0.90]

Table 29: Participant-clustered paired bootstrap intervals for the component ablations on *iAUC-MAE*. Δ is the ablated population model minus the full population model, so positive values indicate degradation from removing the component. Bold entries have an interval excluding zero.

Dataset	Ablation	Δ	95% CI
CGMacros	MAE point loss	**+81.56**	**[30.39, 135.27]**
CGMacros	Raw polynomial post-meal time	**+35.03**	**[6.93, 66.75]**
CGMacros	Unstructured meal-response head	+38.71	[-0.68, 83.76]
CGMacros	w/o clock time	**+78.43**	**[32.84, 129.29]**
CGMacros	w/o frozen CGM-JEPA tokens	**+87.20**	**[48.76, 124.09]**
CGMacros	w/o meal response	**+300.62**	**[136.30, 472.79]**
CGMacros	w/o nutrient modifiers	**+37.31**	**[1.03, 82.00]**
CGMacros	w/o post-meal time	**+47.85**	**[10.60, 84.77]**
CGMacros	w/o short-horizon statistics	**+67.17**	**[10.49, 127.15]**
CGMacros	w/o state response	**+66.41**	**[11.14, 127.65]**
BIG IDEAs	MAE point loss	**-35.54**	**[-80.78, -0.04]**
BIG IDEAs	Raw polynomial post-meal time	**+18.05**	**[5.36, 32.92]**
BIG IDEAs	Unstructured meal-response head	-12.27	[-34.86, 8.59]
BIG IDEAs	w/o clock time	-7.75	[-44.76, 21.47]
BIG IDEAs	w/o frozen CGM-JEPA tokens	-19.99	[-66.44, 17.33]
BIG IDEAs	w/o meal response	**+179.16**	**[46.43, 295.91]**
BIG IDEAs	w/o nutrient modifiers	**-31.06**	**[-50.20, -12.56]**
BIG IDEAs	w/o post-meal time	-4.49	[-23.05, 12.66]
BIG IDEAs	w/o short-horizon statistics	-14.72	[-84.29, 42.43]
BIG IDEAs	w/o state response	-2.70	[-23.97, 21.61]
T1D-UOM	MAE point loss	-28.66	[-80.55, 31.28]
T1D-UOM	Raw polynomial post-meal time	-20.84	[-43.11, 7.75]
T1D-UOM	Unstructured meal-response head	-11.69	[-40.87, 24.31]
T1D-UOM	w/o clock time	+8.39	[-9.05, 32.91]
T1D-UOM	w/o frozen CGM-JEPA tokens	-5.38	[-74.44, 83.48]
T1D-UOM	w/o meal response	+37.15	[-91.52, 221.66]
T1D-UOM	w/o nutrient modifiers	**+71.01**	**[3.15, 194.04]**
T1D-UOM	w/o post-meal time	-18.60	[-48.26, 17.04]
T1D-UOM	w/o short-horizon statistics	-10.81	[-150.29, 173.24]
T1D-UOM	w/o state response	+7.66	[-19.82, 49.84]

Table 30: Participant-clustered paired bootstrap intervals for the component ablations on *Peak-MAE*. Δ is the ablated population model minus the full population model, so positive values indicate degradation from removing the component. Bold entries have an interval excluding zero.

Dataset	Ablation	Δ	95% CI
CGMacros	MAE point loss	-0.33	[-0.99, 0.35]
CGMacros	Raw polynomial post-meal time	**+0.63**	**[0.27, 1.06]**
CGMacros	Unstructured meal-response head	-0.06	[-0.53, 0.41]
CGMacros	w/o clock time	**+0.74**	**[0.24, 1.30]**
CGMacros	w/o frozen CGM-JEPA tokens	**+0.83**	**[0.45, 1.19]**
CGMacros	w/o meal response	**+9.40**	**[7.34, 11.58]**
CGMacros	w/o nutrient modifiers	+0.44	[-0.03, 1.01]
CGMacros	w/o post-meal time	**+1.06**	**[0.53, 1.61]**
CGMacros	w/o short-horizon statistics	**+0.77**	**[0.24, 1.36]**
CGMacros	w/o state response	**+1.65**	**[1.00, 2.37]**
BIG IDEAs	MAE point loss	**-1.19**	**[-2.05, -0.44]**
BIG IDEAs	Raw polynomial post-meal time	+0.24	[-0.04, 0.47]
BIG IDEAs	Unstructured meal-response head	**-0.74**	**[-1.26, -0.25]**
BIG IDEAs	w/o clock time	+0.01	[-0.60, 0.51]
BIG IDEAs	w/o frozen CGM-JEPA tokens	-0.56	[-1.36, 0.08]
BIG IDEAs	w/o meal response	**+3.91**	**[2.34, 5.41]**
BIG IDEAs	w/o nutrient modifiers	**-0.32**	**[-0.59, -0.08]**
BIG IDEAs	w/o post-meal time	+0.23	[-0.11, 0.55]
BIG IDEAs	w/o short-horizon statistics	-0.12	[-0.78, 0.62]
BIG IDEAs	w/o state response	-0.08	[-0.35, 0.20]
T1D-UOM	MAE point loss	**-0.97**	**[-2.04, -0.29]**
T1D-UOM	Raw polynomial post-meal time	+0.30	[-0.16, 0.87]
T1D-UOM	Unstructured meal-response head	**-0.87**	**[-2.14, -0.19]**
T1D-UOM	w/o clock time	+0.23	[-0.05, 0.45]
T1D-UOM	w/o frozen CGM-JEPA tokens	-0.49	[-1.67, 0.95]
T1D-UOM	w/o meal response	**+3.51**	**[0.90, 6.74]**
T1D-UOM	w/o nutrient modifiers	+0.46	[-0.28, 1.41]
T1D-UOM	w/o post-meal time	+0.06	[-1.17, 1.50]
T1D-UOM	w/o short-horizon statistics	-0.11	[-1.37, 1.53]
T1D-UOM	w/o state response	+0.84	[-0.12, 1.96]

Table 31: Participant-clustered paired bootstrap intervals for the component ablations on *Near-Peak Distance*. Δ is the ablated population model minus the full population model, so positive values indicate degradation from removing the component. Bold entries have an interval excluding zero.

Dataset	Ablation	Δ	95% CI
CGMacros	MAE point loss	**-2.38**	**[-3.35, -1.41]**
CGMacros	Raw polynomial post-meal time	**-0.79**	**[-1.40, -0.18]**
CGMacros	Unstructured meal-response head	+0.19	[-1.16, 1.67]
CGMacros	w/o clock time	-0.50	[-1.43, 0.41]
CGMacros	w/o frozen CGM-JEPA tokens	-0.17	[-0.97, 0.64]
CGMacros	w/o meal response	**+6.91**	**[4.24, 9.60]**
CGMacros	w/o nutrient modifiers	**-0.90**	**[-1.62, -0.14]**
CGMacros	w/o post-meal time	**-3.35**	**[-4.84, -1.83]**
CGMacros	w/o short-horizon statistics	-0.16	[-0.87, 0.46]
CGMacros	w/o state response	**-2.29**	**[-4.02, -0.40]**
BIG IDEAs	MAE point loss	-0.85	[-2.26, 0.16]
BIG IDEAs	Raw polynomial post-meal time	-0.33	[-0.97, 0.28]
BIG IDEAs	Unstructured meal-response head	+0.62	[-0.13, 1.47]
BIG IDEAs	w/o clock time	-0.05	[-0.76, 0.79]
BIG IDEAs	w/o frozen CGM-JEPA tokens	+1.00	[-0.13, 2.29]
BIG IDEAs	w/o meal response	+0.33	[-0.87, 1.64]
BIG IDEAs	w/o nutrient modifiers	+0.44	[-0.01, 0.93]
BIG IDEAs	w/o post-meal time	-0.04	[-0.67, 0.62]
BIG IDEAs	w/o short-horizon statistics	-0.36	[-1.72, 0.95]
BIG IDEAs	w/o state response	-0.52	[-1.59, 0.50]
T1D-UOM	MAE point loss	-0.93	[-3.80, 3.32]
T1D-UOM	Raw polynomial post-meal time	-0.18	[-1.25, 0.80]
T1D-UOM	Unstructured meal-response head	+1.19	[-1.57, 5.35]
T1D-UOM	w/o clock time	**+0.81**	**[0.31, 1.54]**
T1D-UOM	w/o frozen CGM-JEPA tokens	+0.89	[-0.94, 2.72]
T1D-UOM	w/o meal response	+1.11	[-0.80, 3.17]
T1D-UOM	w/o nutrient modifiers	-0.79	[-2.08, 0.04]
T1D-UOM	w/o post-meal time	-1.02	[-2.87, 1.49]
T1D-UOM	w/o short-horizon statistics	+0.84	[-2.06, 3.30]
T1D-UOM	w/o state response	-1.61	[-4.01, 1.75]

Table 32: Participant-clustered paired bootstrap intervals for the component ablations on *Recovery-MAE*. Δ is the ablated population model minus the full population model, so positive values indicate degradation from removing the component. Bold entries have an interval excluding zero.

Dataset	Ablation	Δ	95% CI
CGMacros	MAE point loss	**-1.18**	**[-1.84, -0.51]**
CGMacros	Raw polynomial post-meal time	+0.18	[-0.14, 0.53]
CGMacros	Unstructured meal-response head	**-0.67**	**[-1.29, -0.06]**
CGMacros	w/o clock time	-0.10	[-0.43, 0.23]
CGMacros	w/o frozen CGM-JEPA tokens	-0.36	[-0.71, 0.01]
CGMacros	w/o meal response	**+3.80**	**[2.89, 4.71]**
CGMacros	w/o nutrient modifiers	**-0.33**	**[-0.59, -0.04]**
CGMacros	w/o post-meal time	**-1.85**	**[-2.52, -1.18]**
CGMacros	w/o short-horizon statistics	-0.03	[-0.39, 0.34]
CGMacros	w/o state response	**-1.62**	**[-2.35, -0.89]**
BIG IDEAs	MAE point loss	**-1.11**	**[-1.94, -0.40]**
BIG IDEAs	Raw polynomial post-meal time	**-0.78**	**[-1.38, -0.26]**
BIG IDEAs	Unstructured meal-response head	**-1.21**	**[-1.95, -0.57]**
BIG IDEAs	w/o clock time	+0.34	[-0.08, 0.69]
BIG IDEAs	w/o frozen CGM-JEPA tokens	**+0.33**	**[0.04, 0.60]**
BIG IDEAs	w/o meal response	**+1.41**	**[0.54, 2.24]**
BIG IDEAs	w/o nutrient modifiers	+0.12	[-0.46, 0.62]
BIG IDEAs	w/o post-meal time	**-0.54**	**[-1.12, -0.05]**
BIG IDEAs	w/o short-horizon statistics	+0.46	[-0.23, 1.17]
BIG IDEAs	w/o state response	+0.38	[-0.38, 1.10]
T1D-UOM	MAE point loss	+1.06	[-0.48, 3.20]
T1D-UOM	Raw polynomial post-meal time	+0.09	[-0.29, 0.53]
T1D-UOM	Unstructured meal-response head	+0.60	[-1.33, 2.03]
T1D-UOM	w/o clock time	**+0.38**	**[0.14, 0.75]**
T1D-UOM	w/o frozen CGM-JEPA tokens	-0.18	[-0.59, 0.20]
T1D-UOM	w/o meal response	**+0.52**	**[0.12, 1.13]**
T1D-UOM	w/o nutrient modifiers	-0.07	[-0.28, 0.11]
T1D-UOM	w/o post-meal time	-0.03	[-0.69, 0.80]
T1D-UOM	w/o short-horizon statistics	**+1.10**	**[0.39, 2.07]**
T1D-UOM	w/o state response	+0.05	[-0.35, 0.68]